Signed Network Dataset Repository: Extracting Signed Relations From Social Networks

Aikta Arya*, Pradumn Kumar Pandey†, Tyler Derr‡
*†Dept. of Computer Science and Engineering, Indian Institute of Technology Roorkee, Uttarakhand, India.
‡Dept. of Computer Science, Vanderbilt University, Tennessee, United States.
Emails: [*a_arya, †pradumn.pandey]@cs.iitr.ac.in, ‡tyler.derr@vanderbilt.edu

***Abstract*—The proliferation in the use of social media and networking platforms has enhanced frequent interactions among individuals. The signed networks can be utilized to best perceive such complex relations. However, these real-world signed network datasets are scarce as the real-world signed datasets often involve dealing with sensitive information about individuals and their relationships. The ethical and privacy concerns of various social media platforms can further limit the availability of datasets for research and academic purposes. Furthermore, the availability of signed bipartite network datasets is even scarcer than signed network datasets. Existing signed network datasets also lack node/edge level attributes, yet recent trends in graph analytics have shown the significance of these features, for example, with GNNs, to advance benchmarking. To address such issues, we present three datasets in this paper: Reddit Posts (bipartite network with primary node level attributes and secondary level edge attributes), Bitcoin-OTC, and Bitcoin-Alpha datasets (signed networks with primary node and edge level attributes). Additionally, we explore potential research applications the dataset opens up across various domains. Various datasets discussed in this paper, along with already existing signed network datasets, are made publicly available using Signed Network Dataset Repository (SigNET Repo) at: https://aikta-arya.github.io/SigNET-/**

***Index Terms*—Signed network datasets, Real-world networks.**

I. Introduction

The recent proliferation in use of social media and networking platforms has facilitated increased interactions among individuals, giving rise to diverse relationships like friendships and animosities. To represent these intricate relations, researchers utilize signed networks, where nodes represent individuals and signed edges depict the nature of their interactions. These networks find application in various domains including link and sign prediction [1], [2], community detection [3], [4], node and edge classification [5], [6], spam classification [7], [8], triangle counting [9], [10], network reconstruction and modeling [11]–[14], and trust/distrust prediction [15]. Real-world datasets play an essential role in developing and evaluating algorithms and theories, offering insights into complex social dynamics [16]. However, accessing such datasets is challenging due to ethical and privacy issues surrounding sensitive information about individuals and their relationships, thus requires collaborative efforts to gather, share, and curate them.

Moreover, while traditional signed networks have gained attention, signed bipartite networks, featuring distinct node sets with signed links between them, remain relatively understudied despite their prevalence in diverse fields. Scarce availability of signed bipartite network datasets effect research in areas such as, product recommendations, congressional voting analysis [17], and balance theory analysis [18], [19]. Addressing this scarcity requires collective endeavors to compile more comprehensive datasets. Additionally, existing datasets often lack sufficient features, hindering bench-marking on tasks like node and label prediction. Incorporating node and edge attributes into datasets is crucial for accurate comparisons among methods. Furthermore, a fundamental challenge in network analysis arises from the absence of true community information in existing signed network datasets. True community information is vital for understanding underlying social dynamics, highlighting the need for datasets that include such information from the outset.

To address these issues discussed above, this paper presents one signed bipartite network dataset, the Reddit Posts dataset, and two signed network datasets, The Bitcoin-OTC and Bitcoin-Alpha datasets. The Reddit Posts dataset comprises 80938 distinct Reddit post data extracted from the Reddit social media platform starting from 2020 to 2023. The dataset mainly constitutes the posting behavior of 59352 distinct authors that posted their posts in at least one of the 103 different communities (subreddits) of the Reddit platform for the purpose of building a bipartite signed network. Furthermore the Reddit Posts dataset is equipped with true community information of 103 communities. Additionally, the posts posted by authors in a community are further discussed by other users in the form of comments and replies to those comments (comment forest). We have extracted these comments on various posts, keeping the crucial true community information intact with the dataset. To the best of our knowledge, we are the first to consider preserving true community information along with the edges of the signed bipartite network.

