SAME ITEMS, DIFFERENT ROLES: CONTEXT-AWARE FACET DISENTANGLEMENT GRAPH NETWORK FOR DISCRIMINATIVE BUNDLE RECOMMENDATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Discriminative bundle recommendation aims to rank predefined item collections (i.e., bundles) for users. Multi-view graph methods integrate user-item, user-bundle, and bundle-item relations. However, they do not use co-member-conditioned facet roles and explicit bundle structure to dynamically select from history-derived interests when ranking candidate bundles. To address this gap, we propose a context-aware facet disentanglement graph network for discriminative bundle recommendation, namely CFDGN. First, a co-member-conditioned router combines intrinsic item facets with bundle context to infer bundle-specific roles. Second, a role-aware component summarizes role-derived bundle structure as three unlabeled soft states (i.e., Similar, Complementary, and Noise). Third, historical bundles provide candidate-independent interest values for each user. The current candidate selects their mixture, which is then evaluated through four structural matching paths. A positive pointwise mutual information (PPMI) prior derived exclusively from training-observed bundles supplies auxiliary relation evidence. A calibrated structural residual augments the backbone MultiCBR's collaborative score. Experiments on three benchmark datasets show that CFDGN exceeds eleven competing baselines on both Recall and NDCG. Its relative Recall@20 gains over MultiCBR are 3.8%, 1.7%, and 8.3%, respectively.

1 INTRODUCTION

The same item can play different roles across collections. A hoodie anchors a sporty outfit yet serves as a mid layer in a winter look. Online services organize items into ready-made collections (i.e., bundles), such as playlists, outfits, and product sets. Users then obtain a group of related items in one decision. Discriminative bundle recommendation ranks these predefined bundles for each user rather than generating new ones (Sun et al., 2022; 2026; Tu et al., 2026). Accurate ranking requires both user preference and the roles of member items.

Early bundle recommenders combine user-bundle and user-item evidence through factorization or attention (Cao et al., 2017; Chen et al., 2019; Li et al., 2021). Later studies turn to the member side by contextualizing items within a bundle, enriching item-level representations, or modeling bundling strategies (Brosh et al., 2022; Du et al., 2023; Wei et al., 2023; Nguyen et al., 2024). Recently, graph-based models propagate sparse signals over relations among users, bundles, and items and have achieved strong performance (Deng et al., 2020; Chang et al., 2020; 2023). CrossCBR and MultiCBR further align or fuse these interaction views (Ma et al., 2022; 2024; Li et al., 2026b). Along another line, MIDGN and HIDGN disentangle user intents, so one item can express different intents across bundles (Zhao et al., 2022; Zou et al., 2024). More recent studies add distillation, adaptive graph fusion, or diffusion-based denoising (Ren et al., 2023; Tao et al., 2025; Li, 2025; Zhang et al., 2026a). These advances strengthen collaborative and intent representations. However, an item's role in a bundle is still not inferred by contrasting its intrinsic facets with its actual co-members. History-derived interests are also matched to every candidate in the same way. Two bundle-specific gaps therefore remain:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

(1) Bundle-specific structural roles of member items. An item carries multiple intrinsic facets. Only part of them align with a given bundle. The hoodie should read as an anchor in the sporty outfit yet as a plain mid layer in the winter look. Telling such cases apart requires comparing the item with its actual co-members. Existing bundle models contextualize items within a bundle or assign them to intent factors (Brosh et al., 2022; Zou et al., 2024), yet the role itself is never inferred by contrasting intrinsic salience with the co-member context. Without this contrast, the activation of secondary facets is left unexplained, and bundle composition cannot be built from member roles.

(2) Candidate-conditioned selection over history-derived interests. A user's historical bundles reveal several stable interests. A new candidate usually matches only part of them. Target-aware methods select among history-derived interests for each candidate (Zhou et al., 2018; Li et al., 2019), but the candidate in those models is a single item. For a bundle, the informative selection signal lies in the structure formed by its members. Throughout, the interest values must remain candidate-independent. Only their selection and matching should follow the current bundle.

To address these two challenges, we propose the Context-aware Facet Disentanglement Graph Network (CFDGN). CFDGN keeps MultiCBR as the collaborative backbone and adds a separate structural branch. Specifically, a co-member-conditioned router contrasts each item's intrinsic facet salience with the context of its actual co-members. An intrinsically secondary facet can therefore become active in a particular bundle. Then, role-weighted members yield explicit bundle semantics with three soft structural states (i.e., Similar, Complementary, and Noise). This avoids a separately learned structural bundle-ID embedding. Subsequently, historical bundles provide candidate-independent interest values, and each candidate selects their mixture according to the structure of its members. A positive pointwise mutual information (PPMI) prior derived from training-observed bundles supplies auxiliary relation evidence. Finally, four structural matching paths evaluate the selected mixture. The structural score enters the final ranking as a calibrated bounded residual over the backbone.

