A Multi-perspective Approach for Analyzing Long-running Live Events on Social Media. A Case Study on the "Big Four" International Fashion Weeks[⋆]

Alireza Javadian Sabet[a,*], Marco Brambilla[a] and Marjan Hosseini[b]

[a] Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Via Giuseppe Ponzio, 34/5, I-20133 Milano, Italy.
[b] University of Connecticut, Computer Science and Engineering Department, 369 Fairfield Way, Storrs, CT 06268.

ARTICLE INFO

Keywords:
Digital Ecosystems
Long-running Live Event
Popularity Prediction
Fashion Week
Big Data
Social Media
Brand Analysis
User Generated Content
Instagram
Feature Selection

ABSTRACT

In the last few years, thanks to the emergence of Web 2.0, social media has made the concept of *online live events* possible. Users participate more and more in *long-running* recurring events in social media by sharing their experiences and desires. This work introduces long-running live events (LRLEs), as a type of activity that span physical spaces and digital ecosystems, including social media. LRLEs encompass several individuals, organizations, and brands collaborating/competing in the same event. This provides unprecedented opportunities to understand the dynamics and behaviour of event-oriented participation, through collection and analysis of data of user behaviours enabled by the Web platform, where most of the digital traces are left by users. What makes this setting interesting is that the behaviours that are traced are not focused only on one individual brand or organization, and thus allows one to understand and compare the respective roles and influence in a defined setting. In this paper we provide a high-level and multi-perspective roadmap to mine, model, and study LRLEs. Among the various aspects, we develop a multi-modal approach to solve the problem of post popularity prediction that exploits potentially influential factors within LRLE. We employ two methods for implementing feature selection, together with an automated grid search for optimizing hyper-parameters in various regression methods.

1. Introduction

Over the last few decades, social media (SM) has dominated day to day people's life by providing platforms for users to share content easily. It has become a consolidated and reliable source of information since it encompasses data originated from feelings and experiences of large groups of people, due to its accessibility and user-friendliness. In particular, SM has demonstrated a huge potential in communication, interactions, and community building during large scale events [42], such as the Arab spring in 2010 [49] and the 2008 U.S. presidential elections [52].

[⋆] This work was partially funded by: Regione Lombardia POR-FESR Project "FaST (Fashion Sensing Technology) - ID 187010". We thank the Fashion In Process group of Politecnico di Milano, and especially Paola Bertola, Chiara Colombi and Federica Vacca, who supported us in the definition of the domain-specific knowledge related to the event; WeManage Group, who helped providing the requirement specification; and Studio 4SIGMA and CGNAL for collaboration in the implementation of the prototypes.

[*] Corresponding author

alireza.javadian@polimi.it (A. Javadian Sabet); marco.brambilla@polimi.it (M. Brambilla); marjan.hosseini@uconn.edu (M. Hosseini)

www.arjsabet.com (A. Javadian Sabet); www.marco-brambilla.com (M. Brambilla); https://marjanh.com/ (M. Hosseini)

ORCID(s): 0000-0001-9459-2411 (A. Javadian Sabet); 0000-0002-8753-2434 (M. Brambilla); 0000-0002-0927-6658 (M. Hosseini)

https://twitter.com/ArjSabet (A. Javadian Sabet), https://twitter.com/MarcoBrambi (M. Brambilla), https://twitter.com/MarjHosseini (M. Hosseini)

https://www.linkedin.com/profile/view?id=arjsabet (A. Javadian Sabet), https://www.linkedin.com/profile/view?id=marcobrambi (M. Brambilla), https://www.linkedin.com/profile/view?id=marjhosseini (M. Hosseini)

1.1. Context and Motivations

In the following, we briefly discuss which categorization approaches exist to distinguish between different types of events and why Long-Running Live Events should be treated as distinct events. Researchers mostly follow two different event categorization approaches in the literature – namely, Topic-based and Crisis/Extreme event detection.

Topic-based categorization is the most widely used approach in the literature, which categorizes SM events according to their topics (*e.g.*, politics, sport, and so on [99]). According to the topic-based categorization approach, an event (*i.e.*, a topic) might be an actual event that is held in a location and has a predefined schedule; for example, [5, 87] monitor and investigate the participants' reaction to a cultural event. Also, it is possible that some events are not necessarily held physically and do not have a predefined calendar. For example, [84] explores online conversations of Italian users around vaccines on Twitter[1]. In general, considering the nature of topic-based categorization, events are not necessarily organized by any organizations, and the events might be about *Bursty topics*. Guille *et al.* [42] defines *Bursty topic* in social network sites as *"a behavior associated with a topic within a time interval in which it has been extensively treated but rarely before and after the event's duration"*.

Alternatively, some events can be characterized as **crisis/extreme** situations. In this case the discussion on SM is typically oriented toward the critical issues related to the event. In a way, this can be considered as a bursty topic about an actual event that is not organized in advance.

In this work, we define **Long-Running Live Events (LR-**

[1] https://twitter.com/

Table 1
The main characteristics of different approaches in categorizing social media events.

	Characteristics					
	Topic	**Schedule**	**Recurring**	**Location**	**Organizer**	**Bursty**
Topic-based event	Single	Maybe	No	Maybe	Maybe	Maybe
Crisis event	Single	No	No	Yes	No	Yes
Long-running Live event (LRLE)	Multiple	Yes	Yes	Yes	Yes	Yes

LEs) as *events that have a significantly long duration, are explicitly organized by some host entity, are held physically in some locations, and are periodically repeated for a long period of time*. Examples include festivals, fairs, and so on.

The user-generated content (UGC) related to LRLEs mostly reflects the participants' in-person experience during the actual event. Moreover, since they are well-established events in society, some governmental and/or non-governmental organizations organize such events. Unlike the other categorization approaches of SM events, LRLEs can cover more than a single topic. Table 1 summarizes the key characteristics of different approaches in SM event categorization. In the rest of the work, whenever we refer to events, we are explicitly targeting LRLEs.

Studying and understanding the dynamics of LRLEs is potentially appealing to many communities. For the public sector, it is a new door of exploring several unexplained topics such as societies' behavioral patterns and improving communication [59, 83]. For brands and businesses, it is a profitable place for commerce, in which they could attract extra public attention with lower cost compared to other forms of advertising [44, 86]. These advantages are achieved by Word-of-Mouth [100], as SM is a modern type of media governed by individuals who actively produce UGC [3]. Furthermore, the events that are covered by SM are more likely to promote the users' engagement, due to the information diffusion [7, 37, 68], specifically if the events are well-established in the society. From the research point of view, LRLEs provide potential topics such as *content popularity prediction* [95, 76], *measuring the profile influence* [90], *opinion formation* [2], *crowd preferences identification* [34, 103], *user profiling* [32, 75, 106], *recommender systems applications* [22, 104], *event detection* [28], *behavioral patterns recognition* [27, 70], and *urban resources allocation* [85]. By enjoying the rewarding properties like being held periodically and extensive coverage by SM, LRLEs are typically populated by users who would have paved for the mentioned communities to benefit the opportunities even more.

1.2. Objectives

Given the importance and potentials of LRLEs, in this work **we aim to design and propose a high-level conceptual model and a method that define the main elements of LRLEs and the procedure to study them based on data collected from social media**. To illustrate the feasibility and usefulness of studying these events, we provide a multi-perspective experiment on the Instagram's[2] posts related to the Big Four Fashion Weeks (New York, Paris, London, and Milano). Using this as a case study, we design an experiment that has the following objectives:

- Understanding the temporal dynamics of users' behavior during the events.
- Understanding the geographical distribution of the users and their posting activities.
- Studying brand popularity during the events.
- Detecting dependencies between the events' locations and the posting dynamics of participants targeting brands.
- Determining the main factors influencing the popularity of the user-generated content and design a model to predict the popularity of such content.

1.3. Structure of the Work

The rest of the work is organized as follows. Section 2 provides the background and state of the art. Section 3 introduces LRLEs and provides a high-level overview of their main elements and a conceptual procedure as a road map to study them. Section 4 provide some details about our experiment on the case study of Big Four's Fall/Winter Fashion Weeks 2018, including Explanatory Data analysis and modeling post popularity. Finally, Section 5 is conclusion.

2. Related Work

As mentioned earlier, the aim of this work is to design a framework for extracting the knowledge and designing predictive model to predict the popularity of SM posts during such events. Section 2.1 provides a literature review regarding the various exploratory data analysis techniques on SM. Section 2.2 discusses related work on predicting the popularity prediction of SM posts.

2.1. Exploratory Data Analysis on Social Media

Spatial analysis of users' response to Milano Fashion Week event on Instagram and the event coverage was previously investigated by [14]. They observed the propagation of the brand's related SM events, using posts' geolocation information and various time windows with different duration. Another geographical analysis on Instagram studied the data consisting of users' information collected from celebrities' accounts whose posts were viral at a particular time and some random selection of their followings and followers, along with their posts. They investigated to what

[2]https://www.instagram.com/

extent Instagram posts are geo-tagged and how they are geographical spread in the world. Besides, they provided the world's most popular locations frequently used in the posts [71].

Users analysis on Twitter data attempted to determine whether a user is a member of a community [88] using a set of semantic features of the tweets, mainly the vocabularies, based on the idea that a set of users sharing similar vocabulary can form a community. [6] provides users behavioral analysis from different aspects, such as users' preferences in posting time and their reaction to other posts. They extracted information from more than 1M images and videos from Instagram posts and showed that users tend to like posts that already have a considerable number of likes. Concerning the users' network, [71] plotted the correlation between the users' followers and followings count. They found that these counts are linearly correlated when users have a fair number of followers. [51] studied users' posting behavior and the most popular hashtags and camera filters. They focused on the analysis of the individual posts. For instance, they investigated Instagram posts by applying computer vision techniques on the post images, categorized them, and provided relative popularity of categories and clusters of users.

Brands analysis has been previously done by [14] and investigated the popularity of 65 active brands in the Milano Fashion Week event. They applied Principal Components Analysis (PCA) [31] to extract the most influential factors on the brands' coverage, then clustered the brands into four groups based on the extent to which the responses are dispersed and demonstrated that the brands' popularity is not highly correlated to the geographical features. [15] explored the communities that arise around commercial brands on SM. They aim at understanding the meaning of similarity, collaboration, and interaction among the members of these communities. To do so they encoded the communities network into a graph model which contains the user nodes and friendship relations. Then, they compare the communities graph model with a heterogeneous graph model composed of posts and hashtags. Finally, by inducing direct user-to-user connections through the posts and hashtags as the intermediate nodes, they build a reduced network.

2.2. Post Popularity Prediction on Social Media

Multi-modal approaches are the predominant choice for predicting SM content popularity. Some examples are [4, 40, 57, 61, 64, 73, 74, 98, 102, 108], which incorporate different aspects of posts, particularly characteristics of the post generator and posts' content and information which are the features encoded in the posted media. For instance, khosla *et al.* [61] incorporated social context and image-related features of a post in a multi-modal approach. They define popularity as log-normalized view counts in Flickr[3]. They applied Support Vector Regression (SVR) [8] using different types of features and reported Spearman's rank correlation [94] (up to 0.81) of the obtained popularity and the actual popularity by using both the mentioned features categories.

Since the focus of the work was on understanding image characteristics and limited information about the social signal, they didn't consider the context in which those images were taken; for example,

Totti *et al.* [98] studied Pinterest[4] data with *number of reshare* as the popularity score. They divided the features into three groups of visual, semantic, and social-network properties. They considered two extreme classes by applying a binary classification on whether an image will be highly popular or unpopular by using a random-forest ensemble [107] of 200 tree estimators.

Another multi-modal approach [50], done on 10k samples from Yahoo Flickr dataset [96]. They considered tags and visual features. They then built SVR with linear and Radial Basis Function (RBF) [23] kernels and Multiple Kernel Learning (MKL) [38] and reported the results in terms of Spearman's rank correlation among the true and predicted output. The best method was SVR with an RBF kernel.

Multimodal context-aware recommender for post popularity prediction in Social Media [73], is a research conducted in order to predict the popularity of items (*i.e.*, places) considering individuals' preferences regarding the items in the model. In their study they used a dataset containing 600K posts collected from Instagram which are related to different touristic places in The Netherlands (as items). The predictor is designed based on Factorization Machine (FM) [89] which has been extended in their case employing visual and textual contents as information. The results suggest that it is beneficial to apply multi-modal context-aware recommender to model the post popularity.

Jaakonmaki *et al.* [56] studied the impact of content, context, and user engagement in marketing on Instagram. They considered *likes* as an indicator of interest, and *comments* as the degree of verbal interaction. They included creator-related, contextual, and content features. The later was extracted through Natural Language Tool Kit (NLTK) [13] and Clarifai[5]. They utilized least absolute shrinkage and selection operator (LASSO) [97] as the model and found that 40% of the deviance in engagement can be explained by only 10 predictors, while to reach 50%, half of the total number of predictors were needed. The most impactful features were reported mainly to be creator-related ones such as the number of followers, age, and sex.

Zohourian *et al.* [110] explored information from videos and images collected from three Instagram business accounts consisting of 271 instances for predicting popularity as the number of likes. They regarded the prediction problem as both regression and classification problems. As regression methods, they applied linear regression, local polynomial regression, and SVR. They achieved the lowest root-mean-square error (RMSE) of 0.002 using local polynomial regression. Before applying the classification, they categorize popularity into three classes: low, medium, high, then utilized k-nearest neighbor, random-forest, Naive Bayes, C4.5, and decision tree [65]. They achieved an accuracy of 90.77%.

[3] https://www.flickr.com/

[4] https://www.pinterest.com/

[5] https://www.clarifai.com/

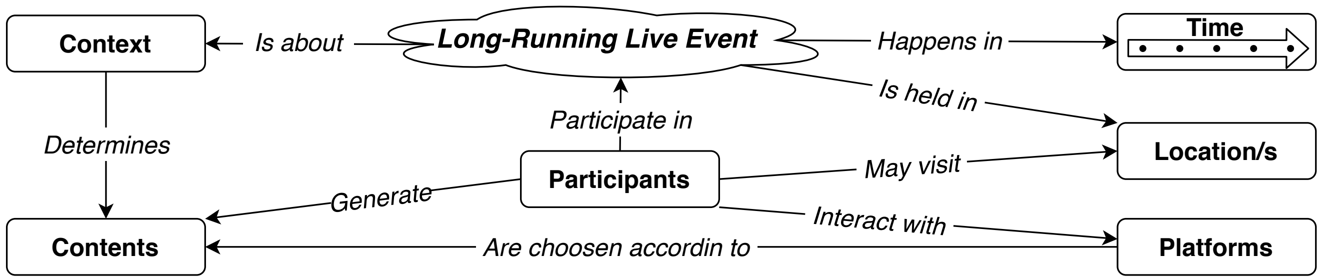


Figure 1: Conceptual overview of the main LRLE's elements and their relationships.

