Wireless Pers Commun (2018) 102:2183–2198
https://doi.org/10.1007/s11277-018-5499-6

Positive and Negative Link Prediction Algorithm Based on Sentiment Analysis in Large Social Networks

Debasis Das[1]

Published online: 12 February 2018

Abstract Signed network analysis being one of the greatest disruptive innovations within the last decade has assembled a vast amount of attention of the citizenry. The positions of the users of the signed networks are used by several societies in the world to see the mentality of the users, the current movement of the grocery store and many more things. But even so, in that location is a latent potential of social nets. Ace of the facial expressions that, we were able to determine was about seeing the relationship between the users (i.e., especially, the negative (i.e., −Ve) link in social networks) on the signed network using the stakes that the users work and the reaction of the other users towards it. The anticipation of a negative link (i.e., −Ve) can be applied in the information security field, to observe the aberrations in the largest social networks and further discover the malicious nodes in the larger social network; say, if two nodes are doing things together even though in that respect is no intercourse between them. It can likewise be utilized in improving the recommendation system in social networks as if there is some probability between the two the nodes of being an enemy or disliking each other then we can slay them from each other's recommendation list or could assign a lesser weight to them in a recommended technique. To accomplish all this relationship between the nodes we first need to determine whether the user is posting posts with positive emotion (like happy, excited, etc.) or negative emotion (like angry, sad, and so on), and then that we can further examine the learning ability of the user and utilize it to recommend the people who we have previously separated with the similar personality. For that we have applied the sentiment analysis in social networks, which splits up the users into five simple categories: Highly Positive (i.e., Highly +Ve), Positive (i.e., +Ve), Neutral, Negative (i.e., −Ve) and Highly Negative (i.e., Highly −Ve).

✉ Debasis Das
debasisd@goa.bits-pilani.ac.in

[1] Department of Computer Science and Information Science, Bits Pilani, K. K. Birla Goa Campus, Zuarinagar, Goa 403726, India

Keywords Signed networks · Large social networks · Negative link prediction algorithm · Sentiment analysis · Recommendation system

1 Introduction

With the rapid growth in social network [1–3], there are large scale data of social media [4] available to us. These large data involve the synergy amidst its members, for instance comments and connections. The commentary is always a short paragraph with only a few convictions [5], which are transmitted from one user to another. The link is usually a label with sign value that stands for one user's certain kind of opinion to another, such as conveying support or contradict. We need the social members as vertexes and links as directed edges, the link network can be planned into a graph. Granting to the existing studies the link prediction problem [6–9] can be categorized into two classes: user similarity based matrix and machine learning based methods. The user similarity based matrix is used for recommendation systems [10, 11]. The computing cost is down and can infer top k recommendations. Only one of the major causes behind its drawback is that its performance decreases as the k value becomes large. Hence, in this paper, we are working to have basically five types of links, Highly Positive (i.e., Highly +Ve), Positive (i.e., +Ve), Neutral, Negative (i.e., −Ve) and Highly Negative (i.e., Highly −Ve). Neutral links [12] can be viewed as interactions that hold no diagonal.

We take the social members as vertexes and the interactions between them as links or directed edges, so that the link network can be designed into a graph, as shown in Fig. 1. Signed network analysis [13–17] has attracted increasing attention in the recent years. The Negative links [1, 2] are unwanted properties in online worlds. Most social networking [18–20] sites allow positive links (i.e., Friendships in Facebook and Following on Twitter) and few social web sites allow negative links. In other words, a gap exists between the importance of negative links and their accessibility in real data sets. Consequently, the

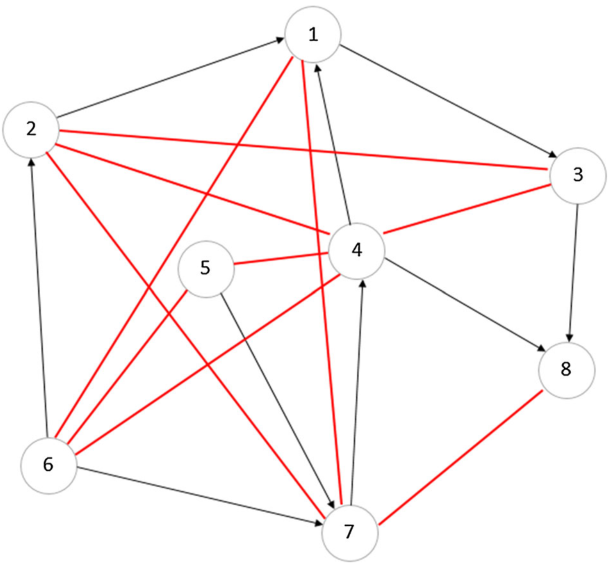


Fig. 1 Black links can be predicted using link prediction algorithms

Negative links could be as important as positive links and it has added value over positive links recent years [21, 22]. The availability of negative links allow various social media applications [23–26] for positive link prediction, recommendation and classification and clustering.

The paper is organised as follows: Sect. 2 provides the related works of negative link prediction in social network. Section 3 presents the objectives for negative link prediction in the social media. Section 4 presents the problem formulation for negative link prediction. Section 5 provides sentiment analysis for negative link prediction. Section 6 presents the Proposed Schemes for negative link prediction. The results is discussed in Sect. 7. Section 8 finally concludes the work.

