Article

A Space-Time Framework for Sentiment Scope Analysis in Social Media

Gianluca Bonifazi , **Francesco Cauteruccio** , **Enrico Corradini** , **Michele Marchetti** , **Luigi Sciarretta**, **Domenico Ursino *** and **Luca Virgili**

Department of Information Engineering (DII), Polytechnic University of Marche, Via Brecce Bianche 12, 60121 Ancona, Italy
* Correspondence: d.ursino@univpm.it

Abstract: The concept of scope was introduced in Social Network Analysis to assess the authoritativeness and convincing ability of a user toward other users on one or more social platforms. It has been studied in the past in some specific contexts, for example to assess the ability of a user to spread information on Twitter. In this paper, we propose a new investigation on scope, as we want to assess the scope of the sentiment of a user on a topic. We also propose a multi-dimensional definition of scope. In fact, besides the traditional spatial scope, we introduce the temporal one, which has never been addressed in the literature, and propose a model that allows the concept of scope to be extended to further dimensions in the future. Furthermore, we propose an approach and a related set of parameters for measuring the scope of the sentiment of a user on a topic in a social network. Finally, we illustrate the results of an experimental campaign we conducted to evaluate the proposed framework on a dataset derived from Reddit. The main novelties of this paper are: (i) a multi-dimensional view of scope; (ii) the introduction of the concept of sentiment scope; (iii) the definition of a general framework capable of analyzing the sentiment scope related to any subject on any social network.

Keywords: spatial scope; temporal scope; sentiment analysis; social network analysis; Reddit

1. Introduction

Suppose we are at the shore of a lake on a becalmed day with a flat lake surface. Suppose now that we throw a stone into it. We can see how, starting from the point where the stone falls, the water begins to ripple, and small waves are created. These waves are higher near the point where the stone fell while they become smaller and smaller as we move away from it, until they disappear. The heavier the stone thrown, the higher the waves and the farther they propagate. As time passes, the height of the waves tends to decrease until, if no more stones are thrown into the lake, they disappear and the lake surface becomes motionless again. In our opinion, this image describes better than any other what is meant by "scope". From a more formal point of view, in the Concise Oxford Dictionary (Concise Oxford Dictionary—https://en.oxforddictionaries.com (accessed on 15 September 2022)), *scope* is defined as "*the extent of the area or subject matter that something deals with or to which it is relevant*".

Certainly, there are several similarities between the concept of scope and some other ones used in sociology. Consider, for example, the concepts of centrality, reliability, power, reputation, influence, trust, diffusion, etc. In fact, scope goes beyond these concepts and, at the same time, embraces all of them. In fact, they can be seen as different aspects of scope, which certainly exert their influence on it.

Scope has already been studied in past literature. For example, Ref. [1] analyzed the scope of users and hashtags in Twitter, while [2] proposed an approach to compute the scope of a smart object in a Multi-IoT context. In Refs. [3–8], the authors presented

approaches to analyze some aspects of scope (e.g., reliability, trust, and influence) for users and/or hashtags. In Ref. [9], the authors studied the distribution of the influence of a user across the network, while in Ref. [10], the authors analyzed the attractiveness of users in networks.

In the social network context, which we focus on in this paper, another much analyzed concept is that of user sentiment. Sentiment analysis is one of the most active research strands regarding social networks and, more generally, artificial Intelligence and data analysis [11–14]. In fact, nowadays, millions and millions of people express their sentiments on the most disparate topics through social networks [15–18]. The knowledge of such sentiments and their evolution in space and time is a valuable source of information for various professionals, such as marketers, politicians, journalists, decision makers, and so on. Finally, knowing how a user's sentiment about a topic can propagate to her neighbors, the neighbors of her neighbors, etc., and how that propagation evolves over time, and being able to measure this through appropriate techniques and metrics, represent challenging issues with enormous practical implications.

One way to address these issues might be integrating the concepts of scope and sentiment of one or more users on a certain topic. In fact, to the best of our knowledge, there is not yet an approach that integrates the concepts of scope and sentiment and treats scope from both a spatial and a temporal point of view (with further points of view, or dimensions, likely to emerge in the future). This paper aims to address this issue by proposing a model and a related approach to investigate the spatial and temporal scope of a user's sentiment about a topic in a social platform.

Our proposed model is based on two graphs. The first is a bipartite one that stores all available information about users, their posts, comments, and sentiments on certain topics. It can be employed as an information source from which it is possible to extract all the necessary data for the processing required by the approach proposed in this paper and any future approaches. The second is derived from the first; it is a single-mode graph representing users and their interactions and is employed for the analyses proposed in this paper. Our model also includes several complement functions that can be exploited to obtain specific information from the first graph or to perform certain supporting processing.

Our approach consists of several steps. First, it identifies the topics emerging from the posts and comments published by users. Then, for each topic, it determines the sentiments of the various users who covered it. For these two activities it employs techniques already proposed in the literature [11,13,19–22]. In other words, for these two activities, our approach is independent of the techniques adopted to perform them. Next, it exploits the concepts, measures, and techniques of Social Network Analysis [23] and graph theory [24–26] to define spatial scope. This definition has a dual nature, since the spatial scope is defined as a set of pairs or a rooted graph. Both definitions aim to indicate the users involved in the spread of sentiment and the intensity degree of the latter for each user involved. The two-fold definition of spatial scope allows us to use both set theory and graph theory for its analysis. Using them, our approach provides a set of metrics and measures for assessing the spatial scope of a user's sentiment on a topic.

In a completely similar way, our approach acts to ensure the temporal scope assessment of a user's sentiment on a topic. In this case, the temporal scope is defined as an ordered list of pairs. As mentioned above, spatial and temporal scopes are orthogonal and represent two views or dimensions that could be integrated, and possibly enriched with other views in the future.

To evaluate the potential of the proposed framework, we conducted a series of experiments on a dataset derived from Reddit, one of the most popular social networks. As we will see in the paper, our model proved to be capable of extracting interesting information that may be useful to various professionals interested in the knowledge of scope.

In summary, the gap in the literature that this paper aims to fill concerns the lack of a general framework capable of supporting a multi-dimensional analysis of the scope of the sentiment of users on any topic in any social network. In fact, as will become

clearer in the following, some papers in the past literature analyze the scope of a user or a smart object, others investigate user sentiment, but none analyze the scope of the sentiment of one or more users. Moreover, all previous papers that analyze scope take only the spatial dimension into account and none of them consider other dimensions, such as the temporal one. Finally, most of the studies on sentiment refer to either a well-defined social network or a well-defined subject, and were not designed to operate on any social network and any subject. Our paper aims to fill this gap. Specifically, its contributions are as follows: (i) it introduces the concept of scope of a user's sentiment on one or more topics in a social platform; (ii) it proposes a multi-dimensional definition of scope; and (iii) it presents a framework for studying the scope of a user's sentiment on one or more topics and extracting information from the corresponding data; this framework can operate on any social platform and can evaluate the scope of the user sentiment on topics concerning any subject.

The outline of this paper is as follows: In Section 2, we illustrate the Related Literature. In Section 3, we describe our model for scope representation. In Section 4, we present our approach. In Section 5, we illustrate the experimental campaign we conducted. Finally, in Section 6, we draw our conclusions and look at possible future developments.

2. Related Literature

2.1. Preface

In recent years, social networks have been exploited by groups of users to discuss a variety of topics, from politics to gossip, from health to sport, and so on. The pervasive spread of social networks has prompted researchers to study the behavior of users who join social networks in various reference contexts [27–29]. Alongside the analysis of the structure of networks, which already provides very interesting knowledge patterns about the behavior of users accessing them, researchers have also begun to examine the content posted and exchanged between users [30–33]. Regarding the latter, elements of particular interest to them are the extraction of topics from texts and the assessment of the sentiment that users have about a given topic. Our work aims to integrate these two research streams because it aims to analyze how the scope of the sentiment of a user on a specific topic propagates both spatially and temporally across a social network.

To better perform our analysis, we divide this section into two parts. In the first, we analyze works dealing with scope and related concepts, while in the second we focus on the analysis of sentiment diffusion within a community.

2.2. Related Literature on the Concept of Scope

In the past, the scope of users in social networks has been studied in Ref. [1]. In this paper, the authors analyze the scope of an entity on Twitter. Specifically, they define a framework to measure various aspects of scope (e.g., influence, reliability, popularity) simultaneously and for multiple entities (e.g., users, hashtags). In this way, they can measure the overall scope and several aspects of it by comparing the latter with each other and for different entities. Such comparisons make it possible to extract knowledge patterns (indicating, for example, the presence of anomalies and outliers) that can then be used in several application domains (e.g., information dissemination). In Ref. [2], the authors extend the concept of scope from people to smart objects in a multiple IoT context. In particular, they formalize two scope definitions for smart objects and illustrate some real-world applications of the knowledge patterns thus extracted. Returning to the people-to-people social network context, many authors analyze single aspects of the concept of scope, such as reliability, trust, and influence for users and/or hashtags [3,5–7]. Regarding user influence, the authors of Ref. [9] use the PageRank algorithm to analyze the distribution of influence across the network. In Ref. [10], the authors analyze the attractiveness of users in a social network. The approach they propose characterizes a user based on the new users she is able to interrelate with over time. The authors propose to perform influence maximization, but they do not consider the topics covered by users.

Other approaches investigate the evolution of topic trends in Twitter by analyzing the use of hashtags. Indeed, the latter allow a natural division of posts according to their topics [4,8,34,35]. In particular, the authors of Ref. [8] measure the topic-sensitive influence of users on Twitter by means of an approach based on PageRank. They analyze how a social network user can influence other ones on a topic. Moreover, they propose a metric to compute the influence of users on specific topics. Our approach differs from the one proposed in Ref. [8] in that the latter does not aim to describe the features characterizing the influence of one user on another but only gives a quantitative estimate of that influence. Furthermore, the approach proposed in Ref. [8] does not consider the value of sentiment in its analysis. In Ref. [4], the authors propose an approach that, given a hashtag on Twitter, uses the corresponding comments to construct its profile in order to predict its popularity. This approach has some similarities with ours. In fact, it can be seen as a method for predicting the spatial scope of a hashtag, and thus how far the latter will spread. Our approach differs from the one described in Ref. [4] in that, in investigating the spread of a user's sentiment toward a topic, it analyzes the contribution made by other users of her neighborhood. Therefore, our analysis is more user-centric than the one proposed in Ref. [4]. Furthermore, the approach of Ref. [4] does not consider sentiment when analyzing the expansion of a topic. Finally, it is not intended to propose a multi-dimensional analysis of a hashtag's popularity unlike our approach, which proposes both a spatial and a temporal scope and leaves open the possibility of further dimensions of scope in the future.

In Ref. [36], the authors present a study of the spread of negative sentiment in Online Social Media. The main similarity between the approach of Ref. [36] and ours concerns the idea of studying the spread of user sentiment in online platforms. However, the approach proposed in Ref. [36] focuses on hate speech and does not consider topics, unlike our approach, which precisely analyzes the scope of user sentiment on topics. In Ref. [37], the authors study the propagation of negative sentiment in messages exchanged on the Chinese Sina microblog. This approach is therefore specific to a social platform, and, in this feature, it differs from our approach that it has been designed to operate on any social platform. In Ref. [38], the authors investigate how the spread of sentiment can cause a viral spread of fake news in social media. This paper focuses on the analysis of a specific phenomenon (i.e., the relationship between sentiment and fake news). Instead, our paper aims to propose a general framework capable of studying the space-time evolution of the scope of user sentiment on one or more topics.

2.3. Related Literature on the Sentiment of Users

The second part of our analysis on related literature concerns the sentiment of a user on one or more topics. Many approaches to face this problem have been proposed in the past literature. Some of them address this issue from a static point of view; in particular, they generally employ opinion mining techniques to understand the sentiment emerging from a given text [39–42]. In contrast, other techniques address this problem from a dynamic point of view; in fact, given the characteristics of a sentiment on a topic, they want to understand how those characteristics affect the spread of that sentiment both among users and over time [43–49]. Specifically, in Ref. [43], the authors propose a model that combines sentiment and opinion propagation to assess the global sentiment on a given topic. Similarly to our approach, the one of Ref. [43] considers the sentiment emerging from a text and wants to understand how it propagates. However, the authors of Ref. [43] do not aim to provide a formalization of such propagation.

In Ref. [44], the authors present the MISNIS framework, which aims to identify the most influential users on specific topics. It also divides user messages into three categories, based on the results obtained after performing a sentiment analysis task. To carry out topic mining, it analyzes all the words in the message and not only the hashtags; thus, it is able to achieve a higher accuracy. MISNIS and our approach share the goal of analyzing how influential a user is with respect to a specific topic. However, the concepts of topic and sentiment are kept separate in MISNIS, while they are integrated into our approach because

it aims to assess the sentiment of users on a topic. Topic and sentiment analysis is also the core of the approach described in Ref. [50]. This approach considers topics and sentiments from Reddit posts and comments and aims to analyze them for extracting information without creating a model.

In Ref. [51], the authors analyze and represent the spread of microblogging topics and sentiments through two graphs and propose metrics to measure the influence of stakeholders in that spread. In pursuing this goal, the approach of Ref. [51] has several similarities with ours. For instance, both are graph-based and define metrics to measure sentiment diffusion. Unlike Ref. [51], which considers topics and sentiments separately, our approach integrates these two entities because it analyzes the spread of users' sentiments on topics. The approach of Ref. [51] is based on a global analysis that examines the whole network to identify the stakeholders of interest. In contrast, our approach tends to work on partitions of the network and not on the global one; in fact, it analyzes how the sentiment of a user on a topic propagates to its neighborhood. Finally, it considers two orthogonal types of scope diffusion, namely spatial and temporal ones. This concept is not present in the approach of Ref. [51].

