A Hilbert-Schmidt Independence Criterion

The Hilbert-Schmidt Independence Criterion (HSIC) [22] is a statistical measure of dependency. Defined as the squared Hilbert-Schmidt norm, it quantifies the dependence between two variables by evaluating the cross-covariance operator within the Reproducing Kernel Hilbert Spaces. More specifically, for two random variables $\mathbf{S}$ and $\mathbf{X}$, the HSIC is rigorously defined as:

$$\begin{aligned}\texttt{HSIC}(\mathbf{S},\mathbf{X}) &= \|C_{\mathbf{SX}}\|^2_{\text{HS}} \\ &= \mathbb{E}_{\mathbf{S},\mathbf{S}',\mathbf{X},\mathbf{X}'}\left[k_1(\mathbf{S},\mathbf{S}')k_2(\mathbf{X},\mathbf{X}')\right] \\ &\quad + \mathbb{E}_{\mathbf{S},\mathbf{S}'}\left[k_1(\mathbf{S},\mathbf{S}')\right]\mathbb{E}_{\mathbf{X},\mathbf{X}'}\left[k_2(\mathbf{X},\mathbf{X}')\right] \\ &\quad - 2\mathbb{E}_{\mathbf{S},\mathbf{X}}\left[\mathbb{E}_{\mathbf{S}'}\left[k_1(\mathbf{S},\mathbf{S}')\right]\mathbb{E}_{\mathbf{X}'}\left[k_2(\mathbf{X},\mathbf{X}')\right]\right],\end{aligned}$$

where $\mathbb{E}$ denotes the expectation, and k_1, k_2 are two kernel functions for variables $\mathbf{S},\mathbf{X}$. $\mathbf{S}'$ and $\mathbf{X}'$ are two independent copies of $\mathbf{S}$ and $\mathbf{X}$.

To develop a formal independence test grounded in the HSIC, we are required to approximate the HSIC value with a finite collection of observations. This approximation, referred to as the Empirical HSIC, is defined by the following formula:

$$\texttt{HSIC}(\mathbf{S},\mathbf{X}) = \frac{1}{(|\mathcal{U}|-1)^2}\texttt{trace}(\mathbf{K}^{(s)}\mathbf{E}\mathbf{K}^{(x)}\mathbf{E}), \tag{19}$$

where $|\mathcal{U}|$ is the number of independent observations. $\mathbf{K}^{(s)}$ and $\mathbf{K}^{(x)}$ are the Gram matrices with elements as $\mathbf{K}^{(s)}_{i,j} = k_1(\mathbf{S}_i,\mathbf{S}_j)$ and $\mathbf{K}^{(x)}_{i,j} = k_2(\mathbf{X}_i,\mathbf{X}_j)$, respectively. $\mathbf{E} = \mathbf{I} - \frac{1}{|\mathcal{U}|}\mathbf{1}$ is to center the Gram matrix to have zero mean.

To enhance computational efficiency, instead of multiplying the representations $\mathbf{S}$ and $\mathbf{X}$ by $\mathbf{E}$, we can normalize them and adopt the linear kernel. Consequently, we derive the following simplified form [73]:

$$\texttt{HSIC}(\mathbf{S},\mathbf{X}) = \texttt{trace}(\mathbf{S}^\top\mathbf{X}\mathbf{X}^\top\mathbf{S}). \tag{20}$$

B Theoretical Proofs

Proof of Lemma 4.1. In the community recommendation task, a user can join multiple communities. Therefore, we adjust $\delta(u_i,u_j)$ to be the number of communities that both users have joined. Then, we obtain $\delta(u_i,u_j) = \sum_{k=1}^{|C|}\mathbf{\Phi}_{i,k}\cdot\mathbf{\Phi}_{j,k}$.

Next, denote the degree vector as $\mathbf{d}$, where $\mathbf{d}_i = \mathbf{d}_i$ for each index i, and $\sqrt{\mathbf{d}}$ is the element-wise square root of $\mathbf{d}$. According to Eq.(1), we can derive that

$$\begin{aligned}Q(\tilde{\mathbf{A}}) &= \sum_{k=1}^{|C|}\sum_{(u_i,u_j)\in\mathcal{E}}\left(\tilde{\mathbf{A}}_{i,j} - \frac{\sqrt{\mathbf{d}}_i\cdot\sqrt{\mathbf{d}}_j}{|\mathcal{E}|}\right)\cdot\mathbf{\Phi}_{i,k}\cdot\mathbf{\Phi}_{j,k} \\ &= \sum_{k=1}^{|C|}\mathbf{\Phi}^\top_{\cdot,k}\left(\tilde{\mathbf{A}} - \frac{\sqrt{\mathbf{d}}\cdot\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right)\mathbf{\Phi}_{\cdot,k} \\ &= \texttt{trace}\left(\mathbf{\Phi}^\top\left(\tilde{\mathbf{A}} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right)\mathbf{\Phi}\right).\end{aligned}$$

□

Proof of Theorem 4.2. First, we need the following lemmata.