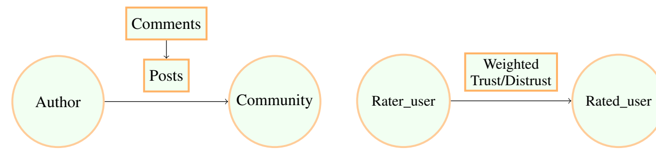
The Bitcoin-OTC network constitutes a web of trust among its users from 2010 to 2023 that consists of 5927 nodes and 36029 edges, representing the weight of trust among two users based on their previous transactions. The weighted edges of the Bitcoin-OTC signed network are supported by the timestamp and comment attributes. The timestamp attribute is a required field in those applications where the edge insertion and deletion depends on the time, for instance, network diffusion and modeling [11], [12]. Furthermore, the nodes of Bitcoin-OTC dataset are attributed with several node level features, making it suitable for applications such as node classification and spam classification [5], [20], [21]. Another Bitcoin-Alpha signed dataset consists of a total of 3755 nodes and 24207 edges representing the weighted trust among two users of the Bitcoin-Alpha platform. These weighted edges are also attributed with features such as timestamps and comments. Similar to the Bitcoin-OTC dataset, the Bitcoin-Alpha dataset is also equipped with several node-level features. In conclusion, the datasets presented in this paper address most of the issues discussed above.

Our contributions are summarized as:

- **Novel Datasets:** We present Signed Network Dataset Repository (SigNET Repo[1]) as a alternative dataset repository for a diverse collection of existing as well as new signed network datasets- (1) Reddit Posts dataset[2] with primary node level attributes and secondary level edge attributes. (2) Bitcoin-OTC[3] and (3) Bitcoin-Alpha[3].
- **Availability and Utility:** The datasets are cataloged in a searchable repository that can be easily retrieved using their unique Digital Object Identifier (DOI):
  - Reddit Posts Dataset: 10.5281/zenodo.10508388
  - Bitcoin Network Datasets: 10.5281/zenodo.10478502
- **Predicted Impact and Beneficial Groups:** For researchers, academicians and data scientists working in domain of network science and social media analysis, our contributions include: (i) an open-source dataset repository for signed network datasets. (ii) a new Reddit Posts dataset enriched with primary node level attributes and secondary level edge attributes having diverse applications in community detection, sign prediction and link prediction etc. (iii) updated Bitcoin network datasets with primary node and edge level attributes (iv) a range of potential applications that could enhance current methods and solutions to a practical level that can further contribute in proliferation of research work in signed networks.

II. Related Datasets

The behavior of commenting and posting in online communities, as well as trust-distrust dynamics among users, is studied through real-world signed networks from social media. The significant analyses include assessments of U.S. Congressional speeches for support or opposition to legislation [22], initial theories of signed networks using datasets like Wiki Election Slashdot, and Epinions [23], and public datasets such as Wiki-Conflict, WikiSigned, and chess from the KONECT project [24]. Studies also explore individual assessments through signed networks and sentiment analysis [25], with datasets such as Wiki Rfa, BitcoinA and BitcoinO [26] used for



(a) Reddit Posts Dataset. (b) Bitcoin network Datasets.

Fig. 1: A visualization of the diverse attributes in both the (a) Reddit Posts dataset and (b) Bitcoin network datasets. The node-level features are represented in circles and edge-level features are represented in rectangles.

predictive tasks. While existing signed network datasets have laid foundational groundwork, they often lack crucial elements necessary for comprehensive analysis. For instance, in the case of Bitcoin networks, existing datasets (available at the SNAP[4]) have been collected earlier without considering crucial factors such as time and attributes. Subsequently, time information was added, but attributes are still lacking. However, recent trends in graph analytics, particularly with the rise of Graph Neural Networks (GNNs), have emphasized the significance of node and edge features. Therefore, having signed networks with comprehensive features is essential to advance benchmarking and analysis in this evolving field.

