IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 3, Ver. IV (May-Jun. 2016), PP 116-124
www.iosrjournals.org

Machine Learning techniques for filtering of unwanted messages in Online Social Networks

Pinniboyina Venkateswarlu[1], R Lakshmi Tulasi[2]
#1 Student of M.Tech (CSE) and Department of Computer Science & Engineering,
#2 Prof, Head of the Department in Computer Science & Engineering, QIS Instititute Of Technology, Ongole, AP, India.

Abstract: *As of Recent Years, Online Social Networks have transformed into a key bit of step by step life. One key issue in today user wall(s) is to give users the ability to control the messages posted in solitude private space to keep up a vital separation from that undesirable substance is appeared. Up to now userwalls give small support to this need. To fill the fissure, I propose a structure allowing user divider users to have a prompt control on the messages posted on their walls. This is refined through a versatile principle based structure, that allows users to change the filtering criteria to be associated with their walls, and Machine Learning based sensitive classifier actually checking messages in moving of substance based isolating. T-OSN expect a urgent part in regular life. User can relate with other user by sharing a couple sorts of substance such as picture, sound and video substance. Main problem in OSN (Online Social Network) is to hindering security in posting undesirable messages. Ability to have a prompt control over the messages posted on user divider is not gave. Undesirable post will be particularly posted on general society divider. Simply the undesirable messages will be blocked not the user. To keep up a vital separation from this issue, BL (Black List) part is proposed in this paper, which avoid undesired producers messages. BL is used to make sense of which user should be inserted in BL and pick when the support of the user is finished. Machine Learning Text Categorization is in like manner used to arrange the short texts.*

Keywords: *Online Social Networks (OSN), Blacklist, online informal organization, Machine learning content arrangement, short content order.*

I. Introduction

Information and correspondence development expect a tremendous part in today's organized society. It has affected the online relationship between users, who think about security applications and their recommendations on individual insurance. There is a need to develop more security frameworks for different correspondence advancements, particularly online casual groups. OSNs offer alongside no sponsorship to foresee undesirable messages on userwalls. With the nonattendance of plan or filtering gadgets, the user gets all messages posted by the users he takes after. Generally speaking, the user get a noisy stream of upgrades. In this paper, an information Filtering structure is introduced. The system focuses on one kind of supports: Lists which are a physically picked assembling of users on OSN. Rundown sustains tend to be fixated on specific subjects, despite it is still uproarious in view of insignificant messages. In this way, we propose a web filtering system, which isolates the focuses in a summary, filtering through unessential messages [1]. In OSNs, information filtering can moreover be used for a substitute, more fragile, reason. This is a direct result of the route that in OSNs there is the probability of posting or commenting diverse posts on particular open/private areas, brought generally speaking walls. In the proposed structure Information isolating can thus be used to give users the ability to normally control the messages formed in solitude walls, by filtering through undesirable messages. The purpose of the present work is along these lines to propose and likely survey a motorized system, called Filtered Wall (FW), prepared to channel undesirable messages from OSN userwalls. We enterprise Machine Learning (ML) content request frameworks [2] to actually dole out with each short text a game plan of groupings considering its substance. The noteworthy attempts in building an incredible short substance classifier are amassed in the extraction and decision of a game plan of depicting and discriminant highlights. Regardless, the purpose of most of these recommendation is generally to give users a game plan framework to keep up a vital separation from they are overwhelmed by worthless data. In OSNs, information isolating can moreover be used for a substitute, more fragile, reason. This is a direct result of the path that in OSNs there is the probability of posting or commenting distinctive posts on particular open/private regions, brought all around walls. Information filtering can thusly be used to give users the ability to normally control the messages made in solitude walls, by filtering through undesirable messages. We believe this is a key OSN organization that has not been given all things considered. Without a doubt, today OSNs give no sponsorship to stay away from undesirable messages on userwalls. Case in point, Facebook grants users to state why ought to allowed expansion messages in their walls (i.e., buddies, partners of sidekicks, or portrayed get-togethers of colleagues).

Regardless, no substance based slants are reinforced and subsequently it is impossible to keep away from undesired messages, for instance, political or profane ones, paying little heed to the user who posts them. Giving this organization is not simply a question of using heretofore described web substance burrowing techniques for a substitute application, rather it requires to blueprint off the cuff request frameworks.