The study in [6] has collected 1.2M posts of 200K Instagram users, including both the popular and the ordinary ones. They focused on user-related and post-related features. They have concluded that users with more followers receive more attention and popularity, as *the rich get richer* phenomenon. Moreover, they discovered that the use of more hashtags could result in attracting more audiences. Finally, they realized that users tend to post during the weekend and in the afternoon while their temporal investigation did not provide information on whether such behavior is the same for LRLEs or not.

Yamaguchi *et al.* [105] explored Chictopia[6], which is a fashion-focused online community. They modeled the popularity of outfits pictures by applying a multi-modal approach incorporating visual, social, and textual factors. Their data is a collection of around 320k images, from 34k unique users, and popularity is defined as the logarithmic number of votes, comments, and bookmarks. By applying linear regression on the log-votes, they confirmed that users' social features dominate the popularity of posts. Also, they modeled the problem as a binary classification to identify whether the post will be among the top $k\%$ of the most popular ones or not. It resulted in knowing that recognizing the most popular posts is easier than the least popular.

3. Long-Running Live Events (LRLEs)

Guille *et al.* [42] defined *Bursty topic* in online social networks as *"a behavior associated with a topic within a time interval in which it has been extensively treated but rarely before and after the event's duration"*. Inspired by this concept, we consider two categories of bursty topics, namely, *irregular* and *regular*. The topics that happen without any specific schedule are irregular bursty topics such as natural hazards (*e.g.*, [91]), accidents [81], and rainy days [92], while regular bursty topics happen regularly and have pre-defined calendars, such as Earth Day. In this section, we introduce LRLEs as examples of regular bursty topics, provide a conceptual overview of their elements, and a high-level procedure as a road map to studying them.

There are many potential research topics in the context of LRLEs, including but not limited to: *content popularity prediction* [76, 95], *measuring the profile influence* [90], *opinion formation* [2], *crowd preferences identification* [34, 103], *user profiling* [32, 75, 106], *recommender systems applications* [22, 104], *event detection* [28], *behavioral patterns recognition* [27, 70], and *urban resources allocation* [85].

[6]http://www.chictopia.com/

3.1. Definition

Long-running live events (LRLEs) are periodically repeated events like festivals that are held physically in some locations and are covered on SM. The UGC related to LRLEs mostly reflect the in-person experience of the participants in the actual event. Moreover, since they are well-established events in the society, there are some governmental and/or non-governmental organizations that organize such events. Unlike the other categorization approaches of SM events (see Table 1), LRLEs can be about more than a single topic.

3.2. Elements

Figure 1 depicts a conceptual overview of LRLE's principal elements and the relationships among them.

- **Location:** Events can be held in a place or in many places. If the event is held in multiple locations, one should investigate the possible interactions among places. Furthermore, exploring the events' location will hint about the cultural preferences and behavior of the participants.

- **Time:** Schedule and frequency of the events can vary occasionally. LRLEs might have multiple sub-events happening at different times. The key point is that the events are holding periodically. If the schedules of events overlap, the possible interactions among them should be inspected too.

- **Context:** Understanding the context of the LRLE is crucial. Not only, the event's topics, participants, and the type of UGC are determined according to the context. But also, the study's goals are determined according to the context. The context can be related to a single or multiple industries. Identifying the engaged industries provides an insight into the target participants and the potential content to be analyzed. Some examples of events from different industries are EXPO, Comic-Con, and Fashion Week (FW).

- **Content:** Content is the shared media by participants such as videos, audios, images, and texts. The type of preferable content depends on the context of the LRLE. For instance, in fashion week events, generating images could be more favorable for the users, while in political events, textual participation might be required.

- **Platforms:** Sources of data for studying LRLEs. The choice of target platform/s is dependent on many factors. The main factors can are:

  - The main preferred type/s of content by participants of the target LRLE. For example, in case of FWs, the most informative platforms are those that provide visual contents (*i.e.*, image and video).
  - The topology of the platform which can be either unilateral (*e.g.*, Twitter and Instagram) or bilateral (*e.g.*, Facebook) [42].
  - The policies and limitations provided by the platform's application programming interface (API).

- **Participants:** Participants of the event can be categorized into two groups. The first group is the organizers of the event like brands and industries, and the second group is people who attend the events. The first group of the participants' motivation is maximizing the second group's attendance, and the second group is the ones who share the online content.

3.3. Procedure

Figure 2 shows a high-level road-map to extract knowledge from LRLEs.

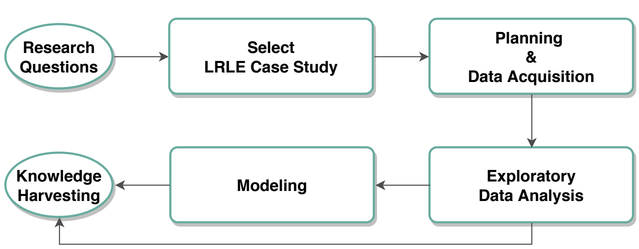


Figure 2: High-level road map showing the main steps to harvest knowledge from LRLEs.

Considering the potential of LRLEs, *research questions* might arise to seek characteristics, behavior, and the interactions of LRLE elements during the event. Research questions can also be hypotheses about the cause of phenomena. LRLEs are characterized by their elements, and a case study is a combination of specific choice for each of those characteristics. Having in mind the research questions, the best *LRLE case study* can be selected according to the requirements to answer the research question. For instance, if the research questions are about the outcome of elections and the influential factors for the popularity of the candidates, Twitter might be a better platform to study than Instagram or YouTube[7]. Likewise, if the research questions are related to the fashion industry, the platforms with more visual options such as Instagram and Flickr might be desirable. Other elements can be set according to the research need; whether we intend to study a phenomenon on a particular group of people or worldwide, do we need to have information about recent events, or should we consider more repetitions of the same event. Answering these questions requires in-depth domain knowledge, and this step is the prerequisite for *planning and data acquisition* step, which includes data collection, preparation, and cleaning. Then interesting statistical analysis can be performed, either on LRLE elements or other related subjects. The insights of the *Exploratory Data Analysis* step might be adequate to answer the initial questions or raise new questions. Having more information about the case study, one can proceed to model other interesting phenomena. In this regard, employing machine learning (ML) techniques potentially results in extracting the *knowledge* for that purpose. For example, utilizing feature engineering techniques provide information about the most influential factors. One would expect that investigations focused on temporal drift, temporal analysis and evolution (*e.g.*, concept, sentiment, *etc.*) would be the central point of interest given the definition of LRLEs.

3.4. Challenges

In fact, most of the issues and challenges that one may face during studying LRLEs, are the ones when dealing with SM in general. Here we name some of the most important ones.

3.4.1. Platform related challenges

Some of the main challenges specific to the SM platform are as follow.

- Limitations due to the policies and design of platforms, which result in the lack of information. For example, Instagram does not provide the temporal evolution of the post's number of likes.
- The noise caused by the platform's design. For instance, in the case of Instagram, when requesting the posts related to a specific hashtag, even if the post's caption does not include that hashtag, but it is mentioned in the comments by anyone, the post appears in the result.
- API's constraints like the allowed number of requests for data collection.
- Legal issues which are mostly related to the copyright.
- Privacy and regulations modifications that may occur in the time of the data collection phase.

[7] https://www.youtube.com/

3.4.2. Four V's of Big Data

The data driven approach of users' behavioral analysis is based on the concept of big data paradigm [11, 19]. Understanding this paradigm is mandatory for studying LRLEs due to the large number of mobile sensors that need to be governed and continuously produce data. According to [109], the four V's of big data are:

- Volume: The amount (or size) of data is exponentially increasing.
- Variety: (*a.k.a*, heterogeneity) The types of data that need to be analyzed are many, spanning textual, numerical, and multimedia content, which might hinder applying statistical methods.
- Velocity: Sensors data is continuously generated and communicated, making the whole approach a continuous quasi-realtime solution.
- Veracity: The amount and speed of data pose challenges to the quality or correctness of the data, which need to be monitored and validated to avoid wrong perceptions and actions.

4. Experiments

This section investigates FW events as the case study.

4.1. Case Study

FWs encompasses all the main elements of LRLE and is an excellent example to study. It is internationally accepted and participated by many users, covered well in SM, and engages many brands and organizations. We selected Big Four's Fall/Winter Fashion Weeks 2018 as our case study since it is the latest FW event before starting our experiment. London hosts two FWs events[8] in multiple time slots. *Mens London FW* (LFW(MENS)) started in January 6th, 2018 and lasts for three days. *London FW* (LFW) started in February 16th, 2018, and lasts for five days. However, *New York FW*[9] (NYFW) does not split its events into time slots with a gap in between. Its events started in February 2nd, 2018 and ended in February 20th. Being nineteen days long makes it the longest event. Milan FW[10] consists of two separate events. The first one *Mens Milan FW* (MFW(MENS)) started in January 12th, 2018 and continued for four days while the second one which is *Milan FW* (MFW) started in February 20th, 2018 which is the last day of both LFW and NYFW, and finished in February 26th, 2018 which is the starting day of Paris fashion week. Paris FW[11] hosts the second longest event, and follows the same strategy as London and Milan for splitting the events into two sub-events; *Haute Paris Fw* (PFW(HAUTE)) and *Paris FW* (PFW) each of which lasts for nine days, starting from January 17th, 2018 and February 26th, 2018 respectively.

[8] https://londonfashionweek.co.uk/
[9] https://nyfw.com/home/
[10] https://www.cameramoda.it/en/
[11] https://fhcm.paris/en/paris-fashion-week-en/

In the rest of the work, the terms "event" and "city" are used interchangeably. In addition, the term "caption" refers to the textual content of the post.

4.2. Data Collection

We collected our dataset using Instagram API [55], since to the best of our knowledge, there is no benchmark dataset regarding FWs. Data includes event related posts and media shared on Instagram starting from Jan. 1st, 2018 to March 11th, 2018 (five days before the first event *i.e.*, London FW Men and five days after the last event *i.e.*, Paris FW). We found the events' most used hashtags by manually exploring Instagram's search function and other online resources as the hashtag seeds. We collected over 3M related public posts and the user profiles who published those posts. The seeds for four cities are different, but in the end, we merged all the posts as a single dataset and added a categorical attribute showing which post is related to which city event. Unlike many other studies that collected posts of a few users or specific types of users such as celebrities, we added diversity to the data by *hashtag-based* data collection approach. The resulting dataset is composed of $905,726$ posts, $171,078$ correspondent user profiles and $723,831$ images. Due to inherent noise in the data collected based on keyword search [21], data cleaning approaches were exploited. First, we removed duplicated posts that contained multiple hashtag seeds from the data (duplication removal). Then the posts with missing important attributes were eliminated (field error removal). It should be noted that the mentioned data cleaning steps have been done before the analysis presented in Section 4.3.5. As a result, the potential user profiles identified as bot/fake were not removed from the rest of the analysis. The reason for not removing the potential bot/fake accounts and their related posts from the rest of the analysis was that we considered them influential entities in the system that need to be analyzed. The dataset is available at [17] and its description are presented in [18].

4.3. Exploratory Data Analysis

This section discusses the results obtained from the exploratory data analysis.

4.3.1. Hashtags Frequency Analysis

We investigated all the hashtags mentioned in *posts' caption body*; these hashtags include both the hashtag seeds and the extracted ones. The distribution of hashtags usage frequency is extremely *heavy-tailed*. The total number of hashtags used in all the posts is almost 14M, which less than 15% of them have been used more than or equal to 10 times. Figure 3 is the *WordCloud*[12] representation of the top frequently used hashtags in posts' captions. The words bigger in size are the most frequently used ones. The presence of the leading hashtags of fashion week events, despite their absence in the initial hashtag list, confirms that the seeds for data collection have been chosen appropriately.

[12] https://github.com/amueller/word_cloud

Figure 3: Word cloud representation of the most frequent used hashtags in the case study.

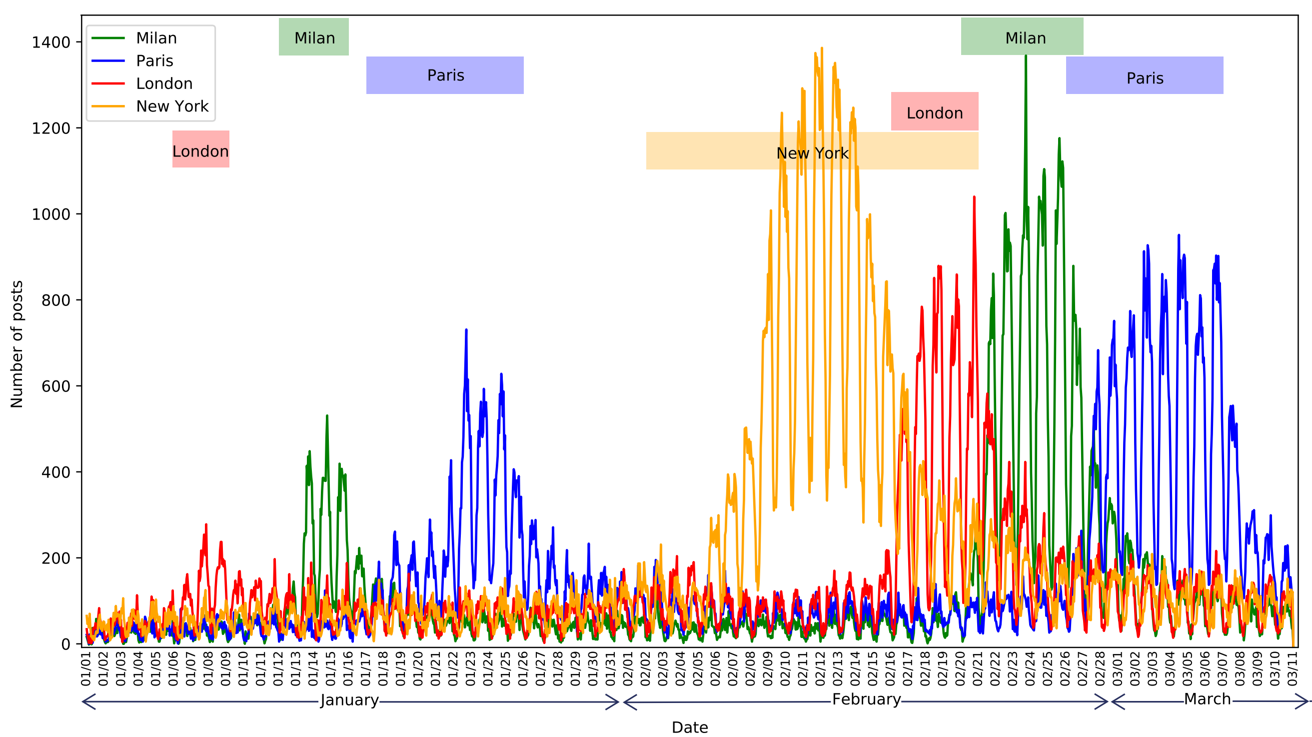


Figure 4: Temporal posting behavior of the Instagram users of the case study with the granularity of 1 hour. Different colors define the cities signals and the colored boxes in the background, specify the official calendar for each sub-events.