Furthermore, the signed bipartite network datasets, such as Bonanza, U.S. Senate, and U.S. House of Representatives used in [19] are rather small and lack detailed information. These limitations constrain the depth of analysis and benchmarking that can be performed. Additionally, existing signed network datasets, often derived from older technology and product review platforms, have properties that differ significantly from today's online social media environments. Reddit, for instance, encompasses more diverse topics and reflects modern social media dynamics, underscoring the need for new, detailed datasets from such platforms. Hence, our proposed datasets significantly advance over previous ones, offering researchers the necessary tools for more comprehensive and insightful analyses in social media and community dynamics.

III. Datasets

A. The Reddit Posts Dataset

This dataset is collected from Reddit[5], a social media platform where millions of users interact by sharing videos, images, or text. Reddit is divided into subreddits (community), each focused on specific topics or interests. Users create posts in these subreddits, which are discussed through *comment threads*. Posts and comments can receive downvotes or upvotes. The Subreddit moderators manage communities, assign flags to comments, and can ban users for inappropriate behavior.

[1] https://aikta-arya.github.io/SigNET-/
[2] https://github.com/Aikta-Arya/Reddit-Posts-Dataset.git
[3] https://github.com/Aikta-Arya/Bitcoin-Network-Datasets.git
[4] https://snap.stanford.edu/data/soc-sign-bitcoin-otc.html
[5] https://www.reddit.com/

TABLE I: Reddit Posts primary edge level features.

Primary Features	Description
Subreddit_id	ID of the subreddit of the comment.
Score	The number of upvotes for the comment.
Upvote ratio	The percentage of upvotes on the post.
Created on	Unix Time the post was created.
Post URL	The URL the post.
Distinguished	Whether or not the post is distinguished.
Edited	Whether or not the post has been edited.
Over 18	Whether or not the post has been marked as NSFW.
Locked	Whether or not the post has been locked.
Original content	Whether or not the post has been set as original content.
Total comments	number of comments on the post.
Author name	name of the submitting author of the post.

TABLE II: Reddit Posts node level features of communities.

Node Features	Description
Id	Id of the subreddit.
Sub_Name	Name of the subreddit.
Full_Name	Fullname of the subreddit.
Subscribers	Count of subscribers.
Description	Description of the subreddit, shown in searches.
Created on	Unix Time when the subreddit was created.
Spoilers_enabled	Whether the spoiler tag feature is enabled.
over18	Whether or not the subreddit has been marked as NSFW.

TABLE III: Reddit Posts node level features of Authors.

Node Features	Description
Id	Id of the redditor.
Name	Username of the redditor.
Karma	The link karma for the redditor.
Created on	Unix Time when the account was created.
Trophies	list of the redditor's trophies.
Moderator	list of subreddits moderated and manged by the redditor.
Employee	Whether or not the redditor is a Reddit employee.
Friend	Whether or not the redditor is friends with the authenticated user.
Premium_user	Whether or not the redditor has active Reddit Premium status.

TABLE IV: Reddit Posts secondary level edge features.

Secondary Features	Description
Id	Id of the comment.
Author	Name of owner of the comment.
Comment	The body of the comment, as Markdown.
Score	The number of upvotes for the comment.
Created on	Unix time when the comment was created.
Replies	Provides an instance of replies of the comment.
Submission	The submission (post ID) that the comment belongs to.
Subreddit	The subreddit that the comment belongs to.
Subreddit_id	The subreddit ID that the comment belongs to.
Edited	Whether or not the comment has been edited.
Is_submitter	Whether or not the comment author is also the author of the post.
link_id	The post ID that the comment belongs to.
parent_id	The ID of the parent comment (prefixed with t1_).
Permalink	A permalink for the comment.

1) Data Extraction: The dataset extraction is performed using PRAW (Python Reddit API Wrapper), which is equipped with Python packages that provides access to Reddit's API. Using PRAW, we extracted 80938 distinct Reddit post data from 2020 to 2023. Moreover, we have extracted and mined the posting behavior of 59352 distinct Authors that posted their posts in at least one of the 103 different communities (subreddits). The node set of the Reddit post dataset constitutes two types of nodes belonging to either the community (subreddit) set or the author set.