II. Related Work

M. Chau and H. Chen [2] depicts as the Web continues creating, it has ended up being dynamically difficult to chase down related information using traditional web lists. Subject specific web records give an alternative way to deal with support giving so as to compel information recuperation on the Web more correct and revamp looking in changed spaces. Regardless, planners of topic specific web look instruments need to address two issues: how to discover relevant files (URLs) on the Web and how to filter through unessential reports from a game plan of records assembled from the Web. This paper reports our investigation in tending to the second issue. We propose a machine-learning-based system that solidifies Web content examination and Web structure examination. We address each Web page by a game plan of substance based and association based segments, which can be used as the data for various machine learning figurings. The proposed technique was realized using both a nourishment forward/back inciting neural framework and a support vector machine. Two examinations were made and coordinated to differentiate the proposed Web-highlight procedure and two existing Web page isolating systems - a watchword based technique and a word reference based strategy. The exploratory results showed that the proposed approach with everything taken into account performed better than the benchmark approaches, especially when the amount of planning records was little. The proposed philosophies can be associated in point specific web crawler change and other Web applications, for instance, Web content organization. R.J. Mooney and L. Roy depict [3] Recommender systems upgrade access to material things and information by making tweaked proposition in perspective of past outlines of a user's inclinations and revultions. Most existing recommender structures use social isolating frameworks that develop recommendations as for other users' slants. By separation, substance based procedures use information around a thing itself to make proposals. This procedure has the advantage of having the ability to endorse as of now unrated things to users with excellent interests and to offer elucidations to its recommendations. We portray a substance based book recommending structure that uses information extraction and a machine-learning figuring for substance course of action. Beginning test outcomes demonstrate this philosophy can convey exact proposals. These examinations rely on upon evaluations from discretionary samplings of things and we discuss issues with past tests that use skewed examples of user picked cases to survey execution. F. Sebastiani depicts The automated categorization[4] (or course of action) of compositions into predefined classes has seen an impacting eagerness for the latest ten years, as a result of the extended openness of files in cutting edge structure and the accompanying need to create them. In the examination bunch the transcendent approach to manage this issue relies on upon machine learning frameworks: a general inductive process actually produces a classifier by learning, from a course of action of pre-requested records, the characteristics of the orders. The advantages of this philosophy over the learning outlining system (involving in the manual importance of a classifier by space experts) are a not too bad reasonability, huge speculation stores similarly as expert work power, and clear conservativeness to different zones. This survey discusses the standard approaches to manage content characterization that fall within the machine learning perspective. We will discuss in unobtrusive component issues identifying with three particular issues, particularly document representation, classifier improvement, and classifier appraisal. M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari [5] this paper proposes a system approving substance build message filtering for as for line Social Networks (OSNs). The structure licenses OSN users to have a quick control on the messages posted on their walls. This is expert through a versatile fundamental based structure, that allows a user to re-try the filtering criteria to be associated with their walls, and a Machine Learning based fragile classifier thus checking messages in maneuvering of substance based isolating.

III. Problem Definition

In the present OSN frameworks obstructing of user is for lifetime. We beat this Problem by utilizing Proposed System. In our framework we plan to hinder the user for specific time period furthermore send notice to them who posted on divider. The use of substance construct separating in light of messages posted on OSN userwalls postures further difficulties given the short length of those messages separated from the wide scope of subjects that might be said. Short content arrangement has gotten up to presently little consideration. Late work highlights trouble in molding solid choices, essentially as a consequence of the very certainty that the depiction of the short content is delicate, with a few incorrect spellings, non-standard lexis. Our work is moreover inspire by the different gets to administration models and associated approach dialects and social control instruments that are anticipated to date for OSNs since separating offers numerous similitudes with access administration.

