



Article

Modeling, Evaluating, and Applying the eWoM Power of Reddit Posts

Gianluca Bonifazi , Enrico Corradini , Domenico Ursino * and Luca Virgili

Department of Information Engineering (DII), Polytechnic University of Marche, 60121 Ancona, Italy

* Correspondence: d.ursino@univpm.it

Abstract: Electronic Word of Mouth (eWoM) has been largely studied for social platforms, such as Yelp and TripAdvisor, which are highly investigated in the context of digital marketing. However, it can also have interesting applications in other contexts. Therefore, it can be challenging to investigate this phenomenon on generic social platforms, such as Facebook, Twitter, and Reddit. In the past literature, many authors analyzed eWoM on Facebook and Twitter, whereas it was little considered in Reddit. In this paper, we focused exactly on this last platform. In particular, we first propose a model for representing and evaluating the eWoM Power of Reddit posts. Then, we illustrate two possible applications, namely the definition of lifespan templates and the construction of profiles for Reddit posts. Lifespan templates and profiles are ultimately orthogonal to each other and can be jointly employed in several applications.

Keywords: Word of Mouth; eWoM; Reddit; eWoM Power; post lifespan; post lifetime; post profiling; post categorization



Citation: Bonifazi, G.; Corradini, E.; Ursino, D.; Virgili, L. Modeling, Evaluating, and Applying the eWoM Power of Reddit Posts. *Big Data Cogn. Comput.* **2023**, *7*, 47. <https://doi.org/10.3390/bdcc7010047>

Academic Editors: Vincenzo Moscato and Giancarlo Sperli

Received: 13 January 2023

Revised: 2 March 2023

Accepted: 7 March 2023

Published: 9 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the Cambridge Business English Dictionary (<https://dictionary.cambridge.org>, accessed on 12 December 2022), Word of Mouth (hereafter, WoM) is defined as “the process of telling people you know about a particular product or service, usually because you think it is good and want to encourage them to try it”. The previous definition links WoM mainly to the business world, in particular to the marketing scenario, and many papers have followed this track in the past literature [1]. Therefore, it is not surprising that the electronic WoM (eWoM) phenomenon has been particularly studied for business-oriented social platforms, such as Yelp (<https://www.yelp.com>, accessed on 12 December 2022) and TripAdvisor (<https://www.tripadvisor.com>, accessed on 12 December 2022) [2–6]. However, eWoM can have a scope that goes beyond business (for example, it may concern news, a political event, a scientific discovery, etc.), and therefore, it can also be studied for more generalist social platforms, such as Facebook (<https://www.facebook.com>, accessed on 12 December 2022), Twitter (<https://www.twitter.com>, accessed on 12 December 2022), and Reddit (<https://www.reddit.com>, accessed on 12 December 2022). In the past, several authors investigated eWoM for the first two of these social platforms [7–12], whereas eWoM for Reddit has been little investigated [13–16].

Actually, we argue that the investigation of eWoM for Reddit could provide new interesting perspectives on this phenomenon, captured thanks to some specificities of this social network. Indeed, Reddit has some peculiarities with respect to Facebook and Twitter [13,17,18]. These two social networks prompt the majority of users to publish their posts very frequently and “on the fly” for several reasons, like the limits on the length of both posts and replies. Instead, Reddit users may write very long posts, which stimulates them to express their ideas in detail. Furthermore, it is possible to write long comments to a post. Both features lead users to not express their ideas “on the fly”, but with “a cool head” on this social network. From the eWoM perspective, this represents a very relevant

difference with respect to Facebook and Twitter. In fact, thanks to this feature, eWoM could have a higher quality because every idea expressed and every answer published are not the result of a passing emotion, but of a reasoned thought. In this paper, we focused exactly on the eWoM Power of a Reddit post, i.e., the potential of a post published on this social medium to be spread and known. Clearly, the investigation we propose cannot be exhaustive, but it is meant to be the first step in that direction. Our ultimate goal is to raise interest so that other authors feel stimulated to continue this investigation, going deeper and deeper. Our aims in the paper were two, namely: (i) proposing an approach for the evaluation of the eWoM Power of a post; (ii) exploiting the eWoM Power for addressing two further issues of Reddit posts, namely the definition of lifespan templates and the construction of profiles.

Our approach assumed that each post, when published, has an initial eWoM Power that can be evaluated based on some parameters. This initial value does not remain constant, but, rather, changes continuously over time. In fact, several parameters can influence it and make it grow. In the meantime, the “time” tends to make the post obsolete and its eWoM Power decrease. The influence of time in the definition of eWoM Power is extremely important, but also complex to investigate. For this reason, we propose two eWoM Power definitions. The first, which we call *naive eWoM Power*, is simpler, because it does not take the “time” factor into account. However, it considers all the other parameters that may affect the value of eWoM Power. As we will see below, these parameters are the number of comments received by the post, the number of times it has been reposted, its score, its possible awards, the karma and the seniority of the author who published it, and the number of subscribers of the subreddit where it was published. The second one, which we call *refined eWoM Power*, starts from the naive eWoM Power, but is more complete and complex. In fact, it also takes the “time” factor into account and, therefore, allows us to make a dynamic evaluation of eWoM Power.

The seven factors that affect the naive eWoM Power, and consequently also the refined one, can remain constant or grow over time. Their growth leads to an increase of the refined eWoM Power. This increase is contrasted by the “time” factor; in fact, as time passes, the post content becomes obsolete [19] and its eWoM Power decreases. Depending on how the seven factors of naive eWoM Power (and their consequent ability to contrast the effects of the “time” factor) vary over time, we have different lifespans of Reddit posts. As a second contribution of this paper, we performed an investigation on this topic to understand if there are lifespan templates, i.e., typical lifespans that can be attributed to the vast majority of posts. As we will see below, we were able to identify six lifespan templates that respond to these characteristics. Each of them is extremely interesting because it traces a possible path that a new post could take, based on the initial value of the refined eWoM Power and its behavior over time.

In addition to this, eWoM Power can play a key role in the effort of profiling Reddit posts. In fact, as a third contribution of this paper, we propose an approach that uses eWoM Power and lifetime to define a set of post profiles. As we will see below, even if our profiling activity starts from the post categorization based on lifespan templates, in the end, we verified that the same profile, under specific circumstances, can be associated with posts related to different lifespan templates. This implies that lifespan templates and profiles are two orthogonal tools categorizing posts.

The knowledge extracted in this paper has several applications. First, the categorization of posts resulting from lifespan templates and profiles allows us to determine the main features of successful posts. This provides useful guidance on how a post should be made in order to reach a high eWoM Power and, therefore, be successful. Of course, this is crucial in a variety of applications, from digital marketing to spreading political and social messages and opinions, from finding the right influencers to convey a particular message, to studying the costs and benefits of using a specific social network (in this case, Reddit) to spread a certain message.

This paper is structured as follows: In Section 2, we examine the related literature. In Section 3, we illustrate the dataset that we used for our tests. In Section 4, we present our approach to evaluate the naive and the refined eWoM Power of Reddit posts. In Section 5, we describe its usage for both the construction of lifespan templates and the definition of possible post profiles. In Section 6, we present a retrospective of the approach proposed. Finally, in Section 7, we draw our conclusions and discuss some possible future developments.

2. Related Literature

Word of Mouth (WoM) is one of the most-ancient mechanisms to convey information in the history of human society [20]. Over the years, it has been defined in many ways: In [21], it is denoted as a person-to-person communication tool between a communicator and a receiver. In [22], it is described as the exchanging of marketing information between consumers. In [23], it is more generally defined as consumer-to-consumer communication. Thanks to the growth of online communication on social media, websites, and blogs [24], a variant of WoM, i.e., electronic WoM (eWoM), is becoming increasingly used and investigated [25,26]. There are several differences between WoM and eWoM [27]. The first one regards credibility; in fact, in WoM, the information receiver generally knows very well the communicator and trusts her/him; instead, in eWoM, the two actors know each other less or do not know each other at all. A second difference concerns privacy; in fact, in WoM, conversation is almost always private, interpersonal, and conducted in real-time, whereas, in eWoM, the shared information is not private and, sometimes, can be viewed by anyone at any time. A final difference regards diffusion speed; in fact, in WoM, messages spread more slowly than in eWoM, where information travels more quickly among users thanks to the Internet.

The appearance of Online Social Networks, such as Yelp, TripAdvisor, Facebook, Twitter, Reddit, etc., has led companies and corporations to monitor the eWoM about their brands [28–32]. For instance, in [33], the authors apply the ego-network analysis in a travel-related eWoM scenario to outline the strength of social ties between the communicator and the receiver; furthermore, they apply the whole-network analysis to investigate the network structure of eWoM communication. In [34], the authors examine whether personality traits of online users accentuate or attenuate the effectiveness of eWoM on social media platforms; they also analyze how companies can leverage these insights for improving their social media advertising. Another interesting field where eWoM is deeply studied by researchers is reputation platforms, such as Yelp or TripAdvisor. For instance, in [2], the author uses speech code theory to explain and evaluate how computer users communicate by posting reviews on Yelp. On the other hand, in [3], the authors examine how eWoM and the content of online reviews, including restaurant reviews on TripAdvisor and Yelp, increase helpfulness perceptions. The authors of [4] exploit TripAdvisor reviews to analyze five factors on hotel management reputation that can impact eWoM perceived credibility and adoption. These factors are eWoM volume, source credibility, rate extremism, consumer involvement, and perceived eWoM credibility. Finally, in [5,6], the authors study why a traveler trusts TripAdvisor reviews; furthermore, they profile people who post helpful reviews in the online travel community.

Researchers have put much effort into understanding how the interest raised by a post decreases over time [35–41]. Indeed, as we can experience every day on any social network, some posts tend to lose interest quickly, while others keep up the attention of users for a longer period of time [42]. In [36], the authors present an approach to predict the lifespans of popular tweets based on their static characteristics and dynamic retweeting patterns. For each tweet, they generated a time series based on the first-hour retweeting information and then compared it with those of historic tweets of the same author and time. They tested their approach on a three-month real dataset from Tencent blog. Starting from this last work, in [35], the authors propose a model to predict the lifespan and retweet time of tweets, with the ultimate goal of measuring their popularity. For this purpose, they collected tweets

from June to October 2012 and extracted data regarding posting times, content, and social features (such as followers, user reliability, tweet informativeness, etc.). In [40], the authors investigate how much time a Facebook post continues to engage users. For this purpose, they followed the posts of 100 participants for 24 h and observed that, on average, these posts received the largest number of social engagements within the first 2 to 4 h and rarely after 4 h. In [37], the authors propose an approach to measure the popularity of online posts concerning specific topics to discover potential drug abusers. They claimed that identifying popular posts online can help to prevent future crimes. To evaluate the popularity of posts, they mainly considered the lifespan and comment frequency pattern of posts; starting from these factors, they propose four general models to compute the popularity of posts in different social platforms.