Lemma B.1 ([25]). *Let $\mathbf{M}$ be a matrix whose dominant eigenvalue λ satisfies $|\lambda| < 1$. Then, $\mathbf{I}-\mathbf{M}$ is invertible, and its inverse $(\mathbf{I}-\mathbf{M})^{-1}$ can be expanded as a Neumann series: $(\mathbf{I}-\mathbf{M})^{-1} = \sum_{t=0}^{\infty}\mathbf{M}^t$.*

Lemma B.2. *$\forall 1 \le i \le |\mathcal{U}|$, $-2 \le \lambda_i(\tilde{\mathbf{A}} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top/|\mathcal{E}|) \le 2$ where $\lambda_i(\cdot)$ returns the i^{th} eigenvalue.*

Denote by O_{SMM} Eq. (5). We take the derivative of O_{SMM} w.r.t. $\mathbf{G}$ and set it to zero, which yields:

$$\begin{aligned}&\frac{\partial O_{\text{SMM}}}{\partial \mathbf{G}} = \alpha\cdot\left(\frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|} - \tilde{\mathbf{A}}\right)\mathbf{G} + (1-\alpha)\cdot(\mathbf{G}-\mathbf{U}^\circ) = 0 \\ &\left((1-\alpha)\mathbf{I} - \alpha\left(\tilde{\mathbf{A}} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right)\right)\mathbf{G} = (1-\alpha)\mathbf{U}^\circ \\ &\left(\mathbf{I} - \frac{\alpha}{1-\alpha}\left(\tilde{\mathbf{A}} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right)\right)\mathbf{G} = \mathbf{U}^\circ \\ &\mathbf{G} = \left(\mathbf{I} - \frac{\alpha}{1-\alpha}\left(\tilde{\mathbf{A}} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right)\right)^{-1}\mathbf{U}^\circ.\end{aligned} \tag{21}$$

When $\alpha < \frac{1}{3}$, we have $\frac{\alpha}{1-\alpha} < \frac{1}{2}$. Thus, the eigenvalues of $\frac{\alpha}{1-\alpha}(\tilde{\mathbf{A}} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|})$ are in the range of $(-1,1)$. Applying Lemma B.1 on Eq. (21) completes the proof. □

Proof of Lemma B.2. First, we need to prove the eigenvalues of $\tilde{\mathbf{A}}$ and $\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top/|\mathcal{E}|$ are in the range of $[-1,1]$.

For $\tilde{\mathbf{A}}$, it is known that its dominant eigenvalue is bounded by 1 [13]. Thus, the range of its eigenvalues is within the interval $[-1,1]$.

For $\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top/|\mathcal{E}|$, we first prove that its Rayleigh quotient satisfies

$$\frac{\mathbf{x}^\top\cdot\frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\cdot\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} \le 1.$$

First, we can derive that

$$\begin{aligned}\mathbf{x}^\top\left(\mathbf{I} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right)\mathbf{x} &= \sum_{u_i\in\mathcal{U}}\mathbf{x}_i^2 - \sum_{u_i,u_j\in\mathcal{U}}\mathbf{x}_i\cdot\frac{\sqrt{\mathbf{d}_i\cdot\mathbf{d}_j}}{|\mathcal{E}|}\cdot\mathbf{x}_j \\ &= \sum_{u_i\in\mathcal{U}}\mathbf{x}_i^2\cdot\frac{\mathbf{d}_i}{|\mathcal{E}|} + \sum_{u_i\in\mathcal{U}}\mathbf{x}_i^2\cdot\left(1-\frac{\mathbf{d}_i}{|\mathcal{E}|}\right) \\ &\quad - \sum_{u_i,u_j\in\mathcal{U}}\mathbf{x}_i\cdot\sqrt{\frac{\mathbf{d}_i}{|\mathcal{E}|}}\cdot\mathbf{x}_j\cdot\sqrt{\frac{\mathbf{d}_j}{|\mathcal{E}|}} \\ &= \frac{1}{2}\sum_{v_i\in C}\left(\mathbf{x}_i\cdot\sqrt{\frac{\mathbf{d}_i}{|\mathcal{E}|}} - \mathbf{x}_j\cdot\sqrt{\frac{\mathbf{d}_j}{|\mathcal{E}|}}\right)^2 \\ &\quad + \sum_{u_i\in\mathcal{U}}\mathbf{x}_i^2\cdot\left(1-\frac{\mathbf{d}_i}{|\mathcal{E}|}\right).\end{aligned}$$

Since $1 - \frac{\mathbf{d}_i}{|\mathcal{E}|} \ge 0$, we have $\mathbf{x}^\top\left(\mathbf{I} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right)\mathbf{x} \ge 0$, meaning that $\frac{\mathbf{x}^\top\cdot\frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\cdot\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} \le 1$. This implies that the dominant eigenvalue of $\frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}$ is less than or equal to 1. Its eigenvalues are in the range of $[-1,1]$.

