

# Towards a more Systematic Analysis of Twitter Data: A Framework for the Analysis of Twitter Communities <sup>\*</sup>

Alessia Antelmi<sup>1,3</sup>, Josephine Griffith<sup>2</sup>, and Karen Young<sup>2,3</sup>

<sup>1</sup> Università degli Studi di Salerno, Fisciano, Italy  
aless.antelmi@gmail.com

<sup>2</sup> National University of Ireland Galway, Galway, Ireland  
{josephine.griffith, karen.young}@nuigalway.ie

<sup>3</sup> Insight Centre for Data Analytics, Data Science Institute, Galway, Ireland

**Abstract.** To date, many studies have used the social media platform Twitter to gather insights into real-life events. The current literature focuses on patterns around isolated case studies and their dynamics happening on the platform, but it still lacks standard techniques for comparing behavioural and interaction patterns within and across Twitter communities. To fill this gap, we present a framework for characterizing online Twitter communities from a quantitative and a semantic point of view. We then discuss an example of the application of the framework to compare two distinct Twitter fan communities. This case study application clearly illustrates the benefits of the framework, while also highlighting potential areas for improvement and further extensions.

**Keywords:** Online Social Network Analysis · Twitter Communities Analysis.

## 1 Introduction

The ever-increasing use of the Internet produces a huge amount of structured and unstructured data that can be mined and analysed to gather insights into several domains. In this context, online social networks represent a rich opportunity to collect real user data, especially from Twitter<sup>4</sup> which is well-suited to the task of discovering opinions, ideas and events [2]. With 335 million monthly active users as reported by the *Statista* website, the microblogging platform Twitter has been widely studied in contexts of political, crisis and brand communication and user engagement around shared experiences such as TV shows and everyday interpersonal exchanges [4]. Bruns and Stieglitz [4] proposed a catalogue of standard, replicable metrics for studying hashtagged Twitter conversations motivated by the absence in previous work of such metrics to compare

---

<sup>\*</sup> This paper has been funded in part by Science Foundation Ireland under grant number SFI/12/RC/2289 (Insight).

<sup>4</sup> <https://twitter.com>

one hashtagged event with another. However, the literature still lacks standard techniques for comparing behavioural and interaction patterns within and across Twitter communities. This prevents researchers from developing a comprehensive perspective about how Twitter is used by brands to engage with fans and critics and how this use changes over time. In this work, we present a framework for characterizing online Twitter communities from a quantitative and a semantic point of view, using data retrieved from both the profile and the timeline of the users. In Section 2 we describe in detail the proposed framework, while also outlining related work. In Section 3 we introduce two use cases showing how the framework can be applied to analyse and compare users' behaviour within and across Twitter communities. In experiments, we ensure that the collected data is cleaned so that any spam/bot content is removed prior to analysis [5, 17]. Section 4 discusses the results obtained and ideas for future work.

## 2 A Framework for Analysing Twitter Communities

In this work, we consider a *Twitter community* as a set of Twitter users who share a common interest (e.g. some followers of a TV series' Twitter account), motivated by the research of Java et al. [10]. Our framework is made up of two principal components to deal with the User Generated Content (UGC) - in terms of topics, sentiment and emotions expressed - and the user's interaction behaviours and posting patterns, as shown in Figure 1. We will describe the *semantic* component in Section 2.1, and the *quantitative* component in Section 2.2.

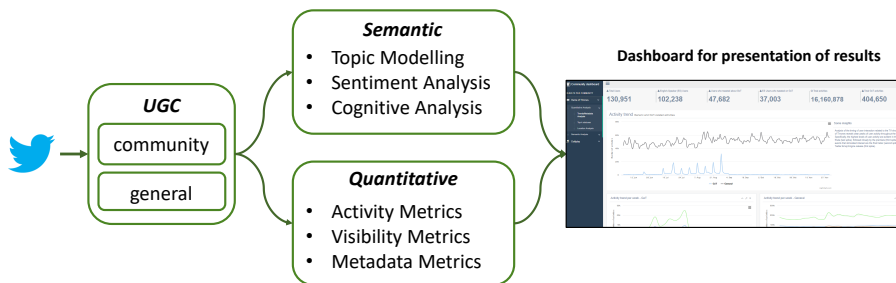


Fig. 1: Proposed framework and its main components

### 2.1 Semantic Analysis

Semantic analysis enables insights into the content produced by the community of interest. In our framework we propose a three-level semantic analysis approach, exploring the topics discussed, the sentiment and the cognitive sphere of the posts. Where the given community is selected according to a specific interest/topic, it can be useful to split the UGC into two subsets: (i) the first containing all the activities related to the interest/topic chosen and (ii) the

