Mining and comparing engagement dynamics across multiple social media platforms

Conference or Workshop Item

Mining and Comparing Engagement Dynamics Across Multiple Social Media Platforms

Matthew Rowe
School of Computing and Communications,
Lancaster University,
Lancaster, UK
m.rowe@lancaster.ac.uk

Harith Alani
Knowledge Media Insitutte
Open University,
Milton Keynes, UK
h.alani@open.ac.uk

ABSTRACT

Understanding what attracts users to engage with social media content is important in domains such as market analytics, advertising, and community management. To date, many pieces of work have examined engagement dynamics in isolated platforms with little consideration or assessment of how these dynamics might vary between disparate social media systems. Additionally, such explorations have often used different features and notions of engagement, thus rendering the cross-platform comparison of engagement dynamics limited. In this paper we define a common framework of engagement analysis and examine and compare engagement dynamics across five social media platforms: Facebook, Twitter, Boards.ie, Stack Overflow and the SAP Community Network. We define a variety of common features (social and content) to capture the dynamics that correlate with engagement in multiple social media platforms, and present an evaluation pipeline intended to enable cross-platform comparison. Our comparison results demonstrate the varying factors at play in different platforms, while also exposing several similarities.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.2.8 [**Database Management**]: Database Applications—*Data Mining*

General Terms

Experimentation, Measurement

Keywords

Social Media, Engagement, Data Mining

1. INTRODUCTION

The rise of the information age has led to the increased need for users to allocate their attention in a more intelligent and considered fashion. Nowhere better is this manifest than on social media, where the rate at which social data is produced and the scale at which it is available has galvanised research into understanding attention through the guise of *engagement dynamics*: understanding how and why users engage with certain pieces of social media content (i.e. status updates) and not others. The free, open and widely-used nature of social media means that several parties have a vested interest in understanding such dynamics for their own needs. For instance, marketing a product or injecting content into a social media platform, with the desire for users to engage with such content, requires understanding what factors are associated with engagement and how these differ between platforms. It could be the case that using familiar language to platform users is important in one context, while not so in another - for instance when advertising a product or event in a dedicated topical web forum.

Despite the emergence of a large-body of literature examining engagement dynamics, such works have thus far focused on different platforms using disparate approaches without considering a unified assessment that spans social media. As a result, findings that have emerged to date have not always been consistent in terms of: (i) examined features; and (ii) the association of such features with engagement in terms of both magnitude and sign - i.e. certain works have concentrated on the role of content features in initiating engagement, whereas other works have only considered social features. Furthermore, the actual action that represents engagement is found to be disparate between such works: for instance, on Twitter this can vary between identifying engagement when an individual *favourites* a status update, *retweets* a message [4, 18, 9, 8], or when a user *replies-to* a message [14, 15, 17], thereby representing differing modalities of engagement.

Social media platforms serve different purposes, offer different capabilities, and cultivate different social norms, therefore one would anticipate the underlying dynamics (communication, interaction, behavioural) on social media to differ from one platform to another. To reach an understanding of the persistent dynamics that emerge on each platform it is therefore necessary to establish a common evaluation framework that can be applied across different social media platforms, and to compare and contrast the results of applying this framework to a range of platforms. This is what we set out to achieve in this paper by presenting a comparative study of the engagement dynamics across five social media platforms: Boards.ie, Twitter, SAP Community Network, Stack Overflow and Facebook. In performing this study we make the following contributions:

- We define a collection of social and content features that are common across the chosen social platforms, chosen from related work intended to capture factors that influence engagement.
- We present a machine-learning based approach for engagement prediction (defined as a binary classification problem) that includes feature standardisation, dataset balancing via under sampling, and time-ordering to enable inter-social media comparison of engagement dynamics.
- We contrast the role of different features on engagement likelihood across five social media platforms, thereby mining and comparing *engagement dynamics*, and contrast these findings with the engagement dynamics from existing studies on individual social media platforms.

To the best of our knowledge this is the first work that sets out to perform inter-social media analysis of engagement dynamics. By building upon related work for the chosen features to analyse, and comparing our results with prior findings, our work contributes to the domain of Web Science with an understanding of how engagement differs across social media. The findings from our work can be used by, and will have implications for, online product marketers and any content publishers keen to increase the potential for users to engage with their published content.

We have structured the paper as follows: section 2 describes the related work within the field of engagement dynamics, with an emphasis on data mining oriented research towards extracting engagement dynamics. Section 3 describes the datasets that we used for our experiments, the collection methods employed, and information describing the magnitude of the collected datasets. Section 4 explains the features that we engineered for the experiments and how they can be implemented by other researchers. Section 5 details the experiments that were performed to mine engagement dynamics including predicting which posts would be engaged with, the experimental setup that we followed, and the dynamics of engagement that we extracted. Section 6 discusses implications of this work and section 7 finishes the paper with conclusions drawn from the work.

2. RELATED WORK

Recent years have seen a large body of research begining to emerge on measuring and predicting attention generation across social media platforms. Attention has been measured in different forms (e.g. retweets, replies to comments, popularity of posts and answers) in search of a better understanding of its dynamics and the features that influence it. For retweets, it was found that content features were more influential than social features for determining whether a tweet will be retweeted or not [4, 18, 9, 8]. For example, the presence of URLs and hashtags in the tweets were often found to be good indicators of retweetability [12, 18]. Some social features, such as #followers, #followees, and account age were found to have some impact on retweet predictions, although less than content features [18]. Previous tweeting activities of users do not seem to encourage retweets. In fact, it was found that the more users tweet and favourite tweets, the less their tweets get retweeted [18]. Hodas et al [11] stress that the rapid visual decay of tweets on Twitter clients is a significant factor in retweetability. The authors also found that tweets are less likely to be retweeted by people who follow many users due to the thin spread of their attention to too many users and tweets. The role of topics in attention generation were also investigated. It was found that people are less likely to retweet on topics that they themselves tweet about [9], and that tweets on topics of general interest are more likely to be retweeted [12]. However, in Boards.ie, users with high topic entropy (i.e. tend to post about the same topics) seem to receive more replies [13].