Then, by Weyl inequality [18, 64] that for any Hermitian matrices $\mathbf{A}$ and $\mathbf{Y}$, their dominant eigenvalues hold $\lambda_{i+j-1}(\mathbf{A}+\mathbf{Y}) \le \lambda_i(\mathbf{A}) +$

$\lambda_j(\mathbf{Y}) \leq \lambda_{i+j-n}(\mathbf{A}+\mathbf{Y})$ with $\lambda_1 \geq \ldots \geq \lambda_n$. Consequently, we can obtain

$$\lambda_1\left(\tilde{\mathbf{A}} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right) \leq \lambda_1(\tilde{\mathbf{A}}) + \lambda_1\left(-\frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right) = \lambda_1(\tilde{\mathbf{A}}) - \lambda_n\left(\frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right) \leq 1-(-1) = 2,$$

and

$$\lambda_n\left(\tilde{\mathbf{A}} - \frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right) \geq \lambda_n(\tilde{\mathbf{A}}) + \lambda_n\left(-\frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right) = \lambda_n(\tilde{\mathbf{A}}) - \lambda_1\left(\frac{\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top}{|\mathcal{E}|}\right) \geq (-1)-1 = -2.$$

Based thereon, we derive $-1 \leq \lambda_i(\tilde{\mathbf{A}} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^\top/|\mathcal{E}|) \leq 1$. □

Proof of Theorem 4.3. For Common Neighbors (CN),

$$p(u_i,u_j) = |\mathcal{N}_\mathcal{G}(u_i) \cap \mathcal{N}_\mathcal{G}(u_j)| = \sum_k \mathbf{A}_{ik}\mathbf{A}_{jk} = \mathbf{A}_i\mathbf{A}_j^\top = f(\mathbf{A})_i \cdot f(\mathbf{A})_j^\top.$$

For Adamic-Adar Index (AAI),

$$p(u_i,u_j) = \sum_{u\in\mathcal{N}(u_i)\cap\mathcal{N}(u_j)} \frac{1}{\log \mathbf{d}_u} = \sum_u \frac{\mathbf{A}_{iu}\mathbf{A}_{ju}}{\log \mathbf{d}_u} = \sum_u \frac{\mathbf{A}_{iu}}{\sqrt{\log \mathbf{d}_u}} \cdot \frac{\mathbf{A}_{ju}}{\sqrt{\log \mathbf{d}_u}} = \left(\mathbf{A}\mathbf{D}_{\log}^{-1/2}\right)_i \left(\mathbf{A}\mathbf{D}_{\log}^{-1/2}\right)_j^\top = f(\mathbf{A})_i \cdot f(\mathbf{A})_j^\top.$$

For Resource Allocation Index (RAI),

$$p(u_i,u_j) = \sum_{u\in\mathcal{N}_\mathcal{G}(u_i)\cap\mathcal{N}_\mathcal{G}(u_j)} \frac{1}{\mathbf{d}_u} = \sum_u \frac{\mathbf{A}_{iu}\mathbf{A}_{ju}}{\mathbf{d}_u} = \sum_u \frac{\mathbf{A}_{iu}}{\sqrt{\log \mathbf{d}_u}} \cdot \frac{\mathbf{A}_{ju}}{\sqrt{\log \mathbf{d}_u}} = \left(\mathbf{A}\mathbf{D}^{-1/2}\right)_i \left(\mathbf{A}\mathbf{D}^{-1/2}\right)_j^\top = f(\mathbf{A})_i \cdot f(\mathbf{A})_j^\top.$$

For Salton Index (SI),

$$p(u_i,u_j) = \frac{|\mathcal{N}_\mathcal{G}(u_i) \cap \mathcal{N}_\mathcal{G}(u_j)|}{\sqrt{\mathbf{d}_i\mathbf{d}_j}} = \sum_u \frac{\mathbf{A}_{iu}\mathbf{A}_{ju}}{\sqrt{\mathbf{d}_i\mathbf{d}_j}} = \sum_u \frac{\mathbf{A}_{iu}}{\sqrt{\mathbf{d}_j}} \cdot \frac{\mathbf{A}_{ju}}{\sqrt{\mathbf{d}_j}} = \left(\mathbf{D}^{-1/2}\mathbf{A}\right)_i \left(\mathbf{D}^{-1/2}\mathbf{A}\right)_j^\top = f(\mathbf{A})_i \cdot f(\mathbf{A})_j^\top.$$