4.3.2. Temporal Analysis

To investigate the information obtained from the posts regarding the date and time they were published, we considered events (cities) as temporal signals with the granularity of dates and hours, the signals start from the first day of the study and continue until the last day. Then, we assigned all the posts in data to one or some of the four signal events according to the presence of representative hashtags of that particular event in the post. In case those posts contain more than one representative hashtags, we assigned multiple labels to them. The magnitude of each event signal at a specific hour is equal to the total number of posts labeled to that event in that particular hour. Moreover, if a post contained hashtags about both events in a single city (*e.g.*, about FW Men and Women) then it is considered a focused post on the entire super-event (FW Milan). It should be noted that the events' representative hashtags are the ones employed for data collection. Figure 4 illustrates the acquired signals depicted in different colors for each event along with the actual time of the events according to the fashion weeks calendar in the background. For better visualization of the relationships between the posts' signals and calendar, the event color's actual time is the same as the signal related to that particular event. The plot approves that the temporal dynamic of the users posting for the events is in direct relationship with the actual events, coinciding with the peaks of events' signals to the middle of actual events. Besides, the signal value increases sharply just before the corresponding events start, continues its growth until the middle of the events, and then decreases moderately after the events conforming temporal dynamics of bursty topics [42]. Figure 4 emphasizes the importance of considering the temporal aspects of the posts with respect to the events. Concerning individual signals, the first peaks in Milan and London related to men FWs in these cities suggest that men FW events were less popular.

4.3.3. *Hashtag Relevancy Analysis*

To inspect the extent to which posts are genuinely related to the event represented by the hashtags in their caption, it is possible to add four extra Boolean fields: *Milan*, *Paris*, *London* and *New York* to each post. Their values represent whether the captions of a specific post contains at least one of the hashtags seeds specific to each city or not. After that, we calculated the percentage of posts' relevancy to the cities by showing the degree in which the posts of cities overlap.

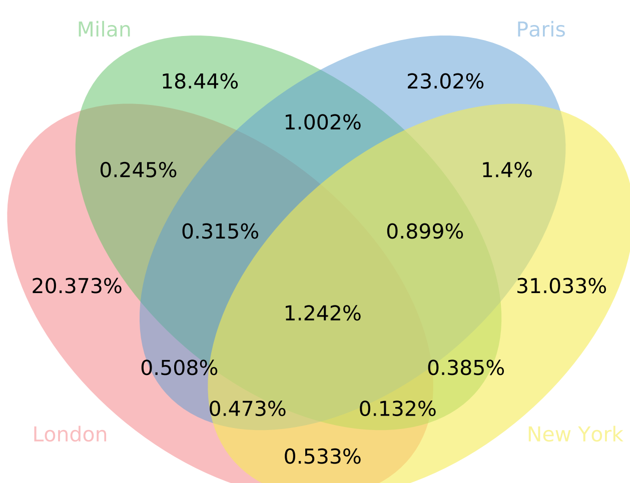


Figure 5: Venn diagram representing the portion of dataset posts contain hashtags of the different combination of cities.

The Venn diagram[13] in Figure 5 reports all the possible logical states of the posts relative to the cities. It suggests that the majority of posts (92.866%) have hashtags that are just related to one city hashtag list. This means that with more certainty, these posts are indeed related to that specific event, while the other posts include hashtags related to more than a single city. The latter could increase the uncertainty about the real association of these posts and the corresponding events. Users who posted these contents might have used these bunch of hashtags just to increase the visibility of their posts since each post logically can occur for one of the events unless the user intentionally aimed to generate a more general content than the scope of a single event, for examples, about *fashion week* in general, or to compare events of multiple cities.

[13]https://github.com/LankyCyril/pyvenn/blob/master/pyvenn-demo.ipynb

Table 2
Instagram users who posted about Big Four's Fall/Winter 2018 fashion weeks categories according to the purity of their posts.

User Category	Percent
Pure Content Generators	94.1%
Mixed Content Generators	1.86%
Pure and Mixed Content Generators	3.23%

4.3.4. *User Categories According to Posting Behavior*

Motivated by the polarization in targeting a single event or multiple events by posts (Section 4.3.3), we want to categorize users according to their posting behavior. Each post is about one or more events. If a post refers to only one city, we consider it a pure post; otherwise, we consider it a mixed post. According to this categorization for the posts, in the following, we can consider a rough estimation of three categories of the users according to their posting behavior *i.e.*, if the user's posts are about a single event or multiple events.

(i) **Pure Content Generators:** This category refers to the users whose posts always target a single event (pure post). A user in this category might target MFW in one post and NYFW in another post, but not both events simultaneously in a single post.

(ii) **Mixed Content Generators:** Unlike pure content generators, the mixed content generators always target more than a single event in each of their posts. In other words, the set of posts made by these users does not contain any pure posts.

(iii) **Pure and Mixed Content Generators:** The users who have both pure and mixed posts in the posts set belong to this group.

The third group (iii) perhaps justifies posts' relevance to the event, because they are the ones having both kinds of posts. This might suggest that they know the difference between these kinds of captioning. If it was just for visibility, they could have posted with multiple related hashtags all the time. Table 2 reports the share of each of the categories of users in the case study.

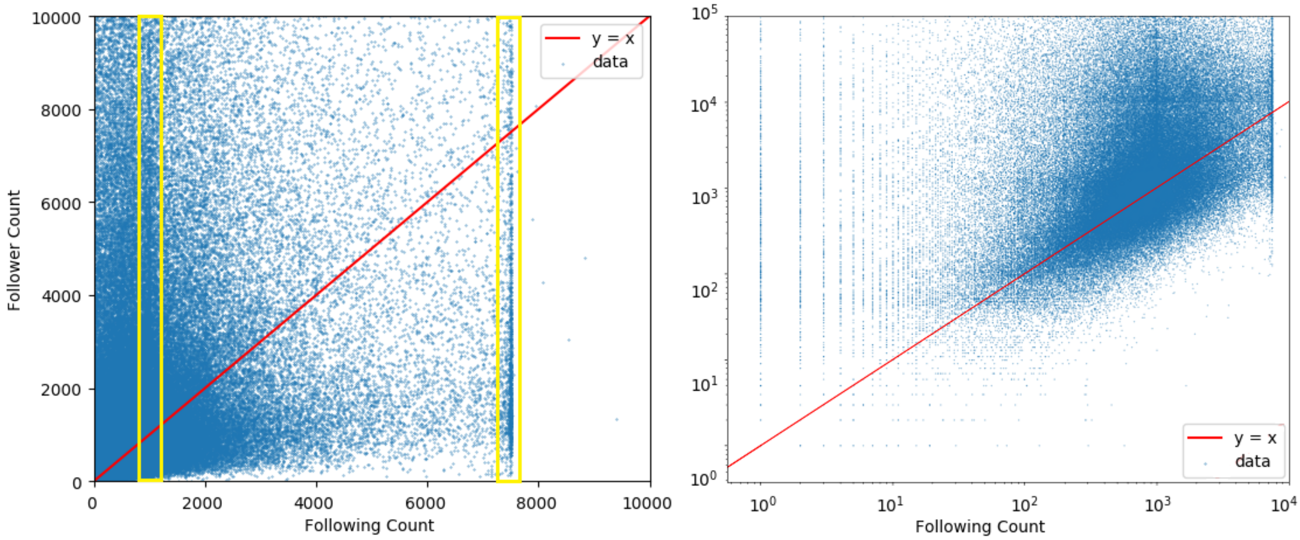


Figure 6: (left) Users' number of followers (*y-axis*) vs. number of followings (*x-axis*) and (right) in logarithmic.

4.3.5. Users Social Network Analysis

We investigated users' following and followers counts. This analysis gives a rough idea about the shape of the users' network in the data. Figure 6 (left), plots the correlation between the number of followers and followings of each user. Each point corresponds to a single user, with the number of followings and followers (limited to 10, 000 for readability purposes. For higher follower vs. following counts please check Figure 16) on the x and y axes, respectively. Figure 6 (right), reports the same plot in log-log scale.

In Figure 6 (left), the most upper-left and the bottom-right corners are related to the extreme types of users, probably celebrities and bots with a high unbalance between followers and followings counts. In addition, the majority of the data points fall above the red line, which suggests that most of the users have more followers than following. Precisely, out of 171, 078 user profiles in the dataset, the number of users having more followers than following is 116, 191 (67.92%), while the users who follow other accounts more than they are being followed are 51, 559 (30.14%). Only 3, 328 (1.94%) accounts have the same number of following as followers, which shows our users are mostly influencers. This is reasonable considering the nature of the users in the dataset, who are the people who published for long-running live international events. From another perspective, if we consider the users as a subset of Instagram network mapping to a directed graph, this diagram suggests that the number of incoming edges (external users who follow the subset of users in the dataset) is more than the number of outgoing edges (the users whom our dataset users follow). Moreover, the scatter plots in Figure 6 reveal some compelling patterns. There are few regions (highlighted in yellow boxes) where the number of followings are much higher than the number of followers. For example, around 1, 000 and 7, 500 on the x-axis. Also, a few users have a number of following more than the peak in 7, 500, even though the plot in the x-axis is limited to 10, 000. We compared our log-log plot (Figure 6 (right)) to the same plot manikonda *et al.* [71] provided; the general distributions in both are the same, but these artifacts are specific to the users in our dataset.

In a closer look at Figure 6 (left), we found three more dense areas, referred to as peaks in histogram (See Figure 7), located in position 0, 999 and 7, 500, corresponding to the followings count of 840, 169 and 234 users in the dataset. The first peak is as a result of 840 users with exactly 0 followings, but different followers count. They are probably celebrities who decided not to follow anyone or unused or fake accounts. There is also a possibility that Instagram has banned some users from following other people. On the other hand, the third peak, which collects all the dots in the very right part of Figure 6 is probably the bots, who follow many accounts to attract followers or increase the visibility of their posts or their account. Their overall success in this regard has been reflected in the same figure. Their followers count on average are less than their followings count. This peak corresponds to 234 users who follow exactly 7, 500 accounts but have different followers count. This specific number could result from a policy by Instagram that forbids users to follow more than 7, 500 profiles.

4.3.6. Geographical Analysis

To understand the users' and posts' geographical distribution, we extracted geo-related metadata. Among 905, 726 collected posts, 42.59% are geo-tagged. To analyze and compare the geographical posts distribution, specific to a single event, sub-figures in Figure 8 provide the heatmap of the posts related to the particular events showing the distribution of the posts worldwide and Figure 9 is a zoom to provide a better comparison of the engaged regions.

Figure 10 illustrates dataset users' spatial distribution. The locations obtained from the registered city of residence in the user profile. Indeed, the red dots in the map account for

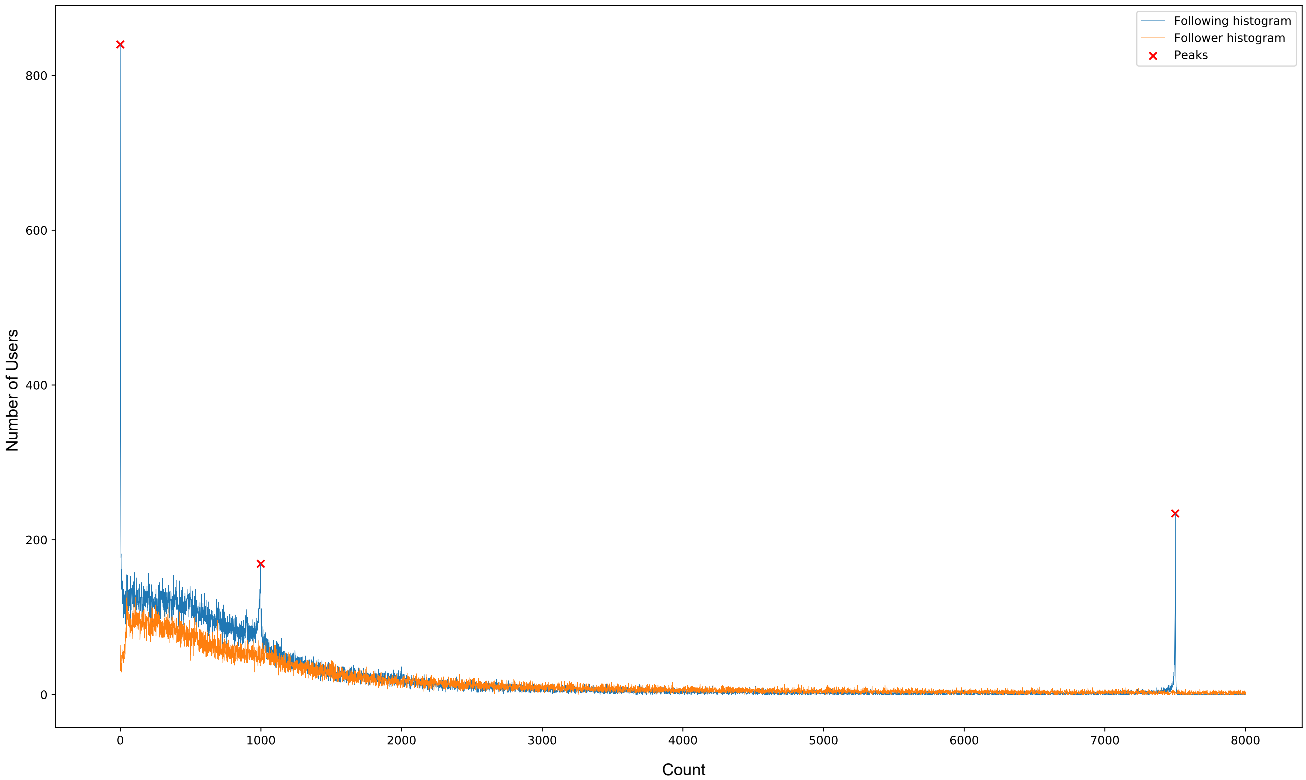


Figure 7: Histogram of the number of following (*blue*) and followers (*orange*) on *x-axis* both limited to 10,000, and number of users with the correspondent numbers on *y-axis* for the Instagram user's profile who posted about the case study. The peaks with high values are crossed in red.

