Disentangling the Information Flood on OSNs: Finding Notable Posts and Topics

Disentangling the Information Flood on OSNs: Finding Notable Posts and Topics

Paola Caso*, Martino Trevisan†, and Luca Vassio‡
*‡Politecnico di Torino, Italy †University of Trieste, Italy
Email: *paola.caso@studenti.polito.it, †martino.trevisan@dia.units.it, ‡luca.vassio@polito.it,

***Abstract*—Online Social Networks (OSNs) are an integral part of modern life for sharing thoughts, stories, and news. An ecosystem of influencers generates a flood of content in the form of posts, some of which have an unusually high level of engagement with the influencer's fan base. These posts relate to blossoming topics of discussion that generate particular interest among users: The COVID-19 pandemic is a prominent example. Studying these phenomena provides an understanding of the OSN landscape and requires appropriate methods. This paper presents a methodology to discover notable posts and group them according to their related topic. By combining anomaly detection, graph modelling and community detection techniques, we pinpoint salient events automatically, with the ability to tune the amount of them. We showcase our approach using a large Instagram dataset and extract some notable weekly topics that gained momentum from 1.4 million posts. We then illustrate some use cases ranging from the COVID-19 outbreak to sporting events.**

***Index Terms*—Online Social Networks; Engagement; Anomaly Detection; Graph Modeling; Community Detection; Instagram.**

I. INTRODUCTION

Online Social Networks (OSNs) are the preferred forum for disseminating information, counting billions of users worldwide. In OSNs, users can interact with each other and share their ideas, experiences, and news in the form of text, image, video, and audio content: Users are not only consumers of content, but also its creators. In OSNs, users can follow other profiles and thus subscribe to each other's posts. Some profiles can achieve a considerable fan base, sometimes reaching millions of *followers*. These profiles, often referred to as *influencers*, include not only celebrities such as musicians or athletes, but also profiles of personalities famous because of the specific content they offer on OSNs. This is the case for fashion bloggers, food bloggers or lifestyle influencers.

Influencers discuss and produce content on a variety of topics, and their posts attract users who respond by commenting or *liking* them. The level of engagement with a post depends on various factors, such as the post creator and content. Some posts may have a particularly high level of engagement compared to what might be expected, given the influencer's recent history. We can define them as *notable posts*. The high engagement of notable posts can be due to various factors. The post may be an isolated outlier, as is the case with posts about memorable moments in an influencer's life (e.g., the birth of a child). In other cases, notable posts target topics that suddenly gain prominence in OSNs, such as important events (e.g., a natural disaster or a war). The recent COVID outbreak is a textbook case, where the debate on OSNs suddenly polarized around the pandemic and the resulting social distancing and lockdown measures. We refer to these as *notable topics*, i.e., topics whose posts have an unusually high level of engagement. The discovery of notable posts and topics is critical to the study of debate in OSN, as it allows us to identify new events, trends, and social phenomena. It can also support marketing strategies that are increasingly based on the study of engagement factors and mechanisms. Recent literature has shown how new topics of discussion in OSNs can suddenly blossom and disappear. The COVID-19 outbreak is known to have caused a social infodemic [1], [2]. Other examples include revolutionary waves [3], [4] or natural disasters [5], [6]. In this literature, the volume of content on certain topics is measured by the number of posts, likes, or comments, usually using a list of keywords or requiring other manual intervention. Thus, there is a need for automated methods to detect notable posts and topics, while manual analysis of OSN posts falls short due to the information flood to be analyzed.

In this work, we propose a methodology for automatically detecting notable posts and topics. It is based on a combination of techniques for data processing and consists of four steps. First, we model the expected engagement (in terms of likes or comments) that posts from different influencers usually receive. Second, we extract notable posts using anomaly detection techniques, with the goal of filtering out uninteresting or ordinary posts. Third, we model notable posts as a *network* in which posts are connected if they share one or more topics that we identify based on the hashtags. Finally, we obtain the notable topics by extracting *communities* of posts from graphs using the Louvain method. Our approach allows us to tune the number of notable posts, and thus of identified notable topics.

We showcase our methodology on a large dataset from Instagram containing ≈1.4 million posts from 1 400 influencers over a period of 5 years. Our results show that our methodology successfully identifies notable topics during these years by detecting posts about the COVID-19 outbreak, TV shows and sports events. Although our methodology combines established anomaly and community detection techniques, our results show that our approach is suitable for disentangling the complex ecosystem of OSNs. To allow the improvement of our methodology and its reuse on other datasets, we make our code available online[1].

