Detecting Linkedin Spammers and its Spam Nets

Víctor M. Prieto*, Manuel Álvarez* and Fidel Cacheda*
*Department of Information and Communication Technologies, University of A Coruña, Coruña, Spain 15071
Email: {victor.prieto, manuel.alvarez, fidel.cacheda}@udc.es

***Abstract*—Spam is one of the main problems of the WWW. Many studies exist about characterising and detecting several types of Spam (mainly Web Spam, Email Spam, Forum/Blob Spam and Social Networking Spam). Nevertheless, to the best of our knowledge, there are no studies about the detection of Spam in Linkedin. In this article, we propose a method for detecting Spammers and Spam nets in the Linkedin social network. As there are no public or private Linkedin datasets in the state of the art, we have manually built a dataset of real Linkedin users, classifying them as Spammers or legitimate users.**

The proposed method for detecting Linkedin Spammers consists of a set of new heuristics and their combinations using a kNN classifier. Moreover, we proposed a method for detecting Spam nets (fake companies) in Linkedin, based on the idea that the profiles of these companies share content similarities. We have found that the proposed methods were very effective. We achieved an F-Measure of 0.971 and an AUC close to 1 in the detection of Spammer profiles, and in the detection of Spam nets, we have obtained an F-Measure of 1.

I. Introduction

Currently, the WWW is the biggest information repository ever built, and it is continuously growing. According to the study presented by Gulli and Signorini [1] in 2005, the Web consists of thousands of millions of pages. In 2008, according to Official Blog of Google[1], the Web contained 1 trillion unique URLs.

Due to the huge size of the Web, search engines are essential tools in order to allow users to access relevant information for their needs. Search engines are complex systems that allow collecting, storing, managing, locating and accessing web resources ranked according to user preferences. A study by Jansen and Spink [2] established that approximately 80% of search engine users do not take into consideration those entries that are placed beyond the third result page.

This fact, together with the great amount of money that the traffic of a web site can generate, has led to the appearance of persons and organizations that use unethical techniques to try to improve the ranking of their pages and web sites. Persons and organizations that use these methods are called spammers, and the set of techniques used by them, are called Spam techniques.

There are different types of Spam based on the target client: Web Spam [3] [4] or Email Spam [5]. Web Spam contains several Spam types such as: Blog/Forum Spam, Review/Opinion Spam and Social Networking Spam. Blog/Forum Spam is the Spam created by posting automatically random comments or promoting commercial services to blogs, wikis, guestbooks. Review/Opinion Spam tries to mislead readers or automated opinion mining and sentiment analysis systems by giving undeserving positive opinions to some target entities in order to promote them and/or by giving false negative opinions to some other entities in order to damage their reputations. Finally, Spam is also becoming a problem in social networks. There are existing studies in the literature about Spam in Video Social Networks [6] or Twitter [7] [8]. There are several features of Social Networks that could make Spam even more attractive:

- The target client is directly the final user. Web Spam is focused on content, so the Web Spammers try to improve the relevance of a web site by, for example, keyword stuffing. When the user conducts a search, it is likely that a Web Spam page will appear. However, it depends on the user clicking on this Web Spam page. Social Networks allow direct Spam, therefore the user will receive the Spam no matter what.
- It is focused on specific user profiles. In the case of Email Spam the content and the products of the email are generic because Spammers do not have information about target users. However, Social Networks (Facebook [9], Twitter [10] or Linkedin[2]) allow us to know a great amount of user data, so spammers use this data to aim each type of content or product at a specific audience.
- Social networks contain social network search tools to target a certain demographical segment of users.

This article focuses on Spam in the Linkedin social network. Linkedin is a social networking web site for people in professional occupations. It was founded in 2002 and in 2013, Linkedin had more than 200 million registered users in more than 200 countries.

Although some existing works have been performed to detect Spam in some well-known social networks, to the best of our knowledge this is the first one focused on the Linkedin social network and it presents a different approach to detect Spam in Social Networks. First, due to the lack of public or private Linkedin Spam datasets in the state of the art, we have generated one by means of a honeypot profile and searches of the Spam phrases. The process used to create the dataset is explained in Section V.

Second, we have created a method for detecting Linkedin Spammers. For that, we have analysed the Spammers profiles on Linkedin, and we have proposed a set of new heuristics to characterise them. Finally, we have studied the combination of these heuristics using different types of classifiers (Naïve Bayes [11], SVM [12], Decision Trees [13] and kNN [14]).

[1] http://googleblog.blogspot.com.es/2008/07/we-knew-web-was-big.html

[2] http://press.linkedin.com/about

Third, we present a method for detecting Spam Linkedin nets, that is, to detect sets of fake users created to send Spam messages to the real users connecting with them. This allows filtering those legitimate companies among fake companies, which creates a large amount of profiles for the unique purpose of generating Spam. The method is based on the similarity of their profiles and contacts. For that, the method uses distance functions (Levenshtein [15], Jaro-Winkler [16], Jaccard [17], etc.) which calculate the similarity value of each company.

After performing the experiments, we have determined that for the Spammers detection method the best classifier is kNN, and for the Spam nets detection method the best distance function is Levenshtein.

In short, these are the main contributions of this article: a) a detection method for Linkedin Spammers, b) a detection method for Spam nets and c) the first Linkedin Spam dataset.

The structure of this article is as follows. In Section II we comment on the works presented in the literature regarding the different types of Spam, and Spam techniques, as well as the ones that deal with the distinct detection methods. Section III shows the presence of Spam in social networks, specifically in Linkedin. Section IV explains the two proposed detection methods. In Section V the Linkedin Spam dataset we have created is explained. Section VI analyses the results obtained detecting Spammers profiles and Spam nets by applying the proposed methods. Finally, in sections VII and VIII we comment on our conclusions and the future works respectively.

II. Related Work

Spam has existed since the Web appeared and it has been growing in importance with the expansion of the Web. Currently, Spam is present in various applications, such as email servers, blogs, search engines, videos, opinions, social networks, etc. Different approaches to Spam detection have appeared [18] [19], however, the best results have been obtained by the methods based on the machine learning approach. Below, we analyse some of the more important articles for the different types of Spam.

