Universal Graph Contrastive Learning with a Novel Laplacian Perturbation (Supplementary Material)

Taewook Ko[1] **Yoonhyuk Choi**[1] **Chong-Kwon Kim**[2]

[1]Department of Computer Science and Engineering, Seoul National University
[2]Research Institute of Energy AI, Korea Institute of Energy Technology

A NOTATIONS

We provide a summary of the notations used in this paper and their descriptions in the following tables for reader convenience.

Notation	Description
$\mathcal{G}$	Graph
$\mathcal{G}^+, \mathcal{G}^-$	Positive and negative graph
V , E	Node and edge set
$\mathbf{S}$	Sign matrix
$\boldsymbol{Z}$	Node representation matrix
$\mathbf{A}$, $\mathbf{A}_s$	Adjacency matrix and symmetric adjacency matrix
$\mathbf{D}$, $\mathbf{D}_s$	Degree matrix and symmetric degree matrix
$\mathbf{L}$	Laplacian matrix
$\mathbf{L}^q$	Magnetic Laplacian matrix with parameter q
$\widetilde{\mathcal{G}}$, $\widetilde{\mathbf{L}}^q$	Structure perturbed graph and perturbed magnetic Laplacian
$\mathbf{P}^q$	Phase matrix
$\mathbf{H}^q$	Complex Hermitian adjacency matrix
q	Phase control parameter
$\mathbf{X}$	Input graph signal
$\mathbf{M}$	Projected representations
g	Projected head
$\mathbf{W}$, $\mathbf{b}$	Learnable weight matrix and bias

Table 1: Notations of this paper and its descriptions

B EXPERIMENT DETAILS

B.1 LINK SIGN PREDICTION TASK

B.1.1 Dataset and Metric

We used four signed-directed graph dataset, Bitcoin-Alpha, Bitcoin-OTC, Epinions, and Slashdot which are widely used in signed-directed graph research. Bitcoin-Alpha[1] and Bitcoin-OTC[2] [Kumar et al., 2016] are extracted from Bitcoin trading

[1]http://www.btc-alpha.com
[2]http://www.bitcoin-otc.com

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

platforms. Nodes are users, and edges are user relationships. Users can score the others on a scale of -10 to +10. Edges higher than 0 are treated as positive edges, otherwise negative edges. Epinions[3] [Guha et al., 2004] is a who-trust-whom network crawled from a consumer review site. Users can notate trust or distrust to reviews of other users. Slashdot[4] [Kunegis et al., 2009] is a social network of user community site. Especially they share new information. Users tag others as friends or foes, and we can construct positive and negative edges with this information. The preprocessed datasets can be found at Standford Network Analysis Project (SNAP)[5]. Some papers [Li et al., 2020, Derr et al., 2018] used sub-networks of the originals due to the large network size. We use the whole graph structure for the experiments. In the training phase, we sample positive and negative edges at a ratio of 3:1, but in the validation and test sets, we maintain a natural ratio. It is because the ratio of the positive and negative edges is highly unbalanced. If we train a model with 90 percent of positive samples, a model can easily improve its performance by simply predicting all links are positive. Then, we adopt four metrics, AUC, macro-F1, micro-F1, and binary-F1, for unbiased evaluation.

B.1.2 Implementation Details

Since some graph contrastive baselines are intended for self-supervised learning, we train them with the same semi-supervised loss of the proposed model. Moreover, we removed the read-out process of GraphCL and SimGRACE, which are designed for graph embedding. We ran ten times of experiments with different seed sets for a fair comparison. The seeds are [0, 10, 20, 30, 40, 50, 60, 70, 80, 90]. We apply early stop conditions by comparing the training and validation losses. The model parameters with the lowest validation loss are saved during the training. If validation loss goes up consecutively for more than ten epochs, we stop training and get performance with a test set. We follow the hyperparameter settings of the original papers of each model. The node embedding dimension is set to 128 for all the baselines to make the same learning capacity. The edges are split into 60:20:20 for training, validation, and test sets. However, we did not use all the positive edges as training instances during the training stage. The structure perturbing ratio p and r are set to 0.1 for all datasets. The magnetic Laplacian phase q is perturbed by adding Gaussian noise with a standard deviation of 0.1. The contrastive loss weight α is set to 0.2. Graph encoder stacks two signed-directed spectral convolution layers. We use Adam optimizer with learning rate = 0.001, weight decay = 0.001. All experiments are run 10 times with different seed sets to avoid randomness and get the average score. The experiments are conducted on Xeon E5-2660 v4 and accelerated via Nvidia Titan XP 12G GPU. The software is implemented via Ubuntu v16.4 with python v3.7 and Pytorch v1.12.1.