In the literature, few works investigate and categorize posts through their lifespan [43,44]. Specifically, in [43], the authors traced the lifecycle of 17 popular political rumors that circulated on Twitter over 13 months during a U.S. presidential election. They figured out that false rumors (i.e., misinformation) tend to come back multiple times after their initial publication, while true rumors (i.e., facts) do not. They observed that rumors revive, sometimes with textual changes, until the tension around the corresponding topic dissolves. In [44], the authors propose an approach that employs the frequent pattern mining method to extract daily occurring frequent patterns, used to extract events such as elections, economic crises, emergencies, and so on. Finally, we point out that several studies have investigated the user posting behavior on Reddit. For instance, in [45], the authors study anxiety disorders through personal narratives expressed on Reddit. In this way, they extend to this social platform several analyses carried out on Twitter. In [46], the authors explore user posting behavior on Reddit and demonstrated that the “answer-person” role is present on this social medium. They also showed that Reddit users rarely exhibit significant participation in more than one community.

Our approach has several differences compared to those presented above. First of all, to the best of our knowledge, it is the first one to study eWoM on Reddit. From this point of view, our paper represents an advancement in the eWoM literature. Furthermore, the formulas for the evaluation of eWoM Power we propose here were designed to take into account all possible parameters of this social medium that may have an influence on it. As a consequence, they are tailored to this social platform, even if the ideas underlying them could be applied to other generic platforms, such as Facebook and Twitter. Our approach also couples the study of eWoM Power with two important applications of it, namely the definition of post lifespan templates and the construction of post profiles. These two applications are generic and, in turn, can be the starting point for many other ones. Their presence greatly differentiates our study of eWoM from the ones proposed by other papers in the past, which investigated the ability of a message to stimulate eWoM in very specific applications related to digital marketing.

3. Dataset Description

Before describing the dataset used in this paper, we present a short overview of the Reddit structure. This social medium consists of a set of subreddits. Each subreddit contains a set of posts that can be of different types (text, images, links to external resources, etc.). Each post has a title, and users can comment on it. Our dataset was downloaded from pushshift.io (accessed on 21 November 2022) [47], one of the main data repositories providing content on Reddit. For the analyses of this paper, we considered all posts published on Reddit from 1 February 2020 to 30 April 2020. The length of the time interval we took into consideration and the consequent dataset dimension are in line with the number of posts used in the analyses of Reddit proposed in the past literature [48–51].

The total number of posts was 62,521,691; the total number of authors was 9,140,478; finally, the total number of subreddits was 909,149. The extracted fields for each post (respectively, author, subreddit) are reported in Table 1 (respectively, Tables 2 and 3).

Table 1. The fields extracted for each post.

Name	Description
post_id	The identifier of the post.
author_id	The identifier of the author of the post.
author	The username of the author of the post.
subreddit_id	The identifier of the subreddit in which the post was published.
subreddit	The name of the subreddit in which the post was published.
created_utc	The creation date of the post.
num_comments	The number of the comments received by the post.
num_crossposts	How many times the post was crossposted.
nsfw	If the post contains Not Safe For Work (NSFW) material.
score	The overall score of the post; it is computed as the sum of the positive and the negative votes obtained by the post itself.
title	The title of the post.
is_video	If the post consists of a video.
has_image	If the post contains an image.
total_awards	The number of awards received by the post; an award is an acknowledgment of a user indicating that she/her liked the post.
is_part_of_collection	If the post is part of a collection.
is_self_post	If the post contains a link outside of Reddit.

Table 2. The fields extracted for each author.

Name	Description
author_id	The identifier of the author.
author	The username of the author.
created_utc	The date of the author's registration to Reddit expressed in Coordinated Universal Time.
is_gold	The author purchased a premium subscription to Reddit.
is_suspended	If the author has been suspended.
karma	The author's karma on Reddit; the karma of an author is a score representing her/his contribution to the Reddit community, evaluated on the basis of her/his posts and comments.

Table 3. The fields extracted for each subreddit.

Name	Description
subreddit_id	The identifier of the subreddit.
subreddit	The name of the subreddit.
created_utc	The creation date of the subreddit.
description	The description of the subreddit.
over_18	If the subreddit is an NSFW one.
subscribers	The number of users subscribed to the subreddit.

4. Modeling and Evaluating the eWoM Power of Reddit Posts

4.1. Determining the Support Theoretical Tools

Our approach to evaluate the eWoM Power of Reddit posts takes into account the information about the posts, authors, and subreddits provided by Reddit through pushshift.io (accessed on 21 November 2022) and described in detail in Section 3. Clearly, not all the fields of our dataset contribute directly to the estimation of the eWoM Power of a post. Therefore, we first had to analyze the semantics of these fields to determine which of them can contribute to this estimation. At the end of this analysis, we identified the following candidate fields: (i) for posts: `num_comments`, `num_crossposts`, `score`, `total_award`, and `title` (evaluated through its sentiment polarity); (ii) for authors: `karma` and `created_utc`; (iii) for subreddits: `subscribers` and `description` (evaluated through its sentiment polarity). In order to decide whether these candidate fields can contribute to the eWoM Power

of a post and, in the affirmative case, to identify how they contribute, we studied the distribution of the posts, authors, and subreddits of our dataset against all these fields. This study is essential to define a function that models this contribution. In fact, suppose that, for a candidate field, say the score of posts, the distribution follows a power law. This distribution describes the relationship between two variables, where one variable is proportional to a power of the other one. It can be expressed as $y = k \cdot (x + \delta)^\alpha$. In this formula, the exponent α is a measure of the steepness of the power law, and so, the higher α is, the higher the steepness of the curve. Instead, δ is an offset parameter that shifts the distribution along the x-axis. When δ is positive, the distribution is shifted to the right; when δ is negative, the distribution is shifted to the left. Power law distributions are characterized by a long tail of rare events, meaning that the distribution has a high frequency of small values and a low frequency of large values. In our scenario, taking the score parameter into consideration, this means that many posts have a low score, while a few posts have a high score. Therefore, with this type of distribution, having a high score is extremely difficult for a post. On the other side, a post that achieves this goal acquires a “competitive advantage” over other posts. This advantage ultimately results in an increase of its eWoM Power.

The distributions of the posts, the authors, or the subreddits against the candidate fields are shown in Figures 1–5. From the analysis of these figures, we can observe the following:

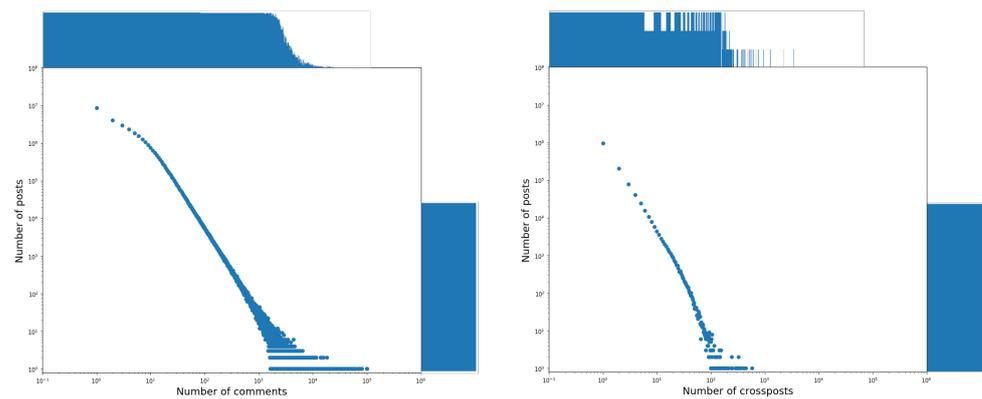


Figure 1. Distributions of posts against comments (left, log–log scale) and crossposts (right, log–log scale) enriched with marginal distributions.

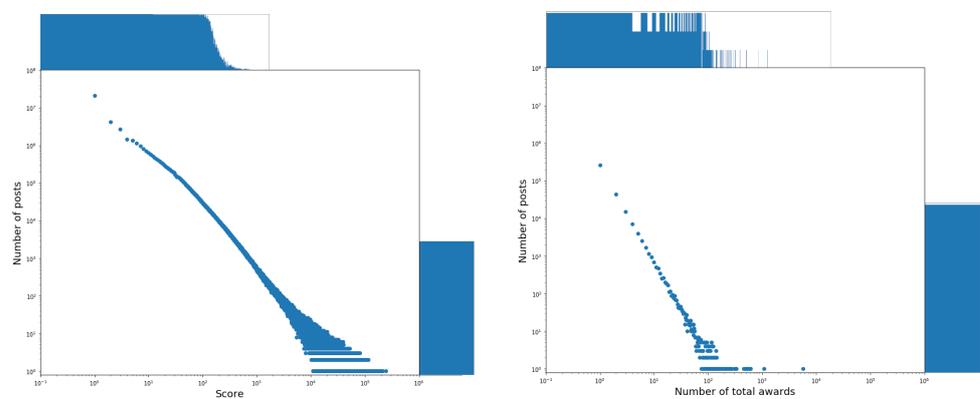


Figure 2. Distributions of posts against scores (left, log–log scale) and awards (right, log–log scale) enriched with marginal distributions.

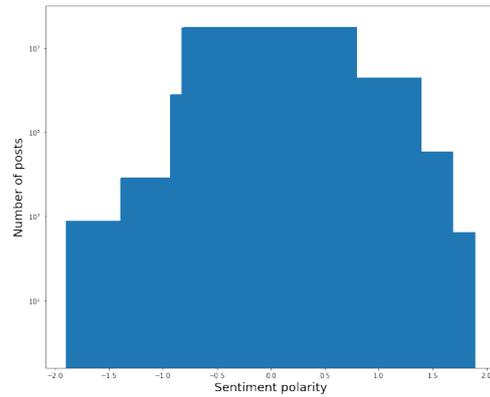


Figure 3. Distribution of posts against the sentiment polarity of their title (semi-logarithmic scale).

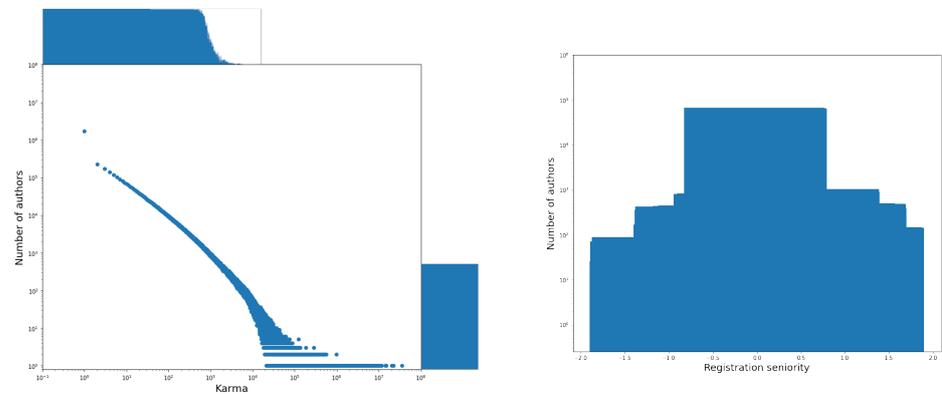


Figure 4. Distributions of authors against their karma (left, log–log scale, enriched with marginal distributions) and registration dates (right, semi-logarithmic scale).

