Revisiting LightGCN: Unexpected Inflexibility, Inconsistency, and A Remedy Towards Improved Recommendation (Rebuttal Details)

Anonymous Author(s)

1 THEORETICAL ANALYSIS

In this section, we provide the theoretical analysis that supports our empirical observations. We first examine the norm of the aggregated embedding of uncorrelated (i.e., independent) neighbors. Then, we generalize this to neighbors with correlations. Lastly, we discuss the space complexity of LightGCN++.

1.1 Norm of Aggregated Embedding of Uncorrelated Neighbors

To demonstrate that our Observation 1 is not mathematically trivial, we show that embeddings form simple distributions do not exhibit the observation. For example, if we simply assume that neighbor embeddings are independently sampled from a normal distribution, the expected norm of the unscaled aggregated embeddings is not linear w.r.t. the number of neighbors.

Theorem 1. *Let M be a set of d-dimensional embeddings $\{\mathbf{x}_i\}_{i=1}^{|M|}$, where each dimension of $\mathbf{x}_i \in M$ is independently drawn from a normal distribution $N(0, \sigma^2)$. If the embeddings are uncorrelated (i.e., independent) sample-wise, the expected L2 norm of the unscaled aggregated embeddings in M is proportional to $\sqrt{|M|}$:*

$$\mathbb{E}\left[\left\|\sum_{\mathbf{x}_i \in M} \mathbf{x}_i\right\|\right] \propto \sqrt{|M|}$$

Proof. Let $\mathrm{X}_j^{(M)} = \sum_{\mathbf{x}_i \in M} \mathbf{x}_{ij}$. Based on the property of the sum of normal distributions, $\mathrm{X}_j^{(M)} \sim N(0, |M|\sigma^2)$. Thus, $\mathrm{X}_1^{(M)}/\sqrt{|M|\sigma^2}$, $\cdots, \mathrm{X}_d^{(M)}/\sqrt{|M|\sigma^2}$ are d independent random variables from $N(0, 1)$. This implies that the following statistic is distributed according to the chi distribution with d degrees of freedom:

$$\sqrt{\sum_{j=1}^{d}\left(\frac{\mathrm{X}_j^{(M)}}{\sqrt{|M|\sigma^2}}\right)^2} = \frac{1}{\sqrt{|M|\sigma^2}}\left\|\mathrm{X}^{(M)}\right\|.$$

From the expected value of the chi distribution:

$$\mathbb{E}\left[\frac{1}{\sqrt{|M|\sigma^2}}\left\|\mathrm{X}^{(M)}\right\|\right] = \sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)},$$

where $\Gamma(\cdot)$ is the gamma function. Thus, the expected value of L2 norm $\left\|\mathrm{X}^{(M)}\right\|$ of $\mathrm{X}^{(M)}$ is:

$$\mathbb{E}\left[\left\|\mathrm{X}^{(M)}\right\|\right] = \sqrt{2|M|\sigma^2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \propto \sqrt{|M|}.$$

This completes the proof. □

From Theorem 1, it is evident that embeddings that are simply sampled independently from a normal distribution do not exhibit the linear relationship in Eq. (9), implying that Observation 1 is indeed non-trivial. Next, we examine the variance of the L2 norm of the aggregated embedding.

Theorem 2. *Let M be a set of d-dimensional embeddings $\{\mathbf{x}_i\}_{i=1}^{|M|}$, where each dimension of $\mathbf{x}_i \in M$ is independently drawn from a normal distribution $N(0, \sigma^2)$. If the embeddings are uncorrelated (i.e., independent) sample-wise, the variance of the L2 norm of the unscaled aggregated embeddings in M is:*

$$\mathbb{V}ar\left[\left\|\sum_{\mathbf{x}_i \in M} \mathbf{x}_i\right\|\right] = |M|\sigma^2\left(d - \left(\sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\right)^2\right)$$