remaining ones. Splitting the dataset in this way enables the comparison of behavioural patterns across the same set of users regarding the topic of interest and the other remaining activities. The analyses described can then be run independently on both subsets indicating differences and similarities that exist within a community in comparison to general discussions. This division can be done using a **keyword list** based on the chosen topic.

**Topic Modelling Level.** Topic modelling is a machine learning technique that looks for patterns in the use of words and attempts to inject semantic meaning into vocabulary, whereby a *topic* consists of a cluster of words that frequently occur together [6]. Previous work [1, 9] presents a survey of tools and approaches for topic detection from Twitter streams, exploring different types of topic detection techniques and evaluating their performance. Lau et al. [12] and Jonsson et al. [11] focus on the evaluation of the Latent Dirichlet Allocation (LDA) topic modelling algorithm and its variants, while Musto et al [14] implement a pipeline of entity linking algorithms. *Entity linking algorithms* automatically incorporate stopword removal, bigram recognition, entity identification and disambiguation. They can also enrich the representation with features which do not explicitly occur in a text: for example, if an entity is mapped to a Wikipedia page, it is possible to browse a Wikipedia category’ tree to further enrich content representation introducing the most relevant ancestor categories of that page. Discovering the topics discussed in the UGC is the first step in detecting the interests of a community.

**Sentiment Analysis Level.** The study of the tweets’ polarity (examination of the sentiment of the tweets) can give important insights into what is happening in the real world and what people think about a given event. Bollen et al. [3] found that events in the social, political, cultural and economic sphere do have a significant, immediate and highly specific effect on the various dimensions of public mood, suggesting that large-scale analyses of mood can provide a solid platform to model collective emotive trends in terms of their predictive value with regards to existing social as well as economic indicators. Martinez et al. offer a survey [13] of the state of the art techniques used to explore sentiment analysis on Twitter. It is worthwhile highlighting that due to the nature of the tweets, i.e rich in emojis, some studies focus on extracting the sentiment using this piece of information [20, 19].

**Cognitive Analysis Level.** Exploiting the cognitive sphere of the UGC gives deeper knowledge about the emotional aspects of the content and the personality of its author [15]. Several works explore this dimension on the Twitter platform. Qiu et al. [15] study the relationship between personality and the microblog, pointing out the potential of using social media for personality research. Tumasjan et al. [16] use cognitive analysis to investigate whether Twitter is used as a forum for political deliberation and whether online messages on Twitter validly mirror offline political sentiment. The synergistic use of Twitter and the

analysis of the cognitive sphere can also help in the health domain, where simple natural language processing can yield insights into specific disorders [7] and into the level of stress during workdays and weekends [18]. Linguistic Inquiry and Word Count<sup>5</sup> (*LIWC*) - text analysis software developed to assess emotional, cognitive and structural components of text samples using a psychometrically validated internal dictionary - is one of the most used tools in cognitive analysis, thanks to its ease of use and its broad range of social and psychological insights. Other tools are *ANEW*<sup>6</sup>, a dictionary focused on academic text, and *GI*<sup>7</sup>, a computer-assisted approach for content analyses of textual data.

## 2.2 Quantitative Analysis

While the semantic analysis provides a way to analyse the content posted by the users, *quantitative analysis* of user tweets provides insights into user behaviours and interaction patterns. We identify three typologies of quantitative metrics: *activity*, *visibility* and *metadata*.