2 Related Work

Tang et al. [1, 2] used the precision mathematics to predict the negative links in social networks. Their Support Vector Machine(SVM) derivations for their NEgative Link Prediction (NeLP) framework is very extensive. Tang et al. [2] presented a review of signed networks analysis in large social networks and discuss some promising research directions and new frontiers. Tang et al. [3] proposed a new analytical model in social networks which is built stochastically from a node level up.

Beigi et al. [6] developed a model for predicting user relationships from scratch. Usually, and even in their paper, they consider an input, and then predicts links. They established a model to classify the predictions as credible or not credible, based on features extracted from the users. While in [7], the authors have discussed on how social media can be used for disaster relief and situational awareness. The research [8] was quite useful for us, as they simply talked about scaling up link prediction in social networks, because of the huge data sets.

Another very recent and intriguing research in [9]. They innovatively followed social theories in link analysis in unsigned networks. The authors of papers [10] have focused on trust prediction, the reason being the advent of on-line interactions on e-commerce websites, which has increased the importance of predicting user-trust relationship. The paper currently discusses the community detection approach which influence the accessible trust relations and rating similarities in the network to recompense for the lack of labels (known relationships in the network). The novel purpose of the algorithm being able to propagate trust between users, even when they are not closely connected via existing links. Their framework lays emphasis on the concept that the trust values of the central community members can be used as a predictor for relationship between other community members. Overall the whole paper is divided into two phases where first phase discusses about game-theoretic approach and the second phase discusses about SLM (Smart Local Moving) community detection. As a result, their algorithm surpasses any other trust predicting methods.

Unlike any other research on signed network analysis where the sole focus is only on positive links the authors of paper [11] have given their attention to both positive and negative links. Their main aim is the prediction of the sign of links in a newly formed network, which they referred to as the target network, in a reliable and efficient way. Generally in the newly formed network quite less amount of edge sign information is available which is not sufficient to train a good classifier and hence to overcome this issue they have taken the aid from an existing mature signed network which they referred to as the source network that has ample amount of edge sign information within it. They have smartly used the transfer learning approach to pull out the edge sign information from the source network and constructed a feature that can transfer the topological knowledge from the source network to target network. Finally, they have adapted an AdaBoost like transfer learning algorithm which has an instance weighting for model learning with the utilization of more convenient training instances in the source network. Their experimental results of three real large signed social networks have proven that prediction accuracy can be improved by 40% on using transfer learning algorithm over baseline methods.

Chiang et al. [12] have concentrated their research on micro blogging site Twitter. They have considered opinion mining on twitter data and their contributions in the paper include POS specific prior polarity feature and analyse the use of a tree kernel to restrain the need for tedious feature engineering. Further, they have probed a contrasting method of data representation which has proved to be significant over the previously used uni-gram model. Another addition to the paper is that they have declared their results from manually annotated data that is not biased unlike the data collected by using specific search queries. The main aim of the paper is to evaluate the use of linguistic feature like POS tags, and they have used two new resources for pre-processing twitter data namely (a) an emotion dictionary and (b) an acronym dictionary. Overall the conclusion is that with the newly added feature and the tree kernel the performance is almost same, but surpasses the state-of-the-art-base-line. Symeonidis et al. [13] have studied the co-action between the positive and negative relationship and how they affect the structure of on-line social network. They have connected their analyses of theories of signed network from social psychology. They have used classical balanced theory and developed an alternate status theory. Their overall approach is to acclimate theories of social psychology and analyse the signed networks as they began to appear in social computing applications.

3 Objective

Our primary objective is to predict negative link in the social media via Sentiment Analysis or Opinion Mining. The negative link prediction problem in machine learning based method is handled as a categorization task. Support vector is applied for examining how the link's values are affected by the link network structure features. The Sentiment analysis or opinion mining is the computational study of a person's opinion, appraisals, attitude, and emotions towards entities, people, events, events, themes and their attributes which is further sorted out as positive, negative and indifferent. In this procedure the link connection from one user to another user is submitted as an opinion and hence the link prediction becomes similar to that Sentiment Analysis or opinion mining. We offer a novel framework for negative and positive link prediction for large dataset in social networks using Sentiment Analysis to predict the negative links by integrating positive links and content centric user interaction using mathematical formulation. This proposed framework evaluates in real world social media dataset.

4 Problem Statement

Let $G(V, E)$ be a social network, where $\mathrm{V} = v_1, v_2, \ldots\ldots, vm$ be the set of m users and E is the social relationship between m users. A social network can be decomposed into a positive network component $G_p(V, E_p)$ and a negative network component $G_n(V, E_n)$ where E_p and E_n are the sets of positive and negative links. For a given node v_s and a candidate set $C = v1, v2, \ldots, v|C|$, our goal is to predict whether there is a link between v_s and $v_i(v_i \epsilon C)$ and to find a predictive function for v_s. Here we use sentiment analysis to overcome the challenges faced by pre-existing methods.

5 Sentiment Analysis

We will use it to classify the comments of people on social media to five categories:

- Strongly positive (strongly +Ve)
- Positive (+Ve)
- Neutral
- Negative (−Ve), and
- Strongly negative (strongly −Ve)

Since, SVM is a supervised learning algorithm we will be providing a manually classified dataset for it. Since, sometimes it happens that the classification that our algorithm does tend to over-fit or under-fit our dataset which definitely does not give us the correct results. Over-fitting tends to provide more specialized results whereas, under-fitting provides very generalized results. Hence, the over-fitting and under-fitting both are negatives for a classification technique. To overcome this issue, we will also be using regularization in our algorithm to increase generalization in over-fitting cases, while reduce it in the case of under-fitting to provide the classifications with a good amount of approximations.