[1]https://github.com/SmartData-Polito/notable-posts-topics-OSN

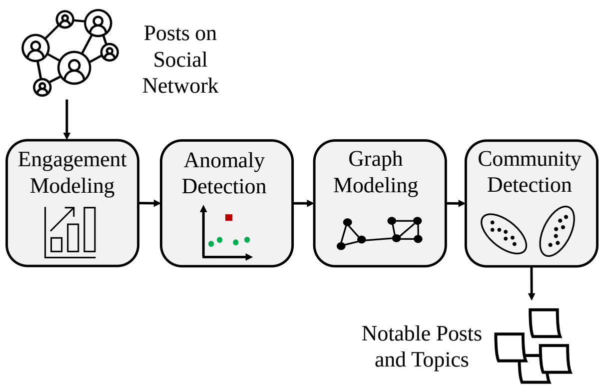


Fig. 1: Overview of the building blocks of our methodology.

II. METHODOLOGY

In this section, we illustrate our methodology. Our goal is twofold: i) we aim at identifying notable posts that obtain an unexpected high number of reactions, and ii) we want to group those posts to obtain notable topics referring to a particular event. We expect to operate on OSN data but do not tie our methodology to any particular OSN. We suppose to consider a set of *profiles* that produce *posts*. Profiles have *followers*, i.e., other profiles that subscribe to visualize in their feed a profile's posts. Each post consists in a textual and multimedia content, and it receives *reactions* (e.g., comments or likes) by the creator's followers.

Our methodology operates in near real time. At every step (e.g., one hour, one day or one week) it detects notable posts and topics by modeling the expected engagement based on the past history. Our approach consists on a four-step process, sketched in Figure 1. We implemented all code in Python, using the Pandas and NetworkX libraries.[1]

A. Engagement Modeling

The first step of our methodology is the modeling of the expected engagement of posts. We define the engagement as the total number of reactions (likes, comments or other metrics, depending on the OSN) obtained by a post. As the number of reactions of a post increase over time, we sample them at a fixed interval, which is 24 hours after post creation in our experiments.

To identify the number of likes and comments the post is expected to receive, we take the last 100 posts from a given profile and consider the reactions they obtained. In case the profile created less than 100 posts, we consider all of them. We drop the top and bottom 25% of those 100 posts in terms of reactions and compute the average number of reactions of the middle 50% of the posts. We consider this quantity as the *expected* number of reactions a post would get. In literature, this is the classical interquartile mean (or 25% trimmed mean). The main advantage of the trimmed mean is robustness and higher efficiency for skewed distributions like the ones we are studying (see Section III): the trimmed mean is less sensitive to outliers than the mean but it will still give a reasonable estimate of central tendency or mean. When a new post of the profile is published, we consider the number of reactions it obtains and compare it to the expected value.[2] In this way, we define the *Engagement Score* as:

$$\text{Engagement Score} = \frac{\text{Post Reactions}}{\text{Expected Reactions}}$$

A score greater than 1 indicates that the post is performing better than usual for a post of the given profile. A score below 1 means it is faring worse than expected.

B. Anomaly Detection

Following the definition of Engagement Score, we are now interested in discovering notable posts. These *overperforming* posts have a high Engagement Score, i.e., they receive more reactions than expected. Thus, we want to detect posts whose score deviates significantly from the scores normally received by the profile that created the post, i.e., the outliers.

To detect notable posts, we analyze the time series of engagement scores of the last 100 posts for a given profile.[3] Indeed, we want to detect anomalies for each individual time series, therefore with respect to the context of each profile's engagement score history. To detect notable posts, we use the Boxplot Rule method [7], [8], which is widely used in the anomaly detection field, based on verifying that the tested values exceed a certain threshold. The upper limit is calculated using the following formula:

$$\text{Upper Limit} = Q_3 + \alpha(Q_3 - Q_1)$$

where Q_1 and Q_3 are, respectively, the first and third quartile and $(Q_3 - Q_1)$ is the Interquartile Range. Notice that we are interested in notable posts, that received an abnormally high number of reactions. Thus we are not interested in the lower limit, that with the Boxplot Rule is defined as $Q_1 - \alpha \cdot (Q_3 - Q_1)$. Using the above formula, we define notable posts those having an Engagement Score greater than or equal to Upper Limit. As our methodology works in near real-time, the Upper Limit must be dynamically computed using the 100 most recent posts of the given profile. The constant α is a parameter of the method and is typically set to 1.5. Here, we prefer not restricting to a specif value, rather we use α to regulate the number of notable posts to be passed to the next steps.

We are aware of more advanced anomaly detection techniques [7], [8]. Indeed, we also tested on our dataset the z-score method, ARIMA, and Isolation Forest. Results, not reported here for lack of space, are available in our technical report [9]. Here, we restrict to the results of the Boxplot Rule, given it is a simple method that obtained slightly better results based on modularity performance (see Section II-D).

[2] Again we sample reactions after the same fixed amount of time.

[3] If the profile created fewer posts, we consider them all.