**Activity metrics.** Activity metrics describe the daily activity pattern of the community - in terms of the content posted or liked - and the number of different types of activities, i.e. the total amount of tweets, quotes, retweets, comments and likes. Evaluating the number of daily activities is useful in identifying any spike in the interaction pattern and its potential reason (e.g. political election, movie premiere). Assessing the number of different activities that exist can help in finding out the proportional distribution between *information providing* and *information seeking* users. This information can be helpful when evaluating information propagation strategies.

**Visibility metrics.** Visibility metrics count the number of retweets and likes received and they can help in understanding the visibility of the users within the community and Twitter in general.

**Metadata metrics.** Metadata metrics are evaluated on the metadata field retrieved from the Twitter JSON object describing a user's activity. These metrics enable the identification of the most used *hashtags*, *posting devices* and attached *media* (in terms of photos, videos and gifs) as well as the location of the activity posted.

## 2.3 Summary of the Framework

We have outlined the organization of our framework, designed to provide a structured overview of approaches and techniques that best suit the analysis of the community of interest. The framework serves to **guide** the choice of these, exploring some of the most common tools and techniques used and describing whether they can be useful or not according to the insights they offer about the

---

<sup>5</sup> <http://liwc.wpengine.com>

<sup>6</sup> <http://www.newacademicwordlist.org>

<sup>7</sup> <http://www.wjh.harvard.edu/inquirer>

data (e.g., understanding the cause of a peak in the user activity and the associated reaction). The high modularity that distinguishes the framework allows the use of some, or all, of its components based on the desired depth of analysis required. A visualization component, where we outline several techniques to plot the outcomes obtained from the analyses, is also included in the framework. We do not discuss this component in this work, but we provide a link<sup>8</sup> to a dashboard visualizing all results from the use cases described. To understand how the framework can be used to answer specific research questions we tested it by applying it to two real Twitter communities. These two use cases are described in the following section.

### 3 Use Cases

The application of the proposed framework to Twitter communities is trialed in two separate use cases – individually initially, followed by a comparative analysis across the two communities. While both communities were analysed using all six components of the framework, we will present only the most interesting outcomes here. It is worth noting that there are many tools available for each task described; we chose the ones that have been widely used in the literature and that are, at the same time, both easy to obtain and do not require a significant coding effort - thus making the framework more accessible to more people. Furthermore, thanks to the modular design of the framework, different tools can be added, replaced and removed as new technologies are developed.

We consider two Twitter fan communities - one related to the TV show Game of Thrones (GoT), the other to the British rock band Coldplay - and we apply the same methodology to both of them. We chose the GoT community due to the fact that the Game of Thrones TV show has established a reputation for being widely discussed on social media channels, specifically on the Twitter platform, which allows dynamic, real-time engagement and interaction with viewers. We picked a Coldplay community for a similar reason: with a gross of 523 million dollars and 5.39 million fans attending their tour in 2017, the pop/rock band Coldplay is one of the most famous in the last decade. In both cases, Twitter plays the role of an important medium to engage fans. We retrieved the complete followers list of the GoT Twitter account and then randomly picked 350,000 users from among them. We collected data - timeline and profile information - from the 130,951 users among this random group whose timeline was publicly shared over a 6-months timeframe from June 3<sup>rd</sup> to December 3<sup>rd</sup>, 2017. We followed the same process to select a random subset of the Coldplay official Twitter account followers obtaining a final 121,306 users. We first divided both datasets into two different subsets, one containing all posts about either a GoT or Coldplay topic, the other two consisting of all the remaining activities within a dataset. To accomplish this task we used a **keyword list** and we searched for these keywords in the *text*, *hashtags* and *user mentions* status fields. The GoT keyword list was based on GoT world in general, HBO.com (e.g. promotional

---

<sup>8</sup> <http://www.alessiaantelmi.it/framework/production>

campaigns, episodes' titles) and GoT books (e.g. titles, characters), while the Coldplay keyword list was based on the band members' names, Coldplay songs and albums' titles and tours' names. This process yielded a final total of 404,650 GoT user activities (with 47,682 users who posted an update about the TV show) and 16,160,878 generic GoT-unrelated posts within the GoT dataset, and 6,814 Coldplay user activities (with 2,103 users who posted an update about the band) and 4,270,077 generic posts within the Coldplay dataset. It is interesting to note that there is a considerable imbalance in the number of activities posted by the two communities, probably due to the different nature of the community-related events analysed (TV series and concerts). All the analyses that follow have been run independently on all four sub datasets - the semantic analysis approach has only been applied to the English tweets.