Our main contributions are summarized as follows:

- We present CFDGN, which introduces context-aware facet disentanglement into discriminative bundle recommendation.
- We devise a co-member-conditioned router that infers bundle-specific structural roles of member items.
- We introduce candidate-conditioned interest selection and a bounded structural residual over the collaborative backbone.
- Extensive experiments on three benchmark datasets show that CFDGN consistently outperforms eleven competing baselines.

2 Related Work

Bundle Recommendation. Bundle research covers two related tasks: ranking ready-made sets and constructing new ones. Recent construction methods learn multimodal bundling strategies (Nguyen et al., 2025). Ranking methods instead predict which existing bundles a user will prefer. General set-representation learning has also been tested on bundle ranking (Chen et al., 2026). CoHeat and EpicCBR address newly introduced bundles with sparse interactions (Jeon et al., 2024; Li et al., 2026c). MoDiffE uses diffusion experts for cold-start bundles (Li et al., 2026a). RDiffBR studies changing membership, and CoReSBR models a different, sequential bundle task (Lee & Kang, 2025). These advances address creation, cold start, or changing collections, but do not infer an item's structural role by contrasting its intrinsic facets with the actual co-members of each ranked bundle.

Graph-based Bundle Recommendation. Graph models connect users, bundles, and items so that sparse interactions can share information. AGF learns how to filter bundle neighborhoods, and dual-supervised contrastive learning adds interaction-based supervision (Li et al., 2023; Wu et al., 2024). Recent CCRL-MGF models meta-path relations across views, while MGC-BR augments multi-graph learning with language-model-derived signals for sparse users (Zhang et al., 2026b; Jiang et al., 2026). These methods improve graph propagation or view alignment. They do not jointly contrast intrinsic item facets with current co-members, construct bundle structure from the resulting roles, and use that structure to select among fixed history-derived interest values.

3 CFDGN FRAMEWORK

Figure 1 outlines CFDGN. MultiCBR supplies the collaborative score. The detached structural branch combines PPMI-enriched facets, contextual roles, and member-derived bundle semantics with candidate-conditioned selection over history-derived interests to form a four-path residual.

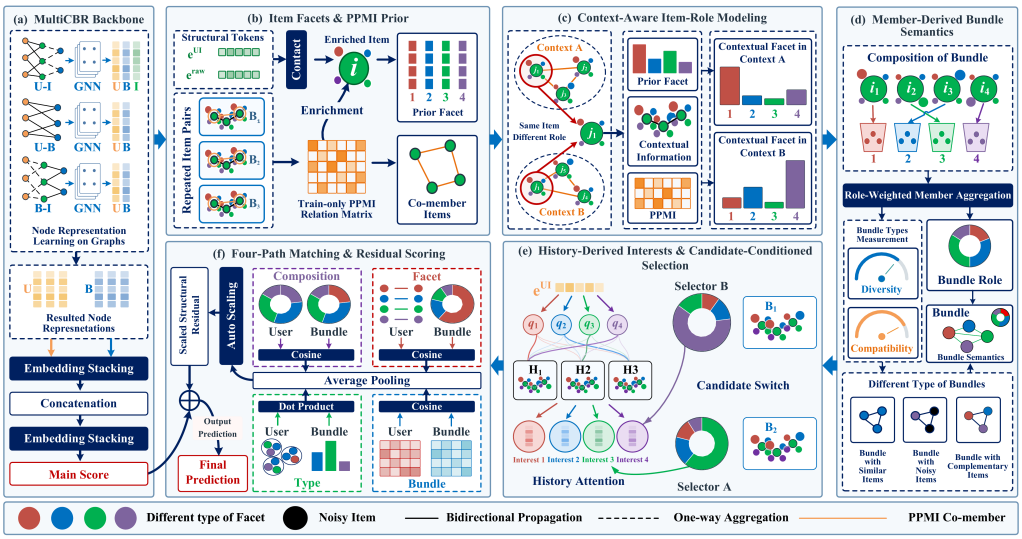


Figure 1: CFDGN augments MultiCBR with contextual item roles, member-derived bundle semantics, and candidate-conditioned selection over history-derived interests.