Nowadays, human language has become quite complex and is very difficult to make the computer learn what is the emotions behind it. For example: "My flight got delayed, Wonderful! #hard-luck #sarcasm #ridiculous", this sentence (without hash-tags) would be classified as positive, but if we look closely we would be able to infer that there is more information present about the emotion of the sentence in the hash-tags. So, if we provide more weightage to the hash-tags rather than the sentence itself while classifying the sentences, we will get a better classification.

6 Prediction of Negative Links using Sentiment Analysis in Social Media

Figure 2 above accurately describes our entire proposed approach in a simplified form. We first collected all the data from social networks, i.e., comments, tweets, etc. We will then categorise them as specified, using Sentiment Analysis, into the 5 categories, namely Highly Positive, Positive, Neutral, Negative and Highly Negative, since human emotions are not very discreet and a range of categories will actually help preserve all the emotions with high accuracy. After categorising them successfully, we remove and add links according our algorithm. The existing and predicted links must satisfy the tolerance ranges of the Extended Structural Balance Theory to maintain balance, as explained earlier. We then separate all the negative links from the obtained network, because accurately finding all the negative links was our primary aim. We calculate the reliability weight of these separated negative links, to show the reliability and accuracy of the obtained negative links. The algorithm and calculations have been explained in the following section in detail.

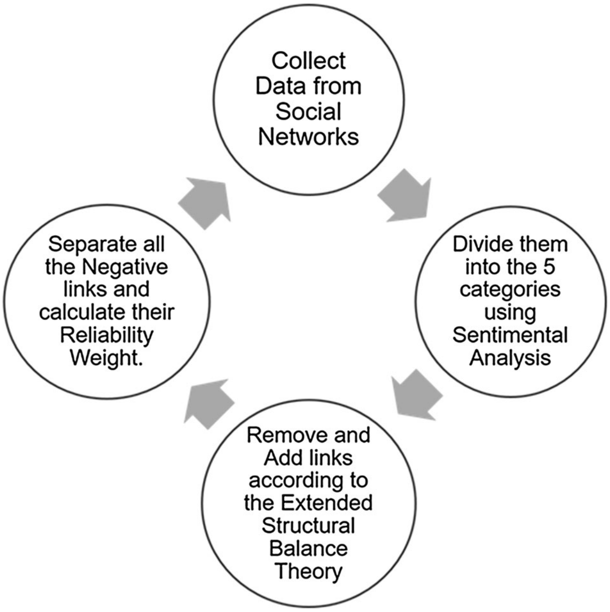


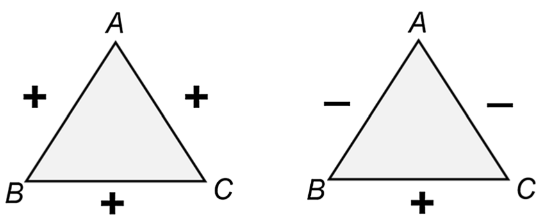
Fig. 2 Flow chart for proposed algorithm

6.1 Structural Balance Theory and Social Status Theory

As mentioned earlier, there are two theories, Structural Balance Theory(SBT) and Social Status Theory(SST), that are support systems for this link prediction problem. They provide rules, that are results of common human behavior, following which can result in balanced networks after prediction. In a social system, relations among individuals characterize the interactions. So the overall sentiments among different agents/users, determines the balance of a social system. There are mainly two types of links considered, positive and negative.

In SBT, when we consider a triad of three nodes, as shown in Fig. 3, the polarity of two links can cause an effect on the third link, i.e., if we know AB and AC, we can determine BC based on Structural Balance Theory. The SBT works on the popular belief that "my friend's friend is my friend" and "my enemy's enemy is my friend", as shown in Fig. 3. But since human emotions are not absolute, as in our case, we would have to modify this for the 5 categories (shown in Fig. 4) we are considering, which is explained in the following sections.

Fig. 3 Balanced triads according to structural balance theory



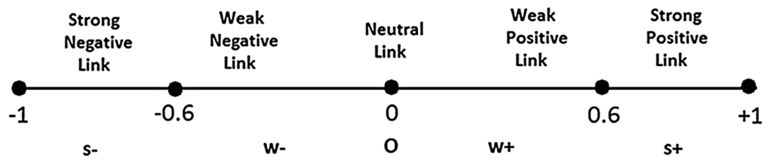


Fig. 4 The strength values of different types of links

6.2 Extended Structural Balance Theory

We were giving a lot of thought on how this can be done, when we found a research on what they called an extended balance theory [1, 11]. The authors of [11] did an extensive research with experimental results on the Extended Structural Balance Theory (ESBT), for the arbitrary relationship strengths, as we were looking for. The only assumption they considered to be that, in resolving tensions within imbalanced relationships, people tend to avoid the effort of changing relationships if possible.