In Refs. [52,53], the authors analyze changes in topics and sentiment during emergencies. For this purpose, they consider different types of emergencies. In addition, they analyze the spread and propagation of topics and sentiment for different user stereotypes, e.g., governments, celebrities, and media. Our paper differs from Refs. [52,53] because it does not analyze sentiment but the scope of sentiment. Moreover, it proposes a multidimensional (in particular, spatio-temporal) view of scope. Finally, our framework is general and can be applied to any social network for analyzing the sentiment on topics under any circumstances. Instead, Refs. [52,53] focus on emergencies.

In particular, as far as temporal scope is concerned, we point out that various studies have been proposed in the literature, which aim to assess how several single aspects of scope (e.g., influence, reliability, popularity) evolve over time [54–57]. However, none of these approaches comprehensively consider the concept of scope but only assesses individual aspects of it. Moreover, in the reference contexts, the goals they set and the techniques they use to achieve those goals are very different from the ones adopted in our approach.

3. The Proposed Model

3.1. A Formal Representation of the Context of Interest

Before presenting our model, it is necessary to provide a formalization of the context in which it operates. This context concerns a social platform whose users can publish posts and comments. We assume that both posts and comments consist mainly of text; if there are other types of content, these are only to accompany the text. A user publishes a comment when she wants to reply to a previously published post or comment.

We employ the symbol $\mathcal{U} = \{u_1, \cdots, u_l\}$ to represent the set of users operating in our context, the symbol $\mathcal{P} = \{p_1, \cdots, p_m\}$ to denote the set of posts published by the users of $\mathcal{U}$ in a time interval T, and the symbol $\mathcal{C} = \{c_1, \cdots, c_n\}$ to indicate the sets of comments posted by these users in T. Given a user $u_i \in \mathcal{U}$, we adopt the symbol $\mathcal{P}_i$ (resp., $\mathcal{C}_i$) to denote the subset of the posts (resp., comments) of $\mathcal{P}$ (resp., $\mathcal{C}$) published by her.

As specified in the Introduction, one of the most important factors to consider in our analysis is time. Therefore, a way to model it is in order. To this end, given an overall time interval of interest, we can think of modeling it as an ordered sequence of z time slices, $T = T_1, \cdots, T_z$. For example, T could be a certain month, say August 2022, and it could be represented as a succession of 31 time slices, one for each day. It is advisable that our representation of time should allow the indexing of the sequence of time slices. In other words, it should be possible to select only a particular interval of contiguous time slices of T (e.g., the second decade of August 2022). To this end, our time model uses the notation $T[x..y]$, $1 \leq x \leq y \leq z$, to denote the interval of contiguous time slices in T that begins at T_x and ends at T_y. If $x = y$, then it means that we want to take a single time slice; in this case,

we will use the abbreviated notations T_x or $T[x]$ to represent $T[x..x]$. If $x = 1$ and $y = z$, then it means we are considering the overall interval of interest; in this case, we will use the abbreviated notation T, instead of $T[1..z]$, to denote that interval.

The previous notation about time intervals and slices can be extended to the other sets of the model. Specifically, we denote by $\mathcal{P}[x..y] \subseteq \mathcal{P}$ (resp., $\mathcal{C}[x..y] \subseteq \mathcal{C}$) the subset of posts (resp., comments) published in the time interval $T[x..y]$ and by $\mathcal{P}[x]$ (resp., $\mathcal{C}[x]$) the subset of posts (resp., comments) published in the time slice x. Finally, we use the abbreviated notation $\mathcal{P}$ (resp., $\mathcal{C}$) to indicate the overall set of posts $\mathcal{P}[1..z]$ (resp., comments $\mathcal{C}[1..z]$) published in the overall time interval T.

Two additional concepts that play a key role in our context are the ones of topic and sentiment tag. A topic is an abstract concept discussed in one or more posts and comments. Natural Language Processing (NLP) researchers have long studied the issues of topic modeling and extraction, and proposed a variety of interesting solutions [19,20]. A sentiment tag is a keyword used to summarize the sentiment expressed on a particular topic. Typical sentiment tags are "pos", "neg", and "neu", to indicate a positive, a negative, and a neutral sentiment, respectively. In the following, we denote by $\mathcal{T} = \{t_1, \cdots, t_q\}$ the set of topics extracted from the posts of $\mathcal{P}$ and the comments of $\mathcal{C}$, while we denote by $\mathcal{S} = \{s_1, \cdots, s_r\}$ the set of available sentiments (tags). In the following, to simplify the discussion, we will use the term "sentiment" instead of "sentiment tag". Given a topic $t_j \in \mathcal{T}$ and a sentiment $s_k \in \mathcal{S}$, we use the pair (t_j, s_k) to indicate that t_j has been tagged with s_k, i.e., that s_k has been associated with t_j.

3.1.1. Identifying Topics from Posts and Comments

Our framework is independent of the technique adopted for constructing the set $\mathcal{T}$ of topics related to posts and comments. Recall that, in our context, the latter consist mainly of text and that any other content is only an accompaniment to the text. Consequently, to construct $\mathcal{T}$, we can use any approach for identifying topics from a given text proposed in the past literature (see Refs. [19–21] for some surveys on it). Therefore, in the following, we assume that, given a post $p \in \mathcal{P}$ (resp., a comment $c \in \mathcal{C}$), our framework can employ a technique capable of deriving topics from p (resp., c), adding them to the overall set $\mathcal{T}$ of topics and associating them with the post p (resp., comment c) from which they were derived.

3.1.2. Identifying the Sentiments Characterizing Posts and Comments

Our framework is also independent of the technique used for identifying the sentiments associated with a text. This issue has been extensively studied in sentiment analysis research. Here, researchers have proposed several techniques capable of defining, characterizing, and extracting the sentiment expressed in a text (see Refs. [11,13,22] for some surveys about this topic). In this context, the terms "sentiment tag", "sentiment value", or, simply, "sentiment" have been used equivalently [58].

The technique adopted for identifying sentiments from posts and comments must examine each post (resp., comment) $p \in \mathcal{P}$ (resp., $c \in \mathcal{C}$), acquire the sentiments that emerge in it and associate them with the corresponding topics of $\mathcal{T}$ referring to p (resp., c). In more detail, it proceeds as follows: let p (resp., c) be a post (resp., comment) of $\mathcal{P}$ (resp., $\mathcal{C}$). It could consist of a simple text, expressing a single sentiment, or a complex text, expressing several sentiments that may even conflict with each other. Clearly, the former hypothesis is a special case of the latter; so, in the following, we will consider the latter directly. In this case, we assume that p (resp., c) consists of a succession $p_1, p_2, \cdots, p_w$ (resp., $c_1, c_2, \cdots, c_w$) of texts such that each of them expresses a single sentiment. In what follows, we will use the term "fragment" to refer to each textual content p_k (resp., c_k), $1 \le k \le w$; in addition, we will use the symbol f_k to generalize p_k and c_k. Now, the technique described in Section 3.1.1 can be applied on the fragment f_k to obtain the set $\mathcal{T}_{f_k}$ of topics considered in f_k. Then, any technique proposed in the literature to derive the sentiment expressed in a simple text (see [11,13,22]) can be applied on f_k to determine the sentiment s_k characterizing it. At this

point, for each topic $t_j \in \mathcal{T}_{f_k}$, we have a pair (t_j, s_k) indicating that s_k is the sentiment on t_j expressed in f_k. The set of all the sentiments extracted from all the posts of $\mathcal{P}$ and all the comments of $\mathcal{C}$ form the set $\mathcal{S}$ of the sentiments that characterize our reference context.

3.2. The Proposed Model

After formalizing the context of interest, in this section we describe our proposed model. The first element of it is a bipartite support graph $\mathcal{B}$, aiming to enable the storage of the key information of the reference context. As we will see below, from this rather rich and complex graph, it is possible to derive more agile ones, which allow us to perform our analyses more effectively and efficiently. $\mathcal{B}$ is defined as follows:

$$\mathcal{B} = \langle N' \cup N'', E' \rangle \tag{1}$$

$N' \cup N''$ represents the set of nodes of $\mathcal{B}$. Specifically, the nodes in N' are associated with users in $\mathcal{U}$. Indeed, each node $n_i \in N'$ corresponds to a user $u_i \in \mathcal{U}$, and vice versa. Since there is a biunivocal correspondence between a node of N' and a user of $\mathcal{U}$, we will use these two terms interchangeably in the following. Each node $n_{jk} \in N''$ corresponds to a pair (t_j, s_k), where t_j is a topic of $\mathcal{T}$ and s_k is a sentiment of $\mathcal{S}$. It indicates that t_j has been tagged with s_k in at least one post of $\mathcal{P}$ or comment of $\mathcal{C}$.

E' represents the set of edges of $\mathcal{B}$; an edge $(n_i, n_{jk}) \in E'$ between a node $n_i \in N'$ and a node $n_{jk} \in N''$ indicates that the user u_i published at least one post or comment in which she expressed the sentiment s_k on the topic t_j. Since u_i may have carried out this task more times in the time interval T, we associate a label l_{ijk} with (n_i, n_{jk}). It indicates the list of timestamps of the posts and/or comments published by u_i in which she expressed the sentiment s_k on the topic t_j.

$\mathcal{B}$ contains all potentially useful information to enable us to investigate the context of our interest. However, being a two-mode graph, it is not easy to analyze and manipulate. Graph theory suggests constructing one or more one-mode graphs from it, each focusing on a single aspect of interest and operating on them [23]. Now, the object of our analysis is the spatial and temporal evolution of the scope of the sentiment of users in a social platform. Consequently, the key aspect to focus on consists of users and their interactions; given these premises, it is reasonable to construct a user-centered single-mode graph from $\mathcal{B}$ in such a way as to operate directly on it instead of $\mathcal{B}$. This graph is defined as follows:

$$\mathcal{A} = \langle N, E \rangle \tag{2}$$

N represents the set of nodes of $\mathcal{A}$. A node $n_i \in N$ corresponds to a user $u_i \in \mathcal{U}$, and vice versa. Again, since there is a biunivocal correspondence between a node $n_i \in N$ and a user $u_i \in \mathcal{U}$, we will employ these two terms interchangeably in the following. Clearly, N is equivalent to the set of nodes N' of $\mathcal{B}$. E is the set of edges of $\mathcal{A}$. An edge $e_{ih} = (n_i, n_h)$ belonging to E indicates that the users u_i and u_h published at least one post/comment on the same topic and, at least once, u_i published a comment on a post/comment of u_h, or vice versa.

As can be seen from its definition, the graph $\mathcal{A}$ is very agile and streamlined so that the analyses performed on it are effective and efficient. In some of these analyses there may be a need to use data present in $\mathcal{B}$ that we do not deem necessary to report in $\mathcal{A}$ so as not to burden this graph (for example, because such data are only rarely used). In these cases, we define ad hoc functions that retrieve the necessary information from $\mathcal{B}$ and complement our model. For example, if the set of posts on a given topic t_j published by a certain user u_i in the time interval $T[x..y]$ were needed, we could define a function that receives u_i, t_j and $T[x..y]$, and returns the desired set of posts. In Section 3.2.1, we list the functions needed for our approach and that complement our model.

Given the graph $\mathcal{A}$ and a topic t_j of $\mathcal{T}$, we define the projection $\overline{\mathcal{A}^j}$ of $\mathcal{A}$ onto t_j as the graph obtained from $\mathcal{A}$ by considering only the nodes corresponding to users who published at least one post or comment having t_j as their topic. More formally:

$$\overline{\mathcal{A}^j} = \langle \overline{N^j}, \overline{E^j} \rangle \qquad (3)$$

$\overline{N^j}$ represents the set of nodes of $\overline{\mathcal{A}^j}$. A node $n_i \in \overline{N^j}$ corresponds to a user $u_i \in \mathcal{U}$ who published at least one post or comment on the topic t_j. $\overline{E^j}$ represents the set of edges of $\overline{\mathcal{A}^j}$. There exists an edge (n_i, n_h) in $\overline{E^j}$ if there exists a corresponding edge (n_i, n_h) in the graph $\mathcal{A}$.

Given the graph $\mathcal{A}$ and the time interval $T[x..y]$, we denote by $\mathcal{A}[x..y]$ the "projection of $\mathcal{A}$" in $T[x..y]$:

$$\mathcal{A}[x..y] = \langle N[x..y], E[x..y] \rangle \qquad (4)$$

A node n_i belongs to $N[x..y]$ if the corresponding user u_i published at least one post/comment in $T[x..y]$. An edge $e_{ih} \in E[x..y]$ indicates that u_i and u_h published at least one post/comment on the same topic in the time interval $T[x..y]$ and that, in the same interval, at least once, u_i published a comment on a post or a comment of u_h, or vice versa. Clearly, $\mathcal{A}[x] = A[x..x]$ is the "projection of $\mathcal{A}$" in the time slice T_x and $\mathcal{A}[1..z]$ is equivalent to $\mathcal{A}$.

3.2.1. Functions Complementing Our Model

In this section, we present some support functions that complement our model. They will be used to formalize the activities performed by our approach. Before describing them, we feel it is appropriate to introduce some concepts concerning the prevalence or ambivalence of the sentiment of a user or a community on a topic.

Let u_i be a user of $\mathcal{U}$ and let t_j be a topic of $\mathcal{T}$. We define the *positive* (resp., *negative*, *neutral*) *sentiment degree* of u_i on t_j as the fraction of posts and/or comments on t_j published by u_i with which a positive (resp., negative, neutral) sentiment was associated after the application of the approach described in Section 3.1.2.

