Energy Levels Based Graph Neural Networks for Heterophily

Qian Wang¹, Youfa Liu².*
¹College of Science, Huazhong Agricultural University
²College of Informatics, Huazhong Agricultural University
Corresponding author: Youfa Liu, e-mail: liuyoufa@mail.hzau.edu.cn

Abstract. The representation power of graph neural networks (GNNs) under heterophily has drawn much attention recently, in which some connected nodes do not have common labels or similar features. Previous approaches typically incorporate high-order neighbourhoods’ information so as to reinforce aggregation operation. However, they do not directly face with heterophily in first-order neighbours. To deal with this problem, we propose the energy levels guided GNNs in which energy levels of classes are incorporated for better expressiveness. We prove that the label with higher energy level would imply higher energy in the spectrum of the normalized graph Laplacian. Experiments on common benchmark datasets show the effectiveness of the proposed approach.

1. Introduction
Graph neural networks (GNNs) can effectively deal with structural data. There are rich application fields of GNNs, such as social networks[1], bioinformatics[2], knowledge graph[3] and many others. Various architectures (GCN[4], GraphSAGE[4]) of GNNs are devised for semi-supervised node classification tasks in which we only know part of label information of nodes in a graph and they can outperform traditional graph learning approaches.

The expressiveness power of the GNNs is one of the attracted topics. For example, GCN[4] and GraphSAGE[4] cannot distinguish some simple graph structures. Different network architecture design may result in the significant difference of practical performance. According to the degree of the edge homophily ratio[6], graphs could be coarsely categorized into two types, i.e. homophily and heterophily graph. We give an example of two kinds of graphs in Figure1.

Figure 1. The left and right graph of vertical dashed line belong to homophily and heterophily graph respectively.
This paper concerns the heterophilic setting in which connected nodes in a graph does not share the same label or have similar feature representation. For example, in dating network, connected nodes almost with different labels because people often likes dating with the opposite gender.

Few works focus on heterophily setting, for example, H2GCN[6]. H2GCN utilizes three different strategies to enhance the expressive power of network under heterophily setting guided by perturbation analysis and graph signal processing. These strategies include ego- and neighbourhood-embedding separation, high-order neighbourhoods and combination of intermediate representations. Although H2GCN is effective on some realistic datasets, the work[6]has a main drawback: Classes in the neighbourhoods are equally treated and lack of discrimination in the process of aggregation.

To deal with these drawbacks, we propose the energy levels guided graph neural networks (ELG2NN) for heterophily. The contributions of the proposed approach are as follows:

1. We introduce energy levels of classes. This induces the discriminant (or hierarchical) information of different classes, which can make the aggregation more discriminative. We prove that energy levels have an affinity with energy in graph signal processing.

2. Experiments on synthetic and real datasets demonstrate the effectiveness of the proposed approach.

The rest of this paper is organized as follows. In Section 2, we briefly formulate some related works. In section 3, we give problem setup and notations. In Section 4, we propose energy levels based GNNs. Then we conduct experiments on real datasets. Finally, we conclude this paper.

2. Related Works

GCN[4] is proposed as an extension of deep convolutional neural networks to graph-structural data. The different aggregation operation over neighbourhoods could produce different kinds of architecture, such as GraphSAGE[4], GIN[7] and so on. Exploring effective graph pooling techniques also gives rise to a large amount of works, such as SOPOOL[8], ATTNPOOL[9], CONVPOOL[10] and so on. Although these strategies are empirically effective, the expressive power of networks still draw researchers attention. GIN[7] analyzes the expressive of GNNs to distinguish different graph structures. It relates GNNs with Weisfeiler-Lehman (WL) graph isomorphism test[11]. Recent work H2GCN[6] considers the expressive power of GNNs under heterophily. Many popular GNNs do not work under this case and are even worse than models that do not take the graph structure into account. It takes three strategies and works well under heterophily with theoretical justification.

Our proposed approach ELG2NN has four strategies: (S1) ego- and neighbor-embedding separation; (S2) higher-order neighborhoods; (S3) Combination of intermediate representations; (S4) Emphasize the importance of classes in aggregation which is measured by the energy levels of classes. The following table displays the difference among the proposed approach and recent works including GCN[4], GCN-Chey[12], GraphSAGE[4], MixHop[13] and H2GCN[6].