Contradictory to the case with retweets, when predicting replies on Twitter, it was found that social features (e.g., #followers, #lists) play a more important role than content features and tweet topics [14, 15, 17]. Sousa and colleagues showed that Twitter replies by users with smaller social networks are more driven by social aspects than users with larger ego-networks [17]. Nevertheless, as with retweets, combining social and content features produced the best predictions of replies on Twitter [14, 20, 15]. However, when analysing replies on Boards.ie, it was found that content features are better for predictions than social features, thus contradicting the findings obtained from predicting replies on Twitter, and instead agreeing with the predictions of retweets [13]. This highlights the role played by the type and goal of the communities on their engagement dynamics and associated features. For example, the presence of a URL in posts on Boards.ie is only good for generating a reply in general forums, and users with low forum entropy, account age, and #posts, are less likely to get replies in support communities [20]. Other variations in dynamics were observed across topics on Yahoo! Answers [1]. Here it was found that lower entropy of the answerer is good for predicting best answers, but only in technical topics and others where factual information is needed. For Yahoo! Answers, combining content and social features also provided the best predictions of best answers [3], which matches the findings from other social media platforms.

Models for predicting the popularity of comments on various social platform have also been explored in the literature. For example, features such as sentiment, and ratings of new comments were shown to be sufficient for predicting popularity of comments on YouTube [16]. In a study on the popularity of comments on Digg, Hsu and colleagues [7] showed that social features were better than content features for such predictions, which corresponds to the results obtained from predicting Twitter replies. Similar to the studies on Twitter and Boards.ie, combining social and content features also produced best predictions in Digg. Digg comments' popularity dynamics produced a few more patterns that contradict the findings obtained from Twitter. For example, unlike the case with retweets, previous activities of Digg users proved to be important for predicting popularity of their comments [7]. On the other hand, account age, which was deemed important for predicting retweets [18], is one of the least influential on Digg predictions. Furthermore, topic entropy, which was found to be good for predicting tweets' popularity[18], is much less useful for predicting Digg comments' popularity [7]. As for predicting the popularity of *articles* on Digg, it was found that content feature are most valuable if no *click* data is available yet, whereas social features are more useful for predicting popularity when the visibility of a Digg article or a YouTube video is limited to a small number of users [19] . On Slashdot, content features (mainly quality of content) had a significant impact on the number of replies to comments, whereas the influence of the reputation of users was relatively weak [5]. This resembles the case with retweets, and replies on Boards.ie, but not replies on Twitter, or popularity of Digg comments [7]. In Facebook, it was found that social ties between users are good for predicting the length of reply chains [2]; the authors also found that the time it takes for the first reply to arrive is another good indicator of the length of the thread.

The above studies clearly demonstrate the many variations in how the dynamics of engagement differ across the various social platforms. In this paper we concentrate on the process of a user *replying*) to content as a modality of engagement and apply a unified collection of features to datasets collected from five social media platforms. In doing so, we aim to reach a better understanding of how general or specific some of these findings are to individual social platforms.

3. SOCIAL MEDIA DATASETS

For our experiments we used data collected from five distinct social media platforms, each platform providing a different type of functionality and set of community forming features. Descriptive statistics for each dataset are shown in Table 1, these datasets were obtained from the following platforms:

3.1 Boards.ie

Boards.ie is the most-popular Irish community message board and provides a large number (>600) of discussion forums where each forum contains posts related to a distinct topic (e.g. Football, Xbox). Users do not build social networks on the platform, and instead the reply graph is used to construct social networks based on implicit edges from which we derive the social features described below. We define a seed post on this platform as a post that is the first in a discussion thread and receives a reply from another user, conversely a non-seed is a post that is not engaged with (has no replies). We use one dataset for our experiments from Boards.ie as described in Table 1.

3.2 Twitter

Twitter is a microblogging platform that allows users to post messages (Tweets) up to 140 characters in length. Users *follow* other users such that they subscribe to their content (Tweets) and receive them in their timeline, users can also *Retweet* other users' messages which then propagates those message through their follower network. We provide three datasets collected for our experiments: a random corpus (Twitter Random), a corpus of Tweets collected during the Haiti earthquake (Twitter Haiti) and a corpus of Tweets collected during president Obama's state of the union address in 2010 (Twitter Union), all of which are described in Table 1. For each dataset we were originally provided with a collection of Tweet IDs that were collected during the event (Haiti earthquake, State of the Union address) or time period (random collection). Some Tweets were replies to others, therefore we had to collect the seed posts that originally started the chain: we iteratively moved up the reply chain - i.e. from reply to parent post - until we reached the seed post in the discussion by querying Twitter's REST API[1] for the original post that was replied-to. Therefore in the Twitter datasets, collected Tweets that received no reply are non-seeds, and the root post that initiated the discussion chain are seeds.

3.3 SAP Community Network (SAP)

The SAP Community Network is a community question answering system related to SAP technology products and information technologies. Users sign up to the platform and post questions related to technical issues, other users then provide answers to those questions and should any answers satisfy the original query, and therefore solve the issue, the answerer is awarded points. Hence, on SAP, there is prestige attached to the accruing of points over time as a large cumulation indicates expertise within a technical domain. Similar to Boards.ie, SAP provides no explicit means for a user to befriend or follow another user, therefore to construct social networks, and hence derive our social features that we will define below, we use the reply-to graph to form implicit connections between users. We define a seed post on this platform as any post that is the first in a discussion thread, and therefore a post asking for help or a question, that is engaged with by a community member, while conversely, a non-seed post is post that receives no engagement. We use one dataset for our experiments from SAP.

