
CE7490 Project: Benchmarking Algorithms for Weight Prediction in Weighted Signed Networks

Ruihang Wang

School of Computer Science Engineering
Nanyang Technological University
ruihang001@e.ntu.edu.sg

Meng Shen

School of Computer Science Engineering
Nanyang Technological University
meng005@e.ntu.edu.sg

Yihang Li

School of Electrical and Electronic Engineering
Nanyang Technological University
leah9704@gmail.com

Abstract

The abstract paragraph should be indented 1/2 inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

With the rapid development of information technology, social network has become a significant part of modern life. As remarkable number of messages reflecting opinion and attitude flowing on the network, social network analysis gradually gains popularity in a wide range of fields including political science, social psychology and business analysis. Social network analysis offers a tool to seek for insight through the structure of social networks but is limited by the complexity. Users are developing rich relationships through networks while analysis tools are generally reducing the relationships to simple links. To cover the gap between simplicity of network simulation and complexity of real relationships is a remarkable research topic.

In social network, people show different attitudes towards connection to others. The richness of a social network contains positive and negative interactions. The main part of connections are positive ones, including connecting with friends, fans, and followers. There also exist negative connections, including scammers, political enemy and opponents. These connections can be used to predict relationships between independent users. For example, a politician might look at such people as potential voters for him if he has strength in that particular topic of interest. In the same vein, we might wish to estimate how much a person P1 agrees/disagrees with people who are already positive or negative about a product in order to quantify if they are more likely to be target customers of the product. Moreover, edge weight prediction may be useful to improve traditional tasks in signed networks such as node ranking[1], anomaly detection[2][3], network analysis[4][5], community detection[6], information diffusion[7] [8] and sentiment prediction[9]. Therefore, the prediction of edge weights in WSNs can be advantageous in various tasks, both when WSNs are explicit and implicit.

To better understand theories and master practical skills, we investigate and implement a common set of algorithms and evaluate their performance on real-world datasets. The rest of the paper is structured as follows: Section 2 presents related work in this topic. Section 3 describes the overview of our project. Section 4 formulates the problem of weight prediction in weighted signed networks. Section

5 conducted experiments on real-world datasets using different algorithms. Section 6 evaluates the performance of all tested methods. The conclusion is summarized in Section 7.

2 Related Work

Edge Sign Prediction in SSNs The very early work in SSN mainly based on balance and status. The first of these theories is structural balance theory, which originated in social psychology in the mid-20th-century. As formulated by Heider in the 1940s[10], and subsequently cast in graph-theoretic language by Cartwright and Harary[11], structural balance considers the possible ways in which triangles on three individuals can be signed, and posits that triangles with three positive signs and those with one positive sign are more plausible — and hence should be more prevalent in real networks — than triangles with two positive signs or none. Balanced triangles with three positive edges exemplify the principle that “the friend of my friend is my friend,” whereas those with one positive and two negative edges capture the notions that “the friend of my enemy is my enemy,” “the enemy of my friend is my enemy,” and “the enemy of my enemy is my friend.” Structural balance theory has been developed extensively in the time since this initial work[12], including the formulation of a variant “weak structural balance” proposed by Davis in the 1960s [13] as a way of eliminating the assumption that “the enemy of my enemy is my friend”. Weak structural balance posits that only triangles with exactly two positive edges are implausible in real networks, and that all other kinds of triangles should be permissible.

Balance theory can be viewed as a model of likes and dislikes. However, as Guha et al. observe in the context of Epinions[14], a signed link from P1 to P2 can have more than one possible interpretation, depending on P1’s intention in creating the link. In particular, a positive link from P1 may mean, “P2 is my friend,” but it also may mean, “I think P2 has higher status than I do.” Similarly, a negative link from P1 to P2 may mean “P2 is my enemy” or “I think P2 has lower status than I do.”

Later work have developed features and models to predict the sign of an edge random-walk and trust-propagation algorithms[15][14], and using social interaction information for prediction[16][17]. Metric Multidimensional Scaling (MDS) assigns an m-dimensional position to vertices in a SSN or WSN to minimize ‘stress’, based on extended balance theory[18]. It then uses the metric distance between two vertices to predict the sign of an edge between them. All of these papers predict edge sign in unweighted SSNs, while fairness and goodness method predicts the weight of an edge along with its sign.