3. Problem Setup and Notations

Let $G = (V; E)$ be an undirected graph whose vertex set is $V$ and edge set is $E$. Each node $v$ has a feature description $x_v \in \mathbb{R}^m$. For any vertex $v$ of $G$, $N(v)$ denote its (first-order) neighbourhood. $N^2(v)$ denotes the second-order neighborhood of $v$, i.e. those vertices with minimum distance 2 to $v$. Given a set of samples $\{(v_i, y_i)\}_{i=1}^M$ from a graph $G$, where $v_i$ is the $i$-th node mark and $y_i$ is its label. The aim is to train a model based on these training samples and predict labels for residual nodes $\{v_{M+1}, v_{M+2}, \ldots, v_N\}$ in the graph $G$, where $\#V$ denotes the number of nodes in the graph $G$.

For any $v \in V$, let $N^0(v) = N(v) \setminus \{v\}$ and $N^2(v) = N^2(v) \setminus \{v\}$ respectively. $d_v$ denotes the degree of the node $v$. $K$ denotes the set of classes. Let $A$ be the adjacent matrix of graph $G$. Let $L$
be the graph Laplacian. \( L = D - A \), where \( D \) is a diagonal matrix whose diagonal elements are represented by \( D_{ii} = \sum_j A_{ij} \). The normalized graph Laplacian is \( \tilde{L} = I_{\#V} - \tilde{A} \), where \( \tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \) and \( I_{\#V} \) is the identity matrix of order \( \#V \). \( \lambda_{\text{max}}(\tilde{L}) \) denotes the largest eigenvalue of \( \tilde{L} \).

### 4. The Proposed Approach

In this section, we introduce the energy levels of classes which emphasize the importance of classes in a neighborhood for aggregation. We also provide the relevant theoretical justification.

#### 4.1. Energy Levels

Graph signal over graph \( G \) is a \( \#V \)-dimensional vector whose components lie in the set \( \{0, 1\} \). We use the following one-hot encodings for labels. Let

\[
S = [s^{(1)}, \ldots, s^{(k)}]
\]

where \( s^{(k)} \) denotes the one-hot class encoding of class \( k \). That is, if the label is \( k \), its corresponding component is 1; otherwise, its component is 0. Assume that the eigenvalues of \( \tilde{L} \) satisfy \( 0 < \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{\#V-1} \). Let \( a_1, a_2, \ldots, a_{\#V-1} \) be orthonormal basis which makes \( \tilde{L} \) diagonalizable.

For any class \( k \), we can represent \( s^{(k)} \) as \( s^{(k)} = \sum_{i=0}^{\#V-1} s_{k,i} a_i \) where \( s_{k,i} \) represents the frequency component. Recall from[6] that the graph signal \( s^k \) has higher energy that \( s^l \) on high-frequency components if there exists an integer \( 0 < M \leq \#V-1 \) such that

\[
\sum_{i=M}^{\#V-1} s_{k,i}^2 > \sum_{i=M}^{\#V-1} s_{l,i}^2
\]

(2)

In this case, we say that the class \( k \) has higher energy on high-frequency components. In what follows, we propose the energy levels for classes and then reveal how they induce energy level on high-frequency components in Theorem 1.

The energy level for node label \( k \) is

\[
E_k = 1 - \frac{1}{2\lambda_{\text{max}}(\tilde{L}) \#T} \sum_{v \in V} \sum_{u \in N(v)} \left( \frac{s_u^{(k)}}{\sqrt{d_u}} - \frac{s_v^{(k)}}{\sqrt{d_v}} \right)^2,
\]

(3)

where

\[
T = \max_{k \in K} \#\{v \in V : y_v = k\},
\]

(4)

\( s^{(k)} \in \mathbb{R}^{\#V} \) is the graph signal corresponding to the class \( k \) and \( \tilde{L} \) is the normalized graph Laplacian.

We give an example on the computation of energy levels in Figure 2. Then we prove that the energy levels have an affinity with energy in graph signal processing.

**Theorem 1.** If the energy levels of class \( k \) and \( l \) satisfy \( E_k > E_l \), then the class \( k \) has higher energy on high-frequency components than \( l \) in the spectrum of the normalized graph Laplacian \( \tilde{L} \).
Figure 2. The illustration of (group) aggregation over $N^v(\nu)$ under heterophily. This aggregation is different from traditional aggregation operation in which all classes are equally treated and hence is not discriminant under heterophily. ‘$\oplus$’ denotes weighted average. The weight is determined by the energy levels of classes.

