



ISSN: 0975-833X

Available online at <http://www.journalcra.com>

*International Journal of Current Research*  
Vol. 8, Issue, 02, pp. 26737-26742, February, 2016

**INTERNATIONAL JOURNAL  
OF CURRENT RESEARCH**

## RESEARCH ARTICLE

# MACHINE LEARNING TECHNIQUES FOR FILTERING NOISY CONTENTS IN ONLINE SOCIAL NETWORK

<sup>1</sup>Kiruthika, P. and <sup>2</sup>Kalaiprasath, R.

<sup>1</sup>Selvamm Arts & Science College, Namakkal, TamilNadu, India

<sup>2</sup>Aksheyaa College of Engineering, Chennai, India

### ARTICLE INFO

#### Article History:

Received 12<sup>th</sup> November, 2015  
Received in revised form  
23<sup>rd</sup> December, 2015  
Accepted 12<sup>th</sup> January, 2016  
Published online 27<sup>th</sup> February, 2016

#### Key words:

Short Text Classification,  
Content based filtering,  
Policy based personalization,  
Online Social Networks.

### ABSTRACT

Today online social networks has provide only little support for prevent the displaying of noisy contents on users space. To address this issue, we propose directly to control the noisy contents in user private/public space. This is achieved by using machine learning techniques to automatically assign with each short text messages under the categories based on its content with the help of content based filtering and users customize to filter without displaying of noisy content from their own private space by applying of filtering rules using rule based system.

*Copyright © 2016 Kiruthika and Kalaiprasath. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

**Citation:** Kiruthika, P. and Kalaiprasath, R. 2016. "Machine Learning Techniques For Filtering Noisy Contents in Online Social Network", *International Journal of Current Research*, 8, (02), 26737-26742.

## INTRODUCTION

Online Social Network (OSN) is one of the popular for made communication between users. It offer users to share some information to others by using several contents such as text, image, audio, video data. However, the majority aim of the proposals is mainly to provide a classification mechanism to avoid they are overwhelmed by useless data. A main part of social network content is constituted by short text, a notable example are the messages permanently written by OSN users on particular public/private areas, called in general walls. Today OSNs provide very little support to prevent unwanted messages on user walls. For example, Facebook allows users to state who is allowed and who is not allowed to insert messages in their private/public space. It has no content-based preferences are supported and therefore it is not possible to prevent undesired messages. In traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences. The aim of the current paper is to propose and experimentally evaluate an automated

system, called Filtered Wall (FW), able to filter noisy contents from OSN user space. The key idea of the proposed system is the support for content based user preferences. We exploit Machine Learning (ML) text categorization techniques (Sebastiani, 2002) based soft classifier automatically labelling messages with the help of content based filtering. To classify the text for labelling messages by using the short text classifier (STC). Short text classifiers (Vanetti *et al.*, 2010) are concentrated in the extraction and selection of a set of characterizing features. The overall short text classification strategy on Radial Basis Function Networks (RBFN) for their proven capabilities in acting as soft classifiers, in managing noisy data. We insert the neural model within a hierarchical two level classification strategy. In the first level, the RBFN categorizes short messages as Neutral and Nonneutral. In the second level, Nonneutral messages are classified producing gradual estimates of appropriateness to each of the considered category. In contrast, by means of the proposed mechanism, a user can specify what contents should not be displayed on his/her wall, by specifying a set of filtering rules. Filtering rules are allowed to specify filtering conditions based on user profiles.

**\*Corresponding author: Kalaiprasath, R.**  
*Aksheyaa College of Engineering, Chennai, India.*

In addition, the system provides the support for user-defined Blacklists (BLs), that is, lists of users that are temporarily prevented to post any kind of messages on a user wall.

### Related Work

Our related work has using content-based filtering and policy-based personalization to filtering noisy contents in Online Social Network user's space.

### Content Based Filtering

Information filtering systems also involve a large amount of data and streams of incoming data, whether broadcast from a remote source or sent directly by other sources. Filtering is based on descriptions of individual or group information preferences or profiles. Filtering also implies removal of data from an incoming stream (Belkin and Croft, 1992).

