Balancing the trade-off between privacy and profitability in Social Media using NMSANT

Rahul Ranjan
IIIT Allahabad
rr.rahul@aol.com

Charul
IIIT Allahabad
charul.agrl@gmail.com

Devina Vyas
SSIPMT Raipur
dvnessie81091@gmail.com

Durga Prasad Guntoju
IIIT Allahabad
durga.guntoju@gmail.com

***Abstract*—Social Media is being used as a key platform by advertisers to improve business by providing targeted and personalized advertising. There exist a trade-off between productivity in advertising and invasion of user's privacy in the existing approaches. Due to these privacy concerns, there were many law suits filed against the Beacon[1] advertising model used by Facebook resulting in its discontinuation. The new approaches need to address targeted audience while preserving the privacy in order to build appropriate revenue model for their operations. In this paper we propose an innovative model that leverage the trade-off. The model effectively interacts with user's linked data present in the web structured format, retrieve it and integrate data from marketing partners. Then it broadcasts the advertisement in social graph with flexible sentences For targeted advertising and privacy our model maintains interaction records among users in virtual containers for finding tie-strength[2] and also active friends [1] by Association Rules Mining [ARM] algorithm. We applied and validated our approach using a real data set obtained from 506 active social media users.**

***Keywords* - Social Media Advertising, Data Mining, Business intelligence, Privacy, NMASNT[2]**

I. INTRODUCTION

Social media provides an easily accessible platform to create, share, and exchange information and ideas in virtual communities and networks[3]. In a recent survey it has been found that consumer spent most of their time on the social networks compared with any other category of websites, roughly estimating; out of total online time 60 percent of their time goes to social network[5]. People are engaging themselves more and more on these sites which has resulted in 37 percent increase in total time spent on social media[5].

Many enterprises are seeking to integrate social media into business marketing [4]. Social media is changing consumer's behavior around the globe and now it seems obvious. Current market statistics state that worldwide 86% of companies have presence on Facebook and Twitter, while just over half of these use YouTube and LinkedIn and only slightly more than a third have a presence on Pinterest or Google+[17].

With revolution in personal data generation, gives a new opportunity to business parties for targeted marketing. But due to exorbitant use of user data without their consent lead violation[27] of privacy rights. Now it's become mandatory that personal information shared by users on social media must only be used with consent of him/her unless it will be termed as constitutional rights violation[18]. Also when a user creates account in any social media he/she is made to agree upon some terms and conditions which he/she agrees without reading it and this allows to use the personal details of the user to be used for data mining, behavioral advertising and data sharing. [19]

So in this paper we present a new revenue model that will defend the privacy of the user as well as will be productive for targeted advertising. We model the system as social graph in which each node represents a user and have a corresponding virtual container that stores all the information about the user activities. In virtual container we replace personal identity with unique global identifiers, in-order to safeguard the privacy of users and then we use the data in it for targeted advertising. Our model is divided into two phases, First we collect user data in virtual container and apply advance ARM to find the tie-strength[28] between nodes. Second, we broadcast advertisement using flexible sentences only to those node which have tie-strength attribute greater than threshold value.

The rest of the paper is organized as follows. Section 2 is about literature review. Section 3 describes our approach to the problem at hand in detail. Section 4 presents the implementation of the model. Section 5 summarizes survey report. Section 6 discusses our evaluation, while the implications of the results and possible future extensions are described in Section 7.

II. RELATED WORK

Social Media sites let anyone to register or sign up and allow user to use the site free of cost. So advertising is the main source[29] of revenue generation for these sites. The important point is their main source of revenue is through ads and sponsored stories. Under data protection, privacy and

[1] Active friend – Node having regular interaction
[2] NMASNT – Novel Method for advertising in social media network

competition laws, social media companies have to notify users or seek consent data mining, behavioral advertising and data sharing[21]. Therefore in spite of the popularity of the social media, there is nexus between privacy of the user and the profitability of the vendor[e.g. 7, 8, 9, 10]. Unfortunately there is trade-off in current Facebook advertising policy [e.g. 11, 12, 13, 14].