For Leicht-Holme-Newman Index (LHNI),

$$p(u_i,u_j) = \frac{|\mathcal{N}_\mathcal{G}(u_i) \cap \mathcal{N}_\mathcal{G}(u_j)|}{\mathbf{d}_i\mathbf{d}_j} = \sum_u \frac{\mathbf{A}_{iu}\mathbf{A}_{ju}}{\mathbf{d}_i\mathbf{d}_j} = \sum_u \frac{\mathbf{A}_{iu}\mathbf{A}_{ju}}{\mathbf{d}_i\mathbf{d}_j} = \left(\mathbf{D}^{-1}\mathbf{A}\right)\left(\mathbf{D}^{-1}\mathbf{A}\right)^\top = f(\mathbf{A})_i \cdot f(\mathbf{A})_j^\top.$$

The theorem is then proved. □

Proof of Lemma 4.4. By the definition of $s(u_i,u_j)$ in Eq. (11),

$$\begin{aligned}
&\sum_{k=1}^{|C|}\left(\hat{\mathbf{Y}}_{i,k} - \sqrt{\frac{\boldsymbol{\delta}_i}{|\mathcal{Y}|}}\cdot\sqrt{\frac{\boldsymbol{\sigma}_k}{|\mathcal{Y}|}}\right)\cdot\left(\hat{\mathbf{Y}}_{j,k} - \sqrt{\frac{\boldsymbol{\delta}_j}{|\mathcal{Y}|}}\cdot\sqrt{\frac{\boldsymbol{\sigma}_k}{|\mathcal{Y}|}}\right)\\
&=\sum_{k=1}^{|C|}\hat{\mathbf{Y}}_{i,k}\cdot\hat{\mathbf{Y}}_{j,k} - \hat{\mathbf{Y}}_{i,k}\cdot\sqrt{\frac{\boldsymbol{\delta}_j}{|\mathcal{Y}|}}\cdot\sqrt{\frac{\boldsymbol{\sigma}_k}{|\mathcal{Y}|}} - \hat{\mathbf{Y}}_{j,k}\cdot\sqrt{\frac{\boldsymbol{\delta}_i}{|\mathcal{Y}|}}\cdot\sqrt{\frac{\boldsymbol{\sigma}_k}{|\mathcal{Y}|}}\\
&\quad+\frac{\sqrt{\boldsymbol{\delta}_i}\sqrt{\boldsymbol{\delta}_j}\cdot\boldsymbol{\sigma}_k}{|\mathcal{Y}|^2}\\
&=\sum_{k=1}^{|C|}\hat{\mathbf{Y}}_{i,k}\cdot\hat{\mathbf{Y}}_{j,k} + \sum_{k=1}^{|C|}\frac{\sqrt{\boldsymbol{\delta}_i}\sqrt{\boldsymbol{\delta}_j}\cdot\boldsymbol{\sigma}_k}{|\mathcal{Y}|^2}\\
&\quad-\sum_{k=1}^{|C|}\mathbf{Y}_{i,k}\cdot\frac{\sqrt{\boldsymbol{\delta}_j}}{\sqrt{\boldsymbol{\delta}_i}\cdot|\mathcal{Y}|} - \sum_{k=1}^{|C|}\mathbf{Y}_{j,k}\cdot\frac{\sqrt{\boldsymbol{\delta}_i}}{\sqrt{\boldsymbol{\delta}_j}\cdot|\mathcal{Y}|}\\
&=\sum_{k=1}^{|C|}\hat{\mathbf{Y}}_{i,k}\cdot\hat{\mathbf{Y}}_{j,k} + \frac{\sqrt{\boldsymbol{\delta}_i}\sqrt{\boldsymbol{\delta}_j}}{|\mathcal{Y}|} - \frac{2\sqrt{\boldsymbol{\delta}_i}\sqrt{\boldsymbol{\delta}_j}}{|\mathcal{Y}|}\\
&=\sum_{k=1}^{|C|}\hat{\mathbf{Y}}_{i,k}\cdot\hat{\mathbf{Y}}_{j,k} - \frac{\sqrt{\boldsymbol{\delta}_i}\sqrt{\boldsymbol{\delta}_j}}{|\mathcal{Y}|}.
\end{aligned} \tag{22}$$