Proof. Let $\mathrm{X}_j^{(M)} = \sum_{\mathbf{x}_i \in M} \mathbf{x}_{ij}$. Then, from the proof of Theorem 1, $\mathrm{X}_1^{(M)}/\sqrt{|M|\sigma^2}, \cdots, \mathrm{X}_d^{(M)}/\sqrt{|M|\sigma^2}$ are d independent random variables from $N(0, 1)$. This implies that the following statistic is distributed according to the chi distribution with d degrees of freedom. From the variance of the chi distribution:

$$\mathbb{V}\mathrm{ar}\left[\frac{1}{\sqrt{|M|\sigma^2}}\left\|\mathrm{X}^{(M)}\right\|\right] = d - \mathbb{E}\left[\frac{1}{\sqrt{|M|\sigma^2}}\left\|\mathrm{X}^{(M)}\right\|\right]^2 = d - \left(\sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\right)^2.$$

From the properties of variance, we have:

$$\mathbb{V}\mathrm{ar}\left[\left\|\mathrm{X}^{(M)}\right\|\right] = |M|\sigma^2\left(d - \left(\sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\right)^2\right).$$

This completes the proof. □

Since the expected value of the chi distribution is close to $\sqrt{d - \frac{1}{2}}$ for large d, the following approximation holds in the high-dimensional space:

$$\mathbb{V}\mathrm{ar}\left[\left\|\sum_{\mathbf{x}_i \in M} \mathbf{x}_i\right\|\right] \approx \frac{|M|\sigma^2}{2}.$$

This indicates that the variance linearly increases with the number of neighbors, i.e., $|M|$.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1.2 Norm of Aggregated Embedding of Correlated Neighbors (Generalization)

We have observed that when embeddings are independently sampled from a normal distribution, the expected norm of the unscaled aggregated embeddings is sublinear w.r.t. the number of neighbors. Here, we demonstrate how the correlation between neighbors affects the linearity between the L2 norm of the aggregated embedding and the number of neighbors. Specifically, we introduce the correlation coefficient ρ between neighbors when computing the L2 norm of the aggregated pairs.

Theorem 3. *Let M be a set of d-dimensional embeddings $\{\mathbf{x}_i\}_{i=1}^{|M|}$ consisting of random variables. Assume that (1) at each j^{th} dimension, $\mathbf{x}_{1j}, \cdots, \mathbf{x}_{|M|j}$ are drawn from a multivariate normal distribution $N(0, \Sigma)$ with $\Sigma_{ii} = \sigma^2\ \forall i$, $\Sigma_{ik} = \rho\sigma^2\ \forall i \neq k$ for some $\rho \geq 0$ (i.e., each $\mathbf{x}_{ij} \sim N(0, \sigma^2)$ and each pair $\mathbf{x}_{ij}$ and $\mathbf{x}_{kj}$ have correlation coefficient ρ), and (2) the dimensions are mutually independent and thus i.i.d. Then, the expected L2 norm of the unscaled aggregated embeddings in M follows the proportionality:*

$$\mathbb{E}\left[\left\|\sum\nolimits_{\mathbf{x}_i \in M} \mathbf{x}_i\right\|\right] \propto \sqrt{|M|(1-\rho) + |M|^2\rho}.$$

Note that if $\rho = 1$, then $\mathbb{E}\left[\left\|\sum_{\mathbf{x}_i \in M} \mathbf{x}_i\right\|\right] \propto |M|$.