3.1 TASK AND COLLABORATIVE BACKBONE

Let $\mathcal{U}$, $\mathcal{B}$, and $\mathcal{I}$ be the user, bundle, and item sets. Binary matrices X, Y, and Z encode user-bundle (UB), user-item (UI), and bundle-item (BI) relations. For user u, we rank bundles absent from u's training interactions. MultiCBR fuses these views (Ma et al., 2024) into score s_{uB}^{main}. The structural branch reads detached embeddings through $\text{sg}(\cdot)$ and directly uses neither BI-propagated item embeddings nor a separate bundle-ID vector. Each structural bundle representation comes from its members.

3.2 CONTEXT-AWARE ITEM-ROLE MODELING

We first map item i to a structural token and $K \geq 2$ facets:

$$\mathbf{t}_i = f_{\text{item}}[\text{sg}(\mathbf{e}_i^{\text{raw}}) \| \, \text{sg}(\mathbf{e}_i^{UI})], \qquad \mathbf{r}_i^k = W_k \mathbf{t}_i, \tag{1}$$

where f_{item} is a bias-free linear-GELU-LayerNorm map to token $\mathbf{t}_i$, $\|$ concatenates the raw and UI embeddings, and W_k produces facet vector $\mathbf{r}_i^k$ for $k = 1, \ldots, K$.

Set $a_i^k = \|\mathbf{r}_i^k\|_2$ and $\mathbf{z}_i^k = \mathbf{r}_i^k/(a_i^k + \varepsilon)$, with $\varepsilon > 0$. Let a_i collects the K facet norms. A low-salience facet can still matter in a bundle.

Each training-observed bundle ID counts once per co-member pair, not per user interaction. Sparse positive pointwise mutual information (PPMI) weights A^{PPMI} retain positive non-self pairs seen at least twice, the top 20 per row, then symmetrize by elementwise maximum. Validation and test interactions are excluded. The row-normalized $\bar{A}$ passes same-index messages:

$$\mathbf{m}_i^k = \text{Norm}\left(\sum_j \bar{A}_{ij} \mathbf{z}_j^k\right), \qquad g_{i,P}^k = \text{sigmoid}\left(\frac{\langle \mathbf{z}_i^k, \mathbf{m}_i^k \rangle}{\tau_P}\right) \mathbf{1}[\|\mathbf{m}_i^k\|_2 > 0], \tag{2}$$

where $\mathbf{m}_i^k$ is the facet-k message, $g_{i,P}^k$ is its similarity gate, τ_P is the temperature, $\text{Norm}(x) = x/(\|x\|_2 + \varepsilon)$, and $\mathbf{1}[\cdot]$ is the indicator.

Fixed PPMI strength β_P gives $\widetilde{\mathbf{z}}_i^k = \mathrm{Norm}(\mathbf{z}_i^k + \beta_P g_{i,P}^k \mathbf{m}_i^k)$. Global PPMI neighbors enrich facets, but only members of B directly enter local routing and aggregation. If item i has no retained PPMI neighbor, $g_{i,P}^k = 0$. Operator $\mathcal{N}$ centers a K-facet vector, adds ε to its mean squared deviation, then divides by the square root of that sum.

For item i in candidate bundle B with $|B| \geq 2$, the router compares its intrinsic salience with that of its co-members:

$$c_{iB}^k = \frac{\sum_{j \in B \setminus \{i\}} a_j^k}{|B| - 1}, \qquad c_{iB}^{P,k} = \sum_{j \in B \setminus \{i\}} A_{ij|B} a_j^k, \tag{3}$$

where c_{iB}^k is the plain co-member signal, $c_{iB}^{P,k}$ is its PPMI-weighted counterpart, and $A_{ij|B} = A_{ij}^{\mathrm{PPMI}} / (\sum_{\ell \in B \setminus \{i\}} A_{i\ell}^{\mathrm{PPMI}} + \varepsilon)$ normalizes the retained prior over co-members.

Define $p_i = \mathcal{N}(\log(1 + a_i))$ as intrinsic salience, and let c_{iB} and c_{iB}^P collect their K signals. Both context sources are masked for singleton bundles. The PPMI source is also masked when no eligible co-member is present. The two context contrasts are

$$s_{iB} = p_i - \mathcal{N}(\log(1 + c_{iB})), \qquad s_{iB}^P = p_i - \mathcal{N}(\log(1 + c_{iB}^P)), \tag{4}$$

where s_{iB} is the plain-context contrast and s_{iB}^P is the PPMI-context contrast.