### 3.1 Semantic Analysis

**Topic Analysis.** The topic modelling tool we chose was the machine learning toolkit MALLET<sup>9</sup>, which provides an efficient way to build up topic models based on the LDA algorithm. We found that 4 was the optimal number of clusters for the GoT-related activities, corresponding to the following topics: **broadcast of a new episode, season première/finale, trailers/scenes** (e.g., videos on YouTube) and **episodes' content** (e.g., lines spoken by GoT TV series characters). To evaluate the topics for the generic activities in the GoT dataset, we split them according to their creation date and we analysed the topics per month. Due to the huge amount of posts to evaluate, we randomly picked three different subsets (around 50,000 posts each) from among them on which to run the topic analysis. We found that the most discussed topics across the whole six months are: **politics** (e.g., Brexit, Trump, Obama, Catalonia), **sport** (e.g., cricket, NBA, tennis, football), **special events** (Father's Day, Thanksgiving) and **news** (e.g., Hurricanes Harvey and Irma, Mexico Earthquake, Las Vegas shooting). To verify if the use of an entity linking algorithm can improve the evaluation of the topics, we also used Tag.me APIs<sup>10</sup> that enabled us to identify Wikipedia entities referred to in the text of the content analysed. The investigation of the Wikipedia entities in the GoT-related dataset helped in sharpening the broader topics obtained through LDA, showing that they mainly refer to (i) **GoTstoryline characters**: Jon Snow was the most discussed, followed by Arya Stark, Daenerys Targaryen and Sansa Stark and (ii) **locations**: either described in the books or used as filming sets (identified under the Wikipedia entity World of A Song of Ice and Fire). The Wikipedia entities retrieved from the generic activities in the GoT dataset didn't add any further information; most common entities found are Father's Day, Theresa May, Israel cricket team, Manchester United F.C., racism, Houston, hurricane season, Thanksgiving, Donald Trump.

Analysis of the Coldplay dataset found the following topics in the Coldplay-related activities: **A Head Full of Dreams Tour, Houston concert and**

<sup>9</sup> <http://mallet.cs.umass.edu>.

<sup>10</sup> <http://tagme.di.unipi.it>

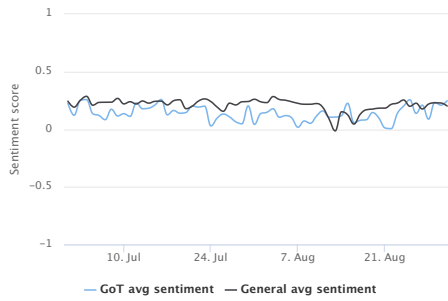
**hurricane, Album anniversary, Albums/songs advertising.** As in the GoT case, the Wikipedia entities collected gave us further information about these topics, such as most tweeted songs (A Sky Full of Stars and Fix You) and the location of the Houston concert (NRG Stadium). Frequently discussed topics in the generic activities within the Coldplay-dataset (evaluated as in the GoT case) are mostly **music-related** with references to the Teen Choice Awards and to the MTV Video Music Awards - the evaluation of the number of the Wikipedia entities found in this dataset highlights the same result. Other topics are related to **politics** (e.g., Obama, Trump, racism), **TV-series** (e.g., Game of Thrones, Gotham), **sport** (e.g., Premier League, NBA, cricket) and **special events/news** (e.g., Harry Potter, Father’s Day, Houston Hurricane, Diwali). It is interesting to note that both communities are engaged with almost the same set of generic topics, such as politics, sports, news and special events.