The edges of the Reddit post dataset represent that the given author (redditor) has posted a post in a given community (subreddit). These edges are further attributed with the primary edge features (refer Figure 1a) such as Score, Upvote ratio, Created on, Post URL, Distinguished, Edited, Over 18, Locked, Original content, Total comments, Author's name. The description of these primary edge features is discussed in Table I. Furthermore, the attribute Upvote ratio is utilized for assigning sign $+1$ or -1 to the edges of the resultant network. Meanwhile, the edges (posts) with an Upvote ratio more than 0.5 are designated with sign $+1$, and the edges (posts) with an Upvote ratio less than or equal to 0.5 are designated with sign -1. The Reddit posts network consists of a total of 80938 distinct edges (posts) among authors (redditors) and communities representing a signed relation on the basis of the Upvote ratio of the post.

The Reddit post dataset consists of two types of nodes, communities and authors, that are also attributed to the node features. The nodes belonging to the community set are attributed with features Id, Sub_Name, Full_Name, Subscribers, Description, Created on, over18, Spoilers_enabled. Meanwhile, the nodes belonging to the author set are attributed with features such as Id, Name, Karma, Created on, Moderator, Trophies, Employee, Friend, and Premium_user. The description of these node features is summarized in Table II and Table III.

The posts posted by authors in a community are further discussed by other users in the form of comments and replies to those comments. The edges (posts) are further attributed with secondary features Id, Author, Comment, Score, Created on, Replies, Submission, Subreddit, Subreddit_id, Editted, Is_submitter, link_id, parent_id, Permalink. The description of these secondary level edge features is given in Table IV

2) Potential Applications:

- The Reddit posts dataset represents interactions among various authors (redditors) across diverse communities (subreddits). These communities represent the true communities on Reddit, which is rarely observed feature in already available datasets. The true community data can be further utilized for studying structural, functional and diffusion dynamics of diverse communities [4], [27].
- The node-level attributes of nodes belonging to either the author or community set in the Reddit posts dataset can be used for various applications, including node classification [5], [20], [21], label prediction, awarding Trophies, and predicting subreddit for redditors.
- The primary and secondary edge-level attributes in the Reddit posts signed bipartite network dataset can be used for link prediction [28], [29], sign prediction, assessing post appropriateness for subreddits, predicting post or comment scores, forecasting redditor behavior, identifying suitable subreddits to prevent moderator removal, and detecting malicious activities.

- All node and edge-level features of the Reddit Posts dataset include the 'Created on' attribute, representing the Unix time of their insertion into the network. This enhances the dataset's suitability for applications where node or edge insertion and deletion depend on time, such as network diffusion and modeling [11], [12].

B. The Bitcoin Network Datasets

Proliferation in the use of online transactions and cryptocurrency has emerged as a transformation of the global financial landscape, reshaping traditional paradigms of commerce and payment systems. Out of various available cryptocurrencies, Bitcoin is widely considered the most famous cryptocurrency elementally employed for anonymous online transactions. The inherent anonymity also introduces complementary risk [30], such as those associated with trading platforms where two parties directly interact without any supervision. This has prompted the online community to establish various platforms where they can all publicly assess the trustworthiness of their counterparts to hopefully reduce the impact of scammers.

We further discuss two proffered Bitcoin web of trust network datasets: (1) Bitcoin-OTC [1] and (2) Bitcoin-Alpha[2]. The Bitcoin-OTC dataset is extracted from its official website, where the dataset is available as the web of trust among its users. The Bitcoin-OTC and Bitcoin-Alpha platforms allows its users to rate other users ranging from $+10$ to -10 (0 excluded). According to the guidelines of the platform, $+10$ designates the highest weighted trust, representing that the user is not a fraudster, whereas -10 designates the lowest weighted trust or total distrust, representing that the user is a fraudster. The other intermediate values among $+10$ and -10 hold the subsequent intermediate meanings.