The router fuses them with p_i:

$$\ell_{iB} = \sum_{s \in \{p_i, s_{iB}, s_{iB}^P\}} w_s \mathcal{N}(s), \qquad g_{iB} = \mathrm{softmax}_k(\ell_{iB} / \tau_{\mathrm{role}}), \tag{5}$$

where ℓ_{iB} is the router logit, w_s is fixed to $0.40/0.45/0.15$ for $p_i/s_{iB}/s_{iB}^P$ in order, τ_{role} is the temperature, $\mathrm{softmax}_k$ normalizes over facets, and g_{iB} is the role distribution.

The source weights are fixed across datasets, not validation-selected. Unavailable sources are masked, and the remaining source weights are renormalized. Dependence on $B \setminus \{i\}$ allows an item to take different roles across bundles.

3.3 Member-derived bundle semantics

Role weights aggregate members separately for each facet:

$$\mathbf{h}_B^k = \sum_{i \in B} \alpha_{iB}^k \widetilde{\mathbf{z}}_i^k, \qquad \rho_B^k = \frac{1}{|B|} \sum_{i \in B} g_{iB}^k, \tag{6}$$

where $\alpha_{iB}^k = g_{iB}^k / (\sum_{j \in B} g_{jB}^k + \varepsilon)$ normalizes facet-k role weights within B, $\mathbf{h}_B^k$ is its member-derived facet, and ρ_B is its mean role composition.

Let H_{n} denote Shannon entropy divided by $\log K$. Let $\bar{H}_B$ be the average normalized entropy of member roles and H_ρ the normalized entropy of ρ_B. Role diversity and certainty are

$$D_B = \frac{\max(0, H_\rho - \bar{H}_B) \log K}{\log \min(|B|, K)}, \qquad S_B = 1 - \bar{H}_B, \tag{7}$$

where $D_B = 0$ for singleton bundles, D_B is normalized Jensen-Shannon role dispersion (Lin, 1991), and S_B measures how sharply each role is assigned.

We next estimate member compatibility. For $|B| \geq 2$, Cov_B^P is the share of $\binom{|B|}{2}$ member pairs linked in the retained A^{PPMI}. Over those pairs, Str_B^P averages bounded strength $1 - \exp[-\mathrm{PPMI}(i,j)]$ and is zero if none exist. Their product is $E_B^P = \mathrm{Cov}_B^P \mathrm{Str}_B^P$. Semantic and prior evidence combine:

$$d_B = \frac{\| \sum_{i \in B} \mathbf{v}_i \|_2^2 - |B|}{|B|(|B| - 1)}, \qquad C_B = 1 - (1 - C_B^{\mathrm{sem}})(1 - E_B^P), \tag{8}$$

where $\mathbf{v}_i = \mathrm{Norm}(W_v \mathbf{t}_i)$ is a normalized item vector with learned W_v, d_B approximates mean pairwise cosine by treating the stabilized $\mathbf{v}_i$ as unit vectors, $C_B^{\mathrm{sem}} = \mathrm{sigmoid}(\gamma d_B + b)$ is semantic compatibility with bias b and softplus-constrained $\gamma > 0$, and C_B is final compatibility that can only increase with positive prior evidence.

For bundles with at least two items, we summarize three possible structures with

$$\mathbf{p}_B = \mathrm{softmax}\left(\frac{[S_B(1-D_B)C_B,\ S_B D_B C_B,\ S_B D_B(1-C_B)]}{\tau_{\mathrm{type}}}\right), \tag{9}$$

where τ_{type} is a temperature, the entries are unlabeled Similar/Complementary/Noise tendencies, and S_B changes softmax sharpness but not state order when $S_B > 0$.

For singleton bundles, set $d_B = D_B = E_B^P = \mathrm{Cov}_B^P = 0$, $C_B = 0.5$, and $\mathbf{p}_B = (1/3, 1/3, 1/3)$. No pairwise relation is inferred. Because Noise also contains D_B, a low-diversity, low-compatibility bundle can remain ambiguous. Finally, we concatenate the member-derived signals:

$$\mathbf{z}_B = \mathrm{Norm}(W_B \mathbf{x}_B), \tag{10}$$

where $\mathbf{x}_B$ concatenates $\mathbf{h}_B^1, \ldots, \mathbf{h}_B^K$, then ρ_B, $\mathbf{p}_B$, D_B, C_B, S_B, and Cov_B^P in that order, and W_B projects $\mathbf{x}_B$ into the shared d_f-dimensional facet space without a separate bundle-ID vector.