(a) (b) (c) (d)

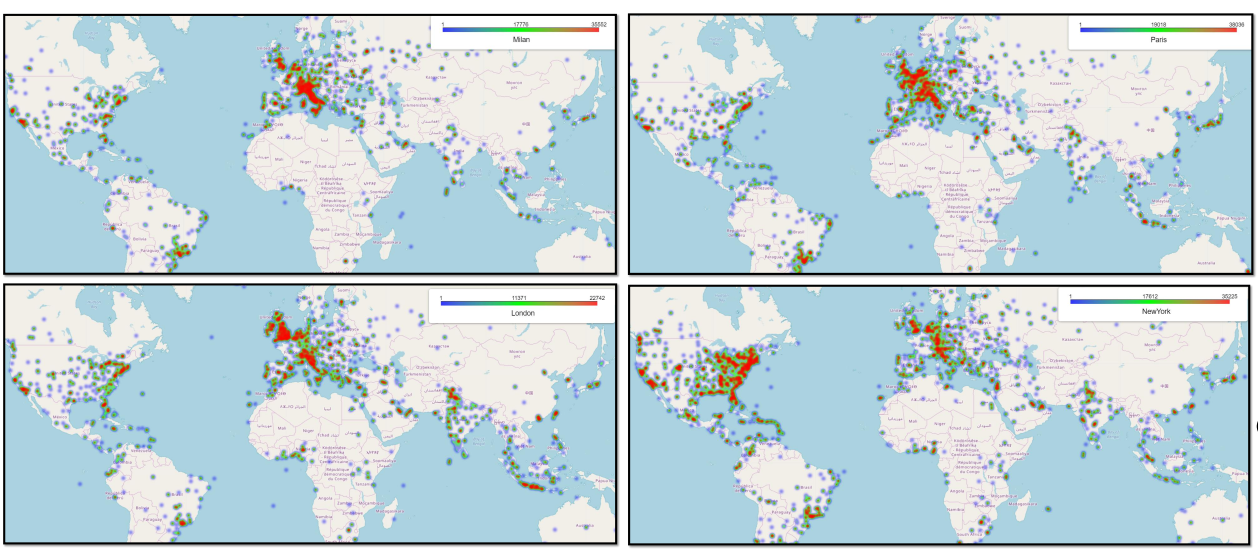
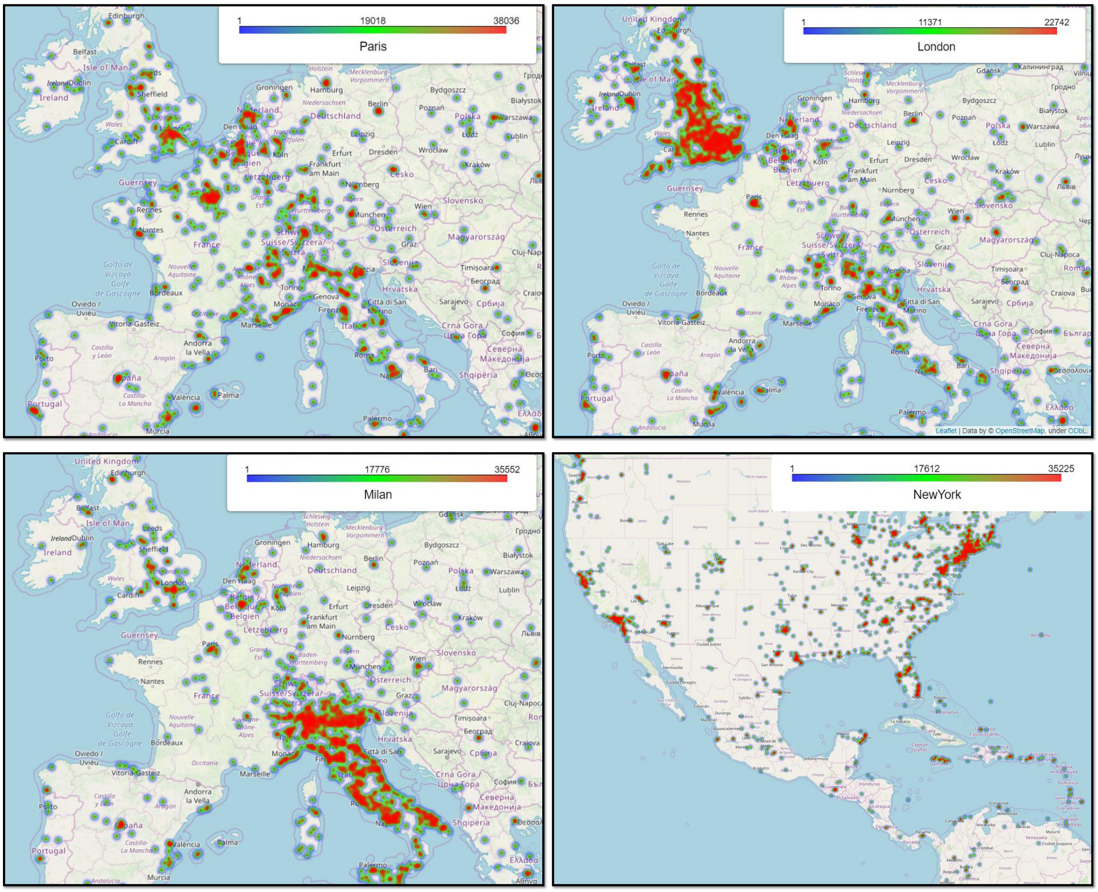


Figure 8: Heatmap representing the worldwide geographical density of the Geo-located case study posts for (a) Milan, (b) Paris, (c) London, and (d) New York.

53.16% of the users for whom location metadata was available at the time of data collection.

From Figures 8 and 10 it can be grasped that most of the posts have been published in Big Four's regions and mostly by the users who are living in the same regions, or in the other cities which host other fashion week events, *i.e.*, the case study events are mainly attended and/or talked about by local people and attract less international attention than expected. The reason for this claim is that the distribution of the data in these figures are very similar.



(a) (b) (c) (d)

Figure 9: Heatmap representing the regional geographical density of the geo-located case study posts for (a) Paris, (b) London and (c) Milan in Europe, and (d) New York in U.S.

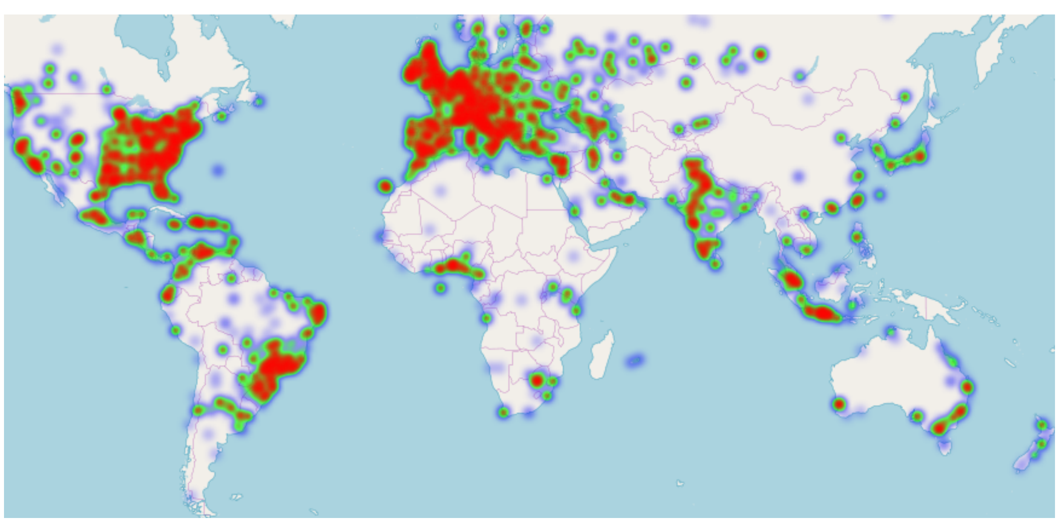


Figure 10: Geographical dispersion of the Instagram users in the case study.

4.3.7. Brand Analysis

To identify the brands participated in the FWs, we manually extracted the name of the brands from FW online website[14]. In order to build a list of hashtags for each of the brands, we investigated the posts published by their official Instagram account. We examined some of the brands in the case study that gained more attention from the users. Considering the distribution of posts related to each brand in four cities, we would like to know if the city would affect the

[14]https://fashionweekonline.com/

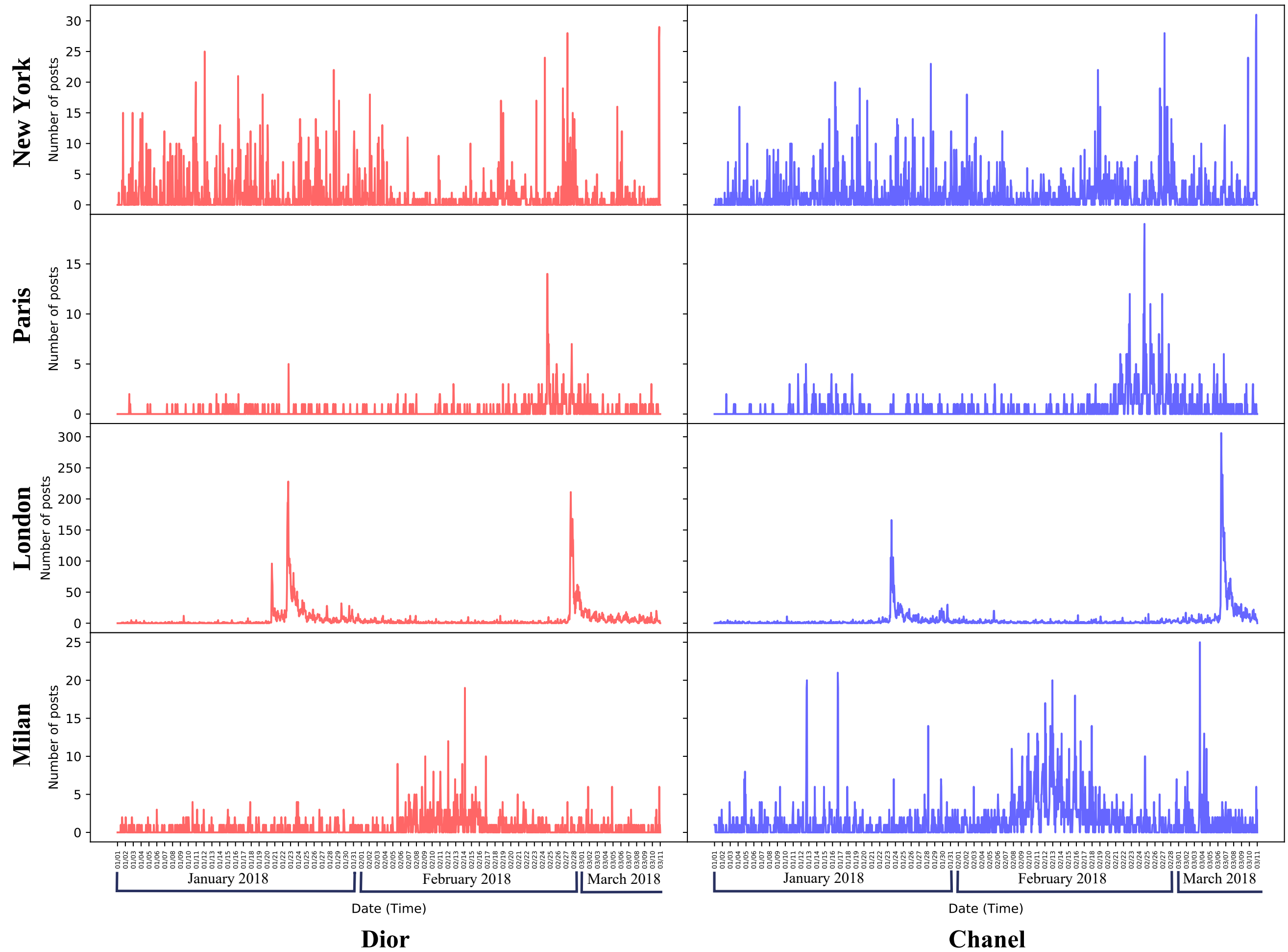


Figure 11: Users' responses to (*left*) Dior in red vs. (*right*) Chanel in blue in the case study for the entire experiment period and each city separately (on *y-axis*) with the granularity of 1 hour.

coverage of a brand or not, *i.e.*, whether the distribution of the number of posts related to a specific brand is dependent on the city hosting Fashion Week or not. To do so, we divided the duration of FWs to hours and extracted temporal signals for leading brands, such that the magnitude of each signal is equal to the number of posts containing the brands' hashtags published in that hour. Then we focused on two of the most frequently tagged brands, *Chanel* and *Dior*, by further decomposing their signals to include just the posts tagged in one city at a time and obtained eight signals. Figure 11 compares the temporal dynamics of Dior and Chanel categorized based on the posts related to each of the cities. Please note that in order to make the peaks for each city the y-ranges are different per city. The shapes suggest that the dynamic of brands in the same city are more similar than the dynamic of the same brand in different cities. Besides, we computed pair-wise Spearman's rank correlation among the acquired signals and provided a heatmap in Figure 12. Even though the events' signals were not shifted to be aligned with each other, the results confirm that the city is more impactful than the brand in terms of both dynamic shape and magnitude. Let us take Dior posts in Paris as an example; its correlation score with Chanel in Paris is 0.59, which is higher than Dior's scores in other cities such as Milan, which is 0.1. Thus, one may interpret given the temporal dynamic of a brand in a city, it is possible to predict the temporal dynamic of other brands in the same city.

4.4. Modeling Post Popularity

To the best of our knowledge, this is the first work aim to model the UGC's popularity on SM during LRLEs. State-of-the-art in predicting the popularity of UGC on SM discussed in Section 2.2, has one or more of the following shortcomings:

The first is related to ignoring the potential relation between the post and event. Since most of them focus on predicting the popularity of UGC on SM, they ignored considering the information related to the events. For example, while some work considers only the days of the week as an indicator of the posting time, we consider the post's timestamp with respect to the time of events to see if the post is published *before*, *during*, or *after* the event. The sec-

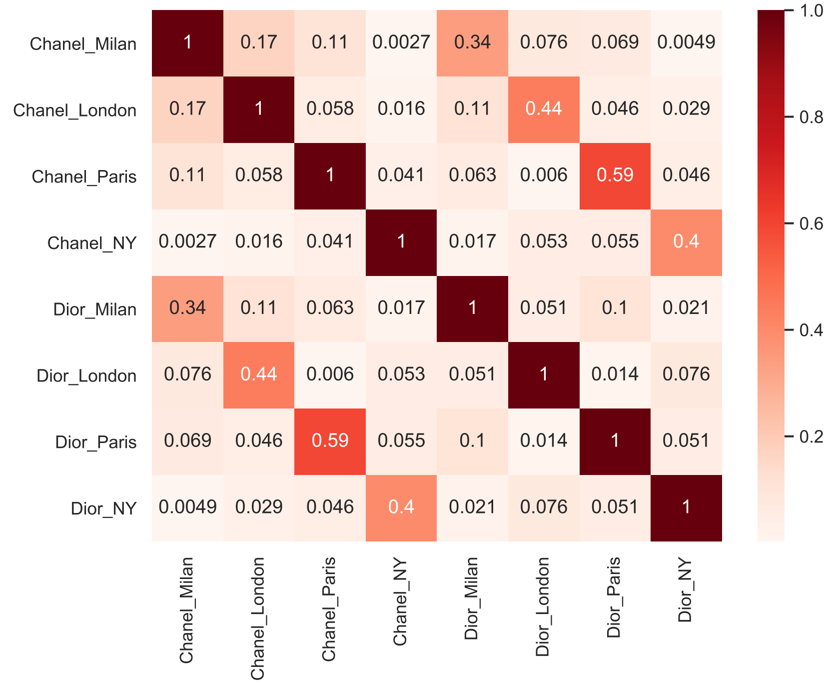


Figure 12: The Heatmap matrix of the Spearman's correlation analysis showing the correlation coefficients among the values obtained from the responses to the brands (Dior and Chanel) for each of the Big Four cities.

ond is related to the profile-oriented collection of the information related to the user activities. For instance, in most cases, they collect some of the recent posts by the user directly from the profile and ignore the context in which each of those posts has been published. In this work, for each user, we investigate their posts related to the events of the study. The third is related to the *interpretability* of the predictive model provided in the literature. To reduce the predictive model's complexity and provide the set of most influential features, we utilize the proposed pipeline with feature selection techniques. The fourth is related to the evaluation of the predictive model itself. We evaluate the effectiveness of the proposed model using the number of likes.