Proof. Let $\mathbf{X}_j^{(M)} = \sum_{\mathbf{x}_i \in M} \mathbf{x}_{ij}$. Based on the property of the sum of normal distributions, the expected value of $\mathbf{X}_j^{(M)}$ is 0. The variance of $\mathbf{X}_j^{(M)}$ is:

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\mathbf{X}_j^{(M)}\right] &= \sum_{\mathbf{x}_i \in M} \mathbb{V}\mathrm{ar}\left[\mathbf{x}_{ij}\right] + \sum_{\substack{\mathbf{x}_i, \mathbf{x}_k \in M \\ i \neq k}} \mathbb{C}\mathrm{ov}\left(\mathbf{x}_{ij}, \mathbf{x}_{kj}\right) \\ &= |M|\sigma^2 + |M|(|M|-1)\rho\sigma^2 \\ &= \sigma^2\left(|M|(1-\rho) + |M|^2\rho\right).\end{aligned}$$

Thus, $\mathbf{X}_j^{(M)} / \sqrt{\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}$ for $j = 1 \cdots d$ are d independent random variables from $N(0,1)$. This implies that the following statistic is distributed according to the chi distribution with d degrees of freedom:

$$\sqrt{\sum_{j=1}^{d}\left(\frac{\mathbf{X}_j^{(M)}}{\sqrt{\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}}\right)^2} = \frac{\left\|\mathbf{X}^{(M)}\right\|}{\sqrt{\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}}.$$

From the expected value of the chi distribution:

$$\mathbb{E}\left[\frac{\left\|\mathbf{X}^{(M)}\right\|}{\sqrt{\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}}\right] = \sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)},$$

where $\Gamma(\cdot)$ is the gamma function. Thus, the expected value of L2 norm $\left\|\mathbf{X}^{(M)}\right\|$ of $\mathbf{X}^{(M)}$ is:

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{X}^{(M)}\right\|\right] &= \sqrt{2\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \\ &\propto \sqrt{|M|(1-\rho) + |M|^2\rho}.\end{aligned}$$

This completes the proof. □

Theorem 3 suggests that if the embeddings are independent of each other (i.e., $\rho = 0$), as assumed in Theorem 1, the norm of the aggregated embedding is sublinear w.r.t. the number of neighbors, i.e., $\mathbb{E}\left[\left\|\mathbf{X}^{(M)}\right\|\right] \propto \sqrt{|M|}$. Conversely, if the embeddings exhibit complete linear relationships with each other (i.e., $\rho = 1$), the norm of the aggregated embedding is linear with the number of neighbors, i.e., $\mathbb{E}\left[\left\|\mathbf{X}^{(M)}\right\|\right] \propto |M|$. This indicates how the degree of correlation between embeddings influences the linearity of their aggregated norm. We conjecture that Observation 1 is attributed to the strong correlations between user/item embeddings. Next, we examine the variance of the L2 norm of the aggregated embedding.

Theorem 4. *Let M be a set of d-dimensional embeddings $\{\mathbf{x}_i\}_{i=1}^{|M|}$ consisting of random variables. Assume that (1) at each j^{th} dimension, $\mathbf{x}_{1j}, \cdots, \mathbf{x}_{|M|j}$ are drawn from a multivariate normal distribution $N(0, \Sigma)$ with $\Sigma_{ii} = \sigma^2\ \forall i$, $\Sigma_{ik} = \rho\sigma^2\ \forall i \neq k$ for some $\rho \geq 0$ (i.e., each $\mathbf{x}_{ij} \sim N(0, \sigma^2)$ and each pair $\mathbf{x}_{ij}$ and $\mathbf{x}_{kj}$ have correlation coefficient ρ), and (2) the dimensions are mutually independent and thus i.i.d. Then, the variance of the L2 norm of the unscaled aggregated embeddings in M is:*

$$\mathbb{V}ar\left[\left\|\sum\nolimits_{\mathbf{x}_i \in M} \mathbf{x}_i\right\|\right] = \sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)\left(d - \left(\sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\right)^2\right)$$

Proof. Let $\mathbf{X}_j^{(M)} = \sum_{\mathbf{x}_i \in M} \mathbf{x}_{ij}$. Then, from the proof of Theorem 3, $\mathbf{X}_j^{(M)} / \sqrt{\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}$ for $j = 1, \cdots, d$ are d independent random variables from $N(0,1)$. This implies that the following statistic is distributed according to the chi distribution with d degrees of freedom. From the variance of the chi distribution:

$$\begin{aligned}&\mathbb{V}\mathrm{ar}\left[\frac{1}{\sqrt{\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}}\left\|\mathbf{X}^{(M)}\right\|\right] = \\ &d - \mathbb{E}\left[\frac{1}{\sqrt{\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}}\left\|\mathbf{X}^{(M)}\right\|\right]^2 = d - \left(\sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\right)^2.\end{aligned}$$

