A Survey on Identification and Analysis of Poor Quality Content on Facebook.

Prateek Dewan, Indraprastha Institute of Information Technology - Delhi (IIITD)

Online social media services like Facebook witness an exponential increase in user activity when an event takes place in the real world. This activity is a combination of good quality content like information, personal views, opinions, comments, as well as poor quality content like rumors, spam, and other malicious content. Although, the good quality content makes online social media a rich source of information, consumption of poor quality content can degrade user experience, and have inappropriate impact in the real world. In addition, the enormous popularity, promptness, and reach of online social media services across the world makes it essential to monitor this activity, and minimize the production and spread of poor quality content. Multiple studies in the past have analyzed the content spread on social networks during real world events. However, little work has explored the Facebook social network. Two of the main reasons for the lack of studies on Facebook are the strict privacy settings, and limited amount of data available from Facebook, as compared to Twitter. With over 1 billion monthly active users, Facebook is about five times bigger than its next biggest counterpart Twitter, and is currently, the largest online social network in the world. In this literature survey, we review the existing research work done on Facebook, and study the techniques used to identify and analyze poor quality content on Facebook, and other social networks. We also attempt to understand the limitations posed by Facebook in terms of availability of data for collection, and analysis, and try to understand if existing techniques can be used to identify and study poor quality content on Facebook.

Categories and Subject Descriptors: J.4 [Computer Applications] Social and Behavioral Sciences

General Terms: Human Factors, Measurement, Security

1. INTRODUCTION

Over the past decade, online social media (OSM) has stamped its authority as one of the largest information propagators on the Internet. OSM services have defied all regional, cultural, and language boundaries, and provided every Internet user on the planet with an equal opportunity to speak, and be heard. Nearly 25% of the world’s population uses at least one social media service today. People across the globe actively use social media platforms like Twitter and Facebook for spreading information, or learning about real-world events these days. A recent study revealed that social media activity increases up to 200 times during major events like elections, sports, or natural calamities [Szell et al. 2014]. This swollen activity contains a lot of information about the events, but is also prone to severe abuse like spam, misinformation, and rumor propagation, and has thus drawn great attention from the computer science research community. Since this stream of information is generated and consumed in real time, and by common users, it is hard to extract useful and actionable content, and filter out unwanted feed. Twitter, in particular, has been widely studied by researchers during real-world events [Becker et al. 2011; Hu et al. 2012; Kwak et al. 2010; Sakaki et al. 2010; Weng and Lee 2011]. However, few studies have looked at the content spread on social media platforms other than Twitter to study real-world events [Chen and Roy 2009; Hille and Bakker 2013; Osborne et al. 2012]. Surprisingly, there has been little work on studying content on Facebook during real world events [Westling 2007], which is five times bigger than Twitter in terms of the number of monthly active users. 2

In this survey, we look at the existing work done in the space of identification and analysis of malicious content on Facebook, and event analysis on online social media in general. The aim of is report is to look at a


2http://www.diffen.com/difference/Facebook_vs_Twitter
range of research attempts which would help to explore malicious content spread on Facebook during events. In particular, we look at three distinct areas, viz. a) the Facebook social graph, b) attack and detection techniques with respect to malicious content on Facebook, and c) analysis of events using online social media data. Then, we look at the various limitations that Facebook poses, which makes event analysis, and detection of malicious content on this network a hard problem. Towards the end, we discuss the implications and research gaps in identifying and analyzing malicious user generated content on Facebook during events.

1.1 Why Facebook?

Facebook is currently, the largest online social network in the world, having more than 1.28 billion monthly active users. Over the past decade, its popularity has seen such a monumental surge, that “checking” their Facebook accounts has become an addiction for Internet users. Figure 1 depicts a road sign in New York prompting users to avoid using Facebook on their mobile devices while walking on the street. Researchers have even proposed a “Facebook addiction scale” to measure the level of users’ obsession with Facebook [Andreassen et al. 2012]. Apart from keeping in touch with friends and family, a big proportion of users also resort to Facebook for getting their daily dose of news and updates about what is going on around the world. A recent survey of 5,173 adults suggested that 30% of people get their news from Facebook, while only 8% receive news from Twitter and 4% from Google Plus [Holcomb et al. 2013]. These numbers suggest that researchers need to look beyond Twitter, and study other social networks to get a better understanding of the flow of information and content on online social media during major real world events. Although there has been some work related to events on Wikipedia [Osborne et al. 2012] and Flickr [Chen and Roy 2009], there hardly exists any work on studying Facebook content during real world events.

The sheer volume of the publicly available Facebook content (approx. 1.33 billion posts per day [Facebook et al. 2013]) makes it a potentially rich source of information. In addition, recent introduction of features like hashtag support [Lindley 2013] and Graph search for posts [Room 2013], have largely increased the level of visibility of public content on Facebook, either directly or indirectly. Users can now search for topics and hashtags to look for content, in a fashion highly similar to Twitter; thus making the public Facebook content more visible and consumable by its users. This increasing public visibility, and an enormous user-base, potentially makes Facebook one of the largest and most widespread sources of information on the Internet, especially during real world events, when social media activity swells significantly.

1.2 Facebook activity during events

Social media activity rises manifold during events like sports, natural calamities etc. [Szell et al. 2014]. Unfortunately, spammers make use of this behavior to spread malicious content on social media platforms including Facebook, especially, whenever an event takes place. The FIFA World Cup in 2014, for example, saw a record breaking 350 million users generating over 3 billion posts on Facebook over a period of 32 days. Such colossal magnitude of activity makes online social media platforms an even more attractive venue for spammers during events. Recently, spammers exploited the famous biting incident during the 2014 FIFA World Cup, where an Uruguayan player was banned for biting an opponent. Spammers used the viral nature of this incident to spread links on Facebook, pointing to a phishing page prompting visitors to sign a petition in defence of an Uruguayan player. In another recent incident of Malaysian Airline MH17 flight crash, scammers placed dozens of so-called ‘community pages’ on Facebook, dedicated to victims of the tragedy. On the page, Facebook users were tricked into clicking links showing extra or unseen footage of the

---

3http://newsroom.fb.com/company-info/, as recorded on July 1, 2014.