In content-based filtering system selects information items based on the correlation between the content of the items and the user similar preferences. The most common filtering systems are difficult for automatic multilabel text categorization. Content based filtering has using machine learning techniques (Sebastiani, 2002) automatically to classify the text by learning from the preclassified set. The feature extraction procedure maps text into a compact representation of its content and is uniformly applied to training. Several experiments prove that Bag-of-Words (BoW) approaches yield good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statistical quality (Dumais et al., 1998; Lewis, 1992). In our scenario, we consider gradual membership to classes a key feature for defining flexible policy-based personalization strategies.

### Policy-Based Personalization

To Exploiting classification mechanisms for personalizing access in OSNs. A classification method (Sriram et al., 2010) has been proposed to categorize short text messages in order to avoid overwhelming users of micro blogging services by raw data. The system described in (Sriram et al., 2010) focus on micro blogging services such as twitter, the users may become overwhelmed by the raw data. To avoid this problem has using classification of short text messages. It classifies incoming tweets into categories. The classification method has using Bag-of -Words (BoW) approach to classify the text to a predefined set of generic classes. The user can then view only certain types of tweets Based on their interests. However, such systems do not provide a filtering policy layer by which the user can. Exploit the result of the classification process to decide how and to which extent filtering out unwanted information. In contrast, our filtering policy language allows the setting of FRs according to a variety of criteria that do not consider only the results of the classification process but also the relationships of the wall owner with other OSN users as well as information on the user profile. In our work has inspired by the many access control Models and related policy languages. To enforce mechanisms that has been proposed so far for OSNs (Bonchi and Ferrari, 2010).

In content filtering can be considered and extension of access control, since it can be used both to protect objects from unauthorized subjects, and subjects from inappropriate objects. In OSN has the majority of access control models proposed so far enforce topology-based access control. Our system is the availability of a description for the message contents to be exploited by the filtering mechanism.

### Filtered wall Conceptual Architecture

The aim of this paper is to develop a method that allows OSN users to easily filter undesired messages, according to content based criteria. In particular, we are interested in defining an automated language-independent system providing a flexible and customizable way to filter and then control incoming messages. In general, the architecture in support of OSN services is a three-tier structure. The first layer commonly aims to provide the basic OSN functionalities (i.e., profile and relationship management). According to this reference layered architecture, the proposed system has to be placed in the second and third layers, as it can be considered as a SNA. In particular, users interact with the system by means of a GUI setting up their filtering rules, according to which messages have to be filtered. Moreover, the GUI provides users with a FW that is a wall where only messages that are authorized according to their filtering rules are published.

The Fig.1 core components of proposed system are the Content-Based Messages Filtering (CBMF) and the Short Text Classifier (STC) modules. The latter component aims to classify messages according to a set of categories. The first component exploits the message categorization provided by the STC module to enforce the filtering rules specified by the user. Note that, in order to improve the filtering actions, the system makes use of a blacklist (BL) mechanism. By exploiting BLs, the system can prevent messages from undesired users. The system is able to detect who are the users to be inserted in the BL according to the specified user preferences, so to block all their messages and for how long. they should be kept in the BL

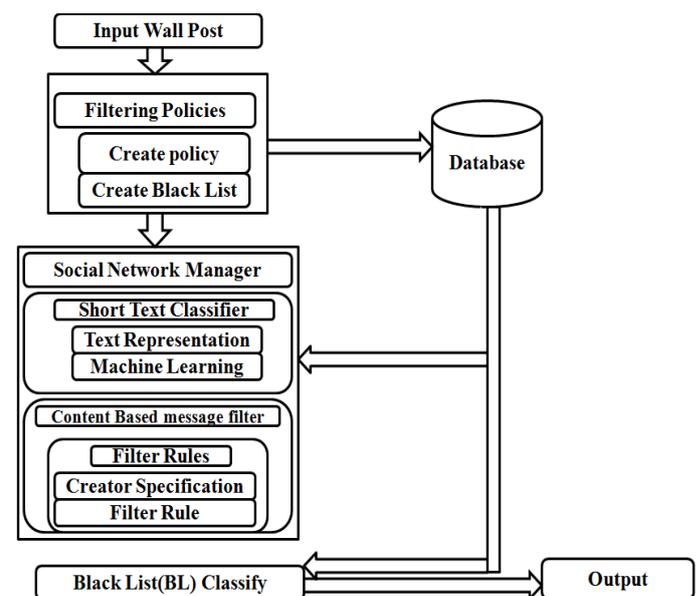


Fig 1. System Architecture

The way followed by a message, from its writing to the other wall/space in final publication can be summarized as follows:

- The user tries to post a message, which is intercepted by Filtered Wall.
- A ML-based text classifier extracts metadata from the content of the message.
- Filtered Wall uses metadata provided by the classifier, together with data extracted from the social graph and users' profiles, to enforce the filtering and BL rules.
- Depending on the result of the previous step, the message will be published or filtered by FW.

### Short text Classifier

The task of semantically categorizing short texts is conceived in our approach as a multi-class soft classification process composed of two main phases: text representation and Machine Learning-based classification.

### Text Classification

The extraction of an appropriate set of features by which representing the text of a given document is a crucial task strongly affecting the performance of the overall classification strategy. Different sets of features for text categorization have been proposed in the literature (Dumais *et al.*, 1998), however the most appropriate feature types and feature representation for short text messages have not been sufficiently investigated. Proceeding from these considerations and basing on our experience documented (Lewis, 1992). We consider two types of features, Bag of Words (BoW) and Document properties (Dp), that are used in the experimental evaluation to determine the combination that is most appropriate for short message classification. We introduce CF modelling information that characterizes the environment where the user is posting.

- Correct words: express the amount of terms  $t_k \in T \setminus K$  where  $t_k$  is a term of the considered document  $d_j$  and  $K$  is a set of known words for the domain language.
- Bad words: are computed similarly to the correct words feature, whereas the set  $K$  is a collection of "dirty words" for the domain language.
- Capital words: express the amount of words mostly written with capital letters, calculated as the percentage of words within the message, having more than half of the characters in capital case. For example the value of the feature for the document "To be OR NOt to BE" is 0.5 since the words "OR" "NOt" and "BE" are considered as capitalized.
- Punctuations characters: calculated as the percentage of the punctuation characters over the total number of characters in the message. For example the value of the feature for the document "Hello!!! How're u doing?" is 5/24.
- Exclamation marks: calculated as the percentage of exclamation marks over the total number of punctuation characters in the message. Referring to the aforementioned document the feature value is 3/5.
- Question marks: calculated as the percentage of question marks over the total number of punctuations characters in

the message. Referring to the aforementioned document the feature value is 1/5.

### Machine Learning Based classification

We address the short text categorization as a hierarchical two-level classification process. The first-level classifier performs a binary hard categorization that labels messages as Neutral and Non-Neutral. The first-level filtering task facilitates the subsequent second-level task in which a finer-grained classification is performed. The second-level classifier performs a soft-partition of Non-neutral messages assigning with a given message a gradual membership to each of the non-neutral classes. The first and second-level classifiers are then structured as regular RBFN, conceived as hard and soft classifier respectively. Its non-linear function maps the feature space to the categories space as a result of the learning phase on the given training set constituted by manually classified messages.

In the second level of the classification stage, we introduce a modification of the standard use of RBFN. Its regular use in classification includes a hard decision on the output values: according to the winner-take-all rule, a given input pattern is assigned with the class corresponding to the winner output neuron which has the highest value. In our approach, we consider all values of the output neurons as a result of the classification task and we interpret them as gradual estimation of multimembership to classes. ML-based classifier needs to be trained with a set of sufficiently complete and consistent preclassified data. The difficulty of satisfying this constraint is essentially related to the subjective character of the interpretation process with which an expert decides whether to classify a document under a given category.

### Filtering Rules and Blacklist Management

In this section, we introduce the rules adopted for filtering unwanted messages. In defining the language for filtering rules specification, we consider three main issues that, in our opinion, should affect the filtering decision. The first aspect is related to the fact that, in OSNs like in everyday life, the same message may have different meanings and relevance based on who writes it. As a consequence, filtering rules should allow users to state constraints on message creators. Thus, creators on which a filtering rule applies should be selected on the basis of several different criteria; one of the most relevant is by imposing conditions on user profile's attributes. In such a way it is, for instance, possible to define rules applying only to young creators, to creators with a given religious/ political view, or to creators that we believe are not expert in a given field (e.g. by posing constraints on the work attribute of user profile).