C. Graph Modeling

We now define a network or *graph* of notable posts for each time step. This block is instrumental to later group them into notable topics. In our graphs, each notable post forms a node, and we desire to have an edge between two posts if they refer to a similar topic. Although there exist a wide corpus of techniques to extract topics from text (and images), in the context of OSNs, these approaches struggle to cope with short texts and sarcasm. In this work, we opt to use *hashtags* as a practical instrument to define the topic(s) of a post. Almost all OSNs nowadays allow users to include one or more hashtags when creating posts in the form of of a short string prefaced by the hash (#). Hashtags are a form of user-generated tagging used to indicate the content of a post so that it can be also found by other users.In OSNs, there is no check of coherence between a post's hashtags and actual content, thus they can potentially mismatch. As future work, we plan to integrate the hashtags with other techniques for topic modeling, customized to work on OSNs, to possibly fill this gap.

We build a graph for each time step, where each node represents a post, connected to other posts by a weighted edge if it has at least one hashtag in common with those posts. For the computation of the weight of the arcs, we opt to use the metric based on the Jaccard Index similarity measure. Each pair of nodes (posts) is characterized by the sets of hashtags s_1 and s_2, and their Jaccard Index similarity is defined as the ratio between the size of intersection and the size of the union of the two sets:

$$\text{Jaccard Index} = \frac{|s_1 \cap s_2|}{|s_1 \cup s_2|}.$$

Thus, two posts sharing a high fraction of their hashtags result in having a high index. Two posts without hashtags in common do not even form an edge.

D. Community Detection

In the last step, we group similar notable posts using the defined network graph. The rationale is that our similarity metric allows us to find groups of posts related to the same notable topic. Since we are working with a network that connects posts with one or more common hashtags, we need a community detection algorithm to find subgraphs with densely connected nodes.

To extract communities, we use the widely used Louvain algorithm [10], [11]. The goal of the Louvain algorithm is to maximize the modularity of communities, where modularity quantifies the quality of the assignments of nodes to communities. Intuitively, the modularity captures how densely connected nodes are within a community compared to how connected they would be in a random network with the same degree sequence. Modularity is defined in the range of -0.5 to +1, and modularity values of 0.5 or higher are considered strong evidence of well-formed communities. The Louvain method operates by finding small communities optimizing modularity locally at all nodes. Then, each small community is merged into a meta-vertex and the first step is repeated.

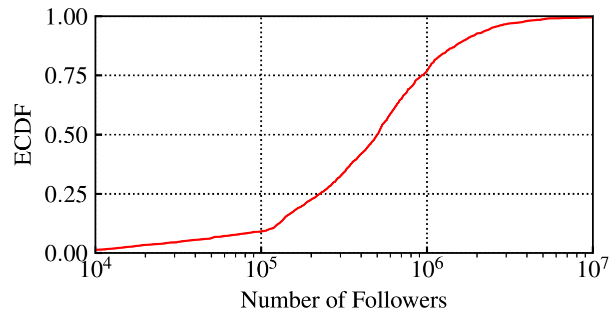


Fig. 2: Distribution of followers per profile in December 2020. Notice the log scale on the x axis.

The final result of our methodology is, for each time step, the list of notable posts grouped into communities. Each community represents a notable topic that we identify using the hashtags related to the posts in the community.

III. Dataset

We evaluate our methodology using a dataset of posts obtained from the Instagram social network. To obtain the posts we rely to the CrowdTangle tool and its API[4]. CrowdTangle is a content discovery and social analytics tool owned by Meta, which is open to researchers and analysts worldwide to support research, upon having a partnership agreement.

We selected the posts of the top ranked Italian profiles according to the "Influencer Italia" ranking website[5] over the course of 5 years, from January 1^{st}, 2016 to December 31^{st}, 2020, covering 261 weeks (time frames). The list of profiles includes politicians, athletes, musicians and Instagram influencers (food bloggers, travel bloggers, etc.). The list is publicly available online on our repository.[1] After removing erroneous or incomplete data, our cleaned dataset contains a total of 1 400 influencers and 1 400 697 posts. Various metadata is available for each post, including the text (with possible hashtags) and the number of reactions (likes and comments) it received in different time steps. For our analyses, we always look at the number of reactions the post received one day after it was created (24 hours).

The profiles included in our dataset have different levels of activity. Some create new posts daily or even several times a day, while others, interestingly, post only a few times over the entire 5-year dataset. The latter case is especially true for celebrities who attract many followers without posting frequently. Profiles in the dataset created a median of 46 posts. Those who created more than 100 (1000) posts are 31.7% (2.5%). The profiles also have different fan bases, although they are taken from the top Italian ranked in terms of follower count. In Figure 2 we show the distribution of followers per profile (sampled on the last day of our dataset) in the form of an Empirical Cumulative Distribution Function

[4]https://github.com/CrowdTangle/API
[5]https://www.influenceritalia.it

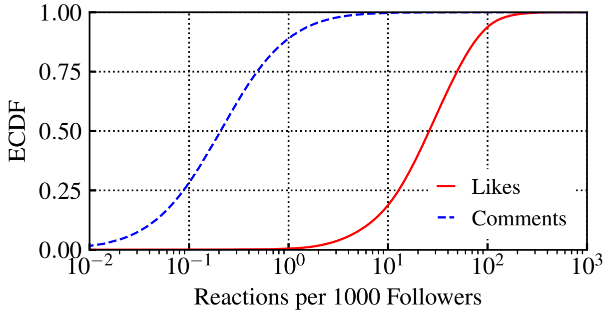


Fig. 3: Distribution of post reactions per 1000 followers. Notice the log scale on the x axis.