As noted earlier, to model the distinction, we conceive a scenario where relationships take in varying intensity levels. A close friend or family tie will be interpreted by a strong positive link (s+), i.e. strong trust constructs, and utter dislike or hatred will be interpreted as a strong negative link (s−), i.e., strong distrust constructs. Nevertheless, there will be many other types of trust relationships in between the spectrum of (strong) trust and (strong) distrust. In our example, we will be considering three more types in between them, namely weak positive links (w+), weak negative links (w−), and neutral links (o). Weak positive links/ties can be considered as partial trust, or the feeling for an acquaintance. I might know that person, and also the fact that he is courteous, only it may not be the “strong friendship/trust” case. The Same can be studied with the weak negative links/affiliations. I might experience a minor disagreement, or some negative bias, but not really “hatred” towards him and neutral relationships are quite relatable too, as the one which is

Table 1 The tolerance ranges for the triad to be balanced

(A, B)	(A, C)	(B, C)’s Tolerance
s+	s+	<s+, w+>
s+	w+	<s+, 0>
s+	0	<s+, w−>
s+	w−	<0, s−>
s+	s−	<w−, s−>
w+	w+	<s+, w−>
w+	0	<s+, s−>
w+	w−	<w+, s−>
w+	s−	<0, s−>
0	0	<s+, s−>
0	w−	<s+, s>
0	s−	<w+, s−>
w−	w−	<s+, s->
w−	s−	<s+, s−>
s−	s−	<s+, s−>

unbiased, i.e., non-negative, non-positive relationship with no view at all. Immediately that we defined the categories, our relationship spectrum would see something like Table 1.

As we know, the Structural Balance Theory [11, 12] considers triads, and two links of the triad always cause an influence over the third one, be it good or bad, making the triad balanced or unbalanced (shown in Table 2). Since we had only positive and negative links earlier, it was easier to pinpoint the unbalanced states, as there were only a few. Now when we have five different types of links, instead of just positive and negative, the number of balanced and unbalanced combinations will significantly increase. The authors of [1] have verified that if we know any two links of the triad, the third link will have a range of values according to the spectrum. It the third link goes out of the range, there will be stress in the relationships, and the triad will be unbalanced (shown in Table 2). So we define tolerance as the range of relations the third link of triad can take, for the entire triad to be balanced. The authors of [1] have experimentally verified this, and proposed the tolerance values in Table 1 for the spectrum in Fig. 4.

Now that we have the ESBT, we will combine the output from Sentiment Analysis with this, and determine the negative links. The Sentiment Analysis will initially analyze the links from the data we provide, and give values ranging from -1 to $+1$, as output, corresponding to each link (Lets call this value SA_{Value}). We will be assigning the values of the different types of links we have to categorize the links, as shown in Fig. 4, and then separate the negative links, and calculate the reliability weight matrix [2, 9, 10], according to the algorithm.

6.3 Balanced Negative Sample Set Construction

Social networks have enabled a vast diversity of relations between users, such as follower relations in twitter. The increasing availability of large scale online social network data (i.e., twitter data set) is useful not only various tasks in social network analysis, such as community detection in social network and negative link prediction in social networks [29] but it is also helpful for traditional data mining tasks such as feature selection and recommendation. As we have stated, instead of just positive or negative, we are considering 5 different types of links (Strongly positive, Weakly positive, neutral, weakly negative, strongly negative). In the previous sections, we obtained the SA_{value} from the sentiment analysis, which was used for the categorization of all the links. Now we have processed the categorized data, which is where this algorithm comes into play. The network G with categorized links is our input. We first copy all the links from G to NS for further changes (Step 1–4). Now we analyze all the links, and remove the links which do not satisfy the ESBT (Step 5). Then we add new links that might exist according to ESBT (Step 6). Since our aim was to effectively obtain negative links, we now separate all the negative links from the network by deleting positive and neutral links (Step 7–9), and calculate reliability

Table 2 Examples for unbalance triads

The argument fot stress in a triad
(s+s+s−) (s+s+w−) (s+w+s−) (w+w+s−): My two friends cannot get along with each other
(s+s+O): My two close friends don't friend each other
(s+s−O): My enemy's close don't pick a side

weight for the resulting links (Step 10–12). The details of reliability weight calculation are explained in detail in the next section.

Algorithm 1 Balanced Negative Sample Set Construction

Input : The network G with different signs from Sentiment Analysis

Output : Balanced Negative sample set NS and the reliability weight matrix W **Step 1.** Initialize $NS = \emptyset$

Step 2. for all $G_{ij} \neq \emptyset$ do

$NS = NS \cup < u_i, u_j > fromG$

end for

Step 3. Remove samples $< u_i, u_j >$ from NS if $u_i - u_j$ is in any triads of G that does not satisfy the Extended Balance Theory.

Step 4. Add samples $< u_i, u_k >$ into NS if $u_i - u_k$ can make all triads that involve

u_i and u_k in G satisfying the Extended Balance Theory (the lower limit of the tolerance range)

Step 5. for all $< u_i, u_j > \in NS$ do

Step 6. Separate Negative links by deleting all neutral, weak positive, and strong positive links from NS

Step 7. end for

Step 8. for all $< u_i, u_j > \in NS$ do

Step 9. Calculate a reliability weight $W_i j$

Step 10.end for

The proposed algorithm for negative links and positive link construction using a mathematical formulation to construct link prediction framework. The strong correlation between strong negative interaction and strong negative links suggest that strong negative interactions are likely to have negative links.