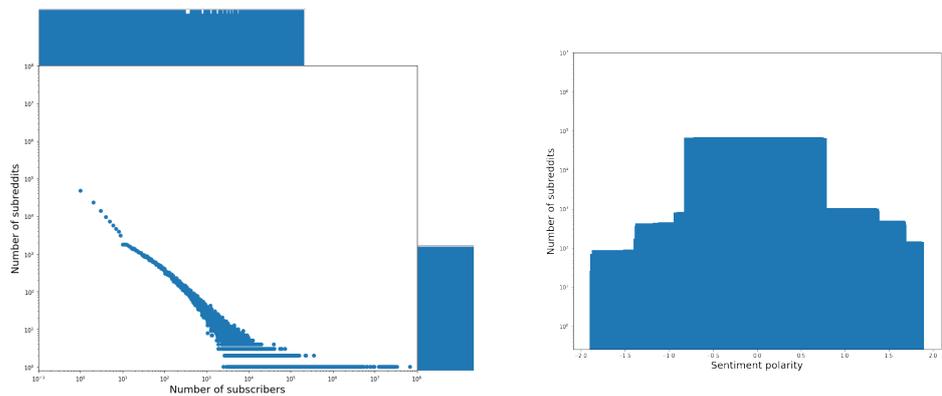


Figure 5. Distributions of subreddits against subscribers (left, log–log scale, enriched with marginal distributions) and sentiment polarity of their description (right, semi-logarithmic scale).

- The distributions of posts against the fields `num_comments`, `num_crossposts`, `total_awards`, and `score` follow a power law. The values of the parameters α and δ of these distributions, together with the maximum values of the corresponding fields in our dataset, are shown in Table 4.
- The distribution of posts against the sentiment polarity of the field `title` is centered on the value of 0. It highlights that almost all the titles of the posts (in particular, 91% of them) have a null sentiment polarity. Furthermore, the very few posts having a different behavior are distributed quite uniformly on the right and left of 0, and in any case, most of the corresponding sentiment polarity values are very close to 0. This allowed us to conclude that the sentiment polarity of the titles does not contribute

significantly to characterizing the posts and, ultimately, that the field `title` does not provide a significant contribution to the eWoM Power of a post.

- The distribution of authors against the fields `karma` and `created_utc` follows a power law (see Table 4 for the values of α and δ). The distribution of authors against `created_utc` should be interpreted as follows. On the abscissae axis, we put the registration seniority of the authors to Reddit, grouped by bimesters. As for author seniority, we point out that, in our dataset, the youngest author registered on 8 March 2020 (and, therefore, has a seniority of 0 bimesters), whereas the oldest one registered on Reddit on 19 January 2006 (and, therefore, has a seniority of 84 bimesters).
- The distribution of subreddits against the field `subscribers` follows a power law (see Table 4 for the values of α and δ).
- The distribution of subreddits against the field `description` is centered on the value of 0 of its sentiment polarity; its overall trend is analogous to the one of the distribution of posts against the field `title`. Therefore, also in this case, we can conclude that this field does not contribute to eWoM Power.

Table 4. Values of α and δ of the power law distributions and maximum values of the fields of interest for the eWoM Power computation.

Field	α	δ	Maximum Value
<code>num_comments</code>	1.425	0.012	100,002
<code>num_crossposts</code>	1.386	0.027	576
<code>score</code>	1.595	0.009	242,971
<code>total_awards</code>	1.371	0.032	5638
<code>karma</code>	1.873	0.018	35,249,636
<code>created_utc</code>	2.028	0.089	89
<code>subscribers</code>	1.862	0.022	67,092,933

In accordance with the above reasoning, we can conclude that two of the field candidates for the contribution to eWoM Power should be discarded. All the others, instead, should be selected. Interestingly, each selected field shows a power law distribution. We conclude this parameter selection activity with the consideration that we believe important because it highlights the strengths and limitations of our way of proceeding in this paper. In the Introduction, we mentioned that posts and comments on Reddit are generally made with a cool head and not impulsively. Therefore, already, the parameters (which, ultimately, reflect the frequency of posts, the seniority of their publishers on Reddit, and their polarity) are very good indicators for eWoM Power evaluation on Reddit. Actually, we are aware that a very in-depth semantic analysis of post contents could provide additional insights for an even more accurate evaluation of the eWoM Power of Reddit posts. However, as we said, this paper wants to be just a first step in this direction. We hope that it can serve as encouragement to other authors to deepen this study.

Finally, we point out that some of the parameters identified are specific to Reddit, while others are more generic and can be found to be almost identical on other social networks. Some of the specific parameters may have corresponding ones, with a similar or close semantic, on social networks such as Facebook and Twitter. This makes our approach replicable on other social networks, after having modified it appropriately. However, we want to emphasize again that the real difference between the adoption of these parameters, or similar ones, on Facebook and Twitter, on the one hand, and on Reddit, on the other hand, lies in the different way users behave on this social medium, as we discussed in the Introduction. This gives these parameters much more weight and relevance on Reddit than on platforms that encourage users to submit many short posts and comments on the fly.

In order to model the contribution that each selected field can provide to the eWoM Power of a post, it is necessary to start from the main feature of the power law distribution, according to which most of the elements involved in it have a low value and only few of them have a high value. To model the contribution to the eWoM Power of a field, say x ,

characterized by this type of distribution, we need a function that returns low values and has a slow growth against x for low values of this parameter. On the contrary, when x is higher than a certain threshold, this function must start to return higher values, and its growth against x must be faster. A function with these characteristics, well known in the literature, especially in the context of deep learning, is the Rectified Linear Unit (ReLU) function. A very general formulation of it is the following:

$$R(x, z, w) = \begin{cases} f(z) \cdot x & \text{if } x < h(w) \\ g(z) \cdot x & \text{if } x \geq h(w) \end{cases} \quad (1)$$

The functions f , g , and h are used to decide the two growth rates of R and the point at which this growth increases. z and w are suitable parameters that can be related to the trend of x ; they could also coincide. For example, if x has associated a power law distribution, z and/or w could coincide with the corresponding α parameter. The definition of the ReLU function introduced above is extremely general. A specific formulation of the ReLU function typically used in the deep learning context is Leaky ReLU [52], which is defined as follows:

$$R(x) = \begin{cases} 0.01 \cdot x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (2)$$

We chose to adopt Leaky ReLU in our approach because it is a function that has a small negative slope for negative inputs, while, in the classical definition of ReLU, all the input values below zero are set to zero. In this way, using Leaky ReLU, the signal of x , when $x < 0$, is mitigated through $R(x)$ and is not discarded. Starting from it, a specialization of R , which takes all the previous observations and needs into account, and which, therefore, may be able to estimate the contribution of a parameter x to the eWoM Power of a post, is the following:

$$R(x, \alpha) = \begin{cases} 0.01 \cdot \nu(x) & \text{if } \nu(x) < \frac{1}{\alpha} \\ \nu(x) & \text{if } \nu(x) \geq \frac{1}{\alpha} \end{cases} \quad (3)$$

In this formula, ν is a normalization function for x ; it receives an actual value of this parameter and returns a normalized value of it, belonging to the real interval $[0, 1]$. A possible formulation of ν is the following:

$$\nu(x) = \frac{x}{x_M} \quad (4)$$

where x_M is the maximum value of it in the dataset.

The proposed definition of R reaches the goals we were aiming for. In fact, in the power law distribution, the steeper the curve, the higher the value of α is. On the other hand, as the curve steepness increases, the highest values of x should be increasingly “rewarded” with a higher eWoM Power. However, in order to counterbalance the tendency to excessively reward a very small number of elements, as α grows, ν moves increasingly to the left of the point where $R(x, \alpha)$ passes from a low slope to a high one. In the definition of R , we also decided to use the normalized value of x , and not the actual one, because we will adopt R to model the contribution to the eWoM Power of many fields with very heterogeneous ranges of values. Normalization allowed us to homogenize the values returned by the Leaky ReLU functions associated with the various fields, so that they can be combined in appropriate expressions in the next steps of our approach.

4.2. A Naive Formulation of the eWoM Power of Reddit Posts

After choosing to use R for evaluating the contribution that x can give to the eWoM Power of a post, we can define a first “naive” version of the function $\overline{W\mathcal{P}}(p)$, which computes the eWoM Power of a post p on Reddit. This function can be defined as:

$$\overline{\mathcal{WP}}(p) = \frac{\omega_c \cdot R(c, \alpha_c) + \omega_{cr} \cdot R(cr, \alpha_{cr}) + \omega_s \cdot R(s, \alpha_s) + \omega_a \cdot R(a, \alpha_a) + \omega_k \cdot R(k, \alpha_k) + \omega_u \cdot R(u, \alpha_u) + \omega_{sb} \cdot R(sb, \alpha_{sb})}{\omega_c + \omega_{cr} + \omega_s + \omega_a + \omega_k + \omega_u + \omega_{sb}} \quad (5)$$

Here: (i) $R(c, \alpha_c)$ is the contribution provided by num_comments; (ii) $R(cr, \alpha_{cr})$ is the contribution given by num_crossposts; (iii) $R(s, \alpha_s)$ is the contribution of the score of p ; (iv) $R(a, \alpha_a)$ is the contribution given by the parameter total_awards; (v) $R(k, \alpha_k)$ is the contribution provided by the karma of the author of p ; (vi) $R(u, \alpha_u)$ is the contribution given by the parameter created_utc of the author of p (we recall that this parameter denotes the seniority of the author of p in Reddit); (viii) $R(sb, \alpha_{sb})$ is the contribution given by the parameter subscribers.

If a post has just been created, the corresponding values of $R(c, \alpha_c)$, $R(cr, \alpha_{cr})$, $R(s, \alpha_s)$, and $R(a, \alpha_a)$ are set to 0. If an author has just registered to Reddit, the corresponding values of $R(k, \alpha_k)$ and $R(u, \alpha_u)$ are set to 0. If a subreddit has just been created, the corresponding value of $R(sb, \alpha_{sb})$ is set to 0. Finally, if a post p is new, its author has just registered to Reddit, and the subreddit in which it is published has just been created, it is assumed that the eWoM Power of p is 0. These last observations highlight a very important property of eWoM Power that we mentioned in the Introduction. In fact, the eWoM Power of a post is not static and immutable over time. On the contrary, it is in continuous evolution. As the values of the parameters mentioned above increase, the eWoM Power grows because it receives positive contributions from them. However, there is a fundamental aspect that works in the opposite direction: it is the “time” factor. In fact, as time goes by, a post becomes obsolete, and this provides a continuous negative contribution to its eWoM Power. Therefore, the positive and negative contributions represent two contrasting and dynamic forces acting in the eWoM Power of a post. As a consequence, the naive version of $\overline{\mathcal{WP}}(p)$ must be refined, taking the contribution of time into account.