3.4 User interests from history and candidate-conditioned selection

Let $\mathcal{H}_u$ contain the bundles observed for user u in training. A learned W_q^k projects the UI-view user embedding $\mathbf{e}_u^{UI}$ into candidate-independent query $\mathbf{q}_u^k = \mathrm{Norm}(W_q^k\, \mathrm{sg}(\mathbf{e}_u^{UI}))$. The history supplies its interest values:

$$\beta_{uB_h}^k = \frac{\exp(\cos(\mathbf{q}_u^k, \mathbf{h}_{B_h}^k)/\tau_{\mathrm{hist}})}{\sum_{B_h' \in \mathcal{H}_u} \exp(\cos(\mathbf{q}_u^k, \mathbf{h}_{B_h'}^k)/\tau_{\mathrm{hist}})}, \qquad \mathbf{h}_u^k = \sum_{B_h \in \mathcal{H}_u} \beta_{uB_h}^k \mathbf{h}_{B_h}^k, \tag{11}$$

where $\beta_{uB_h}^k$ attends to a historical bundle $B_h \in \mathcal{H}_u$, τ_{hist} is the attention temperature, and $\mathbf{h}_u^k$ is the candidate-independent value for interest k.

The same weights summarize historical soft states and role compositions:

$$\boldsymbol{\pi}_{u,\mathrm{type}}^k = \sum_{B_h \in \mathcal{H}_u} \beta_{uB_h}^k \mathbf{p}_{B_h}, \qquad \boldsymbol{\pi}_{u,\rho}^k = \sum_{B_h \in \mathcal{H}_u} \beta_{uB_h}^k \rho_{B_h}, \tag{12}$$

where $\boldsymbol{\pi}_{u,\mathrm{type}}^k$ and $\boldsymbol{\pi}_{u,\rho}^k$ are the history-weighted state and role summaries, respectively, and both sums use historical bundles B_h, not the current candidate B.

The reliability R_u^k equals $[1 + K \sum_{B_h \in \mathcal{H}_u} (\beta_{uB_h}^k)^2]^{-1}$ and grows with effective history size.

Candidate B changes the weights ω_{uB}^k, not the historical values $\mathbf{h}_u^k$:

$$\omega_{uB}^k = \mathrm{softmax}_k\big(\cos(\mathbf{q}_u^k, \mathbf{h}_B^k)/\tau_{\mathrm{match}}\big), \tag{13}$$

where ω_{uB}^k is the active weight for interest k and τ_{match} is a temperature.

Let c_{uB}^k and $\widetilde{c}_{uB}^k$ be the raw and reliability-adjusted composition matches. They are $c_{uB}^k = \cos(\boldsymbol{\pi}_{u,\rho}^k, \rho_B)$ and $\widetilde{c}_{uB}^k = 0.5 + R_u^k(c_{uB}^k - 0.5)$.

In training, both B^+ and B^- use the same history excluding B^+. Triples with no remaining history contribute no residual loss. At inference, history-free users keep the backbone score.

3.5 Four-path matching and learning

The selector compares history-derived interests with the current bundle through four structural paths. Facet and bundle matching are

$$s_{\mathrm{facet}} = \sum_k \omega_{uB}^k \cos(\mathbf{h}_u^k, \mathbf{h}_B^k), \qquad s_{\mathrm{bundle}} = \cos(\mathbf{z}_{uB}, \mathbf{z}_B), \tag{14}$$

where $\mathbf{z}_{uB} = \sum_k \omega_{uB}^k \mathrm{Norm}(\mathbf{h}_u^k)$ forms a candidate-weighted mixture of history-derived interests, s_{bundle} compares this mixture with the member-derived bundle vector, and s_{facet} compares corresponding facets.

State and composition matching are

$$s_{\mathrm{type}} = \sum_k \omega_{uB}^k \langle \boldsymbol{\pi}_{u,\mathrm{type}}^k, \mathbf{p}_B \rangle, \qquad s_{\mathrm{comp}} = \sum_k \omega_{uB}^k \widetilde{c}_{uB}^k, \tag{15}$$

where s_{type} compares soft structural states and s_{comp} compares role compositions.

The cosine paths lie in $[-1, 1]$. We map the two distribution-matching paths from $[0, 1]$ to $[-1, 1]$ within a fixed equal-weight fusion:

$$r_{uB}^{\text{raw}} = \tfrac{1}{4}\left[s_{\text{facet}} + s_{\text{bundle}} + (2s_{\text{type}} - 1) + (2s_{\text{comp}} - 1)\right], \tag{16}$$

where $r_{uB}^{\text{raw}} \in [-1, 1]$ is the fused raw structural score.