Proof. Since $E_k > E_l$, we have

$$\frac{1}{2} \sum_{u \in V} \sum_{v \in N(v)} \left( \frac{s^{(k)}_u}{d_u} - \frac{s^{(k)}_v}{d_v} \right)^2 > \frac{1}{2} \sum_{u \in V} \sum_{v \in N(v)} \left( \frac{s^{(l)}_u}{d_u} - \frac{s^{(l)}_v}{d_v} \right)^2.$$ (5)

Hence

$$(s^{(k)})^T \tilde{L}s^{(k)} > (s^{(l)})^T \tilde{L}s^{(l)}$$ (6)

Regard $s^{(k)}$ and $s^{(l)}$ as two graph signals, we prove that signal $s^{(k)}$ has higher energy than $s^{(l)}$ on high-frequency components than $s^{(l)}$. This is done by contradiction. Assume that, for any integer $0 < M \leq \#V$, \( \sum_{i=M}^{\#V-1} s^{(k)}_{i,j} > \sum_{i=M}^{\#V-1} s^{(l)}_{i,j} \) always holds. Then

$$\sum_{i=1}^{\#V-1} (\lambda_i - \lambda_{i-1}) \sum_{j=1}^{\#V-1} s^{(k)}_{i,j} \leq \sum_{i=1}^{\#V-1} (\lambda_i - \lambda_{i-1}) \sum_{j=1}^{\#V-1} s^{(l)}_{i,j}$$ (7)

Note that

$$\lambda_0 \sum_{i=1}^{\#V-1} s^{(k)}_{i,j} = 0 = \lambda_0 \sum_{i=1}^{\#V-1} s^{(l)}_{i,j}.$$ (8)

Combining these two inequalities, we have

$$\sum_{i=0}^{\#V-1} \lambda_i s^{(k)}_{i,j} \leq \sum_{i=0}^{\#V-1} \lambda_i s^{(l)}_{i,j}$$ (9)

However, by $(s^{(k)})^T \tilde{L}s^{(k)} > (s^{(l)})^T \tilde{L}s^{(l)}$, we have

$$\sum_{i=0}^{\#V-1} \lambda_i s^{(k)}_{i,j} > \sum_{i=0}^{\#V-1} \lambda_i s^{(l)}_{i,j}.$$ (10)

This contradicts with the assumption.

4.2. ELG$_2$NN
Considering hierarchy from the order of neighbourhood and energy level, we propose a hierarchical graph neural networks (ELG$_2$NN) framework for heterophily. There are three steps in all.
First, we perform feature embedding. Specifically, we utilize a fully-connected layer to map initial features \( \{x_v\}_{v \in V} \) into a latent low-dimensional subspaces and produce latent representations \( \{m_v^{(0)}\}_{v \in V} \). Second, layer-by-layer aggregation and combination are constructed. We apply the iterative formula \( m_v^{(l)} = C(m_v^{(l-1)}, \Phi_1, \Phi_2) \), where

\[
\Phi_1 = \mathcal{A}\{m_u^{(l-1)} \mid u \in N_1(v)\}, \quad \Phi_2 = \mathcal{A}\{m_u^{(l-1)} \mid u \in N_2(v)\},
\]

in which \( \mathcal{A} \) and \( C \) denote the aggregation and combination under hierarchy of the order of neighbourhood and energy level respectively. Specifically, \( \Phi_1 = \sum_{I \in \mathcal{J}_1} \gamma_I \sum_{u \in N_1(v), y_u = y_I} \frac{m_u^{(k-1)}}{d_u \sqrt{d_v}} \), where \( \mathcal{J}_1 \subseteq \mathcal{K} \) denotes the set of classes w.r.t \( N_1(v) \) and the weight is given by energy levels, i.e. \( \gamma_I = E_I \). Hence, nodes from those class with higher energy would make more contribution to the aggregation operation. Similarly, \( \Phi_2 = \sum_{I \in \mathcal{J}_2} \gamma_I \sum_{u \in N_2(v), y_u = y_I} \frac{m_u^{(k-1)}}{d_u \sqrt{d_v}} \), where \( \mathcal{J}_2 \subseteq \mathcal{K} \) denotes the set of classes w.r.t. \( N_2(v) \), \( \gamma_I \) is the weighted value determined by ratio of energy levels of classes, \( d_{v,2} \) denotes the degree in the sense of second-order neighbourhood. \( C \) is a concatenation operator which combines information from neighbourhoods of different orders. Finally, we combine all the intermediate representations and get the final representation

\[
\tilde{m}_v = \text{COMBINE}(m_v^{(0)}, m_v^{(1)}, \ldots, m_v^{(K)}),
\]

where \( K \) is an iterative parameter specified by user. We empirically set ‘COMBINATION’ operation as concatenation. Finally, we use multi-layer perceptrons (MLP) with softmax operator to predicate the labels. In experiments, MLP is instantiated by a fully connected layer.