3.4 Server Fault

Similar to SAP, Server Fault is a platform that is part of the Stack Overflow question answering site collection.[2] The platform functions in a similar vein to SAP by providing users with the means to post questions pertaining to a variety of server-related issues, and allowing other community members to reply with potential answers. Answers are then voted by the community as being the best one, and the original question poster can also select his chosen best answer. Similar to SCN, Server Fault also lacks explicit edge-creation features, therefore we use the reply-to graph (i.e. where a user has replied to another user's question) to form an implicit edge between the users. A seed post on Server Fault is any question that is engaged with by community members (i.e. is replied to with answers), while a non-seed post is any post that fails to receive engagement. We use one dataset from Server Fault.

3.5 Facebook

For our final dataset we use data obtained from Facebook groups related to university course discussions. The groups enable users to connect and discuss all kinds of issues with their degree course material and potential avenues for solving any related problems. Although Facebook provides the ability to collect social network data for users, we opted to using the reply-to graph within the groups to build those social networks for individual users. In doing so we would constrain the social dynamics at play to those within the context of the groups. We define a seed post in this context

[1]`http://dev.twitter.com`
[2]`http://stackoverflow.com/`

Table 1: Statistics of the collected social media datasets that we used for our experiments. The seeds and non-seeds counts differ for various platforms indicating the extent to which class imbalance is evident.

Platform	Time Span	Post Count	User Count	Seeds	Non-seeds	Replies
Boards.ie	[01-01-2005,13-02-2008]	6,120,008	65,528	398,508	81,273	5,640,227
Twitter Random	[24-03-2007,25-07-2011]	1,468,766	753,722	144,709	930,262	390,795
Twitter Haiti	[28-05-2009,13-10-2010]	65,022	45,238	1,835	60,686	2,501
Twitter Union	[03-08-2009,28-01-2010]	81,458	67,417	11,298	56,135	14,025
SAP	[15-12-2003,20-07-2011]	427,221	32,926	87,542	7,276	332,403
Server Fault	[01-08-2008,31-03-2011]	234,790	33,285	65,515	6,447	162,828
Facebook	[18-08-2007,24-01-2013]	118,432	4,745	15,296	8,123	95,013

as any post that starts a thread in a discussion group where users engage with the post, while a non-seed is any post that fails to attract engagement from community members.

4. FEATURE ENGINEERING

For our experiments we wanted to see how existing social and content dynamics function across social media. As we have pointed out in the related work section, there are a variety of works that have examined different features on different platforms. However, in several cases the portability of such features to different platforms is limited (e.g. hashtags on Boards.ie) and therefore considerations must be made when compiling the feature sets to enable cross-platform inspection of engagement dynamics. This section therefore provides descriptions of *intersecting* features that function across all five platforms under inspection. Investigating platform-specific features is out of the scope of this study, where we only focus on common and comparable features. We begin by defining the social features, before going on to explain content features, and at each step highlighting the computational aspects of the features.

4.1 Social Features

Social features capture the social network properties of the author of a post and his activity and time on the platform. These features have been used extensively in previous works when examining the effect that the social network position and audience size has on a user's ability to initiate engagement with platform users. We define five social features as follows:

- *In-degree:* For the author of each post (seed or non-seed), this feature measures the number of incoming connections to the user. On platforms where edges are explicitly defined between users (Twitter) we count the number of followers a user has, on platforms where edges are implicit - i.e. via the reply-to graph where user B replied to a post by user A then we say that a directed edge from user B connects to user A - this is the total number of repliers to a given user (in this instance we use 6-month window prior to when the post was made based on prior work [13]).
- *Out-degree:* This feature measures the number of outgoing connections from the user. In a similar manner to in-degree we use the explicit edges from Twitter and the implicit edges gleaned from the reply-to graph for the other platforms (again a 6-month prior window from the post data is used).
- *Post Count:* Measures the number of posts that the user has made over the previous 6-months.
- *User Age:* Measures the length of time that the user has been a member of the community in days.
- *Post Rate:* Measures the number of posts made by the user per day.

4.2 Content Features

Content features capture the qualities and characteristics of a given post and have also been used throughout the related work. We found that existing attempts to characterise engagement through content dynamics often include features which are not portable to different platforms (e.g. @mentions on Twitter, hashtags, etc.), therefore we have seven common features that can be computed across our five social media platforms. These are defined as follows:

- *Post Length:* Number of word tokens in the post.
- *Complexity:* Measures the cumulative entropy of terms within the post to gauge the concentration of language and its dispersion across different terms. Let $T(p)$ denote a function that returns the unique terms in post p and $tf(t,p)$ denote a function that returns the frequency of token $t \in T(p)$ in post p. The complexity of post p is defined as follows:

$$complexity(p) = \frac{1}{|T(p)|} \sum_{t \in T(p)} tf(t,p)(\log |T| - \log tf(t,p)) \tag{1}$$

  This measure returns a high entropy if the post contains many terms which are not repeated often, and thus the random variable's entropy is increased, while a low entropy denotes repetition of terms from a limited vocabulary.
- *Readability-Fog:* Gunning fog index using average sentence length (ASL) [6] and the percentage of complex words (PCW): $0.4*(ASL+PCW)$ This feature gauges how hard the post is to parse by authors.
- *Readaibility:* LIX Readability metric. As opposed to Gunning Fog, this metric determines complexity of words based on the number of letters rather than on the number of syllables. The Readability of a post is computed as:

$$Readability(p) = \frac{|Words|}{|Sentences|} + \frac{|Words > 6letters|}{|Words|} * 100 \quad (2)$$

- *Referral Count:* Count of the number of hyperlinks within the post. This measure is sometimes used as a naive spam measure, where posts with many hyperlinks could by for advertising some product or event and might therefore be less likely to lead to engagement.