Before 2009 Facebook was using Beacon[1] revenue model but it was removed later on because of privacy violations. Beacon broadcasted users' stories to their friends without their permission. Many security (privacy) suits were filed against Beacon so it was shutdown in 2009 and Facebook co-founder Mr. Mark Zukerberg had apologized for this in his personal blog[30]. After injection of COPRA rule privacy violation can cause a big loss to these companies and can shutdown in future.

So there is an immense need of an alternate model which can filter all user stories as well as remove personal identifiers to ensure the privacy of user, and which can take the advantage of data mining to trace the common interests of the users. Also extends it to locate future customers based on the pattern extracted from previous listeners of the notification and should be functioned to make the advertisement model richer. Social Networking sites employ analytics for user generated information to provide targeted and specific advertisements[20]. But amidst all this excitement of such revolutionary advertising, the privacy of social media users cannot be compromised.

For instance, there are five ways to capitalize on social media utilization. (Constantinides et al., 2008) like direct commutation with customer on social media, advertising etc. and most of the companies uses that only. There are a lot of literatures discussing business respective of social media. Melakoski et al. (2007) state the term of "participatory economy", which refers to commodities that are produced, distributed, shared and consumed in social media. There are some basic categories of business model for online social media services discussed by Rappa (2010). Nevertheless, various unforeseen issues and challenges have surfaced by these models (Kim et al., 2009) and privacy is foremost.

So far, there are the potential problems in advertising in social network which are to be addressed[22]. Applying data mining techniques like association rule mining and clustering in advertising in semantic web leads to good results. So The concept of "Injecting Social Networking Data Into Semantic Web" and use of data-mining techniques for finding strongest associations motivated us to develop the NMASNT.

We have modified the previously used advertising model (Beacon) by Facebook and used Facebook datasets for simulation. Also we have introduced the concept of virtual container to safe-guard users privacy. Finally instead of sending a user's activity to all his/her friends, we proposes a new way of sending refined activity notification to specific group of people.

III. ARCHITECTURE

From Social network point of view dealing the trade-off problem is biggest challenge. Considering Facebook as our example, when it had switched from Beacon to existing model it lead to huge revenue loss to Facebook. The problem with beacon approach was that it was an opt-out system instead of opt-in, means if someone decline to share something Beacon still went ahead and shared it with their friends.

A. Notification Generation

Disclosing of the user information while broadcasting his action might cause privacy issues, to avoid this problem NMASNT recommends broadcasting a refined notification with flexible sentences[31].

Notification: Mr X has read "hacking Facebook passwords" article from Washington post social reader.

Refined Notification using flexible sentences: One of your friend tracked "hacking Facebook passwords" via Washington Post social reader.

Use of **flexible sentences** from new open graph protocol to enhance user engagement as well as privacy.

B. Avoiding un-necessary broadcasting

Current analytics broadcasts the advertisement notification to those nodes which are not active friends leading unnecessary advertisement. So we are using tie-strength to determine strongest association. Consider a politician or company that wants to broadcast a message through the network such that it only passes through trusted friends. Because strongly tied friends often reconcile their interests..

C. Common Interest mining by using data mining

NMASNT performs modified association rule mining on the users data stored in virtual containers who are involving in various events. Mining an association in the context of social media means an intelligent program sniffing common habits like intimacy, likes etc. and generating the tie-strength. of users who are actively involving in various events, this strategy helps to enhance customer base and targeted marketing in social media.

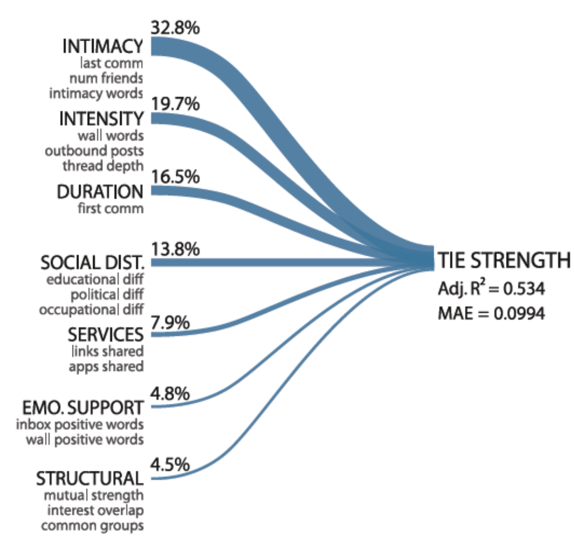


Figure 1. Showing tie-strength calculations[32]

D. Virtual Containers

Main aim of virtual container is safeguard user privacy in other way means that social network companies provide their data to third part vendors e.g. Facebook did with datalogix. Also virtual container contains information like news feed, messages, status updates, recommendations etc. in anonymous form.