There are many articles about Web Spam. Henzinger *et al.* [4] discuss the importance of this phenomenon and the quality of the results that search engines offer. Gyöngyi and Garcia-Molina [3] propose a taxonomy of this type of Spam. Ntoulas *et al.* [20] highlight the importance of analysing the content to detect this type of Spam.

On the other hand, there are studies focused on the detection of Email Spam. Among them, we highlight the work performed by Ching-Tung *et al.* [21]. This article presents a new approach for detecting Email Spam based on visual analysis, due to Spam emails embedding text messages in images to get around text-based anti-spam filters. They use three sets of features: a) embedded-text features (text embedded in the images), b) banner and graphic features (ratio of the number of banner images and ratio of the number of graphic images) and c) image location features. One of the first studies which focused on the detection of Email Spam based on machine learning, was that proposed by Sahami *et al.* [5].

Currently, Spam in social networks is booming, due to the wide use and the easy access to user data. There are several articles focused on this type of Spam.

Gao *et al.* [22] present an initial study to quantify and characterize Spam campaigns launched using accounts on online social networks. They analyze 3.5 million Facebook users, and propose a set of automated techniques to detect and characterize coordinated Spam campaigns. Grier *et al.* [23] present a characterization of Spam on Twitter. The authors indicate that 8% of 25 million URLs studied point to phishing, malware, and scams listed on popular blacklists. However their results indicate that blacklists are too slow at identifying new threats. In 2010, Wang presented an article [24], where he proposed a Spam detection prototype based on content and graph features. Another interesting articles focused on Twitter Spam, are the ones carried out by Yardi *et al.* [25] and Stringhini *et al.* [26], which study the behavior of Twitter Spammers finding that they exhibit different behavior (tweets, replying tweets, followers, and followees) from normal users (non-spammers).

With respect to the forum/blog Spam, it is necessary to highlight the study performed by Youngsang *et al.* [27]. In this work, the authors study the importance of forum Spam, and the detection of these web pages by using several new heuristics and an SVM classifier. Another study presented by Mishne [28], describes an approach for detecting blog Spam by comparing the language models used in different posts.

In the literature there are other articles related to other types of Spam. An example is the article by Jindal and Liu [29], where they present a detailed analysis about the Spam in the context of product reviews. Mukherjee *et al.* [30] presented an article focusing on detecting fake reviews. The authors propose an effective technique to detect such groups, using the following features: ratio of group size, group size, support count, time window between fake reviews, etc. Lim *et al.* [31] presented another interesting study about this type of Spam. The authors propose a supervised method to discover review Spammers. To achieve that, they identify several characteristic behaviors of review spammers and model these behaviors to perform a ranking of the different reviewers.

Finally, we want to highlight an interesting article performed by Benevenuto *et al.* [6], where the authors propose a method for detecting Spam in video social networks. They use three sets of features: a) quality of the set of videos uploaded by the user, b) individual characteristics of user behavior and c) social relationships established between users via video response interactions.

In this Section, we have presented a wide set of articles related with the different types of Spam and detection methods. Among them are several focused on Social Networking Spam, however, to the best of our knowledge, this is the first study that analyses and detects Spam in Linkedin. Due to significant differences with other social networks, it is necessary to analyse in detail its characteristics and propose new heuristics to detect Spammers and their Spam nets. Some of its major differences are: it is a professional network, premium accounts allow access to detailed user data or users can be filtered (using search tools) to conduct Spam campaigns. Moreover, its interesting and different characteristics, from the Spammers point of view, make this study both useful and necessary.

	Google+	Facebook	Linkedin	Twitter
Users	+500	+1100	+200	+500

TABLE I: Number of users, in millions, for the main social networks

III. Motivation

Spam is one of the most important challenges on the Web. Currently, due to the boom in social networks, the Spam generated in them is growing constantly. Probably, one of the main reasons for this growth is the large amount of users that they contain. In Table I we show the number of users for: Google+[3], Facebook[4], Linkedin[5] and Twitter[6].

There are several reasons why Spammers are using the social networks. These reasons can be divided into 3 topics:

- Audience:
  - Huge audience (see Table I), this means big profits for spammers, even if only a small percent visit the page or buy the product.
  - It is very easy to create Spam, because social networks allow direct Spam, that is, the Spammer knows the name and data (job, contacts, skills, etc.) of each of its victims. That is the difference with Web Spam or Email Spam where the Spammer does not know these data about its victims, only the email and perhaps the name.
- Different options and tools to create Spam:
  - Fast distribution of Spam, due to user trust and their curiosity. The users trust anything that they see posted by one of their contacts. An example of this, is the use of popular hashtags in Twitter to lure users to their Spam sites.
  - This type of Spam allows the creation of fake relationship contacts, to make the user's profile appear more real on the social network.
  - Spammers can send messages to the users and include embedded links to pornographic or other product sites designed to sell something.
  - Internet social networks contain common fan pages or groups that allow people to send messages to a lot of users even if the Spam user does not have these users as contacts.
- Little investment. Unlike other types of Web Spam, which requires investing in domains, hosting, developers, etc., social network Spam only needs accounts in social networks such as Linkedin, Facebook or Twitter.

A. How much Spam is there in Linkedin?

Currently, Linkedin Spam presents a significant problem for its users. A large amount of forums and blogs show their grievances regarding this type of Spam and offer some advice to users. Public reports of Panda Security[7] about the intense Spam campaigns in Linkedin [32], or the fake emails of Linkedin to exploit Java and Adobe vulnerabilities, show the importance of the problem [33].

One approach to determine the incidence of one problem or topic in society is to measure its impact in Google searches (or other important search engines). This method has already been used to study the presence of certain diseases in society [34].