- *Informativeness:* The novelty of the post's terms with respect to other posts. We derive this measure using the Term Frequency-Inverse Post Frequency (TF-IDF) measure:

$$informativeness(p) = \sum_{t \in T(p)} tf(t,p) \times ipf(t) \quad (3)$$

  This measure will return high informativeness if the post contains unique terms with respect to the platform's vocabulary, while if the post contains terms that are familiar to the platform's users it will return a low informativeness value.

- *Polarity:* This measure assesses the average polarity of a post using SentiWordnet.[3] Our inclusion of this feature is to assess whether either positive or negative post polarity is associated with seeds or non-seeds, or whether subjective or objective posts also have an association. Let $T(p)$ denote a function returning the set of unique terms in post p, the function $pos(t)$ returns the positive weight of the term t from the lexicon and $neg(t)$ returns the negative weight of the term. We therefore define the polarity of p as:

$$polarity(p) = \frac{1}{|T(p)|} \sum_{t \in T(p)} pos(t) - neg(t) \quad (4)$$

5. MINING ENGAGEMENT DYNAMICS

Identifying which factors correlate with engagement across different social media platforms requires examining the contribution of individual features to predictive performance and then inspecting the effects of those features. In this section we describe our experiments to predict which posts will be seeds and which will be non-seeds, and compare our results and the findings in relation to existing work from the state of the art.

5.1 Experimental Setup

To uncover engagement dynamics across disparate social media systems we first derived the set of posts that would constitute the instances in each platform's dataset ($D = \{(\mathbf{x}_i, y_i)\}$ - this is to train a machine learning classifier. As shown in Table 1, there are large class imbalances between the seeds and non-seeds in the differing datasets - sometimes where there are more seeds than non-seeds and vice versa in other times. In order to ensure that we have a balanced class distribution in each dataset, we performed random *undersampling* from the dominant class (seed or non-seed) from each respective dataset. This resulted in a 50:50 split between seeds and non-seeds in our datasets - the resultant number of instances within each dataset is shown in Table 2. This method enables the incorporation of several baselines - as we will discuss shortly - that in turn enhance the assessment of model performance and make the process more straightforward.

After balancing the datasets' seeds-to-non-seeds distribution, we then constructed each post's instance features using the previously described features; this resulted in a vector representation of each post ($\mathbf{x} \in \mathbb{R}^{12}$). Within each dataset we then *standardised* each feature by normalising the respective feature value from each instance to have unit variance (i.e. $N(0,1)$) and thus converting it to a z-score according to the feature distribution. By performing this conversion we were provided with standardised datasets from which model coefficients can be compared, once induced, without the limitation of outlier values skewing the coefficients. This final process resulted in the construction of each platform's dataset($D = \{(\mathbf{x}_i, y_i)\}$) as a set of pairs mapping each instance to its class label, where $y_i \in \{0,1\}$ - with 0 denoting a non-seed and 1 denoting a seed. We maintained time ordering of the datasets such that posts, both seeds and non-seeds, were kept in an ascending publication date order and segmented each dataset into a training and test split using the standard 80/20% splits respectively.

For our prediction experiment we induced a logistic regression model using the training split and applied it to the test split. We trained the model using different feature sets (e.g. social, content, social+content) to see which feature set performed best and how this differed between the various datasets and platforms. We then inspected the coefficients of the logistic regression model of each platform to see how a change in each feature was associated with the likelihood of engagement. By performing this inspection we could see how engagement dynamics in each of the studied platforms contrasted against the related work - i.e. how a change in the magnitude of a feature would impact the log-odds of the classifier, and hence the likelihood of the post being engaged with.

5.1.1 Evaluation Measures and Baselines

To assess the performance of our models we used the standard classification performance measures of precision, recall and f-measure (F1: with $\beta = 1$ to count precision and recall equally). We also measured the Matthews' Correlation Coefficient (MCC) as a means to contrast our performance against a random guesser baseline. An MCC of +1 indicates perfect performance (i.e. matching predicted labels with observed labels), while a value of -1 indicates complete disagreement between the predictions and observed labels, and a value of 0 indicates that the performance is on a par with a random guesser.[4] Therefore, models should aim to surpass MCC=0, and thus beat a random guessing model.

[3] http://sentiwordnet.isti.cnr.it/

[4] MCC is calculated from classification contingency tables and the χ^2 test statistic divided by the set size and then the square root is taken.

Table 2: Statistics of the social media datasets that we used for our experiments once under sampling has been applied

Platform	Time Span	Seeds	Non-seeds	Instance Count
Boards.ie	[01-01-2005,13-02-2008]	398,508	81,273	162,546
Twitter Random	[24-03-2007,25-07-2011]	144,709	930,262	289,418
Twitter Haiti	[28-05-2009,13-10-2010]	1,835	60,686	3,670
Twitter Union	[03-08-2009,28-01-2010]	11,298	56,135	22,596
SAP	[15-12-2003,20-07-2011]	87,542	7,276	14,552
Server Fault	[01-08-2008,31-03-2011]	65,515	6,447	12,894
Facebook	[18-08-2007,24-01-2013]	15,296	8,123	16,246
Total				521,922

5.2 Results

We begin by examining the performance of different feature sets on predicting seed posts and how these feature sets differ across the platforms. Table 3 presents the performance that the logistic regression model achieves when trained on isolated feature sets (i.e. social features), and then all features together. We note that for the isolated feature sets **content** achieves the best performance (in terms of F1) for Boards.ie, Twitter Random, Twitter Union and Facebook, while **social** features perform best for Twitter Haiti, Server Fault, and Facebook. When we combine the features together we find that for every platform we exceed the performance of using solitary feature sets. This indicates that different platforms have factors influencing users' engagement with content, however the importance of both social and content dynamics is paramount. We will delve into how such dynamics differ below.