Since $\mathbf{Y}_{i,k} \in \{0,1\}$, we have

$$\|\hat{\mathbf{Y}}_i\|_2^2 = \sum_{k=1}^{|C|}\hat{\mathbf{Y}}_{i,k}^2 = \sum_{k=1}^{|C|}\frac{\mathbf{Y}_{i,k}^2}{\boldsymbol{\delta}_i\cdot\boldsymbol{\sigma}_k} = \sum_{k=1}^{|C|}\frac{\mathbf{Y}_{i,k}}{\boldsymbol{\delta}_i\cdot\boldsymbol{\sigma}_k}.$$

Here, we disregard the community with $\boldsymbol{\sigma}_k = 0$, since they convey no information. Therefore, with $\boldsymbol{\sigma}_k \geq 1$,

$$\|\hat{\mathbf{Y}}_i\|_2^2 = \sum_{k=1}^{|C|}\frac{\mathbf{Y}_{i,k}}{\boldsymbol{\delta}_i\cdot\boldsymbol{\sigma}_k} \leq \sum_{k=1}^{|C|}\frac{\mathbf{Y}_{i,k}}{\boldsymbol{\delta}_i} = 1.$$

For any two vectors $\mathbf{a},\mathbf{b} \in \mathbb{R}^n$, the Cauchy-Schwarz Inequality states $|\mathbf{a}\cdot\mathbf{b}| \leq \|\mathbf{a}\|_2 \cdot \|\mathbf{b}\|_2$. By Cauchy-Schwarz,

$$\left|\sum_{k=1}^{|C|}\hat{\mathbf{Y}}_{i,k}\cdot\hat{\mathbf{Y}}_{j,k}\right| \leq \|\hat{\mathbf{Y}}_i\|_2\cdot\|\hat{\mathbf{Y}}_j\|_2 \leq 1\cdot 1 = 1.$$

Since $\hat{\mathbf{Y}}_{i,k} \geq 0$, the dot product is non-negative

$$\sum_{k=1}^{|C|}\hat{\mathbf{Y}}_{i,k}\cdot\hat{\mathbf{Y}}_{j,k} \geq 0.$$

Then, we can derive $0 \leq \sum_{k=1}^{|C|}\hat{\mathbf{Y}}_{i,k}\cdot\hat{\mathbf{Y}}_{j,k} \leq 1$, which completes the proof by plugging it into Eq. (22). □

C Experimental Details

C.1 Datasets and Hyperparameter Settings

We describe the details of each dataset used in the experiments in what follows:

- *BlogCatalog* [46]: This is a network of social relationships of bloggers from the BlogCatalog website. In this scenario, we define the topic categories provided by the users as their preferred communities.
- *Flickr [46]*: The data retrieved from a well-known site that serves as a hub for users to showcase their personal pictures and video snippets. The community is defined as the interest groups that are formed by users.
- *Deezer-HR* & *Deezer-RO* [52]: The data are collected from the music streaming service Deezer. These datasets represent friendship networks of users from Croatia and Romania, respectively. In this dataset, we define the distinct music genres that users prefer as communities, which can be regarded as interest groups.
- *DBLP [71]*: The data is sourced from the DBLP computer science bibliography. Specifically, the social network under consideration is a co-authorship network where two authors are linked if they have co-authored at least one paper. Subsequently, the publication venue, such as a journal or a conference, serves as a ground-truth community. The community membership network represents authors who have published in a specific journal or conference.
- *Youtube [71]*: The data is collected from YouTube, a popular video-sharing website that incorporates a social network. In the YouTube social network, users can establish friendships with one another. Additionally, users have the ability to create groups to which other users can apply to join. For the purposes of this study, these user-defined groups are regarded as ground-truth communities.

We introduce the parameters that we did not mention in the main text. Some parameters are fixed for each dataset since it did not make a big difference for the experiment results, e.g., α in Eq. (6) is fixed as 0.33 for all datasets, the iteration rounds T of Eq. (6) is fixed as 2, and γ in Eq. (2) is fixed as 0.3. The hyperparameters λ in Eq. (14), β in Eq. (3) and θ in Eq. (18) are the optimal parameters obtained through grid search on each dataset.

Table 4: Parameter setting in CASO

	Datasets					
Parameter	*BlogCatalog*	*Flickr*	*Deezer-HR*	*Deezer-RO*	*DBLP*	*Youtube*
α	0.33	0.33	0.33	0.33	0.33	0.33
γ	0.3	0.3	0.3	0.3	0.3	0.3
T	2	2	2	2	2	2
λ	0.01	0.01	0.1	0.05	0.7	0.1
β	1	0.4	0.6	0.6	1	0.9
θ	1	1	0.05	0.5	0	0.01

C.2 Additional Experimental Results

To comprehensively evaluate the algorithm's performance, we recorded Recall@K and NDCG@K results from top-1 to top-5 for each dataset. The best result is highlighted in **bold**, and the runner-up is underlined.