One might argue that this formula is simple. Regarding this potential objection, we would like to point out the following:

- It represents only the starting point for the computation of the refined eWoM Power, which, instead, is characterized by a more sophisticated formulation.
- The adoption of a weighted mean gives our definition of naive eWoM Power a great flexibility because changing the weights makes it possible to model very different situations, starting from the same simple common formulation.
- The adoption of a weighted mean gives our definition of naive eWoM Power a great extensibility, making both the addition of new parameters that one wants to consider and the modeling of scenarios different from those initially envisaged easy (see Section 5.1 for some of them). In this last case, as we will see below, it will be enough to define a new combination of weights that is well fit to the scenario that one wants to model.

4.3. A Refined Formulation of eWoM Power of Reddit Posts

To perform this task, it is first necessary to understand how the value of a post changes over time. To the best of our knowledge, in the past literature, no attempt to perform this analysis for Reddit posts has been made. However, some authors analyzed this problem with reference to content on Twitter and found that their value decays according to a power law function [8]. They also measured the value of the coefficient α of this function and found that it is equal to 1. This result is totally in line with the data analytics theory, according to which the value of the data decreases exponentially over time [19].

The results described in [8] and especially the data analytics theory [19] allow us to assume that, in the long run, the value of a post always decays exponentially. Actually, it could happen that, in some cases, this value stops decaying and even starts to increase again, for a certain period [53–55]. This fact could also happen several times in the lifetime of a post. However, if we analyze the phenomenon in the long run, without considering this “noise” that could be present in some moments, we can certainly say that the value of a post decays over time [8,19]. This theoretical result was also confirmed experimentally

on Reddit, as we will see later in this paper. Indeed, we will see that there are several templates modeling the possible trends of the value of a post on Reddit, but all of them, eventually, in the long run, lead to a decrease toward zero. In data analytics theory, the concept of the “value” of data is generic. However, it is clear that there is a relationship between eWoM Power and value, and it is exactly on this idea that we will rely following this discussion. Now, since:

- The decay law of the value of a post over time is similar to all the distribution functions related to the other parameters of interest for eWoM Power;
- We have seen that, in those cases, the Leaky ReLU function is well suited to express the contribution of each of those parameters to eWoM Power;

It seems reasonable to use the Leaky ReLU function also to model the role of the “time” factor in determining the eWoM Power of a post.

A final reasoning is necessary before defining a refined version of eWoM Power. In fact, the naive version $\overline{\mathcal{WP}}(p)$, previously defined, assumes that all parameters cannot vary over time. Instead, a more refined modeling of eWoM Power must consider that these parameters can change over time. For this reason, the refined eWoM Power should be computed through a function $\mathcal{WP}(p, t)$, which returns the eWoM Power of a post p at the time instant t .

In order to formalize the function $\mathcal{WP}(p, t)$, it is necessary to introduce two support functions, which we call $\widehat{\mathcal{WP}}(p, t)$ and $R(t, \alpha_t)$. They play a key role in the definition of $\mathcal{WP}(p, t)$. Let t be a time instant:

- The function $\widehat{\mathcal{WP}}(p, t)$ returns the value of $\overline{\mathcal{WP}}(p)$ at t .
- $R(t, \alpha_t)$ is the Leaky ReLU function at t . As for it, we note that:
 - When $\nu(t)$ is smaller than $\frac{1}{\alpha_t}$, the decrease caused by the “time” factor is low and grows slowly over time.
 - When $\nu(t)$ is higher than or equal to $\frac{1}{\alpha_t}$, the decrease becomes consistent and grows quickly over time.
 - In the definition of the normalization function ν associated with $R(t, \alpha_t)$, we assumed that the maximum value t_M of t is equal to the number of seconds elapsed between t and the time instant in which p was published.

We now have all the necessary elements to formalize $\mathcal{WP}(p, t)$. In particular:

$$\mathcal{WP}(p, t) = \begin{cases} \widehat{\mathcal{WP}}(p, t) - R(t, \alpha_t) & \text{if } \widehat{\mathcal{WP}}(p, t) > R(t, \alpha_t) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

5. Experiments and Possible Applications of the eWoM Power of Reddit Posts

In this section, we propose some experiments aimed at evaluating the goodness of the proposed approach, as well as at illustrating two possible applications of it. Before starting such discussions, we highlight that, at the address <https://github.com/daisy-univpm/eWoMPower> (accessed on 1 March 2023) we created a GitHub repository, which contains all the code implementing our approach, as well as a file with the instructions for extracting from pushshift.io (accessed on 21 November 2022) the datasets we used.

5.1. Preliminaries: Evaluation of the Parameters’ Impact on Naive eWoM Power

In Section 4, we saw that the formula for $\overline{\mathcal{WP}}(p)$ is a weighted mean of the contributions provided by seven different parameters. We deliberately left our approach flexible, and we have not defined fixed values for the weights of this mean, so that the user can set them autonomously. We also defined a default combination, which sets all the weights to one, denoting that all the parameters have the same importance in the definition of the eWoM Power. The system implementing our approach proposes this default configuration to the user, but leaves her/him free to modify it in the way she/he considers most appro-

appropriate. Therefore, based on the scenario of interest, a user can identify the most-suitable set of weights. To simplify this task for her/him, we designed some predefined weight combinations; she/he can choose one of them or define a new one. Our weight combination proposals are reported in Table 5. As said above, the flexibility of our definition allows users to define new combinations, based on the reference scenario and the issues to address.

Table 5. Our proposals of weight combinations.

Combination	Weight Values	Characteristics
C_0	All weights are set to 1.	This is the default combination. It gives the same importance to all parameters.
C_1, C_2	$\omega_c = \omega_{cr} = 2$ (for C_1); $\omega_c = \omega_{cr} = 10$ (for C_2); all the other weights are set to 1.	It privileges comments and crossposts and, therefore, those posts for which a discussion has been started and which have been crossposted on more subreddits. The privilege degree is moderate in C_1 and high in C_2 .
C_3, C_4	$\omega_s = \omega_a = 2$ (for C_3); $\omega_s = \omega_a = 10$ (for C_4); all the other weights are set to 1.	It privileges scores and awards and, therefore, posts with a high quality. The privilege degree is moderate in C_3 and high in C_4 .
C_5, C_6	$\omega_k = \omega_u = 2$ (for C_5); $\omega_k = \omega_u = 10$ (for C_6); all the other weights are set to 1.	It privileges the karma and seniority of the post authors and, therefore, the posts published by famous authors on Reddit. The privilege degree is moderate in C_5 and high in C_6 .
C_7, C_8	$\omega_k = \omega_u = \omega_{sb} = 2$ (for C_7); $\omega_k = \omega_u = \omega_{sb} = 10$ (for C_8); all the other weights are set to 1.	It privileges not only the karma and seniority of the post authors, but also the number of subscribers to the subreddit where the post was published. The privilege degree is moderate in C_7 and high in C_8 .
C_9, C_{10}	$\omega_c = \omega_{sb} = 2$; $\omega_{cr} = \omega_a = 4$ (for C_9); $\omega_c = \omega_{sb} = 10$; $\omega_{cr} = \omega_a = 100$ (for C_{10}); all the other weights are set to 1.	It privileges those parameters that, according to their semantics, should play the most-important roles in the definition of eWoM Power. The privilege degree is moderate in C_9 and high in C_{10} .
C_{11}, C_{12}	$\omega_{sb} = 2$; $\omega_c = 4$; $\omega_{cr} = 6$; $\omega_a = 8$ (for C_{11}); $\omega_{sb} = 10$; $\omega_c = 100$; $\omega_{cr} = 1,000$; $\omega_a = 10,000$ (for C_{12}); all the other weights are set to 1.	It privileges those parameters characterized by rare occurrences; it assumes that the rarer an occurrence, the more its presence in a post contributes to increasing the post's eWoM Power. The privilege degree is moderate in C_{11} and high in C_{12} .

We used these 13 combinations (which correspond to 13 different scenarios) to perform a series of tests for studying the contribution that each parameter gives to the naive eWoM Power. The first test we performed aimed to compute the average eWoM Power of the posts of our dataset for each scenario. The result obtained is shown in Table 6. From the analysis of this table, we can derive several pieces of information on the role and weight of the various parameters influencing the naive eWoM Power. In this activity, we refer to the combination C_0 as the default one and drew our conclusions by comparing the other combinations with respect to it.

First of all, we note that the shift from the linear scale to the logarithmic one strengthens the differences that can be observed with respect to the default combination, both those involving an increase of the average eWoM Power and those involving a decrease of it. However, the interesting information here is another, i.e., that the logarithmic scale has a much greater impact for the combinations having an average eWoM Power higher than C_0 than for the ones with an average eWoM Power lower than it. In fact, the maximum variation in the increase of the average eWoM Power with respect to the default one caused by the logarithmic scale can be observed when passing from C_5 to C_6 and is equal to 57.66%. Instead, the maximum variation in the decrease happens when passing from C_3 to C_4 and is equal to 36.28%.

Table 6. Average eWoM Power of posts for the predefined weight combinations.

Combination	Average eWoM Power of Posts	Combination	Average eWoM Power of Posts
C_0	0.02984	C_2	0.00843
C_1	0.02323	C_4	0.00856
C_3	0.02327	C_6	0.06335
C_5	0.04018	C_8	0.06123
C_7	0.04170	C_{10}	0.00326
C_9	0.01766	C_{12}	0.00059
C_{11}	0.01155		

We also note that the choice of weights has enough influence on the computation of the eWoM Power. However, this impact is not to alter the final result. In fact, adopting the linear scale, the maximum change compared to the default value is 39.74%, in the case that there is an increase of the average eWoM Power. Instead, the maximum change is equal to 61.30%, in the presence of a decrease. As far as this result is concerned, we believe it should be emphasized that, once a combination has been chosen, it should be maintained in the computation of the eWoM Power of all the posts to analyze. Therefore, the increases or decreases caused by this combination with respect to the default one represent a sort of “bias” that the end user decides to apply on all the posts she/he will examine. However, this “bias”, just because it is applied to all the posts, does not affect the correctness of the analysis regarding how much one post is better or worse than another one as concerns eWoM Power.

Examining the combinations that return a higher value of the eWoM Power than the default one, we can observe that this happens for $C_5 - C_8$. This implies that, given a post, the author publishing it and the subreddit in which it is published play a fundamental role in its eWoM Power and, ultimately, in its success. We observed that C_7 (respectively, C_8) differs from C_5 (respectively, C_6) due to the fact that, besides privileging the author, it also privileges the subreddit. Note that the increase of the value of the eWoM Power of C_5 , compared to the default one, is equal to 34.67%, while the increase of the value of the eWoM Power of C_7 , compared to the default one, is equal to 39.74%. This implies that, in determining the eWoM Power of a post, its author has much more than the subreddit where it is published.

As for the other combinations, we can see that $C_1 - C_4$ return average values of eWoM Power quite close to the default ones, while $C_9 - C_{12}$ return much more pessimistic values than the default ones. Recall that these combinations generally privilege the history of the single post (for example, the awards or the comments it received) with respect to the author who published it and the subreddit where it was published. This allowed us to say that the author who published a post and the subreddit where it was published play an extremely relevant role for its spreading capability. Actually, this result is in line with the past literature of social network analysis, which highlights the role played by influencers in information diffusion on social media [10,56,57].

