

# wsGAT: Weighted and Signed Graph Attention Networks for Link Prediction

Marco Grassia<sup>1</sup> and Giuseppe Mangioni<sup>1</sup>

Department of Electric Electronic and Computer Engineering, University of Catania,  
Catania 95100, ITALY  
{marco.grassia, giuseppe.mangioni}@unict.it

**Abstract.** Graph Neural Networks (GNNs) have been widely used to learn representations on graphs and tackle many real-world problems from a wide range of domains. In this paper we propose *wsGAT*, an extension of the Graph Attention Network (GAT) [24] layers, meant to address the lack of GNNs that can handle graphs with signed and weighted links, which are ubiquitous, for instance, in trust and correlation networks. We first evaluate the performance of our proposal by comparing against GCNII [6] in the weighed link prediction task, and against SGCN [8] in the link sign prediction task. After that, we combine the two tasks and show their performance on predicting the signed weight of links, and their existence. Our results on real-world networks show that models with *wsGAT* layers outperform the ones with GCNII and SGCN layers, and that there is no loss in performance when signed weights are predicted.

**Keywords:** network science, link prediction, geometric deep learning, geometric representation learning, graph neural networks

## 1 Introduction

Graphs (or networks) are a very flexible formalism that can be used to represent many real-world systems made up of interacting entities [20]. In a network, nodes represent entities and links model how they interact. For instance, in a social network, nodes can be people while links among them represent some kind of social interaction, such as friendship or acquaintance relations [3]. Depending on the modeled system, links can be also directed [2] and/or weighted [21]. Directionality means there is an asymmetric relation among nodes, as it happens in a social media, like Twitter, where a user *Bob* may follow *Alice*, while *Alice* is not required to follow *Bob*'s Twitter account. Moreover, relations can be characterized by different strength levels, that are represented in a network by labelling links with weights expressed in an appropriate scale.

Signed networks [16] are another class of networks, where links can be positive or negative. They are especially used to model good or bad relations among nodes. A notable example are trust networks [1,4,5], where nodes are users and positive/negative links among them are used to model trust/distrust relations.

In general, signed networks can also be weighted, so relations can be positive or negative and with a given strength. Considering again trust networks, each link can express more or less strong relationships of trust or distrust. Weighted and signed networks are also commonly used to represent correlations networks [19,7], where links among entities express the level of correlation that, in general, can be a positive or negative real number.

To solve many (hard) problems on networks, deep learn techniques have recently been used [26,25,11,10]. In particular, they employed Graph Neural Networks (GNNs) [22] to learn representations on graphs by abstracting from the specific application domain. GNNs are powerful tools and their applicability has been successfully demonstrated even to solve very complex problems, such as link prediction in complex networks. Among different GNN layer model, the Graph Attention Networks (GATs) [24] are one of the most promising, both in terms of performance and flexibility in solving problems in different domains. The original GAT formulation only took into account (un)directed un-weighted networks. In this paper we propose *wsGAT*, an extension of the GAT to cope with signed and weighted networks. We show wsGAT applicability to real-world signed and weighted networks by solving the link prediction task.

We compare wsGAT performance by solving the same task with GCNII [6] and SGCN [8] models, respectively used to perform weighted and signed link prediction. Our results show that models with wsGAT layers outperform the ones with GCNII and SGCN layers.

The paper is organized as in the following. Section 2 formally introduces our proposed model and how it works, whereas Section 3 illustrates the experiments carried out on real-world datasets together with a comparison with other approaches. Finally, Section 4 provides some concluding remarks and ideas for future developments.

## 2 Formulation

In this paper, we extend the *Graph Attention Networks* (GAT) [24] by modifying the computation of the attention coefficient to also account for the (signed) link weight.

As common in the literature, we indicate with  $\mathbf{h}_n^{(k)} \in \mathbb{R}^{F_k}$  the node embedding of node  $i$  after the  $k$ -th GNN layer, where  $F_k$  is the number of features. According to this notation,  $\mathbf{h}_n^{(0)}$  are the node’s input features  $\mathbf{x}_n$ .