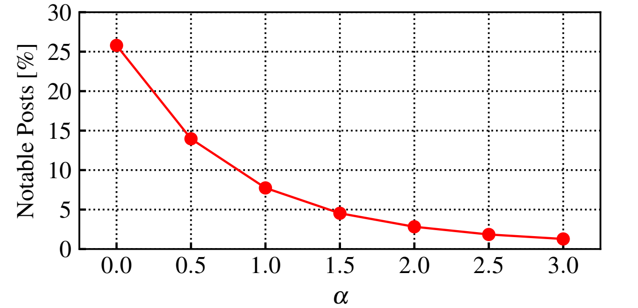
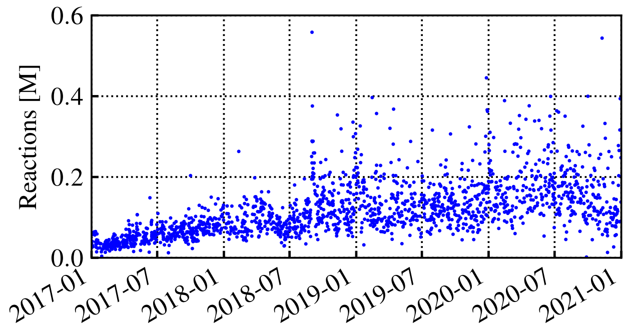


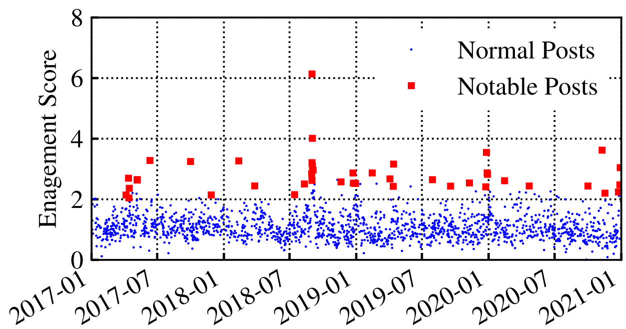
Fig. 4: Fraction of notable posts with different values of α for the boxplot rule for anomaly detection.

(ECDF). All profiles have more than 10 k followers and 91% have more than 100 k. In median, a profile in our dataset has 502 k followers. According to the common classification of influencers [12], profiles with more than 1 million followers are called *mega influencers* and account for 23% of the profiles in our dataset.

The number of a profile's followers clearly has a decisive impact on the number of reactions her posts receive. The larger the fan base, the more followers will see posts in their feed. On Instagram, profiles can react to posts by liking or commenting on them. In Figure 3, we show the ECDF of the number of *likes* and comments of posts in our dataset, normalized over the number of followers of the creator (at the time the post was created). We first note that *likes* are two orders of magnitude more common than comments. Posts typically receive between 10 and 100 likes per 1000 followers, while they receive more than one comment only in 11.3% of cases. In median, posts receive 25 (0.2) likes (comments) per 1000 followers. Interestingly, 0.9% of posts receive more than 200 *likes* per 1000 followers, more than one fifth of the profile's fan base. We found that these distributions are well fit by a log-normal distribution. Indeed, comparing the empirical distribution with the log-normal fit, we obtain a very small Kolmogorov distance. Specifically, we obtain 0.035 and 0.013, respectively for likes and comments (with parameters of the log-normal μ = 3.17, σ = 1.04 for likes and μ = -1.59, σ = 1.33 for comments). In the remainder of the paper, we are interested in notable posts, that are those at the right tails of these distributions.



(a) Reactions (in millions). Each blue dot represents a post.



(b) Engagement Score. Notable posts are marked in red.

Fig. 5: Reactions and Engagement Score for the Italian influencer Valentina Ferragni.

IV. Results

In this section, we present the results obtained with our methodology on the Instagram dataset described in the previous section. First, we discuss the Engagement Score and the resulting notable posts. Then, we show the community detection results and some examples of the notable topics we found.