Having made these premises, we can now introduce our complement functions. They are:

- $\sigma^+(u_i, t_j)$: It receives a user u_i and a topic t_j and computes the positive sentiment degree δ_{ij}^+ of u_i on t_j. δ_{ij}^+ ranges in the real interval $[0, 1]$; the higher its value, the higher the strength of the positive sentiment. $\sigma^+(u_i, t_j)[x..y]$ represents the "projection of $\sigma^+(u_i, t_j)$" in the time interval $T[x..y]$; it performs the same computation as $\sigma^+(u_i, t_j)$ but considers only the posts and comments published in the time interval $T[x..y]$. We indicate by $\delta_{ij}^+[x..y]$ the corresponding result. Clearly, $\delta_{ij}^+[x] = \delta_{ij}^+[x..x]$ is the "projection of δ_{ij}^+" in the time slice T_x and $\delta_{ij}^+[1..z]$ is equivalent to δ_{ij}^+;

- $\sigma^=(u_i, t_j)$: It receives a user u_i and a topic t_j and computes the neutral sentiment degree $\delta_{ij}^=$ of u_i on t_j. $\delta_{ij}^=$ ranges in the real interval $[0, 1]$; the higher its value, the higher the strength of the neutral sentiment. $\sigma^=(u_i, t_j)[x..y]$ represents the projection of $\sigma^=(u_i, t_j)$ in the time interval $T[x..y]$; $\delta_{ij}^=[x..y]$ denotes the corresponding result;

- $\sigma^-(u_i, t_j)$: It receives a user u_i and a topic t_j and computes the negative sentiment degree δ_{ij}^- of u_i on t_j. δ_{ij}^- ranges in the real interval $[0, 1]$; the higher its value, the higher the strength of the negative sentiment. $\sigma^-(u_i, t_j)[x..y]$ represents the projection of $\sigma^-(u_i, t_j)$ in the time interval $T[x..y]$; $\delta_{ij}^-[x..y]$ denotes the corresponding result;

- $\nu(n_i, \lambda, gr)$: It receives a graph $gr = \langle \hat{N}, \hat{E} \rangle$, a node $n_i \in \hat{N}$ and a positive integer λ and returns a set of nodes representing the neighborhood of level λ of n_i in gr. Formally speaking:

$$\nu(n_i, \lambda, gr) = \{n_h | n_h \in \hat{N}, \langle n_i, n_h \rangle = \lambda\} \qquad (5)$$

Here, $\langle n_i, n_h \rangle$ represents the length of the shortest path from n_i to n_h in gr;

- $\bar{v}(n_i, gr)$: It receives a graph $gr = \langle \hat{N}, \hat{E} \rangle$ and a node $n_i \in \hat{N}$ and returns the set of nodes directly connected to n_i in gr. In other words, $\bar{v}(n_i, gr) = v(n_i, 1, gr)$;
- $size(gr)$: It receives a graph gr and returns its size, i.e., the number of its nodes;
- $diameter(gr)$: It receives a graph gr and returns its diameter, i.e., the length of the longest shortest path between any pair of nodes in gr;
- $sps(u_i, t_j)$: It receives a user u_i and a topic t_j and returns true if u_i has a strongly positive sentiment on t_j. This happens when $\sigma^+(u_i, t_j) > \sigma^-(u_i, t_j)$ and $\sigma^+(u_i, t_j) \geq \sigma^=(u_i, t_j)$. In all the other cases it returns false. $sps(u_i, t_j)[x..y]$ represents the projection of $sps(u_i, t_j)$ in the time interval $T[x..y]$;
- $wps(u_i, t_j)$: It receives a user u_i and a topic t_j and returns true if u_i has a weakly positive sentiment on t_j. This happens when $\sigma^+(u_i, t_j) > \sigma^-(u_i, t_j)$ and $\sigma^+(u_i, t_j) < \sigma^=(u_i, t_j)$. In all the other cases it returns false. $wps(u_i, t_j)[x..y]$ represents the projection of $wps(u_i, t_j)$ in the time interval $T[x..y]$;
- $sns(u_i, t_j)$: It receives a user u_i and a topic t_j and returns true if u_i has a strongly negative sentiment on t_j. This happens when $\sigma^-(u_i, t_j) > \sigma^+(u_i, t_j)$ and $\sigma^-(u_i, t_j) \geq \sigma^=(u_i, t_j)$. In all the other cases it returns false. $sns(u_i, t_j)[x..y]$ represents the projection of $sns(u_i, t_j)$ in the time interval $T[x..y]$;
- $wns(u_i, t_j)$: It receives a user u_i and a topic t_j and returns true if u_i has a weakly negative sentiment on t_j. This happens when $\sigma^-(u_i, t_j) > \sigma^+(u_i, t_j)$ and $\sigma^-(u_i, t_j) < \sigma^=(u_i, t_j)$. In all the other cases it returns false. $wns(u_i, t_j)[x..y]$ represents the projection of $wns(u_i, t_j)$ in the time interval $T[x..y]$.

At the end of this section, we observe that the four functions $sps()$, $wps()$, $sns()$, and $wns()$ are *mutually exclusive*, in the sense that at most one of them must be true, and *complete*, in the sense that at least one of them must be true. It follows that, in a given time interval $T[x..y]$, given a user $u_i \in \mathcal{U}$ and a topic $t_j \in \mathcal{T}$, exactly one of these functions returns true and all the others return false. This allows us to determine the concept of *sentiment type* of u_i on t_j in the time interval $T[x..y]$; it represents the sentiment type associated with that function among the four indicated above that returns true in the time interval $T[x..y]$. Clearly, the possible sentiment types are: (i) strongly positive (hereafter, *sp*); (ii) weakly positive (hereafter, *wp*); (iii) weakly negative (hereafter, *wn*); (iv) strongly negative (hereafter, *sn*).

4. The Proposed Approach

4.1. Objective and Research Questions

The main objective of this paper is to introduce a multi-dimensional view of the scope of the sentiment of a user on one or more topics on any social platform. The paper also wants to define a framework for sentiment scope evaluation. Getting more specific, the main research questions the paper wants to answer are the following:

- **RQ1**: Is it possible to introduce the concept of sentiment scope? In fact, in the past, scope was defined for users and smart objects, but never for sentiments.
- **RQ2**: Is it possible to define a temporal view of scope? In fact, in the past literature, the only view of scope considered was the spatial one.
- **RQ3**: Is it possible to define a framework for evaluating the space-time scope of a sentiment of one or more users on any topic in any social platform?
- **RQ4**: Are there any differences between the scope of negative and positive sentiments?
- **RQ5**: Are there any differences between the scope of strong and weak sentiments?
- **RQ6**: How does the scope of the sentiment of a user on one or more topics propagate to her neighbors?
- **RQ7**: What kind of behavior do users generally exhibit with respect to a sentiment on a topic? In other words, is their sentiment stable or swinging?
- **RQ8**: In showing their sentiments on topics, are users posed and balanced or are they biased toward positive sentiments or negative ones?

In the next sections, we aim to answer all these research questions.

4.2. Determining the Spatial Scope of the Sentiment of a User on a Topic

In this section, we illustrate our approach to determine the spatial scope of the sentiment of a user $u_i \in \mathcal{U}$ on a topic $t_j \in \mathcal{T}$. In Section 3.2.1, we have seen that there exist four possible sentiment types. Consequently, it is possible to determine four kinds of scope, one for each sentiment type. In this section, we examine all of them starting with the scope associated with a strongly positive sentiment.

First, let us specify how the spatial scope of u_i on t_j can be represented. A first possibility consists of a set Σ_{ij}^+ of pairs, as shown in Equation (6):

$$\Sigma_{ij}^+ = \{(u_1, \delta_{1j}^+), (u_2, \delta_{2j}^+), \cdots, (u_g, \delta_{gj}^+)\} \tag{6}$$

Each pair (u_h, δ_{hj}^+), $h \neq i$, belongs to Σ_{ij}^+ and consists of a user u_h, directly or indirectly connected to u_i, and the corresponding positive sentiment degree δ_{hj}^+ on t_j.

A second representation consists of a subgraph $\overline{\mathcal{A}^{ji}} = \langle \overline{N^{ji}}, \overline{E^{ji}} \rangle$ of $\mathcal{A}^j$. A node n_h belongs to $\overline{N^{ji}}$ if the corresponding user u_h is present in Σ_{ij}^+. Furthermore, n_i belongs to $\overline{N^{ji}}$. An arc (n_h, n_u) belongs to $\overline{E^{ji}}$ if an arc (n_h, n_u) also exists in $\overline{E^j}$. We call *origin of* $\overline{\mathcal{A}^{ji}}$ the node n_i corresponding to the user u_i.

At this point we can define our approach for computing the spatial scope associated with a strongly positive sentiment (hereafter referred to as *strongly positive spatial scope*) of u_i on t_j. We represent this approach by defining a function $\psi^+()$ shown in Equation (7). It receives u_i, t_j and the initially empty set Σ_{ij}^+ as parameters. It basically performs a depth-first search on $\overline{\mathcal{A}^j}$, starting from u_i and selecting a node only if certain constraints are satisfied for it. It can be formalized as shown in Equation (7):

$$\psi^+(u_i, t_j, \Sigma_{ij}^+) = \begin{cases} \{(u_i, \delta_{ij}^+)\} \cup \bigcup_{n_h \in \overline{v}(n_i, \overline{\mathcal{A}^j})} \psi^+(u_h, t_j, \Sigma_{ij}^+ \cup \{(u_i, \delta_{ij}^+)\}) & \text{if } sps(u_i, t_j) = \texttt{true} \\ & \text{and } (u_i, \delta_{ij}^+) \notin \Sigma_{ij}^+ \\ \varnothing & \text{otherwise} \end{cases} \tag{7}$$

In other words, the function $\psi^+()$, when applied on u_i and t_j, first checks whether u_i has a strongly positive sentiment on t_j. If this is true and the pair (u_i, δ_{ij}^+) is not already present in Σ_{ij}^+, then $\psi()^+$ adds this pair to Σ_{ij}^+. Afterwards, it recursively calls itself by passing as input each node directly connected to n_i in $\overline{\mathcal{A}^j}$. In contrast, if u_i has not a strongly positive sentiment on t_j or the pair (u_i, δ_{ij}^+) is already present in Σ_{ij}^+, then $\psi()$ simply returns $\varnothing$ and the recursion stops.

The *strongly negative spatial scope* can be defined in a similar way (see Equation (8)). Again, we can introduce a function $\psi^-()$ that receives u_i, t_j and the initially empty set Σ_{ij}^-. It has an identical behavior to the function $\psi^+()$, except that δ_{ij}^+ is replaced by δ_{ij}^- and the function $sps()$ is replaced by the function $sns()$, defined in Section 3.2.1. Its formalization is shown in Equation (8):

$$\psi^-(u_i, t_j, \Sigma_{ij}^-) = \begin{cases} \{u_i, \delta_{ij}^-\} \cup \bigcup_{n_h \in \overline{v}(n_i, \overline{\mathcal{A}^j})} \psi^-(u_h, t_j, \Sigma_{ij}^- \cup \{(u_i, \delta_{ij}^-)\}) & \text{if } sns(u_i, t_j) = \texttt{true} \\ & \text{and } (u_i, \delta_{ij}^-) \notin \Sigma_{ij}^- \\ \varnothing & \text{otherwise} \end{cases} \tag{8}$$

The *weakly positive* (resp., *negative*) *spatial scope* is defined similarly to the strongly positive (resp., negative) spatial scope. In this case, we introduce a function $\xi^+()$ (resp., $\xi^-()$) that receives u_i, t_j and an initially empty set Π_{ij}^+ (resp., Π_{ij}^-). Its behavior is identical to the one of the function $\psi^+()$ (resp., $\psi^-()$) except that the function $sps()$ (resp., $sns()$) is replaced by the function $wps()$ (resp., $wns()$). The formalization of $\xi^+()$ and $\xi^-()$ is shown in Equations (9) and (10):

$$\xi^+(u_i, t_j, \Pi_{ij}^+) = \begin{cases} \{u_i, \delta_{ij}^+\} \cup \bigcup_{n_h \in \overline{\nu}(n_i, \overline{\mathcal{A}^j})} \xi^+(u_h, t_j, \Pi_{ij}^+ \cup \{(u_i, \delta_{ij}^+)\}) & \text{if } wps(u_i, t_j) = \texttt{true} \\ & \text{and } (u_i, \delta_{ij}^+) \notin \Pi_{ij}^+ \\ \varnothing & \text{otherwise} \end{cases} \quad (9)$$

$$\xi^-(u_i, t_j, \Pi_{ij}^-) = \begin{cases} \{u_i, \delta_{ij}^-\} \cup \bigcup_{n_h \in \overline{\nu}(n_i, \overline{\mathcal{A}^j})} \xi^-(u_h, t_j, \Pi_{ij}^- \cup \{(u_i, \delta_{ij}^-)\}) & \text{if } wns(u_i, t_j) = \texttt{true} \\ & \text{and } (u_i, \delta_{ij}^-) \notin \Pi_{ij}^- \\ \varnothing & \text{otherwise} \end{cases} \quad (10)$$

At this point, we have defined the functions for computing the strongly positive (resp., negative) spatial scope Σ_{ij}^+ (resp., Σ_{ij}^-) and the weakly positive (resp., negative) spatial scope Π_{ij}^+ (resp., Π_{ij}^-)). We have also previously seen that it is possible to provide a graph-based representation of such a scope. In the following, in order not to burden the notation, we will use the symbol $\mathcal{SG}^+$ (resp., $\mathcal{SG}^-$, $\mathcal{WG}^+$, $\mathcal{WG}^-$) to denote the graph-based representation corresponding to Σ_{ij}^+ (resp., Σ_{ij}^-, Π_{ij}^+, Π_{ij}^-). Its formalization is reported in Equation (11):

$$\begin{aligned} \mathcal{SG}^+ &= \langle SN^+, SE^+ \rangle \\ \mathcal{SG}^- &= \langle SN^-, SE^- \rangle \\ \mathcal{WG}^+ &= \langle WN^+, WE^+ \rangle \\ \mathcal{WG}^- &= \langle WN^-, WE^- \rangle \end{aligned} \quad (11)$$

By studying some properties of these graphs, it is possible to define a variety of information regarding the scope of the sentiment of u_i on t_j.

In what follows we will perform all our analyses with regard to the graph $\mathcal{SG}^+$, although everything we will see can be straightforwardly extended to the other three graphs.