5. Experiments

In this section, we evaluate the effectiveness of the proposed ELG\(_2\)NN on real datasets.

5.1. Evaluation on Real Datasets

1) Datasets: We conduct experiments on widely used domain datasets [16][17][17][19][20][21] which include Texas, Wisconsin, Actor, Squirrel, Chameleon, Cornell, Cora Full, Citeseer, Pubmed and Cora.

2) Experimental Setup: For Cora-Full, the training, validation and test splits are 25%, 25% and 50% for each class, respectively. For residual datasets, we randomly select 48%/32%/20% of nodes per class for train/validation/test provided by [22].

3) Analysis: We compare the proposed ELG\(_2\)NN with existing approaches, such as H\(_2\)GCN[6], GCN-Cheby[12], GraphSAGE[4], MixHop[13], GCN[4], ChebyJK[23], GAT[24], and MLP with only one hidden layer. From Table 1, we can see that, except for Squirrel and Chameleon dataset, the proposed ELG\(_2\)NN can outperform state-of-the-art approaches. The cause may be attributed to the introduction of energy levels of classes which gives rise to discriminative (group) aggregation. The node features for Squirrel and Chameleon are not high-quality so that the performance of all methods are not well.
Table 1. Graph classification accuracies (%) of ELG\textsuperscript{2}NN on real datasets.

| Methods     | Texas     | Wisconsin | Actor  | Squirrel | Chameleon | Cornell | CoraFull | Citeseer | Pubmed | Cora |
|-------------|-----------|-----------|--------|----------|-----------|---------|----------|----------|--------|------|
| MLP         | 81.89±4.78 | 85.29±3.61 | 35.76±0.98 | 29.68±1.81 | 46.36±2.52 | 81.08±6.37 | 58.76±0.50 | 72.41±2.18 | 86.65±0.35 | 74.75±2.22 |
| GCN         | 59.46±5.25 | 59.01±6.99 | 30.26±0.79 | 36.89±1.34 | 59.82±2.58 | 57.03±4.67 | 68.39±0.32 | 76.68±1.64 | 87.38±0.66 | 87.28±1.26 |
| GAT         | 58.38±4.45 | 55.29±8.71 | 30.62±1.11 | 54.69±1.95 | 59.82±2.58 | 57.03±4.67 | 68.39±0.32 | 76.68±1.64 | 87.38±0.66 | 87.28±1.26 |
| Cheby4K     | 78.38±6.37 | 82.55±4.57 | 35.14±1.37 | 45.03±1.73 | 63.79±2.27 | 74.59±7.87 | 66.87±0.29 | 76.98±1.18 | 89.07±0.30 | 85.49±2.27 |
| MixHop      | 77.84±7.73 | 75.88±4.90 | 32.22±2.34 | 43.80±1.48 | 50.50±2.53 | 73.51±6.34 | 65.59±0.34 | 76.26±1.33 | 85.31±0.61 | 87.61±0.85 |
| GCN-Cheby   | 77.30±4.97 | 79.41±4.46 | 34.11±1.09 | 43.86±1.64 | 55.24±2.76 | 74.32±7.46 | 67.41±0.69 | 75.82±1.53 | 88.72±0.55 | 86.76±0.95 |
| GraphSAGE   | 82.43±6.14 | 81.18±5.56 | 34.23±0.99 | 41.61±0.74 | 58.73±1.68 | 75.95±5.01 | 65.14±0.75 | 76.04±1.30 | 88.45±0.30 | 86.90±1.04 |
| H2GCN-1     | 84.86±6.77 | 85.67±4.69 | 35.86±1.03 | 36.42±1.89 | 57.11±1.58 | 82.16±4.80 | 68.15±0.49 | 77.07±1.64 | 89.40±0.34 | 86.92±1.37 |
| H2GCN-2     | 82.16±5.28 | 85.88±4.22 | 35.62±1.30 | 37.90±2.02 | 59.39±1.98 | 82.16±4.00 | 69.05±0.37 | 76.88±1.77 | 89.59±0.33 | 87.81±1.35 |
| ELG\textsuperscript{2}NN | 85.21±6.64 | 86.92±4.71 | 37.41±1.05 | 40.57±1.85 | 62.17±1.72 | 83.32±5.82 | 71.37±0.38 | 78.49±1.69 | 90.12±0.36 | 88.46±1.58 |

6. Conclusion

Graph with heterophily occurs frequently but few works on GNNs could deal with it. In this paper, we propose the energy levels guided graph neural networks (ELG\textsuperscript{2}NN) framework for heterophily. We then introduce the energy levels for classes which can distinguish the importance of different classes and prove that energy levels have affinity with the energy on frequency components. This induces a new kind of aggregation operation which emphasizes on the important of different classes. Experiments real datasets verify the effectiveness of the proposed approach.