Comprehensive Examination Survey Report, Vol. 1, No. 1, Article 1, Publication date: October 2014.
Identifying and Analyzing Poor Quality User-generated Content on Facebook

Fig. 1. Street artist in New York, USA puts up road signs telling mobile users to pay attention while walking.

crash. Instead of seeing a video, they were led to various pop-up ads for porn sites or online casinos.  

This kind of activity not only violates Facebook’s terms of service, but also degrades user experience. According to a research, Facebook spammers make $200 million just by posting links. 

Facebook itself has confirmed spam as a serious issue, and taken steps to reduce spam content in users’ newsfeed recently [Owens and Turitzin 2014]. Identifying spam on Facebook, however, evidently remains a hard problem. Despite of Facebook having a high performance immune system of their own [Stein et al. 2011], users still encounter an enormous number of spam and malicious content on regular basis. Existing approaches to detect spam in other online social media services like Twitter [Benevenuto et al. 2010; Grier et al. 2010; McCord and Chuah 2011; Wang 2010], cannot be directly ported to Facebook due to multiple issues. These include the public unavailability of critical pieces of information like profile, and network information, age of the account, no limit on post length, etc. There exists dire need to study spam content on Facebook, and develop techniques to identify it efficiently, and automatically.

7http://www.nltimes.nl/2014/07/22/flight-17-spam-scams-facebook-twitter/ 
1.3 Scope of this report

The focus of this survey is to cover the existing work lying in the intersection space of three concepts, viz. online social media, malicious user generated content, and events (Figure 2). In particular, we concentrate our focus around the work done on Facebook, but we do not, however, restrict ourselves to Facebook related literature only. We first explore the Facebook network structure. Since there does not exist a lot of work in the space of identification and analysis of malicious user generated content on Facebook during events, we look at malicious content on Facebook, and event analysis on online social media, separately.

The rest of the report is structured as follows. Section 2 discusses the work done on characterization of the Facebook social graph. Section 3 focuses on the attack and detection techniques with respect to malicious content on Facebook. Event analysis on Twitter, and other online social media services is discussed in Section 4. Section 5 highlights the various limitations and challenges in collection and analysis of data from Facebook. Lastly, we discuss the research gaps in the existing literature in Section 6.

2. EXPLORING THE FACEBOOK NETWORK

Facebook is a bidirectional network, where two users cannot connect without mutual consent. Each “object” on this network, like a user, picture, post, video, album etc. can be considered as a node, and each “action”, for example, a like, comment, share, friend connection etc. is an edge, which connects two nodes. Figure 3 depicts a pictorial representation of this graph.

2.1 Characterization of the Facebook network

Facebook was made publicly available in September 2006. Within a couple of years, the rapid growth rate of Facebook’s user base made it a center of attraction for research around the world. Researchers started off by collecting and trying to understand subsets of the Facebook network. Lewis et al. [Lewis et al. 2008]

Fig. 2. Focus of this survey report. The emphasis is on, but not limited to highlighting the work done on the Facebook social network.
provided a first of its kind dataset of Facebook users, and made it publicly available. Authors downloaded and characterized the profile and network data of 1,640 freshmen students enrolled at a diverse private college in the Northeast U.S. in 2009 by requesting permission from Facebook, and the university in question. Their findings revealed some interesting characteristics about these students, for example, the average pair of students whether or not they shared ties or demographics were found to have a higher percentage of favorite books/authors in common (2.%) than movies (1.5%) or music (1.5%). Similarly, they observed the highest similarity among friends who both appeared in each other’s photo albums. Closely related to this work was Amanda et al.’s work [Traud et al. 2011], where authors studied the structure of social networks of students by examining the graphs of Facebook “friendships” at five American universities at a single point in time. The primary aim of this paper was to use an unsupervised algorithm to compute the community structure consisting of clusters of nodes of these universities and to determine how well the demographic labels included in the data correspond to algorithmically computed clusters. Following up, authors extended their study to one hundred American colleges and universities, and examined homophily and community structure for each of the networks, and compared the community structure to partitions based on the given categorical data [Traud et al. 2012]. Both these studies also obtained their datasets from Facebook directly.

All the aforementioned studies showed promising results, derived by use sound methodologies, but suffered from intrinsic sampling bias. The sample datasets obtained for all these studies were fairly small, and represented a confined set of Facebook users, which were American college students. Although these datasets were obtained directly or indirectly from Facebook itself, and were complete in most aspects, the results obtained by these studies cannot be extended and generalized to the common Facebook audience. To overcome this sampling bias, researchers either needed to analyze the entire Facebook network at once, or to employ better techniques for sampling, to obtain more representable subsets of the Facebook graph. We discuss the former in the next paragraph, and latter in Section 2.2.

By the year 2011, over 500 millions users around the world had become a part of the Facebook network.\footnote{http://www.digitalbuzzblog.com/facebook-statistics-stats-facts-2011/} This was the time when Ugander et al. [Ugander et al. 2011] first studied the complete Facebook graph. Authors of this work brought out some major insights about Facebook’s network by conducting a large scale analysis on the entire Facebook network. They confirmed the ‘six degrees of separation’ phenomenon on a global scale, and found that the social network is nearly fully connected, with 99.91% of individuals
belonging to a single large connected component. In addition, authors observed a strong effect of age on
friendship preferences as well as a globally modular community structure driven by nationality, but did not
find any strong gender homophily. In a follow up study, Backstrom et al. [Backstrom et al. 2012] found the
average distance between two Facebook users to be 4.74, corresponding to 3.74 intermediaries or “degrees of
separation” instead of six. Prior to this work, most of the research involving Facebook studied it from either
from a privacy standpoint [Acquisti and Gross 2006; Boyd 2008; Hargittai et al. 2010; Liu et al. 2011], or
a social science perspective, looking at why users use Facebook [Joinson 2008; Ellison et al. 2007; Pempek
et al. 2009; Ross et al. 2009], and what are the mental, social, and emotional affects of Facebook usage on
individuals [Gonzales and Hancock 2011; Kim and Lee 2011].