LRLE is a valuable source of data in terms of reliability since it encompasses the information that originated from the real feelings and experiences of large groups of people. Moreover, many organizations and brands are involved in these events, which provides an unbiased environment for them to evaluate their objectives by analyzing users' behavior and preferences. It is often in the interest of brands to quantify their popularity and compare it to other peers. Researches such as [29, 30, 46, 62, 63] provide statistics confirming the numerous opportunities for brand management by making the best use of SM. Consequently, investigating the popularity of content is rewarding for many communities. This is a way for them to build and maintain brand loyalty [10, 24, 80], which in return provides benefits; for example, market share, sales revenues, and so on and so forth [1, 58, 60].

As a result of the mentioned reasons, we aim to model and estimate UGC's popularity on SM during LRLEs by designing a *multi-modal strategy*.

Considering the nature of the problem, the *interpretability* of the predictive model is a crucial demand since not only a reasonable estimation should be provided, but we are also interested in the most important factors that contribute to the popularity. Accordingly, *relevant features* should be identified through machine learning techniques.

The main elements and steps to predict the post popularity are presented in Procedure 1.

Procedure 1 Post Popularity Prediction

Input: $\mathcal{O}riginal\ \mathcal{D}ata$, $\mathcal{FS}\ param.$, $\mathcal{R}egressors$, $\mathcal{H}yperparameters$, k.
Output: $\mathcal{M}_r^\star$, $\mathcal{A}ccuracy_r$.

```
Sampled Data = SAMPLING(Original Data)
Data = FEATURE EXTRACTION(Sampled Data)
Prepared Data = PREPARE(Data)
(TR, TE) = SPLIT(Pprepared Data)
F* = FEATURE SELECTION(TR, FS param.)
for r in Regressors do
    for h in Hyperparameters do
        H* = k-FOLD-CV(k, F*, h,r)
    end for
    M*_r = BUILD MODEL( H*, F*,r)
    Accuracy_r = EVALUATE( M*_r, TE)
end for
```

4.4.1. Popularity Definition

In most of the research so far, SM's popularity is related to the amount of attention a post receives. Depending on the type of SM, this attention is quantified in different manners. In some of the studies on Instagram [61, 101, 102], a log-normalized form of popularity has been used considering the temporal aspects. In many studies such as [110], the popularity score is the ratio between the number of likes and the number of followers. In multiple research, the number of likes of a post (*likes count*) has been considered as the unnormalized popularity score. We follow the last definition of the popularity and ignore the temporal aspects of popularity since data is collected far enough from the event to assume that the number of likes of the posts have been reached to a stable number and would not be changed anymore. Since we expect to see more likes for the posts made by accounts with more followers, we ignore normalizing the likes according to the number of followers. Instead, we consider the followers count as a potential factor to find popularity.

4.4.2. Sampling

Due to the underlying challenges in many big data studies, it might be useful to perform sampling as a pre-processing step. In our data, the presence of noise and missing values, may cause poor model performance.

We sampled $5,583$ posts from the dataset to address these problems, which contain user-, post- and event-related attributes using Random Under Sampling [9], while preserving the target distribution using stratification on the number of likes. We extended the data to include visual content features, as popularity might also depend on the post's image.

After sampling, we split the dataset into training and test parts, we have implemented a custom random under-sampling by stratification on the likes count to have the distribution of the output the same in training and test set.

4.4.3. Feature Extraction

Applying a multi-modal approach requires considering post-related information, the content of the posts, and user characteristics. Moreover, some aspects of posts such as hashtags information and generation time are extracted and employed as potentially influential factors for building a predictive model depending on the type of SM. Besides, since the proposed method is to design a predictive model for a particular event, other features related to the event might correlate with the post's popularity. In general all the mentioned features fall into four main categories; *user-related*, *content-related*, *post-related* and *event-related*. Figure 13 reports the features hierarchy and details about the groups and sub-groups that we extracted in this work.

User-related features, also referred to as *social context* properties, gather information concerning the user who posted, such as their followers and followings count, profile type, *etc*. Depending on the target platform, they may be available as part of the provided meta-data embedded in the collected posts, or they may require further data collection actions to be obtained.

Content-related information also depends on the type of SM (*e.g.*, Video on YouTube and text on Twitter). On Instagram, image is the primary content, so this category should aggregate visual features in the images. Visual features can be low-level (basic) image-related features, mostly resulted from statistical data extracted from pixel-level operations, like image brightness or dominant color. Additionally, visual features include high-level properties that give information about the semantic of the post, such as the presence of particular objects or the image's topic. We extracted some of the visual features by using Microsoft Azure's Computer Vision services[15].

Post-related features are simply the meta-data of the post which have been provided through the API, such as the temporal and geographical attributes, hashtags, tags, number of comments, *etc*.

Event-related group of features is the statistical information, such as the average of the likes that each user obtained regarding the target event. It should be noted that, unlike some studies which require collecting the recent or even all the activities of each user from its profile, we consider only the posts of the same user as a representative of its activity in that context. As an example of a feature in this study, we included whether the post has been published before, during, or after the event period.

For a detailed explanation of the procedure followed for extracting the features, we refer the reader to Appendix B.

4.4.4. Feature Construction

To improve the performance of the predictive model, extra features can be created from the existing ones [77], because some features might be insignificant in terms of prediction capabilities, while if combined with other features, they could be highly correlated to the output. Since they are built based on existing features, we call them higher-level features. For example, in the case of LRLEs, while the post timestamp is one of the raw features, other attributes can be inferred using temporal aspects of the posts. We can extract meaningful attributes from it, such as whether the post is published in the weekdays or on the weekend, or other relevant encodings. These higher-level features can potentially make the model more interpretable by adding some semantics that makes sense. Regarding images, even though deep neural networks (DNNs) [39] can be applied to extract high-level features, another layer can still be added for interpretability purposes. For instance, given the number of faces in an image, we could add another feature encoding the number of male or female faces or the average age of people in that image. Other quantitative features can be added in different cases according to the context. To the best of our knowledge, this work is the first employing these *higher* level features in multi-modal approaches.

Besides some raw features cannot be directly inserted into the dataset, and need some modifications before using in post popularity predictive model, the following reports the preprocessing techniques used:

[15] https://azure.microsoft.com

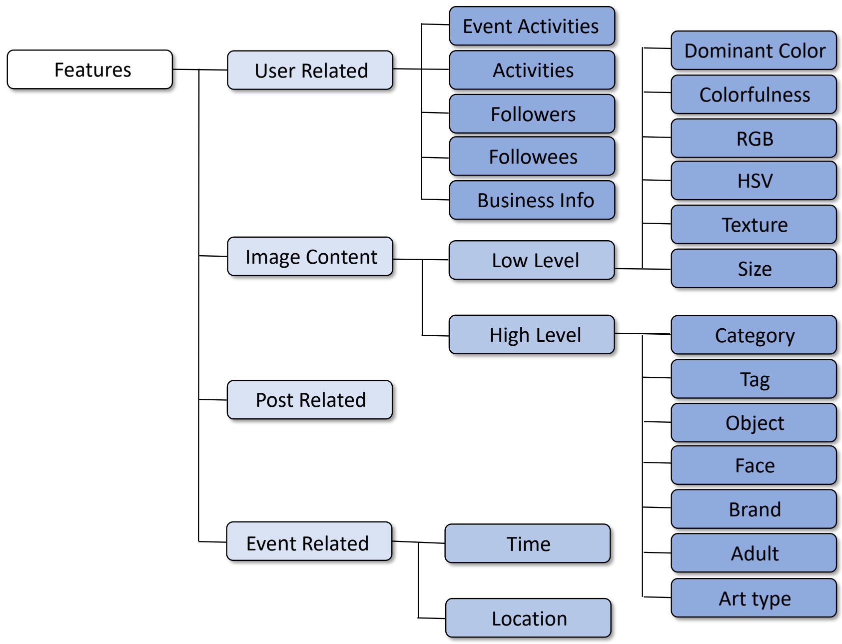


Figure 13: Hierarchical representation of the case study's features types.

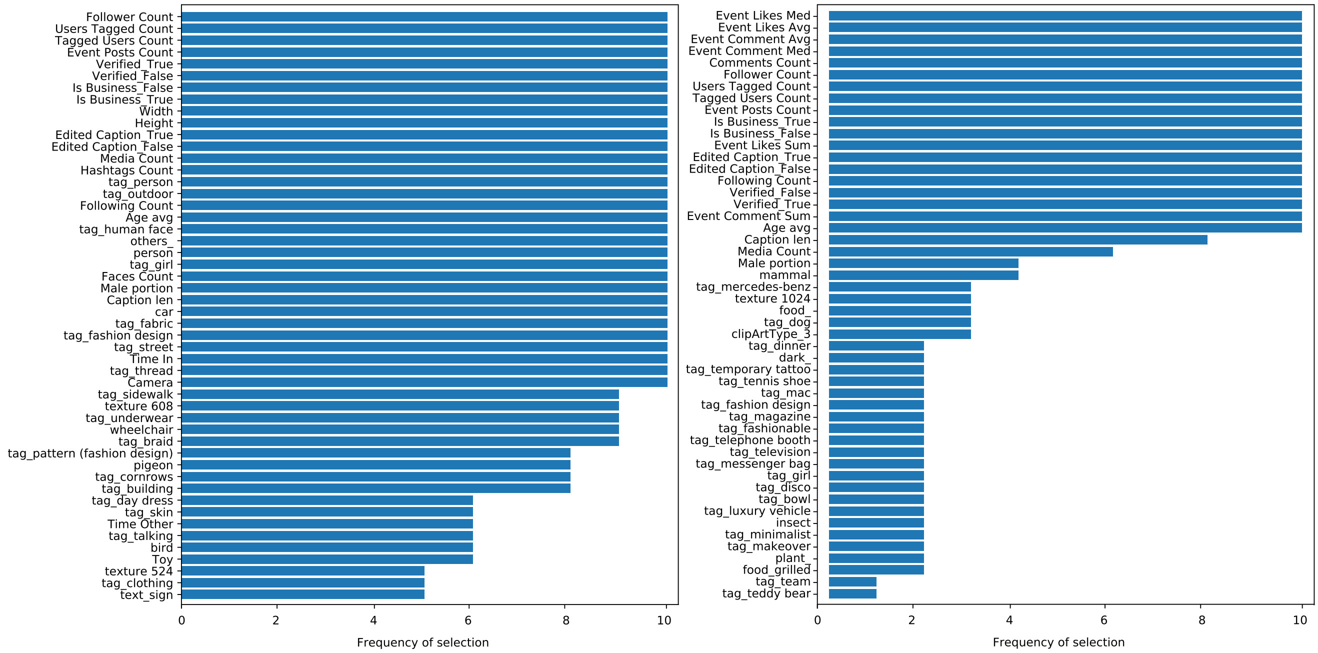
- **One-hot Encoding** for categorical variables has been employed to transform these kinds of features to numerical features for building the model. Using dummy variables are not preferable because they impose logical orders for different values of the feature. This process is applied to *edited caption, verified, isBusiness, clip art type, line drawing type* features.
- **Normalization** step is the action in which all the features in the dataset have been normalized using Python *sklearn* [82] standard scaler package to have zero mean and standard deviation equal to 1, to accelerate training of the regressors and hyper-parameters optimization.
- **Descritizing** Discretizing is another modification we applied for the high-level features with continuous value. For example, in case of high-level concepts obtained from images, like the presence of objects, faces, or brands, ML methods mostly provide a confidence level between 0 and 1, called score. In this case, the option is discretizing scores, to be either 0 or 1. Thus an optimal threshold must be found using cross-validation techniques. However, in some cases, such as the score of *being positive* of a text, one may choose to discretize scores or use the score as it is.

4.4.5. Feature Selection

Due to the abundant number of extracted features in the dataset, if the regressors fit all of them, it would be unnecessarily complicated and subject to overfitting. Accordingly, we have implemented two modes of feature selection (FS), both filter (*a.k.a.*, ranking or screening) methods, using Spearman's rank correlation and distance correlation (dCor) metric for evaluating the dependence of the *likes count* to a single or a subset of features. Spearman's rank correlation is a univariate metric formerly used in [61]; however, it only can measure the correlation of a single feature to the output, while dCor, previously applied for FS by [20, 48], is a multivariate index and evaluates the dependence of a subset of variables to the output. Then applying cross-validation on the training set and finding the mean-square error (MSE) in the regressors (Ridge [72], SVR [8], extreme gradient boosting (XGBoost) [25], and DNN [69]), we found 40 as the optimal number of features to be used in the final model.

Figure 14 shows the first 50 frequently selected features in 10 runs of two modes of FS on the training set. The top features are mainly user-related such as the ones related to a business profile and the number of followers and followings, which emphasizes the importance of influencers' network of connections on the popularity of the posts they publish, and indirectly implies the visibility of the users can have a significant impact on the popularity. Other features that have often been selected were among event-related and high-level image-related features, such as the presence of particular objects or human faces in the images and the semantic of the images like outdoor scenes. Among post-related features, the number of used hashtags has been selected more often.

In both FS modes, the average age of faces in the images, if any, that was added on top of the high-level image features is revealed in these plots that have been unanimously selected in all the runs. To our knowledge, there has been no effort to retrieve and add these kinds of features for studying post popularity in SM.



(a) dCor (b) SRC

Figure 14: Top 50 frequently selected features in 10 runs by FS phase of the proposed method using a) *dCor* and b) *SRC* indexes. The *y-axis* lists the top 50 features ordered by their frequency, while the *y-axis* reports the corresponding number of selection.

4.4.6. Base Model

After finding the most relevant features, we applied a simple regressor using all the features to obtain the base error estimate, on top of which we can evaluate the improvement of the performance of the model when FS and four supervised methods are applied. The detail about the implementation of the base method is provided in Appendix D. The best correlation of the predicted and actual likes count is 0.765 and 0.830 using Spearman and dCor in 10 runs of the base model.