The first two properties of the scope of the sentiment of u_i on t_j that we consider are its breadth and its depth. Regarding the breadth, it is immediate to think that it can be obtained by considering the size of $\mathcal{SG}^+$, that is, the number $|SN^+|$ of its nodes. As far as the depth is concerned, we recall that $\mathcal{SG}^+$ derives from a depth-first search performed on $\overline{\mathcal{A}^j}$ starting from the node n_i, which we have also called the origin of $\mathcal{SG}^+$. Therefore, the depth of the scope can be determined by computing the diameter of $\mathcal{SG}^+$, that is, the maximum length of the minimum paths from n_i to any other node of $\mathcal{SG}^+$.

An important investigation consists in determining how the strongly positive sentiment degree varies as we move away from n_i in $\mathcal{SG}^+$. To do this, we can consider the neighborhood of level λ, $1 \le \lambda \le d$, $d = diameter(\mathcal{SG}^+)$, obtained by applying the function ν on n_i, λ and $\mathcal{SG}^+$. For each neighborhood, it is then possible to compute the average strongly positive sentiment degree of the nodes belonging to it. Generally, if there were no interference, as we move away from n_i, the average strongly positive sentiment degree of a neighborhood should decrease because the influence that n_i exerts on nodes tends to decrease. However, it could be the case that, once we move away from n_i, there is another node different from it that exerts an influence on the nodes of the neighborhood of n_i. If the new "influencer" has a discordant sentiment with n_i, we might see a steep decrease in the average strongly positive sentiment degree, or even a reversal of sentiment polarity. By contrast, if the new "influencer" has a concordant sentiment with n_i, we may see a slowdown in the decline of the average strongly positive sentiment degree, or even a new growth of it. The correlation that can arise between two scopes is a challenging topic that is, however, beyond the objective of this paper. Here, we simply provide a tool for computing the variation in the average strongly positive sentiment degree as we move away from n_i.

Let $\nu(n_i, \lambda, \mathcal{SG}^+)$ be the neighborhood of level λ, $1 \le \lambda \le d = diameter(\mathcal{SG}^+)$ of n_i in $\mathcal{SG}^+$. The average positive sentiment degree $\overline{\delta_{ij_\lambda}^+}$ of $\nu(n_i, \lambda, \mathcal{SG}^+)$ is defined in Equation (12):

$$\overline{\delta_{ij_\lambda}^+} = \frac{\sum_{n_h \in \nu(n_i, \lambda, \mathcal{SG}^+)} \delta_{hj}^+}{size(\nu(n_i, \lambda, \mathcal{SG}^+))} \tag{12}$$

In other words, it is obtained by computing the average strongly positive sentiment degree of all the nodes belonging to $\nu(n_i, \lambda, \mathcal{SG}^+)$. $\overline{\delta_{ij_\lambda}^+}$ ranges in the real interval $[0, 1]$; the higher its value, the higher the strength of the average positive sentiment.

At this point, we have at our disposal a succession of values $\varrho_0^+, \varrho_1^+, \cdots, \varrho_d^+$ such that $\varrho_0^+ = \delta_{ij}^+, \varrho_h^+ = \overline{\delta_{ij_h}^+}, 1 \leq h \leq d, d = diameter(\mathcal{SG}^+)$. The examination of that succession can give us some interesting insights into how the average strongly positive sentiment degree evolves as we move away from n_i. It takes into account the decreasing influence of n_i as we move away from it, as well as the possible presence of any interference from other "influencers".

By plotting the values of $\varrho_0^+, \varrho_1^+, \cdots, \varrho_d^+$, we get a "spectrum" of the trend of the strongly positive sentiment degree in the spatial scope of u_i. In fact, several interesting pieces of information can be derived from that spectrum. These include:

- The variation in the average strongly positive sentiment degree in the hth section of the spectrum, defined in Equation (13):

$$\Delta_h^+ = \varrho_h^+ - \varrho_{h-1}^+ \tag{13}$$

- The relative variation in the average strongly positive sentiment degree in the hth section of the spectrum, defined in Equation (14):

$$\overline{\Delta_h^+} = \frac{\varrho_h^+ - \varrho_{h-1}^+}{\varrho_{h-1}^+} \tag{14}$$

- The mean variation in the average strongly positive sentiment degree in the hth section of the spectrum, defined in Equation (15):

$$\widehat{\Delta_h^+} = \frac{\varrho_h^+ - \varrho_0^+}{h} \tag{15}$$

- The maximum variation in the average strongly positive sentiment degree in the hth section of the spectrum, defined in Equation (16):

$$\Delta^{M+} = \max_{h=1..v} |\Delta_h^+| \tag{16}$$

- The minimum variation in the average strongly positive sentiment degree in the hth section of the spectrum, defined in Equation (17):

$$\Delta^{m+} = \min_{h=1..v} |\Delta_h^+| \tag{17}$$

Finally, we can analyze the monotonicity of the succession $\varrho_0^+, \varrho_1^+, \cdots, \varrho_d^+$. In particular, we are interested in whether it is monotonically non-increasing. This occurs when $\varrho_h^+ \leq \varrho_{h-1}^+$. In fact, if such a condition is not satisfied, we can say that, as we move away from n_i, there is at least one further "influencer" with a sentiment concordant with the one of n_i that is acting on the nodes of the neighborhoods of n_i. Otherwise, it could be that there is no other "influencer" interfering with n_i or that such an "influencer" is present but with a discordant sentiment with the one of n_i.

What we have seen now are just some of the analyses we can perform on spatial scope. They allow us to give an idea of the potential of this concept. Many other analyses could be thought of simply by applying concepts from mathematical analysis to the succession $\varrho_0^+, \varrho_1^+, \cdots, \varrho_d^+$ or concepts from graph theory to the graph $\mathcal{SG}^+$.

Finally, it is worth emphasizing again that all the analyses we have previously done on $\mathcal{SG}^+$ could be straightforwardly extended to $\mathcal{SG}^-$, $\mathcal{WG}^+$ and $\mathcal{WG}^-$.

4.3. Determining the Temporal Scope of the Sentiment of a User on a Topic

In this section, we illustrate our approach to determine the temporal scope of the sentiment of the user $u_i \in \mathcal{U}$ on a topic $t_j \in \mathcal{T}$. In Section 3.2.1, we introduced two concepts on sentiment scope, namely *sentiment type* and *sentiment degree* of u_i on t_j. These two concepts play a key role in the analysis of temporal scope. Recall that the sentiment type of u_i on t_j can be strongly positive (*sp*), weakly positive (*wp*), weakly negative (*wn*), and strongly negative (*sn*). Instead, the sentiment degree of u_i on t_j is given by the value of the parameter δ_{ij}^+, in case the sentiment type is *sp* or *wp*, or the value of the parameter δ_{ij}^-, in case it is *wn* or *sn*.

The temporal scope of u_i on t_j in the time interval $T[x..y]$ can be represented by an ordered list of pairs, as shown in Equations (18)–(20).

$$\Theta_{ij}[x..y] = [(\tau_x, \theta_x), (\tau_{x+1}, \theta_{x+1}), \cdots, (\tau_y, \theta_y)] \tag{18}$$

$$\tau_b = \begin{cases} sp & \textit{if } sps(u_i, t_j)[b] = \texttt{true} \\ wp & \textit{if } wps(u_i, t_j)[b] = \texttt{true} \\ wn & \textit{if } wns(u_i, t_j)[b] = \texttt{true} \\ sn & \textit{if } sns(u_i, t_j)[b] = \texttt{true} \end{cases} \tag{19}$$

$$\theta_b = \begin{cases} \delta_{ij}^+[b] & \textit{if } (sps(u_i, t_j)[b] = \texttt{true}) \text{ or } (wps(u_i, t_j)[b] = \texttt{true}) \\ \delta_{ij}^-[b] & \textit{if } (wns(u_i, t_j)[b] = \texttt{true}) \text{ or } (sns(u_i, t_j)[b] = \texttt{true}) \end{cases} \tag{20}$$

Recall that $sps(u_i, t_j)[b]$ (resp., $wps(u_i, t_j)[b]$, $wns(u_i, t_j)[b]$, $sns(u_i, t_j)[b]$) represents the projection of $sps(u_i, t_j)$ (resp., $wps(u_i, t_j)$, $wns(u_i, t_j)$, $sns(u_i, t_j)$) in the time slice T_b (see Section 3.2.1). Analogously $\delta_{ij}^+[b]$ (resp., $\delta_{ij}^-[b]$) denotes the value returned by the function $\sigma^+(ui, t_j)$ (resp., $\sigma^-(ui, t_j)$) when projected in the time slice T_b (see, again, Section 3.2.1).

Clearly, by moving from a time instant T_b to a time instant T_{b+1} the value of τ can remain unvaried or change and the value of θ can increase, decrease, or remain constant. Each combination of the trend of these two parameters at the transition from T_b to T_{b+1} gives us interesting information about the time trend of the sentiment degree of u_i on t_j. For example:

- If both τ_b and τ_{b+1} are equal to *sp*:

  - if $\theta_{b+1} > \theta_b$, it means that the sentiment degree is strengthening;
  - if $\theta_{b+1} = \theta_b$, it means that the sentiment degree is static;
  - if $\theta_{b+1} < \theta_b$, it means that, although a strongly positive sentiment still characterizes u_i, it is weakening.

- If $\tau_b = sp$ and $\tau_{b+1} = wp$, it means that the posts and comments on t_j published by u_i in which she shows a neutral sentiment, are increasing. This increase is such that they exceed the ones in which u_i shows a positive sentiment. The number of posts/comments with a positive sentiment continues to be greater than the number of posts/comments with a negative sentiment. However, at the time slice T_{b+1} we are seeing a weakening of the positivity of the sentiment of u_i on t_j, compared to the time slice T_b.

- If $\tau_b = sp$ and $\tau_{b+1} = wn$, it means that u_i is changing her sentiment on t_j. This change is not yet radical, since there is a prevalence of neutral posts/comments over negative ones.

- If $\tau_b = sp$ and $\tau_{b+1} = sn$, it means that u_i has completely changed her sentiment on t_j. The greater the gap between θ_b and θ_{b+1} and the greater the change occurred.

Similarly, suitable information can be extracted in case $\tau_b = wp$, $\tau_b = wn$ or, finally, $\tau_b = sp$.

Analogously to what we have seen for spatial scope, several measures can also be defined for temporal scope. They allow us to get a quantitative view of the changes in the sentiment degree of u_i on t_j over a time interval. Some of these measures are the following (in defining these measures, we will refer to the time slices T_b and T_{b-1}, instead of the time slices T_b and T_{b+1}, to bring their definition in line with that of the metrics for spatial scope, explained in Section 4.2.):

- The variation in the sentiment degree between the time slices T_{b-1} and T_b. It can be defined in Equation (21):

$$\Lambda_b = \theta_b - \theta_{b-1} \tag{21}$$

- The relative variation in the sentiment degree between the time slices T_{b-1} and T_b. It can be defined in Equation (22):

$$\overline{\Lambda_b} = \frac{\theta_b - \theta_{b-1}}{|\theta_{b-1}|} \tag{22}$$

- The mean variation in the sentiment degree in the time interval $T[x..y]$. It is defined in Equation (23):

$$\widehat{\Lambda} = \frac{\theta_y - \theta_x}{y - x} \tag{23}$$

- The maximum variation in the sentiment degree in the time interval $T[x..y]$. It is defined in Equation (24):

$$\Lambda^{M+} = \max_{b=x..y} |\Lambda_b| \tag{24}$$

- The minimum variation in the sentiment degree in the time interval $T[x..y]$. It is defined in Equation (25):

$$\Lambda^{m+} = \min_{b=x..y} |\Lambda_b| \tag{25}$$

In addition to defining appropriate metrics to measure the change in the sentiment of u_i on t_j, we can check whether the succession of the values of the sentiment degree in the interval $T[x..y]$ is monotonic or not. This information must be closely coupled with that related to sentiment type. In particular, if the succession of values $\theta_x, \theta_{x+1}, \cdots, \theta_y$ is monotonically non-increasing, it means that, in the time interval $T[x..y]$, the sentiment of u_i on t_j is not strengthening and, rather, it is presumably decreasing. Such a decrease could cause the sentiment type to go from strongly positive to weakly positive, weakly negative, or even strongly negative. On the other hand, if the previous succession of values is monotonically non-decreasing, it means that, in the time interval $T[x..y]$, the sentiment of u_i on t_j is not weakening, and, rather, it is presumably strengthening. In this case, we might see reverse transitions from the previous case, e.g., from strongly negative to weakly negative, weakly positive, and strongly positive.

The previous succession may also not be monotonic. In this case, the measures on changes in sentiment degree defined above could be extremely useful. It might also be useful to determine how often the change from one type of sentiment to another occurs, or how often the change from an increasing to a decreasing trend occurs, or vice versa.

Analogously to the spatial scope, those seen above are just some of the analyses that can be performed on the temporal scope. Many other analyses could be performed by applying the concepts of mathematical analysis or time series analysis to the succession of values $\theta_x, \theta_{x+1}, \cdots, \theta_y$.

5. Experimental Campaign

5.1. Dataset Description

To build a dataset capable of supporting our experiments, we chose Reddit as the reference social platform. We carried out such a choice because: (i) Reddit is very popular (in fact, it currently ranks 11th among the most visited sites according to Visual Capitalist

(www.visualcapitalist.com (accessed on 12 September 2022))); (ii) it allows posts and comments on any topic; and (iii) its data are easily accessible through `pushshift.io` [59]; the latter is a data repository that allows people to download data related to Reddit comments and posts through a suitable API.

In building our dataset we focused on the posts and comments of one particular subreddit, namely /r/worldnews. The reasons for this choice lie in the fact that it has already been used as a reference subreddit in previous analyses (see Refs. [60–62]) and in the fact that it is one of the most complete and neutral news-related subreddits.

Specifically, through `pushshift.io`, we retrieved all posts and comments, along with the corresponding metadata, published in this subreddit from 25 February 2022 to 25 March 2022. The number of posts taken into account is equal to 9884 while the number of comments is equal to 633,371.