6.4 Reliability Weight Calculation

Observations from data analysis indicate that negative sample candidates [22–24] with more negative interactions are more likely to have negative links, and are therefore more likely to be reliable. Therefore, we associate each $< u_i, u_j >$ with a reliability weight WT_{ij}, which is defined as follows:

$$WT_{ij} = f(NS_{ij} * |SA_{Value}|)$$

if the pair $<u_i, u_j> \in NS$ has negative interactions, we define the reliability weight as a function f of the number of negative interactions NS_{ij} where $f(x) \in [0, 1]$ is a non-decreasing function of x, multiplied by the absolute value of SA_{Value}, i.e., the value obtained from Sentiment Analysis. We did this because the strong negative links will obviously contribute more to the reliability than a weak positive link. So by multiplying SA_{Value} we are ensuring it. Because more negative interactions [34–37] two users have, the more likely it is that a negative link exists between them.

7 Results

The positive link prediction [32, 33] is dependent on "typical" behaviour of social networks such as triadic closure, it is natural to explore similar properties of negative links [25, 26] with respect to other positive links, and content-centric interactions. Such an understanding lays the groundwork for a meaningful link-prediction model [29–31]. For the purpose of this study, we collected datasets [27] from twitter, that explicitly allow users to express the Highly Positive (i.e., Highly +Ve), Positive(i.e., +Ve), Neutral, Negative(i.e., −Ve) and Highly Negative(i.e., Highly −Ve) [28].

For our experimentations we have taken the datasets of twitter. We applied our sentiment analysis algorithm on these datasets to classify the data sets into the five defined categories (i.e., Highly Positive (i.e., Highly +Ve), Positive(i.e., +Ve), Neutral, Negative(i.e., −Ve) and Highly Negative(i.e., Highly −Ve)).

As per the experiments we conducted, we plotted the graph below and were able to infer the results which is shown in Fig. 5.

In our original graph (shown in Table 3) we have the following attributes:

Node1 The node that generates the tweet on the social network.

Node2 The node that reacts to the tweet on the social network by say, re-tweeting it or giving negative comments.

Sentiment It is the quantified probable sentiment of the link between node1 and node2 holds, based upon node1s tweet.

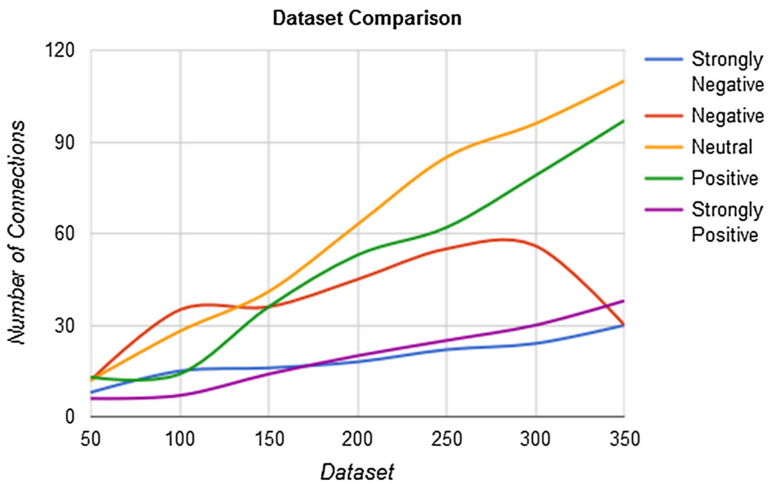


Fig. 5 Dataset comparison

Table 3 Original graph

Node 1	Node 2	Sentiment	Tweet
1	5	2	@swithchfoot: http://twitpic.com/2ylzl - Awww, that's a bummer. You shoulda got David carr of Third Day do it. ;D
1	5	2	is upset that he can't update his Facebook by texting it... and might cry as a result School today also. Blah!
1	2	4	@Kenichan 1 dived many times for the ball. Managed to save 50% The rest go out of bounds
2	3	2	my whole body feels itchy and like its on fire
2	1	2	@nationwideclass no, it's not behaving at all. i'm mad. why am i here? because 1 can't see you all over there.
2	5	3	@kwesidei not the whole crew
3	1	4	Need a hug
3	4	4	@LOLTrish hey long time no see! Yes... Rains a bit, only a bit LoL, l'm fine thanks, how's you?
3	5	3	@Tatiana_K nope they didn't have it
3	5	3	@twittera que me muera ?
3	1	3	spring break in plain city... it's snowing
3	2	3	l just re-pierced my ears
4	5	2	@caregiving 1 couldn't bear to watch it. And l thought the UA loss was embarrassing.....
4	1	2	@octolinz16 It it counts, idk why l did either. you never talk to me anymore
5	4	4	@smarrison i would've been the first, but i didn't have a gun. not really though, zac snyder's just a doucheclown
5	2	4	@iamjazzyfizzle l wish l got to watch it with you!! l miss you and @iamlilnicki how was the premiere?!
5	4	1	Hollis' death scene will hurt me severely to watch on film wry is directors cut not out now?
5	4	3	about to file taxes

Tweet It is the node1s tweet upon which the sentiment of the link is being predicted.

7.1 Methodology

We consider the original graph as the directed graph such that the link originates from $node_1$ and points to $node_2$. We use the following methodology to achieve our predictions (Table 4):

Step1 We take a row of the original graph, table (without tweet) says $node_1$ (A) and $node_2$ (B).

Step2 We take the next row of the same table (without tweet) and check that the $node_1$ (A) should be same while the $node_2$ (C) should be different in the rows. This is done to ensure that we consider a triad and not a duo, as there can be many instances of the two nodes interactions.