Another relevant issue to be taken into account in defining a language for filtering rules specification is the support for content-based rules. This means filtering rules identifying messages according to constraints on their contents. More precisely, the idea is to exploit classes of the first and second level as well as their corresponding membership levels to make users able to state content-based constraints.

Another issue we believe it is worth being considered is related to the difficulties an average OSN user may have in defining the correct threshold for the membership level. To make the user more comfortable in specifying the membership level threshold, we believe it would be useful allowing the specification of a tolerance value that, associated with each basic constraint, specifies how much the membership level can be lower than the membership threshold given in the constraint.

### Filtering Rules

#### Definition 1 (Creator specification)

A creator specification creator Spec implicitly denotes a set of OSN users. It can have one of the following forms, possibly combined:

- A set of attribute constraints of the form  $an \text{ OP } av$ , where  $an$  is a user profile attribute name,  $av$  and  $OP$  are, respectively, a profile attribute value and a comparison operator, compatible with an domain.
- A set of relationship constraints of the form  $(m; rt; \text{min Depth}; \text{max Trust})$ , denoting all the OSN users participating with user  $m$  in a relationship of type  $rt$ , having a depth greater than or equal to  $\text{min Depth}$ , and a trust value less than or equal to  $\text{max Trust}$ .

#### Definition 2 (Filtering rule)

A filtering rule FR is a tuple (author, creator Spec, contentSpec, action)

- Where author is the user who specifies the rule;
- creatorSpec is a creator specification, specified according to Definition 1;
- ContentSpec is a Boolean expression defined on content constraints of the form  $(c, ml)$ , where  $C$  is a class of the first or second level and  $ml$  is the minimum membership level threshold required for class  $C$  to make the constraint satisfied;
- Action  $\in \{\text{block notify}\}$  denotes the action to be performed by the system on the messages matching contentSpec and created by users identified by creatorSpec.

### Online Setup Assistant for FRs Thresholds

We address the problem of setting thresholds to filter rules, by conceiving and implementing within FW, an Online Setup Assistant procedure. OSA presents the user with a set of messages selected from the data set. For each message, the user tells the system the decision to accept or reject the message. The collection and processing of user decisions on an adequate set of messages distributed over all the classes allows computing customized thresholds representing the user attitude in accepting or rejecting certain contents. Such messages are selected according to the following process. A certain amount of non neutral messages taken from a fraction of the data set and not belonging to the training/test sets, are classified by the ML in order to have, for each message, the second-level class membership values

### Example

These filtering criteria can be easily specified through the following FRs:

- ((Bob; friend Of; 2; 1), (Vulgar; 0:80), block)
- ((Bob; friend Of; 1; 0:5), (Vulgar; 0:80), block)

Eve, a friend of Bob with a trust value of 0.6, wants to publish the message “G\*d d\*mn f\*ck\*ng s\*n of a b\*tch!” on Bob’s FW. After posting the message, receives it in input producing the grade of membership 0.85 for the class Vulgar. Therefore, the message, having a too high degree of vulgarity, will be filtered from the system and will not appear on the FW.

### Blacklists

A further component of our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs is directly managed by the system, which should be able to determine who are the users to be inserted in the BL and decide when users retention in the BL is finished. To enhance flexibility, such information are given to the system through a set of rules, hereafter called BL rules.

In contrast, to catch new bad behaviours, we use the Relative Frequency (RF) that let the system be able to detect those users whose messages continue to fail the FRs. The two measures can be computed either locally, that is, by considering only the messages and/or the BL of the user specifying the BL rule or globally, that is, by considering all OSN users walls and/or BLs. A BL rule is therefore formally defined as follows:

#### Definition 3 (BL rule)

A BL rule is a tuple (author, creatorSpec, creatorBehavior, T) where

- Author is the OSN user who specifies the rule, i.e., the wall owner;
- CreatorSpec is a creator specification, specified according to Definition 1;