As a second test, we selected the top 1000 posts with the highest eWoM Power for each combination. Our goal was to check if they were completely different from each other or if, instead, there were posts that fell in the top 1000 ones regardless of the selected combination. To make this verification, we computed the intersection of the top 1000 posts of all the combinations $C_0, C_1, C_3, C_5, C_7, C_9$, and C_{11} . We obtained that this intersection included 446 posts. This result is extremely interesting because it tells us that a considerable number of posts have a very high value of eWoM Power, regardless of the chosen combination of parameter weights. This is an important guarantee of the objectivity of our approach. On the other hand, about half of the top 1000 posts of each combination depend on it. This is also an important result because it provides the user with a certain flexibility to determine which eWoM Power is best suited to her/his reference scenario.

After this, we computed the intersection of the top 1000 posts for each possible pair of weight combinations. The results obtained are shown in Table 7. From the analysis of this table, we can detect various interesting knowledge patterns. In particular, we can observe the following:

- \mathcal{C}_0 , \mathcal{C}_1 , and \mathcal{C}_3 have a very high number of top 1000 posts in common. This result is in line with the ones reported in Table 6, where we can see that the average values of eWoM Power for these three combinations are very close.
- \mathcal{C}_9 and \mathcal{C}_{11} have a very high number of top 1000 posts in common. In this case also, the result is in line with the ones on the average values of eWoM Power, reported in Table 6. In fact, the values of the eWoM Power of \mathcal{C}_9 and \mathcal{C}_{11} are very close to each other and very far from those of the other combinations.
- \mathcal{C}_5 differs considerably from all the other combinations. This is not surprising for \mathcal{C}_0 , \mathcal{C}_1 , and \mathcal{C}_3 , while it is unexpected for \mathcal{C}_7 . In fact, \mathcal{C}_5 and \mathcal{C}_7 differ only by the fact that the latter privileges not only the author (as \mathcal{C}_5 does), but also the subreddit. While this fact does not produce substantial differences on the average values of eWoM Power, it causes great differences in the lists of the top 1000 posts.
- There is a remarkable overlap between the top 1000 posts of \mathcal{C}_7 and the top 1000 posts of \mathcal{C}_0 , \mathcal{C}_1 , and \mathcal{C}_3 . This is quite surprising because this overlap is not reflected in the average values of eWoM Power, shown in Table 6; in fact, in this table, \mathcal{C}_7 is the combination that differs most from \mathcal{C}_0 , \mathcal{C}_1 , and \mathcal{C}_3 .

Table 7. Cardinality of the intersection of the top 1000 posts for each possible pair of weight combinations.

	\mathcal{C}_0	\mathcal{C}_1	\mathcal{C}_3	\mathcal{C}_5	\mathcal{C}_7
\mathcal{C}_0	1000				
\mathcal{C}_1	994	1000			
\mathcal{C}_3	978	982	1000		
\mathcal{C}_5	595	596	594	1000	
\mathcal{C}_7	991	985	969	590	1000
\mathcal{C}_9	803	802	789	449	806
\mathcal{C}_{11}	803	805	791	450	805

5.2. Evaluation of the Goodness of the Results Returned by Our Approach

As specified in the Introduction, eWoM has been much studied both on specialized social platforms, such as Yelp and TripAdvisor, and on some generalist platforms, such as Facebook and Twitter. Unfortunately, such studies are essentially absent for Reddit. Therefore, testing the adequacy of our approach cannot be performed using a classical way of proceeding based on comparing the results it returns with those returned by other closely related approaches proposed in the past. In a situation like this, it seemed to us that the most-correct (though very laborious and expensive) way of proceeding was through the support of a human expert.

To this end, for each combination $\mathcal{C}_0 - \mathcal{C}_{12}$, we randomly selected 200 posts from the top 1000 returned by it, for a total number of posts equal to 2600 (It is worth pointing out that this sample size is consistent with similar studies in the literature [58,59].) and asked a human expert to verify whether, indeed, each of those posts had been shown to have high eWoM Power in reality. In Table 8, we report, for each combination, the percentage of posts that, according to the human expert, had met that condition. From the analysis of this table, it can be seen that the human expert judged the results returned by our approach as really excellent. In fact, for any combination, more than 90% of the posts judged to have high eWoM Power by our approach were confirmed as such by the human expert. Moreover, in many cases, this percentage is close to or reaches 100%. We can also note that: (i) the combinations with a “moderate” privilege degree (i.e., $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_3, \mathcal{C}_5, \mathcal{C}_7, \mathcal{C}_9, \mathcal{C}_{11}$) returned better results on average than those with a high privilege degree (i.e., $\mathcal{C}_2, \mathcal{C}_4, \mathcal{C}_6, \mathcal{C}_8, \mathcal{C}_{10}, \mathcal{C}_{12}$); (ii) there is a strong correspondence between the results in Table 8 and those in Table 6, in

that the greater the overlap between the posts of a pair of combinations, the more similar the corresponding results returned by the human expert are.

Table 8. Percentage of posts returned by our approach that the human expert evaluated to have high eWoM Power in reality.

Combination	Percentage of Posts	Combination	Percentage of Posts
C_0	100%	C_2	99%
C_1	100%	C_4	98%
C_3	99.5%	C_6	93.5%
C_5	95%	C_8	97.5%
C_7	99%	C_{10}	97%
C_9	98%	C_{12}	96.5%
C_{11}	98%		

As the second task in this experiment, we asked the human expert to give us hints as to which combinations were best and whether there was a specific area in which one combination was better than the others. The human expert, after a careful examination of the posts available, concluded the following:

- The best combination is C_0 . It can be used in all circumstances remaining confident that the results it returns are correct and satisfactory.
- C_1 is suitable for scenarios where the popularity of the post is the most-important factor. It can favor the diffusion of the most-popular and -widely shared content.
- C_2 is suitable for scenarios where the engagement of the post is the most-important factor. It can favor the diffusion of highly commented on or upvoted posts.
- C_3 is suitable for scenarios where the quality of the post is important. It can favor the diffusion of high-quality posts.
- C_4 is suitable for scenarios where the expertise of the author is important. It can favor the diffusion of posts authored by individuals with relevant professional or personal experience.
- C_5 is suitable for scenarios where the social influence of the authors of posts is important. It can favor the diffusion of posts published by user with high social influence.
- C_6 is suitable for scenarios where the author's activity level is important. It can favor the diffusion of posts authored by users who frequently contribute to the community.
- C_7 is suitable for scenarios where the number of subscribers to the subreddit where the post was published is important as much as the karma and seniority of the post author. It can favor the diffusion of posts that are written by authoritative users on authoritative subreddits.
- C_8 is suitable for scenarios where the number of subscribers to the subreddit and the karma and seniority of the author are of utmost importance. It can favor the diffusion of the posts published by influencers in the community.
- C_9 is suitable for scenarios where the comments to the post and the subreddit where the post was published are more important than the other parameters and awards and shares are even more important. It can favor the diffusion of posts stimulating many interactions and whose quality has been recognized by users.
- C_{10} is suitable for scenarios where the interactions with and the awards to the posts are of utmost importance. It can favor the diffusion of posts with the highest number of interactions by users and that received the highest number of awards with respect to the other posts.
- C_{11} is suitable for scenarios where the rarest parameters (such as awards and cross-posts) should play a more important role. It can favor the diffusion of posts that are different from the majority of the other ones as they have higher values on these parameters.

- C_{12} is suitable for scenarios where the rarest parameters are of utmost importance. It can favor the diffusion of the rarest posts, which have the highest values of awards, crossposts, and number of comments.

5.3. A First Application: Determining the Lifespan of a Reddit Post

As we have seen in Section 4, the “time” factor plays a fundamental role in eWoM Power. Examining the formula of refined eWoM Power, reported in Equation (6), we see that it has two components, namely $\widehat{\mathcal{WP}}(p, t)$ and $R(t, \alpha_t)$. The former contributes to an increase of the refined eWoM Power, while the latter contributes to its decrease. According to this reasoning, the possible dynamics of eWoM power could be many, especially taking into account that seven different factors can act simultaneously on $\widehat{\mathcal{WP}}(p, t)$. For example, a post can be submitted by an author with a high score and karma, resulting in a very high initial refined eWoM Power. However, it may not be attractive, which would result in a fast decay of eWoM Power. On the contrary, a post could start with a very low eWoM Power, but shortly after, it could be reposted by a user with a high karma on a subreddit with many subscribers. This would result in a significant increase of its eWoM Power; this value could remain stable for some time and, then, decay because the content of the post becomes obsolete. These are just two examples of the possible lifespans of Reddit posts. Actually, the possibilities could be many. Therefore, an interesting challenge is the definition of a set of lifespan templates in which most of the posts can fall.

There are several possible ways to deal with this problem. For example, one might think of taking all the available posts (which, in our case, were 62,521,691), considering the corresponding lifespans, performing a clustering task on them, and identifying a lifespan template for each cluster. This way of proceeding is certainly very interesting. However, it presents some problems. First, in this case, one would have to cluster trends of functions that vary over time. This type of clustering is extremely difficult (e.g., it requires the prior alignment of temporal features before their clustering) and is very space- and time-consuming. In particular, for the number of posts available to us, the problem may not be solvable in a reasonable amount of time and space. Furthermore, once clusters are identified, it is not straightforward to define a representative lifespan for each of them [60]. To overcome some of these difficulties, in the past, it was considered to represent the temporal functions in alternative formats (e.g., in the form of strings [61]) and to later apply clustering on the new formats. This way of proceeding solves some problems such as, for example, the alignment of functions. However, it is also time-consuming and raises new issues. For example, once the string representative of each cluster is identified, it is not easy to perform the inverse transformation required to obtain the trend template corresponding to that cluster. Taking into account all these difficulties and wishing to obtain the lifespan templates as accurately as possible, we thought of proceeding differently, requiring the support of a human expert. The latter was asked to examine a number of posts and identify possible post lifespan templates from them. Clearly, the presence of the human expert greatly reduces the number of post lifespans that can be examined to obtain templates, but it allows a very accurate result to be potentially obtained. It is then necessary that this result can be confirmed in a subsequent validation phase involving all available posts.