Edge Weight Prediction in Social Networks There is substantial work on predicting edge weights in unsigned social networks. Communication based features have been shown to be very important in quantifying the edge weights between users[19][20][21][22], but since we only have the WSN without any communication information, we compare with non-communication based techniques. These include baselines such as reciprocal edge weight[19] and triadic balance and status measures[20][23]. Two popular unsupervised algorithms are EigenTrust[24] and TidalTrust[25]. EigenTrust calculates a global value of trust for each vertex by finding the left eigenvector of a normalized trust matrix. TidalTrust calculates the trust from a source to a sink, by recursively propagating trust via the sink’s predecessors till it reaches the source. A very recent work in this direction is trustiness and trustworthiness[26]. However, these papers deal with edge weight prediction in unsigned social networks, while WSN is a new problem deriving from SSN. So in this paper, these techniques are implemented and compared with fairness and goodness method as benchmark models.

3 Project Overview

In this project, we extensively investigate and experiment methods for edge weight prediction in weighted signed networks. All algorithms are tested and evaluated on published real-world datasets. Moreover, we try our best to improve the performance on some traditional methods. The finished works are summarized as follows:

1. Literature review on OSN and select a topic about predicting weight of edges for weighted signed network.
2. Investigate the state-of-the-art algorithms *fairness-goodness* in[27] and studied a common set of baselines for weight prediction.

3. Conducted experiments on each algorithm and reproduce the results mentioned in original paper using real-world dataset.

Experimental 1 - Removing one edge prediction : When we remove one of the edges from the network and try to predict it from the rest of network structure one at a time, we compared fairness and goodness metrics with other typical methods and noticed that fairness and goodness method produces almost the best result.

Experimental 2 - Removing N %-out edge prediction : When we remove N% of the edges from the network and try to predict them from each of the individual features, we see that in all cases, fairness and goodness metrics are the best.

4. Evaluate results of different methods.

4 Problem formulation

5 Experiments implementation

5.1 Reciprocal

As a baseline of weight prediction, reciprocal algorithm takes the weight of (v, u) as equal to the weight of (u, v) . When the reciprocal edge doesn't exist, the weight of that edge is set to 0. Therefore, the weight of (u, v) is predicted by:

$$W(u, v) = \begin{cases} W(v, u), & \text{if } u \rightarrow v \text{ exist} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

5.2 Triadic balance

This definition is derived directly from balance theory according to []. The algorithm takes the average product of edge weights for all incomplete triads that the edge (u, v) is a part of. Incomplete triads are triads that would form involving edge (u, v) after it is created. To be specific, let U_n and V_n denotes the set of neighbors of vertex u and v , respectively. To find all possible triads, vertexes with both connections of u and v are obtained by $N = U_n \cap V_n$. When $N = \emptyset$, the weight of (u, v) is set to 0. Then, the weight of (u, v) is predicted by:

$$W(u, v) = \begin{cases} \frac{\sum_{n \in N} W(u, n) + W(n, u) + W(v, n) + W(n, v)}{M}, & \text{if } N \neq \emptyset \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where M is number of vertexes in set N , $W(u, n)$, $W(n, u)$, $W(v, n)$, $W(n, v)$ are weights of each edge connected to u or v of vertexes in set N , respectively.

5.3 Triadic Status

Imaging there are two nodes named A and B , and they are unlinked. What we want to do is to predict the weight of the edge from A to B . Suppose that there is another node X , and (A, B, X) compose a triangle once A and B are linked. Overall, there are four different triangle-links, as the edge between X and A can go in either direction and similarly for the edge between X and B , if the sign of weight is not considered.