**Sentiment Analysis.** To evaluate the sentiment of the collected activities we used the freely available lexicon and rule-based classifier Vader [8], “specifically attuned to sentiments expressed in social media”<sup>11</sup>. We studied the sentiment expressed in the whole GoT dataset from the 1<sup>st</sup> July until the 31<sup>st</sup> August, when the majority of the GoT-related activities happened (see Figure 4a), while we observed the sentiment expressed in the Coldplay dataset from August 15<sup>th</sup> until August 31<sup>st</sup>, corresponding to a peak in the number of Coldplay-related activities (see Figure 2d). Figures 2a and 2c show the average daily sentiment within the GoT and the Coldplay dataset, respectively. This is prevalently positive for both communities with few points touching zero, meaning an increasing number of negative posts that lower the average sentiment value in those days. To find the events identified with these valleys in the sentiment, we combined the GoT-related activities and the associated sentiment in Figure 2b. We discovered that lower values in the sentiment are related to the third, fourth, fifth and seventh episodes of the GoT TV series (where an episode is represented by a spike in the activities). Lower sentiment values are also observable (Figure 2a) for the generic activities posted during the second week of August; they are mostly related to the nationalist march in Charlottesville and to the racism question in general. Figure 2d shows the same analysis for Coldplay-related activities, where a peak in the activities corresponds to the lowest daily sentiment; this event is related to the cancelled Coldplay concert in Houston because of the hurricane Harvey. This also explains the many tweets related to the NRG stadium, the location of this concert.

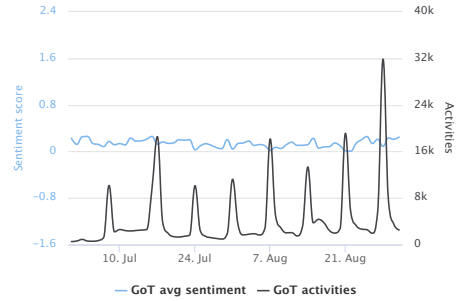
**Cognitive Analysis.** We used the LIWC dictionary as a cognitive analysis tool. The analysis identified (see Figures 3a and 3c) that both communities have a positive and confident style (a high value for the *Tone* and *Clout* variables, respectively), expressed with a distanced form of discourse (a low *Authentic* value). Figures 3b and 3d show the outcomes for the other LIWC dimensions analysed. The GoT-related activities are not only the more negative in terms of sadness,

---

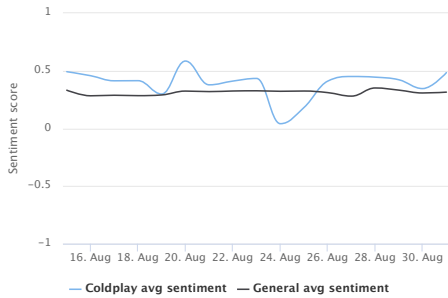
<sup>11</sup> <https://github.com/cjhutto/vaderSentiment>



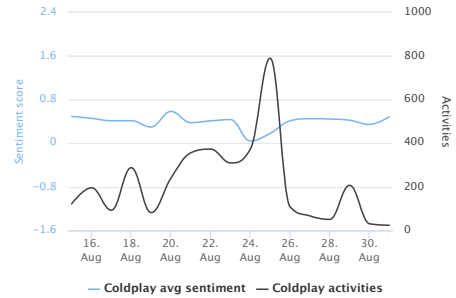
(a) Average daily sentiment from the 1<sup>st</sup> July until the 31<sup>st</sup> August (GoT dataset).



(b) Comparison between the GoT-related activities and the sentiment expressed.



(c) Average daily sentiment from the 15<sup>th</sup> August until the 31<sup>st</sup> August (Coldplay dataset).



(d) Comparison between the Coldplay-related activities and the sentiment expressed.

Fig. 2: Sentiment analysis results for both communities

anxiety and anger, but many of them also refer to status, dominance and social hierarchies (a high value for *Power* variable) and they include several references to other people (high value for *Affiliation* variable). Both these values are reflected in the *Drives* dimension. This result is not surprising due to the topics on which the GoT TV series is based (e.g., battles for power). Both communities are focused on present events, given the use of present-tense verbs; while the GoT one refers also to past events with past-tense verbs. Moderately-high values for the *Informal*, *Netspeak* and *Assent* variables reflect the writing style of social media, such as the use of basic punctuation-based emoticons and abbreviations like *LOL*. Examples of Assent words are: *agree*, *OK*, *yes*. The moderately-high value for the *CogProc* dimension highlights how users are willing to express their opinions in their tweets; this aspect is supported by the use of verbs like *think*, *consider* and *should*. It is also interesting to note the absence of *Perceptual processes*, i.e. the absence of the massive use of verbs indicating seeing, hearing and feeling, even though many activities within the GoT community refer to the TV show and many posts in the Coldplay community refer to music.