4.4.7. Fit the Model/Hyper-parameters Tuning

For the further possible improvements on top of the base model, we considered only the selected features and applied four supervised methods, namely Ridge, SVR, XGBoost, and DNN. The reason for selecting these regressors is analyzing the data using different kinds of learning methods. Optimizing these complex learning models requires hyper-parameters tuning. Examples of hyper-parameters tuning algorithms are stochastic such as iterated racing[12] method, or random algorithms. Moreover, some of the sequential and greedy algorithms can be applied if the evaluation of the objective function is expensive [53, 54]. In this experiment, we considered the configuration space as a tree structure in which the leaf nodes represent a unique combination of hyper-parameters for each method. Then, we set up a grid search mechanism by which all the leaf nodes for all the regressors are tested [47, 66, 67]. We also included ridge regression to the grid search and ran the algorithm for 10 independent runs. In each run, the shuffling of the samples are the same for all the regressors, *i.e.*, the same samples are in the test set and training set for all the regressors in a single run, but we changed the seed for independent runs to increase the certainty of the results. For the evaluation of each combination, we applied k-fold cross-validation [35] where ($k = 5$) in which the initial training set is divided into train and validation parts so that we ensure all the training set samples are included in the validation part once. The model with the combination of the hyper-parameters is fit to train part for each k, and the performance metrics are measured for the validation part. Then the final performance is the average of k measurements. The grid search output for each regressor would be the best subset of parameters giving the best performance (the least RMSE) on the validation set among the other combinations (details are provided in Appendix E).

4.4.8. Results and Discussion

The best result is obtained by training a SVR model with linear kernel, $C = 2$ and $\epsilon = 0.9$. In this configuration mean absolute error (MAE) is equal to 14.895, and Spearman's rank correlation between the actual like count and the one predicted by the model is 0.952. Unfortunately, few studies were done on Instagram data, and as far as we know, none of them was done on a case study similar to ours. Most of the researches on the post popularity prediction were conducted either on Flickr or other SM platforms or on incomparably smaller scales in terms of the number of instances or users' profile types. However, in this work, our main criteria for collecting posts was their potential relations with four events held in different countries. Moreover, in most of the other works, the initial seeds were particular group(s)

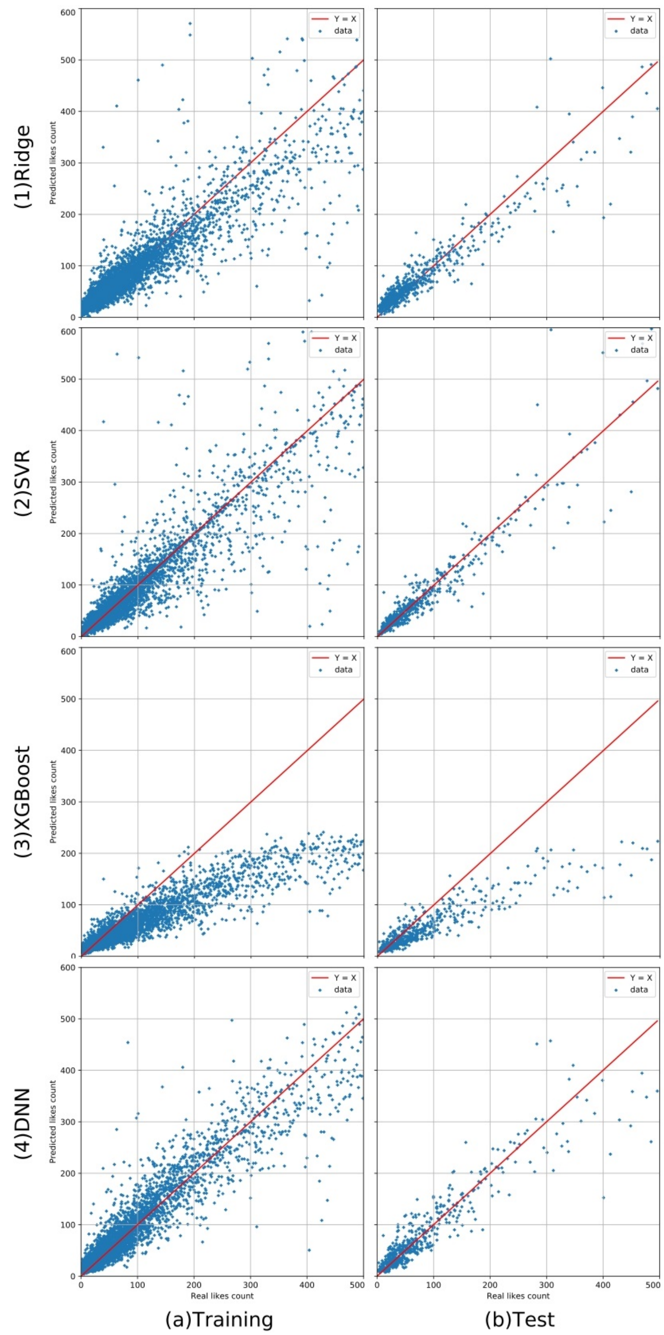


Figure 15: Predicted likes count (*y-axis*) vs. true likes count (*x-axis*) resulted from *(a)* training and *(b)* test datasets by Ridge, SVR, XGBoost and DNN considering top 50 features selected by the FS method using dCor index.

of users, based on which post data were collected, while regarding this case study, we first collected the posts, then we extracted the users who published them. Because of the reasons mentioned above, it was impossible to compare our results with others. The only metric which was provided in case of regression on popularity prediction was mostly the correlation among predicted value and true values of popularity, which we provided here using Spearman rank correlation and dCor. In all the regressors, the obtained correlation in different runs of the method has a very small standard deviation, suggesting that the correlation value is reliable.

Figure 15 depicts the true likes count versus the predicted values in applying the Ridge, SVR, XGBoost, and DNN regressors, which are built using tuned hyper-parameters. The bests results obtained for each regressor and their corresponding hyper-parameters during the test are presented in 3 (For the detailed information see Table 6 in Appendix E).

One of the most critical points uncovered by Figure 15 is that in none of the methods, overfitting has happened during the learning process, as the error rate in the training set is comparable to the test set, even in case of XGBoost regressor which is the worst in terms of predicting the out-

Table 3
Detailed information of the best hyper-parameters achieved during training models for Ridge, SVR, XGBoost and DNN regressors using top 50 features according to dCor index along with their corresponding best performance metrics on the test dataset.

Ridge	$\alpha = 0.1$					
		MAE	RMSE	MSE	Spearman	dCor
	best	18.893	31.030	962.883	0.914	0.939
SVR	*kernel* = linear, $C = 2$, $\varepsilon = 0.9$					
		MAE	RMSE	MSE	Spearman	dCor
	best	**14.895**	**30.851**	**951.794**	**0.941**	**0.952**
XGBoost	*learning rate* = 0.1, *reg lambda* = 1, *min child weight* = 1, *max depth* = 6					
		MAE	RMSE	MSE	Spearman	dCor
	best	32.543	59.102	3493.001	0.895	0.901
DNN	*learning rate* = 0.001, *batch size* = 512					
		MAE	RMSE	MSE	Spearman	dCor
	best	19.112	33.480	1120.896	0.908	0.930

put. As demonstrated in Figure 15 (3)XGBoost, the trained XGBoost model shows a tendency towards predicting popularity less than its true value and most of the data is below the line $y = x$; this could be an indication of insufficiency in grid search configuration space and to investigate more, other XGBoost hyper-parameters should be tuned. Moreover, the absence of any similar benchmark prevents from judgment about the overall performance of the method. For example, we are not sure whether the obtained value for RMSE is sufficiently good or not, considering the fact that the range of likes count is between 0 and 500, and the regression errors in samples having higher values would extensively deteriorate the results, even if just one sample, as a result of the absence of fair performance metric capable of dismantling this effect. Besides, the imbalanced distribution of the output would further degrade the results by introducing bias in the regressors' learning phase, and should be tackled by applying other sampling methods.

5. Conclusion

In this work, we provided a *high-level roadmap* for performing big data analysis on long-running live events (LRLEs). To put the proposed road map in practice, we chose the International Big Four Fashion Weeks events held in Milan, New York, Paris, and London as a case study, and we analyzed the respective posts on Instagram. We collected a large dataset containing about 1M relevant posts and 172K users who generated those posts and published it online [18] for the researchers for future research. To the best of our knowledge, it is the first dataset containing a complete duration of such combinations of events, and the diversity of the user profiles provides a better understanding of the majority of the participants. We performed a comprehensive quantitative and qualitative analysis on the case study. The analysis has been done such that it considers the interaction between the elements of such events. Temporal analysis of the posting behavior in each city indicates that the temporal dynamics of the events' responses have a direct relationship with the time of the events. We categorized posts based on if the hashtags they used in their posts are relevant to one or more events. This showed that there is a share of users that post on multiple events for increasing visibility and a share of "genuine" posts about one specific event at a time. These posts should be regarded as useful content for collecting knowledge and understanding the events. We extracted geographical distribution of the posts, and information about the following and followers' distribution of users, thus profiling users who participate in the discussion. Additionally, although the case study consists of international events from different countries, the online participants were mostly from those countries, and the events did not attract participants from other places in the world. We highlighted the most visible brands in the events according to the frequency of related hashtags and showed that brand dynamic is more affected according to the event locations rather than brand.

In order to enable transfer of knowledge for future events, we model and predicted the popularity of user-generated content during LRLEs using a multi-modal methodology. Besides, a wide variety of feature types were extracted hierarchically, including attributes related to posts, users, content, and event plus some extra semantic features obtained from high-level image properties. Unlike other studies in this context, instead of considering the group of feature types separately, we applied two feature selection methods. We found out that the semantic (higher level) type of features could be potentially useful in predicting popularity, which could be a possible future study in this context. We chose four types of regression methods, namely ridge, support vector regression, gradient tree boosting, and neural networks, and performed hyper-parameters tuning, utilizing a grid search mechanism. The final predictive models were evaluated using the test data by several performance metrics. Last but not least, by using feature selection methods in the procedure, the model interpretation would potentially facilitate understanding the potential factors concerning the problem of post popularity prediction and provides useful insights for the beneficiaries.

Future work concerns applying the proposed framework on other LRLEs from different context such as Comic-Con to evaluate the effectiveness of the framework. Moreover, motivated by the study on conversation graphs in online social media [16], we are designing *conversation agents* capable of participating in some discussions [93] during LRLEs. Such conversation agents would be beneficial for the LRLEs organizers to facilitate the customer relationship management [36, 78].

A. Users Social Network Analysis

Figure 16 plots the correlation between the number of followers and followings of each user. Each point corresponds to a single user, with the number of followings and followers (limited to 20,000 for a better comparison with Figure 6) on the x and y axes, respectively.

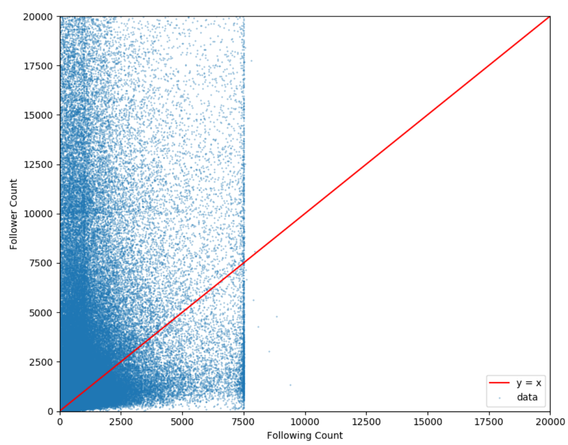


Figure 16: Users' number of followers (y-axis) vs. number of followings (x-axis) limited to 20 K.

B. Feature Extraction

In this study, we extracted information from different sources if they potentially influence popularity. User-related features include information about the user who published the posts and is directly accessible through the Instagram API. Additionally, we included a post-related category of features such as hashtags, tags, number of comments, *etc.* We also analyze factors correlated to the popularity of the posts in specific events, such as whether the post has been published before, during, or after the event period. Image-related features investigate post images from different aspects, which are listed in the following sections. We considered the higher-level image related attributes in the same category of high-level features. It should be noted that, *Dominant color* and *high-level* features have been extracted using Microsoft Azure's Computer Vision services[16].

[16] https://azure.microsoft.com

C. Image-related Features Extraction

Apart from the social context (user-related characteristics), the existing correlation between the popularity of a post and its content suggests that for improving the predictive capability of the model, these features, namely image-related or visual, should be considered. These attributes are categorized into high and low-level features, which are extracted as follows.

Low-level features are the features that can be acquired by pixel-level operations and present some statistical information about the image.

- *Colorfulness* score would be assigned to each image as the quantification of how wide is the range of colors in the image. The score is a real number between 0 and 1. The more an image has different colors, the more the score tends to 1. If the image is black and white, it would be considered 0. The score has been implemented, as suggested in [45].
- *Dominant color* feature is another type of low-level attribute that we added to the data which detects dominant color from a list containing 12 colors and after one-hot encoding adds 12 extra columns to the data.
- *HSV* channel consists of *hue*, *saturation* and *value*, and for each image the average of the pixels for the image channels have been computed as different features, which ranges between 0 and 1. Totally it adds 3 columns to the dataset.
- *RGB* is another set of channels for describing colors in images, in which channel *red*, *green* and *blue* represent the intensity of presence of these colors in each pixel. Like HSV, we added 3 columns to the data containing the average value of each channel throughout the image. The values are originally between 0 and 255, but we normalized them to be consistent with other added features.
- *Entropy* determines to what extent pixel values are similar. It quantifies the grayness of the image by providing a value between 0 and 1.
- *Texture* is another characteristic of the image describing gradient, which could impact the popularity. We included it because the human brain has shown different reactions to images with different textures, and many researchers have considered it a potential descriptive set of features [61]. Consequently, we implemented *Local Binary Patterns* (LBP) [79] which is a gray scale and rotation invariant texture descriptor and provides $1,024$ columns within the range of 0 and 1 to the dataset. After the images are converted to grayscale, each pixel's annular space in the obtained images is quantized into $1,022$ bins, and the spatial resolution of the LBP operator equals 8.

High-level features are human-perceivable properties of the images. The list of them is detailed as follows:

- *Category* feature distinguishes and categorizes the semantic topic(s) of the image among 77 existing taxonomies with parent/child hereditary hierarchy, by giving a score which is the probability (between 0 and 1) of belonging an image to the potential detected categories.
- *Moderate content* features detect racy and adult content in the images and provide two columns *Adult* and *Racy* as a confidence score with the values between 0 and 1.
- *Faces* related attributes determine the presence of faces in images and return some properties about the detected faces' genders and ages. In our approach, the provided faces features have been converted into four columns as follows. *Faces Count* presents the number of detected faces, *Age avg* calculates the average age of the detected faces, *Female Portion* and *Male Portion*, show the ratio of the female and male faces among all identified ones accordingly. It should be noted that these features were added on top of the original features provided by the service.
- *Tag* features identify potential tags as the subject of the image from a set of 1,620 tags, including objects, actions, and scenery. They provide a confidence score between 0 and 1, which is the probability of the image being related to the detected tags.
- *Brands* feature detects the presence of commercial global brands' logos with a confidence score between 0 and 1. In our approach, we added one column which contains the count of presented brands in the image, if any, zero otherwise.
- *Object* features provide a confidence score between 0 and 1 if they distinguish any objects from a set of 205 in the images.
- *Types* features provide two columns namely *line drawing type* and *clip art type*, the former is 1 if the image is line drawing, 0 otherwise. The later takes one of the integer values between 0 and 3, represented by non-clip-art, ambiguous, normal-clip-art, and good-clip-art, respectively.