Table 3: Performance of the logistic regression classifier trained over different feature sets and applied to the test set.

(a) Boards.ie

Features	P	R	F1	MCC
Social	0.592	0.591	0.591	0.092
Content	0.664	0.660	0.658	0.162
Social+Content	0.670	0.666	0.665	0.168

(b) Twitter Random

Features	P	R	F1	MCC
Social	0.561	0.561	0.560	0.061
Content	0.612	0.612	0.611	0.112
Social+Content	0.628	0.628	0.628	0.128

(c) Twitter Haiti

Features	P	R	F1	MCC
Social	0.968	0.966	0.966	0.482
Content	0.752	0.747	0.747	0.250
Social+Content	0.974	0.973	0.973	0.493

(d) Twitter Union

Features	P	R	F1	MCC
Social	0.542	0.540	0.539	0.042
Content	0.650	0.642	0.639	0.147
Social+Content	0.656	0.649	0.646	0.153

(e) SAP

Features	P	R	F1	MCC
Social	0.650	0.631	0.628	0.142
Content	0.575	0.541	0.521	0.063
Social+Content	0.652	0.632	0.629	0.144

(f) Server Fault

Features	P	R	F1	MCC
Social	0.528	0.380	0.319	-0.014
Content	0.626	0.380	0.275	0.032
Social+Content	0.568	0.407	0.359	0.012

(g) Facebook

Features	P	R	F1	MCC
Social	0.635	0.632	0.632	0.133
Content	0.641	0.641	0.641	0.140
Social+Content	0.660	0.660	0.660	0.158

5.2.1 Feature Effects

Fig. 1 presents bar plots of the feature coefficients in the logistic regression model. A positive value coefficient for a given feature (i.e. appearing above the x-axis) indicates that an increase in the magnitude of this feature has a positive bearing on the probability of a post initiating engagement. Conversely, a negative value (i.e. appearing below the x-axis) indicates that the feature has a negative effect on engagement probability, in essence the coefficients are log-odds ratios. Therefore by inspecting the coefficients of the model we can examine how engagement dynamics differ between social media platforms and across the features. The logistic regression model also includes significance probabilities for each calculated coefficient under the null hypothesis a given coefficient is 0 (and thus has no effect on the engagement likelihood). We only report on features whose inclusion in the model is significant at the **5% significance** level - Fig. 2 shows the plot of these significance probabilities for each dataset's features.

Fig. 1 indicates that there are clear differences in the engagement patterns between the examined social media platforms and also within the platforms themselves (i.e. the differing effects for the Twitter datasets between the random corpus, the Haiti-specific corpus and the State of the Union Address corpus). For instance, when we examine the social features we see that for the **in-degree** of users, an increase in the in-degree is associated with an increase in engagement likelihood for all datasets except for Twitter Union, suggesting that the number of followers in this context could have a negative effect on the probability of a user replying to Tweets - i.e. more popular or listened-to individuals were ignored when discussing the political topic of the State of the Union address.

For the **out-degree** of a user a reduced value is associated with an increase in engagement likelihood for all social media datasets except for Twitter Union and Twitter Haiti, suggesting that the propensity of users to follow (Twitter Random) and reply-to (Boards.ie, Server Fault, SAP, Facebook) other users have a negative impact on engagement probability. For the **post count** of users we observe consistent effects for the Twitter datasets: an increase in the number of Tweets that a user publishes is associated with an increase in engagement likelihood. However, for the other datasets we find the opposite to be true: increased posting is associated with a decrease in engagement probability. This suggests that for conversation and discussion-oriented social media (Boards.ie, Server Fault, SAP, Facebook) an increase in a user's activity can have a detrimental effect on the probability of their future posts being engaged with by community members. It could be that in such contexts a user's activity is picked up as an annoyance and therefore leads to more of their posts being ignored.

Assessing the content feature effects we also note marked differences, we now pick out the salient findings. A reduction in the **referral count** (i.e. number of hyperlinks) in a post was found to seed engagement for all platforms, suggesting that URLs have a detrimental effect on yielding replies from other users. One can imagine that a URL posted within a message could denote a website, product, service or event advertisement, thereby not requiring replies (might instead be retweeted, Liked, etc) or leading users to ignoring such content. In terms of **complexity**, an increase in term-entropy was found to be positively associated with engagement for Boards.ie, Twitter Union, SAP and Facebook, while a decrease was found for the remaining datasets. This finding is interesting as it suggests differences in the engagement dynamics between the two question-answering platforms: SAP and ServerFault. For the former, it appears that users respond to posts which have a more varied vocabulary and are longer in describing their issue (shorter **post length** is associated with seeds on SAP), while for the latter a more terse post is preferred on ServerFault (limited vocabulary and longer). Turning now to the **informativeness** of post content we find that unique terminology with respect to the platform, and thus higher **informativeness**, is preferential for engagement on Twitter (Random and Union), however a reduction of unique terminology is preferential on Boards.ie, Server Fault and Facebook. This finding suggests that for discussion-oriented platforms, users are more likely to reply to posts that contain language which they are familiar with, while for microblogging platforms, which restrict the post length, unique terms lead to engagement.

5.2.2 *Comparison with Related Work*

Analysing the diversity of engagement dynamics both across social media and within the same platform suggests the presence of variance across studies and the disparity between findings from the related work. To ground our findings with that from the related work we compared the engagement dynamics derived from our experiments with findings from the literature. Due to the multidimensional nature of such a comparison, encompassing different feature sets, individual features, and pieces of work, we compiled the table shown in Table 4 to enable a coherent comparison. This table shows how a certain feature performed in our studies versus in other studies that were mostly done on different datasets and platforms. The aim is to highlight general differences and similarities, irrespective of the analysed data and platforms.