IV. Content-Based Filtering

Data sifting frameworks are intended to characterize a flood of progressively created data dispatched no concurrently by a data maker and present to the user those data that are prone to fulfill his/her prerequisites [3]. In substance based sifting every user is accepted to work autonomously. Thus, a substance based separating framework chooses data things in view of the relationship between the substance of the things and the user inclinations instead of a communitarian sifting framework that picks things in light of the connection between kin with comparable inclinations [4]. While electronic mail was the first space of ahead of schedule work on data separating, consequent papers have tended to enhanced areas including newswire articles, Internet "news" articles, and more extensive system assets [5], [6]. Reports prepared in substance based sifting are for the most part literary in nature and this makes content-based separating near content grouping. The movement of separating can be demonstrated, truth be told, as an instance of single mark, twofold arrangement, dividing approaching records into significant and non-applicable classifications [7]. More mind boggling separating frameworks incorporate multi-mark content classification naturally naming messages into fractional topical classifications. In [4] a point by point examination investigation has been directed affirming predominance of Boosting-based classifiers, Neural Networks and Support Vector Machines over other prominent strategies, for example, Rocchio and Naive Bayesian. Notwithstanding, it is worth to note that the vast majority of the business related to message sifting by ML has been connected for long-frame content and the evaluated execution of the content order techniques entirely relies on upon the way of printed reports.

V. Policy-Based Personalization of OSN Contents

There have been a few proposition misusing grouping components for customizing access in OSNs. For example, in [8] an arrangement strategy has been proposed to classify short instant messages with a specific end goal to abstain from overpowering users of microblogging administrations by crude information. The user can then view just certain sorts of tweets in light of his/her advantage. Interestingly, Golbeck and Kuter [2] propose an application, called FilmTrust, that adventures OSN trust connections and provenance data to customize access to the site. In any case, such frameworks don't give a sifting approach layer by which the user can abuse the aftereffect of the grouping procedure to choose how and to which degree sifting through undesirable data. Conversely, our sifting approach dialect permits the setting of FRs as per an assortment of criteria, that don't consider just the consequences of the arrangement prepare additionally the connections of the divider proprietor with other OSN users and also data on the user profile. Additionally, our framework is supplemented by an adaptable component for BL administration that gives a further chance of customization to the sifting strategy. The methodology embraced by MyWOT is very diverse. Specifically, it bolsters separating criteria which are far less adaptable than the ones of Filtered Wall. Content separating can be considered as an expansion of access control, since it can be utilized both to shield objects from unapproved subjects, and subjects from wrong questions. In the field of OSNs, the larger part of access control models proposed so far implement topology-based access control, as indicated by which get to control necessities are communicated regarding connections that the requester ought to have with the asset proprietor. We utilize a comparable thought to distinguish the users to which a FR applies. Notwithstanding, our sifting arrangement dialect develops the dialects proposed for access control approach particular in OSNs to adapt to the amplified necessities of the separating area. To be sure, since we are managing separating of undesirable substance as opposed to with access control, one of the key elements of our framework is the accessibility of a portrayal for the message substance to be abused by the sifting instrument. Conversely, nobody of the entrance control models already refered to misuse the substance of the assets to authorize access control. In addition, the thought of BLs and their administration are not considered by any of the aforementioned access control models. At last, our strategy dialect has a few associations with the approach structures that have been so far proposed to bolster the particular and requirement of arrangements communicated as far as imperatives on the machine justifiable asset portrayals gave by Semantic web dialects. Illustrations of such structures are KAoS and REI, concentrating for the most part on access control, Protune [3], which gives bolster additionally to trust arrangement and protection approaches, and WIQA [4], which gives end users the capacity of utilizing sifting strategies as a part of request to signify given "quality" prerequisites that web assets must fulfill to be shown to the users. Be that as it may, albeit such structures are effective and sufficiently general to be redone and/or stretched out for various application situations they have not been particularly imagined to address data separating in OSNs and thusly to consider the user social diagram in the approach determination process.

VI. Algorithmic Strategy

These are the 3 algorithms used in proposed system:
1. Short Text Algorithm
2. Stemmer Algorithm
3. Stop Word Algorithm