In the original GAT formulation, the authors borrow the *attention* mechanism [23], defined to handle variable length sequences and used successfully in the Natural Language Processing (NLP) field, to assign a (relative) importance score to each of the neighbors of the target node. Specifically, they compute the *attention* coefficient  $\alpha_{ij}$  of a node  $i$  for each neighboring node  $j$  as in equation 1.

$$\alpha_{ij}^k = \text{softmax}(\text{LeakyReLU}(\mathbf{e}_i^k))_j \quad (1)$$

$$e_{ij}^k = \mathbf{a}_k^T(\mathbf{W}^k \mathbf{h}_i \parallel \mathbf{W}^k \mathbf{h}_j) \quad (2)$$

where  $\mathbf{W}^k \in \mathbb{R}^{F_k \times F_{k+1}}$  is a learned weight matrix,  $\mathbf{a}_k^\top \in \mathbb{R}^{2 \cdot F_{k+1}}$  is a learned weight vector and  $\parallel$  is the concatenation operator.

The attention coefficient is then used to scale the incoming node embedding of the neighbours as in equation 3.

$$\mathbf{h}_i^{(k+1)} = f\left(\sum_{j \in \mathcal{N}_i \cup \{i\}} \alpha_{ij}^k \mathbf{h}_j^{(k)}\right) \quad (3)$$

where  $f$  is an activation function,  $\mathcal{N}_i$  is the neighbourhood of node  $i$ , and may include the node  $i$  itself if self-loops are added to the network.

The main limitation of this formulation is that the same weight matrix  $\mathbf{W}^k$  is applied independently to both of the embeddings of the target and neighbouring nodes, i.e., they are combined linearly. To achieve better attention scores, other approaches that use Multi-Layer Perceptrons have been proposed [12].

We follow this trend and also account for the (signed) link weight  $w_{ij}$  in the attention computation. In detail, we first modify the computation of  $e_{ij}^k$  as follows:

$$e_{ij}^k = \text{MLP}^k(\mathbf{h}_i^{(k)} \parallel \mathbf{h}_j^{(k)} \parallel w_{ij}) \quad (4)$$

where  $\text{MLP}^k$  is a Multi-Layer Perceptron with the only requirement that the last layer can also produce negative values (e.g., a zero-centred activation function is used) and  $\mathbf{w}_{ij}$  is the weight of the link.

The attention coefficients are then computed as:

$$\alpha_{ij}^k = \text{sign}(e_{ij}) \cdot \text{softmax}(\text{abs}(\mathbf{e}_i^k))_j \quad (5)$$

That is, in our formulation  $\alpha_{ij} \in [-1, 1]$ , meaning that the contribution of each neighbouring node to equation 3 can be positive or negative.

The choice of a Multi-Layer Perceptron allows the network to learn the relative importance of the features of the neighbouring nodes  $j$ , with respect to the ones of the target node  $i$ , and is also affected by the weight and sign of the link between them.

As in the original GAT formulation, *wsGAT* also support multi-head attention, meaning that multiple embeddings for a node can be computed — each using a different set of parameters — and concatenated/sum together.

### 3 Results

To validate the proposed *wsGAT* layer, we test it in the link and weight prediction task on real-world trust networks.

Since, to the best of our knowledge, no other GNN layer can handle both signed and weighted links, we first decompose the final task of signed and weighted link prediction in two sub-tasks and compare our proposal against the state-of-the-art layers. Specifically, we first compare on the link sign prediction with *Signed Graph Convolutional* (SGCN), and on the (unsigned) link weight prediction with *GCNII*.

### 3.1 Dataset

We test our proposal on 4 real-world trust networks. More in detail, we test on the who-trusts-whom networks from the Advogato online community, where trust 4 trust levels can be assigned (corresponding weights are from 0.4 to 1.0 with 0.2 step), from the Bitcoin Alpha and OTC platforms, where scores are on a scale of -10 (total distrust) to +10 (total trust), and from the Epinions.com community, where users can assign a positive or negative trust score to each other. We summarize the networks used for the experiments in Table 1.

Network	V	E	Positive Links	Min. Link Weight	Max. Link Weight	Refs
advogato	6,541	51,127	100%	0	1	[17]
bitcoin-alpha	3,783	24,186	89.98%	-10	10	[14]
bitcoin-otc	5,881	35,592	93.64%	-10	10	[14]
epinions	131,828	841,372	85.29%	-1	1	[18]

**Table 1. Dataset.** Details about the networks used in this paper.

### 3.2 Sign prediction

In the first sub-task, we perform sign prediction — i.e. prediction of the kind of relationship (positive or negative) between two nodes in trust networks — and compare against *SGCN* [8].