Creator Behavior consists of two components RF Blocked and min Banned. RF Blocked = (RF, mode, window) is defined such that

- $RF = \frac{\#bMessages}{\#tMessages}$ , where  $\#tMessages$  is the total number of messages that each OSN user identified by creatorSpec has tried to publish in the author wall (mode = my Wall) or in all the OSN walls (mode = SN); whereas  $\#bMessages$  is the number of messages among those in  $\#tMessages$  that have been blocked;
- Window is the time interval of creation of those messages that have to be considered for RF computation; minBanned = (min, mode, window), where min is the minimum number of times in the time interval specified in window that OSN users identified by creator Spec have to be inserted into the BL due to BL rules specified by author wall (mode = my Wall) or all OSN users (mode =SN) in order to satisfy the constraint.

- T denotes the time period the users identified by CreatorSpec and creator Behavior have to be banned from author wall.

**Example:** The BL rule

(Alice,(Age < 16), (0:5,myWall,1 week), 3 days) inserts into the BL associated with Alice's wall those young users (i.e., with age less than 16) that in the last week have a relative frequency of blocked messages on Alice's wall greater than or equal to 0.5.

## Evaluation

### Problem and Data Set Description

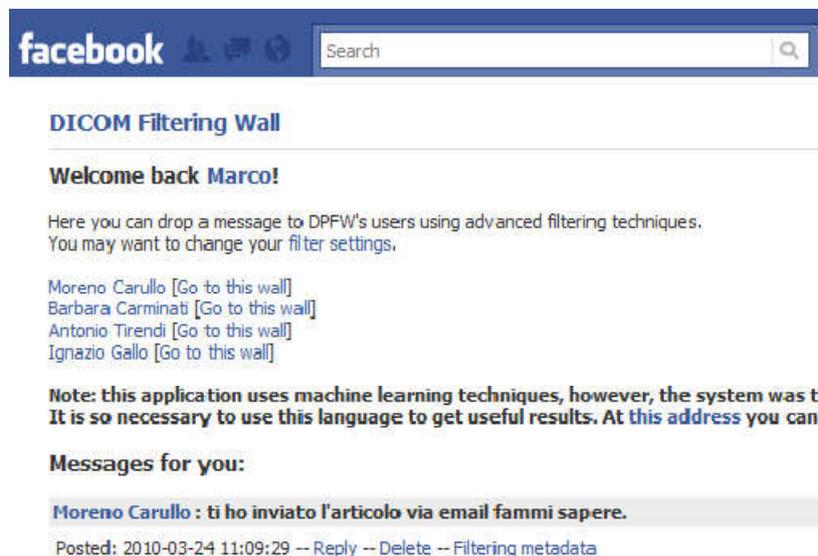
The set of classes considered in our experiments is  $\Omega = \{\text{Neutral, Violence, Vulgar, Offensive, Hate, Sex}\}$  where  $\Omega = \{\text{Neutral}\}$  are the second-level classes. The percentage of elements in D that belongs to the Neutral class is 31 percent.

## Short Text Classifier Evaluation

Two different types of measures will be used to evaluate the effectiveness of first-level and second-level classifications. In the first level, the short text classification procedure is on the basis of the contingency table approach. In particular, the derived well-known Overall Accuracy (OA) index capturing the simple percent agreement between truth and classification result.

At second level, we adopt measures widely accepted in the Information Retrieval and Document Analysis field, that is, Precision (P) that permits to evaluate the number of false positives, Recall (R), that permits to evaluate the number of false negatives, and the overall metric F-Measure defined as the harmonic mean between the above two indexes (Information Retrieval, 1992). The F-Measure is commonly defined in terms of a coefficient  $\beta$  that defines how much to favour Recall over Precision. We chose to set  $\beta = 1$ .

## Expected Output



**DICOM Filtering Wall**

Welcome back Marco!

Here you can drop a message to DPFW's users using advanced filtering techniques. You may want to change your filter settings.