2.2 Facebook crawling

As discussed in Section 2.1, results obtained from studies conducted on the Facebook network subsets were
non generalizable, and needed better sampling techniques. Researchers in some more studies crawled the
Facebook network to study user interactions, instead of fetching selective data from Facebook [Viswanath
of the user population into networks to perform a complete crawl of its subsets in an iterative fashion. Their
primary data set comprised of profile, Wall and photo data crawled from the 22 largest regional networks
on Facebook between March and May of 2008. In all, this dataset was composed of full profiles of over 10
million Facebook users. A similar technique was used by Vishwanath et al. [Viswanath et al. 2009] to crawl
a partial subset of the New Orleans network. Authors were able to gather information for about 90,269 users
and 3,646,662 friendship links between those users. This accounted for 52% of the users in the New Orleans
network based on the statistics provided by Facebook at that time.

With large scale crawls of the Facebook network, and bigger datasets, researchers were able to reduce
sampling biases to some extent. However, there was still scope for better sampling techniques to enhance
representativeness and generalizability. To this end, Gjoka et al. [Gjoka et al. 2010] implemented several
crawling techniques to obtain a representative and unbiased sample of the Facebook network. Authors found
the Metropolis-Hasting random walk and a re-weighted random walk to work well, whereas the traditional
Breadth-First-Search and Random Walk were found to perform quite poorly, producing substantially bi-
ased results. The collected samples were validated against a true uniform sample obtained using Facebook
user IDs, as well as via formal convergence diagnostics, and were shown to have good statistical properties.
The technique used by the authors for collecting a true uniform sample was then utilized by Catanese et
al. [Catanese et al. 2011], who performed a comparative analysis of two large crawls of Facebook; one using
the Breadth-First-Search technique, and the other using the true uniform sampling technique mentioned
previously. Authors of this work highlighted some distinct differences between the two datasets, including
dergee distribution, clustering coefficient, Eigenvector centrality etc.

Summary In this section, we looked at the existing work done on characterizing the Facebook graph,
and its subsets. We observed that quite a few pieces of early experiments conducted using Facebook’s data
suffer from sampling bias, and the datasets used are not representative of the entire Facebook population. To
counter this sampling bias, we then looked at literature which made use of large scale crawls of the Facebook
network, and found how some crawling techniques worked better than some others. Given that about 20%
of the world’s population uses Facebook, it is essential to obtain research outcomes which are generalizable,
and representative of a large proportion of the Facebook population.

It is important to note that while researchers have statistically proven subsets of the Facebook graphs to
be unbiased, none of the aforementioned work looked at the representativeness of the graph samples obtained
in terms of geographical, gender, or language distribution. We believe that these aspects are important to
ensure that samples picked for analysis from the Facebook network are representative of the entire Facebook population in addition to being mathematically and statistically unbiased.

3. MALICIOUS CONTENT ON FACEBOOK

The popularity and reach of Facebook has also attracted a lot of spam, phishing, malware, and other types of malicious activity. Attackers lure victims into clicking on malicious links pointing to external sources, and infiltrate their network. These links can be spread either through personal messages (chats), or through wall posts. To achieve maximum visibility, attackers prefer to post links publicly. Typically, an attacker initiates the attack by posting memes with attention grabbing previews, which prompt users to like, share, or comment on them in order to view them. The actions of liking, commenting or sharing spread these memes into the victim’s network. Once the meme is spread, the victim is redirected to a malicious website, which can further infect her computer, or friends network through phishing, malware, or spyware. Figure 4 shows an example of such a malicious post, which appears to be a video. Clicking on the link redirects the user to a phishing page as shown in Figure 5, which looks very similar to a genuine Facebook post. This phishing page asks the victim to share this video with their friends in order to view it. However, once the victim shares this video, the page redirects to a random advertisement page. The video corresponding to the preview / thumbnail shown in the post, does not actually exist.

Multiple other sources have cited such examples of scams and malicious posts on Facebook in the past few years. In addition to phishing scams, other malicious activity on Facebook includes unsolicited mass mentions, photo tagging, post tagging, private / chat messages etc. Intuitively, a user is more likely to respond to a message or post from a Facebook friend than from a stranger, thus making this social spam a more effective distribution mechanism than traditional email. This increased susceptibility to such kind of spam has prompted researchers to study, and combat social spam and other malicious activity on Facebook.

We now look at the various attack and detection techniques that have been used in the past to identify and spread malicious content on Facebook respectively.

3.1 Attack techniques

In order to identify and contain malicious posts on Facebook, or any OSM, it is essential to explore and understand the techniques that are, or can potentially be deployed by attackers to spread such content. To this end, Patsakis et al. [Patsakis et al. 2009] described how Facebook can be exploited and converted into an attack platform, in order to gain some sensitive data, which can complete a perfect attacking profile against a user. Authors created a Facebook application for demonstration purposes that on the surface was a simple application, but on the background it collected useful data. This app executed malicious code on the victim’s browser, and collected the IP address of the user-victim, the browser version, the OS platform and whether some specific ports are open or closed. This data was then transmitted to the authors over email. Authors also pointed out that their app was indexed on the main list of Facebook applications, despite the fact that the description of app clearly stated that it was generating malicious traffic, and had been created for penetration testing purposes. Huber et al. presented a friend-in-the-middle attack through hijacking session cookies. Authors explained how it was possible to impersonate the victim using this technique, and interact with the network without proper authorization. However, this technique was proposed in 2011, when using HTTPS to connect to the website was optional. Post 2013, all communication on Facebook uses encryption (HTTPS) by default, which means that such attacks are no more possible.