E. Data model used in proposed solution

Semantic Web is the recent trend in Facebook as well as in other social network sites. It not only provides better structuring of data but also flexibility in information extraction. Proposed model is using RDF[16] (resource description format) model which contains public data of user profiles, RDF belongs to the category of graph data base.

F. Algorithm Approach

→ Start

- Sniff interactions among users.
- Capture groups based on interactions at multiple time intervals
- Dynamic clustering based on timed groups constructed in step 2
- Record the action done by user
- Add the action information in the virtual containers of all real followers in dynamic cluster of that user and removing personal identifiers.
- Provide an interactive environment between virtual container and the user.
- Find tie-strength between the persons who are responded to the previously broadcasted notification
- Capture person whose profile pattern tightly matches with the tie-strength found and adding them as future customers

end;

G. Modified Beacon Model:

For instance, we have a network with users 1, 2,3,4,5 and users 2,3,4,5 are friends of user 1.

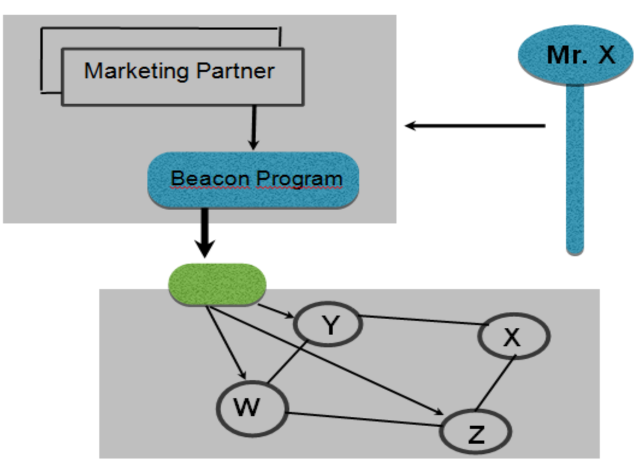


Figure 2. Beacon Model Architecture

- *Old way of advertising:*

Action: "abc" has purchased bottle of wine from xyz.com

Notification :< Mr. abc has purchased wine from xyz.com>

To Friends : 2,3,4,

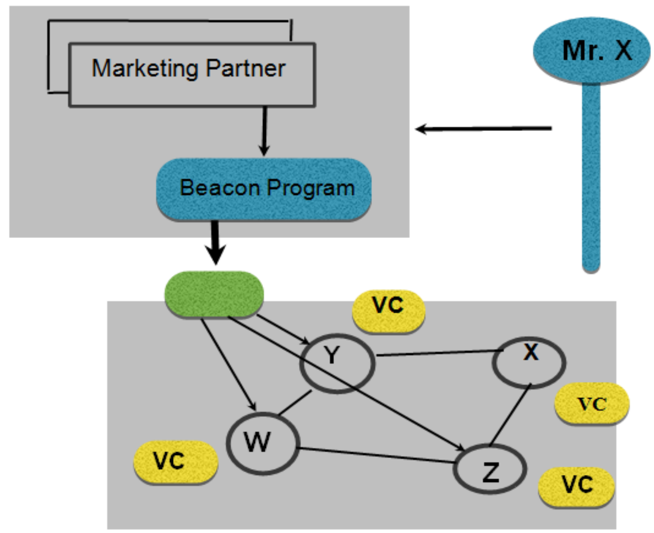


Figure 3. Modified Beacon Model Architecture

- *Current way of advertising:*

Action: "abc" has purchased bottle of wine from xyz.com

User Choice: don't notify

Notification: no message

To Friends: null

- *Proposed way of advertising*

Action: "abc" has purchased bottle of wine from xyz.com

Information: <Mr. abc has purchased wine from xyz.com >

Notification: <Your dear friend has just bought wine from xyz.com>

Static Friends: 2,3,4,5

Dynamic Friends: [($t=1$, {1, 2, 3}), ($t=2$, {1, 2, 3}), ($t=3$, {1, 2, 3, 4}), ($t=5$, {3, 5}), ($t=6$, {3, 5, 4}), ($t=7$, {1, 2, 3})}]