4. clustering Algorithms

1. Short Text Algorithm:

Step 1: For any word of the short text, inquire the word pairs sets.
if there exists a record of a similar concept relationship,
go to step 5
else
if there exists a record of different concept
relationships,
go to step 2 else go to step 9;
Step 2: If there is only one word-pair Ti-tj, go to step 4, e
lse
if there are several word-pairs,
go to step 3;
Step 3: Extract the right words of all the word-pairs related to ti and form into Tx,
if tj TX,
and
tj can be found in the vector space of this short text,
go to step 7, else
extract tj , the right word of the word pair with the
highest strength and go to step 4;
Step 4: Extract tj , the right word of this word-pair,
if tj cannot be found in the vector space of this short text,
go to step 6, else go to step 7;
Step 5: Extract the word set TY in the text,
if there exist tk TY and tk TZ
（attribute set of the word pair（ti,tj））,
go to step 8; else
go to step 10;
Step 6: Calculate the mutual information between tj
and other words in the text,
and
go to step 8 when meeting the requirements,
else go to step 10;
Step 7: Calculate the mutual information between tj
and other words in the text, and
go to step 9 when meeting the requirements,
else go to step 10;
Step 8: Insert tj into the vector space of this short text;
Step 9: Raise the frequency of tj in the vector space of this
Text at λ. ($0<\lambda<1$);
Step 10: Don't extend this word, and input and seek the next word.

2. Stemmer Algorithm

Step 1: Gets rid of plurals and -ed or -ing suffixes
Step 2: Turns terminal y to i when there is another
Vowel in the stem
Step 3: Maps double suffixes to single ones:
-ization, -ational, etc.
Step 4: Deals with suffixes, -full, -ness etc.
Step 5: Takes off -ant, -ence, etc.
Step 6: Removes a final -e
3. Stop words Algorithm
Step 1: $T' \leftarrow Ø$
Step 2: W= the set of all words in the domain
Step 3: $D=T \cap W$
Step 4: Pick word w D
Step 5: For each interface q QI
Step 6: Remove w from the labels of interface q

Step 7: Check the stop word constraints for the labels of sibling nodes
Step 8: if no stop word constraint is violated then
T'←T' U {ω};
Else remove antonyms of w appearing in D from T'
Step 9: goto 4 Step 10: return T'; .

4. Clustering Algorithms

In choosing a subset of clustering algorithms for our study, we explored algorithms with different input information. *Network structure* is the most common input information used by clustering algorithms. This information represents people as nodes and their friendship relations as links. Algorithms with this input information are called structurebased (or structural) algorithms. Other algorithms, called feature-based algorithms, use *nodes' features and attributes* to detect groups. For example, these features can be age, gender, and education of people in OSNs. A third category combines these two inputs *network structure and nodes' features*. In this paper, we focus on structure-based clustering algorithms. One advantage is the ability to interpret why the resulting groupings emerged and to compare algorithms with a consistent evaluation metric across the same network structure [10]. Furthermore, using feature-based algorithms necessitates extracting extra data from an OSN. This extraction results in very high processing time which makes conducting studies in a limited time in the lab difficult or almost impossible.

Structured-based clustering algorithms can be further classified in to three categories based on their membership attribute: (i) '*disjoint clustering*' algorithms where each object can only belong to one group; (ii) '*overlapping clustering*' algorithms where an object can be a member of more than one group. For example, a person may belong to different groups such as 'Family', 'Main East High School', and 'Loves Red Sox'; and (iii) '*hierarchical clustering*' algorithms which categorize objects in a multi-level structure where one group can be a subset of another group [22]. For example, cousin Joe is in a group labelled 'Cousins' which is a subset of a group named 'Family'. '*Hierarchical clustering*' algorithms have been used widely in social network analysis [10]. Figure 1 shows a schematic view of these clustering algorithms based on the defined membership attributes. We chose a representative algorithm from each membership category explained above for a total of three algorithms:

- *Markov Clustering (MCL)*: This algorithm is a *disjoint clustering* algorithm that uses the concept of Markov chains to simulate stochastic flows in graphs and builds a fast and scalable unsupervised clustering algorithm. MCL has a relatively high performance and is scalable [26].
- *OSLOM*: The Order Statistics Local Optimization Method (OSLOM) is an *overlapping clustering* algorithm that is among the first to account for edge weights and overlapping groups. It has a high performance and is scalable to large networks
- *Louvain*: This *hierarchical clustering* algorithm uses modularity as its objective function and maximizes it using multiple heuristics to detect the groups. While this algorithm finds groups in a hierarchical manner, the lowest level of the hierarchies, which are the subgroups, are disjoint; i.e. one person cannot be a member of more than one group in a same level. The Louvain algorithm is highly accurate and has a very low computation time which makes it appealing for our study [5].