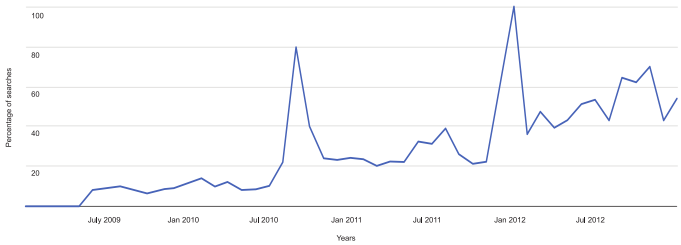

Fig. 1: Trend of the Linkedin Spam

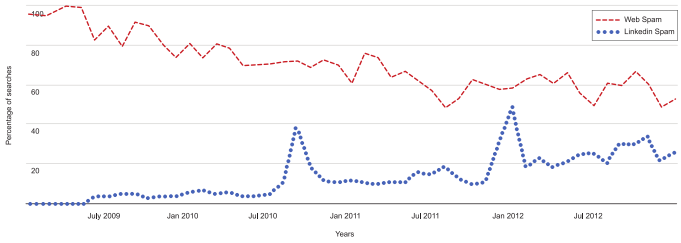

Fig. 2: Web Spam versus Linkedin Spam

In our case, we have used two methods. First, we have measured the trends in the Google searches using Google Trends. Figure 1 shows the obtained results. It depicts the relative amount of searches by year according to the maximum value obtained in 2012 (100%). With the same tool, we have also analyzed the importance of Web Spam versus Linkedin Spam. The results shown in Figure 2, show that the importance of Web Spam (the most important type of Spam) is decreasing compared to the increase in Linkedin Spam.

On the other hand, we have used another approach: searching for the query in Google *Linkein Spam*. The number of results is higher than 53 million pages, which indicates the high presence and concern about social networking Spam.

[3] http://googleblog.blogspot.com.es/2012/12/google-communities-and-photos.html
[4] http://investor.fb.com/releasedetail.cfm?ReleaseID=761090
[5] http://blog.linkedin.com/2013/01/09/linkedin-200-million/
[6] http://techcrunch.com/2012/07/31/twitter-may-have-500m-users-but-only-170m-are-active-75-on-twitters-own-clients/
[7] http://www.pandasecurity.com

B. Can existing Spam detection techniques be used on Linkedin?

Before analysing the existing Spam detection techniques we will explain the different types of Spam. There are different classifications for them, but, the classifications most relevant in the state of the art, performed by Gyongyi and Garcia-Molina [3], and Najork [35], suggest that the main types of Web Spam are: content spam, cloaking and redirection spam, click Spam and link spam.

- Content Spam, that is, a technique based on the modification of the content or the keywords of a web page with the purpose of simulating more relevance to search engines and attract more traffic. To detect this type of Spam the search engines use several algorithms and heuristics based on content analysis. However, there are a lot of Spam pages that avoid these detection algorithms. An example of these heuristics was presented by Ntoulas *et al.* [20].

- Cloaking and Redirection Spam, which consists in dynamically generating different content for certain clients (e.g.: browsers) but not for others (e.g.: crawling systems). There are several techniques to detect this type of Spam, among them we can highlight [36], [37] and [38]. The latter approach, proposed by Wu and Davison [38], where the authors detect this Spam by analysing common words across three copies of a web page.

- Click Spam: this technique is based on running queries against search engines and clicking on certain pages in order to simulate a real interest from the user. Search engines use algorithms to analyze certain logs clicks and detect suspicious behaviours.

- Link Spam, which is the creation of Web Spam by means of the addition of links between pages with the purpose of raising their popularity. It is also possible to create "link farms", which are pages and sites interconnected among themselves with the same purpose. In order to detect Link Spam, the search engines analyse relationships and the graphs between web domains. In this way, they can detect domains with a number of inlinks and outlinks or web graphs suspected of being Spam.

Finally, there is another detection technique which is not focused on a unique type of Spam. It was proposed by Webb *et al.* [39] [40], and the authors analyse the HTTP headers and their common values in Spam pages, to detect Spam.

As we can see, the existing detection techniques cannot be applied to detect Liknedin Spam. Only the algorithms for detecting Content Spam could be used, however, the problem is that the existing heuristics to detect Spam in a typical web page, cannot be applied to web page profiles of Linkedin because their features are completely different. Due to existing detection techniques being impossible to apply, Linkedin has proposed its own particular detection techniques (see Section III-D).

C. How do Spammers create Linkedin Spam and what are the differences between it and other social networks?

The first step for a Spammer is to decide whether the Spamming attack will be a focused or general attack. In the case of it being a focused attack, the Spammer will search for specific users by means of the Linkedin tools. After that, the Spammer can create Spam by the following methods:

- Messages: these are sent by any one of our contacts. However, we often accept contacts because of having contacts in common, or because we think that the job, groups or skills of this user are appropriate for them to be our contact, and perhaps, it could be a job opportunity.

- Groups: that is, those notifications sent to the groups of each of the victims. These messages will be sent by email to the users of the group, and moreover, this post can be seen in the forum of the group.

- Updates: the Spammer makes updates to their profile to invite their contacts to visit their profile.

After we have analysed the operation of Linkedin Spam, we explain the differences in the creation Spam between Linkedin and other social networks.

- Linkedin allows the use of social networks search tools to target a certain demographical segment of the users. This allows the Spam to be made more specific, and therefore, it is likely that the victim will click on the Spam link.

- Linkedin is a social network which focuses on business, companies and professionals, which is very interesting from the point of view of the Spammers. In other words, the possible profit will be higher if the person that visits the Spam site is a businessman instead of a teenager from Facebook. Twitter, Facebook or email can be used for professional ends, however they do not contain the other advantages of Linkedin.

- Linkedin allows direct Spam, such as Twitter, Facebook or email, but in this case, the Spammer knows the name and data (job, contacts, skills, location, etc.) of each of their victims. So, the probability of success is higher in Linkedin than in Web or Email Spam.

- Due to Linkedin being defined as a professional social network, usually the data of its users are real, and the usage of it by them is usually a way to find a job or to find new professional contacts; this is not a game. For this reason, when a Linkedin user receives a message, email or update from Linkedin, they pay more attention to it than to notifications of other social networks. Again, the probability of Spam success is higher.

In summary, the Linkedin Spam is made by means of emails or messages to the victims, in their interest groups or in the updates of the Spammer. So, as we have said, the Spam detection techniques based on content analysis could be used, however new heuristics would have to be created specifically for this enviroment. As a new approach, we propose the detection of Spammers and their Spam nets, instead of Spam

in a particular email, message or comment, which is more difficult. So, if we can detect a Spammer, we can also detect all their Spam messages (emails, comments and updates).