Inspection of the comparison table reveals some interesting similarities and differences. For example, we found that **in-degree** is consistent across the related work with our findings: higher in-degree is associated with an increase in the likelihood of engagement.

Out-degree, however, differs: we find that a reduction is beneficial for all but Twitter Union, where such a reduction is actually contrary to what is found in the literature (where an increase in the **out-degree** of users is associated with increased engagement). For **post count** we observe that a decrease for all datasets but Twitter Random and Twitter Union is associated with increased engagement: this agrees with work of Suh et al. [18] on Twitter. When inspecting the effect of **age** in the related work we found that an increase in user **age** was found to be beneficial for engagement in the work of Suh et al. [18] on Twitter, whereas a decrease was found to be better by Hsu et al. [7], where the authors rank comments on weblogs. We also see differences too: decrease across Twitter (thus disagreeing with Shuh et al.) and increase on SAP and Facebook. For **referral count** we found a complete disagreement with related works, where we find an increase in referrals was found to negatively impact engagement likelihood, while the related work found an increase to be better (in Twitter and Weblogs).

6. DISCUSSION AND FUTURE WORK

Our experiments showed that there is a good deal more work to be done in order to reach a greater understanding of how common features influence engagement in different platforms, and even across different datasets from the same platform. In this section we draw attention to a number of related issues that could guide current and future research.

Our selection of features was inspired by the literature, and consisted of features shared by the five platforms we are investigating. This enabled us to compare how these common features correlate with engagement in multiple social platforms. However, we acknowledge the fact that there could be other features, perhaps specific to certain social platforms, that might have a greater influence on engagement. It is also worth noting that the same feature could be used differently in different social platforms, which could explain any variation in their performance for stimulating engagement. These interesting issues are not however within the scope of this study.

For certain features, such an in/out degree, their calculation was done slightly differently across the platform we are investigating. On Twitter, these features were derived from the number of followers and followees (friends), whereas on the other datasets (Boards.ie, Facebook, SAP) they were derived from the user's reply-to network, which was calculated from all the user's reply actions in the past 6 months. It is not possible to mimic this threshold in Twitter, given that the date when a *follow* relationship is created is not supplied by the Twitter API. In future work we intend to harmonise further the calculation of these particular features by collecting timestamped follow relationships on Twitter.

We considered *replies* as indicators of engagement. Others exist, such as *Likes* and *Retweets*, that could be considered as indicators of some form of engagement. Although our

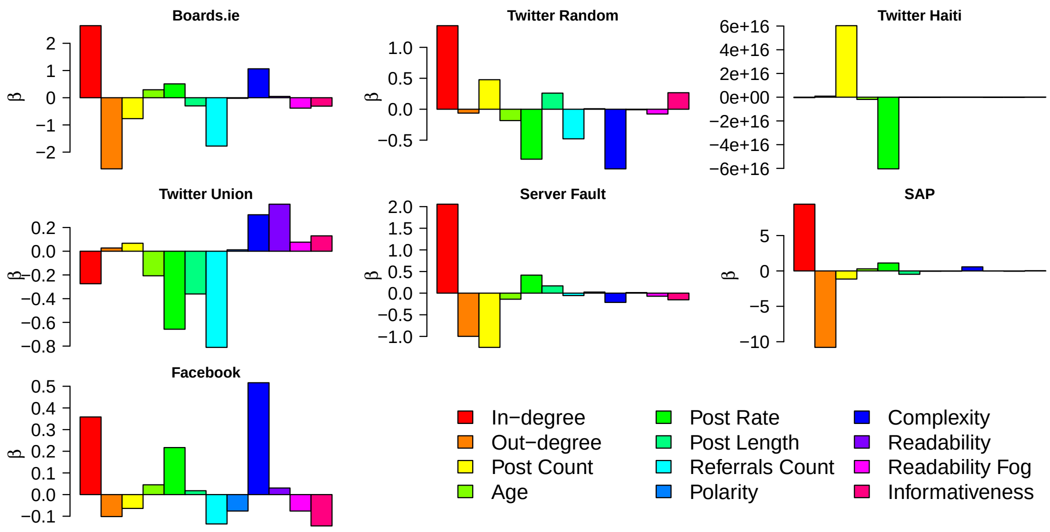

Figure 1: Logistic regression β coefficients for each platform's features. This provides some indication as to the effects of individual features on the response variable (i.e. whether the post seeds engagement or not).

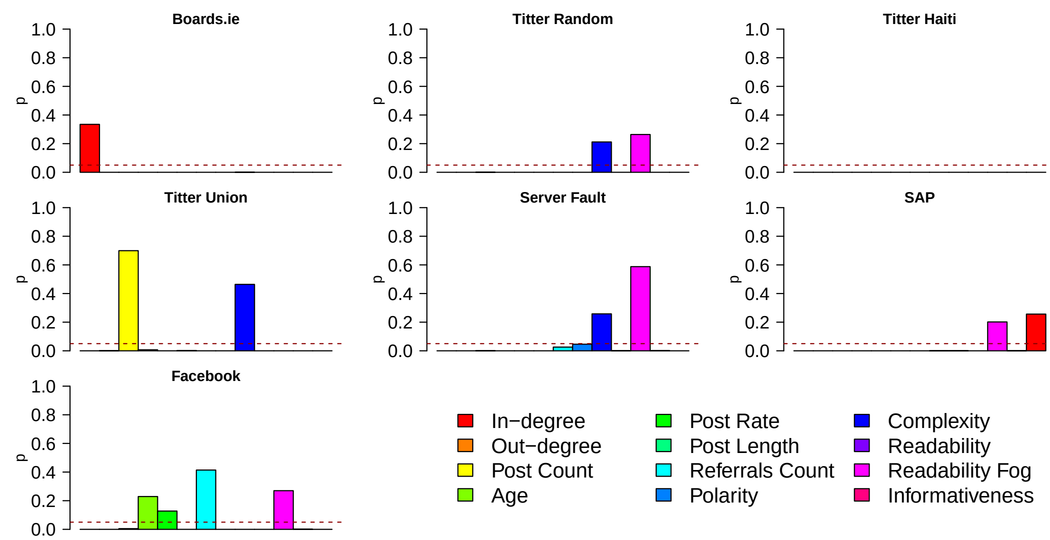

Figure 2: Logistic regression coefficients significance probabilities for each dataset's features. The 5% probability line is marked with the dashed red line.

analysis can be expanded to other indicators, we believe that replies are more likely to indicate a closer or stronger engagement than the other actions - as replying to a person indicates a clear engagement of the replier to the recipient.