**Quantitative Analysis.** A study of the daily **activity pattern** of all datasets



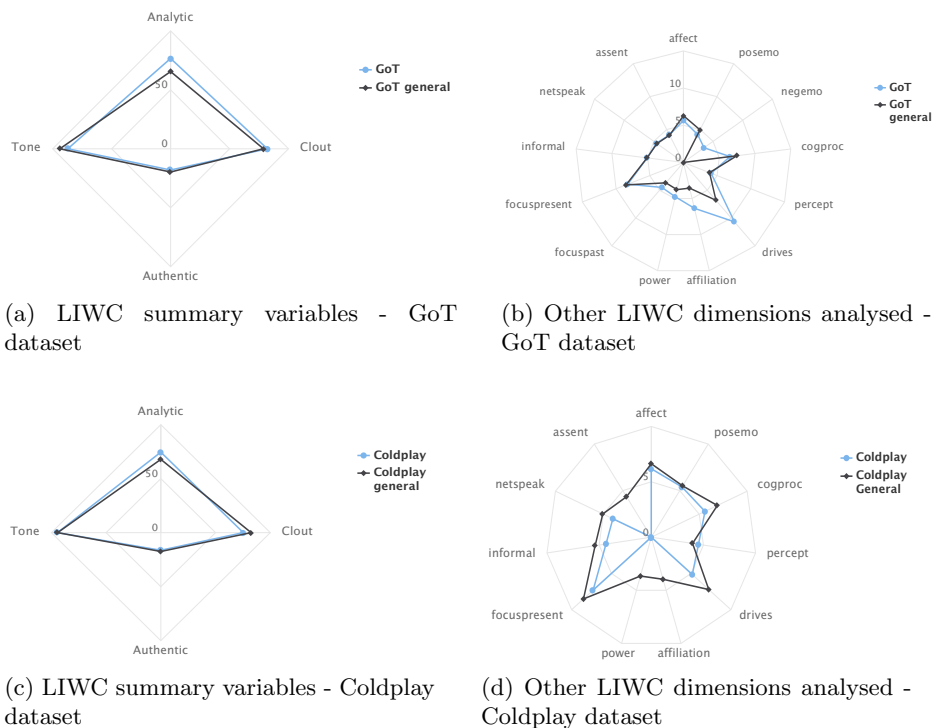
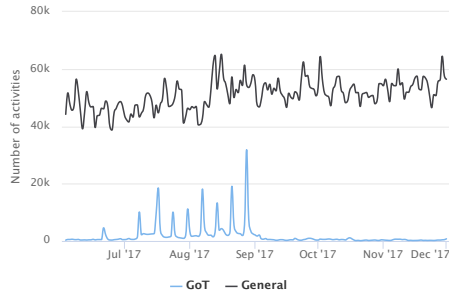


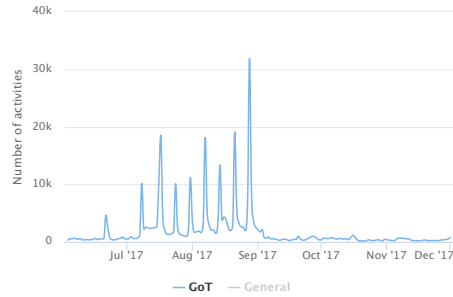
Fig. 3: Cognitive analysis results for the GoT community

reveals that the GoT community is far more active than the Coldplay one. In particular, Figures 4a and 4b evidence clear peaks of user activity related to the GoT show throughout the TV series' seventh season. Specifically, the highest levels of user activity are evident in the season finale (last spike), followed closely by the premiere (third spike). Other events that stimulated interest are the final trailer (second spike) and the Twitter Emoji Engine release (first spike). By contrast, no clear pattern emerges from the Generic activity set. The study of the daily activity pattern for the Coldplay community (Figures 4c and 4d) shows a huge peak of generic activities during August and the second half of the observed period. This is likely due to some events happening in the month of August, as a result of hurricane Harvey (and the following hurricane Irma) and the MTV Video Music Awards held in California. Figure 4d reveals a single major peak in the Coldplay-related activities happening on the 25<sup>th</sup> August corresponding to the cancelled Houston concert. The other highest levels of user activity refer to the concerts performed in Chicago (17<sup>th</sup> August), Cleveland (19<sup>th</sup> August) and Miami (28<sup>th</sup> August), while other minor peaks in October and November refer to concerts as well.