Table 4 Predicted sub-graph

Node1	Node2	Sentiment
1	4	4.0
4	2	3.5
5	1	3.0
5	1	3.0
5	1	2.0
4	2	4.0
2	4	2.5
2	4	3.5

Step3 We then search in our graph for a link between the B & C. If it is present, then we dont bother predicting the link between them , but if the link is not present then we predict the link between them by taking the average of the sentiments of the two extracted rows.

Step4 Repeat the process for all the combinations of the triad in the graph to get the maximum predictions.

7.1.1 Predicted Graph

The predicted graph given below:

Node1 This node is the B, as mentioned in the above methodology.

Node2 This node is the C, as mentioned in the above methodology.

Sentiment It is the average of the sentiments of the extracted rows from which we were able to find the triad which was not present earlier in our graph.

7.1.2 Example

The following comparisons took place in our algorithm to achieve the above predictions:

(1, 5, 5): Not a valid triad, due to presence of 2 nodes only. (1, 5, 2): But (5, 2) edge is already present so, no prediction required. (1, 5, 2): But (5, 2) edge is already present so, no prediction required. (2, 3, 1): But (3, 1) edge is already present so, no prediction required. (2, 3, 5): But (3, 5) edge is already present so, no prediction required. (2, 1, 5): But (1, 5) edge is already present so, no prediction required.

(3, 1, 4): Since, (1, 4) edge is not present so, we predict its sentiment as average of the to put it within the range of the two sentiments. Hence, (4.0 + 4.0) / 2 = 4.0

(3, 1, 5): But (1, 5) edge is already present so, no prediction required. (3, 1, 5): But (1, 5) edge is already present so, no prediction required. (3, 1, 1): Not a valid triad, due to presence of 2 nodes only. (3, 1, 2): But (1, 2) edge is already present so, no prediction required. (3, 4, 5): But (4, 5) edge is already present so, no prediction required. (3, 4, 5): But (4, 5) edge is already present so, no prediction required. (3, 4, 1): But (4, 1) edge is already present so, no prediction required.

(3, 4, 2): Since, (4, 2) edge is not present so, we predict its sentiment as average of the to put it within the range of the two sentiments. Hence, (4.0 + 3.0) / 2 = 3.5 (3, 5, 5): Not a valid triad, due to presence of 2 nodes only. (3, 5, 1): Since, (5, 1) edge is not present so, we predict its sentiment as average of the to put it within the range of the two sentiments. Hence, (3.0 + 3.0) / 2 = 3.0 (3, 5, 2): But (5, 2) edge is already present so, no prediction required. (3, 5, 1): Since, (5, 1) edge is not present so, we predict its sentiment as average of the to put it within the range of the two sentiments. Hence, (3.0 + 3.0) / 2 = 3.0 (3, 5, 2): But (5, 2) edge is already present so, no prediction required. (3, 1, 2): But (1, 2) edge is already present so, no prediction required. (4, 5, 1): Since, (5, 1) edge is not present so, we predict its sentiment as average of the to put it within the range of the two sentiments. Hence, (2.0 + 2.0) / 2 = 2.0 (5, 4, 2): Since, (4, 2) edge is not present so, we predict its sentiment as average of the to put it within the range of the two sentiments. Hence, (4.0 + 4.0) / 2 = 4.0 (5, 4, 4): Not a valid triad, due to presence of 2 nodes only. (5, 4, 4): Not a valid triad, due to presence of 2 nodes only. (5, 2, 4): Since, (2, 4) edge is not present so, we predict its sentiment as average of the to put it within the range of the two sentiments. Hence, (4.0 + 1.0) / 2 = 2.5 (5, 2, 4): Since, (2, 4) edge is not present so, we predict its sentiment as average of the to put it within the range of the two sentiments. Hence, (4.0 + 3.0) / 2 = 3.5 (5, 4, 4): Not a valid triad, due to presence of 2 nodes only.

7.2 Performance Analysis

For our experimentations, we have taken the 350 datasets of twitter. We applied our sentiment analysis technique on these 350 datasets to classify the data sets into the five defined categories (i.e., Highly Positive (i.e., Highly +Ve), Positive(i.e., +Ve), Neutral, Negative(i.e., −Ve) and Highly Negative(i.e., Highly −Ve)). Our simulations were conducted on a PC with Intel Core (TM) i5-5200U CPU @ 2.20 GHz, 16 GB of Installed Memory (RAM) and 2 TB of hard disk. The size of compact twitter datasets of 350 items. As Fig. 6 shows, our proposed scheme (PS) trained on the 350 datasets of twitter formed by sentiment analysis and this scheme achieve the highest prediction accuracy for twitter dataset. To share with large amount of the social network dataset, we nominated a novel scheme that chooses the small amount of training data while keeping high accuracy of prediction. The proposed technique based on the three social theories (i.e., Structural Balance Theory(SBT), Social Status Theory (SST), and Extended Structural Balance Theory (ESBT)) to predict the links (i.e., (i.e., Highly Positive (i.e., Highly +Ve), Positive(i.e., +Ve), Neutral, Negative (i.e., −Ve) and Highly Negative (i.e., Highly −Ve)). These social theories will work better than some existing algorithms because existing