(a) Disjoint Clustering (b) Overlapping Clustering (c) Hierarchical Clustering

Fig.1: Three clustering methods with different membership attributes

VII. Machine Learning Based Classification:

It is said that short content classifier incorporate progressive two level order process. To begin with level classifier execute a twofold hard classification that mark message as unbiased and non-nonpartisan. The primary level separating errand help the succeeding second level assignment in which a better grained arrangement is finished. The second level classifier will do the delicate allotment of non-unbiased messages. Among the assortment of models, RBFN model is chosen. RBFN contain a solitary concealed layer of handling units. Usually utilized capacity is Gaussian capacity. Order capacity is nonlinear, which is the benefit of RBFN. Potential over preparing affectability and potential affectability to info parameters are the downsides.

Architecture Of Proposed System:

Architecture of the proposed system includes filtering rules and blacklist. The whole process will be visible clearly in Architecture. Message will be labeled based on the content, so classification will be over. Then the filtration part, which is done by filtering rules. Analysis of Creating the specification will be done. Finally probability value is calculated and the user who post the unwanted message will be kept in Blacklist. So that the user will be temporarily blocked. Advantage of our proposed System is to have a direct control over the user wall.

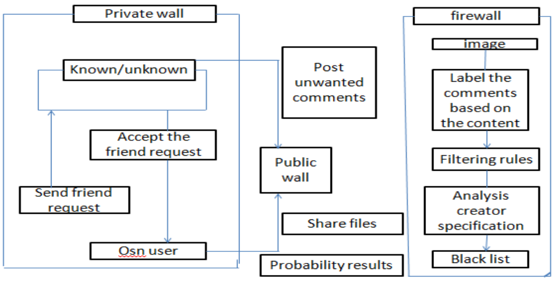

Fig: 2 Architecture diagram

As a result, FR should allow the user to restrict the message creators. Here the type, depth, and trust value are recognized by creator Specification.

Definition1 (Creator specification) A Creator Specification CreaSpec, which denotes a set of OSN users. Possible combinations are 1.Set of attributes in the An OP Av form, where An is a user profile attribute name, Av is profile attribute value and OP is a comparison 2. Set of relationship of the form (n, Rt, minDepth, maxTrust) indicate OSN users participating with user n in a relationship of type Rt, depth greater than or equal to minDepth, trust value greater than or equal to maxTrust.

Definition 2 (Filtering rule) A filtering rule is a tuple (auth,CreaSpec,ConSpec,action) 1. auth is the user who state the rule. 2. CreaSpec is the Creator specification. 3. ConSpec is a boolean expression. 4. Action is the action performed by the system. Filtering rules will be applied, when a user profile does not hold value for attributes submitted by a FR. This type of situation will dealt with asking the owner to choose whether to block or notify the messages initiating from the profile which does not match with the wall owners FRs, due to missing of attributes.

Blacklist:

The main implementation of our paper is to execute the Blacklist Mechanism, which will keep away messages from undesired creators. BL are handled undeviating by the system. This will able to decide the users to be inserted in the blacklist. And it also decide the user preservation in the BL will get over. Set of rules are applied to improve the stiffness, such rules are called BL rules. By applying the BL rule, owner can identify which user should be blocked based on the relationship in OSN and the user's profile. The user may have bad opinion about the users can be banned for an uncertain time period. We have two information based on bad attitude of user. Two principle are stated. First one is within a given time period user will be inserted in BL for numerous times, he /she must be worthy for staying in BL for another sometime. This principle will be applied to user who inserted in BL at least once. Relative Frequency is used to find out the system, who messages continue to fail the FR. Two measures can be calculated globally and locally, which will consider only the message in local and in global it will consider all the OSN users walls.

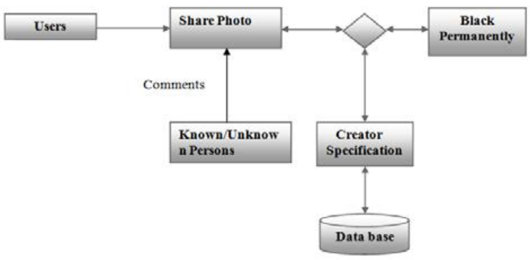

Fig 3: blacklist system

Tightly clustered fringe

Next, we consider the graph properties at the scale of local neighborhoods outside of the core. We first examine clustering, which quantifies how densely the neighborhood of a node is connected.