D. How does Linkedin detect its Spam?

Currently, Linkedin uses two techniques to detect Spam. On the one hand, when a user receives an invitation to become a contact of another user, he can indicate that this person is a Spammer. A Linkedin user has numerous methods of contacting a specific user (as a friend, as a coworker, as a classmate, etc.). Linkedin blocks a contact method to a user profile (Spammer), when it has received 5 requests rejecting said account by the same contact method indicating that it is Spam. Alternatively, the user can report the profile of the Spammer to the following e-mail address: abuse@linkedin.com.

However, from our opinion, these methods are not sufficient, due to two reasons. First, people are lazy, and because of that they will usually not accept this person but also will not usually notify that the user is a Spammer. For the same reason, only in a few cases, the user sends an email to report a Spammer. The second reason is because of the speed and ease with which criminal organizations and Spammers can create a lot of accounts, compared to the slow detection methods used by Linkedin.

In summary, as we have explained, the presence and concern of Spam in social networks is high. Due to this, together with the lack of articles about Linkedin Spam, existing methods for detecting Spam cannot be applied and the need for other methods to complement tools used by Linkedin, we propose a method for detecting Linkedin Spammers, and a method for identifying fake companies (Spam nets).

IV. Detection Methods

We present two detection methods, one for detecting Spammers and another for detecting Spam nets, both in the Linkedin social network. In order to know and understand the behaviour of Linkedin Spammers and Spam nets, and also to test the proposed methods, we have manually built a dataset of Linkedin profiles, classifying them as spammers and legitimate users (see Section V).

The method for detecting Spammers (section IV-A) is based on a set of new heuristics together with the use of machine learning. The heuristics have been obtained by means of the manual and statistical analysis of the legitimate and Spam Linkedin profiles. So, we characterize a Linkedin profile, and then decide whether or not it is Spam. For the appropriate combination of these heuristics, we have tried different classification techniques (decision trees, techniques based on rules, neuronal networks and kNN).

To detect Spams nets in Linkedin (section IV-B), we have focused on the idea that we have observed during the manual analysis. The fake profiles of a fake company, usually share similarities that allow differentiation between legitimate companies and fake companies.

A. Method to detect Linkedin Spammers

We will discuss a set of heuristics that aim to characterize and detect Spam profiles in Linkedin. Some of the features we present below, have appeared because Linkedin is a professional social network, and their users are very careful with the details of their profiles. Linkedin users want to have an updated and complete profile.

The results obtained for each heuristic were tested on the dataset described in Section V. For each non binary feature, we include a figure showing a box and whisker diagram with the feature values, corresponding to Spam and Non-Spam pages. For binary features, we only present the percentage of use for each type of page.

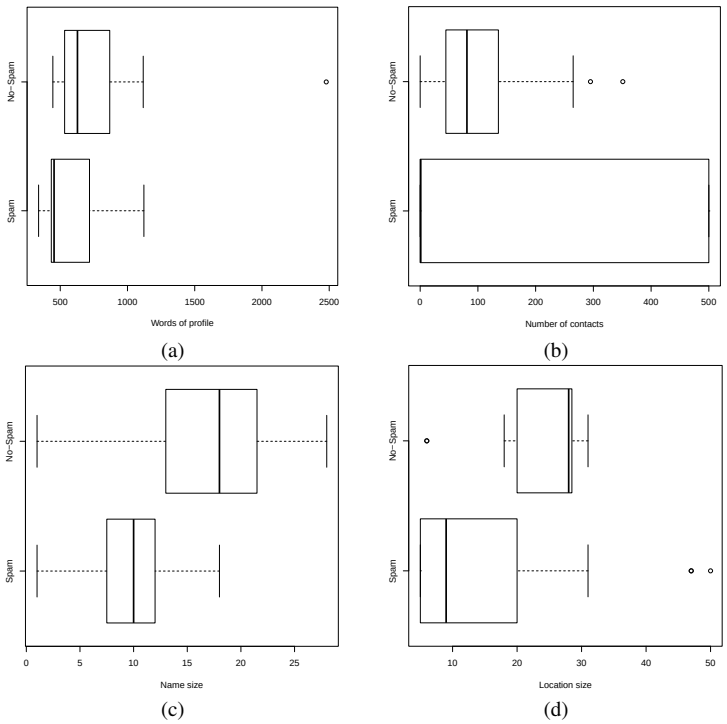

Fig. 3: Number of words (a), contacts (b), name size (c) and location size (d) in the Spam and No-spam profiles

The features analysed are the following:

- Number of words in profile: we have analysed this feature because during the manual labeling of the pages we have observed, that Spam pages usually contain less words than the legitimate Linkedin profiles. Figure 3a shows that the median number of words in Spam profiles is 454 words. In other words, the Spam profiles contain on average 559.3 words and the No-Spam profiles 741.8 words, 24.6% lower.
- Number of contacts: due to the automatic generation of Spam profiles, they present two clear behaviors: profiles with very few contacts or profiles with many contacts. On the other hand, the legitimate profiles follow a uniform distribution of contacts, without these extreme differences. We observe in Figure 3b that in Spam profiles the median is 1 and in the No-Spam profiles it is 81. Moreover, the difference

between averages is very significant. Specifically, Spam profiles contain on average 204.8 contacts, while No-Spam profiles contain only 98.1.

- Name size: in this case, the deficiency of Spam profiles appears in the name of the person. Fake profiles usually contain shorter names and surnames than in legitimate profiles. This is because Linkedin users are very careful and want their data profile to be correct and updated. To achieve this, they use their complete name and do not tend to use short names or nicknames.
  As we thought, the results indicate that legitimate profiles have longer names than fake profiles. Figure 3c shows that the median and average in Spam profiles is 10 and 9.09 letters and in No-Spam profiles is 18 and 17.41, respectively.

- Location size: we have observed that Spam profiles usually contain a simple and smaller location than in the legitimate profiles. Moreover, the location among the fake profiles of the fake companies are very similar.
  In Figure 3d shows the median and average location size in Spam profiles to be 9 letters and 14.72, respectively. In the case of No-Spam profiles, the median is three times higher, than the Spam profiles, 28 letters, and the average is 24.64 letters.