We add it as a bounded residual:

$$s_{uB}^{\text{final}} = s_{uB}^{\text{main}} + \alpha\sigma_{\text{Main}} r_{uB}^{\text{raw}}, \qquad \alpha = 0.5\,\text{sigmoid}(\theta_\alpha), \tag{17}$$

where $\sigma_{\text{Main}} = [\mathbb{E}_{u\in\mathcal{U}} \text{Var}_{B\in\mathcal{B}}(s_{uB}^{\text{main}})]^{1/2}$ is the full-catalog score scale, $\alpha \in (0, 0.5)$ is learned through θ_α, and s_{uB}^{final} is the final ranking score.

The detached scale is refreshed from the non-augmented training graph at the start of each epoch and before evaluation. The four path weights are equal and fixed, not tuned.

For eligible training triples (u, B^+, B^-) with remaining history, B^+ is observed and B^- is sampled from bundles unobserved by u in training. Expectations below average over these triples. Let $m_{\text{main}} = \text{sg}(s_{uB^+}^{\text{main}} - s_{uB^-}^{\text{main}})/\sigma_{\text{Main}}$ be the detached backbone margin and $w = 0.1 + 0.9\,\text{sigmoid}(-m_{\text{main}})$ its hard-pair weight. Raw structural ranking uses

$$\mathcal{L}_{\text{raw}} = \mathbb{E}[\text{sg}(w)\,\text{softplus}(-(r_{uB^+}^{\text{raw}} - r_{uB^-}^{\text{raw}}))], \tag{18}$$

where w retains a 0.1 floor and $\mathcal{L}_{\text{raw}}$ trains structural scores without updating the backbone.

Calibration learns α from the detached raw structural score:

$$\delta_{uB} = \alpha\,\text{sg}(r_{uB}^{\text{raw}}), \qquad m_\delta = \delta_{uB^+} - \delta_{uB^-}, \tag{19a}$$

$$\mathcal{L}_{\text{cal}} = \mathbb{E}[w\,\text{softplus}(-m_\delta)], \tag{19b}$$

$$\mathcal{L}_{\text{preserve}} = \mathbb{E}\left[\tfrac{1-w}{4}(\delta_{uB^+}^2 + \delta_{uB^-}^2 + 2m_\delta^2)\right], \tag{19c}$$

$$\mathcal{L}_{\text{res}} = \mathcal{L}_{\text{raw}} + \mathcal{L}_{\text{cal}} + \mathcal{L}_{\text{preserve}}, \tag{19d}$$

where δ_{uB} is the score-scale-normalized correction used in calibration, m_δ its positive-negative gap, $\mathcal{L}_{\text{cal}}$ rewards positive gaps on harder pairs, $\mathcal{L}_{\text{preserve}}$ limits correction size and gaps on easier pairs, and $\mathcal{L}_{\text{res}}$ combines the residual losses.

Detached raw scores make calibration and preservation update only α.

Fixed-anchor prototype contrastive learning (PCL) and orthogonality (Orth) use $\mathcal{E}_{\text{bundle}}$ (distinct nonempty scored or historical bundles), $\mathcal{E}_{\text{item}}$ (their distinct members), and $\mathcal{E}_{\text{user}}$ (batch rows with nonempty leave-one-out history, repeated users count separately). For $e \in \{\text{item}, \text{bundle}, \text{user}\}$ and $x \in \mathcal{E}_e$, $\bar{\mathbf{z}}_{x,e}^k$ normalizes $\tilde{\mathbf{z}}_i^k$, $\mathbf{h}_B^k$, or $\mathbf{h}_u^k$ at the respective level. The losses are

$$\begin{aligned}
\mathcal{L}_{\text{PCL},e} &= \frac{1}{|\mathcal{E}_e|K} \sum_{x\in\mathcal{E}_e} \sum_{k=1}^{K} \text{CE}\big(\{\cos(\bar{\mathbf{z}}_{x,e}^k, \mathbf{p}_q)/\tau_{\text{PCL}}\}_{q=1}^K, k\big), \\
\mathcal{L}_{\text{Orth},e} &= \frac{1}{|\mathcal{E}_e|} \sum_{x\in\mathcal{E}_e} \frac{\sum_{k\neq l}[(Q_{x,e}Q_{x,e}^\top)_{kl}]^2}{K(K-1)},
\end{aligned} \tag{20}$$

where $\mathbf{p}_q$ is fixed orthogonal prototype q, CE is cross-entropy with 0.05 label smoothing, τ_{PCL} is its temperature, and $Q_{x,e}$ stacks the K unit facets of entity x.

Bundle regularizers detach α_{iB}^k; user regularizers also detach $\beta_{uB_h}^k$. Empty sets give zero loss. PCL aligns facet indices across levels; Orth reduces within-entity facet collapse.