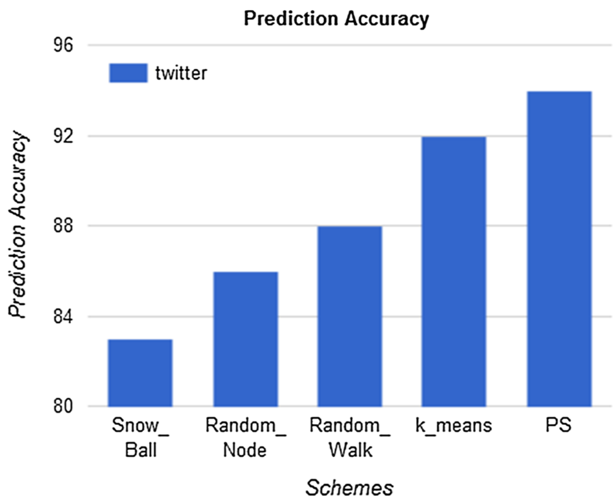


Fig. 6 Prediction accuracy based on different schemes. The size of Twitter training data set is 350

algorithms concerning only one single hypothesis. Experimental results of typical twitter 350 data sets show that our proposed scheme can deal with large signed networks and consistently outperforms other method(i.e., snowball,random node, random walk, and k means).

By sentiment analysis, the user's mentality can divide into five simple categories [37]:(i.e., Highly Positive (i.e., Highly +Ve), Positive(i.e., +Ve), Neutral, Negative(i.e., −Ve) and Highly Negative(i.e., Highly −Ve). These five classifier labels can help to build a negative link between the users in twitter. Moreover, to deal with large social network data sets [37], we employ a classifier that maintains great accuracy of the prediction. Our experimental results of typical large social network data sets and consistently outperforms other schemes ((i.e., snowball,random node, random walk, and k means) because we consider the user mentality can be divide into five categories. We introduce efficient and effective strategies to select training instances from massive data based on the five categories Highly Positive (i.e., Highly +Ve), Positive(i.e., +Ve), Neutral, Negative(i.e., −Ve) and Highly Negative(i.e., Highly −Ve), which allows maintaining the representative link structure and information of the relationships in large-scale social network, and thereby the prediction performance is promised. In the training phase, we use Support Vector Machines (SVM) and sentiment analysis to ensure good generalization performance in large social networks.

8 Conclusion

The social network analysis in the form of the negative links prediction turns out to be useful in a huge number of areas for example, recommendation systems in a societal mesh. Since, sites like Facebook, Twitter, etc. do not provide with a feature of disliking a post of another user, but rather they let us write the comments in relation with that post, it seemed as a potent feature that can be applied for predicting the relations between the users. As traced in this paper that, we read the comments related to a post and we classify them into Highly Positive (i.e., Highly +Ve), Positive (i.e., +Ve), Neutral, Negative (i.e., −Ve) and Highly Negative (i.e., Highly −Ve) categories in order to quantify the relation between the users who posted, the post and the comments. This methodology of extracting the content-centric data along with the previously described social theories, i.e., balance theory and status theory, can be applied to anticipate the type of connections between the users. We ultimately want to propose that this method of predicting the relation between the users works under very close proximity of the real world scenario as the social balance theory, status theory, and the Sentiment Analysis is prevalent in the tangible world and make very exact results of social net analysis. The way we respond to the behavior of another person majorly describes the future relations with that mortal and so does our experiments suggests.

Acknowledgement This research work is fully supported by Early Career Research Award(ECRA) from SERB, DST, Govt. of India, India (Project #: ECR/2015/000256/ES).

References

1. Tang, J., Chang, S., Aggarwal, C., & Liu, H. (2015). Negative link prediction in social media. In *Proceedings of the 8th ACM international conference on web search and data mining* (pp. 87–96).