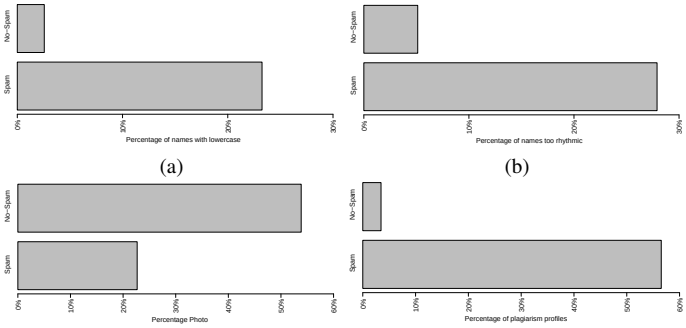

Fig. 4: Percentages of the names written in lowercase (a), percentage of rhythmic names (b), percentage of profiles with photo (c) and percentages of the plagiarism profiles (d) in the Spam and No-spam profiles

- Name written in lowercase: another big weakness of fake profiles, is that their name or surname, are often written in lowercase. Figure 4a indicates that more than 20% of the Spam profiles contain user names in lowercase, and in the legitimate profiles this value is almost 10 times smaller.

- Rhythmic name: a technique used to draw the users attention and build trust in the profile. Specifically, it was observed that often the Spam profiles contained people whose first and last names start with the same two or three letters. Figure 4b shows that the 5.12% of the No-Spam profiles contain a rhythmic name and lastname, but this figure raises up to more than five times, 27.90, in the Spam profiles.

- Profile with photo: we have noted two weaknesses in the fake Linkedin profiles regarding this issue. First, in a social network the users usually have photo, in the case of the Spam profile, they usually do not. And second, if a Spam profile contains a photo, this photo can usually be found by a search engine. In Figure 4c we observe that only 22.67% of the Spam profiles contain photo, and in the case of legitimate profiles this value is more than double, specifically 53.84%.

- Plagiarism in profiles: another weakness of automatically generated Spam profiles, is that their content is small, or, due to the difficulty in generating logical content, the Spammers take texts from the Internet.
  We have used the Grammarly plagiarism checker[8]. This system finds unoriginal text by checking for plagiarism against a database of over 8 billion documents.
  The Figure 4d shows that multiple Spam profiles have copied or automatically generated content. The differences among the results are very significant, specifically 3.45% in No-Spam profiles and 56.53% in Spam profiles.

As we have seen, Spam profiles tend to be simpler and contain less detail than legitimate profiles. Moreover, the results obtained for each type of profile (Spam and No-Spam) show significant differences between them. In multiples cases the results are 2, 3, 5 or even 10 times higher or smaller in Spam profiles than in legitimate profiles. These important differences allow the proposed heuristics to be used to characterize and detect Linkedin Spam.

The detection method proposed uses these heuristics together with machine learning techniques to identify Spammer profiles. The method is to not focus on a specific heuristic but to use all of them. In the case of failure of a particular heuristic, since the method uses all heuristics, the other heuristics will correct this error. For the appropriate combination of heuristics we have tried different machine learning techniques (decision trees, kNN, SVM and Naïve-Bayes). Based on the obtained results (see Section VI), the method for combining the heuristics is kNN.

B. How to Detect Spam Nets?

We have studied a method for detecting fake companies, companies whose unique purpose is to create fake profiles to generate Spam. The proposed method identifies this type of Linkedin companies based on the similarity among the profiles that each company contains.

To measure the similarity between profiles, we have generated a text string that contains the different data of the profile separated by commas ",". A generic example of the text string generated with the Linkedin data profile and the results obtained with the proposed heuristics, is the following:

UserName, ProfileTitle, Location, NumberofContacts, Skills, Education, NumberOfWords, NameSize, LocationSize, RhytmicName, Photo, LowercaseName, Profile-Plagarism

[8] www.grammarly.com/Plagiarism_Check

The value of the user name, title of the profile, user location, number of contacts, skills and education are extracted directly from the profile of the user. However, the variables: NumberOfWords, NameSize, LocationSize, RhythmicName, Photo, LowercaseName and ProfilePlagarism are calculated previously, based on the analysis of the profile. Finally, the value of RhythmicName, Photo, LowercaseName and ProfilePlagarism are boolean.

As a preliminary step, we specify the following concepts to help us formally define our method:

- Let be $pi = Profile\ of\ the\ user\ i.$
- Let be $N_p = Number\ of\ profiles\ of\ a\ company.$
- Let be $distance_{ij} = Similarity\ between\ the\ profiles\ i\ and\ j.$
  To obtain this value, we have studied different functions:
  - Levenshtein [15]: is the minimum number of edits needed to transform one string into the other (using insert, delete or replace operations). This distance is usually denoted as edit distance.
  - Jaro: is a similarity function which defines the transposition of two characters as the only permitted operation to edit. The characters can be a distance apart depending on the length of both text strings.
  - Jaro-Winkler [16]: is a variant of the Jaro metric, which assigns similarity scores higher to those words that share some prefix.
  - Jaccard [17]: it defines the similarity between two text strings A and B as the size of the intersection divided by the size of the union of the corresponding text strings.
  - Cos TF-IDF [41]: given two strings A and B, and, α_1, α_2 ... α_K and β_1, β_2 ... β_L their tokens respectively, they can be seen as two vectors, V_A and V_B, with K and L components. So, the similarity between A and B, can be calculated as the cosine of the angle of these two vectors.
  - Monge Elkan [42]: given two strings A and B, and, α_1, α_2 ... α_K and β_1, β_2 ... β_L their tokens respectively. For each token α_i there is a β_j with maximum similarity. Then the Monge Elkan similarity between A and B, is the average maximum similarity between a couple (α_i, β_j)
- Let be S_{pi} the similarity of a profile, pi, with the other profiles of his company. This value is calculated as the sum of the distances between the text string of the corresponding profile with the text strings of the rest of profiles, divided by the number of profiles, N_p, minus 1.

$$S_{pi} = \frac{\sum_{j=0}^{N_p-1} distance_{ij}}{N_p - 1}$$

We now can define the method we propose to obtain the value, S_c, that summarises the similarity of a specific company. S_c is calculated as the average of the similarities of the profiles, S_{pi}, of the staff of the company.

$$S_c = \frac{\sum_{i=0}^{N_p} S_{pi}}{N_p}$$

After we have obtained the value of similarity of a company, we have to decide if said company is fake or legitimate. In order to do that, we have calculated a threshold for each similarity function. These thresholds allow us to decide when a company contains very similar profiles, and this company will likely be fake, or conversely, the profiles are different enough to be a legitimate company. For that, we have created a training set that contains 4 fake and 4 legitimate companies with the highest number of profiles. In Table II we show the similarity results obtained in this training set. Among the results obtained, we have selected as thresholds those that have obtained the best results (precision, recall and F-Measure) in the training set. The thresholds selected were used to obtain the results (precision, recall and F-Measure) of the method in the full dataset.