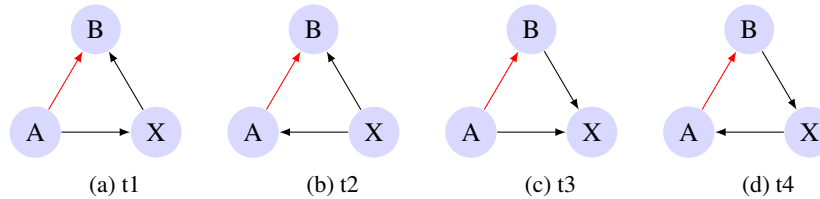


Figure 1: Triadic Status

And we predict the weight of the edge (A, B) using three rules:

1. In Graph t1, the predicted weight of the transitive edge is the sum of the two non-transitive edge weights.
2. In Graph t2 and t3, the predicted weight of a non-transitive edge is the difference of the transitive and the other non-transitive edge weights.
3. In Graph t4, since it is a cyclic triad, the weight of the missing edge is the negative of the sum of other two edge weights.

In the end, we take the average of predicted weights over all triadic-links that these two nodes are a part of.

5.4 Status Theory

Status theory is derived from []. The prediction made by this measure is the difference between the status of vertex u and vertex v . The status $\sigma(i)$ of a vertex i is defined as $\sigma(i) = |W_{in}^+(i)| - |W_{in}^-(i)| + |W_{out}^-(i)| - |W_{out}^+(i)|$. Status increases when receiving positive incoming edges and generating negative outgoing edges to other vertices, while decreases when receiving negative edges and generating outgoing positive edges. Difference in status measures how much 'higher' u 's status is compared to v 's.

5.5 Bias and Deserve

This method is proposed by Mishra and Bhattacharya in []. To compute "bias" and "deserve", we should first normalize the ratings (weights), and keep them in the range of $[-1, 1]$ where 0 is a neutral opinion. Then, we say node u gives a trust-score of w_{ij} to node v for a given rating of (u, v) . The two attributes of a node are defined by:

- *Bias*: This reflects the expected weight of an outgoing edge.
- *Deserve*: This reflects the expected weight of an incoming edge from an unbiased vertex.

Let $d^o(u)$ denotes the set of all outgoing edges from vertex u and likewise, $d^i(u)$ denotes the set of all incoming links to node u . Then, bias (BIAS) and deserve (DES) are iteratively computed as:

$$BIAS^{(t+1)}(u) = \frac{1}{2|d^o(u)|} \sum_{v \in d^o(u)} [W(u, v) - DES^t(v)] \quad (3)$$

$$DES^{(t+1)}(u) = \frac{1}{2|d^i(u)|} \sum_{v \in d^i(u)} [W(v, u)(1 - X^t(v, u))] \quad (4)$$

where $X^t(v, u) = \max\{0, BIAS^t(v) \times W(v, u)\}$. The iterative formulations of bias and deserve allow us to predict the weight of (u, v) based on the deserve value $DES(v)$ of vertex v . Thus, the weight is directly predicted by:

$$W(u, v) = DES(v) \quad (5)$$

5.6 Page Rank

We first compute Page Rank (ωPR) value of every vertex(k) in an unsigned graph according to [28]:

$$\omega PR(k) = \frac{1 - \alpha}{|V|} + \alpha \sum_{z \in in(k)} \frac{\omega PR(z) * W(z, k)}{|out(z)|} \quad (6)$$

α is the probability called damping factor meaning a person will continue clicking on links(travelling in graph), and we set α to 0.85. $|V|$ is the number of all nodes. Eventually, $\omega PR(k)$ will converge to the probability that a person will go and stay at node k .

Then we compute edge weight of (u, v) using weighted average of PageRank values:

$$W(u, v) = \frac{\sum_{z \in \text{out}(u)} \omega PR(z) * W(u, z) + \sum_{z \in \text{in}(v)} \omega PR(z) * W(z, v)}{\sum_{z \in \text{out}(u)} \omega PR(z) + \sum_{z \in \text{in}(v)} \omega PR(z)} \quad (7)$$