D. Base Model

To build the base model, we applied the default implementation of the ridge offered in *sklearn* library in Python [82], and no feature selection method has been applied, so the number of features is 2,244. In the default setting, the ridge uses α hyper-parameter equal to 1, which imposes the strongest regularization. We ran the algorithm for 10 times within each, 5,024 samples are shuffled and separated with the ratio of 90:10 into the training and test part. The average of the runs are depicted in Figure 17 and Table 4.

Table 4
Detailed information about the base model's settings and results of its performance metrics on the training and test sets.

Model	Base model	Ridge regression		
	Parameter	default ($\alpha = 1$)		
	FS	no		
	Features no.	2244		
		Mean	std	best run
Training	MSE	1099.896	29.996	1130.917
	RMSE	33.161	0.452	33.629
	MAE	17.961	0.282	18.088
	Spearman	0.895	0.003	0.894
	dCor	0.933	0.001	0.933
		Mean	std	best run
Test	MSE	3324.35	484.508	2531.067
	RMSE	57.503	4.213	50.310
	MAE	33.358	1.463	31.412
	Spearman	0.755	0.022	**0.765**
	dCor	0.810	0.017	**0.830**

All the obtained metrics in training and test sets are averaged in 10 runs, and the mean and std of the results are reported in Table 4, as well as the best, run among all the runs according to the least value for MSE in test data. Note that the highest degree of regularization in the base model is applied to the features due to the default value of α hyper-parameter.

Table 4 includes the mean value, standard deviation of the obtained performance metrics along with the best run of the method. These values uncover the method's sensitivity to the data partitioning, *i.e.*, how performance might change with having different samples in training and test parts, which is also the indicator of how reliable the method is. For example, in the test set, the RMSE results standard deviation is 4.213, and the best run RMSE is about 7 units different from the mean value of the RMSE, which makes the model less reliable for unseen data since these values in the training set is much lower. On the other hand, comparing RMSE mean value in training and test set (57.503 and 33.161) suggests that the model is overfitted to the train part.

E. Regressors and Hyper-parameter Tuning

To implement SVR, XGBoost and DNN, we exploited *sklearn* [82], *xgboost* [26] Python packages respectively.

In case of DNN, we implemented two sequential architectures [33, 43, 41] which are summarized in Table 5. The results suggest that Architecture 2 performs much better than the first one, indicating that the excess layers in Architecture 1 adds unnecessary complexity to the model, which deteriorates the accuracy. In this regard, we utilized Architecture 2 and did not report the results of the other.

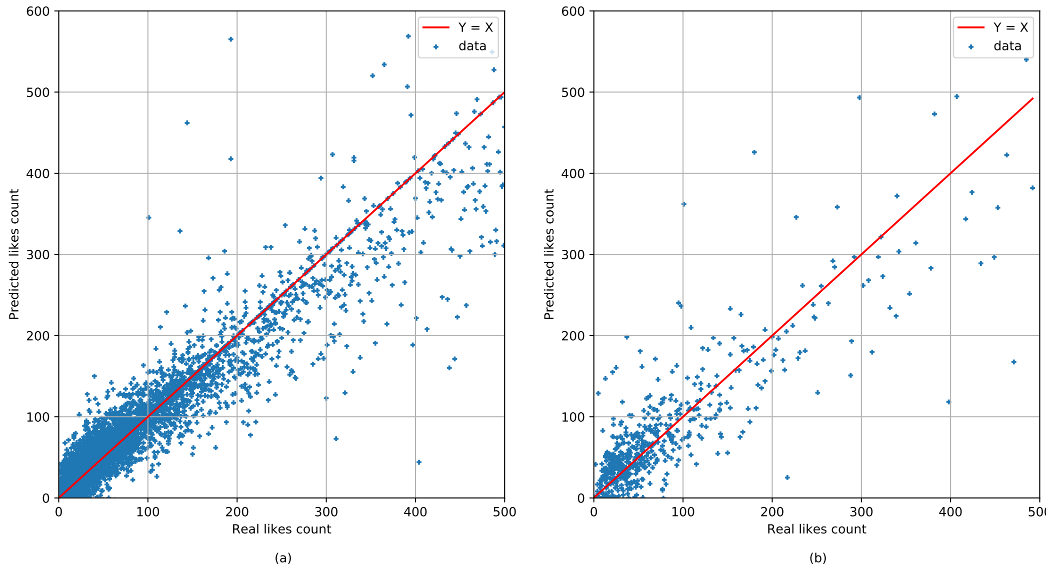


Figure 17: Predicted likes count by the base model (ridge regressor $\alpha = 1$) vs. actual likes count considering all the features. a) For the training and b) for the test sets, both resulted from the sampling phase of the proposed method to sample from the case study's dataset.

Table 5
Summary of the DNN architectures for the proposed method.

Arch.	Layer (type)	Output Shape	# Param
Arch. 1	input (Dense)	(None, 128)	5248
	hidden 1 (Dense)	(None, 256)	33024
	hidden 2 (Dense)	(None, 256)	65792
	hidden 3 (Dense)	(None, 256)	65792
	output (Dense)	(None, 1)	257
	Total params: 170,113		
	Trainable params: 170,113		
	Non-trainable params: 0		
Arch. 2	input (Dense)	(None, 128)	5248
	output (Dense)	(None, 1)	129
	Total params: 5,377		
	Trainable params: 5,377		
	Non-trainable params: 0		

Declaration of Competing Interest

The authors declare that there is no conflict of interest in all aspects of this manuscript preparation and data analysis.

CRediT authorship contribution statement

Alireza Javadian Sabet: Conceptualization, Methodology, Software, Validation, Investigation, Data Curation, Writing - Original Draft, Visualization. **Marco Brambilla:** Supervision, Funding acquisition, Conceptualization, Methodology, Validation, Investigation, Writing - Review & Editing. **Marjan Hosseini:** Software, Validation, Investigation, Data Curation, Writing - Original Draft, Visualization.

References

[1] Aaker, D.A., 2009. Managing brand equity. simon and schuster.

[2] Abid, O., Jamoussi, S., Ayed, Y.B., 2018. Deterministic models for opinion formation through communication: A survey. Online Social Networks and Media 6, 1–17. URL: https://www.sciencedirect.com/science/article/pii/S246869641730085X, doi:https://doi.org/10.1016/j.osnem.2018.02.002.

[3] Alexander, B., 2008. Web 2.0 and emergent multiliteracies. Theory into practice 47, 150–160.

[4] Almgren, K., Lee, J., et al., 2016. Predicting the future popularity of images on social networks, in: Proceedings of the The 3rd Multidisciplinary International Social Networks Conference on SocialInformatics 2016, Data Science 2016, ACM. p. 15.

[5] Arabghalizi, T., Rahdari, B., Brambilla, M., 2017. Analysis and knowledge extraction from event-related visual content on instagram, in: Proceedings of the 3rd International Workshop on Knowledge Discoveryon the WEB-KD-WEB, CEUR-WS. org. pp. 16–27.

[6] Araújo, C.S., Corrêa, L.P.D., da Silva, A.P.C., Prates, R.O., Meira, W., 2014. It is not just a picture: Revealing some user practices in instagram, in: 2014 9th Latin American Web Congress, IEEE. pp. 19–23.

[7] Arnaboldi, V., Conti, M., Passarella, A., Dunbar, R.I., 2017. Online social networks and information diffusion: The role of ego networks. Online Social Networks and Media 1, 44–55. URL: https://www.sciencedirect.com/science/article/pii/S2468696417300150, doi:https://doi.org/10.1016/j.osnem.2017.04.001.

[8] Awad, M., Khanna, R., 2015. Support Vector Regression. Apress, Berkeley, CA. pp. 67–80. URL: https://doi.org/10.1007/978-1-4302-5990-9_4, doi:10.1007/978-1-4302-5990-9_4.

[9] Batista, G.E., Prati, R.C., Monard, M.C., 2004. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD explorations newsletter 6, 20–29.

[10] Bennett, R., Rundle-Thiele, S., 2002. A comparison of attitudinal loyalty measurement approaches. Journal of Brand Management 9, 193–209.

[11] Bessis, N., Dobre, C., 2014. Big data and internet of things: a roadmap for smart environments. volume 546. Springer.

[12] Birattari, M., Yuan, Z., Balaprakash, P., Stützle, T., 2010. F-race and iterated f-race: An overview, in: Experimental methods for the analysis of optimization algorithms. Springer, pp. 311–336.

[13] Bird, S., 2006. Nltk: the natural language toolkit, in: Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, pp. 69–

Table 6
Detailed information presenting the hyper-parameters achieved during training models for Ridge, SVR, XGBosst, and DNN regressors using 50 first ranked features according to dCor index along with their corresponding performance metrics on the training and test datasets.

Ridge		$\alpha = 0.1$					
			MAE	RMSE	MSE	Spearman	dCor
	Training	mean	21.241	41.027	1683.603	0.910	0.915
		std	0.344	0.617	50.277	0.001	0.002
		best	21.653	41.802	1747.417	0.908	0.912
			MAE	RMSE	MSE	Spearman	dCor
	Test	mean	21.254	38.664	1522.367	0.906	0.915
		std	2.152	5.237	419.669	0.008	0.017
		best	18.893	31.030	962.883	0.914	0.939
SVR		*kernel* = linear, $C = 2$, $\varepsilon = 0.9$					
			MAE	RMSE	MSE	Spearman	dCorr
	Training	mean	17.962	44.712	1999.857	0.465	0.919
		std	0.198	0.849	75.274	0.000	0.002
		best	18.255	45.596	2078.974	0.465	0.917
			MAE	RMSE	MSE	Spearman	dCor
	Test	mean	18.200	40.265	1665.247	0.931	0.924
		std	2.173	6.632	542.022	0.006	0.016
		best	**14.895**	**30.851**	**951.794**	**0.941**	**0.952**
XGBoost		*learning rate* = 0.1, *reg lambda* = 1, *min child weight* = 1, *max depth* = 6					
			MAE	RMSE	MSE	Spearman	dCor
	Training	mean	32.281	58.975	3478.770	0.445	0.910
		std	0.358	0.880	104.544	0.001	0.009
		best	32.072	58.485	3420.529	0.445	0.916
			MAE	RMSE	MSE	Spearman	dCor
	Test	mean	34.133	61.875	3830.605	0.881	0.884
		std	1.059	1.461	180.252	0.010	0.016
		best	32.543	59.102	3493.001	0.895	0.901
DNN		*learning rate* = 0.001, *batch size* = 512					
			MAE	RMSE	MSE	Spearman	dCor
	Training	mean	16.375	28.334	813.203	0.926	0.942
		std	1.534	3.221	174.110	0.007	0.009
		best	17.723	30.620	937.577	0.924	0.935
			MAE	RMSE	MSE	Spearman	dCor
	Test	mean	21.619	38.345	1482.473	0.893	0.906
		std	1.645	3.482	270.069	0.012	0.016
		best	19.112	33.480	1120.896	0.908	0.930

72.

[14] Brambilla, M., Ceri, S., Daniel, F., Donetti, G., 2017. Spatial analysis of social media response to live events: The case of the milano fashion week, in: Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee. pp. 1457–1462.

[15] Brambilla, M., Gasparini, M., 2019. Brand community analysis on social networks using graph representation learning, in: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, ACM. pp. 2060–2069.

[16] Brambilla, M., Javadian, A., Sulistiawati, A.E., 2021a. Conversation graphs in online social media, in: Web Engineering. ICWE 2021, Springer International Publishing, Cham. pp. 97–112. URL: https://doi.org/10.1007/978-3-030-74296-6_8, doi:10.1007/978-3-030-74296-6_8.

[17] Brambilla, M., Javadian Sabet, A., 2020. Social media data about the big four fashion weeks - fall-winter 2018. URL: https://doi.org/10.7910/DVN/8BNXES, doi:10.7910/DVN/8BNXES.

[18] Brambilla, M., Javadian Sabet, A., Hosseini, M., 2021b. The role of social media in long-running live events: The case of the big four fashion weeks dataset. Data in Brief 35, 106840. URL: https://www.sciencedirect.com/science/article/pii/S2352340921001244, doi:https://doi.org/10.1016/j.dib.2021.106840.

[19] Brambilla, M., Javadian Sabet, A., Masciadri, A., 2019. Data-driven user profiling for smart ecosystems, in: Smart Living between Cultures and Practices. A design oriented perspective - ISBN: 978-88-7461-496-7. Mandragora, pp. 84–98.

[20] Brankovic, A., Hosseini, M., Piroddi, L., 2019. A distributed feature selection algorithm based on distance correlation with an application to microarrays. IEEE/ACM Transactions on Computational Biology and Bioinformatics 16, 1802–1815. doi:10.1109/TCBB.2018.2833482.

[21] Calisir, E., Brambilla, M., 2018. The problem of data cleaning for knowledge extraction from social media, in: International Conference on Web Engineering, Springer. pp. 115–125.

[22] Campana, M.G., Delmastro, F., 2017. Recommender systems for online and mobile social networks: A survey. Online Social Networks and Media 3-4, 75–97. URL: https://www.sciencedirect.com/science/article/pii/S2468696417300885, doi:https://doi.org/10.1016/j.osnem.2017.10.005.

[23] Chang, Y.W., Hsieh, C.J., Chang, K.W., Ringgaard, M., Lin, C.J., 2010. Training and testing low-degree polynomial data mappings via linear svm. Journal of Machine Learning Research 11.

[24] Chaudhuri, A., Holbrook, M.B., 2001. The chain of effects from brand trust and brand affect to brand performance: the role of brand loyalty. Journal of marketing 65, 81–93.

[25] Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22^{nd} acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

[26] Chollet, F., et al., 2015. Keras. https://keras.io.

[27] Coletto, M., Esuli, A., Lucchese, C., Muntean, C.I., Nardini, F.M., Perego, R., Renso, C., 2017. Perception of social phenomena through the multidimensional analysis of online social networks. Online Social Networks and Media 1, 14–32. URL: https://www.sciencedirect.com/science/article/pii/S246869641630009X, doi:https://doi.org/10.1016/j.osnem.2017.03.001.

[28] Cordeiro, M., Gama, J., 2016. Online social networks event detection: a survey, in: Solving Large Scale Learning Tasks. Challenges and Algorithms. Springer, pp. 1–41.

[29] De Vries, L., Gensler, S., Leeflang, P.S., 2012. Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. Journal of interactive marketing 26, 83–91.

[30] Dholakia, U.M., Durham, E., 2010. One café chain's facebook experiment. Harvard Business Review 88, 26.

[31] Dunteman, G.H., 1989. Principal components analysis. 69, Sage.

[32] Eke, C.I., Norman, A.A., Shuib, L., Nweke, H.F., 2019. A survey of user profiling: State-of-the-art, challenges, and solutions. IEEE Access 7, 144907–144924.

[33] Fausett, L.V., et al., 1994. Fundamentals of neural networks: architectures, algorithms, and applications. volume 3. prentice-Hall Englewood Cliffs.

