

# *Using Sentiment Analysis and context evaluation for preserving Trust-based privacy in Social Networks*

Nadav Voloch  
Department of Computer Science  
Ben-Gurion University of the Negev

*Done as part as the requirements for the course of NLP and Social Dynamics – Spring 2020*

**Abstract**—Online Social Networks (OSN) security and privacy issues have been extensively researched in the past decade. Information is posted and shared by individuals and organizations in social networks in huge quantities. One of the most important unresolved topics are the breaches of privacy that harm OSN users. These breaches occur because of several reasons, one of which is non-trustworthy users. These users, some with malicious intentions, and some with low social media awareness create a problem of spreading data, that is considered private. Users that are not technologically oriented consider their network as a relatively safe space of their friends. This misperception creates a problem if they share information that is considered private. Our research addresses this problem. In our previous research we have devised a comprehensive Trust-based model that handles this problem from the user's Trust aspect. The model involves Access Control for the direct circle of friends and Flow Control for the friends' networks. In this paper we create a context-based model that involves user profiling, OSN features and OSN activities. We then validate this model by analyzing Trust using sentiment analysis on posts in the network. This model creates a much more accurate picture of OSN users and their data and helps revealing the sources of problematic data exposures and can prevent them from happening. Applying this model can help create a much better privacy infrastructure for OSN.

**Keywords**—*Social Networks privacy, Context in Social media, Trust-based privacy, NLP in Social networks.*

## I. INTRODUCTION

Handling Online Social Networks (OSN) security is the subject of many research papers in the past years. The issue of privacy in OSN was handled in early papers such as [1] and [2]. According to [3], there is very little or no actual user-awareness to the spreading of personal data throughout the network and the extent to which the data is spread is seldomly evaluated correctly. The only definite knowledge users have is that their information instances (e.g., pictures, posts, personal details, etc.) are revealed to their direct OSN friends. This situation creates a problem, in which users often post or share inappropriate or private data with their friends, with little awareness of who might see this data, even amongst their OSN friends. An exemplification of such a problematic case is seen in Fig.1, taken from a real OSN. In this example the user would have not

wanted that the vice-principal of his workplace would see this post.

In this case, the user understands the problematic aspect of his post only after he gets the reactions and comments about it. He then, accordingly, adds an apologetic comment about it.

The purpose of our research is to prevent cases like this one and others, in which OSN data privacy is breached due to non-awareness, or, in even more problematic cases, due to malicious use of data done by users that take advantage of this situation and spread it for different purposes.

In our previous work, we have created an OSN security model that is composed of three main phases addressing three of its major aspects: trust, role-based access control ([4], [5]) and information flow, by creating an Information Flow-Control model for adversary detection ([6]), or a trustworthy network ([7]). The main idea of this research is to extend the basic Trust model and make an important separation for different types of data instances, that differ by their subject's category. For example, a political post might be more sensitive for its publisher than a simple "Good morning everyone". Another aspect that is affected from this extension is the users themselves. The OSN user's friends are not homogenic by nature and accommodate different perspectives and views, and accordingly, to the user himself, are trustworthy in several levels. Some are considered close friends, some just acquaintances, or even less.

In the experimental part of this research we used three different datasets, all taken from real Facebook networks. For the first part of context evaluation, we used a network of 917 users, for which we have devised several data categories. We have assessed the Trust level of every user, and in every category. Then we analyzed their posts in every category each Trust-wise. For the second part of sentiment analysis, we took two datasets, one of them contained 61 Facebook posts from 36 different users, and the other contained 75 post from 31 different users. Besides the parameters of the first part of context evaluation, in these datasets there were specific trust scores for the posts, and their sentiment analysis. Our purpose was to find the affect of sentiment in a post to the user's trust in a certain context.

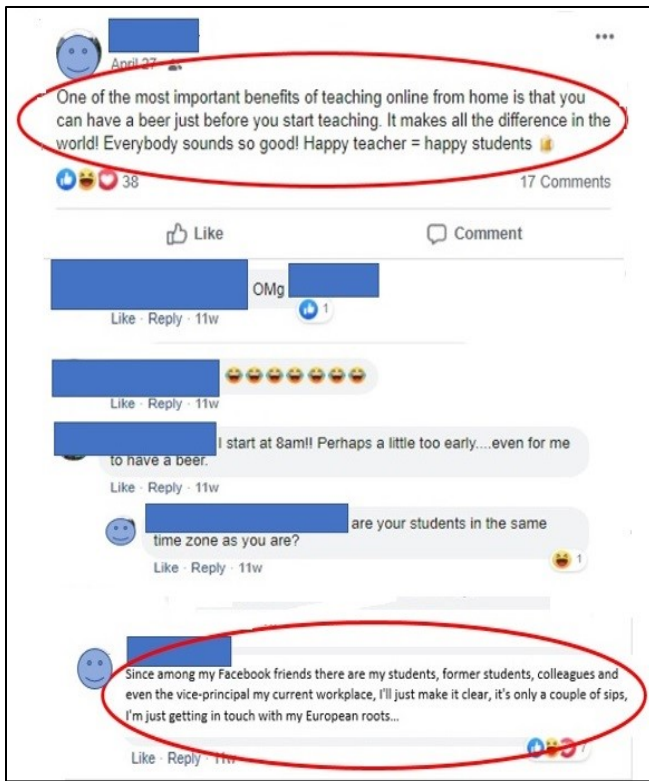


Fig. 1. Inappropriate data sharing due to unawareness in OSN

All the materials, including the different datasets, and the code for the program written for the purpose of this research are all online- A hyperlink is available in the appendix part. The results of these experiments are presented in this paper.