From the properties of variance, we have:

$$\mathbb{V}\mathrm{ar}\left[\left\|\mathbf{X}^{(M)}\right\|\right] = \sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)\left(d - \left(\sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\right)^2\right).$$

This completes the proof. □

Since the expected value of the chi distribution is close to $\sqrt{d - \frac{1}{2}}$ for large d, the following approximation holds in the high-dimensional space:

$$\mathbb{V}\mathrm{ar}\left[\left\|\sum\nolimits_{\mathbf{x}_i \in M} \mathbf{x}_i\right\|\right] \approx \frac{\sigma^2\left(|M|(1-\rho) + |M|^2\rho\right)}{2}.$$

Table 1: Comparison of LightGCN, its enhanced methods (i.e., ALGCN [7], *r*-AdjNorm [9], and SSM [6]), and LightGCN++ reveals that LightGCN++ incorporates key design properties that existing methods have overlooked.

		LightGCN [2]	ALGCN [7]	r-AdjNorm [9]	SSM [6]	**LightGCN++ (Ours)**
Effective weights	Definition	$\lVert\mathbf{e}_u^{(k)}\rVert/\sqrt{\lvert\mathcal{N}_u\rvert}$	$\lVert\mathbf{e}_u^{(k)}\rVert/\lvert\mathcal{N}_u\rvert$	$\lVert\mathbf{e}_u^{(k)}\rVert/\lvert\mathcal{N}_u\rvert^{1-\alpha}$	$\lVert\mathbf{e}_u^{(k)}\rVert/\lvert\mathcal{N}_u\rvert^{\beta}$	$1/\lvert\mathcal{N}_u\rvert^{\beta}$
	P1	✗	✗	✗	✓	✓
	P2	✗	✗	✗	✗	✓
Layer-wise Aggregation	Definition	$\frac{1}{K+1}\sum_{k=0}^{K}\mathbf{e}_i^{(k)}$	$\frac{1}{K+1}\sum_{k=0}^{K}\mathbf{e}_i^{(k)}$	$\frac{1}{K+1}\sum_{k=0}^{K}\mathbf{e}_i^{(k)}$	$\frac{1}{K+1}\sum_{k=0}^{K}\mathbf{e}_i^{(k)}$	$\gamma\mathbf{e}_i^{(0)} + (1-\gamma)\sum_{k=1}^{K}\mathbf{e}_i^{(k)}$
	P3	✗	✗	✗	✗	✓

1.3 Space Complexity of LightGCN++

We analyze the space complexity of LightGCN++. At each k^{th} layer, it has $|V|$ number of d-dimensional embeddings, which requires $O(|V|d)$ space. Given that there are $K+1$ layers, including the initial layer, the total space complexity is $O(|V|Kd)$, which is equivalent to that of LightGCN.

2 ADDITIONAL RELATED WORKS

The propagation rule of LightGCN has been enhanced in various directions. In this section, We begin by reviewing the properties of LightGCN. Then, we examine whether existing methods for enhancing LightGCN address these limitations.

Properties of LightGCN. We begin by reviewing properties of LightGCN that we have identified: **inflexibility** and **inconsistency** which are summarized as follows:

- **P1. Near-uniform neighbor weights.** When $k \geq 1$, effective weights are near-uniform across neighbors (Property 1).
- **P2. Uncontrollable neighbor weights.** When $k = 0$, effective weights are not adjustable (Property 2).
- **P3. Inflexible layer-wise norm scaling.** While norm scaling of embeddings at $k = 0$ and $k \geq 1$ exhibits disparities (Properties 3 & 4), they are aggregated with a fixed ratio (Property 5).

