Experimental Setup

1. Datasets

To evaluate SimRec’s performance, three real-world benchmark datasets from online services are used. The interaction data is divided into training, validation, and test sets with a ratio of 70%:5%:25%. The details of these datasets are as follows:

* Gowalla: This dataset comes from the Gowalla platform and includes user check-in records at various locations, collected between January and June 2010.
* Yelp: Sourced from the Yelp platform, this dataset includes user ratings on venues, gathered from January to June 2018.
* Amazon: This dataset features user rating behaviors for books on the Amazon platform, collected during the year 2013.

A table with numbers and text

AI-generated content may be incorrect.

1. Evaluation Metrics

Similar to previous collaborative-filtering techniques, the authors use all-rank evaluation, in which positive items from test set are ranked with al lun-interacted items for each user. They used Recall@N and NDCG@N metrics[16,32] for evaluation, with a default N=20

1. Models comparison:

In this paper, the authors compare 14 baselines from 4 research lines for comprehensive validation.

**Traditional Collaborative Filtering Technique:**

* BiasMF: A traditional matrix factorization method that incorporates both user and item biases along with trainable embedding vectors.

**Non-GNN Neural Collaborative Filtering Models**

* NCF: One of the earlier collaborative filtering models based on deep learning, using multilayer perceptrons (MLPs) to better capture user-item interactions.
* AutoR: Utilizes a three-layer autoencoder with fully connected layers to represent user interaction data.

**Graph Neural Network-Based Collaborative Filtering Models**

* PinSage: Integrates random walks with graph convolutional networks to scale recommendations to large graphs.
* STGCN: Enhances graph convolutional networks with autoencoding subnetworks to improve inductive reasoning.
* GCMC: A foundational model that introduces graph convolution into matrix completion tasks.
* NGCF: Applies graph convolution operations directly on the user-item interaction graph to learn representations.
* GCCF and LightGCN: Simplify traditional GCNs by removing feature transformations and activation functions to improve performance.

**Disentangled GNN-Based Collaborative Filtering**

* DGCF: Breaks down user-item interactions into multiple latent components during graph message passing.

**Self-Supervised Learning for Recommendation**

* SLRec: Uses contrastive learning and feature-level augmentations in recommendation systems.
* NCL: Enhances self-supervised graph-based collaborative filtering with neighbor-level contrastive learning.
* SGL: Applies various graph and feature augmentations alongside contrastive learning techniques.
* HCCF: Combines global hypergraph neural networks with cross-view contrastive learning to enhance GNN-based collaborative filtering.

RQ1 shows :

1. **Superior Performance of SimRec:**
   * SimRec consistently outperforms baseline methods across various metrics.
   * Statistical significance is confirmed through 𝑝-value calculations over multiple re-training sessions.
   * The MLP-based architecture of SimRec, enhanced by adaptive contrastive knowledge distillation, yields more accurate recommendations than state-of-the-art GNN models.
   * The dual-level knowledge distillation in SimRec captures high-order collaborative signals while mitigating over-smoothing effects.
2. **Advantages Over Self-Supervised GNN Models:**
   * Despite the enhancements from self-supervised learning (SSL) in GNN-based collaborative filtering models, SimRec demonstrates superior performance.
   * Existing SSL frameworks may inadvertently amplify over-smoothing, as seen in models like SGL, which introduce random noise that can degrade embedding quality through high-order propagation.
   * Methods such as NCL and HCCF, which establish connections based on global semantic similarities, might over-smooth nodes that are originally distant in the graph.
   * In contrast, SimRec's graph-less design and adaptive contrastive regularization effectively prevent over-smoothed embeddings.
3. **Limitations of Non-GNN Collaborative Filtering Models:**
   * Non-GNN models like NCF and AutoR exhibit significantly poorer performance, even though they utilize MLP-based architectures similar to SimRec's inference model.
   * This disparity highlights the challenges MLPs face in capturing high-order graph relationships within user and item embeddings.
   * SimRec addresses this by incorporating knowledge distilled from advanced GNN models, enhancing MLP optimization and adaptively filtering over-smoothing signals.
   * The substantial performance gap underscores the effectiveness of SimRec's contrastive knowledge distillation approach.