The *SGCN* layer, to the best of our knowledge, is the only one able to handle signed links. In particular, they use balance theory and compute two feature sets for each node by splitting the node neighbourhood into two sub-neighbourhoods (i.e., one with all the positive links and the other with all the negative ones). That is, each node has a positive and a negative feature sets. This is a limitation from a Network Science perspective, as the two sub-networks may have different characteristics w.r.t. the original network, or disconnected components may emerge (e.g., in the case of unbalanced link signs). However, authors mitigate this issue by influencing each feature set with the other: when computing the positive node features, they also sum a function of the negative ones, and vice-versa. Another limitation of *SGCNs* is that they do not support link weights, which is useful in many contexts, like the trust one.

For a fair comparison with this approach, we use the same input spectral features and the same train methodology proposed in their paper. In detail, we provide the Signed Spectral Embedding (SSE) from [15] as input node features, and use a node classifier that predicts whether, given a pair of node embeddings, the link between the two nodes is positive, negative or non-existent. During the training phase, we provide 80% of existing links as train examples (and remove the remaining ones from the network), plus the same number of non-existing ones sampled randomly. However, in analogy to their methodology, we predict only the sign of existent links during testing.

We employ the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve and the F1-scores to evaluate the prediction performance. According to sign prediction results, shown in Table 2, *wsGAT* outperform the best *SGCN* algorithm on the three signed networks in our dataset. It is worth noting that the *SGCN* results for the *epinions* network differ from the ones reported by the authors in their paper as we do not filter low-degree nodes from the graph.

GNN Layer	bitcoin-alpha		bitcoin-trust		epinions	
SGCN2	0.796	0.917	0.823	0.925	<b>0.842</b>	0.946
wsGAT	<b>0.832</b>	<b>0.967</b>	<b>0.845</b>	<b>0.953</b>	0.839	<b>0.949</b>

Table 2. Sign prediction results (ROC AUC — F1).

### 3.3 Weight prediction

In the second sub-task, we perform link weight prediction – i.e., predict the (unsigned) strength of the relationship between two nodes —. Here, we compare against *GCNII* [6] that were proposed to simplify and improve the Graph Convolutional Networks (GCN) by Kipf et al. [13]. For both *wsGAT* and *GCNII* we employ the same model architecture: after the GNN layers we use two Multi-Layer Perceptrons; while both take a pair of node embeddings as input, one is trained to predict if the existence of the link between the input nodes, the other is trained to predict the weight. Both MLPs in our tests have fixed number of layers (3) and neurons (100 neurons per layer, 1 output).

Regarding the training, we split the network links into training links (80%) and test links (20%, removed before the training). In addition, for each set we sample the same number of non-existing links to provide the negative examples, and assign a 0 weight to them.

This time we use the ROC AUC and the F1 scores to evaluate the link prediction performance, and we measure the error on the weight prediction (only for existing links) with the Mean Absolute Error (MAE). The weight prediction results are reported in Table 3. *wsGAT* outperform the *GCNII* not only in the link prediction task, but also predict more accurate link weights.

### 3.4 Signed weight prediction

Finally, we merge the two sub-tasks discussed previously and predict the existence of links and of their signed weight.

As in the weight prediction sub-task, we use the AUC and the F1 to measure the link prediction performance, and the MAE to measure the error on the weight prediction.

GNN Layer	advogato			bitcoin-alpha			bitcoin-trust		
GCNII	0.880	0.824	0.158	0.912	0.841	0.1470	0.909	0.830	0.179
wsGAT	<b>0.910</b>	<b>0.839</b>	<b>0.142</b>	<b>0.923</b>	<b>0.851</b>	<b>0.130</b>	<b>0.929</b>	<b>0.860</b>	<b>0.154</b>

**Table 3. Absolute weight prediction results (ROC AUC — F1 — MAE).** Note that while the higher the AUC and F1 scores the better, MAE is an error score and lower values represent smaller errors.

The results on the signed and weighted Bitcoin networks, reported in Table 4, show that the link prediction performance is almost the same as the un-signed case, and the mean absolute error (now on a scale from -10 to +10) drops significantly.

GNN Layer	bitcoin-alpha			bitcoin-trust		
wsGAT	0.922	0.839	0.069	0.921	0.852	0.079

**Table 4. Signed weight prediction results (ROC AUC — F1 — MAE).** Note that while the higher the AUC and F1 scores the better, MAE is an error score and lower values represent smaller errors.