Existing LightGCN-enhanced methods. We assess whether existing methods for enhancing LightGCN address the limitations discussed above. Recently, the focus of most research has been on adjusting the embedding norms (i.e., magnitudes) [6, 7, 9]. However, these studies overlook key limitations of LightGCN, which we have empirically discovered in this paper as summarized in Table 1.

Xu et al. [7] generalized the left norm of LightGCN's propagation rule to be a function of the user/item popularity, as follows:

$$\mathbf{e}_i^{(k+1)} = \sum_{u\in\mathcal{N}_i} f(|\mathcal{N}_i|)\frac{1}{|\mathcal{N}_u|}\mathbf{e}_u^{(k)}$$

where $f(|\mathcal{N}_i|)$ is a function of $|\mathcal{N}_i|$, e.g., $f(|\mathcal{N}_i|) = |\mathcal{N}_i|^{0.5}$ or $f(|\mathcal{N}_i|) = \log_2(|\mathcal{N}_i|+1)$. While the function $f(\cdot)$ allows for flexible adjustment of the embedding norms, its effective weights remain inflexible and cannot be adjusted.

Zhao et al. [9] introduced a hyperparameter α as follows:

$$\mathbf{e}_i^{(k+1)} = \sum_{u\in\mathcal{N}_i} \frac{1}{|\mathcal{N}_i|^{\alpha}|\mathcal{N}_u|^{1-\alpha}}\mathbf{e}_u^{(k)},$$

which is essentially equivalent to Eq. (6). This approach provides flexibility in scaling embedding norms by introducing α. However, the effective weights are constant (i.e., uniform) when $k \geq 1$, as discussed in Section 4.1. [1]

Wu et al. [6] introduced another hyperparameter β into the right term of the propagation rule as follows:

$$\mathbf{e}_i^{(k+1)} = \sum_{u\in\mathcal{N}_i} \frac{1}{|\mathcal{N}_i|^{\alpha}|\mathcal{N}_u|^{\beta}}\mathbf{e}_u^{(k)}.$$

While β can be used to flexibly adjust the effective weights of neighbors, its role was overlooked in [6], and it was introduced to control the norm of embeddings along with α. In addition, its effective weight $\lVert\mathbf{e}_u^{(k)}\rVert/|\mathcal{N}_u|^{\beta}$ is uncontrollable when $k = 0$, as discussed in Section 3.4. [2]

Commonly, prior works have primarily focused on adjusting the norms of embeddings. In contrast, LightGCN++ is designed to address two unexpected and unexplored properties of LightGCN: (1) **inflexibility**: the near-uniform weights of neighbors and (2) **inconsistency**: disparities in norm scaling between layers at $k = 0$ and $k \geq 1$. As summarized in Table 1, LightGCN++ extends beyond adjusting embedding norms and addresses the two crucial limitations of LightGCN.

Furthermore, LightGCN has been enhanced through long-range propagation [3, 4]. Specifically, Mao et al. [4] approximated the limit of infinite-layer graph convolution of LightGCN. Kapralova et al. [3] adapted personalized PageRank into LightGCN's propagation rule. These studies have not investigated the unique properties of LightGCN, which we discussed in Section 3.

Attention-based neighbor weighting. It is noted that attention-based methods can address the inflexibility in neighbor weighting of LightGCN. For example, the propagation rule of NGCF [5] is defined as follows:

$$\mathbf{e}_i^{(k+1)} = \sigma\left(\sum_{u\in\mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_i|}\sqrt{|\mathcal{N}_u|}}\,\mathbf{W}\left(\mathbf{e}_u^{(k)} \odot \mathbf{e}_i^{(k)}\right)\right)$$

where we remove self-connections for simplicity. Then, the effective weight of NGCF can be derived as follows:

$$\frac{\mathbf{W}\left(\mathbf{e}_u^{(k)} \odot \mathbf{e}_i^{(k)}\right)}{\sqrt{|\mathcal{N}_u|}}$$