The rest of this paper is structured as follows: Section II discusses the background for our work, with explanations for the related papers it relies on. Section III describes and defines our basic Trust based model. Section IV presents our context-based extension of the model. Section V includes the experimental results of the context-based part, including explanations on the datasets and experimental settings, and Section VI discusses and concludes the paper, along with some future aspects of it.

## II. BACKGROUND AND RELATED WORK

The main problem OSN security models deal with, is the preservation of privacy of users in the network, in which data shared by a certain user can be reached by an unwanted entity, such as a spammer, a data-harvester, or even a real user, that could take advantage of this information, depending on the type of data and the preferences of the data sharing user. In [8] different types of these privacy-breaching scenarios are described. Most of these vulnerabilities occur from discretionary privacy policies of OSN users.

These privacy policies create a misleading knowledge of the number and type of users exposed to this shared data. Most of the solutions suggested demand changes in these specific policies.

There are several approaches of handling OSN Security and privacy, among them are Access Control, Information Flow Control, and Trust.

[9] gives a survey of most of the OSN Access Control models, elaborating the functionalities of the different types.

[10] presents a new model for privacy control based on sharing habits, controlling the information flow by a graph algorithm that prevents potential data leakage. In [11] a relationship-based approach is being handled, giving priority to the users' relationships qualities, on which we have based our initial idea for the model. The social network is usually represented as an undirected graph, where nodes are the OSN users, and edges represent relations between them such as friendship relations.

An Ego node (or Ego user) is an individual focal node, representing a user whose information flow we aim to control. An Ego node along with its adjacent nodes are denoted Ego network. As mentioned in the Introduction, [3] presents the fact that there is very little or no actual user-awareness to the spreading of personal data throughout the network and the extent to which the data is spread is seldomly evaluated correctly. The only definite knowledge users have is that their information instances (e.g., pictures, posts, personal details, etc.) are revealed to their direct OSN friends.

The source of the privacy problem of unawareness of data spreading begins when one of these friends acts upon an information instance, meaning comments on a post, likes or shares a picture, or any other form of OSN action. This action allows any friend of the actor, which is usually not a direct friend of the Ego-node, to see this information instance. Fig. 2 describes an Ego user's data flowing to friends of friends (Users A1, A2 and A3), triggered by an action (a comment in this example) taken by the ego node's direct friend (User A) on the Ego node's data.

Our Trust-based model estimates the friends' Trust in the Ego network, and the friends of friends' Trust. This estimation resembles the problem of spammer detection ([12]) since the misuse of private data is the common ground for both, and the prevention of privacy breaches is their mutual interest. In [13] this detection is also done on Facebook datasets, where it is shown that spammers usually have noticeable differences in values of certain attributes such as number of friends, tags and mentions. For these linked problems of estimating Trust for privacy preservation and spammer detection (or any other malicious or unwanted user) we focus on suggested approaches and solutions of Access Control, Information Flow Control and Trust. The main Access Control model used in OSN is Role-Based Access Control (RBAC) that has many versions, as presented in [14], and limits access by creating user-role assignments. The user must have a role that has permission to access that resource.

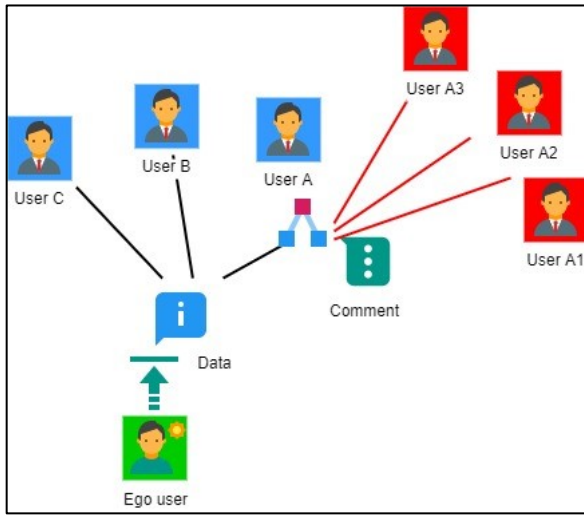


Fig. 2. The source of the privacy problem of unawareness of data spreading

The most prominent advantage of this method is that permissions are not assigned directly to users but to roles, making it much easier to manage the access control of a single user, since it must only be assigned the right role.

An example of using RBAC specifically in Facebook is done in [15], that describes the use of roles in it and the possible breaches that can occur due to the flexible privacy settings of the network. Using Trust in OSN is widely used in different models, and even in relatively early research such as [16], the idea of involving trust in Access Control for OSN user data is handled, in creating Trust criteria for different subjects (users) and objects (data instances). A model that added Trust to RBAC was presented in [17], and it is based on the network users' interactions history, which could be problematic in assessing relatively unknown new connections. We use a similar idea in our Trust-based model. [18] assesses the importance of OSN attributes for similarity-based access control. An important ranking is given to these attributes, based on information gaining from each attribute, figuring their importance in the closeness approximation between users and evaluating their information sharing willingness.