The results clearly demonstrate that our proposed CASO almost consistently outperforms all baseline approaches across all tested datasets, with conspicuous improvements observed from top-1 to top-5 performance metrics. Specifically, on average across NDCG@1 to NDCG@5, CASO achieves considerable improvements over the best baselines, with gains of 2.3%, 4.3%, 3.2%, 3.1%, 5.6%, and 6.4% on the *BlogCatalog*, *Flickr*, *Deezer-HR*, and *Deezer-RO*, *DBLP* and *Youtube* respectively. The improvements achieved by CASO in recall are also significant, with recall gains typically ranging from 1% to as high as 5% across most datasets. For *Flickr* and *DBLP*, CASO demonstrates nearly identical performance to the state-of-the-art models MHCN and GBSR in Recall@5 while showing remarkable superiority across Recall@1 to Recall@5. It is worth noting that CASO often achieves substantial improvements in top-1 performance, with average gains of 5.9% and 5.5% on Recall@1 and NDCG@1, respectively. This indicates that by integrating global and local information, CASO can accurately identify the single community that users are most likely to be interested in.

On smaller datasets like *BlogCatalog* and *Flickr*, even optimal baseline methods such as MHCN integrate multiple triangle structures and self-supervised learning but show notable gaps in extracting global topological information. In contrast, our CASO consistently outperforms such baselines, with typical improvements ranging from 0.005 to 0.03 on *BlogCatalog*. Due to the data structure of *Flickr* possibly being more suited to triangular forms, our CASO shows slightly weaker performance than MHCN in Recall@5, while achieving average increases of 0.014 and 0.026 in other Recall and NDCG metrics, respectively. On medium-sized datasets like *Deezer-HR* and *Deezer-RO* (with dense community structures), our CASO stands out by effectively capturing close user-community connections. While GBSR, the denoise-based optimal baseline, removes interfering edges, it lacks the ability to model such community structures. As a result, GBSR averages a 0.02 decline relative to CASO. On large datasets *DBLP* and *Youtube*, which contain millions of social friendships, CASO achieves at least 1.0% and 4.4% improvements across Recall@1 to Recall@5 and NDCG@1 to NDCG@5 on *Youtube*. On *DBLP*, CASO achieves at least 0.3% and 4.4% improvements across Recall@1 to Recall@4 and NDCG@1 to NDCG@5, respectively, while GBSR occasionally shows better Recall@5 performance. The results highlight the efficacy of our CASO in leveraging and fusing social and collaborative information for community recommendation.

Table 5: Additional experimental results on *BlogCatalog*.

Method	**BlogCatalog**									
	Recall@1	Recall@2	Recall@3	Recall@4	Recall@5	NDCG@1	NDCG@2	NDCG@3	NDCG@4	NDCG@5
SVD++ [37]	0.1721	0.3297	0.5052	0.6686	0.8331	0.1721	0.2711	0.3587	0.4317	0.4854
BPR [51]	0.1628	0.3341	0.5069	0.6753	0.8387	0.1628	0.2720	0.3570	0.4282	0.4921
LightGCN [24]	0.1672	0.3378	0.5035	0.6713	0.8322	0.1672	0.2754	0.3590	0.4287	0.4920
LightGCN-S [74]	0.5762	0.8098	0.9026	0.9550	0.9831	0.5762	0.7254	0.7713	0.7935	0.8037
DiffNet [66]	0.6290	0.8165	0.8962	0.9534	0.9842	0.6290	0.7457	0.7875	0.8102	0.8232
DiffNet++ [65]	0.6195	0.8025	0.8856	0.9398	0.9747	0.6195	0.7329	0.7753	0.7991	0.8131
SEPT [77]	0.1690	0.3489	0.5106	0.6836	0.8251	0.1690	0.2891	0.3664	0.4309	0.4991
MHCN [78]	0.6853	0.8726	0.9419	0.9732	0.9915	0.6853	0.8072	0.8379	0.8539	0.8609
GBSR [74]	0.5822	0.7665	0.8530	0.9205	0.9646	0.5822	0.6976	0.7426	0.7712	0.7873
CASO	**0.7127**	**0.8960**	**0.9517**	**0.9815**	**0.9962**	**0.7127**	**0.8271**	**0.8535**	**0.8676**	**0.8741**
Improv.	**3.993%**	**2.678%**	**1.046%**	**0.852%**	**0.466%**	**3.993%**	**2.463%**	**1.868%**	**1.608%**	**1.527%**

Table 6: Additional experimental results on *Flickr*.