11https://online.wsj.com/news/articles/SB1000142405297020368620457711294273497780
12http://allfacebook.com/facebook-warning-amazon_b74943
13https://www.facebook.com/notes/facebook/a-continued-commitment-to-security/486790652130
14https://www.facebook.com/notes/facebook-engineering/secure-browsing-by-default/10151590414803920

Comprehensive Examination Survey Report, Vol. 1, No. 1, Article 1, Publication date: October 2014.
Fan et al. [Fan and Yeung 2010] proposed a virus model based on the application network of Facebook. Authors also modeled the virus propagation with an email virus model and compared the behaviors of virus spreading in Facebook and email network. Their findings revealed that while Facebook provides a platform for application developers, it also provides the same chance for virus spreading. In fact, the virus was found to spread faster on the Facebook network if users spend more time on it. The result of their simulation showed that, even though a malicious Facebook application attracts only a few users in the beginning, it can still spread rapidly. That is because users may trust their friends of Facebook and install the malicious application.

It is important to understand that in addition to the techniques described above, a large proportion of attacks on Facebook, and even other social networking platforms, make use of social engineering. This is evident since it is hard to initiate the spread of a malicious piece of content on a network without any human involvement. Attackers lure victims into using malicious apps, clicking malicious links, and sharing pieces of content, and in some cases, even pretend to offer various kinds of benefits in return. Since these attacks are well-crafted in most cases, it becomes hard for a legitimate user to be able to comprehend the results of her actions. We now look at the various techniques that have been proposed to detect malicious content on the Facebook social network.

3.2 Detection techniques

Facebook has its own immune system to safeguard its users from unwanted, malicious content [Stein et al. 2011]. Researchers at Facebook built and deployed a coherent, scalable, and extensible real time system to protect their users and the social graph. This system performs real time checks and classifications on every read and write action. As of March 2011, this was 25 billion checks per day, reaching 650K per second at
Fig. 5. A fake Facebook post which looks visually similar to a genuine Facebook video. The URL in the address bar depicts that the page is a fake.

peak. The system also generates signals for use as feedback in classifiers and other components. Facebook’s immune system is based on four design principles, viz. quick detection and response, covering a broad and evolving interface, sharing signals across channels, and classification in real time. Designers of this complex system used an exhaustive set of components and techniques to differentiate between legitimate actions and spam. These components were standard classifiers like Random Forest, Support Vector Machines, Logistic Regression etc., a feature extraction language, dynamic model loading, a policy engine, and feature loops. Figure 6 represents a high level design of the system.

Interestingly, despite this complex immune system deployed by Facebook, unwanted spam, phishing, and other malicious content continues to exist and thrive on Facebook. Although the immune system deployed by Facebook utilizes a variety of techniques to safeguard its users, authors did not present an evaluation of the system in terms of accuracy and efficiency in detecting anomalies. Their work discussed the system and its components in detail, but did not focus on the system evaluation from a research angle.

Gao et al. [Gao et al. 2010], in 2010, presented an initial study to quantify and characterize spam campaigns launched using accounts on Facebook. They studied a large anonymized dataset of 187 million asynchronous “wall” messages between Facebook users, and used a set of automated techniques to detect and characterize coordinated spam campaigns. Authors detected roughly 200,000 malicious wall posts with embedded URLs, originating from more than 57,000 user accounts. They also found that more than 70% of all malicious wall posts advertised phishing sites. Further, their findings revealed that more than 97% of the accounts...
they analyzed, were compromised accounts, rather than “fake” accounts created solely for the purpose of spamming. To the best of our knowledge, this is the only study which addresses the problem of detecting and analyzing malicious content on Facebook via automated means.

Following up their work, Gao et al. [Gao et al. 2012] presented an online spam filtering system that could be deployed as a component of the OSN platform to inspect messages generated by users in real-time. Their approach focused on reconstructing spam messages into campaigns for classification rather than examining each post individually. They were able to achieve a true positive rate of slightly over 80% using this technique, and achieved an average throughput of 1580 messages/sec with an average processing latency of 21.5ms on their Facebook dataset of 187 million wall posts. Although the technique of campaign identification has previously been used for offline spam detection, authors claimed to be able to achieve real time detection using this technique with sufficiently low overhead. Their model stayed accurate for over 9 months after initial training, overcoming the need for frequent re-training.

Stringhini et al. [Stringhini et al. 2010] utilized a honeypot model to collect information about spammers on Facebook. Authors first crawled a set of 2,000 random profiles each, across 16 different regional networks to collect a representative sample which could help them create a representative honey profile for the network. They monitored this profile over a duration of one year, and manually identified 173 spam profiles among a total of 3,831 friendship requests they received. The 173 spam profiles were then crawled to extract features like URL ratio (number of URLs posted per message), message similarity, number of friends, number of messages sent, friend choice (ratio of total number of friend names to total number of distinct friend names). These features were fed to a classifier along with features extracted from 1,000 legitimate profiles. Authors reported a low false positive rate of 2% and a low false negative rate of 1%, but did not report the accuracy measure. Using this model trained on 173 spam, and 1,000 legitimate profiles, authors tested 790,951 profiles from Los Angeles and New York networks, and identified 130 more spam profiles. Seven out of these 130 profiles detected, however, were marked as false positive upon manual inspection.