↓

NMASNT Algorithm

↓

{1, 2, 3} {3, 5}

Task: Loading notifications into virtual-containers of 2 and 3.

- *Explanation of proposed way of advertising*

Person named "abc" has purchased a bottle of wine from xyz.com. So now instead of broadcasting this information to all of his friends our model first apply several algorithm to find the strongest node and this is done on the virtual container to ensure anonymization of person user-id. Not only this, the model then uses the flexible sentences to modify the information string which would be appealing as well as free from personal identifier. Finally this information is broadcasted to only those node which has been found using ARM and tie-strength.

IV. IMPLEMENTATION

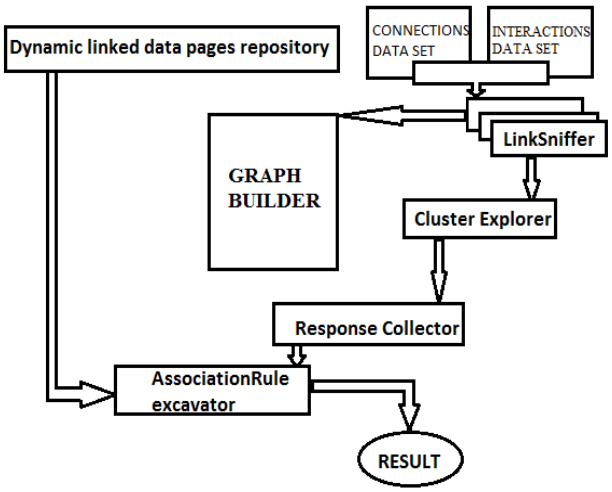
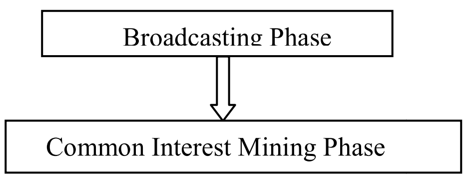


Figure 4. Implementation architecture of NMSANT

NMASNT Algorithm (proposed model) recommends social networking site to broadcast a notification to particular group of people.



A. Broadcasting Phase Algorithm

→ *start*

1. Initialize the friendship matrix
2. Get the friends of target-id
3. Calculate Healthy Communication Threshold
   3(a): Check threshold frequency of advertisement
4. Add friend to target Destination & Id to healthy friends
   4(a): Add healthy friend to friend of friend vector
5. Calculate Results 1 = difference between current time-stamp and recent interaction with common neighbor
6. Calculate Result 2 = difference between recent time-stamp of friend of friend and target-id
7. Check if (virtual edge strength > tie-strength)
   7(a): Add friend of friend in Target Destination List
8. Broadcast Notification to every user in Target List

B. BFA Algorithm in detail

Goal of the algorithm is to find the list of users to whom it has to send the notification about a particular user (event initiator). For this the above mentioned algorithm captures all registered friends, partitions them into two groups named active friends and passive friends. Active friends are those who have good communication (event specific) with the registered friends. Algorithm recommends that notification should be broadcasted to only active friends to avoid un-necessary notifications on passive friend's side. If active friends are very few then, to enhance advertisement destination domain NMASNT recommends considering friends of Active friends whose virtual edge strength is more than 0.5.