The *clustering coefficient* of a node with N neighbors is defined as the number of directed links that exist between the node's N neighbors, divided by the number of possible directed links that could exist between the node's neighbors ($N(N-1)$). The clustering coefficient of a graph is the average clustering coefficient of all its nodes, and we denote it as C.

Table 1 shows the clustering coefficients for all four social networks. For comparison, we show the ratio of the observed clustering coefficient to that of Erdos-R´eyni (ER) random¨ graphs and random power-law graphs constructed with preferential attachment, with the same number of nodes and links. Hence, they provide a point of reference for the degree of local clustering in the social networks. Graphs constructed using preferential attachment also have no locality bias, as preferential attachment is a global process, and they provide a point of reference to the clustering in a graph with a similar degree distribution.

Table 1: The observed clustering coefficient, and ratio to random Erdos-R´eyni graphs as well as random¨ power-law graphs

Network		**Ratio to Random Graphs**	
	C	**Erdos-R´enyi¨**	**Power-Law**
Web [2]	0.081	7.71	-
Flickr	0.313	47,200	25.2
LiveJournal	0.330	119,000	17.8
Orkut	0.171	7,240	5.27
YouTube	0.136	36,900	69.4

The clustering coefficients of social networks are between three and five orders of magnitude larger than their corresponding random graphs, and about one order of magnitude larger than random power-law graphs. This unusually high clustering coefficient suggests the presence of strong local clustering, and has a natural explanation in social networks: people tend to be introduced to other people via mutual friends, increasing the probability that two friends of a single user are also friends.

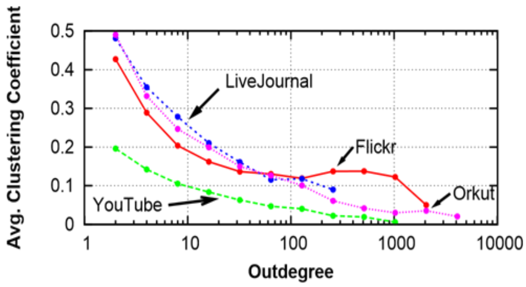

Figure 4: Clustering coefficient of users with different outdegrees. The users with few "friends" are tightly clustered.

Figure 4 shows how the clustering coefficients of nodes vary with node outdegree. The clustering coefficient is higher for nodes of low degree, suggesting that there is significant clustering among low-degree nodes. This clustering and the small diameter of these networks qualifies these graphs as small-world networks [52], and further indicates that the graph has scale-free properties.

Groups

In many online social networks, users with shared interests may create and join groups. Table 5 shows the high-level statistics of user groups in the four networks we study. Participation in user groups varies significantly across the different networks: only 8% of YouTube users but 61% of LiveJournal users declare group affiliations. Once again, the group sizes follow a power-law distribution, in which the vast majority have only a few users each.

Note that users in a group need not necessarily link to each other in the social network graph. As it turns out, however, user groups represent tightly clustered communities of users in the social network. This can be seen from the average group clustering coefficients of group members,

Network	**Groups**	**Usage**	**Avg. Size**	**Avg.** C
Flickr	103,648	21%	82	0.47
LiveJournal	7,489,073	61%	15	0.81
Orkut	8,730,859	13%	37	0.52
YouTube	30,087	8%	10	0.34

Table 2: Table of the high-level properties of network groups including the fraction of users which use group features, average group size, and average group clustering coefficient

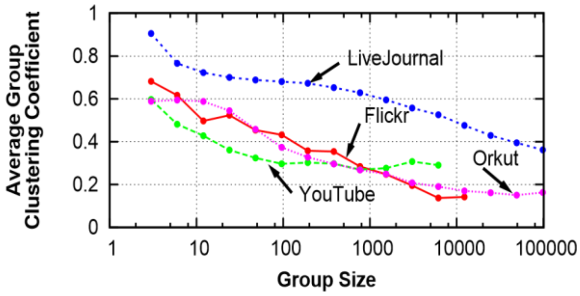

Figure 5: Plot of group size and average group clustering coefficient. Many small groups are almost cliques.