A. Engagement Score and Notable Posts

We first illustrate the effects of the α parameter for the anomaly detection technique. It represents the threshold above which a post is notable, and its value thus allows us to regulate the number of notable posts. With a high value of α, we are more conservative in defining a post as notable, restricting to those with a very high Engagement Score. Thus, α can be used to control the tradeoff between the number of posts to be studied and the abnormality of those posts. For the remainder of the paper, we consider as reactions the sum of Instagram likes and comments. In Figure 4, we show how the number of notable posts varies with different values of α (considering all 261 weeks). For $\alpha = 0$, notable posts are by definition

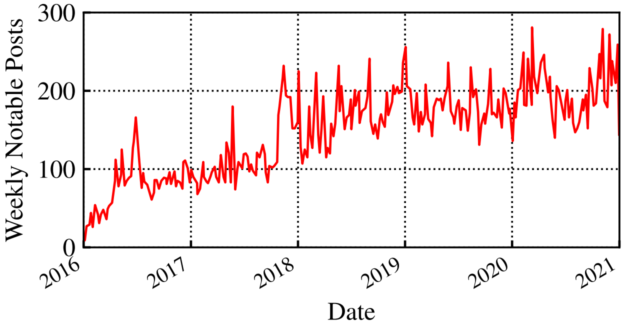


Fig. 6: Number of identified notable posts over the weeks.

all those with an Engagement Score above the 75^{th} percentile of the previous 100 posts. Hence, we obtain a percentage of notable posts around 25% of posts[6]. Increasing α causes this percentage to decrease, and with $\alpha = 1.5$ (the default value of the Boxplot Rule method), notable posts are still 4.5%. Setting α to 2 (3), the percentage of posts decreases to 2.8 (1.3)%. The final choice of α is left to the analyst, while we opt for 2 in the remainder of the paper. The data we want to analyze includes $2-5\,$k posts per week, so looking at the top 2% in terms of Engagement Score (i.e., on the order of a few hundred) seems to be the most reasonable choice.

We illustrate with an example influencer how we identify notable posts. The Engagement Score models the number of reactions we expect from a post, and we look for posts with an unusually high score. We exemplify the whole process by showing the evolution of reactions, Engagement Score, and notable posts for the Italian influencer Valentina Ferragni. In Figure 5a, we show the time series for the number of reactions Valentina Ferragni's posts received over four years of our dataset. We observe a positive trend as the profile gained new followers over the years. The influencer passed from 840 k followers at the beginning of 2017 to almost 4 M followers at the end of 2020, so her posts received more and more reactions over time. The Engagement Score measures how the number of reactions deviates from the expected value by taking a moving average over the last 100 posts. Therefore, it is naturally normalized by the recent size of a profile's audience. Looking at its evolution in Figure 5b, we can see that the Engagement Score assumes values around 1. Values near 0 indicate an under performing post, while high peaks are post eliciting followers' engagement. We are interested in identifying the latter case – i.e., the notable posts - and in the figure they are shown with red markers as identified with $\alpha = 2$. They correspond to the peaks of the Engagement Score. Notice that even though α is constant, working with a sliding window allows us to dynamically change the upper limit over time.

In Figure 6, we show the time series of the number of notable posts for each week of our dataset including years 2016 to 2021. We discover hundreds of notable posts per week, with a slightly increasing trend. The reason is that the total number of posts in our dataset increases over time, caused by the fact that some profiles in our list were created during the period under consideration. In the first months of 2016 the number of notable posts is considerably lower, as our method needs a warm-up period to compute the expected number of reactions and consequently the Engagement Score. We observe many peaks, that we manually analyzed. For example, in March 2020 there is a peak caused by the outbreak of COVID-19. In November 2017 another peak is due to the Italy-Sweden football match, which unexpectedly eliminated Italy from the World Cup qualification. In the following, we show that our methodology can automatically group notable posts on the same notable topic to identify such events.

[6]Not exactly 25% since the process of generation of posts is not i.i.d.

B. Communities and Notable Topics

To find notable topics, we apply the Louvain method for community detection on the graphs created using hashtags extracted from posts. While creating a graph using hashtags is not the only possible choice, it is particularly well suited for Instagram, where we find that 81% of posts contain at least one hashtag. For other OSNs where hashtags are not as popular, we recommend other strategies that we consider future work for now.

First, we characterize the graphs we create at each time step – i.e., taking into account all the posts created in a given week. The number of nodes in these graphs corresponds to the number of posts, which can be derived from Figure 6. In median, we find 297 unique hashtags per week on notable posts, while the 25^{th} and 75^{th} percentiles are 198 and 371 hashtags, respectively. Most nodes (73%) have no edge, so are isolated, meaning that the corresponding posts have no hashtag or none of their hashtags have been used in other trending posts. This is largely expected, since a notable post can easily refer to a topic which is specific to the creator's life and not of interest to other influencers (e.g., a birthday or a wedding). If we exclude these nodes, 63% of the remaining nodes have a single edge, while those with a degree above 5 (10) are 17% (5%).