The model presented in this paper extends our basic model ([4], [5], [6], [7]), and relies on context of data in the OSN, dividing it to different topics with different characteristics. In [19] a contextual social network model that uses personal characteristics as independent social context, and mutual relations is presented. It suggests social context-aware trust inference in OSN, that is used for recommendations on service providers. In [20] there is a very interesting use for context and content analysis in OSN, done in an attempt to predict event attendance in event based OSN (such as Facebook). [21] handles the problem of preserving users' individual privacy when publishing relatively rich information in OSN. This is done by anonymization and context-related Trust, by referring to connections between users in different topics. The use of NLP in information security was done even in relatively early papers such as [22], [23]), and it is, of course, widely used in Social Media ([24]), and Sentiment Analysis is also used for these types of research, such as the one presented in [25].

We use the papers mentioned above ([19-25]) for our context-based model, that will be presented in detail in section IV of this paper.

### III. THE BASIC TRUST-BASED MODEL

#### A. Basic description

We represent a social network as an undirected graph, where nodes are the OSN users, and edges represent relations between them such as friendship relations. An Ego node (or Ego user) is an individual focal node, representing a user whose information flow we aim to control. An Ego node along with its adjacent nodes are denoted Ego network.

The model we present is composed of three main phases addressing three of its major aspects: trust, role-based access control and information flow.

In the First phase, the Trust phase, we assign trust values on the edges connecting direct friends to the Ego node in their different roles. These trust values are calculated based on seven different parameters as explained in the upcoming section. In the second phase, the Role Based Access Control phase, we remove direct friends that do not have the minimal trust values required to grant a specific permission to their roles. After this removal, the remaining user nodes and their edges are also assigned with trust values. In the third and last phase, the Information Flow phase, we remove from the graph edges and nodes that are not directly connected to the Ego-user, by using different graph algorithms, to construct a privacy preserving trusted network.

#### B. Trust parameters, notations and values

The choice of the attributes, for determining the level of trust for the model, is based on our research mentioned in the previous sections (specifically [4], [5], [7]) and here we relate to four main OSN attributes that are presented in Table I. For example,  $p_{MF}$  is the value for the Mutual Friends attribute. A Trust value ranges between 0 and 1 to reflect the probability of sharing information with a certain user: 0 represents total restriction, and 1 represents definite sharing willingness.

The threshold values are denoted here as  $T^{property}$  (e.g. for the  $TF$  attributes the threshold value is  $T^{TF}$ ) and their experimental values, achieved in our previous research mentioned above are presented in Table I. The calculation of a certain property value ( $p_{property}$ ) is done by these thresholds and is as follows:

$$p_{property} = \begin{cases} \frac{property}{T^{property}} & (property < T^{property}), \\ 1 & (property \geq T^{property}). \end{cases} \quad (1)$$

At this point we define the model's User Trust Value ( $UTV$ ), calculated as the weighted average of these properties, taking into consideration the different weights ( $w_i$ ) that were assessed by experimental results in [4] and [5] for the significance (weight) of every attribute-factor. These were of almost similar values in the results, thus can be treated equally.

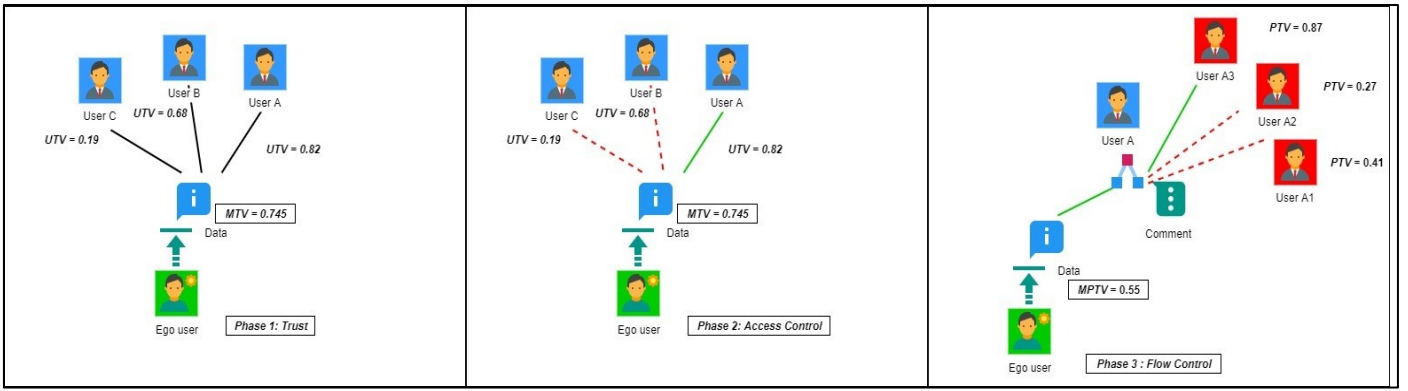


Fig. 3. The basic model's phases for creating a trustworthy network

The  $UTV$  is as follows:

$$UTV = \langle w_i p_i \rangle = \frac{\sum_{i=1}^{|p|} w_i p_i}{\langle w \rangle |p|} \quad (2)$$

The threshold value for the determining whether to give a certain access to a data instance in this model is the Minimal Trust Value ( $MTV$ ) and it is based on the experimental results of [4] and [5] and presented in Table I.

### C. The Access and Flow control for the creation of a trustworthy network

The decision used in [4] and [5] for Access Control for the direct friends involves granting permissions per OSN Role, with the input of the User's Trust Value ( $UTV$ ).

In this phase we prevent access from direct friends in different roles that do have the minimal trust values of specific permissions to these specific adequate roles.