B.2 NODE CLASSIFICATION TASK

B.2.1 Dataset and Metric

Our experiments utilized five datasets, including three directed citation networks (Cora, Citeseer, and Pubmed) and two undirected co-author networks in Computer Science (CS) and Physics. In the citation networks, nodes correspond to scientific publications and edges represent citations, while in the co-author networks, nodes correspond to researchers and edges represent co-author relations, which are bidirectional. All datasets were preprocessed and made available through the DGL library [6]. To evaluate the performance of our models on the node classification task, we used prediction accuracy as our primary metric.

Dataset	# node	# edge	# features	# class
Cora	2,708	10,556	1,433	7
Citeseer	3,327	9,228	3,703	6
Pubmed	19,717	88,651	500	3
CS	18,333	163,788	6,805	15
Physics	34,493	495,924	8,415	5

Table 2: Dataset statistics.

[3] http://www.epinions.com
[4] http://www.slashdot.com
[5] https://snap.stanford.edu/data/index.html#signnets
[6] https://docs.dgl.ai/

B.2.2 Baselines

We implemented nine baselines to compare the model performance. There are five graph convolution models and four constative learning models.

- **GCN** [Kipf and Welling, 2016] is a spectral graph convolution model with Laplacian matrix.
- **GAT** [Veličković et al., 2017] is a spatial graph convolution model utilizing attention mechanism.
- **APPNP** [Gasteiger et al., 2018] utilize PageRank for efficient propagation scheme.
- **MagNet** [Zhang et al., 2021] defined a magnetic Laplacian for directed graph convolution.
- **DiGCN** [Tong et al., 2020] is a directed graph convolution with directed Laplacian matrix.
- **DiGCL** [Tong et al., 2021] is a graph contrastive model for directed graphs, which perturbs the Directed Laplacian matrix by changing the teleport probability of the transition matrix.

And also used contrastive learning models, **GraphCL**, **GCA**, and **SimGRACE**.

B.2.3 Implementation Details

We followed the same settings as the link-sign prediction experiments, conducting ten runs with different seed sets, applying early stopping criteria, and the same computing resources. The hyper-parameters used were consistent with those of the original papers.

B.2.4 Prediction Performance

The proposed UGCL and its variants consistently demonstrate superior performance across various datasets, with the exception of the pubmed dataset. Despite this, the overall results highlight the wide applicability and effectiveness of UGCL in comparison to other approaches. These findings emphasize the competitive performance of UGCL and its potential as a powerful tool for graph-related tasks.

	Method	Directed			Undirected	
		CORA	CITESEER	PUBMED	CS	Physics
Convolution	GCN	0.761	0.657	0.740	0.818	0.906
	GAT	0.780	0.658	0.771	0.827	0.912
	APPNP	0.769	0.664	0.768	0.823	0.920
	MagNet	0.789	0.683	0.765	0.845	0.914
	DiGCN	0.770	0.669	0.776	0.857	0.914
Contrastive	GraphCL	0.782	0.681	0.763	0.887	0.935
	GCA	0.786	0.688	<u>0.794</u>	0.889	0.940
	SimGRACE	0.791	0.673	**0.795**	0.897	0.941
	DiGCL	<u>0.794</u>	0.672	0.757	0.902	0.927
	UGCL	**0.796**	**0.699**	0.762	**0.916**	**0.955**
	UGCL-S	0.787	0.658	0.764	0.893	<u>0.951</u>
	UGCL-L	0.791	<u>0.692</u>	0.751	<u>0.907</u>	0.940

Table 3: Node classification performance. **Bold** and <u>underline</u> indicate the best and the second performance respectively. The performances are the average score of 10 experiments with different seed sets.