5.7 Signed-Hits

The prediction is computed by using a modified version of HITS for signed network, called Signed-HITS[1]. Signed-HITS will compute the hub and authority scores of every node separately on positive graph(all edges' weights are positive) and negative graph, using the equation:

$$\begin{aligned} h^+(u) &= \sum_{v \in \text{out}^+(u)} a^+(v); a^+(u) = \sum_{v \in \text{in}^+(u)} h^+(v) \\ h^-(u) &= \sum_{v \in \text{out}^-(u)} a^-(v); a^-(u) = \sum_{v \in \text{in}^-(u)} h^-(v) \end{aligned} \quad (8)$$

after convergence, we assign the authority score $a(u) = a^+(u) - a^-(u)$ and hub score $h(u) = h^+(u) - h^-(u)$ to each vertex u . Authority scores estimate the node value based on the incoming links, while hub scores estimate the node value based on outgoing links. However, the paper[27] didn't deliver the method to compute edge weight by using the authority score and hub score. Again, we use weighted average to compute the weight prediction:

$$W(u, v) = \frac{h(u) * \sum_{z \in \text{out}(u)} W(u, z) + a(v) * \sum_{z \in \text{in}(v)} W(z, v)}{h(u) * |V_{z \in \text{out}(u)}| + a(v) * |V_{z \in \text{in}(v)}|} \quad (9)$$

In this equation, we use $h(u)$ to be the weight of all out-going edges and $a(v)$ to be the weight of all in-coming edges. The reason is that $h(u)$ represents the node value based on out-going edges. So the predicted weight of out-edges from u should be higher if $h(u)$ is high. Likewise, the predicted weight of in-edges into v should be higher if $a(v)$ is high.

5.8 Linear Regression

In this experiment, we use a linear regression model to predict the weight of edge. Our features are generated from some algorithms mentioned above. First, the test edge e_t whose weight is to be predicted is removed from the network. Second, we remove the second edge e_s and predict the weight of e_s from the rest of the edges. And the predicted value of e_s will be the feature and the true value of e_s will be the ground truth. The second step is done multiple times to generate enough samples for training. After training, we use this model to predict the value of the removed edge e_t .

6 Performance Evaluation

7 Conclusion

References

- [1] Moshen Shahriari and Mahdi Jalili. Ranking nodes in signed social networks. *Social network analysis and mining*, 4(1):172, 2014.
- [2] S. Krumar, F. Spezzano, and V. S. Subrahmanian. Accurately detecting trolls in slashdot zoo via decluttering. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pages 188–195, 2014.
- [3] Zhaoming Wu, Charu C. Aggarwal, and Jimeng Sun. The troll-trust model for ranking in signed networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, WSDM '16, pages 447–456, New York, NY, USA, 2016. ACM.

Table 1: Performance of different algorithms for edge weight prediction in leave-one-out experiment on different datasets. Each cell is (RMSE, PCC). Lower RMSE and higher PCC are better.

Algorithms	BTCAlphaNet	OTCNet	RFArNet
Existing Algorithms			
PageRank	(0.29,0.25)	(0.35,0.30)	(0.21,0.48)
Bias Deserve	(0.23,0.60)	(0.28,0.61)	(0.21,0.48)
Reciprocal	(0.28,0.44)	(0.33,0.47)	(0.35,0.00)
Signed HITS	(0.23,0.66)	(0.27,0.69)	(0.23,0.44)
Status Theory	(0.45,0.00)	(0.48,-0.01)	(0.39,-0.15)
Triadic Balance	(0.31,0.21)	(0.36,0.28)	(0.24,0.29)
Triadic Status	(0.35,0.24)	(0.37,0.42)	(0.32,0.27)
Fairness and Goodness			
Fairness and Goodness	(0.24,0.60)	(0.28,0.63)	(0.22,0.46)
Linear Regression			
Linear Regression	(0.22,0.67)	(0.25,0.72)	(0.20,0.54)