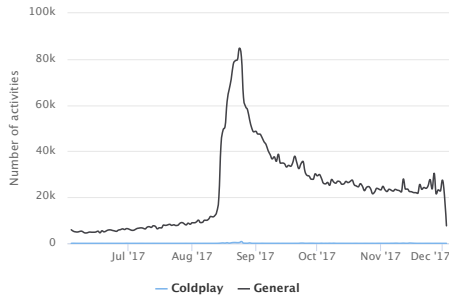
The final **metadata analysis** reveals that the most used *hashtags* related to GoT are those standard, generic hashtags like the name of the show: #game-ofthrones, and different abbreviated versions of it, like #got and #got7, followed



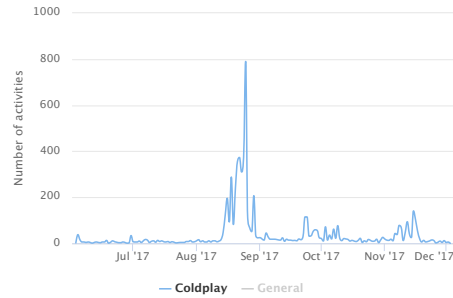
(a) Daily activity pattern for the GoT community



(b) Daily activity pattern for the GoT-related activities



(c) Daily activity pattern for the Coldplay community



(d) Daily activity pattern for the Coldplay-related activities

Fig. 4: Daily activity pattern for both communities

by #winterishere, #thronesyall, #gotmvp, #gameofthronesfinale and #prepareforwinter. The top five hashtags found in the generic activities within the GoT dataset are #giveaway, #win, #tvtime, #mufc (Manchester United F.C.) and #sdcc (San Diego Comic-Con). The most used hashtags related to the Coldplay activities are associated with the *A Head Full of Dreams Tour*, like #coldplay-toronto, #coldplaychicago and #coldplayhouston, in addition to generic ones such as #coldplay and #chrismartin. The top five hashtags for the generic activities are: #pushawardskathniels (it refers to the Push Awards 2017 contest which recognises top online influencers), #mersal, #missuniverse (Miss Universe Philippines), #philippines and #newprofilepic. Both communities share the clear predominance for mobile activity - as the majority of them are generated from an iPhone and from an Android device (more than 70% in total). The most common language is English (more than 60%), followed by Spanish, Portuguese and French in both cases.

### 3.2 Discussion

The proposed framework allows a larger problem, namely the analysis of behavioural and interaction patterns of a Twitter community, to be broken into

sub-problems so that some or all of the different components described in Section 2 can be considered when analysing data from a new community. Its application to two use cases illustrates how a standard approach in analysing communities makes it easier to acquire insights into them, especially when combining and/or comparing the several outcomes obtained. The quantitative analysis can offer an overview of the user behaviour, in terms of interaction patterns and typology of content posted and clearly illustrates the activity level of a community, while the evaluation of the most used hashtags provides insights into the most discussed events within a community, like the San Diego Comic-Con for the GoT community and Miss Universe Philippines for the Coldplay one. Merging the outcomes from the metadata analysis (such as tweeting locations, most used languages and posting devices) provides insights regarding any events happening in a community, as in the case, for example, in the Coldplay community, where most of the activities were from the USA during the band’s tour. Further insights can be obtained by merging the quantitative information with the outcomes from the semantic analysis: Figure 2 shows how combining the daily activity pattern with the sentiment and the topics expressed in the posts yields the most discussed topics and the reaction associated with them. The study of the cognitive dimension can further improve the analysis of the emotional component extending the binary positive/negative sentiment categorization to several categories, for instance in terms of anger, anxiety or happiness. This can be useful to compare the reaction to different events or the way communities express themselves; for example, we found that the GoT community was more negative in terms of anger and anxiety in comparison to the Coldplay dataset.