However, LightGCN++'s simple effective weight of $1/|\mathcal{N}_u|^{\beta}$ offers the following advantages over attention-based neighbor weighting:

[1] Recall our empirical observation that, in $\mathbf{e}_i^{(k+1)} = \sum_{u\in\mathcal{N}_i}\frac{1}{|\mathcal{N}_i|^{\alpha}|\mathcal{N}_u|^{\beta}}\mathbf{e}_u^{(k)}$, the effective weight $\lVert\mathbf{e}_u^{(k)}\rVert/|\mathcal{N}_u|^{\beta}$ becomes a constant if $\beta = 1-\alpha$ (Section 4.1).

[2] The linearity $\lVert\mathbf{e}_u^{(k)}\rVert \mathrel{\underset{\sim}{\propto}} |\mathcal{N}_u|^{1-\alpha}$ is observed only when $k \geq 1$ (Property 2).

Table 2: LightGCN++ and its variant, LightGCN++ (MP), which incorporates mean-pooling layer-wise aggregation, outperform LightGCN-enhanced baselines in terms of NDGC@20.

Dataset	LastFM	MovieLens	Gowalla	Yelp	Amazon
LightGCN	0.2427	0.3010	0.1426	0.0449	0.0274
ALGCN [7]	0.2318	0.3050	0.1376	0.0435	0.0259
r-AdjNorm [9]	0.2434	0.3054	0.1426	0.0450	0.0281
SSM [6]	0.2523	0.3124	0.1426	0.0491	0.0278
UltraGCN [4]	0.2359	0.3172	0.1408	0.0507	0.0278
LightGCN++ (MP)	0.2614	0.3217	0.1436	0.0507	0.0282
LightGCN++	**0.2624**	**0.3275**	**0.1469**	**0.0529**	**0.0294**

- **Enhanced performance.** The simplified neighbor weighting has been empirically shown to yield better performance. LightGCN++ consistently and significantly outperforms NGCF across all datasets. Similarly, LightGCN, despite its inflexibility in neighbor weights, also outperforms NGCF. Due to its empirical superiority, recent state-of-the-art methods (e.g., SimGCL [8] and LightGCL [1]) have adopted non-attention-based strategies for neighbor aggregation.
- **Better interpretability.** The effective weight $1/|\mathcal{N}_u|^{\beta}$ in LightGCN++ directly takes neighbor u's popularity $|\mathcal{N}_u|$ into account, making it more interpretable than the black-box nature of attention-based weights.
- **Explicit controllability.** While attention-based weights are determined by the model (e.g., $\mathbf{W}$), LightGCN++'s effective weight is explicitly adjustable by users.

3 ADDITIONAL EXPERIMENTAL RESULTS

In this section, we provide additional experimental results.

3.1 Additional Baselines

We conduct a comparison of LightGCN++ with existing methods that enhance LightGCN, particularly those discussed in Section 2, i.e., ALGCN [7], r-AdjNorm [9], SSM [6], and UltraGCN [4]. As shown in Table 2, LightGCN++ consistently outperforms these baselines across all datasets. Additionally, its variant, LightGCN++ (MP), which employs a mean pooling for aggregating layer-wise embeddings, demonstrates superior performance compared to the baselines. This demonstrates the effectiveness of the propagation rule of LightGCN++ (Eq. (7)).

3.2 Learnable Embedding Pooling

To evaluate the effectiveness of the adaptive layer-wise embedding aggregation approach used in LightGCN++, we compare it with two intuitive pooling methods. Specifically, mean pooling assigns equal importance across all layers, whereas learnable pooling learns individual weights for each layer. As shown in Table 3, adaptive pooling, upon fine-tuning γ, achieves the best results in terms of NDCG@20. This indicates that addressing the disparities in norm scaling between agg-free and agg-based embeddings by properly balancing them is effective for generating the final embeddings.