Once the data of interest were downloaded from `pushshift.io`, we performed ETL (Extraction, Transformation, and Loading) activities on them. Specifically: (i) we removed all posts and comments published by users who had left Reddit; (ii) we removed all posts and comments that did not have textual content or were written in a language other than English; (iii) we selected only those posts and comments related to a specific discussion theme. Regarding the latter, the choice was complex as it was important to select a specific, but sufficiently broad, theme with many facets, and thus many topics. Based on this reasoning, our choice fell on the armed conflict in Ukraine that began on 24 February 2022.

After filtering and other ETL activities, the final number of posts in the dataset is 2703, which is 27.12% of the initial ones. In contrast, the final number of comments is 82,617, which is 13.21% of the initial ones. In Table 1, we report some of the main characteristics of the final dataset. In addition to the information mentioned above, this table reports some further interesting information. In particular, we can see that the number of authors in our dataset is 4219. Among them only 119 published both posts and comments. This number is clearly very low; in particular, it is 26.50% of the authors publishing posts and 3.14% of those publishing comments.

Table 1. Some main parameters of the dataset adopted for our experiments.

Parameter	*Value*
No. of posts	2703
No. of comments	82,617
No. of (distinct) authors	4219
No. of (distinct) authors publishing posts	449
No. of (distinct) authors publishing comments	3787
No. of (distinct) authors publishing both posts and comments	119

In Figure 1, we show the distribution of comments against posts, while in Figure 2 we report the distribution of comments against score. Both figures are in log-log scale. By examining them we can observe that both distributions follow power laws. Table 2 reports the values of the corresponding coefficients α and δ.
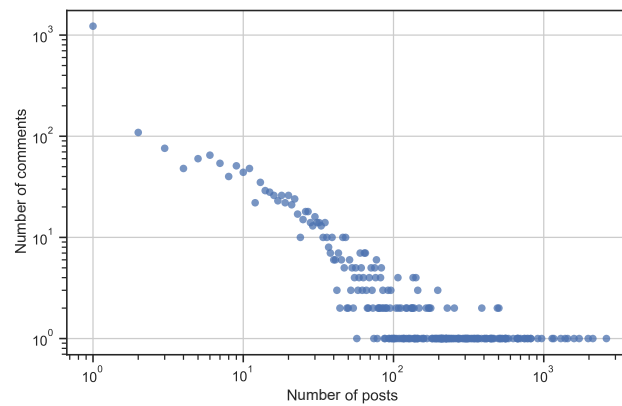
Figure 1. Distribution of comments against posts (log-log scale).
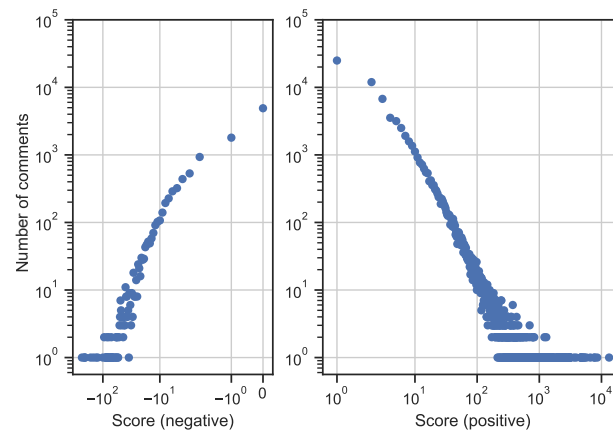

Figure 2. Distribution of comments against score (log-log scale).

Table 2. Values of α and δ of the power law distributions for the considered dataset—* These values were computed considering the absolute values of scores.

Distribution	α	δ
Figure 1	1.8408	0.0419
Figure 2 (left) *	2.9262	0.0418
Figure 2 (right)	2.0383	0.0136

5.2. Identification of Topics and Sentiments

In Section 3.1.1, we saw that our framework is independent of the technique used for constructing the set $\mathcal{T}$ of topics. In our experimental campaign, we adopted BERTopic [63] to obtain $\mathcal{T}$. BERTopic is based on BERT (Bidirectional Encoder Representation from Transformers). The latter is a powerful deep learning-based framework for performing NLP tasks on texts. More specifically, it is a topic modeling technique that exploits transformers [64] and c-TF-IDF [65] to create dense clusters from which easily interpretable topics can be derived. BERTopic receives a set of documents as input and returns the list of topics covered in them. It also associates each topic thus obtained with a description and a count. The former consists of the set of words characterizing the topic. The latter indicates the number of documents mentioning it. Given a document, BERTopic is always able to determine a set of topics that characterize it.

We applied BERTopic to the 2703 posts and 82,617 comments in the dataset and obtained a set $\mathcal{T}$ of 101 topics. Table 3 shows some examples of the extracted topics.

Table 3. Some examples of the topics and their descriptions extracted by BERTopic.

Topic	Description
t_1	{invasion, invade, mission}
t_2	{nato, defence, member, treaty}
t_3	{bunker, underground}

After constructing the set $\mathcal{T}$ of topics, we turned to consider the sentiments characterizing the posts and comments published by users. In this activity, we used roBERTa-base [66]. This system was trained on approximatively 124 million tweets published from January 2018 to December 2021. Next, it was expressively fine-tuned for sentiment analysis using the TweetEval benchmark [67]. We decided to use roBERTa-base because there is a strongly similarity between the shape of texts characterizing tweets and the one of texts in posts and comments of Reddit. In fact, in both cases, we are in the presence of fast-paced messages employed to express opinions and thoughts in general.

The set of sentiments that can be derived by roBERTa-base are those typically used in sentiment analysis, namely "pos", "neg", and "neu". They are also the sentiments considered in our model, as we have seen in Section 3.1. Therefore, the set $\mathcal{S}$ of sentiments is $\mathcal{S} = \{$ "pos", "neg", "neu" $\}$. Table 4 shows some examples of fragments, along with the corresponding sentiments, derived by roBERTa-base. Let f_k be a fragment of a comment or a post (as mentioned in Section 3.1.2, f_k can coincide with a whole comment or a whole post, if these are characterized by a single sentiment.) characterized by a single sentiment. Let s_k be the sentiment that roBERTa-base derived for f_k. Finally, let $\mathcal{T}_{f_k}$ be the set of topics of f_k identified by BERTopic. Then, the joint use of BERTopic and roBERTa-base on f_k allows us to extract a pair (t_j, s_k) for each element t_j of $\mathcal{T}_{f_k}$. Such a pair indicates that the sentiment s_k was associated with the topic t_j in f_k. As previously pointed out, 101 topics were identified in our dataset. From them, 302 pairs of the type (t_j, f_k) were obtained.

Table 4. Some examples of fragments and their sentiments derived by roBERTa-base (swear words are partially masked).

Fragment	Sentiment
"It makes me hopeful too. We need to find a way to get NATO forces engaged."	pos
"But it's a f***ing kid that got killed by that c**t"	neg
"Anyone know when this interview took place? NBC has no time stamp on the video"	neu

5.3. Descriptive Analysis of the Graphs $\mathcal{A}$, $\mathcal{SG}^+$, $\mathcal{SG}^-$, $\mathcal{WG}^+$, and $\mathcal{WG}^-$

In this section, we present a descriptive analysis of the graphs $\mathcal{A}$, $\mathcal{SG}^+$, $\mathcal{SG}^-$, $\mathcal{WG}^+$. and $\mathcal{WG}^-$ obtained from our dataset. This analysis allows us to identify some features of these graphs that will be useful in the next experiments. It also allows us to identify the first differences among the four graphs $\mathcal{SG}^+$, $\mathcal{SG}^-$, $\mathcal{WG}^+$, and $\mathcal{WG}^-$, and thus among the trends of the various sentiment types that we defined in this paper.

We begin our analysis from the graph $\mathcal{A}$. Recall that this is a user-centered, single mode graph representing user interactions. Specifically, an edge in $\mathcal{A}$ indicates that the users associated with the corresponding nodes published at least one post or comment on the same topic and, further, that one of the two users commented on at least one post or comment of the other.

In Table 5, we report the values of some features of the graph $\mathcal{A}$. In particular, we consider the number of nodes, the number of arcs, the density, and the clustering coefficient. Clearly, the number of nodes of $\mathcal{A}$ is equal to the number of distinct authors in the dataset, and thus to 4219. The number of arcs of $\mathcal{A}$ is 32,648 and, consequently, the density is 0.0018, which is a very low value. This can be explained both by taking into account the average number of comments posted by each user, which is 19.58, and by considering

that the condition of existence of an arc in $\mathcal{A}$ is very stringent. In fact, an arc exists in $\mathcal{A}$ if at least one of the comments of one of its nodes refers to a post or comment of the other node. The clustering coefficient is equal to 0.0349. This value is quite high if we consider the low density of $\mathcal{A}$. It implies that this graph consists of several components strongly internally connected and weakly coupled together. Some of these components may also be disconnected from all the others. This is already an interesting result found through our analysis. Indeed, it tells us that in the r/worldnews subreddit, users tend to organize themselves into high cohesive and weakly coupled communities.

Table 5. Some basic properties of the graph $\mathcal{A}$.

Property	*Value*
Number of nodes	4219
Number of edges	32,648
Density	0.0018
Clustering coefficient	0.0349

Having analyzed the graph $\mathcal{A}$, we now turn to the analysis of the graphs $\mathcal{SG}^+$, $\mathcal{SG}^-$, $\mathcal{WG}^+$, and $\mathcal{WG}^-$. As we saw in Section 4.2, each of these graphs is related to a pair (u_i, t_j), where u_i is a user and t_j is a topic. The four graphs are associated with the four possible sentiment types; in particular, the graph $\mathcal{SG}^+$ (resp., $\mathcal{SG}^-$, $\mathcal{WG}^+$ and $\mathcal{WG}^-$) is associated with the strongly positive (resp., strongly negative, weakly positive, weakly negative) spatial scope. Essentially, these graphs represent the spatial spread of the scope related to a user u_i discussing a topic t_j. Recall that, given the pair (u_i, t_j), only one of the four graphs can exist in a given time interval, depending on the sentiment type that u_i had shown for t_j in that time interval. Since each graph is associated with a pair (u_i, t_j), in our analysis we considered all possible pairs of users (u_i, t_j) in the various time slices of the dataset and, for each of them, we calculated its breadth (which coincides with the number of its nodes) and its depth (which coincides with its diameter). Finally, we aggregated the results based on sentiment type, obtaining average values for each graph types. These are shown in Table 6.

Table 6. Average values of breadth and depth for the graphs $\mathcal{SG}^+$, $\mathcal{SG}^-$, $\mathcal{WG}^+$, and $\mathcal{WG}^-$.

Property	$\mathcal{SG}^+$	$\mathcal{SG}^-$	$\mathcal{WG}^+$	$\mathcal{WG}^-$
Average breadth	143	187	89	124
Average depth	7.8	8.4	6.9	7.3

The examination of this table reveals additional interesting insights. First of all, the differences among the four graphs under examination mainly concern the average breadth, while the values of the average depth are more similar. In addition, we can observe that for both the average breadth and the average depth the graphs associated with negative sentiments have higher values than the corresponding graphs associated with positive sentiments. This is in line with several researches proposed in the past literature whose authors found that negative sentiments tend to spread more easily than positive ones [68–72]. Finally, we can observe that, for both average breadth and average depth, the graphs associated with weak sentiments have lower values than the corresponding graphs associated with strong ones. This is in line with other studies proposed in the past literature where it has been shown that the stronger a sentiment is, the more people resonate with it, and the likelier it is they will spread it to others [71,73,74].

5.4. Experiments on Spatial Scope

5.4.1. Variation of the Spatial Scope against the Neighborhood Level

We began our experiments on spatial scope by analyzing how it varies against the neighborhood level and whether this variation differs for the different sentiment types. To conduct this analysis we proceeded as follows.

Let us first consider the case in which the sentiment type is strongly positive. In Section 4.2, we have seen that, in this case, the graph representing the scope is $\mathcal{SG}^+$ and the average positive sentiment degree of the neighbors of level λ of the user u_i on the topic t_j is $\overline{\delta_{ij_\lambda}^+}$, as shown in Equation (12). We have also seen that the trend of this degree against ν is given by a succession of values $\varrho_0^+, \varrho_1^+, \cdots, \varrho_d^+$ such that $\varrho_0^+ = \delta_{ij}^+$, $\varrho_h^+ = \overline{\delta_{ij_h}^+}$, $1 \leq h \leq d$, $d = diameter(\mathcal{SG}^+)$. In other words, this succession measures the variation of the u_i's capability of influencing the sentiment on t_j as we move away from her in the social platform, also taking into account the possible interference of other users.

In Section 5.3, we have seen that the average depth (which coincides with the average diameter) of $\mathcal{SG}^+$ is 7.8. Therefore, in the current analysis, we consider a value of h ranging from 0 to 7.

Consider, now, all possible pairs of users (u_i, t_j) such that u_i showed a strongly positive sentiment on t_j. For each of these pairs, we performed all the computations specified above and constructed the succession $\varrho_0^+, \varrho_1^+, \cdots, \varrho_d^+$. The latter tells us how the average value of the sentiment degree on t_j of the neighbors of level h, $0 \leq h \leq 7$, of the users showing a strongly positive sentiment on t_j varies against the increase of h. In other words, it shows how the influence of the users having a strongly positive sentiment on t_j varies as we move away from them in the social platform, also taking into account the possible interference of other users. Finally, we computed the mean of all the values of $\varrho_0^+, \varrho_1^+, \cdots, \varrho_d^+$ over the possible pairs of users (u_i, t_j). These mean values are graphically reported in Figure 3.
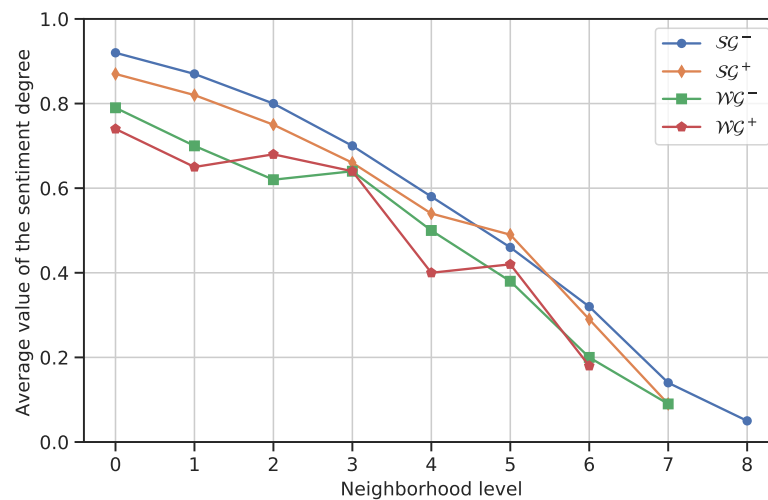

Figure 3. Variation of the mean value of the sentiment degree on t_j of users against the neighborhood level they belong to.