### 3.5 Code availability

*wsGAT* were implemented on top of PyTorch Geometric [9] v1.6.3. Code will be publicly available after the publication of the paper at the following URL <https://github.com/NetworkScienceLab/wsGAT>.

## 4 Conclusions

In this paper we present *wsGAT*, an extension of *Graph Attention Networks* (GAT) to signed and weighted networks. The results in link prediction tasks on signed and weighted real-world trust networks and the comparison with state-of-the-art algorithms confirm the validity of our approach, that provides a useful tool for many research and application scenarios, not limited to link prediction.

## References

1. Bachi, G., Coscia, M., Monreale, A., Giannotti, F.: Classifying trust/distrust relationships in online social networks. In: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing. pp. 552–557. IEEE (2012)
2. Bang-Jensen, J., Gutin, G.Z.: Digraphs: theory, algorithms and applications. Springer Science & Business Media (2008)

3. Borgatti, S.P., Everett, M.G., Johnson, J.C.: Analyzing social networks. Sage (2018)
4. Carchiolo, V., Longheu, A., Malgeri, M., Mangioni, G.: Trust assessment: a personalized, distributed, and secure approach. *Concurrency and Computation: Practice and Experience* 24(6), 605–617 (2012)
5. Carchiolo, V., Longheu, A., Malgeri, M., Mangioni, G.: Users’ attachment in trust networks: reputation vs. effort. *International Journal of Bio-Inspired Computation* 5(4), 199–209 (2013)
6. Chen, M., Wei, Z., Huang, Z., Ding, B., Li, Y.: Simple and deep graph convolutional networks (2020)
7. Chen, Z.J., He, Y., Rosa-Neto, P., Germann, J., Evans, A.C.: Revealing modular architecture of human brain structural networks by using cortical thickness from mri. *Cerebral cortex* 18(10), 2374–2381 (2008)
8. Derr, T., Ma, Y., Tang, J.: Signed graph convolutional network (2018)
9. Fey, M., Lenssen, J.E.: Fast graph representation learning with PyTorch Geometric. In: *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019)
10. Grassia, M., De Domenico, M., Mangioni, G.: Machine learning dismantling and early-warning signals of disintegration in complex systems. *Nature Communications* 12(1), 5190 (Aug 2021), <https://doi.org/10.1038/s41467-021-25485-8>
11. Grassia, M., Lauri, J., Dutta, S., Ajwani, D.: Learning multi-stage sparsification for maximum clique enumeration. *arXiv preprint arXiv:1910.00517* (2019)
12. Hamilton, W.L.: Graph representation learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 14(3), 1–159
13. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks (2017)
14. Kumar, S., Spezzano, F., Subrahmanian, V.S., Faloutsos, C.: Edge weight prediction in weighted signed networks. In: *2016 IEEE 16th International Conference on Data Mining (ICDM)*. pp. 221–230 (2016)
15. Kunegis, J., Schmidt, S., Lommatzsch, A., Lerner, J., Luca, E.W.D., Albayrak, S.: Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization, pp. 559–570. <https://epubs.siam.org/doi/abs/10.1137/1.9781611972801.49>
16. Leskovec, J., Huttenlocher, D., Kleinberg, J.: Signed networks in social media. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. pp. 1361–1370 (2010)
17. Massa, P., Salvetti, M., Tomasoni, D.: Bowling alone and trust decline in social network sites. *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing* pp. 658–663 (2009)
18. Massa, P., Avesani, P.: Controversial users demand local trust metrics: an experimental study on epinions.com community. In: *Proc. American Association for Artif. Intell. Conf.* pp. 121–126 (2005)
19. Mizuno, T., Takayasu, H., Takayasu, M.: Correlation networks among currencies. *Physica A: Statistical Mechanics and its Applications* 364, 336–342 (2006)
20. Newman, M.E.: The structure and function of complex networks. *SIAM review* 45(2), 167–256 (2003)
21. Newman, M.E.: Analysis of weighted networks. *Physical review E* 70(5), 056131 (2004)
22. Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G.: The graph neural network model. *IEEE transactions on neural networks* 20(1), 61–80 (2008)
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need (2017)

24. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks (2018)
25. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S.Y.: A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems* (2020)
26. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M.: Graph neural networks: A review of methods and applications. *AI Open* 1, 57–81 (2020)