Adding more datasets to our analysis, from the same and different platforms, would enrich our experiment and findings. For example, it would be interesting to include additional Twitter and Facebook datasets, to further our study of the impact of topics and non-randomness on engagement dynamics. Adding more datasets could also expand our comparison to the literature, while in this paper the work was intentionally not constrained to like-by-like comparisons (e.g. comparing results across Twitter datasets alone) given that our goal was to study portability of results across multiple platforms.

One of the aims of our comparison to the literature is to highlight any inconsistencies. Nevertheless, it is worth acknowledging that results could vary for numerous reasons, such as due to idiosyncrasies of the used datasets or applied analysis. This emphasises the need for reproducing these types of experiment over multiple datasets and platforms.

7. CONCLUSIONS

Much research has been carried out in recent years to better understand the dynamics of user engagement in various social media platforms. This paper is one of the first to tackle

Table 4: Comparison of the derived engagement dynamics with significant feature findings from the related work. The table is read per-row such that a dot (.) indicates no comparison, ⇑ indicates that the feature is positively associated with engagement while ⇓ being a negative association for a given dataset. The colour coding indicates whether this effect agrees with the finding from the paper in the above column: green for agreement and red for disagreement.

		Naveed et al. [12] (**Twitter** / RT)	Gomez et al. [5] (**Slashdot** / Comments)	Rowe et al. [14] (**Twitter** / Replies)	Adamic et al. [1] (**Yahoo! Answers**/ Replies)	Cha et al. [4] (**Twitter** / RT)	Hsu et al. [7] (**Weblogs** / Comments)	Suh et al. [18] (**Twitter** / RT)	Hodas & Lerman. [11] (**Twitter** / RT)	Kwak et al. [8] (**Twitter** / RT)	Mishne et al. [10] (**Weblogs** / Comments)	Sousa et al. [17] (**Twitter** / RT)	Wu et al. [21] (**Twitter** / RT)
Boards.ie	In-Degree	.	⇑	⇑	.	⇑	.	⇑	.	⇑	.	⇑	⇑
	Out-Degree	.	.	⇓	.	.	.	⇓	⇓	.	.	.	.
	Post Count	.	.	.	.	.	.	⇓	.	.	.	.	.
	Age	.	.	.	.	.	. ⇑	⇑	.	.	.	.	.
	Post Rate	.	.	⇑	.	.	.	.	.	.	.	.	.
	Referral Count	⇓	.	.	.	.	.	⇓	.	.	⇓	.	.
Twitter Rand	In-Degree	.	⇑	⇑	.	⇑	.	⇑	.	⇑	.	⇑	⇑
	Out-Degree	.	.	⇓	.	.	.	⇓	⇓	.	.	.	.
	Post Count	.	.	.	.	.	.	⇑	.	.	.	.	.
	Age	.	.	.	.	.	. ⇓	⇓	.	.	.	.	.
	Post Rate	.	.	⇓	.	.	.	.	.	.	.	.	.
	Referral Count	⇓	.	.	.	.	.	⇓	.	.	⇓	.	.
Twitter Haiti	In-Degree	.	⇑	⇑	.	⇑	.	⇑	.	⇑	.	⇑	⇑
	Out-Degree	.	.	⇑	.	.	.	⇑	⇑	.	.	.	.
	Post Count	.	.	⇑	.	.	.	⇑	.	.	.	.	.
	Age	.	.	.	.	.	. ⇓	⇓	.	.	.	.	.
	Post Rate	.	.	⇓	.	.	. ⇓	.	.	.	.	.	.
	Referral Count	⇓	.	.	.	.	.	⇓	.	.	⇓	.	.
Twitter Union	In-Degree	.	⇓	⇓	.	⇓	.	⇓	.	.	.	⇓	⇓
	Out-Degree	.	.	⇑	.	.	.	⇑	⇑	.	.	.	.
	Post Count	.	.	⇑	.	.	.	.	.	.	.	.	.
	Age	.	.	.	.	.	. ⇓	⇓	.	.	.	.	.
	Post Rate	.	.	.	.	.	. ⇓	.	.	.	.	.	.
	Referral Count	⇓	.	.	.	.	.	⇓	.	.	⇓	.	.
	Polarity	⇑	.	.	.	.	.	.	.	.	.	.	.
Server Fault	In-Degree	.	⇑	⇑	.	⇑	.	⇑	.	⇑	.	⇑	⇑
	Out-Degree	.	.	⇓	.	.	.	⇓	⇓	.	.	.	.
	Post Count	.	.	⇓	.	.	.	⇓	.	.	.	.	.
	Age	.	.	.	.	.	. ⇓	⇓	.	.	.	.	.
	Post Rate	.	.	⇑	.	.	. ⇑	.	.	.	.	.	.
	Referral Count	⇓	.	.	.	.	.	⇓	.	.	⇓	.	.
	Polarity	⇑	.	.	.	.	.	.	.	.	.	.	.
SAP	In-Degree	.	⇑	⇑	.	⇑	.	⇑	.	⇑	.	⇑	⇑
	Out-Degree	.	.	⇓	.	.	.	⇓	⇓	.	.	.	.
	Post Count	.	.	⇓	.	.	.	⇓	.	.	.	.	.
	Age	.	.	.	.	.	. ⇑	⇑	.	.	.	.	.
	Post Rate	.	.	⇑	.	.	. ⇑	.	.	.	.	.	.
	Referral Count	⇓	.	.	.	.	.	⇓	.	.	⇓	.	.
Facebook	In-Degree	.	⇑	⇑	.	⇑	.	⇑	.	⇑	.	⇑	⇑
	Out-Degree	.	.	⇓	.	.	.	⇓	⇓	.	.	.	.
	Post Count	.	.	⇓	.	.	.	⇓	.	.	.	.	.
	Age	.	.	.	.	.	. ⇑	⇑	.	.	.	.	.
	Post Rate	.	.	.	.	.	. ⇑	.	.	.	.	.	.
	Referral Count	⇓	.	.	.	.	.	⇓	.	.	⇓	.	.
	Polarity	⇓	.	.	.	.	.	.	.	.	.	.	.