Similarly, we computed the succession of the mean values of the average sentiment degree on t_j of the neighbors of level h of the users showing a strongly negative (resp., weakly positive, weakly negative) sentiment on t_j. It indicates the variation of the influence of the users having a strongly negative (resp., weakly positive, weakly negative) sentiment on t_j as we move away from them in the social platform, also taking into account the possible interference of other users. In this case, based on Table 6, h should range from 0 to 8 (resp., 6, 7). The values of this succession are graphically reported in Figure 3.

From the analysis of this figure, we can deduce several useful information about the trend of the spatial scope for the different sentiment types. As we might expect, whatever the sentiment type, as the neighborhood level increases, the average sentiment degree (and, thus, the influence of the corresponding users regarding the sentiment on a topic) decreases. As for the different types of scope, we can observe that the users with negative sentiment have a greater influence than the ones with positive sentiment, and the users with strong sentiment have a greater influence than the ones with a weak sentiment. This is in line with the results described in Section 5.3 and those found in the past literature [68–74].

The analysis of Figure 3 shows that the influence of users with strongly negative sentiment degree, besides being generally strong, decreases smoothly. This suggests that it is not affected by any interference from other users. When we turn to users with strongly positive sentiment degree, we can see that there is always a decrease of their influence, but this is somewhat more irregular. This indicates that the influence of this type of users may be affected, although not decisively, by the interference from other users. At some time slices, this interference can accelerate the influence decrease, while, at other time slices, it is able to slow it down. However, it is not able to reverse the trend. As for users with weak sentiment degree, we can observe that the trend is more irregular. Overall, the values are lower than the corresponding ones of the users with strongly negative sentiment degree. In addition, the interference from other users is stronger. In fact, it does not only make the decrease irregular, but is also able to reverse the trend at some points, although only for short stretches. All the peculiarities characterizing the influence of users with weakly negative sentiment degree occur even more strongly for the influence of users with weakly positive sentiment degree. In this case, the trend is even more irregular and its inversions are more frequent and pronounced.

As we have seen in Section 4.2, starting from the successions shown in Figure 3, we can derive several other interesting information. In particular, as for the succession corresponding to the average strongly positive sentiment degree, we have that:

- $\Delta_1^+ = -0.05$; $\Delta_2^+ = -0.07$; $\Delta_3^+ = -0.10$; $\Delta_4^+ = -0.12$; $\Delta_5^+ = -0.12$; $\Delta_6^+ = -0.14$; $\Delta_7^+ = -0.18$.
- $\overline{\Delta_1^+} = -\frac{0.05}{0.92} = -0.05$; $\overline{\Delta_2^+} = -\frac{0.07}{0.87} = -0.08$; $\overline{\Delta_3^+} = -0.13$; $\overline{\Delta_4^+} = -0.17$; $\overline{\Delta_5^+} = -0.21$; $\overline{\Delta_6^+} = -0.30$; $\overline{\Delta_7^+} = -0.56$.
- $\widehat{\Delta_1^+} = -\frac{0.87-0.92}{1} = -0.05$; $\widehat{\Delta_2^+} = -\frac{0.80-0.92}{2} = -0.06$; $\widehat{\Delta_3^+} = -0.07$; $\widehat{\Delta_4^+} = -0.09$; $\widehat{\Delta_5^+} = -0.09$; $\widehat{\Delta_6^+} = -0.10$; $\widehat{\Delta_7^+} = -0.11$.
- $\Delta^{M+} = max(0.05, 0.07, 0.10, 0.12, 0.12, 0.14, 0.18) = 0.18$.
- $\Delta^{m+} = min(0.05, 0.07, 0.10, 0.12, 0.12, 0.14, 0.18) = 0.05$.

Similarly, we can compute the corresponding parameter values for the other successions we examined above.

Furthermore, we can say that the successions related to $\mathcal{SG}^-$ and $\mathcal{SG}^+$ are monotonically non-increasing. In contrast, the successions related to $\mathcal{WG}^-$ and $\mathcal{WG}^+$ are non-monotone.

Finally, as specified in Section 4.2, many other analyses can be conducted on the successions and on the graphs $\mathcal{SG}^+$, $\mathcal{SG}^-$, $\mathcal{WG}^+$ and $\mathcal{WG}^-$ for extracting more information. For example, we can observe that the succession corresponding to $\mathcal{WG}^-$ presents only one trend reversal while the succession corresponding to $\mathcal{WG}^+$ shows two trend reversals, which also have a larger magnitude. This also allows us to say numerically and objectively that the latter succession is more irregular than the former.

5.4.2. Relationship between Density and Clustering Coefficient and Spatial Scope

In the previous sections, we have seen some analyses allowing us to derive information on a spatial scope from its representation through a graph. In particular, we have illustrated what information on a spatial scope can be derived from the breadth and depth of the corresponding graph, as well as from the analysis of the variation of the sentiment against the neighborhood levels. In this section, we want to continue in this direction by considering

some Social Network Analysis and graph theory parameters and seeing if and how they can support us in gaining a deeper understanding of scope. In particular, we will focus on density and average clustering coefficient.

In Section 5.3, we have computed the values of these two parameters for the graph $\mathcal{A}$, and we have seen that they are low; then, we have provided an explanation for this behavior. In this section, we want to see what happens for the graphs $\mathcal{SG}^+$, $\mathcal{SG}^-$, $\mathcal{WG}^+$, and $\mathcal{WG}^-$ associated with the various sentiment types.

To answer that question, we computed the density and the average clustering coefficient for all the graphs of types $\mathcal{SG}^+$ (resp., $\mathcal{SG}^-$, $\mathcal{WG}^+$, and $\mathcal{WG}^-$) associated with the pairs (u_i, t_j) such that u_i showed a strongly positive (resp., strongly negative, weakly positive, weakly negative) sentiment on t_j. Then, we averaged the values obtained for each graph type. In Table 7, we report the corresponding results.

Table 7. Average values of density and average clustering coefficient for the graphs of type $\mathcal{SG}^+$, $\mathcal{SG}^-$, $\mathcal{WG}^+$ and $\mathcal{WG}^-$.

Property	$\mathcal{SG}^+$	$\mathcal{SG}^-$	$\mathcal{WG}^+$	$\mathcal{WG}^-$
Average Density	0.0242	0.0288	0.162	0.0184
Average Clustering Coefficient	0.2215	0.2417	0.1918	0.2012

From the analysis of this table, we can see that the values of the density and the average clustering coefficient of the graph $\mathcal{SG}^+$ (resp., $\mathcal{SG}^-$, $\mathcal{WG}^+$ and $\mathcal{WG}^-$) are much higher than those of the graph $\mathcal{A}$. This can be explained by considering how the graph $\mathcal{SG}^+$ (resp., $\mathcal{SG}^-$, $\mathcal{WG}^+$ and $\mathcal{WG}^-$) is constructed. In fact, such a construction starts from a node serving as the root and gradually adds nodes belonging to the various neighborhoods of the root, along with the corresponding arcs, as long as the conditions expressed in the function $\psi^+()$ (respectively, $\psi^-()$, $\xi^+()$, $\xi^-()$) in Equation (7) (resp., (8)–(10)) are satisfied. This way of proceeding tends to favor the construction of dense and compact graphs obtained as subgraphs of the connected component of $\mathcal{A}$ on which their root node is located. When the boundary of the connected component is reached, the construction of $\mathcal{SG}^+$ (resp., $\mathcal{SG}^-$, $\mathcal{WG}^+$ and $\mathcal{WG}^-$) stops. Such a construction also tends to stop when arriving at sparse areas of the graph $\mathcal{A}$.

Another important information we can derive from examining Table 7 concerns the fact that the density of the graphs associated with strong sentiments is greater than that of the graphs associated with weak sentiments. This difference becomes much less marked if we consider the average clustering coefficient instead of the density. In contrast, there is no great difference between the parameters of the graphs associated with positive sentiments and those of the graphs associated with negative sentiments. This result, coupled with the ones obtained in the previous sections, suggests to us that the negativity of a sentiment is able to increase the intensity of its transmission but it is not able to increase, except marginally, the number of connections activated by users for its transmission.

5.5. Experiments on Temporal Scope

5.5.1. Variation of the Scope over Time for Each Sentiment Type

This test is dual to the one we conducted for the spatial scope in Section 5.4.1. In fact, it aims to evaluate the trend of the sentiment degree over time and how it differs for different sentiment types. The time interval we considered is the reference interval for our dataset, which is the interval from 25 February 2022 to 25 March 2022.

In Section 4.3, we have seen that, given a user u_i and a topic t_j, the temporal scope of u_i on t_j in the time interval $T[x..y]$ is represented by an ordered list of pairs (see Equation (18)), one for each time slice in the interval. The generic pair (τ_b, θ_b) denotes the sentiment type (τ_b) and the sentiment degree (θ_b). Recall that our model associates only one sentiment type with a user u_i and a topic t_j in a time slice T_b. Both values can vary when passing from one time slice to another.

We carried out this experiment as follows: given a time slice T_b (which coincided in practice with a day of the time interval relative to our dataset), we identified all possible pairs (u_i, t_j) such that, in the time interval T_b, the user u_i expressed a sentiment on the topic t_j. Then, for each of these pairs, we determined the sentiment type τ_{ij_b} expressed by u_i on t_j in T_b and the corresponding sentiment degree θ_{ij_b}.

At this point, we partitioned the pairs (u_i, t_j) based on the corresponding sentiment types in T_b and, for each partition, we computed the average value of the sentiment degree. In this way, we obtained four average values of sentiment degree, i.e., $\overline{\theta_b^{sn}}$, $\overline{\theta_b^{sp}}$, $\overline{\theta_b^{wn}}$ and $\overline{\theta_b^{wp}}$, one for each sentiment type. Finally, we repeated these tasks for each time slice of the considered interval. The results obtained are shown in Figure 4, while in Table 8 we report the values of some statistical measures computed over the whole time period of interest for the four cases under consideration.
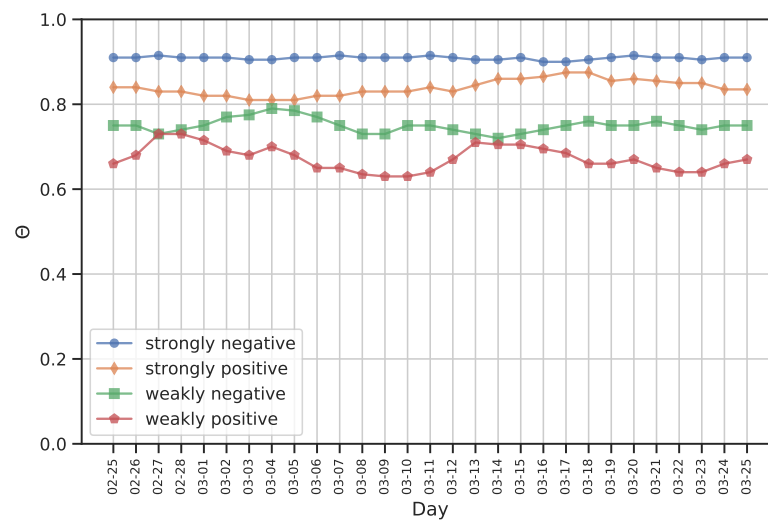

Figure 4. Variation over time of the average value of the sentiment degree associated with each sentiment type.

Table 8. Values of some statistic measures computed over the whole time period for *sn*, *sp*, *wn*, and *wp*.

Parameter	*sn*	*sp*	*wn*	*wp*
Max	0.92	0.88	0.79	0.73
Min	0.90	0.81	0.72	0.63
Mean	0.91	0.84	0.75	0.67
Standard deviation	0.003	0.018	0.017	0.029

From the analysis of Figure 4 and Table 8 we can derive some interesting knowledge patterns on temporal scope. First, we observe that the values of the average sentiment degree are generally very high, since they range from a maximum of 0.92 to a minimum of 0.63.

In addition, we can observe that the trend of the average sentiment degree for strong sentiments is generally higher than that for the corresponding weak sentiments. In fact, Table 8 shows that the average sentiment degree is equal to 0.91 and 0.84 for strong sentiments, while it is equal to 0.75 and 0.67 for weak ones. This confirms what we had already found in Section 5.4.1 for spatial scope. In addition to this, we can observe that the trend over time for strong sentiments is more constant than for weak ones. In fact, in Table 8, we can see that the standard deviation of the sentiment degree is equal to 0.003 and 0.018 for strong sentiments, while it is equal to 0.017 and 0.029 for weak ones. This is a

new knowledge pattern about the trend of sentiment degree that we were able to obtain thanks to the introduction of temporal scope in this paper. It is in line with the previous results in the literature regarding strong and weak sentiments [71,73,74]. It can be explained by considering that strong sentiments correspond to very marked polarizations, and thus are unlikely to change over time, contrary to what happens for weak sentiments.