Using the Louvain method, we extract communities of notable posts from the weekly graphs. The rationale is that posts (nodes) that share hashtags (edges) relate to the same notable topic. In median, we find 10 communities per week, excluding isolated nodes. The maximum number of communities was found during the peak in November 2017, where we identified 18 communities. The goodness of a community structure is usually measured in terms of modularity (see Section II), and values above 0.5 are considered to indicate a strong and cohesive community structure. In Figure 7, we show the distribution of modularity. We calculate it for each week on the community structure we obtain using the Louvain method. In 81% of the weeks it is above 0.5 and its median value is 0.68. We are not surprised to obtain such high modularity values since the graphs are rather small (on the order of a

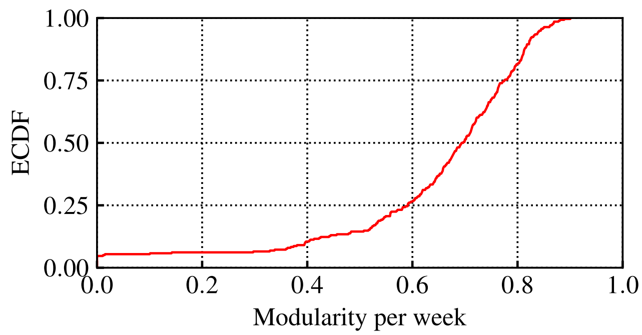
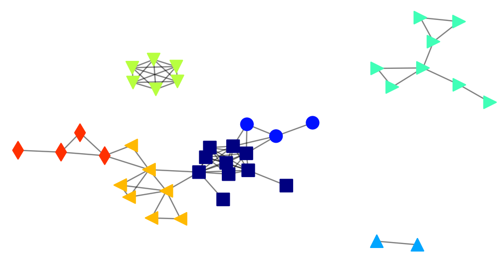
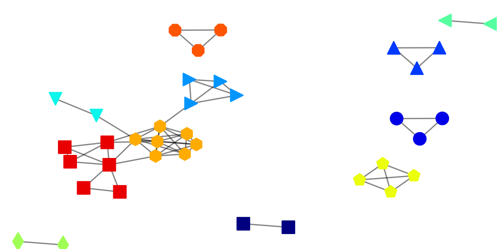


Fig. 7: Distribution of modularity of the identified notable topics on the weekly graphs.



(a) Year 2019, Week 7



(b) Year 2019, Week 36

Fig. 8: Example graph and identified communities for two weeks.

few hundred nodes) and their manual inspection reveals clear groups of well-connected nodes.

We show two examples of these graphs in Figure 8, for two weeks in 2019. We selected on purpose two weeks with a small number of notable posts, so that the resulting graphs have a reduced size, improving the visualization. Starting from the first graph in Figure 8a (nodes are positioned using Fruchterman-Reingold force-directed algorithm), we observe that, on the given week, we identified 84 notable posts. Only 42 of them shared one or more hashtags with other posts, and the figure only plots this subset, as the remaining nodes are isolated components with size 1. The graph has 79 edges, and nodes are arranged in 4 connected components. Using the Louvain method, we identified 7 communities (nodes belonging to different communities are identified with different markers and colors), one for each connected component, except for the biggest one, which was split into 4 communities. In this example, the largest connected component is related to 2019 *Festival di Sanremo*, which is the most popular Italian music competition, and its identified communities are talking about different artists participating in the same competition. Given the clear cohesive community structure of this graph, the Louvain method generated a partitioning with modularity of 0.87. Similarly, in the second graph in Figure 8b, we find 71 notable posts. Among them, only 38 have with one or more edges, which are 61 in total. There are 8 connected components, and Louvain found 11 communities with a modularity of 0.85.

C. Examples of Notable Topics

We finally discuss some example interesting findings emerging from running our methodology on 5 years of Instagram posts. Over 261 weeks, we identified 2 721 communities, referring to diverse topics. In the following, we pinpoint some interesting cases we observed. In summary, these results show that our methodology successfully identify notable topics on OSNs, allowing an analyst to go through a limited number of notable posts and topics to obtain an overview of the trends in a given period. We first show our results related the COVID-19 outbreak, then we illustrate a few other use cases.

The COVID-19 Outbreak. We first focus on the outbreak of COVID-19 in Italy. Italy was among the first countries hit by COVID-19. The first case was identified on February 19^{th}, and on March 11^{th}, the "#IoRestoACasa" decree imposed a total lockdown throughout Italy. People were only allowed to leave the house for valid and proven reasons. The great impact of the pandemic on the OSN debate has already been analyzed in the literature [1], [2]. Using our methodology, we can easily identify this phenomenon. During the week of March 11^{th}, when the lockdown began, almost all notable posts refer to the pandemic. The largest community consists of 32 posts, whose hashtags are shown in Figure 9. We observe how all related somehow to COVID-19 and the lockdown. Examples are *Io Resto A Casa* (the decree imposing the lockdown), *Andrà tutto bene* (everything will be alright) and similar wordings.

Because COVID had tremendous influence throughout 2020, we analyze how many identified Notable Topics included hashtags related to COVID (we manually selected hashtags such as *Coronavirus* and *COVID*). Figure 10 shows the time series of the number of these Notable Topics for each week in 2020. Note that the first topics related to COVID appeared in February 2020 when the first cases were appearing around the globe. Then, when the lockdown was imposed in Italy (week of March 11^{th} 2020), we observed as many as 8 topics related to COVID (the main one already shown in Figure 9). In the following weeks, we observe a decreasing trend in the number of notable topics, reaching zero in the summer of 2020. However, as the year progresses, some new trending topics emerge. See, for example, in November 2020, when a new lockdown was imposed in almost all of Italy.