The removal is also done for their connected networks. After this removal, the remaining user nodes and their edges get also assigned with trust values. For this phase we have devised Algorithm 1 for the access decision:

#### Algorithm 1. GrantPermission

**Input:** Minimal Trust value:  $MTV$ , User  $U$ , Role  $R$ , Permission  $P$

**Output:** granted or denied

```

if  $P \in R$ 
    if  $U.UTV \geq P.MTV$ 
        return granted
    else
        return denied
else
    return denied

```

After this initial screening, we go on to the Flow Control phase, that involves the users that are not directly connected to the Ego-node. For that purpose, we have defined in [6] the following notations:

Let  $G = (V, E)$  be an undirected graph that describes the OSN where  $V$  represents the set of users and  $E$  represents the set of social connections between them.  $v_{src} \in V$  is the Ego source node, that holds the information to be shared, and  $v_{tgt} \in V$  is the Target node, that may or may not get the information from  $v_{src}$ .

A Path from source to a Target node denoted  $PATH^{src \rightarrow tgt}$  is a set  $\{v_{src}, E^{src \rightarrow 1}, v_1, E^{1 \rightarrow 2}, \dots, v_k, E^{k \rightarrow tgt}, v_{tgt}\}$ , where the number of intertwined user-nodes is  $k$ .

We calculate the Trust value of a path  $PATH^{src \rightarrow tgt}$  defined by  $\{v_{src}, E^{src \rightarrow 1}, v_1, E^{1 \rightarrow 2}, \dots, E^{k \rightarrow tgt}, v_{tgt}\}$  by multiplying the trust values of each node and each edge on the path, where the trust of a node  $v_i$  is determined by user-credibility attributes  $ui$  (that are  $TF$  and  $AUA$ ) and the trust of an edge  $i \rightarrow j$  is determined by connection attributes  $ci$  (that are  $MF$  and  $FD$ ). The Path Trust Value is denoted as  $PTV$ .

$$PTV(PATH^{src \rightarrow tgt}) = \prod_{i=1}^k ci \cdot ui \quad (3)$$

In the Flow Control algorithm, described in [6] we begin by finding all possible paths between the source and target nodes and then we calculate the  $PTV$  value for each path and if the  $PTV$  of at least one path is higher than the threshold ( $MPTV$ ), it returns true indicating that the node is an acquaintance. If no path has a sufficient trust value, it returns false, indicating that the target node is an adversary.

The model's result of these phases gives us a trustworthy network of users to which the Ego – user can safely share information with. These three phases are seen in Fig. 3.



TABLE I. EXPERIMENTAL RESULTS FOR TRUST VALUES FOR THE MODEL'S PARAMETERS.

Parameter	Attribute	Experimental value
$T^{AUA}$	Age of User Account (OSN seniority)	23.82
$T^{TF}$	Total Friends	244.34
$T^{MF}$	Mutual Friends	37
$T^{FD}$	Friendship Duration	17.12
$MTV$	Minimal Trust Value	0.745

#### IV. CONTEXT-BASED MODEL FOR PRIVACY IN THE NETWORK

The basic model treats all the users equally, and as we demonstrated in the example not all users are homogenous by nature in their preferences, specifically in different topics and data categories. A certain friend of the Ego user can be very trustworthy, but with very different political opinions from the him, a fact that might make the Ego user wish not to share political posts with him. The main idea of this research is to extend the basic Trust model defined above and make an important separation for different types of data instances. Some, inherently, are more sensitive than others, thus, need to be treated differently. The perspective on the data is very subjective, and for that matter, the opinion that matters the most is the Ego user's one. Some Ego users might see the same content in different lights. This happens naturally because of different political views, different preferences, different types of personality and more.

When a certain user creates a post, a comment or any other type of data instance in the OSN, this action can be used to learn the user's tendencies in different topics and categories, thus we can create a contextual, dynamic trust level per user per topic or data category.

For the purpose of context evaluation, we categorize different users in the Ego network by their Trust per context.

An Ego user gives his friends different trust values for every category, meaning that they have a Subjective Trust Value denoted here as  $STV_{\kappa}$ .

These  $\kappa$  categories can be various by nature, and can include politics, sales, sports, social friendship interactions, etc. The Ego user is the one deciding the value of  $STV_{\kappa}$  of a user in his network. This process is done in the beginning of the friendship (approval of friend request). A user's Trust value can dynamically change over time. This Trust value changes because of different actions the user does in the OSN. For that purpose, we included in the model the Subjective Trust Value of Actions ( $STVA_{\kappa}$ ) in the OSN, that is consisted of the Trust values given by the Ego user to his friends actions (posts, shares, etc.). Every action's Trust value is accordingly is

denoted as  $STV_{iA_{\kappa}}$ . For every action there is a weight ( $w_i$ ) that represents the effect of the action (some actions can affect the user's subjective trust value more than others). These weights can change according to the importance given to them by the user. In this model they are equal by default.  $STVA_{\kappa}$  is computed as a weighted average of the trust values given to each of the actions, normalized by the number of actions, and therefore it is:

$$STVA_{\kappa} = \langle w_i STV_{iA_{\kappa}} \rangle = \frac{\sum_{i=1}^{|STVA_{\kappa}|} w_i STV_{iA_{\kappa}}}{\langle w \rangle |STVA_{\kappa}|} \quad (4)$$

At this point we can calculate the total Trust level of the user in a certain  $\kappa$  category. We denote it as  $fUTV_{\kappa}$ , and it is consisted of the basic model's  $UTV$  and  $STVA_{\kappa}$ , and, considering the weight ( $w$ ) of every factor, is as follows:

$$UTV_{\kappa} = \frac{w_{UTV}UTV + w_{STVA_{\kappa}}STVA_{\kappa}}{\langle w \rangle} \quad (5)$$

We can see an example for such a set of  $UTV_{\kappa}$ 's and access granting for certain data instances in Fig. 4, where the Minimum Trust Value of a certain category of a data instance is presented as  $MTV_{\kappa}$ . Three out of four users hold the necessary trust value ( $UTV_{\kappa}$ ) for  $\kappa=Politics$ , thus have access to it, while only one user hold the necessary  $UTV_{\kappa}$  for  $\kappa=Sales$  and has access to it.