Analysing the results we can see that there are differences about the similarity values obtained by each technique. Levenshtein and Cos TF-IDF have obtained the lower values of similarity, around 0.6 and 0.5 respectively. In the other hand, Jaccard obtains the highest results, close to 1. Jaro and Monge Eklan have obtained intermediate results, with values between 0.7 and 0.8. The results obtained by Jaro and Jaro-Winkler, as we expected, are different. This is because Jaro-Winkler scores words which share some prefix higher and we had observed that the string created contains prefixes that increase the similarity result.

If we study the results of the legitimate companies and the fake companies separately, we can observe that, as we thought, fake companies display more similarity between them than the normal companies. This fact can easily be seen in the results obtained by Levenshtein and Cos TF-IDF measures.

In Section VI-C we present the precision, recall and F-Measure applying the proposed method on the created dataset.

V. LINKEDIN SPAM DATASET

To the best of our knowledge, there is no public dataset of Linkedin profiles. The dataset we have built contains legitimate and Spam profiles. The creation of the dataset was carried out during 30 days, from October 30th to November 30th, 2012.

For the legitimate profiles we have used profiles of users in well known companies (Google, Microsoft, Oracle, Twitter, IBM, etc.). The gathering process of the legitimate profiles was made automatically by means of the Linkedin API[9]. We have used our Linkedin profiles to obtain user profiles of the legitimate companies. The process starts in the public profile of an employee of a legitimate company and continues by the contacts of this user who works in the same company.

On the other hand, we have identified a set of Spam users, based on the Spam messages that they send to other users, and the profiles obtained by searching for words that

[9]https://developer.linkedin.com/

		Levenshtein	Jaro	Jaro-Winkler	Jaccard	Cos TF-IDF	Monge Elkan
Dinowill	Fake	0.745	0.838	0.903	0.993	0.640	0.840
Innovabiz	Fake	0.693	0.768	0.861	0.995	0.551	0.791
Online Pharm	Fake	0.739	0.706	0.824	0.994	0.630	0.819
Pharmacity	Fake	0.723	0.743	0.845	0.991	0.566	0.792
Cisco	Legitimate	0.571	0.730	0.838	0.996	0.403	0.778
Google	Legitimate	0.640	0.736	0.841	0.994	0.477	0.791
Hp	Legitimate	0.639	0.745	0.825	0.995	0.489	0.756
Motorola	Legitimate	0.662	0.756	0.850	0.995	0.539	0.759

TABLE II: Similarity of the analysed fake and legitimate companies

commonly appear in Spam comments in Google, such as "viagra", "growth hormone", "cialis", etc. The list of these words has been obtained by searching Spam words in the WordPress Codex[10], the online manual for WordPress. After this, we created a fake profile on Linkedin, as a honeypot, and sent contact requests to the Spam profiles. All the contact requests were accepted, and usually the Spammers responded 1 or 2 days after the request. Once accepted, as we thought, we could detect new fake profiles among their contacts. The labeling and gathering process of each fake profile was made manually due to the need to check whether each profile was really a fake.

We want to clarify that to obtain legitimate profiles we have not associated the created fake profile with the reputed companies. Both legitimate profiles and fake profiles have not been published anywhere and they have been stored encrypted.

To know more details about the generated dataset, we explain its structure:

- Its size is 1,4 GB.
- It contains 750 profiles, 250 Spammers and 500 legitimate users.
  To decide the size of the dataset, and the subsets (Spammers and legitimate users), we have had a problem because there are no studies about how much Spam there is in Linkedin. So, we have decided that the number of Spam profiles is similar to Spam contained in .com domain [20]. Moreover, we have used a percentage of Spam higher than .com domain, because we want to increase the variability of the profiles in order to make the detection process more difficult.
- The profiles are divided among 150 companies, of which 50 are fake companies and 100 are legitimates.
- Among the fake companies are companies focused on different drugs like Viagra or Cialis, and on Chinese products.
- 80% of the legitimate companies are focused on technology and computer science area.
- Companies are mainly located in USA.
- The size of the legitimate companies is variable, ranging from companies with 500 employees to those with more than 20,000.

The provided data about the used dataset allows other researchers to test and compare our methods. It is likely that their dataset does not contain the same profiles as our dataset, however, in our opinion, they can create a dataset with the same structure and features, and therefore, the results should be very similar.

[10]http://codex.wordpress.org/Spam_Words

VI. Experimental Results

In this Section, we discuss all the issues we found for both the execution and the assessment stages, and we show and analyse the results obtained. First, we show the results obtained when detecting Spammers in Linkedin using the proposed heuristics (see Section IV-A), and then, the results obtained to detect Spam groups or companies, analysing the similarity between the company user profiles.

A. Experimental Setup

To execute the different classifiers, we used WEKA [43], a tool for automatic learning and data mining, which includes different types of classifiers and different algorithms for each classifier. The techniques tested were: SVM, Naïve Bayes, Decision Trees and Nearest Neighbour. To obtain the results, we have used the default parameters in WEKA for each of the machine learning algorithms, specially the value k used in kNN has been 1.

To evaluate the classifier we used the "cross validation" technique [44], that consists in building k data subsets. In each iteration a new model is built and assessed, using one of the sets as a "test set" and the rest as "training set". We used 10 as the value for k ("ten-fold cross validation").

The dataset used to obtain the results was created by us, because there is no public dataset of Linkedin profiles. The method used to generate it and its characteristics were explained in Section V.