1) Data Extraction: We have extracted the web of trust among users of the Bitcoin-OTC platform from 2010 to 2023, consisting of 5,927 nodes and 36,029 edges. These edges represent the weight of trust between two users based on their previous transactions (as shown in Figure 1b). The weighted edges are supported by timestamp and comment attributes. The timestamp is crucial for applications involving time-dependent edge insertion and deletion, such as network diffusion and modeling [11], [12]. The comment attribute provides additional validation information for the trust ratings. Other attributes in the Bitcoin-OTC signed network are listed in Table V with their descriptions.

The users of the Bitcoin-OTC signed network have the node level attributes as shown in Table VI with their corresponding descriptions. According to the data extracted, the earliest users registered on Bitcoin-OTC platform are from 2010 and the latest node is from 2023.

Additionally, the Bitcoin-Alpha dataset extracted from Bitcoin-Alpha platform consists of a total of 3755 nodes and 24207 edges representing the weighted trust among two users of the Bitcoin-Alpha platform. These weighted edges are also attributed with features such as timestamps and comments. The attributes and structure of the Bitcoin-Alpha dataset are similar to the Bitcoin-OTC. The nodes of Bitcoin-Alpha dataset are attributed with node level features such as User, Registered_time, Last_seen_time, Fingerprint, Avg_rating, Total_rating. The description of these attributes is given in Table VII.

[1]https://bitcoin-otc.com/trust.php
[2]https://btc-alpha.com/en

TABLE V: Various attributes of Bitcoin Network Datasets.

Attributes	Description
Rater_user	ID of the Rater User.
Rated_user	ID of the Rated User.
Rating_time	Timestamp when the user was rated.
Rating	Weighted trust rating of the rated user.
Comment	Comment supporting the rating of the user.

TABLE VI: Various node level features of Bitcoin-OTC signed network dataset.

Node Features	Description
User	Name of the user of Bitcoin-OTC platform.
Registered_time	Time when the user was registered
Key_id	Unique Hexadecimal Key.
Total_rating	Total rating of the user.
Num_pos_ratings_recv	Total positive rating received by the user.
Num_neg_ratings_recv	Total negative rating received by the user.
Num_pos_ratings_snt	Total positive rating sent by the user.
Num_neg_ratings_snt	Total negative rating sent by the user.

TABLE VII: Various node level features of Bitcoin-Alpha signed network dataset.

Node Features	Description
User	Name of the user of Bitcoin-Alpha platform.
Registered_time	Time when the user was registered
Last_seen_time	Time when the user has last involved in any activity.
Fingerprint	Digital fingerprint of the user.
Avg_rating	Average rating of the user.
Total_rating	Total rating of the user.

2) Potential Applications: Bitcoin signed network datasets are the signed networks representing the weight of trust or distrust among various users of Bitcoin-OTC and Bitcoin-Alpha platforms.

- These signed datasets can be used in various downstream applications, including signed link prediction and link analysis [28], [29], signed community detection [4], [27], triangle counting and enumeration algorithms [9], edge classification [31] and spam classification [7].
- Additionally, the node level attributes and features available with the dataset make these datasets more appropriate for node classification applications [5], [20], [21].
- The features such as 'Rating_time' allow us to keep track of the time when an edge is inserted into the network, making it more practical to applications sensitive to temporal data such as network modeling and diffusion algorithms [11], [12] and, event detection algorithms [32].

TABLE VIII: Comparison of Precision (P@k), Recall (R@k), F1-Score (F1@k), and Hit Ratio (H@k) of MF-BPR, NGCF, LightGCN, DirectAU, and GraphAU for the **Reddit Posts**.

Models	**P@5**	**R@5**	**F1@5**	**H@5**
MF-BPR [37]	0.0671	0.3185	0.1102	0.3356
NGCF [36]	0.0260	0.1153	0.0418	0.1301
LightGCN [35]	**0.1055**	**0.4954**	**0.1726**	**0.5137**
DirectAU [34]	0.0247	0.1153	0.0402	0.1233
GraphAU [33]	0.0945	0.4372	0.1540	0.4589
Models	**P@20**	**R@20**	**F1@20**	**H@20**
MF-BPR [37]	0.0428	0.7820	0.0809	0.7877
NGCF [36]	0.0332	0.6027	0.0627	0.6164
LightGCN [35]	0.0476	0.8847	0.0901	0.8904
DirectAU [34]	0.0243	0.4566	0.0460	0.4726
GraphAU [33]	**0.0479**	**0.8870**	**0.0906**	**0.8904**

TABLE IX: Comparison between diverse link prediction methods using **Bitcoin-OTC** and **Bitcoin-Alpha** datasets.