Moreno Carullo [Go to this wall]  
Barbara Carminati [Go to this wall]  
Antonio Tirendi [Go to this wall]  
Ignazio Gallo [Go to this wall]

**Note: this application uses machine learning techniques, however, the system was trained on a dataset of Italian text. It is so necessary to use this language to get useful results. At this address you can find more information.**

**Messages for you:**

**Moreno Carullo : ti ho inviato l'articolo via email fammi sapere.**  
Posted: 2010-03-24 11:09:29 -- Reply -- Delete -- Filtering metadata

(a)



**DICOM Filtering Wall**

Marco Vanetti's wall

Your message can't be posted because it was filtered!

Non-Neutral = 1  
Neutral = 0  
Violence = 0.316783  
Vulgar = 1  
Offensive = 0.221593  
Hate = 0  
Sex = 0.601016

**Marco Vanetti : Cazzabuffo!!**  
**!! Suspicious message !!** - Posted: 2010-04-16 05:36:29 -- Reply -- Delete -- Filtering metadata

**Moreno Carullo : ti ho inviato l'articolo via email fammi sapere.**  
Posted: 2010-03-24 11:09:29 -- Reply -- Delete -- Filtering metadata

(b)

Fig. 2. Two relevant use cases of the DicomFW application. Start page proposes the list of walls the OSN user can see (a). A message filtered by the wall's owner filtering rules (b)

## Overall Performance and Discussion

The Precision and the Recall value computed for FRs with (Neutral, 0.5) content constraint. The Precision and the Recall value computed for FRs with (Vulgar, 0.5) constraint. Results achieved by the content-based specification component, on the first-level classification, can be considered good enough and reasonably aligned with those obtained by well-known information filtering techniques (51). The contextual information (CF) significantly improves the ability of the classifier to correctly distinguish between non neutral classes. This result makes more reliable all policies exploiting non neutral classes

To summarize, our application (Fig. 2) permits to

1. View the list of users' FWs; (see fig.2 (a))
2. View messages and post a new one on a FW;
3. Define FRs using the OSA tool. When a user tries to post a message on a wall, he/she receives an alerting message (see Fig. 2(b)) if it is blocked by FW.

## Conclusion

In this paper, we have presented a system to filter out undesired messages from OSN walls. The system exploits a ML soft classifier to enforce customizable content-dependent filtering rules. Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of BLs. For instance, the system can automatically take a decision about the messages blocked because of the tolerance, on the basis of some statistical data (e.g., number of blocked messages from the same author, number of times the creator has been inserted in the BL) as well as data on creator profile (e.g., relationships with the wall owner, age, sex). The development of a GUI to make easier BL and filtering rule specification is also a direction we plan to investigate. As future work, we intend to exploit similar techniques to infer BL and filtering rules.

## REFERENCES

- Belkin, N.J. and Croft, W.B. 1992. "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" *Comm. ACM*, Vol. 35, No. 12, pp. 29-38.
- Bonchi, F. and Ferrari, E. *Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques*. Chapman and Hall/CRC Press, 2010.
- Dumais, S., Platt, J., Heckerman, D., and Sahami, M. 1998. "Inductive Learning Algorithms and Representations for Text Categorization," *Proc. Seventh Int'l Conf. Information and Knowledge Management (CIKM '98)*, pp. 148-155.
- Hanani, U., Shapira, B. and Shoval, P. 2001. "Information Filtering: Overview of Issues, Research and Systems," *User Modeling and User-Adapted Interaction*, Vol. 11, pp. 203-259.
- Information Retrieval: Data Structures & Algorithms*, W.B. Frakes and R.A. Baeza-Yates, eds., Prentice-Hall, 1992.
- Lewis, D.D. 1992. "An Evaluation of Phrasal and Clustered Representations on a Text Categorization Task," *Proc. 15th ACM Int'l Conf. Research and Development in Information Retrieval (SIGIR '92)*, N.J. Belkin, P. Ingwersen, and A.M. Pejtersen, eds., pp. 37-50.
- Sebastiani, F. 2002. "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, Vol. 34, no. 1, pp. 1-47.
- Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H. and Demirbas, M. 2010. "Short Text Classification in Twitter to Improve Information Filtering," *Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '10)*, pp. 841-842.
- Vanetti, M., Binaghi, E., Carminati, B., Carullo, M. and Ferrari, E. 2010. "Content-Based Filtering in On-Line Social Networks," *Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10)*.

\*\*\*\*\*