Method	**Flickr**									
	Recall@1	Recall@2	Recall@3	Recall@4	Recall@5	NDCG@1	NDCG@2	NDCG@3	NDCG@4	NDCG@5
SVD++ [37]	0.1086	0.2207	0.3373	0.4514	0.5599	0.1086	0.1813	0.2447	0.2839	0.3264
BPR [51]	0.1077	0.2164	0.3358	0.4446	0.5612	0.1077	0.1810	0.2362	0.2858	0.3299
LightGCN [24]	0.1090	0.2230	0.3352	0.4478	0.5566	0.1090	0.1806	0.2364	0.2851	0.3277
LightGCN-S [74]	0.3679	0.5518	0.6705	0.7605	0.8352	0.3679	0.4841	0.5447	0.5824	0.6124
DiffNet [66]	0.4613	0.6371	0.7532	0.8306	0.8855	0.4613	0.5760	0.6309	0.6643	0.6838
DiffNet++ [65]	0.3535	0.5322	0.6657	0.7605	0.8257	0.3535	0.4639	0.5310	0.5689	0.5974
SEPT [77]	0.1684	0.2933	0.3950	0.4697	0.5741	0.1684	0.2492	0.3086	0.3478	0.3774
MHCN [78]	0.4882	0.6921	0.8069	0.8916	**0.9381**	0.4882	0.6186	0.6789	0.7093	0.7277
GBSR [74]	0.3226	0.4803	0.6004	0.6939	0.7752	0.3226	0.4221	0.4826	0.5227	0.5537
CASO	**0.5241**	**0.7183**	**0.8191**	**0.8917**	0.9333	**0.5281**	**0.6504**	**0.6996**	**0.7289**	**0.7466**
Improv.	**7.355%**	**3.776%**	**1.522%**	**0.015%**	-0.507%	**8.167%**	**5.128%**	**3.050%**	**2.775%**	**2.597%**

Table 7: Additional experimental results on *Deezer-HR*.

Method	**Deezer-HR**									
	Recall@1	Recall@2	Recall@3	Recall@4	Recall@5	NDCG@1	NDCG@2	NDCG@3	NDCG@4	NDCG@5
SVD++ [37]	0.0732	0.1340	0.1864	0.2822	0.3187	0.1387	0.1518	0.1694	0.2017	0.2147
BPR [51]	0.2899	0.4236	0.5213	0.5919	0.6467	0.4921	0.4928	0.5169	0.5400	0.5605
LightGCN [24]	0.2703	0.4084	0.5030	0.5703	0.6308	0.4814	0.4799	0.5018	0.5233	0.5462
LightGCN-S [74]	0.2914	0.4313	0.5262	0.5952	0.6508	0.4895	0.4962	0.5194	0.5422	0.5638
DiffNet [66]	0.1898	0.2801	0.3617	0.4176	0.4813	0.2913	0.3034	0.3351	0.3597	0.3832
DiffNet++ [65]	0.1714	0.2757	0.3578	0.4149	0.4778	0.2620	0.2920	0.3240	0.3475	0.3713
SEPT [77]	0.2670	0.4306	0.5331	0.6071	0.6576	0.4774	0.4964	0.5206	0.5442	0.5635
MHCN [78]	0.0210	0.0804	0.1346	0.2085	0.2229	0.0540	0.0913	0.1144	0.1483	0.1532
GBSR [74]	0.2921	0.4356	0.5342	0.6118	0.6718	0.4884	0.4982	0.5234	0.5503	0.5738
CASO	**0.2946**	**0.4509**	**0.5602**	**0.6382**	**0.6983**	**0.4957**	**0.5159**	**0.5446**	**0.5723**	**0.5949**
Improv.	**0.823%**	**3.514%**	**4.856%**	**4.314%**	**3.939%**	**0.741%**	**3.547%**	**4.054%**	**4.003%**	**3.676%**

Table 8: Additional experimental results on *Deezer-RO*.

Method	Deezer-RO									
	Recall@1	Recall@2	Recall@3	Recall@4	Recall@5	NDCG@1	NDCG@2	NDCG@3	NDCG@4	NDCG@5
SVD++ [37]	0.0000	0.0188	0.1127	0.1595	0.2174	0.0000	0.0160	0.0743	0.0958	0.1189
BPR [51]	0.2862	0.4194	0.5297	0.6090	0.6677	0.4829	0.4850	0.5164	0.5444	0.5665
LightGCN [24]	0.2783	0.4207	0.5152	0.5818	0.6371	0.4828	0.4884	0.5105	0.5312	0.5518
LightGCN-S [74]	0.2816	0.4282	0.5337	0.6035	0.6609	0.4712	0.4862	0.5162	0.5402	0.5624
DiffNet [66]	0.1765	0.2824	0.3569	0.4326	0.4964	0.2725	0.3005	0.3263	0.3550	0.3840
DiffNet++ [65]	0.1719	0.2688	0.3511	0.4246	0.4985	0.2646	0.2863	0.3187	0.3491	0.3800
SEPT [77]	0.2760	0.4332	0.5357	0.6093	0.6644	0.4801	0.4979	0.5243	0.5482	0.5688
MHCN [78]	0.2714	0.4214	0.5351	0.6057	0.6628	0.4707	0.4842	0.5165	0.5396	0.5612
GBSR [74]	0.2824	0.4313	0.5323	0.6111	0.6750	0.4705	0.4880	0.5160	0.5433	0.5683
CASO	**0.2969**	**0.4578**	**0.5643**	**0.6405**	**0.7019**	**0.4915**	**0.5128**	**0.5423**	**0.5683**	**0.5906**
Improv.	**3.729%**	**5.685%**	**5.335%**	**4.809%**	**3.984%**	**1.779%**	**2.992%**	**3.434%**	**3.659%**	**3.837%**