	Models	**AUC@p**	**AUC@n**	**AUC@non**	**macro-F1**
Bitcoin-Alpha	lSNE [39]	0.870	0.588	0.856	0.591
	nSNE [39]	0.786	0.688	0.667	0.543
	SLF [28]	0.884	0.809	0.886	0.580
	STNE [40]	0.828	0.863	0.844	0.496
	SPMF [1]	**0.890**	**0.868**	**0.896**	**0.633**
Bitcoin-OTC	lSNE [39]	0.790	0.734	0.788	0.578
	nSNE [39]	0.781	0.735	0.748	0.598
	SLF [28]	**0.889**	0.778	0.881	0.611
	STNE [40]	0.810	**0.858**	0.832	0.506
	SPMF [1]	0.884	0.854	**0.892**	**0.627**

IV. Experiments

Both link prediction and recommendations can be considered as edge classification task as it involves determining the type or presence of an edge between two nodes in a graph. Hence, we employ recommendations to validate and benchmark Reddit posts dataset (signed bipartite network) and link prediction task for validating and bench-marking Bitcoin-OTC and Bitcoin-Alpha datasets (signed networks). We employ the following recommendation models as the approaches to validate the collected Reddit posts dataset: GraphAU [33], DirectAU [34], LightGCN [35], NGCF [36], MF-BPR [37] and following link prediction models as the approaches to validate the collected Bitcoin-OTC and Bitcoin-Alpha datasets: NetMF [38], lSNE [39], nSNE [39], SLF [28], STNE [40], SPMF [1].

A. Evaluation Metrics

We adopt Precision@k, Recall@k, F1-Score@k [41] and Hit Ratio@k where, $k = \{5, 20\}$ [42] as matrices to validate the precision of the recommended list of communities (subreddits) to the authors of Reddit in Reddit Posts dataset. Additionally, we employ the standard metrics such as macro-F1 score [43] and area under curve (AUC) [44] to validate the link prediction performance for Bitcoin-OTC and Bitcoin-Alpha datasets. These matrices are insensitive to class distribution and have been widely employed to assess and validate the quality of link prediction models.

B. Experimental Results

Table VIII summarizes the experimental results for validating Reddit Posts dataset in terms of recommendations. It is observed that LightGCN surpasses other graph based methods for $k = 5$ whereas GraphAU surpasses other baselines for $k = 20$ showing that pure graph-based methods are sensitive to datasets and their performance may vary as we increase of decrease the number of recommendations or suggestions. This suggests need for more signed bipartite datasets to conduct more comprehensive and insightful analyses. Table IX summarizes the experimental results for validating Bitcoin-OTC and Bitcoin-Alpha datasets in terms of link prediction accuracy. The prediction performance for positive, negative and null social relationships is denoted as AUC@p, AUC@n, and AUC@non, respectively. It is observed that SPMF surpasses other baselines for Bitcoin-Alpha dataset whereas for Bitcoin-OTC dataset the performance of SPMF, STNE and SLF is comparable as SLF outperforms SPMF marginally for AUC@p and STNE outperforms SPMF marginally for AUC@n. whereas for AUC@non and macro-F1 score SPMF outperforms other baselines which shows that pure graph-based methods are sensitive to datasets necessitating the need for diverse signed network datasets.

V. Conclusion

This research is motivated by the need for real-world signed network datasets for diverse research and academic purposes. These real-world signed network datasets are scarce as they often involve dealing with sensitive information about individuals and their relationships, raising user privacy issues of individual platforms. The proposed datasets also address the issues of the unavailability of true community information with the datasets, making them more applicable to applications such as community detection and node classification.