the vital questions of which of those identified patterns are consistent and applicable to multiple platforms. To answer this question, a common set of features and analysis framework are required that apply to several social media platforms.

To this end, we defined a collection of social and content features, chosen from related work, that are common across five social platforms; Twitter, Facebook, Boards.ie, SAP communities, and Server Fault. We then produced and applied a machine-learning based approach for engagement prediction (defined as a binary classification problem) that included standardisation, dataset balancing and time-ordering to enable comparison of engagement dynamics.

We contrasted the role of different features on engagement likelihood across our five social media platforms, thereby comparing engagement dynamics, and contrasting these findings with the engagement dynamics reported in existing studies on individual social media platforms. We went beyond the comparison of results from same platforms (e.g. Twitter vs Twitter) to comparing across multiple and different platforms (e.g. Twitter vs Boards.ie). Our intention was to identify any similarities and differences in the engagement dynamics and feature sets across a variety of platforms.

Our experiments and results demonstrated that different features could have an opposite effect on engagement in different platforms, or across different non-random datasets from the same platform. We hope that the presented evaluation framework will serve as a basis for future work within the social web community and enable further research into the cross-platform examination of engagement dynamics.

8. REFERENCES

[1] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge sharing and yahoo answers: Everyone knows something. In *Proc WWW Conf*, 2008.

[2] L. Backstrom, J. Kleinberg, L. Lee, and C. Danescu-Niculescu-Mizil. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In *Proceedings of the sixth ACM international conference on Web Search and Data Mining (WSDM)*, Rome, Italy, 2013.

[3] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha. Learning to Recognize Reliable Users and Content in Social Media with Coupled Mutual Reinforcement. In *18th Int WWW Conf*, April 2009.

[4] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *Proc. 4th Int. AAAI Conf. on Weblogs and Social Media (ICWSM)*, Washington, DC, 2010.

[5] V. Gómez, A. Kaltenbrunner, and V. López. Statistical analysis of the social network and discussion threads in slashdot. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 645–654, New York, NY, USA, 2008. ACM.

[6] R. Gunning. *The Technique of Clear Writing.* McGraw-Hill, 1952.

[7] C.-F. Hsu, E. Khabiri, and J. Caverlee. Ranking Comments on the Social Web. In *Int Conf Computational Science and Engineering (CSE) 2009*, 2009.

[8] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web (WWW)*, Raleigh, NC, USA, 2010.

[9] S. A. Macskassy and M. Michelson. Why do people retweet? anti-homophily wins the day! In *Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM)*, Menlo Park, CA, USA, 2011.

[10] G. Mishne and N. Glance. Leave a Reply: An Analysis of Weblog Comments. In *Third annual workshop on the Weblogging ecosystem*, 2006.

[11] O. H. Nathan and K. Lerman. How visibility and divided attention constrain social contagion. In *Proceedings of Social Computing Conference (SocialCom)*, Amsterdam, The Netherlands, 2012.

[12] N. Naveed, T. Gottron, J. Kunegis, and A. C. Alhadi. Bad news travel fast: A content-based analysis of interestingness on twitter. In *Proc. 3rd Int. Conf. on Web Science, 2011*, Bon, Germany, 2011.

[13] M. Rowe, S. Angeletou, and H. Alani. Anticipating discussion activity on community forums. In *Proc. 3rd IEEE Int. Conf. Social Computing (SocialCom2011)*, Boston, MA, USA, 2011.

[14] M. Rowe, S. Angeletou, and H. Alani. Predicting discussions on the social semantic web. In *Extended Semantic Web Conference*, Heraklion, Crete, 2011.

[15] J. Schantl, C. Wagner, R. Kaiser, and M. Strohmaier. The utility of social and topical factors in anticipating repliers in twitter conversations. In *Proceedings of ACM Web Science Conference*, Paris, France, 2013.

[16] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro. How useful are your comments?: analyzing and predicting youtube comments and comment ratings. In *19th Int WWW Conf*, NY, USA, 2010.

[17] D. Sousa, L. Sarmento, and E. M. Rodrigues. Characterization of the twitter @replies network: are user ties social or topical? In *Proceedings of the 2nd international workshop on Search and Mining User-generated Contents (SMUC)*, Toronto, Canada, 2010.

[18] B. Suh, L. Hong, P. Pirolli, and E. H. Chi. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *Proc IEEE Second Int Conf on Social Computing (SocialCom)*, 2010.

[19] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88, 2010.

[20] C. Wagner, M. Rowe, M. Strohmaier, and H. Alani. Ignorance isn't bliss: an empirical analysis of attention patterns in online communities. In *Proc. 4th IEEE International Conference on Social Computing (SocialCom)*, Amsterdam, The Netherlands, 2012.

[21] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. In *In Proceedings of the 20th International Conference on World Wide Web (WWW)*, Hyderabad, India, 2011.