Finally, Figure 6 shows how user participation in groups varies with outdegree. Low-degree nodes tend to be part of very few communities, while high-degree nodes tend to be members of multiple groups. This implies a correlation between the link creation activity and the group participation. There is a sharp decline in group participation for Orkut users with over 500 links, which is inconsistent with the behavior of the other networks. This result may be an artifact of our partial crawl of the Orkut network and the resulting biased user sample.

In general, our observations suggest a global social network structure that is comprised of a large number of small, tightly clustered local user communities held together by nodes of high degree. This structure is likely to significantly impact techniques, algorithms and applications of social networks.

Summary

We end this section with a brief summary of important structural properties of social networks which we observed in our data.

- The degree distributions in social networks follow a power-law, and the power-law coefficients for both in-

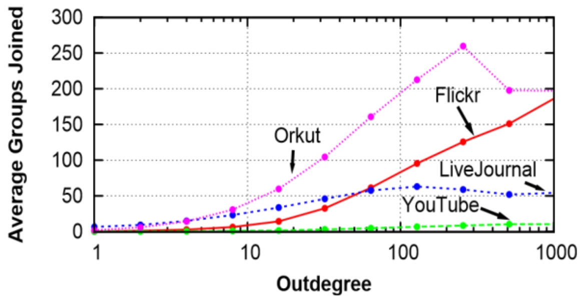

Figure 6: Outdegree versus average number of groups joined by users. Users with more links tend to be members of many groups.

VIII. Conclusion

We have presented an analysis of the structural properties of online social networks using data sets collected from four popular sites. Our data shows that social networks are structurally different from previously studied networks, in particular the Web. Social networks have a much higher fraction of symmetric links and also exhibit much higher levels of local clustering. We have outlined how these properties may affect algorithms and applications designed for social networks.

Much work still remains. We have focused exclusively on the user graph of social networking sites; many of these sites allow users to host content, which in turn can be linked to other users and content. Establishing the structure and dynamics of the content graph is an open problem, the solution to which will enable us to understand how content is introduced in these systems, how data gains popularity, how users interact with popular versus personal data, and so on.

The proposed system may suffer of problems similar to those encountered in the specification of OSN privacy settings. We plan to investigate the development of a GUI and a set of related tools to make easier BL and FR specification, as usability is a key requirement for such kind of applications.

Future Enhancement

I plan to study strategies and techniques limiting the inferences that a user can do on the enforced filtering rules with the aim of bypassing the filtering system, such as for instance randomly notifying a message that should instead be blocked, or detecting modifications to profile attributes that have been made for the only purpose of defeating the filtering system.

References

[1]. "A System to Filter Unwanted Messages from OSN User Walls" Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo, Department of Computer Science and Communication University of Insubria 21100 Varese, Italy IEEE Transactions On Knowledge And Data Engineering Vol:25 Year 2013

[2]. "Content-Based Filtering in On-line Social Networks" M. Vanetti, E. Binaghi, B. Carminati, M. Carullo and E. Ferrari, Department of Computer Science and Communication University of Insubria 21100 Varese, Italy fmarco.vanetti, elisabetta.binaghi, barbara.carminati, moreno.carullo, elena.ferrarig @uninsubria.it

[3]. Adomavicius, G.and Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[4]. F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[5]. M. J. Pazzani and D. Billsus, "Learning and revising user profiles: The identification of interesting web sites," Machine Learning, vol. 27, no. 3, pp. 313–331, 1997.

[6]. N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: Two sides of the same coin?" Communications of the ACM, vol. 35, no. 12, pp. 29–38, 1992.

[7]. P. J. Denning, "Electronic junk," Communications of the ACM, vol. 25, no. 3, pp. 163–165, 1982.

[8]. P. W. Foltz and S. T. Dumais, "Personalized information delivery: An analysis of information filtering methods," Communications of the ACM, vol. 35, no. 12, pp. 51–60, 1992.

Authors:

Pinniboyina Venkateswarlu isa student of Computer Science & Engineering from QIS Institute of Technology, He Presently pursuing M.Tech (CSE) in this college. He received B.Tech from JNTUK.

R Lakshmi Tulasi is a Professor, H.O.D of QIS Institute of Technology, Ongole. She received M.Tech from JNTUCEA. She is pursuing Ph.D. at JNTUH. She is a good Researcher in semantic web, Computer Networks. She attended Various National and International Workshops and Conferences.