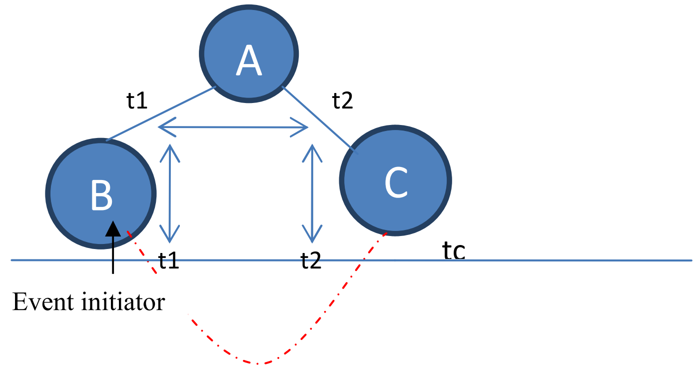


Figure 5. Sample graph (base condition)

Let B has initiated an event, to enhance advertisement destination domain here, NMASNT is considering all Friends of Active Friends of B, but they will be added to destination list if virtual edge strength is more than 0.5. In the above figure for B, A is an active friend and C is friend of Active friend of B. So there is a need to process C for B. Initially NMASNT has calculated a confidence of 0.5 about C. Technically this confidence is termed as Damping Factor in the proposed thesis. Initially damping factor is 0.5 because of inadequate knowledge about C i.e. (yes or no). Now A is common friend of B and C .B and C are communicating with A at times T1, T2 respectively. To check the recent status of these interactions

$$K = \text{current\_time} - \max(T1, T2)$$

Now DampingFactor^K gives us 1 if they have interacted with 'A' most recently. And **DampingFactor^K** value diminishes towards zero if their interaction has happened at long back. With this idea NMASNT sniffs each and every virtual edge (eg: B-C) of interaction temporal status.

To check the cumulative frequency of interactions done by B and C with A, NMASNT considers the harmonic mean of interactions. For example

B to A interactions: n_1

C to A interactions: n_2

Harmonic mean of these interactions is:

$$H_m = 2 / ((1 / n_1) + (1 / n_2))$$

So a consolidated score for the virtual edge B to C based on 1) Temporal behavior and 2) Frequency analysis is

$$\text{SCORE} = \frac{((\text{DampingFactor}^K) * (H_m))}{|T1-T2|+1}$$

In the above figure for B and C there is only one common friend A but in case if there are more than one common friends between B and C. To take more appropriate decision on C about B notification NMASNT sniffs all interactions made by B and C to all their common friends.

In such case SCORE will be (B and C having n common friends)

$$\text{SCORE} = \frac{\sum_{1}^{n} ((\text{DampingFactor}^{K})*(H_m))}{|T1-T2|+1}$$

Now if the score is more than the specified threshold (varies from context to context) than C is allowed to see advertisement notifications about B. In this way NMASNT captures all destination nodes to broadcast notification about event initiator.

C. Analysis of Broadcasting phase Algorithm

Time complexity

Best case: O($\mathbf{n_f}$*T) (when there are no active friends for event initiator)

Worst case: O($\mathbf{n_f}$*$\mathbf{n_{fof}}$*T) (when all friends of event initiator are active friends)

T = O(n) where n is the maximum frequency of interactions

D. Association Rule Mining Algorithm application phase:

Once we have accessed RDF elements of Facebook pages with open graph protocol as shown in the above figures then we can start our algorithm. Here is the detailed explanation.

Extracted elements from Facebook Open Graph Protocol {Id,name,picture,link,likes,category,is_published,website,username,founded,Mission}

e.g. Extracted elements of Coca-Cola:

{id, name, picture, link, likes, cover, category, is_published, website, username, founded, Mission}

- *Step 1 Removing un-necessary data*

Erasing unique attributes (personal identifiers) like id, name, picture, link etc.

- After step1 resultant attribute sets are very useful for data mining in Facebook

Set 1 :{ like, mission, activities}

Coca-Cola

Set 2: {likes, mission}

- NMASNT redirects all generated sets to transaction database T, which will be treated as input to the next step.

Step2 Applying classic ARM(Association Rule Mining) on T

```
Pseudo-code:
    Ck: Candidate itemset of size k
    Lk : frequent itemset of size k

    L1 = {frequent items};
    for (k = 1; Lk !=∅; k++) do begin
        Ck+1 = candidates generated from Lk;
        for each transaction t in database do
            increment the count of all candidates in Ck+1
          that are contained in t
        Lk+1  = candidates in Ck+1 with min_support
        end
    return ∪k Lk;
```

Figure 6 Classic ARM[15]

Above Figure is the snapshot of the classic ARM algorithm, here we are constructing Transaction Database T with attributes sets of web pages of different nodes in an open graph.