C PROOF OF THEOREMS

Theorem 1. *For a signed directed graph $\mathcal{G} = (V, E, \boldsymbol{S})$, both the unnormalized and normalized magentic Laplacian $\boldsymbol{L}_U^q, \boldsymbol{L}_N^q$ are positive semdifinite.*

proof.
The unnormalized magnetic Laplacian $\mathbf{L}_U^q$ is an Hermitian matrix by its definition. Then, we have Imag($\mathbf{x}^\dagger\mathbf{L}_U^q\mathbf{x}$)=0 where $\mathbf{x} \in \mathbb{C}^N$. Now we show Real($\mathbf{x}^\dagger\mathbf{L}_U^q\mathbf{x}$) ≥ 0. The following procedures utilize the definitions of D_s and A_s.

$$\begin{aligned}
&2\text{Real}(\mathbf{x}^\dagger\mathbf{L}_U^q\mathbf{x}) \\
=&2\sum_{u,v=1}^{N}\mathbf{D_s}(u,v)\mathbf{x}(u)\overline{\mathbf{x}(v)} \\
&-2\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)\mathbf{x}(u)\overline{\mathbf{x}(v)}\left[\frac{\cos(i\Theta^q(uv))+\cos(i\overline{\Theta}^q(uv))}{\|\exp(i\Theta^q(uv))+\exp(i\overline{\Theta}^q(uv))\|+\epsilon}\right] \\
=&2\sum_{u=1}^{N}\mathbf{D_s}(u,u)\mathbf{x}(u)\overline{\mathbf{x}(u)} \\
&-2\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)\mathbf{x}(u)\overline{\mathbf{x}(v)}\left[\frac{\cos(i\Theta^q(uv))+\cos(i\overline{\Theta}^q(uv))}{\|\exp(i\Theta^q(uv))+\exp(i\overline{\Theta}^q(uv))\|+\epsilon}\right]
\end{aligned}$$

$$\begin{aligned}
=&2\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)|\mathbf{x}(u)|^2 \\
&-2\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)\mathbf{x}(u)\overline{\mathbf{x}(v)}\left[\frac{\cos(i\Theta^q(uv))+\cos(i\overline{\Theta}^q(uv))}{\|\exp(i\Theta^q(uv))+\exp(i\overline{\Theta}^q(uv))\|+\epsilon}\right] \\
=&\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)|\mathbf{x}(u)|^2+\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)|\mathbf{x}(v)|^2 \\
&-2\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)\mathbf{x}(u)\overline{\mathbf{x}(v)}\left[\frac{\cos(i\Theta^q(uv))+\cos(i\overline{\Theta}^q(uv))}{\|\exp(i\Theta^q(uv))+\exp(i\overline{\Theta}^q(uv))\|+\epsilon}\right] \\
=&\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)\left(|\mathbf{x}(u)|^2+|\mathbf{x}(v)|^2-2\mathbf{x}(u)\overline{\mathbf{x}(v)}\left[\frac{\cos(i\Theta^q(uv))+\cos(i\overline{\Theta}^q(uv))}{\|\exp(i\Theta^q(uv))+\exp(i\overline{\Theta}^q(uv))\|+\epsilon}\right]\right) \\
\geq&\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)(|\mathbf{x}(u)|^2+|\mathbf{x}(v)|^2-2|\mathbf{x}(u)|\left|\overline{\mathbf{x}(v)}\right|) \\
=&\sum_{u,v=1}^{N}\mathbf{A_s}(u,v)(|\mathbf{x}(u)|-|\mathbf{x}(v)|)^2 \\
\geq&0.
\end{aligned}$$

Thus, $\mathbf{x}^\dagger\mathbf{L}_U^q\mathbf{x} \geq 0$ for $\mathbf{x} \in \mathbb{C}^N$, positive semi-definite.

For normalized Laplacian matrix, $\mathbf{L}_N^q = \mathbf{D}_s^{-1/2}\mathbf{L}_U^q\mathbf{D}_s^{-1/2}$.

$$\begin{aligned}
\mathbf{x}^\dagger\mathbf{L}_N^q\mathbf{x} &= \mathbf{x}^\dagger\mathbf{D}_s^{-1/2}\mathbf{L}_U^q\mathbf{D}_s^{-1/2}\mathbf{x} \\
&= \mathbf{y}^\dagger\mathbf{L}_U^q\mathbf{y} \\
&\geq 0.
\end{aligned}$$

where, $\mathbf{y} = \mathbf{D}_s^{-1/2}\mathbf{x}$.
Thus, both unnormalized and normalized magnetic Laplacians are positive semi-definite.