In the future, we plan to enrich our dataset with multimedia information such as images and videos, as well as emoji reactions on Reddit posts. Additionally, we also intend to include signed data extracted from multiple platforms such as Twitter, TikTok, etc. to avoid the limitations that most signed network analysis is performed on older, smaller, and less attribute rich datasets.

VI. Ethical Statement

The datasets contain publicly available information, including potentially offensive or personally identifiable data from user-generated content. No explicit filtering for such information was conducted, as it was sourced from content users posted knowingly in public. The extraction and release of these datasets align with platform terms of service.

Acknowledgment

This research is supported in part by the National Science Foundation under grant number IIS2239881.

References

[1] P. Xu, W. Hu, J. Wu, W. Liu, Y. Yang, and P. Yu, "Signed network representation by preserving multi-order signed proximity," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[2] G. Beigi, S. Ranganath, and H. Liu, "Signed link prediction with sparse data: The role of personality information," in *Companion Proceedings of The 2019 World Wide Web Conference*, 2019, pp. 1270–1278.

[3] P. Esmailian and M. Jalili, "Community detection in signed networks: the role of negative ties in different scales," *Scientific reports*, vol. 5, no. 1, p. 14339, 2015.

[4] R. Sun, C. Chen, X. Wang, Y. Zhang, and X. Wang, "Stable community detection in signed social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 10, pp. 5051–5055, 2020.

[5] J. Tang, C. Aggarwal, and H. Liu, "Node classification in signed social networks," in *Proceedings of the 2016 SIAM international conference on data mining*. SIAM, 2016, pp. 54–62.

[6] J. Tang, Y. Chang, C. Aggarwal, and H. Liu, "A survey of signed network mining in social media," *ACM Computing Surveys (CSUR)*, vol. 49, no. 3, pp. 1–37, 2016.

[7] S. Jeong and K.-h. Lee, "Spam classification based on signed network analysis," *Applied Sciences*, vol. 10, no. 24, p. 8952, 2020.

[8] Y. Liu, "Signed latent factors for spamming activity detection," *arXiv preprint arXiv:2209.13814*, 2022.

[9] A. Arya, P. K. Pandey, and A. Saxena, "Balanced and unbalanced triangle count in signed networks," *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[10] M. N. Kolountzakis, G. L. Miller, R. Peng, and C. E. Tsourakakis, "Efficient triangle counting in large graphs via degree-based vertex partitioning," *Internet Mathematics*, vol. 8, no. 1-2, pp. 161–185, 2012.

[11] A. Arya and P. K. Pandey, "Structural reconstruction of signed social networks," *IEEE Transactions on Computational Social Systems*, 2022.

[12] P. K. Pandey, B. Adhikari, M. Mazumdar, and N. Ganguly, "Modeling signed networks as 2-layer growing networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 7, pp. 3377–3390, 2022.

[13] A. Arya and P. K. Pandey, "Sissrm: Sequentially induced signed subnetwork reconstruction model for generating realistic synthetic signed networks," *IEEE Transactions on Computational Social Systems*, 2024.

[14] A. Arya, "Realistic synthetic signed network generation and analysis," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, pp. 5439–5442.

[15] H. Fang, X. Li, and J. Zhang, "Integrating social influence modeling and user modeling for trust prediction in signed networks," *Artificial Intelligence*, vol. 302, p. 103628, 2022.

[16] P. K. Pandey, A. Arya, and A. Saxena, "X-distribution: Retraceable power-law exponent of complex networks," *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 5, pp. 1–12, 2024.

[17] T. Derr and J. Tang, "Congressional vote analysis using signed networks," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018, pp. 1501–1502.

[18] J. Huang, H. Shen, Q. Cao, S. Tao, and X. Cheng, "Signed bipartite graph neural networks," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 740–749.

[19] T. Derr, C. Johnson, Y. Chang, and J. Tang, "Balance in signed bipartite networks," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1221–1230.

[20] S. Yuan, X. Wu, and Y. Xiang, "Sne: signed network embedding," in *Advances in Knowledge Discovery and Data Mining: 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, Proceedings, Part II 21*. Springer, 2017, pp. 183–195.