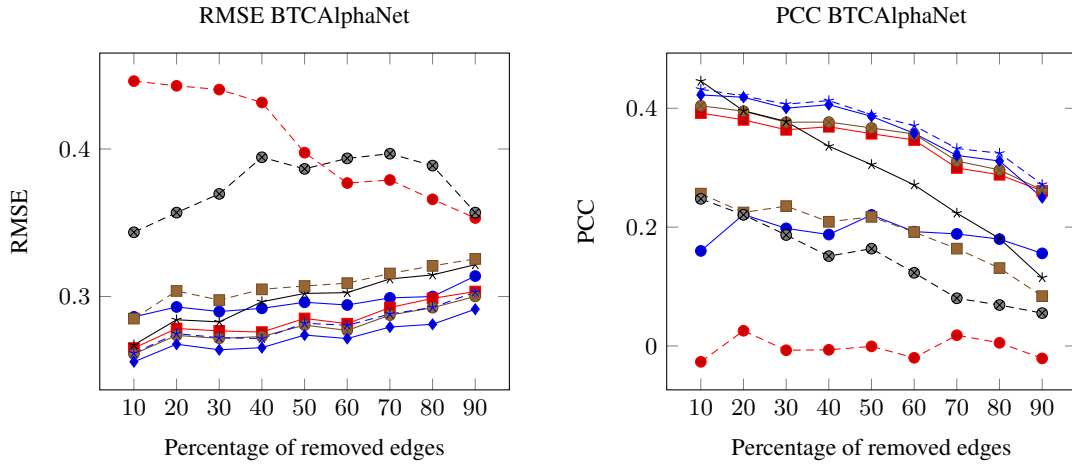
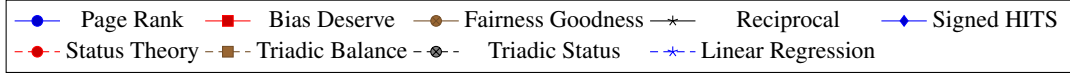