These findings collectively demonstrate SimRec's ability to deliver more accurate and efficient recommendations by effectively addressing the limitations associated with both GNN-based and non-GNN collaborative filtering models.

RQ3: Model Scalability Study

To validate the efficiency of our SimRec in handling large-scale real-world data, we compare SimRec with the best performed baselines on a e-commerce data collected from Tmall platform. The dataset contains around 40 million records of user clicks. To successfully run on this dataset, GNN-based methods have to sample subgraphs for information propagation. In contrast, graph sampling is not required by the MLP-based inference model of our SimRec.

* More Accurate Recommendations: SimRec achieves better recommendation performance in terms of Recall and NDCG. This reflects the higher probability of over-smoothing on the large but sparse interaction graph. Our SimRec avoids this problem without explicit graph message passing. Instead, informative knowledge is distilled from GNNs for model compression.
* Much Higher Efficiency: SimRec greatly reduces the inference time on the large Tmall data. Firstly, the embedding process of our MLP predictor is agnostic to the holistic interaction graph, thus the large-scale graph does not increase much overhead for embedding processing. No graph sampling is required in compar-ison to GNNs. Secondly, SimRec infers user-item relations based on simple MLPs. The computational costs of fully-connected layers in MLPs are much lower than the cost of GNNs.

RQ4: Hyperparameter Study

A graph of different types of data

AI-generated content may be incorrect.

The authors conduct a hyperparameter study related to distillation and regularization strengths to understand how key parameters affect the performance of SimRec. They focus on four main hyperparameters:

Prediction-Level Distillation (λ₁, τ₁):

* λ₁ (loss weight):
  + Small values → insufficient knowledge transfer → poor performance.
  + Large values → distillation loss dominates → harms main training objective.
* τ₁ (temperature):
  + Smaller τ₁ produces larger gradients, making distillation more effective.
* Balanced setting: moderate λ₁ with lower τ₁ yields optimal results.

**Embedding-Level Distillation (λ₂, τ₂):**

* Controls how closely the student’s embeddings (from MLP) match the teacher’s (from GNN).
* Trade-off:
  + Too strong → over-regularization.
  + Too weak → not enough alignment.
* Moderate λ₂ and τ₂ values ensure useful but not overly restrictive guidance.

**Contrastive Regularization (λ₃, τ₃):**

* Prevents over-smoothing of embeddings by encouraging dissimilarity where needed.
* Findings:
  + Small λ₃ or large τ₃ → weak regularization → over-smoothing.
  + Excessive regularization → hurts modeling of node-level relationships.
* Needs careful tuning to strike a balance between distinctiveness and cohesion.

**Per-Batch Distillation Sample Size (|T₁|):**

* Refers to how many prediction-level samples are used in each batch for distillation.
* Observation:
  + Larger batch sizes improve performance by better capturing useful patterns.
  + However, gains plateau after a point — due to diminishing returns and computational cost.

RQ5: Over-Smoothing Investigation

To investigate if graph-less SimRec framework is able to mitigate the over-smoothing effect in graph-structured relation learning for CF, the authors compare representative baselines on SimRec model on the Mean Average Distance (MAD) values over embeddings for the most popular users and items. The evaluation results in table 6 shows that SimRec has higher MAD values on both user and item embeddings for Gowalla and Yelp data, in comparison to not only GCN model GCCF, but also state-of-the-art SSL frameworks.

A table with numbers and text

AI-generated content may be incorrect.

It can be concluded that SimRec framework better addresses the over-smoothing issue, by learning more uniform distributed embeddings for users and items, to better characterize their unique interaction patterns. This should be attributed to the MLP-based inference framework, and the contrastive regularization that adaptively alleviates over-smoothing signals.