## 4 Conclusion

This work presents a framework for the analysis of User Generated Content (UGC) using Twitter communities and applies the framework to two different case studies. The framework comprises two main components - semantic and quantitative - with each component comprising three sub-components. The development of a standard framework for the analysis of Twitter communities provides a simplified approach to compare and correlate outcomes across a range of different case studies. This can be used to find the similarities and differences in behavioural and interaction patterns within and across communities. The presence of a dashboard to interactively visualize the results from the analyses and the user insights produced can be another useful tool to acquire knowledge about the dataset and will be discussed in future work. To further investigate the communities of interest, several additional research methods can be employed. For instance, the work described by Bruns and Stieglitz [4], which focused on hashtagged conversations, can be included to deepen the awareness about how hashtags contribute to share knowledge and discuss events. The framework could also be further extended to consider both a static snapshot of the network structure of the community and its dynamic evolution over time. Finally, another point of interest could be evaluating the framework on different types of datasets,

like real-time data collected from the Twitter stream, data strictly related to an event (e.g. the World Cup) or only retweet data (for instance, to compare retweeted data with only tweet data).

## Acknowledgements

Alessia Antelmi thanks the Erasmus+ grant.

## References

1. Atefeh, F., Khreich, W.: A survey of techniques for event detection in twitter. *Comput. Intell.* **54**(31), 132–164 (2015)
2. Barnaghi, P., et al.: Opinion mining and sentiment polarity on twitter and correlation between events and sentiment. In: 2016 IEEE Second International Conference on Big Data Computing Service and Applications. pp. 52–57 (2016)
3. Bollen, J., Pepe, A., Mao, H.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *CoRR* (2011)
4. Bruns, A., Stieglitz, S.: Towards more systematic twitter analysis: Metrics for tweeting activities **16**, 91–108 (03 2013)
5. Chu, Z., et al.: Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing* **9**(6), 811–824 (2012)
6. Greene, D., et al.: How many topics? stability analysis for topic models. In: *Machine Learning and Knowledge Discovery in Databases*. pp. 498–513 (2014)
7. Harman, G.: Quantifying mental health signals in twitter **2014** (01 2014)
8. Hutto, C.J., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *ICWSM* (2014)
9. Ibrahim, R., et al.: Tools and approaches for topic detection from twitter streams: survey. *Knowledge and Information Systems* **54**(3), 511–539 (2018)
10. Java, A., et al.: Why we twitter: An analysis of a microblogging community. In: *Advances in Web Mining and Web Usage Analysis*. pp. 118–138 (2009)
11. Jonsson, E., Stolee, J.: An evaluation of topic modelling techniques for twitter
12. Lau, J.H., Collier, N., Baldwin, T.: On-line trend analysis with topic models: #twitter trends detection topic model online. In: *COLING* (2012)
13. Martinez-Camara, et.al: Sentiment analysis in twitter. *Natural Language Engineering* **20**(1), 1–28 (2014)
14. Musto, C., et al.: Crowdpulse: A framework for real-time semantic analysis of social streams. *Information Systems* **54**, 127 – 146 (2015)
15. Qiu, L., Lin, H., Ramsay, J., Yang, F.: You are what you tweet : Personality expression and perception on twitter (2013)
16. Tumasjan, A., et al.: Predicting elections with twitter: What 140 characters reveal about political sentiment. In: *ICWSM* (2010)
17. Wang, A.H.: Don’t follow me: Spam detection in twitter. In: 2010 International Conference on Security and Cryptography (SECRYPT). pp. 1–10 (2010)
18. Wang, W., et al.: Twitter analysis: Studying us weekly trends in work stress and emotion **2016**, 355–378 (01 2014)
19. Wolny, W.: Emotion analysis of twitter data that use emoticons and emoji ideograms. In: *ISD* (2016)
20. Wood, I., Ruder, S.: Emoji as emotion tags for tweets. *Emotion and Sentiment Analysis Workshop* (2016)