T looks like Facebook**: {**likes, mission, motto, activities**}**

Coca cola: {mission, motto, likes}

Pepsi: {activities, likes, motto}

Results after applying ARM on Transaction Database T looks like: {likes, motto}

Step 3: Common Interest Mining Algorithm

Result of step2 indicates attributes to be mined to get the more appropriate result about the common interests of the users.

E. Sub Graph Extraction Algorithm

```
DetectCommonInterestSubgraph(G; s; P; f)
    C←P
    while true
        t← null
        for each v 2 (nG(s) - C)
         if f(G;C; v) = true
          if t = null or h(G;C; v) > h(G;C; t)
            t←v
        if t = null
            break
        else
          C ← C U {t}
          return C
```

The algorithms check whether the participants/friends have minimum threshold tie-strength value or not. It takes the Graph, source node, participants nodes and tie-strength as parameter. For each node in graph it compares the tie-strength values. Finally return the filtered set of active friends.

F. Analysis of Algorithm

Algorithm Takes O(T*n^2) time complexity typically.

T: no of Attribute sets(typically we can say that no of nodes involving)

N: attributes of a specific node for which maximum attributes are recorded.

V. SURVEY REPORT

To validate our advertising model where we opt not to disclose the user's identity while broadcasting the advertisement, is acceptable to users, we conducted a survey on active social media users and 506 participants replied to the questions we asked them. All of the participants were active members of Facebook and most of them have their account on 3 to 6 social networking websites. From these questionnaires, we found that of the 506 participants, 315 were males and 191 were females. The mean age was 22.5 years with a standard deviation of 6.35 years. The minimum and maximum ages

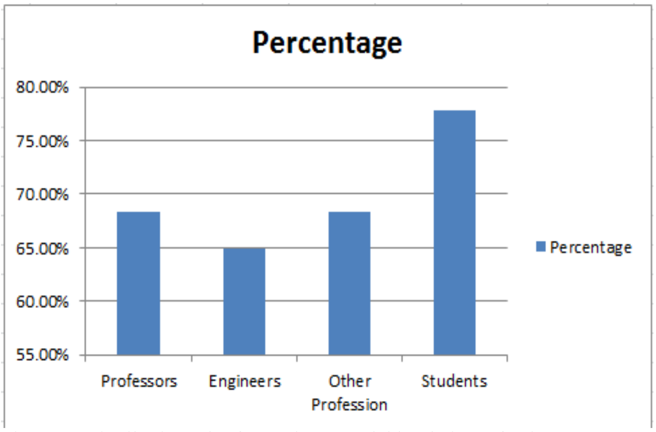


Figure 7. Distribution who favored our model by their profession

were 16 and 62 respectively. Among the participants, 329 were students, 57 were Engineers, 38 were Professors and 82 were of other profession. They were asked to answer a set of questions regarding the privacy on social networking websites based on their usage and experience. From these questionnaires, we found that more than half of the users i.e. around 58.3% users do not read the privacy policies of the social networking website before signing up. More than 77% of the participants agreed that privacy on social networking website is a matter of grave concern to them. About 59.82% of the people were satisfied with the current privacy model used in social network. Total of 74.11% participants have no problem when advertiser uses their activity to broadcast an advertisement to their friends without disclosing their identity (that could identify them personally e.g. name, location).