Let $\mathcal{P}^+$ be retained-PPMI pairs and $\mathcal{P}^-$ an equal-size, seed-fixed bank of distinct zero-co-occurrence pairs. Training cycles through both sets. Using $\ell_{ij} = \gamma\langle \mathbf{v}_i, \mathbf{v}_j\rangle + b$ from Eq. (8), the relation loss is

$$\mathcal{L}_{\text{rel}} = -\tfrac{1}{2}\mathbb{E}_{(i,j)\sim\mathcal{P}^+} \log\text{sigmoid}(\ell_{ij}) - \tfrac{1}{2}\mathbb{E}_{(i,j)\sim\mathcal{P}^-} \log[1 - \text{sigmoid}(\ell_{ij})], \tag{21}$$

where ℓ_{ij} is the semantic logit and excludes the direct PPMI feature E_B^P.

The full objective is

$$\mathcal{L} = \mathcal{L}_{\text{MultiCBR}} + \lambda_1\mathcal{L}_{\text{rel}} + \lambda_2\mathcal{L}_{\text{res}} + \mathcal{L}_{\text{PCL}} + \mathcal{L}_{\text{Orth}}, \tag{22}$$

where $\mathcal{L}_{\text{MultiCBR}}$ is the backbone objective, λ_1 and λ_2 weight relation and residual losses, and $\mathcal{L}_R = \sum_e \lambda_{R,e}\mathcal{L}_{R,e}$ for $R \in \{\text{PCL}, \text{Orth}\}$.

4 EXPERIMENTS

We first introduce the experimental settings and then analyze CFDGN by answering four research questions:

- **RQ1:** How does CFDGN compare with established bundle recommenders?
- **RQ2:** How do its components affect ranking performance?
- **RQ3:** How do contextual roles, soft bundle states, and candidate-conditioned selection over history-derived interests behave?
- **RQ4:** How do the hyper-parameters affect the performance?

4.1 EXPERIMENTAL SETUP

Datasets and evaluation. We use the public Youshu, NetEase, and iFashion datasets. Table 1 gives their statistics. Following MultiCBR, CFDGN adopts the same data preprocessing and train/validation/test splits. We rank the full catalog after masking training interactions and report Recall@K and NDCG@K for $K \in \{20, 40\}$, as in recent bundle recommendation work (Gao et al., 2026). Table 2 includes comparable established methods.

Table 1: Dataset statistics (Train U-B: observed positives).

Dataset	Users	Bundles	Items	Train U-B	U-I edges	B-I edges	Mean size
Youshu	8,039	4,771	32,770	34,416	138,515	176,667	37.03
NetEase	18,528	22,864	123,628	217,864	1,128,065	1,778,838	77.80
iFashion	53,897	27,694	42,563	1,213,676	2,290,645	106,916	3.86

Implementation details. We use Adam (Kingma & Ba, 2015), batch size 2,048, one negative per positive pair, learning rate 10^{-3}, and L2 coefficient 10^{-5}. All datasets use a two-layer backbone ($d = 128$, $K = 4$, $d_f = 64$) and fixed PPMI strength $\beta_P = 0.1$. We select the checkpoint maximizing validation Recall and NDCG at 20 and 40, with @10 metrics breaking ties, and evaluate it once on test.

4.2 OVERALL PERFORMANCE (RQ1)

Table 2 reports test performance, with CFDGN averaged over seeds 2024, 2025, and 2026. The observations are summarized as follows. (1) CFDGN ranks first on every metric and dataset. (2) Compared with MultiCBR, its Recall@20 gains are 3.8%, 1.7%, and 8.3%, while its NDCG@20 gains are 2.1%, 1.8%, and 8.7% on Youshu, NetEase, and iFashion. The same trend holds at @40. (3) iFashion has the smallest mean bundle size (Table 1) and the largest gain. This association does not establish causation.

4.3 ABLATION STUDY (RQ2)

Table 3 reports seed-2024 Youshu test results under fixed settings. Backbone is same-protocol MultiCBR. The w/o variants remove the named source or loss, w/o Type also removes the type path. PriorOnly drops co-member routing, PrimaryOnly fixes the intrinsic top facet, MeanPoolBundle replaces only $\mathbf{z}_B$, and UniformInterest sets $\omega_{uB}^{k} = 1/K$.

The observations are summarized as follows. (1) Full CFDGN exceeds Backbone by 0.0111 Recall@20 and 0.0026 NDCG@20, directly supporting the complete structural branch. (2) PriorOnly and PrimaryOnly respectively lose 0.0010 Recall@20 and 0.0007 NDCG@20, supporting contextual roles and secondary facets. (3) MeanPoolBundle and UniformInterest reduce both metrics, supporting the projected $\mathbf{z}_B$ path and candidate-conditioned weighting, MeanPoolBundle retains the other bundle semantics. (4) w/o PCL has the largest Recall@20 loss (0.0017), while w/o Type and w/o Orth share the largest NDCG@20 loss (0.0008). w/o PPMI is effectively tied, so no PPMI ranking gain is established.