[34] Franch, F., 2013. (wisdom of the crowds) 2: 2010 uk election prediction with social media. Journal of Information Technology & Politics 10, 57–71.

[35] Fushiki, T., 2011. Estimation of prediction error by using k-fold cross-validation. Statistics and Computing 21, 137–146.

[36] Galitsky, B., 2021. Adjusting chatbot conversation to user personality and mood, in: Artificial Intelligence for Customer Relationship Management. Springer, pp. 93–127.

[37] Gomez Rodriguez, M., Leskovec, J., Schölkopf, B., 2013. Structure and dynamics of information pathways in online media, in: Proceedings of the sixth ACM international conference on Web search and data mining, ACM. pp. 23–32.

[38] Gönen, M., Alpaydın, E., 2011. Multiple kernel learning algorithms. The Journal of Machine Learning Research 12, 2211–2268.

[39] Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep learning. volume 1. MIT press Cambridge.

[40] Göring, S., Brand, K., Raake, A., 2018. Extended features using machine learning techniques for photo liking prediction, in: 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), IEEE. pp. 1–6.

[41] Grossberg, S., 1988. Nonlinear neural networks: Principles, mechanisms, and architectures. Neural networks 1, 17–61.

[42] Guille, A., Hacid, H., Favre, C., Zighed, D.A., 2013. Information diffusion in online social networks: A survey. ACM Sigmod Record 42, 17–28.

[43] Hagan, M.T., Demuth, H.B., Beale, M.H., De Jesús, O., 1996. Neural network design. volume 20. Pws Pub. Boston.

[44] Hanna, R., Rohm, A., Crittenden, V.L., 2011. We're all connected: The power of the social media ecosystem. Business horizons 54, 265–273.

[45] Hasler, D., Suesstrunk, S.E., 2003. Measuring colorfulness in natural images, in: Human vision and electronic imaging VIII, International Society for Optics and Photonics. pp. 87–96.

[46] Hausman, A., Kabadayi, S., Price, K., 2014. Consumer–brand engagement on facebook: liking and commenting behaviors. Journal of research in interactive marketing .

[47] Hinton, G., 2010. A practical guide to training restricted boltzmann machines. Momentum 9, 1.

[48] Hosseini, M., 2018. Feature selection for microarray classification problems. Master's thesis. Politecnico di Milano. URL: http://hdl.handle.net/10589/142104.

[49] Howard, P.N., Duffy, A., Freelon, D., Hussain, M.M., Mari, W., Maziad, M., 2011. Opening closed regimes: what was the role of social media during the arab spring? Available at SSRN 2595096 .

[50] Hu, J., Yamasaki, T., Aizawa, K., 2016. Multimodal learning for image popularity prediction on social media, in: 2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), IEEE. pp. 1–2.

[51] Hu, Y., Manikonda, L., Kambhampati, S., 2014. What we instagram: A first analysis of instagram photo content and user types. Proceedings of the International AAAI Conference on Web and Social Media 8. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14578.

[52] Hughes, A.L., Palen, L., 2009. Twitter adoption and use in mass convergence and emergency events. International journal of emergency management 6, 248–260.

[53] Hutter, F., 2009. Automated configuration of algorithms for solving hard computational problems. Ph.D. thesis. University of British Columbia.

[54] Hutter, F., Hoos, H.H., Leyton-Brown, K., 2011. Sequential model-based optimization for general algorithm configuration, in: International Conference on Learning and Intelligent Optimization, Springer. pp. 507–523.

[55] Instagram, 2018. Instagram developer guide. https://www.instagram.com/developer/.

[56] Jaakonmäki, R., Müller, O., vom Brocke, J., 2017. The impact of content, context, and creator on user engagement in social media marketing, in: Proceedings of the 50^{th} Hawaii international conference on system sciences, pp. 1152–1160. URL: http://hdl.handle.net/10125/41289.

[57] Javadian Sabet, A., 2019. Social media posts popularity prediction during long-running live events. A case study on fashion week. Master's thesis. Politecnico di Milano. URL: http://hdl.handle.net/10589/150374.

[58] Kapferer, J.N., 2008. The new strategic brand management: Creating and sustaining brand equity long term. Kogan Page Publishers.

[59] Kayes, I., Kourtellis, N., Iamnitchi, A., 2017. The good, the bad and the deviant in community question answering. Online Social Networks and Media 2, 45–59. URL: https://www.sciencedirect.com/science/article/pii/S2468696417300289, doi:https://doi.org/10.1016/j.osnem.2017.08.002.

[60] Keller, K.L., Parameswaran, M., Jacob, I., 2011. Strategic brand management: Building, measuring, and managing brand equity. Pearson Education India.

[61] Khosla, A., Das Sarma, A., Hamid, R., 2014. What makes an image popular?, in: Proceedings of the 23^{rd} international conference on World wide web, ACM. pp. 867–876.

[62] Kim, A.J., Ko, E., 2010. Impacts of luxury fashion brand's social media marketing on customer relationship and purchase intention. Journal of Global Fashion Marketing 1, 164–171.

[63] Kim, A.J., Ko, E., 2012. Do social media marketing activities enhance customer equity? an empirical study of luxury fashion brand. Journal of Business research 65, 1480–1486.

[64] Kiros, R., Salakhutdinov, R., Zemel, R.S., 2014. Unifying visual-semantic embeddings with multimodal neural language models. arXiv preprint arXiv:1411.2539 .

[65] Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering 160, 3–24.

[66] Larochelle, H., Erhan, D., Courville, A., Bergstra, J., Bengio, Y., 2007. An empirical evaluation of deep architectures on problems with many factors of variation, in: Proceedings of the 24th international conference on Machine learning, ACM. pp. 473–480.

[67] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324.

[68] Liu, C., Zhan, X.X., Zhang, Z.K., Sun, G.Q., Hui, P.M., 2015. How events determine spreading patterns: information transmission via internal and external influences on social networks. New Journal of Physics 17, 113045.

[69] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. Neurocomputing 234, 11–26.

[70] Manca, M., Boratto, L., Morell Roman, V., Martori i Gallissà, O., Kaltenbrunner, A., 2017. Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. Online Social Networks and Media 1, 56–69. URL: https://www.sciencedirect.com/science/article/pii/S2468696416300088, doi:https://doi.org/10.1016/j.osnem.2017.04.002.

[71] Manikonda, L., Hu, Y., Kambhampati, S., 2014. Analyzing user activities, demographics, social network structure and user-generated content on instagram. arXiv preprint arXiv:1410.8099 .

[72] Marquardt, D.W., Snee, R.D., 1975. Ridge regression in practice. The American Statistician 29, 3–20.

[73] Mazloom, M., Hendriks, B., Worring, M., 2017. Multimodal context-aware recommender for post popularity prediction in social media, in: Proceedings of the on Thematic Workshops of ACM Multimedia 2017, ACM. pp. 236–244.

[74] Meghawat, M., Yadav, S., Mahata, D., Yin, Y., Shah, R.R., Zimmermann, R., 2018. A multimodal approach to predict social media popularity, in: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE. pp. 190–195.

[75] Mezghani, M., Zayani, C.A., Amous, I., Gargouri, F., 2012. A user profile modelling using social annotations: A survey, in: Proceedings of the 21st International Conference on World Wide Web, Association for Computing Machinery, New York, NY, USA. p. 969–976. URL: https://doi.org/10.1145/2187980.2188230, doi:10.1145/2187980.2188230.

[76] Moniz, N., Torgo, L., 2019. A review on web content popularity prediction: Issues and open challenges. Online Social Networks and Media 12, 1–20.

[77] Mueller, J.P., Massaron, L., 2016. Machine learning for dummies. John Wiley & Sons.

[78] O'Brien, M., Dyché, J., et al., 2002. The CRM handbook: A business guide to customer relationship management. Addison-Wesley Professional.

[79] Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis & Machine Intelligence , 971–987.

[80] Oliver Richard, L., 1997. Satisfaction: A behavioral perspective on the consumer. New York ' NY: Irwin-McGraw-Hill .

[81] Panagiotou, N., Zygouras, N., Katakis, I., Gunopulos, D., Zacheilas, N., Boutsis, I., Kalogeraki, V., Lynch, S., O'Brien, B., 2016. Intelligent urban data monitoring for smart cities, in: Joint European conference on machine learning and knowledge discovery in databases, Springer. pp. 177–192.

[82] 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. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

[83] Picazo-Vela, S., Gutiérrez-Martínez, I., Luna-Reyes, L.F., 2012. Understanding risks, benefits, and strategic alternatives of social media applications in the public sector. Government information quarterly 29, 504–511.

[84] Pierri, F., Pavanetto, S., Brambilla, M., Ceri, S., 2021. Vaccinitaly: monitoring italian conversations around vaccines on twitter. arXiv preprint arXiv:2101.03757 .

[85] Psyllidis, A., Bozzon, A., Bocconi, S., Bolivar, C.T., 2015. A platform for urban analytics and semantic data integration in city planning, in: International conference on computer-aided architectural design futures, Springer. pp. 21–36.

[86] Qualman, E., 2010. Socialnomics: How social media transforms the way we live and do business. John Wiley & Sons.

[87] Rahdari, B., Arabghalizi, T., Brambilla, M., 2017. Analysis of online user behaviour for art and culture events, in: International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Springer. pp. 219–236.

[88] Ramponi, G., Brambilla, M., Ceri, S., Daniel, F., Di Giovanni, M., 2019. Vocabulary-based community detection and characterization, in: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, ACM. pp. 1043–1050.

[89] Rendle, S., 2010. Factorization machines, in: 2010 IEEE International Conference on Data Mining, IEEE. pp. 995–1000.

[90] Riquelme, F., González-Cantergiani, P., 2016. Measuring user influence on twitter: A survey. Information Processing & Management 52, 949–975. URL: https://www.sciencedirect.com/science/article/pii/S0306457316300589, doi:https://doi.org/10.1016/j.ipm.2016.04.003.

[91] 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, pp. 851–860.

[92] Saravanou, A., Valkanas, G., Gunopulos, D., Andrienko, G., 2015. Twitter floods when it rains: a case study of the uk floods in early 2014, in: Proceedings of the 24th International Conference on World Wide Web, pp. 1233–1238.

[93] Scotti, V., Tedesco, R., Sbattella, L., 2021. A modular data-driven architecture for empathetic conversational agents, in: 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 365–368. doi:10.1109/BigComp51126.2021.00080.

[94] Sedgwick, P., 2014. Spearman's rank correlation coefficient. Bmj 349.

[95] Tatar, A., De Amorim, M.D., Fdida, S., Antoniadis, P., 2014. A survey on predicting the popularity of web content. Journal of Internet Services and Applications 5, 1–20.

[96] Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., Li, L.J., 2015. The new data and new challenges in multimedia research. arXiv preprint arXiv:1503.01817 1.

[97] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.

[98] Totti, L.C., Costa, F.A., Avila, S., Valle, E., Meira Jr, W., Almeida, V., 2014. The impact of visual attributes on online image diffusion, in: Proceedings of the 2014 ACM conference on Web science, ACM. pp. 42–51.

[99] Troncy, R., Malocha, B., Fialho, A.T.S., 2010. Linking events with media, in: Proceedings of the 6th International Conference on Semantic Systems, Association for Computing Machinery, New York, NY, USA. pp. 1–4. URL: https://doi.org/10.1145/1839707.1839759, doi:10.1145/1839707.1839759.

[100] Trusov, M., Bucklin, R.E., Pauwels, K., 2009. Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. Journal of marketing 73, 90–102.

[101] Wu, B., Cheng, W.H., Zhang, Y., Huang, Q., Li, J., Mei, T., 2017. Sequential prediction of social media popularity with deep temporal context networks. arXiv preprint arXiv:1712.04443 .

[102] Wu, B., Mei, T., Cheng, W.H., Zhang, Y., 2016. Unfolding temporal dynamics: Predicting social media popularity using multi-scale temporal decomposition, in: Thirtieth AAAI Conference on Artificial Intelligence, pp. 272–278.

[103] Xiao, S., Wei, C.P., Dong, M., 2016. Crowd intelligence: Analyzing online product reviews for preference measurement. Information &

Management 53, 169–182.

[104] Xu, G., Wu, Z., Zhang, Y., Cao, J., 2015. Social networking meets recommender systems: survey. International Journal of Social Network Mining 2, 64–100.

[105] Yamaguchi, K., Berg, T.L., Ortiz, L.E., 2014. Chic or social: Visual popularity analysis in online fashion networks, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM. pp. 773–776.

[106] Ying, Q.F., Chiu, D.M., Venkatramanan, S., Zhang, X., 2018. User modeling and usage profiling based on temporal posting behavior in osns. Online Social Networks and Media 8, 32–41. URL: https://www.sciencedirect.com/science/article/pii/S2468696418300636, doi:https://doi.org/10.1016/j.osnem.2018.10.003.

[107] Zhang, C., Ma, Y., 2012. Ensemble machine learning: methods and applications. Springer.

[108] Zhang, W., Wang, W., Wang, J., Zha, H., 2018. User-guided hierarchical attention network for multi-modal social image popularity prediction, in: Proceedings of the 2018 World Wide Web Conference on World Wide Web, International World Wide Web Conferences Steering Committee. pp. 1277–1286.

[109] Zikopoulos, P., Deroos, D., Parasuraman, K., Deutsch, T., Giles, J., Corrigan, D., 2012. Harness the power of big data The IBM big data platform. McGraw Hill Professional.

[110] Zohourian, A., Sajedi, H., Yavary, A., 2018. Popularity prediction of images and videos on instagram, in: 2018 4th International Conference on Web Research (ICWR), IEEE. pp. 111–117.

Alireza Javadian Sabet received his master's in Computer Science and Engineering from Politecnico di Milano in 2019. He worked on analyzing social media as his master's thesis and joined Data Science Lab at Politecnico di Milano (POLIMI) afterwards to continue his research. In 2019, he also joined DEpendable Evolvable Pervasive Software Engineering (DEEP-SE) group at POLIMI as a research fellow, and has been studying transportation systems in Europe. His research interests lie in Knowledge Extraction from Social Networks, Business Analysis and Context-aware Recommender Systems.

Marco Brambilla is full professor of Web Science and Software Engineering at Politecnico di Milano, Italy. He leads the Data Science Lab at Politecnico di Milano, DEIB. His current research interests are on Web Science, Big Data Analysis, Social Media Analytics, and Model-driven Development. He is the inventor of the Interaction Flow Modeling Language (IFML) standard by the OMG, and of 2 patents on crowdsourcing and multi-domain search. He started-up Fluxedo and WebRatio.

Marjan Hosseini received the bachelor's degree in Computer Engineering from Shomal University, Iran, and her master's in Computer Science and Engineering from Polytechnic University of Milan in 2018. Afterwards, she joined DEEP-SE Lab in Polytechnic University of Milan and worked on NLP and Machine Learning. Currently, she is working towards her PhD in University of Connecticut in Computer Science and Engineering. Her research interests lie in Bayesian and Probabilistic methods in Machine Learning and Bioinformatics.