Italy-Sweden Football Match. We now analyze is the week of the football match Italy-Sweden already mentioned in

Fig. 9: Word Cloud for a community during the COVID-19 outbreak in Italy.

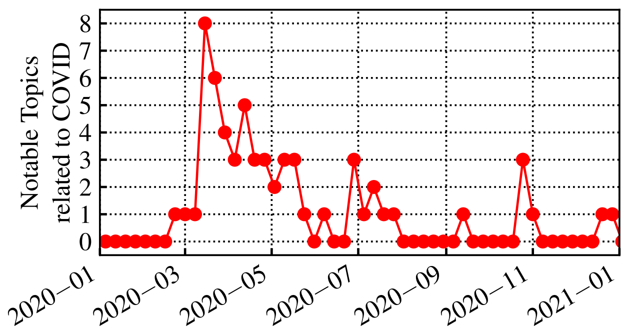


Fig. 10: Number of identified notable communities over the weeks related to COVID-19.

Section IV-A, where we observed a peak in the number of notable posts. The match took place on Monday November 13^{th}, 2017, and the week we are considering begins on that date and ends on Sunday 19^{th}. The match was part of the qualification for the football world cup, and, as a consequence of the defeat, the Italian team was unexpectedly eliminated from the competition. The event had a great echo on the media and on OSNs. During this week, our methodology found a peak in the number of notable posts, which were grouped into 13 communities. We find a large community including 13 notable posts referring to the match, and we show the Word Cloud of the hashtags they contain in Figure 11a. A Word Cloud is a visual representation of the hashtag, where the font size is directly proportional to the frequency of the hashtags in the community. These posts were created by 6 profiles (5 of them are footballer or coaches) and used hashtags related to football, such as *Nazionale* (National team) and *Azzurri* (the Italian national team in slang). On that week, we also find another notable topic related to the popular Italian TV shows (like *GFVIP* and *Rosy Abate*), which were broadcast on the same week (on Italian TV channel *Canale5*) and we identified as a large community of 22 posts, using hashtags related to these shows, shown in Figure 11b.

(a) Community on Football

(b) Community on TV Show

Fig. 11: Word Cloud for two communities for a week in November 2017.

An initial body of work addresses anomaly detection in OSNs. Rahman et al. [13] developed a hybrid anomaly detection system (DT-SVMNB) to classify legitimate and anomalous Facebook users that combines three machine learning algorithms in a cascade: a decision tree, a Support Vector Machine (SVM), and a Naïve Bayesian Classifier (NBC). Similarly, Miz et al. [14] proposed a scalable unsupervised algorithm with a distributed implementation to detect outliers in dynamic graphs, defining an anomaly as a localized abnormal collective behavior of users in a group (cluster) of nodes. Our methodology also uses anomaly detection techniques, but it has the different goal of detecting posts that generate unusual high engagement.

Another set of works proposed methodologies for clustering OSN posts using different approaches. Gao et al. [15] employed clustering algorithms to detect spam campaigns. They then used features from these clusters to train a supervised machine learning model. Savyan et al. [16] clustered posts based on the reactions they received, while Williams et al. [17] have developed GeoContext, a tool for clustering a social media stream into topics based on semantic text analysis. Interdonato et al. [18] combine three different clustering algorithms and natural language processing to obtain a ranking of relevant tweets in emergency situations. Differently from these works, we propose to cluster posts by modeling them as graphs and then applying community detection algorithms. Our approach is thus similar to the methodology proposed by Rubin et al. [19] that detects communities over a network build using hashtags.

V. Related Work

The literature has already studied the presence of anomalous posts in OSN, and how to identify and summarize topics of discussions.

Finally, some works proposed techniques to detect trending post topics and summarize salient events in OSNs. McKelvey et al. [20] developed Truthy, a system that collects

Twitter data to analyze discourse in near real time, specifically designed for the journalistic community that needs a summary of the most important posts on particular topics. Truthy clusters tweets into groups of related messages, defined as all tweets with a common hashtag, stated username, hyperlink, or phrase, which correspond to Twitter users' common conversation subjects, communication routes, or information resources. Similarly, Jin et al. [21] proposed a framework for extracting news from microblogs, while Xia et al. [22] define a methodology for extracting notable events in a given city. Skryzalin et al. [23] achieve a similar goal by combining an information-theoretic analysis to identify time periods whose tweet content differs considerably from usual. In this work we want to go a step further. Our goal is not only to summarize or group a stream of posts. We want to pinpoint only those deserving attention because of the high level of engagement they generate, so clustering is only the final step. To achieve our goal, we need to incorporate different blocks into the pipeline, which we create by taking inspiration from the works mentioned in this section. Thus, our methodology provides a general framework to identify notable posts and group them when they relate to the same notable topic.