Theorem 2. *For a signed directed graph $\mathcal{G} = (V, E, \boldsymbol{S})$, the eigenvalues of the normalized magnetic Laplacian $\boldsymbol{L}_N^q$ lie in [0, 2].*

proof.
$\mathbf{L}_N^q$ has non-negative and real eigenvalues since it is positive semi-definite by Theorem.1. Now, we show the eigenvalues are less than or equal to 2. Here, we use the Courant-Fischer theorem [Golub and Van Loan, 2013],

$$\begin{aligned}\lambda_N &= \max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^\dagger \mathbf{L}_N^q \mathbf{x}}{\mathbf{x}^\dagger \mathbf{x}} \\ &= \max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^\dagger \mathbf{D}_s^{-1/2} \mathbf{L}_U^q \mathbf{D}_s^{-1/2} \mathbf{x}}{\mathbf{x}^\dagger \mathbf{x}} \\ &= \max_{\mathbf{y}\neq 0} \frac{\mathbf{y}^\dagger \mathbf{L}_U^q \mathbf{y}}{\mathbf{y}^\dagger \mathbf{D}_s \mathbf{y}}.\end{aligned}$$

where, $\mathbf{y} = \mathbf{D}_s^{-1/2}\mathbf{x}$. Since $\mathbf{D}_s$ is diagonal,

$$\mathbf{y}^\dagger \mathbf{D}_s \mathbf{y} = \sum_{u,v=1}^{N} \mathbf{D}_s(u,v)\mathbf{y}(u)\overline{\mathbf{y}(v)} = \sum_{u=1}^{N} \mathbf{D}_s(u,u)|\mathbf{y}|^2$$

Similar to Theorem 1, we have

$$\begin{aligned}&\mathbf{y}^\dagger \mathbf{L}_U^q \mathbf{y} \\ =&\frac{1}{2}\sum_{u,v=1}^{N} \mathbf{A}_s(u,v)\left(|\mathbf{y}(u)|^2 + |\mathbf{y}(v)|^2 - 2\mathbf{y}(u)\overline{\mathbf{y}(v)}\frac{\cos(i\Theta^q(uv)) + \cos(i\overline{\Theta}^q(uv))}{\|\exp(i\Theta^q(uv)) + \exp(i\overline{\Theta}^q(uv))\| + \epsilon}\right) \\ \leq&\frac{1}{2}\sum_{u,v=1}^{N} \mathbf{A}_s(u,v)(|\mathbf{y}(u)|^2 + |\mathbf{y}(v)|^2) \\ \leq& \sum_{u,v=1}^{N} \mathbf{A}_s(u,v)(|\mathbf{y}(u)|^2 + |\mathbf{y}(v)|^2) \\ \leq&2\sum_{u,v=1}^{N} \mathbf{A}_s(u,v)|\mathbf{y}(u)|^2 \qquad \text{(since } \mathbf{A}_s \text{ is symmetric)} \\ =&2\sum_{u=1}^{N} |\mathbf{y}(u)|^2 \left(\sum_{v=1}^{N} \mathbf{A}_s(u,v)\right) \\ =&2\sum_{u=1}^{N} |\mathbf{y}(u)|^2 \mathbf{D}_s(u,u) \\ =&2\mathbf{y}^\dagger \mathbf{D}_s \mathbf{y}.\end{aligned}$$

Thus,

$$\lambda_N = \max_{\mathbf{y}\neq 0} \frac{\mathbf{y}^\dagger \mathbf{L}_U^q \mathbf{y}}{\mathbf{y}^\dagger \mathbf{D}_s \mathbf{y}} \leq \max_{\mathbf{y}\neq 0} \frac{2\mathbf{y}^\dagger \mathbf{D}_s \mathbf{y}}{\mathbf{y}^\dagger \mathbf{D}_s \mathbf{y}} = 2.$$

Finally, the eigenvalues of normalized magnetic Laplacian are between [0, 2].

Proposition 1. *Let a $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be a directed graphs on the same vertex set. Then their union $\mathcal{G} = (V, E_1 \cup E_2)$ has entropy $H(\mathcal{G}) \leq H(\mathcal{G}_1) + H(\mathcal{G}_2)$.*

proof.
Let $p_1(x, y)$ and $p_2(x, y)$ be the distributions that minimize $I(X \wedge Y)$ for $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. Then we have a joint distribution with Bayes' rule

$$p(x, y_1, y_2) = p(x) \cdot p_1(y_1|x) \cdot p_2(y_2|x).$$

For a given choice of X, observe the $Y_1 \cap Y_2$ contains X and is an independent set in $\mathcal{G}$. Therefore,

$$
\begin{aligned}
H(\mathcal{G}) &\leq I(X \wedge (Y_1 \cap Y_2)) \\
&\leq I(X \wedge Y_1, Y_2) \\
&= H(Y_1, Y_2) - H(Y_1, Y_2|X) \\
&= H(Y_1, Y_2) - H(Y_1|X) - H(Y_2|X) \\
&\leq H(Y_1) - H(Y_1|X) + H(Y_2) - H(Y_2|X) \\
&= H(\mathcal{G}_1) + H(\mathcal{G}_2).
\end{aligned}
$$