Figure 2: BTCAlphaNet

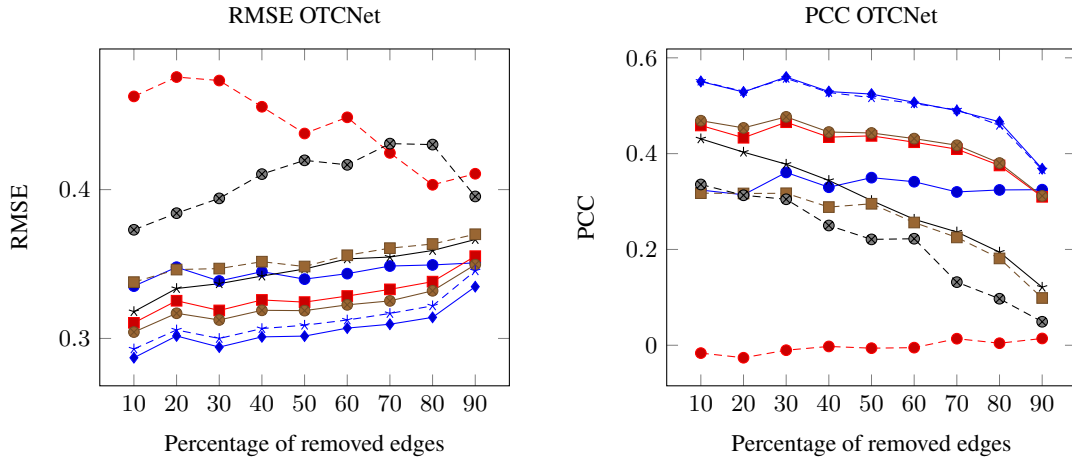


Figure 3: OTCNet

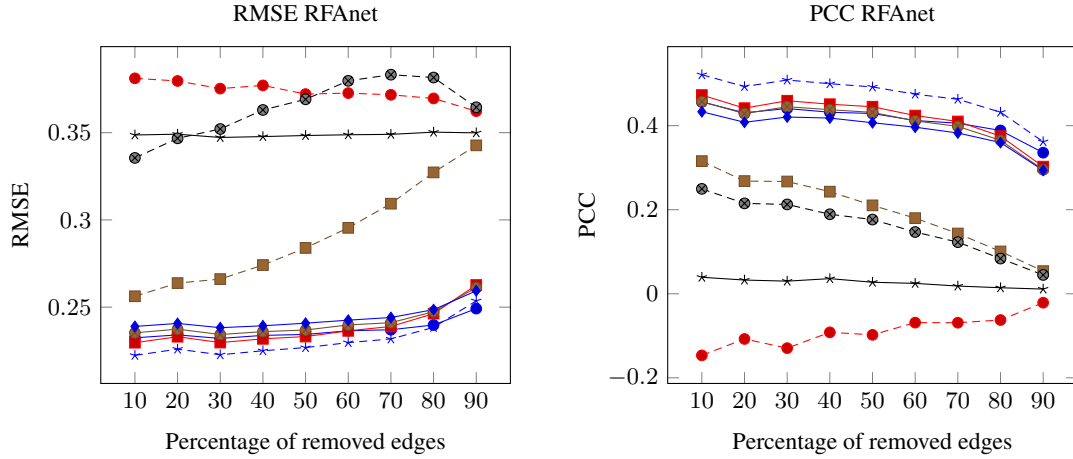


Figure 4: RFAnet

- [4] Srijan Kumar. Structure and dynamics of signed citation networks. In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion*, pages 63–64, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee.
- [5] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1361–1370. ACM, 2010.
- [6] V. A. Traag and Jeroen Bruggeman. Community detection in networks with positive and negative links. *Phys. Rev. E*, 80:036115, Sep 2009.
- [7] Mahsa Shafaei and Mahdi Jalili. Community structure and information cascade in signed networks. *New Generation Computing*, 32(3-4):257–269, aug 2014.
- [8] Dong Li, Zhi-Ming Xu, Nilanjan Chakraborty, Anika Gupta, Katia Sycara, and Sheng Li. Polarity related influence maximization in signed social networks. *PloS one*, 9:e102199, 07 2014.
- [9] Robert West, Hristo Paskov, Jure Leskovec, and Christopher Potts. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, 2, 09 2014.
- [10] Fritz Heider. Attitudes and cognitive organization. *The Journal of Psychology*, 21(1):107–112, 1946. PMID: 21010780.
- [11] Dorwin Cartwright and Frank Harary. Structural balance: A generalization of heider’s theory. *Psychological Review*, 63(5):277–293, 1956.
- [12] S. Wasserman, K. Faust, Cambridge University Press, M. Granovetter, University of Cambridge, and D. Iacobucci. *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.
- [13] James A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20(2):181–187, 1967.
- [14] R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web, WWW '04*, pages 403–412, New York, NY, USA, 2004. ACM.
- [15] Thomas Dubois, Jennifer Golbeck, and Aravind Srinivasan. Predicting trust and distrust in social networks. pages 418–424, 10 2011.

- [16] Shuang-Hong Yang, Alexander Smola, Bo Long, Hongyuan Zha, and Yi Chang. Friend or frenemy? predicting signed ties in social networks. *SIGIR'12 - Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, 08 2012.
- [17] Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. A survey of signed network mining in social media. *ACM Comput. Surv.*, 49(3):42:1–42:37, August 2016.
- [18] Yi Qian and Sibel Adali. Foundations of trust and distrust in networks: Extended structural balance theory. *ACM Trans. Web*, 8(3):13:1–13:33, July 2014.
- [19] Eric Gilbert. Predicting tie strength in a new medium. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12*, pages 1047–1056, New York, NY, USA, 2012. ACM.
- [20] Eric Gilbert and Karrie Karahalios. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09*, pages 211–220, New York, NY, USA, 2009. ACM.
- [21] Rongjing Xiang, Jennifer Neville, and Monica Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pages 981–990, New York, NY, USA, 2010. ACM.
- [22] Indika Kahanda and Jennifer Neville. Using transactional information to predict link strength in online social networks. 01 2009.
- [23] Stavros Sintos and Panayiotis Tsaparas. Using strong triadic closure to characterize ties in social networks. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, pages 1466–1475, New York, NY, USA, 2014. ACM.
- [24] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th International Conference on World Wide Web, WWW '03*, pages 640–651, New York, NY, USA, 2003. ACM.
- [25] Yarden Katz and Jennifer Golbeck. Social network-based trust in prioritized default logic. volume 2, 01 2006.
- [26] A. Roy, C. Sarkar, J. Srivastava, and J. Huh. Trustingness trustworthiness: A pair of complementary trust measures in a social network. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 549–554, Aug 2016.
- [27] Srijan Kumar, Francesca Spezzano, VS Subrahmanian, and Christos Faloutsos. Edge weight prediction in weighted signed networks. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 221–230. IEEE, 2016.
- [28] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.