[21] A. Arya, P. K. Pandey, and A. Saxena, "Node classification using deep learning in social networks," in *Deep Learning for Social Media Data Analytics*. Springer, 2022, pp. 3–26.

[22] M. Thomas, B. Pang, and L. Lee, "Get out the vote: Determining support or opposition from Congressional floor-debate transcripts," in *Proceedings of EMNLP*, 2006, pp. 327–335.

[23] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010, pp. 1361–1370.

[24] J. Kunegis, "Konect: the koblenz network collection," in *Proceedings of the 22nd international conference on world wide web*, 2013, pp. 1343–1350.

[25] R. West, H. S. Paskov, J. Leskovec, and C. Potts, "Exploiting social network structure for person-to-person sentiment analysis," *Transactions of the Association for Computational Linguistics*, vol. 2, pp. 297–310, 2014.

[26] S. Kumar, F. Spezzano, V. Subrahmanian, and C. Faloutsos, "Edge weight prediction in weighted signed networks," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 2016, pp. 221–230.

[27] C. Xia, Y. Luo, L. Wang, and H.-J. Li, "A fast community detection algorithm based on reconstructing signed networks," *IEEE Systems Journal*, vol. 16, no. 1, pp. 614–625, 2021.

[28] P. Xu, W. Hu, J. Wu, and B. Du, "Link prediction with signed latent factors in signed social networks," in *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining*, 2019, pp. 1046–1054.

[29] G. Beigi, J. Tang, and H. Liu, "Signed link analysis in social media networks," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 10, no. 1, 2016, pp. 539–542.

[30] T. Moore and N. Christin, "Beware the middleman: Empirical analysis of bitcoin-exchange risk," in *Financial Cryptography and Data Security: 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers 17*. Springer, 2013, pp. 25–33.

[31] M. Gupta and R. Mishra, "Network projection-based edge classification framework for signed networks," *Decision Support Systems*, vol. 135, p. 113321, 2020.

[32] J. Cadena, A. K. Vullikanti, and C. C. Aggarwal, "On dense subgraphs in signed network streams," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 2016, pp. 51–60.

[33] L. Yang, Z. Liu, C. Wang, M. Yang, X. Liu, J. Ma, and P. S. Yu, "Graph-based alignment and uniformity for recommendation," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 4395–4399.

[34] C. Wang, Y. Yu, W. Ma, M. Zhang, C. Chen, Y. Liu, and S. Ma, "Towards representation alignment and uniformity in collaborative filtering," in *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 2022, pp. 1816–1825.

[35] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "Lightgcn: Simplifying and powering graph convolution network for recommendation," in *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 2020, pp. 639–648.

[36] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, 2019, pp. 165–174.

[37] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," *arXiv preprint arXiv:1205.2618*, 2012.

[38] J. Qiu, Y. Dong, H. Ma, J. Li, K. Wang, and J. Tang, "Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec," in *Proceedings of the eleventh ACM international conference on web search and data mining*, 2018, pp. 459–467.

[39] W. Song, S. Wang, B. Yang, Y. Lu, X. Zhao, and X. Liu, "Learning node and edge embeddings for signed networks," *Neurocomputing*, vol. 319, pp. 42–54, 2018.

[40] P. Xu, W. Hu, J. Wu, W. Liu, B. Du, and J. Yang, "Social trust network embedding," in *2019 IEEE international conference on data mining (ICDM)*. IEEE, 2019, pp. 678–687.

[41] J. Chen, L. Wei, U. Liji, and F. Hao, "A temporal recommendation mechanism based on signed network of user interest changes," *IEEE Systems Journal*, vol. 14, no. 1, pp. 244–252, 2019.

[42] X. Yang, H. Steck, Y. Guo, and Y. Liu, "On top-k recommendation using social networks," in *Proceedings of the sixth ACM conference on Recommender systems*, 2012, pp. 67–74.

[43] M. R. Islam, B. Aditya Prakash, and N. Ramakrishnan, "Signet: Scalable embeddings for signed networks," in *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part II 22*. Springer, 2018, pp. 157–169.

[44] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.