Table 9: Additional experimental results on *DBLP*.

Method	DBLP									
	Recall@1	Recall@2	Recall@3	Recall@4	Recall@5	NDCG@1	NDCG@2	NDCG@3	NDCG@4	NDCG@5
SVD++ [37]	0.0001	0.0002	0.0006	0.0028	0.0087	0.0001	0.0002	0.0004	0.0014	0.0036
BPR [51]	0.0688	0.1218	0.1666	0.2055	0.2339	0.0713	0.1031	0.1256	0.1416	0.1532
LightGCN [24]	0.0948	0.1343	0.1572	0.1731	0.1887	0.1004	0.1218	0.1332	0.1402	0.1447
LightGCN-S [74]	0.6922	0.8446	0.8993	0.9230	0.9366	0.7082	0.7942	0.8215	0.8323	0.8372
DiffNet [66]	0.5387	0.6632	0.7326	0.7704	0.7993	0.5478	0.6206	0.6556	0.6728	0.6836
DiffNet++ [65]	0.6884	0.7783	0.8130	0.8296	0.8436	0.6978	0.7458	0.7633	0.7707	0.7763
SEPT [77]	0.2600	0.3467	0.3925	0.4136	0.4231	0.2655	0.3137	0.3386	0.3474	0.3514
MHCN [78]	0.7220	0.8216	0.8663	0.8885	0.9037	0.7374	0.7840	0.8024	0.8121	0.8170
GBSR [74]	0.6948	0.8491	0.9039	0.9285	**0.9424**	0.7113	0.7977	0.8251	0.8350	0.8398
CASO	**0.7791**	**0.8814**	**0.9146**	**0.9317**	0.9415	**0.7947**	**0.8493**	**0.8656**	**0.8730**	**0.8768**
Improv.	**7.913%**	**3.800%**	**1.182%**	**0.340%**	-0.099%	**7.761%**	**6.460%**	**4.904%**	**4.555%**	**4.404%**

Table 10: Additional experimental results on *Youtube*.

Method	Youtube									
	Recall@1	Recall@2	Recall@3	Recall@4	Recall@5	NDCG@1	NDCG@2	NDCG@3	NDCG@4	NDCG@5
SVD++ [37]	0.0001	0.0003	0.0003	0.0005	0.0006	0.0001	0.0002	0.0003	0.0003	0.0004
BPR [51]	0.0518	0.0828	0.1059	0.1256	0.1414	0.0679	0.0795	0.0903	0.0976	0.1048
LightGCN [24]	0.1068	0.1589	0.1908	0.2159	0.2346	0.1355	0.1541	0.1680	0.1783	0.1861
LightGCN-S [74]	0.3262	0.4436	0.5122	0.5581	0.5929	0.3624	0.4181	0.4501	0.4697	0.4838
DiffNet [66]	0.1818	0.2579	0.3097	0.3479	0.3775	0.1924	0.2359	0.2618	0.2788	0.2907
DiffNet++ [65]	0.2791	0.3460	0.3852	0.4164	0.4445	0.2875	0.3252	0.3435	0.3588	0.3692
SEPT [77]	0.1224	0.1762	0.2133	0.2386	0.2603	0.1483	0.1693	0.1858	0.1969	0.2055
MHCN [78]	0.2246	0.3126	0.3709	0.4156	0.4538	0.2610	0.2970	0.3237	0.3415	0.3543
GBSR [74]	0.3541	0.4815	0.5540	0.6000	0.6348	0.3943	0.4535	0.4870	0.5067	0.5203
CASO	**0.3946**	**0.5066**	**0.5681**	**0.6107**	**0.6413**	**0.4350**	**0.4849**	**0.5132**	**0.5316**	**0.5435**
Improv.	**11.447%**	**5.208%**	**2.540%**	**1.778%**	**1.038%**	**10.335%**	**6.930%**	**5.380%**	**4.907%**	**4.443%**