A second interesting result that can be observed from Figure 4 and Table 8 concerns the trends of negative versus positive sentiments. In fact, we can observe that the values of negative sentiment degrees are on average higher than those of positive sentiment degrees. As evidence of this, Table 8 shows that the average sentiment degree is equal to 0.91 and 0.75 for negative sentiments, while it is equal to 0.84 and 0.67 for positive ones. This represents a confirmation of the results already found for the spatial scope in Section 5.4.1. In addition to this, we can observe that the time trend for negative sentiments is more constant than that for the corresponding positive sentiments. In fact, in Table 8 the standard deviation of the sentiment degree is equal to 0.003 and 0.017 for negative sentiments, while it is equal to 0.018 and 0.029 for positive ones. The latter knowledge pattern is new to the literature and could only be extracted due to the introduction of temporal scope. It is in line with the previous results found in the literature regarding positive and negative sentiments [68–72].

By integrating all the derived information, it follows that the strongest and most stable sentiment is the strongly negative one; it is unlikely to be changed over time. In contrast, the ficklest sentiment is the weakly positive one. Indeed, it can be modified over time by acting appropriately on users. As for the modification possibility, the strongly positive and the weakly negative sentiments lie somewhere between the two extremes.

5.5.2. Analysis of User Stereotypes

In the previous section, we focused on the temporal variation of the average values of sentiment degree. Instead, in this section, we want to analyze the temporal variation of the sentiment degree of single users on specific topics. In particular, we want to define some user stereotypes and check whether and to what extent they are present in our dataset. More specifically, the stereotypes we define are reported in Table 9. It is worth pointing out that these are stereotypes defined by us taking into consideration the semantics of the various sentiment types and the potential usefulness of them. However, new stereotypes may be defined in the future, should the need arise.

Table 9. Some possible user stereotypes.

User Stereotype	Definition
sp-user (strongly positive user) on t_j	This is a user who always showed a sentiment of type sp on t_j when she expressed her opinions during the time interval $T[x..y]$.
sn-user (strongly negative user) on t_j	Similar to the sp-user but with sn instead of sp.
wp-user (weakly positive user) on t_j	Similar to the sp-user but with wp instead of sp.
wn-user (weakly negative user) on t_j	Similar to the sp-user but with wn instead of sp.
nn-user (non-negative user) on t_j	Similar to the sp-user but with sp or wp instead of sp (Recall that a user can show only one sentiment type on a topic t_j in a time slice; however, she can show different sentiments on t_j in different time slices of $T[x..y]$).
np-user (non-positive user) on t_j	Similar to the sp-user but with sn or wn instead of sp.
w-user (weak user) on t_j	Similar to the sp-user but with wp or wn instead of sp.

Table 9. *Cont.*

User Stereotype	Definition
s-user (strong user) on t_j	Similar to the sp-user but with sp or sn instead of sp.
super-sp-user (super strongly positive user)	This is a user who always showed a sentiment of type sp on all the topics she discussed during $T[x..y]$.
super-sn-user (super strongly negative user)	Similar to the super-sp-user but with sn instead of sp.
super-wp-user (super weakly positive user)	Similar to the super-sp-user but with wp instead of sp.
super-wn-user (super weakly negative user)	Similar to the super-sp-user but with wn instead of sp.
super-nn-user (super non-negative user)	Similar to the super-sp-user but with sp or wp instead of sp.
super-np-user (super non-positive user)	Similar to the super-sp-user but with sn or wn instead of sp.
super-w-user (super weak user)	Similar to the super-sp-user but with wp or wn instead of sp.
super-s-user (super strong user)	Similar to the super-sp-user but with sp or sn instead of sp.
sw-user (swinging user) on t_j	This is a user who showed all the four sentiment types on t_j during $T[x..y]$.
ss-user (super swinging user)	This is a user who behaved as a sw-user on every topic she discussed during $T[x..y]$.
p-user (posed user)	This is a user who was sp-user for at least one topic, sn-user for at least a second topic, wp-user for at least a third topic, and wn-user for at least a fourth topic. In other words, she demonstrated the ability to express the full range of sentiments depending on the topic.

After defining stereotypes, we computed how many users in our dataset could be associated with each of them. Recall that our dataset includes 4219 users and 101 topics. The number of potential pairs (u_i, t_j), such that u_i is a user and t_j is a topic, is 426,119, while the number of actual pairs in the dataset is 130,794. The number of users associated with the various stereotypes that we defined is reported in Table 10.

From the analysis of this table, we can deduce some interesting insights. First, we observe that: *(i)* the number of sn-users (resp., super-sn-users) is greater than the one of sp-users (resp., super-sp-users); *(ii)* the number of wn-users (resp., super-wn-users) is greater than the one of wp-users (resp., super-wp-users); *(iii)* the number of np-users (resp., super-np-users) is greater than the one of nn-users (resp., super-nn-users). This is in line with what we have seen in the previous sections and in the literature regarding the trend of positive and negative user sentiments. Similarly, we can observe that: *(i)* the number of sn-users (resp., super-sn-users) is greater than the one of wn-users (resp., super-wn-users); *(ii)* the number of sp-users (resp., super-sp-users) is greater than the one of wp-users (resp., super-wp-users). This result is also in line with what we have seen in the previous sections and in the literature regarding strong and weak sentiments.

On the other hand, it is interesting to note that the number of w-users (resp., super-w-users) is greater than the one of s-users (resp., super-s-users). This might seem a contradiction to the previous results in this section and to the ones in the previous sections. In fact, this is not the case; this phenomenon can be explained by taking into account that the sentiments wp and wn are somewhat "contiguous". Therefore, it is easier for a user to switch from one to the other without ever reaching strong sentiments. In contrast, the sentiments sp and sn are extreme; to be s-user or super-s-user, one could have to oscillate between these two extreme sentiments without ever going through the weak sentiments that lie in between. This is much more difficult than a context in which a user oscillates between two weak sentiments.

A further observation concerns the very low number of swinging users. This is explained by the fact that it is really difficult for a user to have four different sentiment types on the same topic. Even the number of super-s-users is so low that we can assume that their presence is more of a bias than anything else.

Finally, it is worth pointing out that more than half of the users in our dataset are p-users. In our opinion, this is an extremely positive result because it tells us that the users in our dataset were really able to express the full range of possible sentiments depending on the reference topic and time slice.

Table 10. Number of users associated with each stereotype.

Stereotype	**Number of Users**
sp-user (strongly positive user) on t_j	1211
sn-user (strongly negative user) on t_j	1274
wp-user (weakly positive user) on t_j	1058
wn-user (weakly negative user) on t_j	1142
nn-user (non-negative user) on t_j	2119
np-user (non-positive user) on t_j	2497
w-user (weak user) on t_j	1714
s-user (strong user) on t_j	1134
super-sp-user (super strongly positive user)	72
super-sn-user (super strongly negative user)	88
super-wp-user (super weakly positive user)	48
super-wn-user (super weakly negative user)	53
super-nn-user (super non-negative user)	221
super-np-user (super non-positive user)	244
super-w-user (super weak user)	174
super-s-user (super strong user)	118
sw-user (swinging user) on t_j	42
ss-user (super swinging user)	2
p-user (posed user)	2284

5.5.3. Discussion

In this section, we present a discussion regarding the framework proposed in this paper. In particular, we present a brief overview of its main strengths, limitations, and practical applications.

Regarding the first point, the main strengths of our framework are the following: (i) it defines a multi-dimensional view of the concept of scope; (ii) it can operate on any social platform as long as the messages exchanged in the latter are predominantly text-based; and (iii) it can assess the scope of user sentiment on any topic.

Our framework also has some limitations. Specifically: (i) it operates on only one social platform at a time, whereas the various platforms are currently interconnected because many users join simultaneously on multiple platforms, acting as bridges among them; (ii) it can currently handle only two possible dimensions, namely space and time; (iii) it is unable to evaluate and analyze the possible interference that different users may exert on a given user in defining her sentiment on a topic; (iv) it is based on text analysis and, consequently, works on social platforms where the messages exchanged are predominantly textual.

A first possible practical application of our framework involves supporting information diffusion. In fact, the knowledge of scope and its dynamics can facilitate in identifying new strategies to spread certain messages as widely as possible. A second application, dual to the previous one, consists in countering fake news. The latter, in fact, often arouse a strongly negative sentiment. Exploiting this characteristic and the concept of scope makes it possible to define an approach for identifying fake news and countering their spread.

Last but not least, one could use our framework in order to identify certain user stereotypes (think, for instance, of the swinging and posed users introduced in Section 5.5.2).

6. Conclusions

In this paper, we have proposed a framework to determine the spatial and temporal scope of the sentiment of a user on a topic in a social platform. First, we have presented the concept of scope and we have seen that it summarizes several concepts, such as centrality, reputation, and diffusion, introduced in the past Social Network Analysis literature. In fact, all these concepts represent different aspects of the concept of scope. Then, we have introduced the concept of scope of the sentiment of a user on a topic and we have defined a model capable of representing and handling a multi-dimensional view of scope. Afterwards, we have proposed a set of parameters and an approach for evaluating the spatial and the temporal scope of the sentiment of a user on a topic in a social platform. Finally, we have performed a set of tests to evaluate the proposed framework on a real dataset obtained from Reddit.

The main novelties of this paper are the following: (i) it proposes a multi-dimensional view of scope, particularizing it to space and time dimensions; (ii) it introduces the concept of scope of the sentiment of a user on one or more topics; and (iii) it presents a general framework for extracting information about the scope of the sentiment of a user on topics of any subject; this framework is capable of operating on any social platform.

The main results and findings obtained by applying our framework on our Reddit dataset can be summarized as follows: (i) negative sentiments tend to spread more easily than positive ones; (ii) strong sentiments tend to spread more easily than weak ones; (iii) the influence of a user on the sentiment felt by her neighbors tends to decrease when the neighbors' distance from her increases; (iv) users with negative sentiments influence their neighbors more than users with positive sentiments; (v) users with strong sentiments influence their neighbors more than users with weak sentiments; (vi) the influence of users with strongly negative sentiment is not affected by any interference from other users; (vii) the negativity of a sentiment can increase the intensity of its transmission but cannot increase the number of connections activated by users for its transmission; (viii) the temporal trend for strong sentiments is more constant than the one for weak sentiments; (ix) the average degree of strong sentiments over time is generally higher than the average degree of weak sentiments; (x) the average degree of negative sentiments over time is generally higher than the average degree of positive sentiments; (xi) the time trend of negative sentiments is more constant than the one of positive sentiments; (xii) the number of users with negative sentiments is greater than the number of users with positive sentiments; (xiii) the number of swinging users (i.e., users who felt all the four possible sentiment types for the same topic) is negligible; and (xiv) the number of posed users (i.e., users who were capable of feeling the full range of sentiments depending on the topic) is very high.

The ideas proposed in this paper have several possible future developments. First, we plan to extend the concepts proposed here from a single network to a Social Internetworking System, that is, a set of interrelated networks in which each user may join one or more of them. Second, we would like to study further dimensions of the scope of the sentiment of a user on a topic in a social network, in addition to the spatial and temporal ones considered in this paper. Third, we would like to further study the interference of multiple users on the sentiment of a user u_i on a topic t_j. In particular, we would like to analyze the case in which the interfering users are very close, as well as the case in which they have high scope and, therefore, the interference caused by each of them may be significant or, even, decisive. Last but not least, we plan to investigate the possible use of our framework to health economics applications [75] and computational linguistics [76].

Author Contributions: Conceptualization, G.B. and L.V.; methodology, F.C. and D.U.; software, L.S.; validation, E.C. and M.M.; formal analysis, G.B. and F.C.; investigation, E.C. and D.U.; resources, M.M. and L.V.; data curation, G.B. and E.C.; writing—original draft preparation, F.C. and L.V.;