VI. Conclusions and Future Directions

Understanding the evolution of discussion topics on OSNs is critical to identifying emerging trends and the impact of external events, particularly in the social sciences and marketing. The size of OSN data poses a challenge for manual review and requires automated techniques. In this paper, we presented a methodology to identify those posts that elicit unusually high user engagement. Our approach allows filtering notable posts related to topics that receive special attention in OSNs. Our methodology groups these posts by modelling them as a network and extracting communities that relate to the same notable topics. We apply our methodology to a dataset of 1.4 million Instagram posts spanning 5 years. Our results show that we are able to discover remarkable events over the years, and we illustrate the case of COVID-19 as well as a few others.

We believe that our methodology is a simple but effective tool for analysts and practitioners to analyse OSN data, as it allows reviewing a limited number of groups, rather than a flood of posts where manual inspection falls short. So far, our methodology is suitable for Instagram as it largely uses hashtags to find similar posts, and its application to other OSNs requires further investigation and adaptation. In addition, the impact of design tradeoffs, thresholds, and algorithmic decisions (see [9]) still need to be tested on a larger scale, across different OSNs and datasets.

References

[1] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, "The COVID-19 social media infodemic," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.

[3] T. Tudoroiu, "Social media and revolutionary waves: The case of the Arab spring," *New Political Science*, vol. 36, no. 3, pp. 346–365, 2014.

[2] M. Trevisan, L. Vassio, and D. Giordano, "Debate on online social networks at the time of COVID-19: An Italian case study," *Online Social Networks and Media*, vol. 23, p. 100136, 2021.

[4] M. Chan, "Social network sites and political engagement: Exploring the impact of Facebook connections and uses on political protest and participation," *Mass communication and society*, vol. 19, no. 4, pp. 430–451, 2016.

[5] J. Dugdale, B. Van de Walle, and C. Koeppinghoff, "Social media and SMS in the Haiti earthquake," in *Proceedings of the 21st International Conference on World Wide Web*, pp. 713–714, 2012.

[6] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," *International journal of information management*, vol. 38, no. 1, pp. 86–96, 2018.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, jul 2009.

[8] F. Soro, T. Favale, D. Giordano, L. Vassio, Z. Ben Houidi, and I. Drago, *The New Abnormal: Network Anomalies in the AI Era*, ch. 11, pp. 261–288. John Wiley & Sons, Ltd, 2021.

[9] H. reference for double blind review.

[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, oct 2008.

[11] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, p. 026113, 2004.

[12] K. Zarei, D. Ibosiola, R. Farahbakhsh, Z. Gilani, K. Garimella, N. Crespi, and G. Tyson, "Characterising and Detecting Sponsored Influencer Posts on Instagram," in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 327–331, 2020.

[13] M. S. Rahman, S. Halder, M. A. Uddin, and U. K. Acharjee, "An efficient hybrid system for anomaly detection in social networks," *Cybersecurity*, vol. 4, no. 1, pp. 1–11, 2021.

[14] V. Miz, B. Ricaud, K. Benzi, and P. Vandergheynst, "Anomaly detection in the dynamics of web and social networks using associative memory," in *The World Wide Web Conference*, pp. 1290–1299, 2019.

[15] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. N. Choudhary, "Towards online spam filtering in social networks.," in *NDSS*, vol. 12, pp. 1–16, 2012.

[16] P. Savyan and S. M. S. Bhanu, "Behaviour profiling of reactions in facebook posts for anomaly detection," in *2017 Ninth International Conference on Advanced Computing (ICoAC)*, pp. 220–226, IEEE, 2017.

[17] E. Williams, J. Gray, E. Morris, B. Bradshaw, K. Williams, and B. Dixon, "A comparison of two methods for the topical clustering of social media posts," in *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 1–7, IEEE, 2016.

[18] R. Interdonato, A. Doucet, and J. Guillaume, "Unsupervised Crisis Information Extraction from Twitter Data," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 579–580, IEEE Computer Society, aug 2018.

[19] F. Rubin, A. F. Alvim, R. dos Santos, and C. R. de Mello, "Detecting Influential Communities in Twitter during Brazil Oil Field Auction in 2019," in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 897–904, IEEE Computer Society, dec 2020.

[20] K. R. McKelvey and F. Menczer, "Truthy: Enabling the study of online social networks," in *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pp. 23–26, 2013.

[21] P. Jin, L. Mu, L. Zheng, J. Zhao, and L. Yue, "News feature extraction for events on social network platforms," in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 69–78, 2017.

[22] C. Xia, J. Hu, Y. Zhu, and M. Naaman, "What is new in our city? A framework for event extraction using social media posts," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 16–32, Springer, 2015.

[23] J. Skryzalin, R. Field, A. Fisher, and T. Bauer, "Temporal anomaly detection in social media," in *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 505–509, IEEE Computer Society, aug 2017.