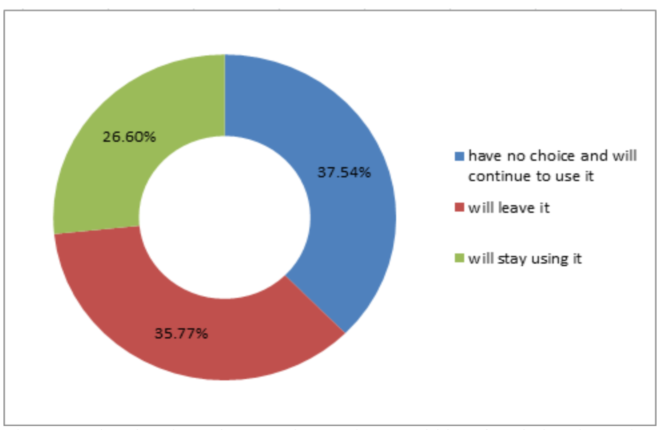


Figure 8. Plot showing what people reaction would be after their privacy is invaded

VI. RESULTS AND DISCUSSIONS

We have used anonymous data-set of 10000 unique Facebook user publically made available by Facebook Social Graph - MHRW & UNI. The data sets includes common attributes like likes, friends information, network id, conversation, privacy setting etc. for each user. We have tested it on a set of 20 users to verify the result more precisely but the model is developed to handle large data sets.

NMASNT has an implicit test case generator which was programmed using the concepts of random distribution and dynamic programming with non-overlapping partition approach. When a tool needs to be tested under different datasets which are having data of millions of users, such type of test case generator really helps in generating efficient test cases.

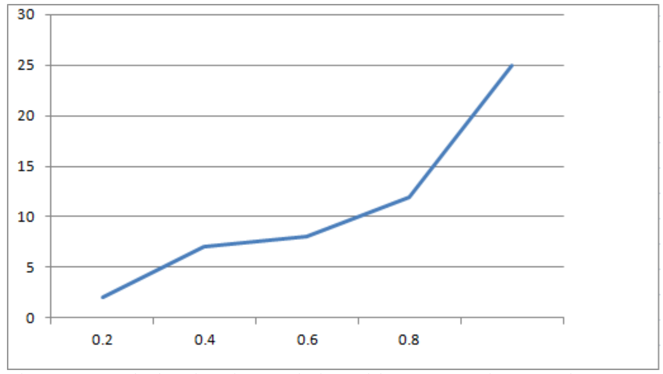


Figure 9. Graph showing time variation with respect to tie-strength

The increasing time with respect to tie-strength shows that model uses more instances of user interactions with its friends. As the time complexity is approximately O(n*2), increasing the tie-strength more than 0.6 results into very steep rise in computation time. Finally we have observed overwhelming output in detection of active friends.

Limitation

Although the model is efficient but its runs out be slow for large data-sets, so future works can be on parallel based architecture of the model. Another limitation is in the survey which we had conducted. The survey was conducted on a relatively small size of sample.

VII. CONCLUSION

This paper proposes a new model for advertising in social media by searching active friends with strong tie-strength. This can be useful in target marketing and avoiding redundant notification to passive friends. From here there are many possible paths that are open. The data can also be used in recommender system and web personalization by improving tie-strength algorithm. Also the method can be implemented for Mobile platform. Moreover the advertising model that we have adopted consist privacy as main feature, which is of paramount importance in the area of research.

We believe this work addresses fundamental challenges for understanding the trade-off problem in revenue models. How can profitability be achieved without hampering the privacy of user's? How to design an advertising model that can address only active friends in social media for effective targeted marketing? Also the work has presented how simple changes in advertising model can increase privacy of user's and their trust. We think this work takes a significant step toward definitively answering these questions.

VII. REFERENCES

[1] Facebook, Beacon and Your Privacy By Katherine Blizard, CSC 300: Professional Responsibilities Dr. Clark Turner, March 2, 2011

[2] Predicting Tie Strength with Social Media by Eric Gilbert and Karrie Karahalios University of Illinois at Urbana-Champaign

[3] Ahlqvist, Toni; Bäck, A., Halonen, M., Heinonen, S (2008). "Social media road maps exploring the futures triggered by social media". *VTT Tiedotteita - Valtion Teknillinen Tutkimuskeskus* (2454): 13.

[4] Social Media's Positive impact on business by Mike Boylan, November 11, 2009.