B. Results for Linkedin Spammers

This section discusses the results obtained applying the proposed method to discover Linkedin Spammers. Table III shows the precision, recall and F-Measure for each of the types of classifiers studied.

	Precision	Recall	F-Measure
Naïve Bayes	0.909	0.908	0.909
SVM	0.837	0.831	0.794
Decision Trees	0.967	0.967	0.967
kNN	0.969	0.979	0.971

TABLE III: Results of the proposed heuristics using different types of classifiers

The analysis of the results shows that the best results are obtained using kNN and the worst with SVM. We also observe that the results of kNN and decision trees are very similar. Specifically, kNN obtains a precision and a recall of 0.979, and an F-Measure of 0.971. These results are similar to those obtained by applying decision trees, where the recall, precision, and therefore, F-Measure is 0.967. This fact, that the three measures are equal occurs because the number of false positives and false negatives are the same. In this case, this occurs because, based on the results, the number of false positives and false negatives are very small, and it is likely that these values match.

On the other hand, SVM achieves 0.837, 0.831 and 0.794 of precision, recall and F-Measure, respectively. Despite these small differences we observe that the results are very satisfactory.

Another issue analysed to determine the right performance of the proposed heuristics is to study the ROC curve of the classifiers. In Figure 5 we show the ROC curve obtained by each of the studied classifier. Again, the best results are obtained using kNN and decision trees. Analysing the area under de ROC (AUC), kNN and decision trees achieves 0.984 and 0.976, respectively, that is, almost a perfect result. On the other hand, Naïve Bayes obtains an AUC of 0.934. Finally, SVM achieves the worst result, with an AUC of 0.629, that means that although, this classifier has obtained a good precision and recall, it is not reliable.

In short, the precision, recall and the ROC curve indicate that the proposed heuristics are very adequate to detect Spammer profiles in Linkedin. Moreover, we have detected that the best classifier for this type of Spam is kNN or decision trees.

In general, the obtained results are very hopeful. In our opinion, this fact is in part due to the fact that Social Networking Spam is still at an early stage.

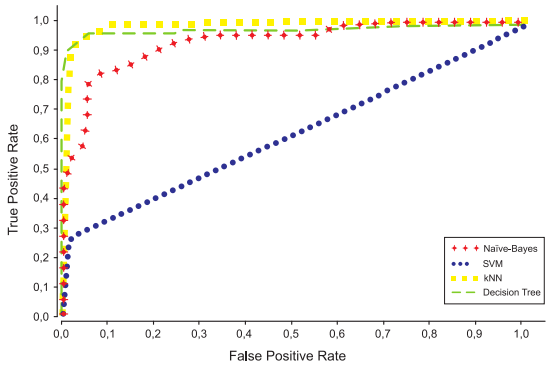

Fig. 5: ROC curve of the tested classifiers

C. Results for Spam Linkedin Nets

In this section we discuss the results obtained to detect fake companies in Linkedin, whose only purpose is to generate Spam. For that, we have applied the method explained in Section IV-B.

The results shown in Table II were used as training set to select the thresholds with best results. The proposed method used the selected thresholds in all labeled companies, legitimate and fake, and we have calculated the precision, recall and F-Measure for each of them. The thresholds used and their corresponding results are shown in Table IV.

	Threshold	Precision	Recall	F-Measure
Levenshtein	0.693	1	1	1
Jaro	0.743	0.750	0.750	0.750
Jaro-Winkler	0.845	1	0.750	0.857
Jaccard	0.993	0.571	1	0.727
Cos TF-IDF	0.551	1	1	1
Monge-Elkan	0.791	0.8	1	0.8

TABLE IV: Results to detect Spam Linkdin nets

The best results are obtained using Levenshtein and Cos TF-IDF. In these two cases the detection is perfect, with an F-Measure of 1. The results indicate that the method used by these two techniques to measure the similarity between profiles, is the most adequate for the text string generated by our method.

In the analysis of the results, we have detected two types of Spam profiles. On the one hand we have very simple Spam profiles and with little content, and, on the other hand, complex Spam profiles with too much content (in most of cases, automatically generated). To detect this second type all the proposed heuristics must be used because, otherwise, it could be skipped.

However, the fake companies only contain one of these two types of Spam profiles and always with very similar content. In short, the method proposed to detect Spam nets has achieved hopeful results, mainly due to two facts: a) the idea proposed is right, and the fake companies contain similar profiles and b) the Social Networking Spam is relatively new and its techniques are unsophisticated. In the future, it is likely that these techniques will improve. However, we have demonstrated that the proposed idea works perfectly, and can be used, in the future, as a base to be complemented with other new techniques.

VII. Conclusions

The presence of different types of Spam (Web Spam, Email Spam, Forum/Blog Spam and Social Networking Spam) on the Web is important and is constantly growing. There are many studies that analyse and present techniques for the detection of different types of Spam. However, to the best of our knowledge, there are no studies about Spam in Linkedin.

In this article, we present a method to detect Spammers and Spam nets in Linkedin social network. We have proposed a set of heuristics that characterize Linkedin Spam profiles and help to identify Linkedin Spammers. These heuristics were used as input to several classification algorithms (Naïve Bayes, SVM, Decision Trees, kNN). The best results are obtained by kNN and decision trees, with an F-Measure of 0.969 and 0.967, respectively, and an AUC close to 1.

Moreover, we have proposed a method for detecting Spam nets in Linkedin. It is based on the idea that the profiles of fake companies share multiple similarities. The method calculates the similarity between different profiles of the companies,

using several distance functions (Levenshtein, Jaro, Smith-Waterman, etc.). The values of similarity obtained are used as thresholds to detect fake companies (Spam nets) among the legitimate companies. Again, the results are also very hopeful. We have achieved an F-Measure of 1 using Levenshtein and Cos TF-IDF.

In short, the results obtained in the study show that, on the one hand, the heuristics proposed are adequate to detect Spammer profiles, and, on the other hand, the new method proposed to detect Spam nets (fake companies) in Linkedin performs very well.

VIII. Future Works

Spam in Linkedin social network is a relatively new Spam type, so our intention is to follow its evolution over time. Due to the continuous changing of Spam techniques, we want to find new and better heuristics to detect this type of Spam. Furthermore, we plan to increase and improve our labeled dataset. Finally, we will test the proposed heuristics and the method for detecting Spam Linkedin nets, in other social networks.