Ahmed et al. [Ahmed and Abulaish 2012] presented a Markov Clustering (MCL) based approach for the detection of spam profiles on Facebook. Authors crawled the public content posted by 320 handpicked Facebook users, out of which, 165 were manually identified as spammers, and 155 as legitimate. Authors then extracted 3 features from these profiles, viz. Active friends, Page Likes, and URLs to generate a weighted
Identifying and Analyzing Poor Quality User-generated Content on Facebook

graph, which served as input to the Markov Clustering model. This work, however, was also targeted at detecting spam campaigns instead of individual posts, similar to Gao et al’s work [Gao et al. 2010]. Although the suggested approach produced reasonable results, there was no evaluation of the model on a bigger dataset.

In an attempt to protect Facebook users from malicious posts, Faloutsos [Faloutsos 2013] designed an efficient social malware detection method which took advantage of the social context of posts. The author was able to achieve a maximum true positive accuracy rate of 97%, with the algorithm requiring, on average, 46 milliseconds to classify a post. This algorithm was then used to develop MyPageKeeper 15, a Facebook app to protect users from malware and malicious posts. Using data from this app, the author analyzed over 40 million posts during a period of 4 months, and found that 49% of the users were exposed to at least one social malware post during this period. This work also showed how social malware significantly differed from traditional email spam or web-based malware. According to Faloutsos, website blacklists were able to identify only 3% of the posts flagged by MyPageKeeper, while 26% of flagged posts pointed to malicious apps and pages hosted on Facebook, which no antivirus or blacklist was designed to detect. In addition to malicious or compromised user accounts, there also exist multiple third party applications which enable / aid the spread of malicious posts on Facebook. To determine if a Facebook application is malicious, Rahman et al. developed FRAppE (Facebook Rigorous Application Evaluator), which was one of the first attempts focused on detecting malicious apps on Facebook [Rahman et al. 2012]. FRAppE worked on a feature set that the author extracted from observing the posting behavior of approximately 111K Facebook apps across 2.2 million Facebook users. The authors found that 13% of the apps in their dataset were malicious, and were able to achieve an accuracy of 99.5% for detecting malicious apps using FRAppE. This was one of the first attempts at identifying malicious apps on Facebook via automated means.

Jin et al. [Jin et al. 2011] in 2011, proposed a scalable online social media spam detection system by utilizing a combination of image content features, text features, and social network features. Their feature sets were independent of the social media platform, but authors chose to test the system on Facebook because of its popularity. Although the feature sets proposed in this work looked more exhaustive than the techniques proposed previously, authors did not present their exact feature set or any evaluation results of their system.

Summary In this section, we looked at multiple research and system level contributions which identify the techniques for spreading malicious content, and address the problem of detection of malicious content on Facebook. However, most of the aforementioned work does not comprehensively address the issues in detail due to multiple reasons. One of the major reasons, as also addressed by Stringhini et al. [Stringhini et al. 2010], is the lack of availability of a substantial amount of data to analyze from Facebook. Large scale studies including studies conducted by Gao et al. [Gao et al. 2010; Gao et al. 2012], and Stringhini et al. [Stringhini et al. 2010] utilized the open nature, and geographically divided crawl-able networks of Facebook prior to October 2009. However, post 2009, the introduction of stringent privacy controls and unification of all geographical networks into one big graph have made it a challenging task to collect a sizable amount of data for analysis on Facebook. We discuss the other challenges in more detail in Section 5.

Most system level contributions towards detection lack vital details including detailed feature set description, source of true positive labeled datasets, details of the algorithms used, comparison of existing spam classification techniques with proposed techniques, or lack of evaluation etc., questioning the reproducibility of their results. In terms of research level contributions, there has been nominal work done in this space, with little, or no follow up studies to look at the evolution of malicious content over time. There does exist some work proposing a variety of attack techniques to propagate malicious content in the network, but the characterization and detection of this content is still an open issue.

15https://apps.facebook.com/mypagekeeper/
There exist multiple other pieces of closely related work, which look at detecting fake profiles / sybil nodes on Facebook, and malicious activity on Twitter. We do not cover those pieces of work, since the focus of this survey is confined to malicious user generated content on Facebook, in particular.

4. EVENT ANALYSIS ON ONLINE SOCIAL MEDIA

Online social media services like Facebook, Twitter, Google+, YouTube, Flickr, Instagram, Pinterest etc. have provided people with a free and open platform to communicate with each other. Today, anything that happens in the real world, is talked about on online social media. From sports to storms, terrorist attacks, bomb blasts, earthquakes, and even elections, users share thoughts and information about literally everything using online social media services. A recent study revealed that social media activity increases up to 200 times during major events like elections, sports, or natural calamities [Szell et al. 2014]. This swollen activity has drawn great attention from the computer science research community. Content and activity on Twitter, in particular, has been widely studied by researchers during events [Becker et al. 2011; Hu et al. 2012; Kwak et al. 2010; Sakaki et al. 2010; Weng and Lee 2011]. However, few studies have looked at social media platforms other than Twitter to study events [Chen and Roy 2009; Hille and Bakker 2013; Osborne et al. 2012]. In this section, we look at the various attempts at analyzing events on online social media. Since there has been substantial amount of work done in this space on Twitter alone, we look at event analysis on Twitter, and event analysis on other social media services, separately.

At the broadest level, analyzing an event on online social media can be broken down into two components. The first component deals with identifying that an event has occurred, and the second component deals with collecting information specific to the event from a social media service. In an ideal scenario, the output of the first component serves as the input to the second component. Detecting that an event has occurred using a stream of user generated content from online social media is, however, a grand challenge in itself. Thus, in practice, most of the work related to event analysis on online social media deals with the two issues separately. For our survey, we do not look at the work which focuses on detection of events. We only look at work done on event-specific online social media data, where the event is already known to have occurred.