The purpose of the estimation of  $UTV_{\kappa}$  for each user in the network is to give an estimation, accurate as possible, to the Ego user's general trust estimation for a user in a category- $STV_{\kappa}$ , that is measured in the first experimental part of this paper.

The evaluation of  $STVA_{\kappa}$  is done in the second experimental part of this paper – by Sentiment Analysis- we wish to assess whether a negative or a positive sentiment has an effect on the Trust a certain user has on a content of a post.

The results of the experimental evaluations for all of these attributes are presented in the upcoming section.

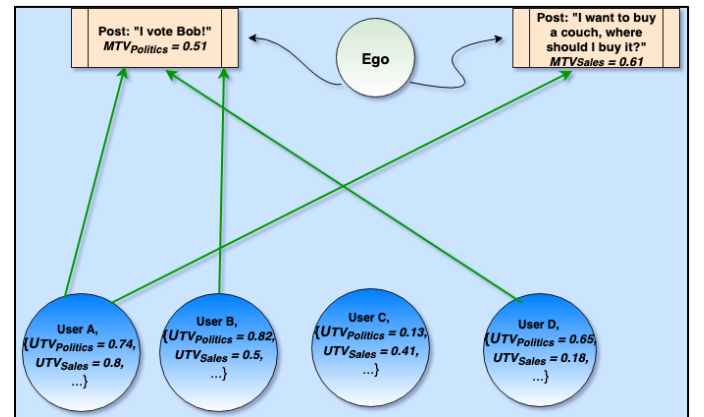


Fig. 4. Access decisions to data instances in different categories and adapted trust values

## V. EXPERIMENTAL SETTINGS, DATASETS AND RESULTS

For the first experimental part of this research we used a dataset of a real Facebook network of 917 users, which are the direct friends of a single Ego user.

The Ego user first collected all of the users' data relevant to the basic  $UTV$  -  $p_{MF}$ ,  $p_{TF}$ ,  $p_{AUA}$ , and  $p_{FD}$ . We then calculated every  $UTV$  accordingly.

There were five different data categories (the  $\kappa$  categories) that were chosen by the Ego user, and for each category four actions (posts, shares, etc.) were documented by him – a total of 20 action per user.

The Ego user then gave a specific Trust estimation for this user in every category ( $STV_{\kappa}$ ). He then went over every action and gave his Trust estimation of every action in every category ( $STV_i A_{\kappa}$ ). After these Trust estimations, we calculated  $STV A_{\kappa}$  as described in Equation 4.

After gathering and calculating all the values, we then reached  $UTV_{\kappa}$  per each category per each user, as described in Equation 5. The comparison that was done is to the  $STV_{\kappa}$  value since our model tries to predict the real Trust estimation of the Ego user for everyone in his network in different categories.

We then divided the users to batches by their  $UTV$ :

- $UTV = 0.7-0.8$ : relatively low trusted users.
- $UTV = 0.8-0.9$ : medium trusted users.
- $UTV = 0.9-1$ : high trusted users.

There we no users below  $UTV=0.7$ , a fact that suggests that this is a relatively strong network.

We did this batch division to examine the different effects of our model's estimation on different types of users.

The results of the average  $UTV_{\kappa}$ 's and  $STV_{\kappa}$ 's of the different batches are presented in Table II.

In Fig.5 we can see a graphic representation of the differences between the  $UTV_{\kappa}$ 's and  $STV_{\kappa}$ 's in the different batches. We can see that the  $UTV_{\kappa}$ 's are generally higher than the  $STV_{\kappa}$ 's. One of the reasons for this is that the networks is relatively strong (high  $UTV$  values). So, accordingly, if we take a closer look, we can see that these differences are even stronger in the high trusted users (0.9-1).

For the second part of sentiment analysis, we took two datasets, both contained all the parameters of the first part. One of them was of the Ego-user from the first part that had contained 61 Facebook posts from 36 different users. The other dataset contained 75 post from 31 different users. Besides the parameters of the first part of context evaluation, in these datasets there were specific trust scores for the posts, and their sentiment analysis. Our purpose was to find the effect of sentiment in a post to the user's trust in a certain context. The topic chosen for the context was Israeli politics – the two datasets of the second part contained only posts from this topic. The sentiment analysis done on the posts of the first dataset was evaluated by the ratio of negative and positive words in the post.

In the second dataset an additional method was checked – besides the ratio of positive/negative words, the post was analyzed by the Python library of Vader-Sentiment, based on [25]. We found a 95% accuracy between these methods.