Acknowledgment

This research was supported by Xunta de Galicia CN2012/211, the Ministry of Education and Science of Spain and FEDER funds of the European Union (Project TIN2009-14203).
The list of domains "es" of 2009, from which we conducted the crawling process, has been provided by the Spanish Business Public Entity Red.es.

References

[1] A. Gulli and A. Signorini, "The indexable web is more than 11.5 billion pages," in *Special interest tracks and posters of the 14th international conference on World Wide Web*, ser. WWW '05. New York, NY, USA: ACM, 2005, pp. 902–903.

[2] B. J. Jansen and A. Spink, "An analysis of web documents retrieved and viewed," 2003.

[3] Z. Gyongyi and H. Garcia-Molina, "Web spam taxonomy," Stanford InfoLab, Technical Report 2004-25, March 2004.

[4] M. R. Henzinger, R. Motwani, and C. Silverstein, "Challenges in web search engines," *SIGIR Forum*, vol. 36, pp. 11–22, September 2002. [Online]. Available: http://doi.acm.org/10.1145/792550.792553

[5] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A bayesian approach to filtering junk e-mail," 1998.

[6] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and K. Ross, "Video interactions in online video social networks," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 5, no. 4, pp. 30:1–30:25, Nov. 2009.

[7] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on twitter," 2010.

[8] K. Lee, J. Caverlee, and S. Webb, "The social honeypot project: protecting online communities from spammers," in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10, 2010, pp. 1139–1140.

[9] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," *CoRR*, vol. abs/1111.4503, 2011.

[10] 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*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 591–600. [Online]. Available: http://doi.acm.org/10.1145/1772690.1772751

[11] H. Zhang, "The optimality of naive bayes," in *FLAIRS Conference*, V. Barr and Z. Markov, Eds. AAAI Press, 2004.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[13] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[14] G. Shakhnarovich, T. Darrell, and P. Indyk, *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing)*. The MIT Press, 2006.

[15] V. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," *Soviet Physics Doklady*, vol. 10, p. 707, 1966.

[16] W. E. Winkler, "String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage," in *Proceedings of the Section on Survey Research*, 1990, pp. 354–359.

[17] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining, (First Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

[18] Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, "Spamming botnets: signatures and characteristics," *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 171–182, Aug. 2008.

[19] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen, "Combating web spam with trustrank," in *Proceedings of the Thirtieth international conference on Very large data bases - Volume 30*, ser. VLDB '04, pp. 576–587.

[20] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in *Proceedings of the 15th international conference on World Wide Web*, ser. WWW '06, pp. 83–92.

[21] C.-T. Wu, K.-T. Cheng, Q. Zhu, and Y.-L. Wu, "Using visual features for anti-spam filtering," in *Image Processing (ICIP 2005) IEEE International Conference on*, vol. 3, pp. 509–12.

[22] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and characterizing social spam campaigns," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, ser. IMC '10. New York, NY, USA: ACM, 2010, pp. 35–47.

[23] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: the underground on 140 characters or less," in *Proceedings of the 17th ACM conference on Computer and communications security*, ser. CCS '10. New York, NY, USA: ACM, 2010, pp. 27–37.

[24] A. H. Wang, "Don't follow me - spam detection in twitter." in *SECRYPT*, S. K. Katsikas and P. Samarati, Eds., 2010, pp. 142–151.

[25] S. Yardi, D. M. Romero, G. Schoenebeck, and D. Boyd, "Detecting spam in a twitter network," *First Monday*, 2010.

[26] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *Proceedings of the 26th Annual Computer Security Applications Conference*, ser. ACSAC '10. New York, NY, USA: ACM, 2010, pp. 1–9.

[27] Y. Shin, M. Gupta, and S. Myers, "Prevalence and mitigation of forum spamming," in *INFOCOM, 2011 Proceedings IEEE*, pp. 2309 –2317.

[28] G. Mishne, "Blocking blog spam with language model disagreement," in *In Proceedings of the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb*, 2005.

[29] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the international conference on Web search and web data mining*, ser. WSDM '08. New York, NY, USA: ACM, 2008, pp. 219–230.

[30] A. Mukherjee, B. Liu, J. Wang, N. Glance, and N. Jindal, "Detecting group review spam," in *Proceedings of the 20th international conference companion on World wide web*, ser. WWW '11. New York, NY, USA: ACM, 2011, pp. 93–94. [Online]. Available: http://doi.acm.org/10.1145/1963192.1963240

[31] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, ser. CIKM '10. New York, NY, USA: ACM, 2010, pp. 939–948. [Online]. Available: http://doi.acm.org/10.1145/1871437.1871557

[32] "Linkedin spam campaign," http://pandalabs.pandasecurity.com/linkedin-spam-campaign/.

[33] "Linkedin spam serving adobe and java exploits,"

http://pandalabs.pandasecurity.com/linkedin-spam-serving-adobe-and-java-exploits/.

[34] J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data." *Nature*, vol. 457, no. 5, pp. 1012–4, 2009.

[35] M. Najork, "Web spam detection." in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, Eds. Springer US, 2009, pp. 3520–3523.

[36] ——, "System and method for identifying cloaked web servers," Patent 20 030 131 048.

[37] K. Chellapilla and A. Maykov, "A taxonomy of javascript redirection spam," in *Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*, ser. AIRWeb '07. New York, NY, USA: ACM, 2007, pp. 81–88. [Online]. Available: http://doi.acm.org/10.1145/1244408.1244423

[38] B. Wu and B. D. Davison, "Cloaking and redirection: A preliminary study," 2005.

[39] S. Webb, J. Caverlee, and C. Pu, "Characterizing web spam using content and http session analysis."

[40] ——, "Predicting web spam with http session information," in *Proceedings of the 17th ACM conference on Information and knowledge management*, ser. CIKM '08, 2008, pp. 339–348.

[41] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, Nov. 1975.

[42] A. Monge and C. Elkan, "The field matching problem: Algorithms and applications," in *In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 267–270.

[43] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, pp. 10–18, November 2009. [Online]. Available: http://doi.acm.org/10.1145/1656274.1656278

[44] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," 1995, pp. 1137–1143.