References

1. Leggio, D.; Marra, G.; Ursino, D. Defining and investigating the scope of users and hashtags in Twitter. In Proceedings of the International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2014), Amantea, Italy, 27–31 October 2014; pp. 674–681.
2. Cauteruccio, F.; Cinelli, L.; Fortino, G.; Savaglio, C.; Terracina, G.; Ursino, D.; Virgili, L. An Approach to Compute the Scope of a Social Object in a Multi-IoT Scenario. *Pervasive Mob. Comput.* **2020**, *67*, 101223. [CrossRef]
3. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the spread of influence through a social network. In Proceedings of the International ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2003), Washington, DC, USA, 24–27 August 2003; pp. 137–146.
4. Ma, Z.; Sun, A.; Cong, G. Will this #Hashtag be Popular Tomorrow? In Proceedings of the ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR 2012), Portland, OR, USA, 12–16 August 2012; pp. 1173–1174.
5. Ma, Z.; Sun, A.; Cong, G. On Predicting the Popularity of Newly Emerging Hashtags in Twitter. *J. Am. Soc. Inf. Sci. Technol.* **2013**, *64*, 1399–1410. [CrossRef]
6. Miller, Z.; Dickinson, B.; Deitrick, W.; Hu, W.; Wang, A.H. Twitter Spammer Detection Using Data Stream Clustering. *Inf. Sci.* **2014**, *260*, 64–73. [CrossRef]
7. Romero, D.; Galuba, W.; Asur, S.; Huberman, B. Influence and passivity in social media. In Proceedings of the International Conference on World Wide Web (WWW'11), Hyderabad, India, 28 March–1 April 2011; pp. 113–114.
8. Weng, J.; Lim, E.; Jiang, J.; He, Q. TwitterRank: Finding Topic-sensitive Influential Twitterers. In Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM 2010), New York, NY, USA, 3–6 February 2010; pp. 261–270.
9. Cataldi, M.; Caro, L.D.; Schifanella, C. Emerging Topic Detection on Twitter Based on Temporal and Social Terms Evaluation. In Proceedings of the International Workshop on Multimedia Data Mining (MDMKDD 2010), Washington, DC, USA, 25 July 2010; pp. 4–13.
10. Qasem, Z.; Jansen, M.; Hecking, T.; Hoppe, H. On the detection of influential actors in social media. In Proceedings of the International Conference on Signal-Image Technology & Internet-Based Systems (SITIS'15), Sorrento, Italy, 26–29 November 2015; pp. 421–427.
11. Yue, L.; Chen, W.; Li, X.; Zuo, W.; Yin, M. A survey of sentiment analysis in social media. *Knowl. Inf. Syst.* **2019**, *60*, 617–663. [CrossRef]
12. Pozzi, F.A.; Fersini, E.; Messina, E.; Liu, B. Challenges of sentiment analysis in social networks: An overview. *Sentim. Anal. Soc. Netw.* **2017**, 1–11. [CrossRef]
13. Yadav, A.; Vishwakarma, D. Sentiment analysis using deep learning architectures: A review. *Artif. Intell. Rev.* **2020**, *53*, 4335–4385. [CrossRef]
14. Birjali, M.; Kasri, M.; Beni-Hssane, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowl.-Based Syst.* **2021**, *226*, 107134. [CrossRef]
15. Cortis, K.; Davis, B. Over a decade of social opinion mining: A systematic review. *Artif. Intell. Rev.* **2021**, *54*, 4873–4965. [CrossRef]
16. Basile, V.; Cauteruccio, F.; Terracina, G. How dramatic events can affect emotionality in social posting: The impact of COVID-19 on reddit. *Future Internet* **2021**, *13*, 29. [CrossRef]
17. Lai, M.; Tambuscio, M.; Patti, V.; Ruffo, G.; Rosso, P. Stance polarity in political debates: A diachronic perspective of network homophily and conversations on Twitter. *Data Knowl. Eng.* **2019**, *124*, 101738. [CrossRef]
18. Ramachandran, D.; Parvathi, R. A novel domain and event adaptive tweet augmentation approach for enhancing the classification of crisis related tweets. *Data Knowl. Eng.* **2021**, *135*, 101913. [CrossRef]
19. Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.; Jian, X.; Li, Y.; Zhao, L. Latent Dirichlet Allocation (LDA) and topic modeling: Models, applications, a survey. *Multimed. Tools Appl.* **2019**, *78*, 15169–15211. [CrossRef]
20. Vayansky, I.; Kumar, S. A review of topic modeling methods. *Inf. Syst.* **2020**, *94*, 101582. [CrossRef]
21. Qiang, J.; Qian, Z.; Li, Y.; Yuan, Y.; Wu, X. Short Text Topic Modeling Techniques, Applications, and Performance: A Survey. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 1427–1445. [CrossRef]
22. Ravi, K.; Ravi, V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowl.-Based Syst.* **2015**, *89*, 14–46. [CrossRef]

23. Tsvetovat, M.; Kouznetsov, A. *Social Network Analysis for Startups: Finding Connections on the Social Web*; O'Reilly Media, Inc.: Newton, MA, USA, 2011.
24. Moazzami, D. Toughness of the Networks with Maximum Connectivity. *J. Algorithms Comput.* **2015**, *46*, 51–71.
25. Khoshnood, A.; Moazzami, D. A Survey on Tenacity Parameter—Part I. *J. Algorithms Comput.* **2021**, *53*, 181–196.
26. Moazzami, D.; Khoshnood, A. A Survey on Tenacity Parameter—Part II. *J. Algorithms Comput.* **2022**, *54*, 47–72.
27. Bonchi, F.; Castillo, C.; Gionis, A.; Jaimes, A. Social network analysis and mining for business applications. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 1–37. [CrossRef]
28. Scott, J. Social network analysis: Developments, advances, and prospects. *Soc. Netw. Anal. Min.* **2011**, *1*, 21–26. [CrossRef]
29. Cantini, R.; Marozzo, F.; Talia, D.; Trunfio, P. Analyzing political polarization on social media by deleting bot spamming. *Big Data Cogn. Comput.* **2022**, *6*, 3. [CrossRef]
30. Bayrakdar, S.; Yucedag, I.; Simsek, M.; Dogru, I.A. Semantic analysis on social networks: A survey. *Int. J. Commun. Syst.* **2020**, *33*, e4424. [CrossRef]
31. Pankong, N.; Prakancharoen, S.; Buranarach, M. A combined semantic social network analysis framework to integrate social media data. In Proceedings of the International Conference on Knowledge and Smart Technology (KST'12), Chonburi, Thailand, 7–8 July 2012; pp. 37–42.
32. Xia, Z.; Bu, Z. Community detection based on a semantic network. *Knowl.-Based Syst.* **2012**, *26*, 30–39. [CrossRef]
33. Ismail, H.; Khalil, A.; Hussein, N.; Elabyad, R. Triggers and Tweets: Implicit Aspect-Based Sentiment and Emotion Analysis of Community Chatter Relevant to Education Post-COVID-19. *Big Data Cogn. Comput.* **2022**, *6*, 99. [CrossRef]
34. Yeasmin, N.; Mahbub, N.I.; Baowaly, M.K.; Singh, B.C.; Alom, Z.; Aung, Z.; Azim, M.A. Analysis and Prediction of User Sentiment on COVID-19 Pandemic Using Tweets. *Big Data Cogn. Comput.* **2022**, *6*, 65. [CrossRef]
35. Poulopoulos, V.; Wallace, M. Social Media Analytics as a Tool for Cultural Spaces—The Case of Twitter Trending Topics. *Big Data Cogn. Comput.* **2022**, *6*, 63. [CrossRef]
36. Gallacher, J.; Bright, J. Hate Contagion: Measuring the spread and trajectory of hate on social media. *PsyArXiv* **2021**. [CrossRef]
37. Yin, F.; Xia, X.; Pan, Y.; She, Y.; Feng, X.; Wu, J. Sentiment mutation and negative emotion contagion dynamics in social media: A case study on the Chinese Sina Microblog. *Inf. Sci.* **2022**, *594*, 118–135. [CrossRef]
38. Pröllochs, N.; Bär, D.; Feuerriegel, S. Emotions explain differences in the diffusion of true vs. false social media rumors. *Sci. Rep.* **2021**, *11*, 22721. [CrossRef]
39. Almars, A.; Li, X.; Zhao, X. Modelling user attitudes using hierarchical sentiment-topic model. *Data Knowl. Eng.* **2019**, *119*, 139–149. [CrossRef]
40. Yang, Z.; Kotov, A.; Mohan, A.; Lu, S. Parametric and non-parametric user-aware sentiment topic models. In Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'15), Santiago, Chile, 9–13 August 2015; pp. 413–422.
41. Naskar, D.; Mokaddem, S.; Rebollo, M.; Onaindia, E. Sentiment analysis in social networks through topic modeling. In Proceedings of the International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia, 23–28 May 2016; pp. 46–53.
42. Liu, B.; Zhang, L. A survey of opinion mining and sentiment analysis. In *Mining Text Data*; Springer: Boston, MA, USA, 2012; pp. 415–463.
43. Neves-Silva, R.; Gamito, M.; Pina, P.; Campos, A.R. Modelling influence and reach in sentiment analysis. *Procedia CIRP* **2016**, *47*, 48–53. [CrossRef]
44. Carvalho, J.; Rosa, H.; Brogueira, G.; Batista, F. MISNIS: An intelligent platform for twitter topic mining. *Expert Syst. Appl.* **2017**, *89*, 374–388. [CrossRef]
45. Ferrara, E.; Yang, Z. Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Comput. Sci.* **2015**, *1*, e26. [CrossRef]
46. Zhao, K.; Yen, J.; Greer, G.; Qiu, B.; Mitra, P.; Portier, K. Finding influential users of online health communities: A new metric based on sentiment influence. *J. Am. Med. Inform. Assoc.* **2014**, *21*, e212–e218. [CrossRef]
47. Cao, N.; Lu, L.; Lin, Y.; Wang, F.; Wen, Z. Socialhelix: Visual analysis of sentiment divergence in social media. *J. Vis.* **2015**, *18*, 221–235. [CrossRef]
48. Kušen, E.; Strembeck, M.; Cascavilla, G.; Conti, M. On the influence of emotional valence shifts on the spread of information in social networks. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '17), Sydney Australia, 31 July–3 August 2017; pp. 321–324.
49. Zafarani, R.; Cole, W.D.; Liu, H. Sentiment propagation in social networks: A case study in livejournal. In Proceedings of the International Conference on Social Computing, Behavioral Modeling, and Prediction (SBP'10), Bethesda, MD, USA, 29 March–1 April 2010; pp. 413–420.
50. Melton, C.A.; Olusanya, O.A.; Ammar, N.; Shaban-Nejad, A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *J. Infect. Public Health* **2021**, *14*, 1505–1512. [CrossRef]
51. An, L.; Zhou, W.; Ou, M.; Li, G.; Yu, C.; Wang, X. Measuring and profiling the topical influence and sentiment contagion of public event stakeholders. *Int. J. Inf. Manag.* **2021**, *58*, 102327. [CrossRef]

52. Cai, M.; Luo, H.; Meng, X.; Cui, Y. Topic-emotion propagation mechanism of public emergencies in social networks. *Sensors* **2021**, *21*, 4516. [CrossRef]

53. Cai, M.; Luo, H.; Meng, X.; Cui, Y. A Study on the Topic-Sentiment Evolution and Diffusion in Time Series of Public Opinion Derived from Emergencies. *Complexity* **2021**, *2021*, 2069010. [CrossRef]

54. Xu, Y.; Li, Y.; Liang, Y.; Cai, L. Topic-sentiment evolution over time: A manifold learning-based model for online news. *J. Intell. Inf. Syst.* **2020**, *55*, 27–49. [CrossRef]

55. Wang, X.; Jin, D.; Musial, K.; Dang, J. Topic enhanced sentiment spreading model in social networks considering user interest. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'20), New York, NY, USA, 7–12 February 2020; Volume 34, pp. 989–996.

56. Tsugawa, S.; Ohsaki, H. Negative messages spread rapidly and widely on social media. In Proceedings of the International Conference on Online Social Networks (COSN'15), Palo Alto, CA, USA, 2–3 November 2015; pp. 151–160.

57. Heimbach, I.; Hinz, O. The impact of content sentiment and emotionality on content virality. *Int. J. Res. Mark.* **2016**, *33*, 695–701. [CrossRef]

58. Majumder, N.; Poria, S.; Peng, H.; Chhaya, N.; Cambria, E.; Gelbukh, A. Sentiment and Sarcasm Classification With Multitask Learning. *IEEE Intell. Syst.* **2019**, *34*, 38–43. [CrossRef]

59. Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; Blackburn, J. The pushshift Reddit dataset. In Proceedings of the International AAAI Conference on Web and Social Media (ICWSM'20), Atlanta, Georgia, USA, 8–11 June 2020; Volume 14, pp. 830–839.

60. Mills, R. Reddit.Com: A Census of Subreddits. In Proceedings of the International Web Science Conference (WebSci'15), Oxford, UK, 28 June–1 July 2015; pp. 49:1–49:2.

61. Guimaraes, A.; Balalau, O.; Terolli, E.; Weikum, G. Analyzing the Traits and Anomalies of Political Discussions on Reddit. In Proceedings of the International Conference on Web and Social Media (ICWSM 2019), Münich, Germany, 11–14 June 2019; pp. 205–213.

62. Horne, B.; Adali, S. The impact of crowds on news engagement: A reddit case study. In Proceedings of the International AAAI Conference on Web and Social Media (ICWSM'17), Montreal, QC, Canada, 15–18 May 2017; p. 11.

63. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.

64. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is All you Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]

65. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API design for machine learning software experiences from the scikit-learn project. In Proceedings of the European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases (ECMP/PKDD 2013), Prague, Czech Republic, 23–27 September 2013; pp. 108–122.

66. Loureiro, D.; Barbieri, F.; Neves, L.; Anke, L.; Camacho-collados, J. TimeLMs: Diachronic Language Models from Twitter. In Proceedings of the Annual Meeting of the Association for Computational Linguistics: System Demonstrations (ACL'22), Dublin, Ireland, 22–27 May 2022; pp. 251–260.

67. Barbieri, F.; Camacho-Collados, J.; Anke, L.E.; Neves, L. TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. In Proceedings of the Findings of the Association for Computational Linguistics (EMNLP'20), Online, 16–20 November 2020; pp. 1644–1650.

68. Yu, H.; Yang, C.; Yu, P.; Liu, K. Emotion diffusion effect: Negative sentiment COVID-19 tweets of public organizations attract more responses from followers. *PLoS ONE* **2022**, *17*, e0264794. [CrossRef]

69. Schöne, J.P.; Parkinson, B.; Goldenberg, A. Negativity spreads more than positivity on Twitter after both positive and negative political situations. *Affect. Sci.* **2021**, *2*, 379–390. [CrossRef]

70. Cinelli, M.; Pelicon, A.; Mozetič, I.; Quattrociocchi, W.; Novak, P.K.; Zollo, F. Dynamics of online hate and misinformation. *Sci. Rep.* **2021**, *11*, 1–12. [CrossRef]

71. Stieglitz, S.; Dang-Xuan, L. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *J. Manag. Inf. Syst.* **2013**, *29*, 217–248. [CrossRef]

72. Suh, B.; Hong, L.; Pirolli, P.; Chi, E.H. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In Proceedings of the International Conference on Social Computing (SOCIALCOM '10), Minneapolis, MN, USA, 20–22 August 2010; pp. 177–184.

73. Rimé, B. The social sharing of emotion as an interface between individual and collective processes in the construction of emotional climates. *J. Soc. Issues* **2007**, *63*, 307–322. [CrossRef]

74. Rimé, B.; Finkenauer, C.; Luminet, O.; Zech, E.; Philippot, P. Social sharing of emotion: New evidence and new questions. *Eur. Rev. Soc. Psychol.* **1998**, *9*, 145–189. [CrossRef]

75. Shin, E. Physician Connectedness and Referral Choice. *Oxford Bulletin of Economics and Statistics*; Wiley Online Library: New York, NY, USA, 2022.

76. Ott, M.; Choi, Y.; Cardie, C.; Hancock, J. Finding deceptive opinion spam by any stretch of the imagination. *arXiv* **2011**, arXiv:1107.4557.