TABLE II. EXPERIMENTAL RESULTS FOR THE CONTEXT MODEL

<b><math>UTV</math> 0.7-0.8 batch</b>				
<b><math>UTV_{\kappa 1}</math></b>	<b><math>UTV_{\kappa 2}</math></b>	<b><math>UTV_{\kappa 3}</math></b>	<b><math>UTV_{\kappa 4}</math></b>	<b><math>UTV_{\kappa 5}</math></b>
0.617	0.62	0.59	0.605	0.608
<b><math>STV_{\kappa 1}</math></b>	<b><math>STV_{\kappa 2}</math></b>	<b><math>STV_{\kappa 3}</math></b>	<b><math>STV_{\kappa 4}</math></b>	<b><math>STV_{\kappa 5}</math></b>
0.53	0.465	0.595	0.422	0.443
<b><math>UTV</math> 0.8-0.9 batch</b>				
<b><math>UTV_{\kappa 1}</math></b>	<b><math>UTV_{\kappa 2}</math></b>	<b><math>UTV_{\kappa 3}</math></b>	<b><math>UTV_{\kappa 4}</math></b>	<b><math>UTV_{\kappa 5}</math></b>
0.665	0.642	0.658	0.657	0.641
<b><math>STV_{\kappa 1}</math></b>	<b><math>STV_{\kappa 2}</math></b>	<b><math>STV_{\kappa 3}</math></b>	<b><math>STV_{\kappa 4}</math></b>	<b><math>STV_{\kappa 5}</math></b>
0.43	0.494	0.586	0.442	0.487
<b><math>UTV</math> 0.9-1 batch</b>				
<b><math>UTV_{\kappa 1}</math></b>	<b><math>UTV_{\kappa 2}</math></b>	<b><math>UTV_{\kappa 3}</math></b>	<b><math>UTV_{\kappa 4}</math></b>	<b><math>UTV_{\kappa 5}</math></b>
0.716	0.721	0.722	0.723	0.717
<b><math>STV_{\kappa 1}</math></b>	<b><math>STV_{\kappa 2}</math></b>	<b><math>STV_{\kappa 3}</math></b>	<b><math>STV_{\kappa 4}</math></b>	<b><math>STV_{\kappa 5}</math></b>
0.419	0.505	0.593	0.464	0.454

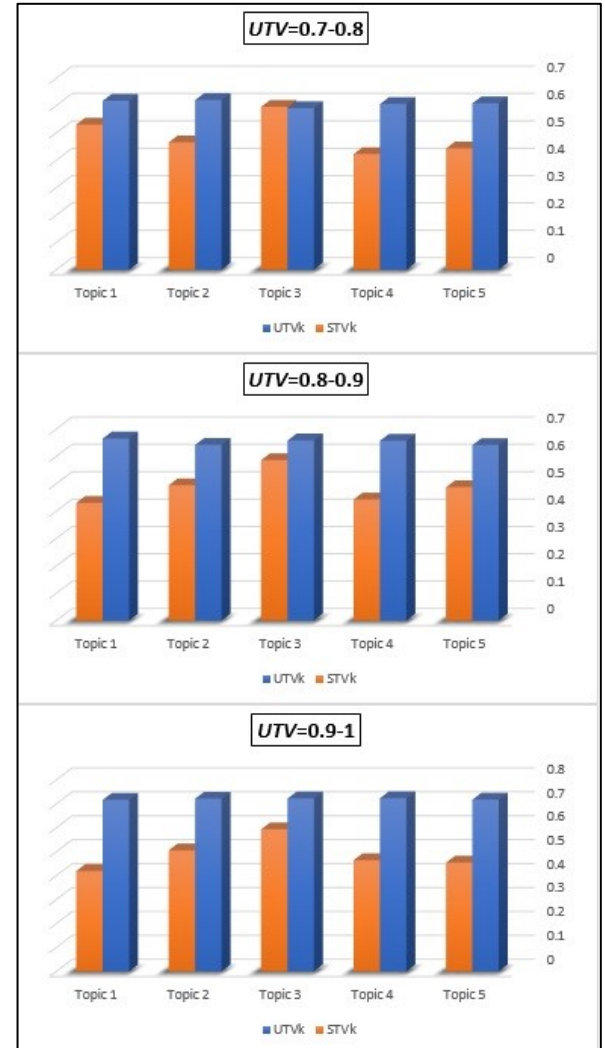


Fig. 5. Comparison of the users' context values in different Trust levels

In both datasets, we divided the results to negative sentiment posts and positive sentiment ones, for the purpose of finding which make a stronger effect on the user trust level. We then calculated the  $\Delta$  (difference) between  $STVA_k$ , which is the trust estimation for posts of a certain user, and the trust level given specifically to the post analyzed, denoted here as  $STV_{post}$ .

The  $\Delta$ 's in the first dataset were relatively very heterogenous, whilst the ones in the second dataset were homogenous. We can see their dispersing in Fig.6.

The results first generally indicated a very negative sentiment regarding the topic of Israeli politics (Dataset A: -0.34; Dataset B: -0.31). In the division between negative sentiment posts and positive ones, we can see that in both datasets there is a relatively bigger effect for positive sentiment posts on the user's trust level. The results of these effects are shown in Table III.

These results can, of course, differ in different Ego-networks and different topics. We can also take into consideration applying different weights to the calculation of the  $UTV_k$ 's, according to the network and preferences of the Ego user, and in topics that are not that vivid or strongly felt such as politics, we may get different sentiment results.