Following this way of proceeding, we initially selected 200 posts at random and asked the human expert to examine them carefully to manually build their lifespans. The choice of 200 random posts in the study was justified as a representative sample size that is large enough to capture the variability and diversity of posts on the platform. The sample size was selected based on the principles of statistical sampling, and it can provide a reliable estimation of the population parameters of interest. Again, a sample size of 200 posts is consistent with similar studies in the literature [58,59]. At the end of this task, we identified six possible templates of post lifespans. They are shown in Figure 6. Let us carefully examine them:

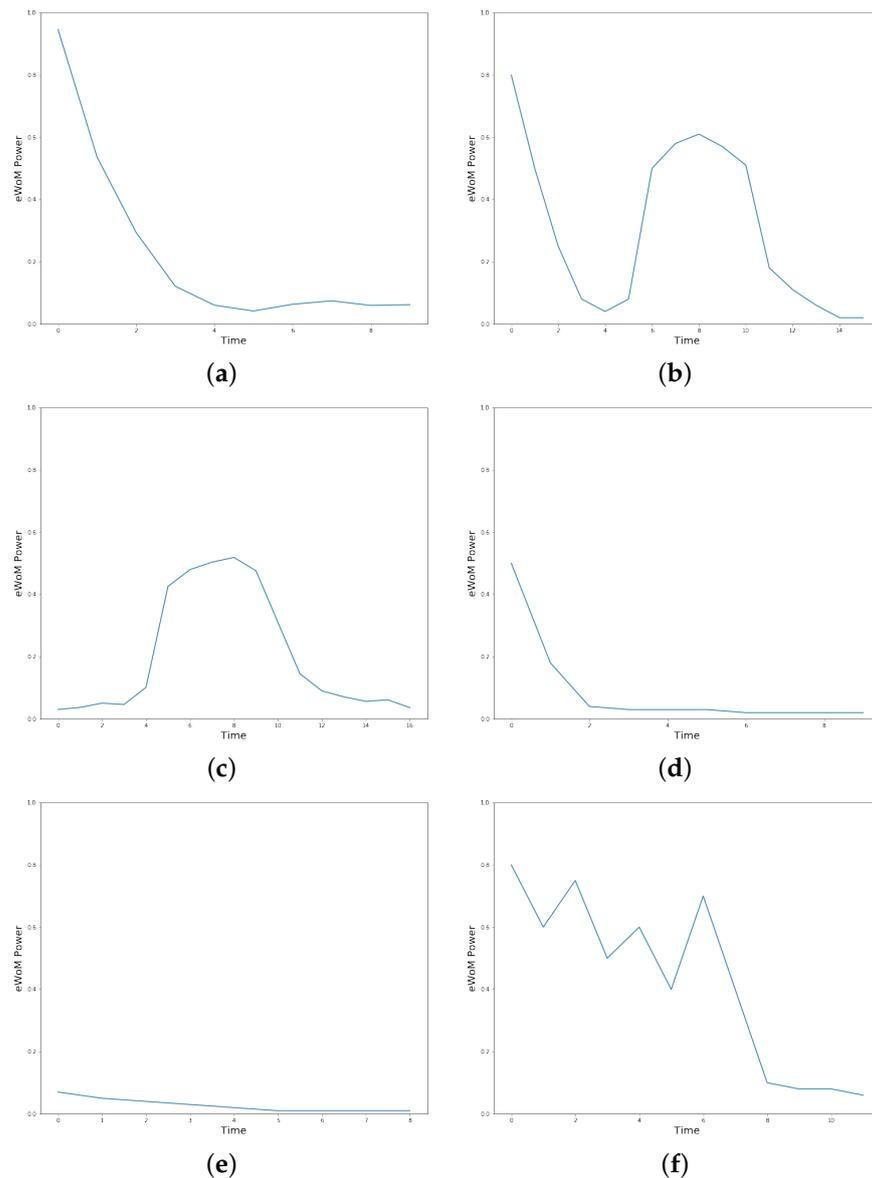


Figure 6. The six lifespan templates we determined: (a) explosive; (b) revived; (c) bell; (d) unlucky; (e) gray; (f) moody.

- The template of Figure 6a could be called “Explosive”. It corresponds to a scenario where an author with high karma and a high score submits a post, possibly in a subreddit with many subscribers. The eWoM Power of this post is initially very high; however, its content does not attract other users, and the post is not successful. As a consequence, the “time” factor leads to a fast decay of the eWoM Power, without being contrasted by positive factors.
- The template of Figure 6b could be called “Revived”. It is similar to the template “Explosive”. However, in this case, the decay caused by the “time” factor is contrasted by the growth of some positive factors. These manage to increase the eWoM Power again and keep it constant for some time until the post content becomes obsolete, the “time” factor prevails, and the eWoM Power decays.
- The template of Figure 6c could be called “Bell”. In this case, the post starts with a low eWoM Power, maybe because its author is not famous, and it is published in a subreddit with few subscribers. At a certain time, the positive factors suddenly make the eWoM Power grow, maybe because the post is reposted by a famous user or

receives an award. At this point, the value of the eWoM Power remains constant for some time until the post content becomes obsolete and it decays.

- The template of Figure 6d could be called “Unlucky”. In this case, the post starts as poorly as in the previous case. However, different from before, there is no positive explosive factor able to significantly increase its eWoM Power, which becomes zero in a short time.
- The template of Figure 6e could be called “Gray”. In this case, the post starts with a medium value of eWoM Power. However, this last parameter does not grow; therefore, in a short time period, the “time” factor lowers its eWoM Power to zero.
- The template of Figure 6f could be called “Moody”. Here, the eWoM Power starts as high as in the “Explosive” template. Similar to what happens for that template, the “time” factor acts negatively, tending to lower the eWoM Power. However, the post is very successful. Therefore, there are periodically positive factors that increase its eWoM Power. These ups and downs are repeated several times until the post content becomes obsolete and its eWoM Power becomes zero.

Once we had identified these six lifespan templates, we ran a test to see if they were really able to cover most of the posts. To this end, we considered all the posts of our dataset, and for each of them, we verified if it fell into a lifespan template and, in the affirmative case, in which template it fell. Clearly, in this task, a certain margin of error was tolerated. In order to handle it, we considered the Root-Mean-Squared Error (RMSE) as a benchmark. Specifically, our need was to find a threshold value of this parameter that simultaneously provided good flexibility and high accuracy. In fact, we must take into account that, since we compared the trends of functions that vary over time, we need some flexibility. Otherwise, the number of possible templates would grow dramatically. On the other hand, we needed to ensure a certain accuracy; otherwise, we risk considering equivalent post lifespans that are very different in reality. In order to define a threshold value for RMSE, we computed the distributions of this parameter for all lifespan templates and posts, studied the characteristics of these distributions, and carried out a series of trial and error by also asking the opinion of the human expert. Due to space limitations, we cannot illustrate all these tasks in detail here. However, we can say that, at the end of them, we saw that the best tradeoff between flexibility and accuracy was obtained by setting the threshold to 0.2. By proceeding with this threshold value, we obtained the distribution of the posts against lifespan templates reported in Table 9. From the analysis of this table, we can observe that, actually, the defined templates were able to capture the lifespans of almost all the posts of the dataset. This result is certainly positive, but we thought it appropriate to make further checks on its stability. In fact, the detected lifespan templates might be well suited only for the posts published in the time period to which the dataset referred, and not for posts published in other time periods.

Table 9. Distribution of the posts of our dataset against lifespan templates.

Lifespan Template	Percentage of Posts
“Explosive”	22.2%
“Revived”	22.6%
“Bell”	7.2%
“Unlucky”	22.5%
“Gray”	18.3%
“Moody”	4.7%
None of them	2.5%

In order to verify the stability of our lifespan templates, we considered a second dataset containing all the posts published from 1 August 2020 to 31 October 2020 and repeated the same tasks carried out on the posts of our first dataset. Our goal was to verify the distribution of the posts against the previously identified lifespan templates. The results obtained are shown in Table 10. As we can see from this table, the distribution of the posts

against the lifespan templates is very close to that of the first dataset reported in Table 9. This is certainly a confirmation that our lifespan templates are not only able to capture almost all post lifespans, but are stable over time. Therefore, they are a valuable tool to classify Reddit posts based on their potential future success and to perform a more refined profiling activity next. We will deal with this last topic in the next subsection.

Table 10. Distribution of the posts of the second dataset against lifespan templates.

Lifespan Template	Percentage of Posts
"Explosive"	21.8%
"Revived"	23.4%
"Bell"	7.5%
"Unlucky"	21.8%
"Gray"	17.6%
"Moody"	4.9%
None of them	3.0%

5.4. A Second Application: Determining the Profile of a Reddit Post

The six lifespan templates defined in the previous section can be the starting point for a more in-depth analysis of Reddit posts, aimed at their profiling. In order to carry out this activity, we initially defined two "dimensions", i.e., two orthogonal parameters necessary to partition the posts with the same lifespan template. The two dimensions identified are the value of the refined eWoM Power (which can be low or high) and the lifetime of the post (which can be short or long). Based on these two dimensions, all posts following a given lifespan template can be partitioned into four subsets. For each of these subsets, we can examine the posts belonging to it and try to understand if there are some features they have in common and, in the affirmative case, if these features, on the whole, define a precise post profile.

Like the previous application, we felt that directly adopting an automated approach would not lead to satisfactory results. Therefore, also in this case, we required the intervention of a human expert. Specifically, for each lifespan template, we randomly selected 100 posts belonging to it. We then computed the distribution for both the eWoM Power and lifetime dimensions and, based on them, asked the human expert to indicate which values of the lifespan in his opinion could be considered short (respectively, long) and which values of the eWoM Power could be considered low (respectively, high). Afterwards, for each of the four categories obtained for each lifespan template, we asked the human expert to examine the posts that fell into that category and to draw a corresponding profile of them. Again, we felt that the presence of the human expert was appropriate because it was necessary to examine the tests of the posts and also understand the presence of aspects, such as irony and sarcasm, for which the interpretative limitations of an automated approach are well known. Clearly, the examination of 600 total posts by a human expert was a laborious and time-consuming work. However, we believe it was worth it, as it allowed us to draw sketches of the posts for each lifespan template and for any type of values of the two dimensions under considerations.

We carried out this analysis for each lifespan template and obtained encouraging results, because, for each partition, we were able to identify very precise profiles for the posts belonging to it. More specifically, the categorization of posts associated with the lifespan template "Explosive" is reported in Table 11. Analyzing it, we can see that all posts with a short lifetime fall in the profile "Meme". The posts with a long lifetime fall in the profile "Niche News", if they have a low value of refined eWoM Power, or in the profile "News", otherwise.

Table 11. Categorization of the posts associated with the lifespan template “Explosive”.

	Short Lifetime	Long Lifetime
<i>Low eWoM Power</i>	Meme	Niche News
<i>High eWoM Power</i>	Meme	News

The three profiles we detected for this lifespan template have the features specified below:

- “Meme” represents a post or a picture that attracts the attention of other users through irony or sarcasm. It can concern the behavior of people or animals or, alternatively, current topics.
- “Niche news” represents news concerning a topic of interest to a small audience. Often, it is local news or news that did not receive great attention from media.
- “News” represents news concerning a current topic. It could regard a chronicle, a celebrity, finance, and so on. In any case, it received great attention from media.

The categorization of the posts associated with the lifespan template “Revived” is shown in Table 12. From the analysis of this table, we observe that all posts with short lifetime fall in the profile “Meme”. Instead, the posts with long lifetime fall in the profile “Cringe/NSFW”, if they have a low value of refined eWoM Power, or in the profile “Digital art”, otherwise.

Table 12. Categorization of the posts associated with the lifespan template “Revived”.

	Short Lifetime	Long Lifetime
<i>Low eWoM Power</i>	Meme	Cringe/NSFW
<i>High eWoM Power</i>	Meme	Digital art

We have already seen the characteristics of “Meme”. Regarding the other two profiles, we note that:

- “Cringe/NSFW” is a post about not safe for work niche topics or, anyway, about topics not suitable for a general audience. Usually, it is an NSFW post published in a subreddit with a lower than average number of subscribers.
- “Digital art” is a post containing a digital drawing, created by the post author herself/himself. The posts of this category are generally published by authors to show off their work and make themselves known.