4.1 Event analysis on Twitter

Twitter has been used widely during emergency situations, such as wildfires [De Longueville et al. 2009], hurricanes [Hughes and Palen 2009], floods [Vieweg et al. 2010] and earthquakes [Earle et al. 2010; Kireyev et al. 2009; Sakaki et al. 2010]. Journalists have hailed the immediacy of the service which allowed “to report breaking news quickly - in many cases, more rapidly than most mainstream media outlets” [Poulsen 2007]. Sakaki et al. [Sakaki et al. 2010] explored the potential of the real-time nature of Twitter and proposed an algorithm to detect occurrence of earthquakes by simply monitoring a stream of tweets in real-time. Here, the authors took advantage of the fact that users tweet about events like earthquakes as soon as they take place in the real world, and were able to detect 96% of all the earthquakes larger than a certain intensity. Their reporting mechanism was able to convey this information to common users through emails, 6 minutes before an official announcement was made by the Japanese Meteorological Agency. In another research work, Sakaki et al. [Sakaki et al. 2011] analyzed tweet trends to extract events that happened during a crisis event from Twitter. They analyzed log of user activity from Japanese tweets on all earthquakes during 2010-2011. Cheong et al. [Cheong and Cheong 2011] performed social network analysis on Twitter data during Australian floods of 2011 to identify active players and their effectiveness in disseminating critical information. More similar work includes an attempt by researchers to identify information from Twitter, that may contribute to enhancing situational awareness during natural hazard real-world events. Authors of this work focussed on communications broadcast on Twitter by people who were “on the ground” during the Oklahoma Grassfires, and Red River Floods in the USA in 2009. They proposed an enhanced set of generic features, which could be used for building systems to improve situational awareness automatically during emergency events [Vieweg
et al. 2010]. Hughes et al. [Hughes and Palen 2009] also studied four high profile, mass convergence events on Twitter and discovered that Twitter messages sent during such events reveal features of information dissemination that support information broadcasting and brokerage.

Varol et al. [Varol et al. 2014] analyzed the Gezi Park movement in Turkey through the lens of Twitter. Authors analyzed 2.3 million tweets about the event produced over a period of 25 days, during May - June, 2013. Authors of this work identified four types of users, viz. common users, rebroadcasters, influentials and hidden influentials, who tweeted during the event. Their analysis revealed that the conversation becomes more democratic as events unfold, with a redistribution of influence over time in the user population. Gupta et al. [Gupta et al. 2013a] studied the public Twitter stream during the Boston marathon bombings in 2013. Their results revealed showed that 29% of the most viral content, during the crisis were rumors and fake content; while 51% was generic opinions and comments; and rest was true information. In addition, authors used regression prediction model, to verify that, overall impact of all users who propagate the fake content at a given time, can be used to estimate the growth of that content in future. Many malicious accounts were also created on Twitter during the Boston event, that were later suspended by Twitter. Authors were able to identify over six thousand such user profiles, and observed that the creation of such profiles surged considerably right after the blasts occurred. In another research work, Gupta et al. [Gupta et al. 2013b] analyzed the public Twitter stream during the hurricane Sandy, which hit the USA in 2012. Here, authors performed a characterization analysis, to understand the temporal, social reputation and influence patterns for the spread of fake images. Their results showed that top thirty users (0.3%) out of 10,215 users resulted in 90% of the retweets of fake images; and network links such as follower relationships of Twitter, contributed very less (only 11%) to the spread of these fake photos URLs. Classification models were used to distinguish fake images from real images spreading during Hurricane Sandy, and authors found the Decision Tree classifier to perform the best, with 97% accuracy in distinguishing fake images from real. Mendoza et al. [Mendoza et al. 2010] used the data from 2010 earthquake in Chile to explore the behavior of Twitter users for emergency response activity. The results showed that propagation of rumor tweets versus true news were different and automated classification techniques can be used to identify rumors. Longueville et al. [De Longueville et al. 2009] analyzed Twitter feeds during forest Marseille fire event in France; their results showed that in location based social networks, spatial temporal data can be analyzed to provide useful localized information about the event.

The aforementioned work clearly highlights the importance of Twitter as an information sharing medium during events. Tweets have been shown to travel faster than the tremors of an earthquake, making it a potential warning system capable of saving thousands of human lives in the event of a major earthquake [Sakaki et al. 2010]. During the Boston Marathon blasts, Twitter came to rescue and helped the Boston Police to track down the two main suspects within hours of the blasts. 16 The popularity, reach, and public nature of Twitter have made it the first choice for almost everyone, including law and order agencies, journalists, and evidently, computer science researchers. However, this intense focus on Twitter has left a gap for similar analysis on other social media services, specially Facebook. Content posted on Facebook and other famous social media services during events has fairly been overlooked. We now look at some of the research work which analyzes events on social media services other than Twitter.

4.2 Event analysis on other online social media

As discussed in Section 4.1, there exists little work which focuses on online social media services other than Twitter to analyze events. Osborne et al. [Osborne and Dredze 2014] examined how Facebook, Google Plus, and Twitter report in breaking news. Authors identified 28 major events which took place in December 2013, and scanned all the three OSM services, viz. Facebook, Twitter, and Google Plus for posts related to these

16https://blog.twitter.com/2013/the-boston-bombing-how-journalists-used-twitter-to-tell-the-story

Comprehensive Examination Survey Report, Vol. 1, No. 1, Article 1, Publication date: October 2014.
events. Their findings revealed that all media carried the same major events, but Twitter continued to be the preferred medium for breaking news, almost consistently leading Facebook or Google Plus. Facebook and Google Plus largely reposted newswire stories and their main research value was that they conveniently packaged multiple sources of information together. This was one of the first attempts towards studying public streams of Facebook, and Google Plus, and comparing them with Twitter. Szell et al. investigated messages from five sources, viz. Facebook, Twitter, app.net, Enron email corpus, and a popular online forum, created in response to major, collectively followed events such as sport tournaments, presidential elections, or a large snow storm [Szell et al. 2014]. They related content length and message rate, and found a systematic correlation during events which can be described by a power law relation - the higher the excitation, the shorter the messages. Authors showed that on the one hand this effect could be observed in the behavior of most regular users, and on the other hand was accentuated by the engagement of additional user demographics who only post during phases of high collective activity. Palen et al. [Palen and Vieweg 2008] studied two real world events, to understand and characterize the wide scale interaction on social networking websites with respect to the events. Authors in this work, presented a qualitative analysis of a group on a popular social networking site as a virtual destination in the aftermath of the Northern Illinois University (NIU) shootings of February 14, 2008 in relation to related activity that happened in response to the Virginia Tech (VT) tragedy 10 months earlier.