3.3 Hyperparameter Tuning Strategies

Here, we share some simple strategies for tuning α, β, and γ in LightGCN++ for its practical usability.

Table 3: LightGCN++'s adaptive pooling for layer-wise embedding aggregation, tuning γ, is more effective than both mean pooling and learnable pooling approaches.

Dataset	LastFM	MovieLens	Gowalla	Yelp	Amazon
Mean Pooling	0.2614	0.3217	0.1436	0.0507	0.0282
Learnable Pooling	0.2590	0.3227	0.1389	0.0524	0.0283
Adaptive Pooling	**0.2624**	**0.3275**	**0.1469**	**0.0529**	**0.0294**

- **Tuning α.** We advise users to adjust α based on the long-tailed characteristics of the item popularity (i.e., degree) distribution. Specifically, increasing α (i.e., $\alpha \rightarrow 1$) leads to a fairer recommendation that is equally likely to recommend both popular and unpopular items. In contrast, decreasing α (i.e., $\alpha \rightarrow 0$) biases the system towards recommending more popular items.
- **Tuning β.** According to our experiments, setting $0 \geq \beta \geq 1$ generally leads to improvements. This suggests that reducing the influence of high-degree neighbors enhances accuracy.
- **Tuning γ.** We recommend that users begin by tuning γ from 0, which excludes the embeddings at the initial layer (i.e., $\mathbf{e}_i^{(0)}$) from the layer-wise aggregation. Then, gradually increasing γ may enhance performance, depending on the dataset.

4 FUTURE DIRECTIONS

Lastly, we share potential future directions of this work as follows:

- **Enhancement of layer-wise pooling.** While we have distinguished between agg-free (k=0) and agg-based (k=1) embeddings for simplicity, more enhanced pooling strategies can potentially improve performance.
- **Enhancement from deeper layers.** Inspired by UltraGCN [4], LightGCN++ could benefit from incorporating propagations from deeper layers.
- **Further analysis on LightGCN.** While LightGCN has empirically shown the effectiveness of removing feature transformation and non-linear activation, it lacks theoretical justification for these improvements. We aim to delve deeper into this topic.
- **Extensions to general GNNs.** The concept of effective weights presents an opportunity to investigate the propagation mechanisms of general GNNs in areas beyond recommender systems.

REFERENCES

[1] Xuheng Cai, Chao Huang, Lianghao Xia, and Xubin Ren. 2022. LightGCL: Simple Yet Effective Graph Contrastive Learning for Recommendation. In *ICLR*.

[2] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *SIGIR*.

[3] Milena Kapralova, Luca Pantea, and Andrei Blahovici. 2023. LightGCN: Evaluated and Enhanced. *arXiv preprint arXiv:2312.16183* (2023).

[4] Kelong Mao, Jieming Zhu, Xi Xiao, Biao Lu, Zhaowei Wang, and Xiuqiang He. 2021. UltraGCN: ultra simplification of graph convolutional networks for recommendation. In *CIKM*.

[5] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural graph collaborative filtering. In *SIGIR*.

[6] Jiancan Wu, Xiang Wang, Xingyu Gao, Jiawei Chen, Hongcheng Fu, Tianyu Qiu, and Xiangnan He. 2022. On the effectiveness of sampled softmax loss for item recommendation. *TOIS* (2022).

[7] Ronghai Xu, Haijun Zhao, Zhi-Yuan Li, and Chang-Dong Wang. 2023. ALGCN: Accelerated Light Graph Convolution Network for Recommendation. In *DASFAA*.

[8] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *SIGIR*.

[9] Minghao Zhao, Le Wu, Yile Liang, Lei Chen, Jian Zhang, Qilin Deng, Kai Wang, Xudong Shen, Tangjie Lv, and Runze Wu. 2022. Investigating accuracy-novelty performance for graph-based collaborative filtering. In *SIGIR*.