The categorization of the posts associated with the lifespan template “Bell” is shown in Table 13. This table highlights that all the posts with a high value of refined eWoM Power fall into the profile “Provocative”. On the other hand, the posts with a low value of this parameter fall in the profile “Cringe/NSFW”, if they have a short lifetime, or in the profile “NSFW”, otherwise.

Table 13. Categorization of the posts associated with the lifespan template “Bell”.

	Short Lifetime	Long Lifetime
<i>Low eWoM Power</i>	Cringe/NSFW	NSFW
<i>High eWoM Power</i>	Provocative	Provocative

We have already seen above the features of “Cringe/NSFW”. As for the other two profiles, we note that:

- “Provocative” is a post that attracts the attention of other users in a provocative way, for example by explicitly attacking a politician or certain categories of people.
- “NSFW” is similar to “Cringe/NSFW”, but while the latter is niche, the former is generic.

The categorization of the posts associated with the lifespan template “Unlucky” is illustrated in Table 14. This table shows that all the posts with a low value of refined eWoM Power fall in the profile “Content Creator”, while the posts with a high value of this parameter fall in the profile “News”.

Table 14. Categorization of the posts associated with the lifespan template “Unlucky”.

	Short Lifetime	Long Lifetime
<i>Low eWoM Power</i>	Content Creator	Content Creator
<i>High eWoM Power</i>	News	News

We have previously seen the features of “News”. Regarding the profile “Content Creator”, we note that it is associated with a creative post of a user, who publishes her/his work online. It has many similarities with “Digital Art” but, different from this last one, its content does not necessarily consist of images.

The categorization of the posts associated with the lifespan template “Gray” is presented in Table 15. This table shows that all the posts with a short lifetime fall in the profile “Doubt”, while all the posts with a long lifetime fall in the profile “Unsuccessful”.

Table 15. Categorization of the posts associated with the lifespan template “Gray”.

	Short Lifetime	Long Lifetime
<i>Low eWoM Power</i>	Doubt	Unsuccessful
<i>High eWoM Power</i>	Doubt	Unsuccessful

As for these profiles, we note that:

- “Doubt” is a post where a user asks a direct question to a category of users to resolve a doubt or know something. Posts with this profile generally attract attention thanks to the answers to this question.
- “Unsuccessful” represents a generic post that failed to attract the attention of other users.

Finally, the categorization of the posts associated with the lifespan template “Moody” is presented in Table 16. This table shows that all the posts with a long lifetime refer to the profile “Politics”, while all the posts with a short lifetime refer to the profile “Meme”, if their refined eWoM Power is low, or to the profile “Provocative”, otherwise.

Table 16. Categorization of the posts associated with the lifespan template “Moody”.

	Short Lifetime	Long Lifetime
<i>Low eWoM Power</i>	Meme	Politics
<i>High eWoM Power</i>	Provocative	Politics

We have already seen the characteristics of “Meme” and “Provocative”. Regarding the profile “Politics”, we note that it is associated with a post about a current political topic. Unlike the posts of “Provocative”, the ones of “Politics” do not attract people’s attention in a negative way; instead, they stimulate the political debate. At the end of this analysis, we can say that the dimensions we identified were able to support us in partitioning the posts belonging to each category of lifespan templates. We were able to associate a precise profile with each partition, allowing us to characterize the content of the corresponding posts.

Finally, it is interesting to highlight that a profile can belong to more than one lifespan template. Therefore, even though we defined the various profiles starting from lifespan templates, in the end, we obtained that profiles and lifespan templates are two orthogonal and independent categorization modes. These modes, therefore, can be used together to better characterize each post in terms of both its evolution over time and its content.

6. Discussion

In this section, we propose a retrospective of our approach. Its objectives are: (i) highlighting the contribution of our approach compared to past literature; (ii) illustrating its specificities, which make it tailored to Reddit; (iii) describing its genericity, which could enable its future application on other social platforms. In the Introduction, we pointed out that the main contributions of this paper are:

- The definition of two parameters, the first naive, but simple to compute, the second refined, but expensive to calculate, for the computation of the eWoM Power.
- The definition of six possible post lifespan templates.
- The definition of a set of post profiles.

First of all, we point out that the second and third contributions are orthogonal to the first. In fact, if we considered a new definition of eWoM Power in place of those provided in this paper or wanted to extend our definition of eWoM Power from Reddit to other social networks, the approaches underlying the second and third contributions would still be valid.

In addition, it is worth pointing out that our approach for defining post lifespan templates and post profiles was based on the intervention of the human expert. As detailed in Section 5.3, in our opinion, this, far from being a flaw, is a factor enhancing the results obtained. Indeed, the presence of the human expert, although costly in terms of time, allowed us to achieve a remarkable level of accuracy and interpretability. Just think of the fact that only the human expert could have correctly interpreted the irony and sarcasm that characterized many of the posts under consideration. The presence of the human expert required us to limit the number of posts taken into consideration to define post lifespan templates and post profiles. However, the validation activities performed subsequently on the whole dataset at hand, with the aim of validating what we had found with the expert's support, allowed us to conclude that not only the soundness, but also the completeness of the results obtained were extremely high.

Regarding the first contribution, it is worth pointing out that the idea behind the two definitions of eWoM Power is extremely general and can be applied to all social platforms. In this paper, we decided to specialize that idea on Reddit. In the Introduction, we explained the reasons for this; they basically regard the fact that Reddit has some peculiarities as a social platform and that the eWoM Power on Reddit was little investigated in the past. The specialization lies in the fact that, in the eWoM Power formulas, we considered some parameters (e.g., the karma of the author of a post) that are typical of Reddit. Other parameters that we considered can also be found on other social networks. In a possible specialization of our approach, the parameters typical of Reddit could be replaced with others that have a close semantics (thus preserving the reasoning that led to the current definition of eWoM Power) or, even, with other parameters that are totally different and represent specificities of the social platform to which they refer. This is possible because: (i) the general approach underlying the two versions of eWoM Power is not platform-dependent; (ii) the definition of $\overline{WP}(p)$ (see Equation (5)) that underlies both definitions of eWoM Power is flexible and extensible, since it is a weighted mean of a number of factors, some of which can be removed, replaced, or complemented by others.

7. Conclusions

In this paper, we studied the eWoM Power of a post on Reddit. First of all, we saw that this parameter defines the ability of a post to be spread and known through this social network. Then, we proposed two definitions of it. The first one, called naive eWoM Power, is simpler because it does not take the "time" factor into account. Actually, although this factor makes the modeling of eWoM Power much more complex, it cannot be ignored. For this reason, we introduced a second definition, called refined eWoM Power, which considers its effects. Afterwards, we saw that the refined eWoM Power is influenced by contrasting parameters, some of which tend to increase its value, while another (i.e., "time")

tends to decrease it. Starting from this consideration, we examined the various possible cases in which one or more parameters prevail over the others. Thanks to this analysis, we were able to derive six lifespan templates in which the vast majority of Reddit posts fall. Finally, taking the eWoM Power and the lifetime of Reddit posts into account, we were able to define a set of possible post profiles. We also showed that lifespan templates and profiles are two orthogonal concepts that can be combined for a detailed categorization of Reddit posts.

As we mentioned in the Introduction, this paper is a first attempt to study the eWoM Power of a post on Reddit. As we saw, this social network has some specificities compared to others, such as Facebook and Twitter, on which several studies about eWoM Power had already been performed. This allowed us to investigate eWoM Power on social networks from different points of view than the past. In particular, we had the opportunity to investigate situations in which users do not post and comment impulsively, but with a cool head. We hope, or rather we are convinced, that this first attempt, even with its limitations, can stimulate other researchers, as well as ourselves, to deepen this analysis. With respect to this, some possible future developments are outlined below. First of all, it could be important to examine the contents of posts and see how they affect eWoM Power. This issue is extremely interesting and alone would require several analyses and papers to be addressed. Second, we would like to investigate whether fake news presents specific lifespans of its eWoM Power; in the affirmative case, this information could be used to define an approach for fake news detection. Finally, we would like to extend the ideas underlying our approach to Twitter and compare the results obtained with those returned by the large number of approaches regarding eWoM Power on Twitter already proposed in the literature. In this way, we could identify possible concordances and discordances in the results, which could give us insight into the strengths and weaknesses of our approach compared to the existing ones proposed for Twitter.

Author Contributions: Conceptualization, G.B. and L.V.; methodology, D.U.; software, L.V.; validation, E.C.; formal analysis, G.B.; investigation, E.C. and D.U.; resources, L.V.; data curation, G.B. and E.C.; writing—original draft preparation, L.V.; writing—review and editing, D.U.; visualization, G.B.; supervision, G.B. and E.C.; project administration, D.U. and L.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: <https://github.com/daisy-univpm/eWoMPower>, accessed on 1 March 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Richins, M.L.; Root-Shaffer, T. The role of evolvment and opinion leadership in consumer word-of-mouth: An implicit model made explicit. *ACR N. Am. Adv.* **1988**, *15*, 32–36.
2. Tucker, T. Online word of mouth: Characteristics of Yelp.com reviews. *Elon J. Undergrad. Res. Commun.* **2011**, *2*, 37–42.
3. Ismagilova, E.; Dwivedi, Y.; Slade, E. Perceived helpfulness of eWOM: Emotions, fairness and rationality. *J. Retail. Consum. Serv.* **2020**, *53*, 101748. [[CrossRef](#)]
4. Reyes-Menendez, A.; Saura, J.; Martinez-Navalon, J. The impact of e-WOM on hotels management reputation: Exploring TripAdvisor review credibility with the ELM model. *IEEE Access* **2019**, *7*, 68868–68877. [[CrossRef](#)]
5. Lee, H.; Law, R.; Murphy, J. Helpful reviewers in TripAdvisor, an online travel community. *J. Travel Tour. Mark.* **2011**, *28*, 675–688. [[CrossRef](#)]
6. Filieri, R.; Alguezaui, S.; McLeay, F. Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tour. Manag.* **2015**, *51*, 174–185. [[CrossRef](#)]
7. Zhang, N.; Campo, S.; Janz, K.; Eckler, P.; Yang, J.; Snetselaar, L.; Signorini, A. Electronic word of mouth on Twitter about physical activity in the United States: Exploratory infodemiology study. *J. Med. Internet Res.* **2013**, *15*, e261. [[CrossRef](#)]