Summary In this section, we looked at the various research on analyzing events on Twitter, and other online social media services. Past research clearly depicts the dominance of Twitter when it comes to analyzing events. However, we also saw evidence of similarity in the information streams between Twitter and other social media services like Facebook and Google Plus during events [Osborne and Dredze 2014]. The dominance of work on Twitter has left a wide gap, and prompts researchers to look at public streams on other social media events during events.

There have been several other closely related pieces of work, which target the problem of event detection on online social media. However, we do not cover this area, since event detection is a big research area in itself, and is beyond the scope of this survey report.

5. FACEBOOK LIMITATIONS AND CHALLENGES

In the previous section, we saw multiple research attempts towards analyzing events on the Twitter social network. It is important to note that there hardly exists any work which focuses on Facebook content for analyzing events. Given that Facebook is much older and bigger than Twitter, the reasons for the lack of work on Facebook need good justification and discussion. In this section, we look at the various limitations and challenges posed by Facebook, which possibly makes it a difficult task to extract, and analyze data from this network. Intuitively, the private nature of Facebook can largely be attributed to the lack of research on Facebook content on a large scale. We now look at these, and other challenges with Facebook research in detail.

5.1 Fine grained privacy settings

Facebook provides its users with an exhaustive set of privacy settings, which enable them to control who can see what information from their profile and posts. Unlike Twitter, majority of content on Facebook, including profile information, content, and network information, is not accessible publicly. Privacy settings at Facebook broadly offer visibility of information at four levels, as follows:

—Only me: Only the authorized user can see this information.
—Friends: Users who are connected to the authorized user via a “friendship” relation can see this information.
Friends of friends: Visibility at this level is increased to one more hop in the network, i.e. users who are friends with the friends of the authorized user, can also see this information.

Public: Anyone on the Internet can see this information.

In addition to these four basic visibility levels, Facebook also provides a custom level of visibility, where a user can choose the audience of her information and content selectively, at an individual level. These visibility levels can be applied to almost all sections of a user’s Facebook account, like profile information, pictures, albums, videos, wall posts, etc. Figure 7 presents a snapshot of the privacy settings page on Facebook. Users can also select who can send them friendship requests, and who can search them.

![Privacy Settings and Tools](image)

**Fig. 7.** A snapshot of the privacy settings page on Facebook. Users can control the visibility level of all their information and content independently.

Although these privacy settings offer users tremendous control over their content, they pose serious implications from a research standpoint. No more than 28% of Facebook users share their content with an audience wider than their friends [magazine 2012]. This leaves researchers with no more than a quarter of the total content on Facebook (probably even lesser), to collect and analyze. Although, with a user base of over a billion users, this proportion of content may convert to millions of posts a day, it is hard to see this content as a good representation of the entire Facebook population. Especially, when it comes to identifying and analyzing malicious content, it is hard to find an effective and scalable solution using only a fraction of content. Attackers can effectively exploit the private nature of the network to target multiple closed networks simultaneously, with no way for outsiders to identify these threats and provide any kind of warnings or protection. Majority of attack vectors may thus, never surface, and go untraceable.

Privacy rules applied on profile and network information of users, make it even harder to analyze the sources of malicious content in Facebook. Apart from gender, name, and username, all other profile information about a user is not available publicly, unless explicitly specified by the user. Vital pieces of information like a user’s work, education, description, location, account creation time, birth date, etc. are virtually
impossible to extract from Facebook. This implies that even if a user is identified as malicious, it is hard to analyze and extract features, which can be used to differentiate a benign user from a malicious one. Lack of network information poses similar implications in analysis of infected subgraphs of the Facebook social network. Networks in Facebook are bidirectional, i.e. two users cannot connect to each other without mutual consent from both. This connection between two users is known as a “friendship” connection in Facebook terminology. Friends of a user are also non-public by default, which eliminates the possibilities of analyzing cascades, communities, infected subgraphs, and the flow of malicious pieces of content in the network. Almost all pieces of work on event analysis on Twitter (Section 4.1), heavily use profile and network features of its users for in-depth characterization and analysis. The absence of these features in Facebook make event analysis on this network a really hard problem for researchers.

Facebook also poses technical challenges and limitations on collection and analysis of the small proportion of data which is publicly available. We now discuss these in Section 5.2.

5.2 Technical limitations

Almost all OSM services, including Facebook, provide a range of Application Programming Interface (API) endpoints for users to interact with the service programmatically, and exchange data. Facebook provides a Search API which can be used to look for public posts containing a particular set of keywords. This API endpoint is restricted to only public content, since getting access to non-public content requires authorization from the owner of the content. In order to collect public data specific to an event, this Search API needs to be supplied with relevant keywords. For example, keywords like fifa, worldcup, fifaworldcup can be used to collect public posts related to the FIFA World Cup. This is very similar to Twitter’s Search API, which is commonly used to search for public tweets. However, a big drawback with Facebook is the absence of an API endpoint similar to Twitter’s Streaming API, which can be used to collect event specific data in real time. In addition, Facebook does not mention if the results returned by its Search API cover the entire public content generated on the network, or only a portion of it. There is no way to verify the total number of posts generated on a topic, and the number of posts returned by the API. This can result in loss of an unknown amount of data. To overcome this issue to some extent, Facebook’s Search API needs to be queried at regular, short intervals of time. If the rate generation of public content during an event is too large, these intervals need to be tuned accordingly, and shortened further.