Acknowledgement

Thanks for the valuable reviews from anonymous reviewer(s). This work is fully supoorted by Project 2662020XXQD002 supported by the Fundamental Research Funds for the Central Universities.

References

[1] Fan W, Ma Y, Li Q, He Y, Zhao E, Tang J and Yin D 2019 Graph neural networks for social recommendation AAAI. pp. 417-426
[2] Bove P, Micheli A, Milazzo P and Podda M 2020 Prediction of dynamical properties of biochemical pathways with graph neural networks Bioinformatics. pp. 32-43
[3] Wang Z, Lv Q, Lan X and Zhang Y 2018 Cross-lingual knowledge graph alignment via graph convolutional networks EMNLP. pp. 349-357
[4] Kipf T. N and Welling M 2016 Semi-supervised classification with graph convolutional networks ArXiv. arXiv:1609.02907
[5] Hamilton W, Ying Z and Leskovec J 2017 Inductive representation learning on large graphs NeurIPS. pp. 1025-1035
[6] Zhu J, Yan Y, Zhao L, M Heimann, Akoglu L and Koutra D 2020 Beyond homophily in graph neural networks: Current limitations and effective designs NeurIPS. pp. 7793-7804
[7] Xu K, Hu W, Leskovec J and Jegelka S 2019 How powerful are graph neural networks? ICLR. pp. 1-17
[8] Wang Z and Ji S 2020 Second-order pooling for graph neural networks IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109
[9] Girdhar R and Ramanan D 2017 Attentional pooling for action recognition NeurIPS. pp. 33-44
[10] Verma S and Zhang Z. L 2018 Graph capsule convolutional neural networks ArXiv. arXiv:1805.08090
[11] Weisfeiler B and Lehman A 1968 A reduction of a graph to a canonical form and an algebra arising during this reduction Nauchno-Technicheskaya Informatsia. 2:12-16
[12] Defferrard M, Bresson X and Vandergheynst P 2016 Convolutional neural networks on graphs with fast localized spectral filtering NeurIPS. pp. 3844-52
[13] Abu-El-Haija S, Perozzi B, Kapoor A, Harutyunyan H, Alipourfard N, Lerman K, Steeg G. V and Galstyan A 2019 Mixhop: Higher-order graph convolution architectures via sparsified neighborhood mixing ICML. 97:21-29
[14] Barabasi A. L and Albert R 1999. Emergence of scaling in random networks Science. pp. 509-512
[15] Karimi F, Genois M, Wagner C, Singer P and Strohmaier M 2017 Visibility of minorities in social networks ArXiv. arXiv:1702.00150
[16] Bojchevski A and Gunnewick S 2018 Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking ArXiv. arXiv:1707.03815
[17] Namata G, London B, Getoor L and Huang B 2012 Query-driven active surveying for collective classification Proc. MLG. pp. 1-8
[18] Rozemberczki B, Allen C and Sarkar R 2019 Multi-scale attributed node embedding ArXiv. arXiv:1909.13021
[19] Sen P, Namata G, Bilgic M, Getoor L, Galligher B and Eliassi-Rad T 2008 Collective classification in network data AI magazine. 29(3):93
[20] Shchur O, Mumme M, Bojchevski A and Gunnewick S 2018 Pitfalls of graph neural network evaluation NeurIPS. pp. 1-11
[21] Tang J, Sun J, Wang C and Yang Z 2009 Social influence analysis in large-scale networks ACM SIGKDD. pp. 807-816
[22] Pei H, Wei B, Chang K. C-C, Lei Y and Yang B 2020 Geom-gcn: Geometric graph convolutional networks ArXiv. arXiv:2002.05287
[23] Xu K, Li C, Tian Y, Sonobe T, Kawarabayashi K and Jegelka S 2018 Representation learning on graphs with jumping knowledge networks ICML. 80:5453-62
[24] Veličković P, Cucurull G, Casanova A, Romero A, Liò P and Bengio Y 2017 Graph attention networks ArXiv. arXiv:1710.10903