[5] State of the media: The social media report 2012". *Featured Insights, Global, Media + Entertainment*. Nielsen. Retrieved 9 December 2012

[6] The Economics of Advertising and Privacy by Catherine Tucker MIT Sloan School of Management, MIT, Cambridge, MA; 617 252-1499 and NBER

[7] Social Network Site Privacy-A comparative analysis of six sites by Jennifer Barrigar, the office of privacy commissioner of Canada Feb 2009

[8] Analysis of Social Networking Privacy Policies by Leanne Wu, Maryam Majedi, Kambiz Ghazinour and Ken Barker, 2010 ACM

[9] Mining Social Media: A Brief Introduction by Pritam Gundecha, Huan Liu, 2012 @INFORMS

[10] Information Revelation and Privacy in Online SocialNetworks (The Facebook case) Pre-proceedings version. ACM Workshop on Privacy in the Electronic Society (WPES), 2005

[11] Facebook's Privacy Trainwreck Exposure, Invasion, and Social Convergence by danah boyd *Harvard University and University of California-Berkeley, USA 2008*

[12] Facebook: Threats to Privacy Harvey Jones, Jos_e Hiram Soltren December 14, 2005

[13] Facebook and Privacy: It's Complicated Maritza Johnson Symposium on Usable Privacy and Security (SOUPS) 2012, July 11–13, 2012, Washington, DC USA

[14] The Political Economy of Privacy on Facebook by Christian Fuchs, Jan 13 2011

[15] Association Rules Mining: A Recent Overview By Sotiris Kotsiantis, Dimitris Kanellopoulos Educational Software Development Laboratory Department of Mathematics, University of Patras, Greece 2006

[16] Querying RDF Data from a Graph Database Perspective by Renzo Angles and Claudio Gutierrez

[17] Marketer still can't tie social to bottomline.http://www.emarketer.com/Article/Marketers-Still-Cant-Tie-Social-Bottom-Line/1009340

[18] Lothar Determann, Freedom Of Communications On the Internet 1 ET. SEQU. (1999).

[19] Rethinking the Role of Consent in Protecting Health Information Privacy, Jan. 2009 at 8 available at http://www.cdt.org/healthprivacy/2009126Consent.pdf

[20] Marketing through Social Networks: Business Consideration from brand to privacy by Katheryn A.Anderson, 2011

[21] Chris Jay Hoofnagle, Ashkan Soltani, Nathaniel Good, Dietrich J. Wambach & Mika D. Ayenson, Can Advertisers Learn That "No Means No"?, Privacy & Sec. L. Rep. (BNA) 1398 (Sept. 26, 2011).

[22] The Online Advertising Industry: Economics, Evolution, and Privacy, David S. Evans, April 10, 2009.

[23] Kim, W., Jeong, O.R. & Lee, S.W. (2010) "On Social Web Sites", Information System, Vol. 35, pp. 215-236.(kim et al)

[24] Rappa, M. (2004) "Business Models on the Web", Managing the Digital Enterprise (Rappa 2010).

[25] Constantinides, E. & Fountain, S.J. (2007) "Web 2.0: Conceptual Foundations and Marketing Issues", Journal of Direct, Data and Digital Marketing Practice, Vol. 9, No. 3, pp. 231-244.

[26] Melakoski et al. (2007) "Himottaa, Mutta Pelottaa", Suomalaisen Sisältötuotantoala näkemyksiä

[27] Lothar Determann & Robert Sprague, Intrusive Monitoring: Employment Privacy Expectations Are Reasonable in Europe, Destroyed in the United States, 26 BERKELEY TECH. L.J. 979 (2011).

[28] Granovetter, M. S. 1973. The Strength of Weak Ties.The American Journal of Sociology, 78(6), 1360–1380.

[29] How social network sites generates revenue.http://computer.howstuffworks.com/internet/social-networking/information/how-social-networking-sites-make-money1.htm

[30] Mark Zuckerberg, Our Commitment to the Facebook Community, Facebook blog , http://blog.facebook.com/blog.php?post=1015037870193711

[31] Better open graph Stories with flexible sentences in Facebook https://developers.facebook.com/blog/post/2013/01/03/better-open-graph-stories-with-flexible-sentences/

[32] Predicting Tie Strength With Social Media Eric Gilbert and Karrie Karahalios University of Illinois at Urbana-Champaign, 2009