8. Asur, S.; Huberman, B.; Szabo, G.; Wang, C. Trends in social media: Persistence and decay. In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM'11), Barcelona, Spain, 17–21 July 2011.
9. Cao, X.; Li, H.; Zhu, R. Analyzing the Online Word of Mouth Dynamics: A Novel Approach. In Proceedings of the Academy of Management, Boston, MA, USA, 7–11 August 2020; Academy of Management Briarcliff Manor: Briarcliff Manor, NY, USA, 2020; Volume 2020, p. 17328.
10. Kim, A.; Johnson, K. Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook. *Comput. Hum. Behav.* **2016**, *58*, 98–108. [[CrossRef](#)]
11. Kumar, A.; Sangwan, S.; Nayyar, A. Rumour veracity detection on twitter using particle swarm optimized shallow classifiers. *Multimed. Tools Appl.* **2019**, *78*, 24083–24101. [[CrossRef](#)]
12. Cho, S.; Cha, M.; Sohn, K. Topic category analysis on twitter via cross-media strategy. *Multimed. Tools Appl.* **2016**, *75*, 12879–12899. [[CrossRef](#)]
13. Medvedev, A.; Lambiotte, R.; Delvenne, J. The anatomy of Reddit: An overview of academic research. In *Dynamics on and of Complex Networks*; Springer: Berlin, Germany, 2017; pp. 183–204.
14. 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](#)]
15. Alnazzawi, N. Using Twitter to Detect Hate Crimes and Their Motivations: The HateMotiv Corpus. *Data* **2022**, *7*, 69. [[CrossRef](#)]
16. Achimescu, V.; Chachev, P.D. Raising the flag: Monitoring user perceived disinformation on reddit. *Information* **2020**, *12*, 4. [[CrossRef](#)]
17. Guidi, B.; Michienzi, A.; Salve, A.D. Community evaluation in Facebook groups. *Multimed. Tools Appl.* **2020**, *79*, 33603–33622. [[CrossRef](#)]
18. Amati, G.; Angelini, S.; Gambosi, G.; Rossi, G.; Vocca, P. Influential users in Twitter: Detection and evolution analysis. *Multimed. Tools Appl.* **2019**, *78*, 3395–3407. [[CrossRef](#)]
19. Erl, T.; Khattak, W.; Buhler, P. *Big Data Fundamentals—Concepts, Drivers & Techniques*; Prentice Hall: Hoboken, NJ, USA, 2015.
20. Dellarocas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Manag. Sci.* **2003**, *49*, 1407–1424. [[CrossRef](#)]
21. Arndt, J. Role of product-related conversations in the diffusion of a new product. *J. Mark. Res.* **1967**, *4*, 291–295. [[CrossRef](#)]
22. Katz, E.; Lazarsfeld, P. *Personal Influence, The Part Played by People in the Flow of Mass Communications*; Transaction Publishers: Piscataway, NJ, USA, 1966.
23. Dean, D.; Lang, J. Comparing three signals of service quality. *J. Serv. Mark.* **2008**, *22*, 48–58 [[CrossRef](#)]
24. Cassavia, N.; Masciari, E.; Pulice, C.; Saccà, D. Discovering User Behavioral Features to Enhance Information Search on Big Data. *ACM Trans. Interact. Intell. Syst.* **2017**, *7*, 1–33. [[CrossRef](#)]
25. Jansen, B.; Zhang, M.; Sobel, K.; Chowdury, A. Twitter power: Tweets as electronic word of mouth. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 2169–2188. [[CrossRef](#)]
26. Zhang, P.; Lee, H.; Zhao, K.; Shah, V. An empirical investigation of eWOM and used video game trading: The moderation effects of product features. *Decis. Support Syst.* **2019**, *123*, 113076. [[CrossRef](#)]
27. Huete-Alcocer, N. A literature review of word of mouth and electronic word of mouth: Implications for consumer behavior. *Front. Psychol.* **2017**, *8*, 1256. [[CrossRef](#)] [[PubMed](#)]
28. Chu, S.; Kim, Y. Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites. *Int. J. Advert.* **2011**, *30*, 47–75. [[CrossRef](#)]
29. Horng, S.; Wu, C. How behaviors on social network sites and online social capital influence social commerce intentions. *Inf. Manag.* **2020**, *57*, 103176. [[CrossRef](#)]
30. Sohaib, M.; Hui, P.; Akram, U.; Majeed, A.; Tariq, A. How Social Factors Drive Electronic Word-of-Mouth on Social Networking Sites? In Proceedings of the International Conference on Management Science and Engineering Management (ICMSEM'19), Toronto, ON, Canada, 5–8 August 2019; Springer: Berlin, Germany, 2019; pp. 574–585.
31. Wang, T.; Yeh, R.; Chen, C.; Tsydypov, Z. What drives electronic word-of-mouth on social networking sites? Perspectives of social capital and self-determination. *Telemat. Inform.* **2016**, *33*, 1034–1047. [[CrossRef](#)]
32. Brown, J.; Broderick, A.; Lee, N. Word of mouth communication within online communities: Conceptualizing the online social network. *J. Interact. Mark.* **2007**, *21*, 2–20. [[CrossRef](#)]
33. Luo, Q.; Zhong, D. Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites. *Tour. Manag.* **2015**, *46*, 274–282. [[CrossRef](#)]
34. Adamopoulos, P.; Ghose, A.; Todri, V. The impact of user personality traits on word of mouth: Text-mining social media platforms. *Inf. Syst. Res.* **2018**, *29*, 612–640. [[CrossRef](#)]
35. Bae, Y.; Ryu, P.; Kim, H. Predicting the lifespan and retweet times of tweets based on multiple feature analysis. *Etri J.* **2014**, *36*, 418–428. [[CrossRef](#)]
36. Kong, S.; Feng, L.; Sun, G.; Luo, K. Predicting lifespans of popular tweets in microblog. In Proceedings of the Special Interest Group on Information Retrieval (SIGIR'12), Portland, OR, USA, 16 August 2012; pp. 1129–1130.
37. Sun, B.; Ng, V. Lifespan and popularity measurement of online content on social networks. In Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI'11), Beijing, China, 10–12 July 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 379–383.

38. Yang, J.; Leskovec, J. Patterns of Temporal Variation in Online Media. In Proceedings of the International Conference on Web Search and Web Data Mining (WSDM 2011), Hong Kong, China, 9–12 February 2011; ACM: New York, NY, USA, 2011; pp. 177–186.
39. Peri, S.; Chen, B.; Dougall, A.; Siemens, G. Towards understanding the lifespan and spread of ideas: Epidemiological modeling of participation on Twitter. In Proceedings of the International Conference on Learning Analytics & Knowledge (LAK'20), Frankfurt Germany, 23–27 March 2020; pp. 197–202.
40. Fiebert, M.; Aliee, A.; Yassami, H.; Dorethy, M. The life cycle of a Facebook post. *Open Psychol. J.* **2014**, *7*, 18–19. [[CrossRef](#)]
41. Bonifazi, G.; Corradini, E.; Marchetti, M.; Sciarretta, L.; Ursino, D.; Virgili, L. A Space-Time Framework for Sentiment Scope Analysis in Social Media. *Big Data Cogn. Comput.* **2022**, *6*, 130. [[CrossRef](#)]
42. Spasojevic, N.; Li, Z.; Rao, A.; Bhattacharyya, P. When-to-post on social networks. In Proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD'15), Sydney, NSW, Australia, 10–13 August 2015; pp. 2127–2136.
43. Shin, J.; Jian, L.; Driscoll, K.; Bar, F. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Comput. Hum. Behav.* **2018**, *83*, 278–287. [[CrossRef](#)]
44. Alkhamees, N.; Fasli, M. Event detection from social network streams using frequent pattern mining with dynamic support values. In Proceedings of the International Conference on Big Data (BigData'16), Washington, DC, USA, 5–8 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1670–1679.
45. Shen, J.; Rudzicz, F. Detecting anxiety through Reddit. In Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality, Vancouver, BC, Canada, 3 August 2017; Association for Computational Linguistics: Cedarville, OH, USA, 2017; pp. 58–65.
46. Buntain, C.; Golbeck, J. Identifying Social Roles in Reddit Using Network Structure. In Proceedings of the International Conference on World Wide Web (WWW'14), Seoul, Republic of Korea, 7–11 April 2014; ACM: New York, NY, USA, 2014; pp. 615–620.
47. 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), Online, 19–20 November 2020; AAAI Press: Palo Alto, CA, USA, 2020; Volume 14, pp. 830–839.
48. Weninger, T. An exploration of submissions and discussions in social news: Mining collective intelligence of Reddit. *Soc. Netw. Anal. Min.* **2014**, *4*, 173–192. [[CrossRef](#)]
49. Newell, E.; Jurgens, D.; Saleem, H.; Vala, H.; Sassine, J.; Armstrong, C.; Ruths, D. User Migration in Online Social Networks: A Case Study on Reddit During a Period of Community Unrest. In Proceedings of the International Conference on Web and Social Media (ICWSM 2016), Cologne, Germany, 17–20 May 2016; AAAI: Palo Alto, CA, USA, 2016; pp. 279–288.
50. Soliman, A.; Hafer, J.; Lemmerich, F. A Characterization of Political Communities on Reddit. In Proceedings of the ACM Conference on Hypertext and Social Media (HT'19), Berlin, Germany, 16–23 September 2019; ACM: New York, NY, USA, 2019; pp. 259–263.
51. 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), Munich, Germany, 11–14 June 2019; AAAI: Palo Alto, CA, USA, 2019; pp. 205–213.
52. Banerjee, C.; Mukherjee, T.; Pasilio, E., Jr. An empirical study on generalizations of the ReLU activation function. In Proceedings of the ACM Southeast Conference (ACM-SE'19), Kennesaw, GA, USA, 18–20 April 2019; ACM: New York, NY, USA, 2019; pp. 164–167.
53. Godes, D.; Silva, J. Sequential and temporal dynamics of online opinion. *Mark. Sci.* **2012**, *31*, 448–473. [[CrossRef](#)]
54. Li, X.; Hitt, L. Self-selection and information role of online product reviews. *Inf. Syst. Res.* **2008**, *19*, 456–474. [[CrossRef](#)]
55. Moe, W.; Schweidel, D. Online product opinions: Incidence, evaluation, and evolution. *Mark. Sci.* **2012**, *31*, 372–386. [[CrossRef](#)]
56. Corradini, E.; Nocera, A.; Ursino, D.; Virgili, L. Defining and detecting k-bridges in a social network: The Yelp case, and more. *Knowl.-Based Syst.* **2020**, *187*, 104820. [[CrossRef](#)]
57. Rios, S.; Aguilera, F.; Nuñez-Gonzalez, J.; Graña, M. Semantically enhanced network analysis for influencer identification in online social networks. *Neurocomputing* **2019**, *326*, 71–81. [[CrossRef](#)]
58. Graves, R.L.; Perrone, J.; Al-Garadi, M.A.; Yang, Y.C.; Love, J.; O'Connor, K.; Gonzalez-Hernandez, G.; Sarker, A. Thematic analysis of reddit content about buprenorphine-naloxone using manual annotation and natural language processing techniques. *J. Addict. Med.* **2022**, *16*, 454. [[CrossRef](#)]
59. Chandrasekharan, E.; Jhaver, S.; Bruckman, A.; Gilbert, E. Quarantined! Examining the effects of a community-wide moderation intervention on Reddit. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2022**, *29*, 1–26. [[CrossRef](#)]
60. Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [[CrossRef](#)]
61. Cauteruccio, F.; Terracina, G.; Ursino, D. Generalizing identity-based string comparison metrics: Framework and Techniques. *Knowl.-Based Syst.* **2020**, *187*, 104820. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.