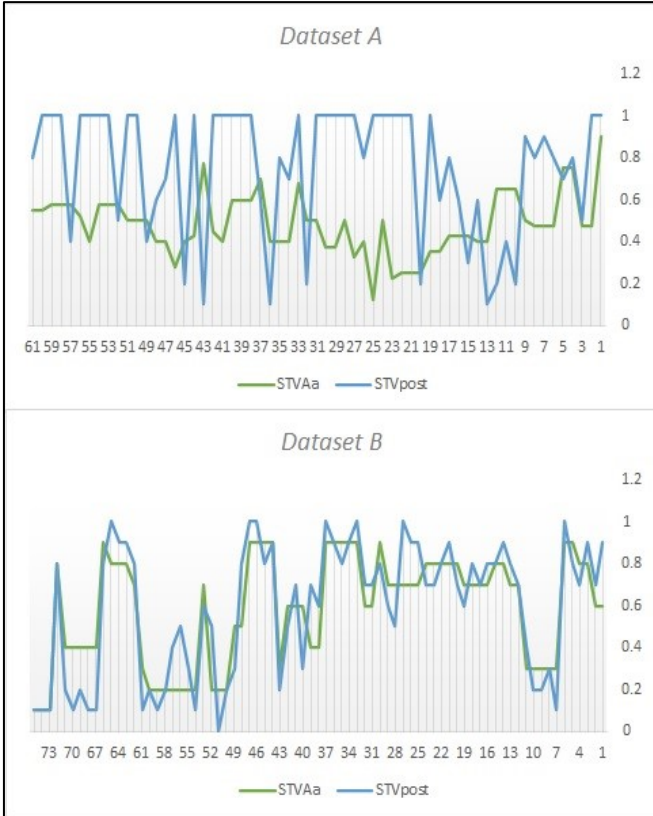


Fig. 6. Difference between  $STV_{post}$  and  $STVA_k$  in the datasets

TABLE III. EXPERIMENTAL RESULTS FOR SENTIMENT ANALYSIS

	Dataset A	Dataset B
No. of users	36	61
No. of posts	61	75
Average Sentiment score (-1 to 1)	-0.34	-0.31
Average effect ( $\Delta$ ) of positive posts	0.3	0.008
Average effect ( $\Delta$ ) of negative posts	0.26	0.004

## VI. DISCUSSION, CONCLUSION AND FUTURE WORK

In this research we presented a Trust-based model that uses context evaluation for preserving privacy in OSN. The model has a main purpose: To every data instance we analyze its proper reliable audience, and users that are less trustworthy in different contexts will be allotted from the data cycle. The results of the experimental part of this paper give us a comprehensive view of the user's subjective trust value, for the purpose of privacy preservation in the Ego network. The less trusted users in different contexts are the ones that the Ego user will most likely to prefer not to show his data in certain categories.

The sentiment analysis on the datasets provided an interesting validation for the choice of parameters chosen for the model- we can see that the course of actions in the OSN can affect the users' trust values. The posts and comments and shares can change the trust values of a certain user, and even a single action can affect this value. It also seems that positive actions effect slightly better than negative ones- they can bring a stronger trust level for a user by the Ego-user. An implementation of such a model will surely provide a much better privacy infrastructure for the network, and a safer environment for its users.

A prospect of future work is clustering category-based trust, in which we aim to create user-clusters based on categories and trust levels. This infrastructure is meant to be a context-safe environment for the Ego-user, in which only context-trusted users will be exposed to data instances of the Ego user in a certain topic.

## VII. APPENDIX- CODE, DATASETS, WEBSITE

For this research, we have conducted an extensive data collection, and also created a software that will help us analyzes the data in the sentiment analysis part.

The program was written in Python, and since the post were in Hebrew, we used the "googletrans" library for translation of the post, and then we used the "vaderSentiment" library for sentiment analysis of the Facebook posts, that were scraped manually, due to the Facebook crawling restrictions. In Fig.7 we can see an example of the program analyzing a post.

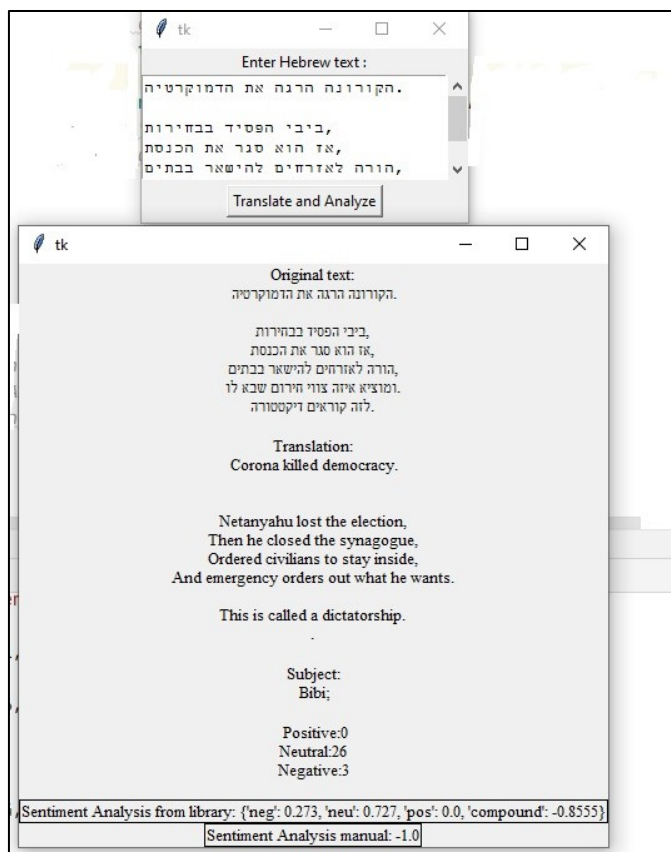


Fig. 7. An example of the program analyzing a sentiment of a post.