As discussed in Section 4, event analysis on online social media has two broad components, viz. event detection, and data collection. Since event detection is a hard problem in itself, identifying events, and feeding them to a data collection framework requires human intervention. In order to automate this task, Twitter provides an API endpoint, which returns the trending topics that are being talked about, in a given locality, and at a given point in time. Past research has shown an overlap of more than 85% between Twitter’s trending topics, and breaking / persistent news on Newswire [Kwak et al. 2010]. Although these trending topics may not be an exact equivalent to events, they can be used to serve as input to the data collection process, thus completely eliminating the human in the loop. Facebook also launched a Trending feature for some of its users recently, but its Trends API isn’t accessible for general public. Again, this imposes a major implication, as the process of collecting Facebook data related to an event cannot be automated, and always requires human intervention, decreasing efficiency.

17https://developers.facebook.com/docs/graph-api/using-graph-api/v2.0/#search
18https://dev.twitter.com/docs/api/1.1/get/search/tweets
19https://dev.twitter.com/docs/api/1.1/post/statuses/filter
20https://dev.twitter.com/docs/api/1.1/get/trends/place
21http://techcrunch.com/2014/01/16/facebook-trending/
22https://developers.facebook.com/docs/trends/v2.0

Comprehensive Examination Survey Report, Vol. 1, No. 1, Article 1, Publication date: October 2014.
Summary There are multiple factors which make working with Facebook data a hard task. In this section, we discussed these factors, and saw some of the limitations and challenges which exist in collection and analysis of public content on Facebook. In addition to the highly private nature of Facebook, we also saw some technical limitations, which hinder data collection on Facebook. However, these limitations and challenges do not completely eliminate the scope of working with Facebook data.

6. DISCUSSION AND RESEARCH GAPS

This survey was aimed at highlighting the existing gaps that exist in studying the Facebook network as one of the biggest channels for information dissemination during events. As evident from the existing literature, there is much need and scope of exploring the potential of the currently unmonitored public stream of Facebook content, and segregating malicious content from useful information during events. Facebook's content, specially during events, can prove to be a vital information channel, and hence needs utmost attention.

Building efficient, scalable solutions for big data services like Facebook, requires a representative sample of data for experimentation, and for drawing valid conclusions. However, as we saw in Section 2, getting a representative sample of the Facebook sub graph is a hard problem in itself. One of the major reasons for researchers being unable to get a convincing data sample is that Facebook’s fine-grained privacy settings make majority of its content private, and publicly inaccessible. About 72% Facebook users set their posts to private [magazine 2012]. This private nature of Facebook has been a major challenge in collecting and analyzing its content in the computer science research community.

In addition, existing techniques related to spread and mitigation of malicious content on Facebook haven’t been studied comprehensively. Most of the techniques proposed for detecting malicious posts on Facebook lack comprehensive evaluation, which is essential to prove their worth and research contribution. There hardly exists any research in the computer science community which characterizes or analyzes malicious content on Facebook on a large scale. The only large scale study on Facebook [Gao et al. 2010] was on a dataset of 187 million wall messages which were collected from a random sample of 3.5 million users by crawling their Facebook walls in 2009. It would be interesting to study how malicious content identified from a random sample of Facebook differs from malicious content on Facebook during events. It is possible that the characteristics of malicious Facebook content vary across different events and differ from malicious Facebook content in general. It would also be interesting to study if malicious content has evolved over time on Facebook.

We also saw some research attempts towards studying events from Facebook data (Section 4). However, Twitter has largely been the focus of researchers for studying events. We saw how Twitter was found to be a vital actor during sporting events, political campaigns, forest fires and even earthquakes. Content on other social networks has, however, not been given much attention in this respect. It is reasonable to assume that other social networks including Facebook also carry event related content, which can be of importance to the population of Internet users where Twitter is not as widely used as some other social networks. Even though researchers found high overlap between Twitter and Facebook streams during events [Osborne and Dredze 2014], we are yet to see dedicated attempts at studying Facebook content during events.

7. CONCLUSION

In this survey, we explored various research attempts towards exploring the Facebook network, analyzing malicious content on it, and analyzing events on online social media in general. The aim of this survey was to look at relevant literature, which could aid in studying and combating malicious user generated content spread on Facebook during events. In order to keep this survey focused, we did not cover a variety of possibly relevant research areas including detection of compromised / fake accounts, and sybil nodes in the Facebook network, detection of spam on other social networks like Twitter, credibility / trustworthiness of information.
of user generated content, and event detection in online social media. We also looked at the various challenges and limitations posed by Facebook (as discussed in Section 5). Apart from technical limitations, there exist various research gaps in existing literature, which are yet to be addressed and explored.

REFERENCES


Hila Becker, Mor Naaman, and Luis Gravano. 2011. Beyond Trending Topics: Real-World Event Identification on Twitter.. In ICWSM.


Bertrand De Longueville, Robin S Smith, and Gianluca Luraschi. 2009. OMG, from here, I can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In LBSN. ACM, 73–80.


Hongyu Gao, Yan Chen, Kathy Lee, Diana Palsetia, and Alok N Choudhary. 2012. Towards Online Spam Filtering in Social Networks,. In NDSS.


Aditi Gupta, Hemank Lamba, and Ponnurangam Kumaraguru. 2013a. $1.00 per RT #BostonMarathon #PrayForBoston: Analyzing Fake Content on Twitter. In eCRS. IEEE, 12.

Comprehensive Examination Survey Report, Vol. 1, No. 1, Article 1, Publication date: October 2014.
Identifying and Analyzing Poor Quality User-generated Content on Facebook


Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media?. In WWW. ACM, 591–600.


Comprehensive Examination Survey Report, Vol. 1, No. 1, Article 1, Publication date: October 2014.


Comprehensive Examination Survey Report, Vol. 1, No. 1, Article 1, Publication date: October 2014.