Theorem 3. *Von Neumann entropy of a signed directed graph can be expressed via two directed graph entropy.*

proof.
For a signed directed graph, $\mathcal{G} = (V, E, \mathbf{S})$, we can split it into two directed graphs via the edge type. Extract positive edges from E and S then construct a directed graph with node set V. Now we have a positive directed graph $\mathcal{G}^+ = (V, E^+)$. Similarly, we have a negative directed graph $\mathcal{G}^- = (V, E^-)$. Therefore, by utilizing Proposition 1.

$$
H(\mathcal{G}) \leq H(\mathcal{G}^+) + H(\mathcal{G}^-).
$$

Proposition 2. *Let $\mathcal{G} = (V, E)$ and $\mathcal{F} = (V, E')$ are graphs with same the same vertex set V and $\mathcal{F}$ is a subgraph of $\mathcal{G}$, $E' \subset E$. Then the entropy is, $H(\mathcal{F}) \leq H(\mathcal{G})$*

proof.
If X, Y are random variables achieving $H(\mathcal{G})$, then Y is also an independent set in $H(\mathcal{F})$. Therefore, $H(\mathcal{F}) \leq I(X \wedge Y) = H(\mathcal{G})$

Theorem 4. ***Perturbation Error of a Signed Directed Graph***

proof.
By Definition 1, we have perturbation error of a graph as:

$$
\Delta H(\mathcal{G}, q, \Delta q) = H(\mathcal{G}, q) - H(\mathcal{G}, q + \Delta q).
$$

Since $H(\mathcal{G}, q) \leq H(\mathcal{G}_D^+, q) + H(\mathcal{G}_D^-, q)$ and $H(\mathcal{G}, q + \Delta q) \leq H(\mathcal{G}_D^+, q + \Delta q) + H(\mathcal{G}_D^-, q + \Delta q)$, we have the following results.

$$
\begin{aligned}
\Delta H(\mathcal{G}, q, \Delta q) &\leq H(\mathcal{G}_D^+, q) + H(\mathcal{G}_D^-, q) - H(\mathcal{G}_D^+, q + \Delta q) - H(\mathcal{G}_D^-, q + \Delta q) \\
&= H(\mathcal{G}_D^+, q) - H(\mathcal{G}_D^+, q + \Delta q) + H(\mathcal{G}_D^-, q) - H(\mathcal{G}_D^-, q + \Delta q) \\
&= \Delta H(\mathcal{G}_D^+, q, \Delta q) + \Delta H(\mathcal{G}_D^-, q, \Delta q)
\end{aligned}
$$

And by Proposition 2,

$$
\begin{aligned}
\Delta H(\mathcal{G}_D^+, q, \Delta q) &\leq \Delta H(\mathcal{G}, q, \Delta q) \\
\Delta H(\mathcal{G}_D^-, q, \Delta q) &\leq \Delta H(\mathcal{G}, q, \Delta q)
\end{aligned}
$$

Therefore, a signed directed graph perturbation error is described in the lower and upper boundaries.

References

Tyler Derr, Yao Ma, and Jiliang Tang. Signed graph convolutional networks. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 929–934. IEEE, 2018.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv preprint arXiv:1810.05997*, 2018.

Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.

Ramanthan Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, pages 403–412, 2004.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

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.

Jérôme Kunegis, Andreas Lommatzsch, and Christian Bauckhage. The slashdot zoo: mining a social network with negative edges. In *Proceedings of the 18th international conference on World wide web*, pages 741–750, 2009.

Yu Li, Yuan Tian, Jiawei Zhang, and Yi Chang. Learning signed network embedding via graph attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 4772–4779, 2020.

Zekun Tong, Yuxuan Liang, Changsheng Sun, Xinke Li, David Rosenblum, and Andrew Lim. Digraph inception convolutional networks. *Advances in neural information processing systems*, 33:17907–17918, 2020.

Zekun Tong, Yuxuan Liang, Henghui Ding, Yongxing Dai, Xinke Li, and Changhu Wang. Directed graph contrastive learning. *Advances in Neural Information Processing Systems*, 34:19580–19593, 2021.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Xitong Zhang, Yixuan He, Nathan Brugnone, Michael Perlmutter, and Matthew Hirn. Magnet: A neural network for directed graphs. *Advances in Neural Information Processing Systems*, 34:27003–27015, 2021.