2. Tang, J., Chang, Y., & Liu, H. (2014). Mining social media with social theories: A survey. *ACM SIGKDD Explorations Newsletter*, *15*(2), 20–29.
3. Tang, J., Chang, Y., Aggarwal, C., & Liu, H. (2016). A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)*, *49*(3), 1–42.
4. Wen, S., Haghighi, M. S., Chen, C., Xiang, Y., Zhou, W., & Jia, W. (2015). A sword with two edges: Propagation studies on both positive and negative information in online social networks. *IEEE Transactions on Computers*, *64*(3), 640–653.
5. Song, D., Meyer, D. A., & Tao, D. (2015). Efficient latent link recommendation in signed networks. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1105–1114).
6. Beigi, G., Tang, J., Wang, S., & Liu, H. (2016). Exploiting emotional information for trust/distrust prediction. In *Proceedings of the 2016 SIAM international conference on data mining* (pp. 81–89).
7. Liu, X., Lin, H., & Yang, Z. (2015). Predicting user relationship from scratch. *Chinese National Conference on Social Media Processing* (pp. 176–183). Singapore: Springer.
8. Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. In *Sentiment analysis and ontology engineering* (pp. 313–340). Springer.
9. Duan, L., Aggarwal, C., Ma, S., Hu, R., & Huai, J. (2016). Scaling up link prediction with ensembles. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining* (pp. 367–376).
10. Beigi, G., Tang, J., & Liu, H. (2016). Signed Link Analysis in Social Media Networks. In *ICWSM* (pp. 539–542).
11. Wang, G. N., Gao, H., Chen, L., Mensah, D. N., & Fu, Y. (2015). Predicting positive and negative relationships in large social networks. *PloS One*, *10*(6), e0129530.
12. Chiang, K. Y., Hsieh, C. J., Natarajan, N., Dhillon, I. S., & Tewari, A. (2014). Prediction and clustering in signed networks: A local to global perspective. *The Journal of Machine Learning Research*, *15*(1), 1177–1213.
13. Symeonidis, P., Tiakas, E., & Manolopoulos, Y. (2010). Transitive node similarity for link prediction in social networks with positive and negative links. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 183–190). ACM.
14. Ye, J., Cheng, H., Zhu, Z., & Chen, M. (2013). Predicting positive and negative links in signed social networks by transfer learning. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1477–1488).
15. Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World wide web* (pp. 641–650).
16. Liben-Nowell, D., & Kleinberg, J. (2007). The link prediction problem for social networks. *Journal of the Association for Information Science and Technology*, *58*(7), 1019–1031.
17. Al Hasan, M., & Zaki, M. J. (2011). A survey of link prediction in social networks. *Social Network Data Analytics* (pp. 243–275). Boston, MA: Springer.
18. Modha, J. S., Pandi, G. S., & Modha, S. J. (2013). Automatic sentiment analysis for unstructured data. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(12), 91–97.
19. Poongodi, S., & Radha, N. (2013). Classification of user Opinions from tweets using Machine Learning Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(9).
20. Yuan, G., Murukannaiah, P. K., Zhang, Z., & Singh, M. P. (2014). Exploiting sentiment homophily for link prediction. In *Proceedings of the 8th ACM Conference on Recommender systems* (pp. 17–24).
21. Liu, F., Liu, B., Sun, C., Liu, M., & Wang, X. (2015). Improving link prediction in social networks by user comments and sentiment lexicon. In *Chinese computational linguistics and natural language processing based on naturally annotated big data* (pp. 356–365). Springer.
22. Cheng, K., Li, J., Tang, J., & Liu, H. (2017). Unsupervised sentiment analysis with signed social networks. In *AAAI* (pp 3429–3435).
23. Wang, S., Tang, J., Aggarwal, C., & Liu, H. (2016). Linked document embedding for classification. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 115–124).
24. Wang, S., Tang, J., Aggarwal, C., Chang, Y., & Liu, H. (2017). Signed network embedding in social media. In *Proceedings of the 2017 SIAM international conference on data mining* (pp. 327–335).
25. Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1361–1370).

26. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015, May). Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on World Wide Web* (pp. 1067–1077).
27. Sharma, P., Singh, U. K., Sharma, T. V., & Das, D. (2016). Algorithm for prediction of links using sentiment analysis in social networks. In *Proceedings of the 7th international conference on computing communication and networking technologies* (pp. 29).
28. Yuksel, M., & Gen, Y. (2017). Adaptive modulation for completion time minimization in wireless broadcast networks. *AEU-International Journal of Electronics and Communications*, *72*, 72–78.
29. de Abreu, C. C. E., Duarte, M. A. Q., & Villarreal, F. (2017). An immunological approach based on the negative selection algorithm for real noise classification in speech signals. *AEU-International Journal of Electronics and Communications*, *72*, 125–133.
30. Zhang, C., Xu, W., Ma, Z., Gao, S., Li, Q., & Guo, J. (2015). Construction of semantic bootstrapping models for relation extraction. *Knowledge-Based Systems*, *83*, 128–137.
31. Gao, S., Luo, H., Chen, D., Li, S., Gallinari, P., Ma, Z., et al. (2013). A cross-domain recommendation model for cyber-physical systems. *IEEE Transactions on Emerging Topics in Computing*, *1*(2), 384–393.
32. Ma, Z., Xie, J., Li, H., Sun, Q., Si, Z., Zhang, J., et al. (2017). The role of data analysis in the development of intelligent energy networks. *IEEE Network*, *31*(5), 88–95.
33. Zhang, C., Si, Z., Ma, Z., Xi, X., & Yin, Y. (2016). Mining sequential update summarization with hierarchical text analysis. In *Mobile Information Systems* (pp. 1–10).
34. Shelke, N., Deshpande, S., & Thakare, V. (2017). Domain independent approach for aspect oriented sentiment analysis for product reviews. In *Proceedings of the 5th international conference on frontiers in intelligent computing: Theory and applications* (pp. 651–659). Singapore: Springer.
35. Ma, Z., Xue, J. H., Leijon, A., Tan, Z. H., Yang, Z., & Guo, J. (2016). Decorrelation of neutral vector variables: Theory and applications. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(1), 129–143.
36. Martnez, V., Berzal, F., & Cubero, J. C. (2017). A survey of link prediction in complex networks. *ACM Computing Surveys (CSUR)*, *49*(4), 69.
37. Das, D., & Sharma, P. (2017). Algorithm for prediction of negative links using sentiment analysis in social networks. In *13th international wireless communications and mobile computing conference (IWCMC)* (pp. 1570–1575).

Debasis Das received his Ph.D. in Computer Science and Engineering from Indian Institute of Technology (IIT) Patna, Bihar, India. His thesis work was on Vehicular Ad-hoc Networks in (VANET). He received his M. Tech in Computer Science and Engineering degree from School of Computer Engineering, KIIT University, Bhubaneswar, Orissa, India. He joined BITS Pilani, K. K. Birla Goa Campus in 2016 and before joining BITS, he was working as an Assistant Professor in NIIT University, India. His research interests include Computer Networks, Vehicular Network, Cloud Computing, internet of Things(IoT), Algorithm, Network Security and Cellular Automata.