The code for the software, including the instructions (in the code remarks) of what necessary libraries to install are in GitHub, and the link to it is in a website page designed for this research- which also presents the project, with the links to the full datasets of the project, and links to previous papers published by us on the model. The URL is: <https://www.cs.bgu.ac.il/~voloch/TrustContextSentimentOSN.html>

## REFERENCES

- [1] Gross, R. and Acquisti, A. "Information revelation and privacy in online social networks." In Proceedings of the 2005 ACM workshop on Privacy in the electronic society, 2005, pp. 71-80. ACM.
- [2] Krishnamurthy, B. and Wills, C. E. "Characterizing privacy in online social networks." In Proceedings of the first workshop on Online social networks. August 2008 , pp. 37-42. ACM.
- [3] Misra, G., and Such, J. M. "How socially aware are social media privacy controls?" Computer, 2016. vol 49(3), pp. 96-99.
- [4] Voloch, N., Levy, P., Elmakies, M., and Gudes, E.. " A Role and Trust Access Control model for preserving privacy and image anonymization in Social Networks". IFIPTM 2019 - 13th IFIP WG 11.11 International Conference on Trust Management, 2019.
- [5] Voloch, N., Levy, P., Elmakies, M., and Gudes, E. " An Access Control model for Data Security in Online Social Networks based on role and user credibility". International Symposium on Cyber Security Cryptography and Machine Learning (CSCML), 2019. Springer, Cham.
- [6] Gudes E. and Voloch N. "An Information-Flow Control model for Online Social Networks based on user-attribute credibility and connection-strength factors", 2nd International Symposium on Cyber Security Cryptography and Machine Learning (CSCML), 2018. Springer, Cham.
- [7] Voloch, N. and Gudes, E. " An MST-based Information Flow Model for Security in Online Social Networks" The 11th IEEE International Conference on Ubiquitous and Future Networks, 2019.
- [8] Li, Y., Li, Y., Yan, Q., and Deng, R. H. "Privacy leakage analysis in online social networks". Computers & Security, 2015. vol. 49, pp. 239-254.
- [9] Sayaf, R., and Clarke, D. "Access control models for online social networks". Social Network Engineering for Secure Web Data and Services, 2012, 32-65.
- [10] Levy, S., Gudes, E., and Gal-Oz, N. "Sharing-habits based privacy control in social networks". In IFIP Annual Conference on Data and Applications Security and Privacy 2016, Springer, Cham. pp. 217-232.
- [11] Cheng, Y., Park, J., and Sandhu, R. "An access control model for online social networks using user-to-user relationships". IEEE transactions on dependable and secure computing, 2016, 13(4), 424-436.
- [12] Cohen, Yehonatan, Daniel Gordon, and Danny Hendler. "Early detection of spamming accounts in large-Scale service provider networks." Knowledge-Based Systems, 2017.
- [13] Zheng, X., Zeng, Z., Chen, Z., Yu, Y., and Rong, C. "Detecting spammers on social networks". Neurocomputing, 159, 27-34., 2015.
- [14] Sandhu, R. S., Coyne, E. J., Feinstein, H. L., and Youman, C. E. "Role-based access control models". Computer, 29(2), 38-47. 1996.
- [15] Patil, Vishwas T., and R. K. Shyamasundar. "Undoing of privacy policies on Facebook." IFIP Annual Conference on Data and Applications Security and Privacy. Springer, Cham, 2017.
- [16] Ali, B., Villegas, W., & Maheswaran, M.. "A trust based approach for protecting user data in social networks." In Proceedings of the 2007 conference of the center for advanced studies on Collaborative research (pp. 288-293). IBM Corp., 2007.
- [17] Lavi, T. and Gudes, E." Trust-based Dynamic RBAC." In Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP) 2016, pp. 317-324.
- [18] Misra, G., Such, J. M., & Balogun, H. "IMPROVE-Identifying Minimal PROfile Vectors for similarity-based access control". In Trustcom/BigDataSE/I SPA, 2016 IEEE (pp. 868-875). IEEE.
- [19] Wang, Yan, Lei Li, and Guanfeng Liu. "Social context-aware trust inference for trust enhancement in social network based recommendations on service providers." World Wide Web 18.1, 2015: 159-184.
- [20] Du, R., Yu, Z., Mei, T., Wang, Z., Wang, Z., & Guo, B.. Predicting activity attendance in event-based social networks: Content, context and social influence. In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing , September, 2014. pp. 425-434.
- [21] Hajian, Sara, Tamir Tassa, and Francesco Bonchi. "Individual privacy in social influence networks." Social Network Analysis and Mining 6.1 2016: 2.
- [22] Atallah, M. J., McDonough, C. J., Raskin, V., & Nirenburg, S. (2000, September). "Natural language processing for information assurance and security: an overview and implementations". In NSPW (pp. 51-65).
- [23] Tsoumas, B., & Gritzalis, D. (2006, April). "Towards an ontology-based security management." In 20th International Conference on Advanced Information Networking and Applications-Volume 1 (AINA'06) (Vol. 1, pp. 985-992). IEEE.
- [24] Louis, A. (2016). Natural language processing for social media.
- [25] Hutto, C. J., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth international AAAI conference on weblogs and social media.