Table 2: Test results (CFDGN: mean over seeds 2024–2026).

Metric	MFBPR	LGCN	SGL	XSim	LGCL	DAM	BNet	BGCN	MIDGN	Cross	Multi	CFDGN
Youshu												
R@20	0.1959	0.2286	0.2622	0.2644	0.2710	0.2082	0.1895	0.2436	0.2682	0.2823	0.2851	**0.2960**
N@20	0.1117	0.1344	0.1551	0.1560	0.1556	0.1198	0.1125	0.1329	0.1527	0.1670	0.1703	**0.1739**
R@40	0.2735	0.3190	0.3630	0.3569	0.3693	0.2890	0.2675	0.3379	0.3712	0.3787	0.3964	**0.3994**
N@40	0.1320	0.1592	0.1825	0.1814	0.1827	0.1418	0.1335	0.1589	0.1808	0.1939	0.2001	**0.2021**
NetEase												
R@20	0.0355	0.0496	0.0716	0.0710	0.0705	0.0411	0.0391	0.0626	0.0678	0.0844	0.0917	**0.0933**
N@20	0.0181	0.0254	0.0386	0.0376	0.0371	0.0210	0.0201	0.0328	0.0343	0.0458	0.0498	**0.0507**
R@40	0.0600	0.0795	0.1104	0.1079	0.1122	0.0690	0.0661	0.0988	0.1085	0.1255	0.1355	**0.1376**
N@40	0.0246	0.0334	0.0488	0.0475	0.0481	0.0281	0.0271	0.0424	0.0451	0.0567	0.0615	**0.0624**
iFashion												
R@20	0.0752	0.0837	0.1000	0.1003	0.0894	0.0629	0.0626	0.0802	0.0856	0.1132	0.1559	**0.1688**
N@20	0.0542	0.0612	0.0739	0.0744	0.0653	0.0450	0.0447	0.0582	0.0473	0.0872	0.1287	**0.1399**
R@40	0.1162	0.1284	0.1498	0.1501	0.1362	0.0995	0.0986	0.1239	0.1299	0.1623	0.2131	**0.2260**
N@40	0.0687	0.0770	0.0915	0.0920	0.0818	0.0579	0.0574	0.0736	0.0593	0.1045	0.1495	**0.1600**

Table 3: Youshu test ablations (seed 2024; same-protocol MultiCBR backbone).

	w/o				Replacement				Backbone	Full
Metric	PPMI	Type	PCL	Orth	Prior	Primary	MeanPool	Bundle	Uniform	MultiCBR	CFDGN
R@20	**0.2963**	0.2959	0.2945	0.2956	0.2952	0.2958	0.2955	0.2956	0.2851	0.2962	
N@20	**0.1729**	0.1721	0.1724	0.1721	0.1724	0.1722	0.1727	0.1726	0.1703	**0.1729**	

4.4 CORE MECHANISM EVIDENCE (RQ3)

We examine contextual roles, member-derived soft tendencies, and candidate-conditioned mixtures of history-derived interests using unlabeled, formula-derived diagnostics. These diagnostics show internal mechanism behavior, but they do not establish externally validated bundle semantics.

Table 4: Mechanism diagnostics (lift: ratio; others: proportions).

Dataset	Role switch	Secondary active	Attn. lift	Interest switch	Router PPMI flip	Fused win
Youshu	0.4012	0.3569	2.7237	0.1869	0.0567	0.8447
NetEase	0.3436	0.4236	2.7406	0.8735	0.0572	0.8068
iFashion	0.4304	0.2762	2.3408	0.9436	0.0524	0.8257

(1) Contextual roles. Table 4 audits known bundle-item memberships. Role switch is the fraction of items with at least two memberships whose $\arg\max_k g_{iB}^k$ differs across bundles. On memberships of these repeated items, Secondary active is the fraction with $\arg\max_k g_{iB}^k \neq \arg\max_k a_i^k$. For the secondary-active subset, Attn. lift averages $\alpha_{iB}^{k^\star}/(1/|B|)$, where $k^\star = \arg\max_k g_{iB}^k$. Table 4 shows role switching, secondary activation, and lift above 2.3, these are internal diagnostics, not external facet labels.

Table 5: Non-singleton soft-state diagnostics: dominant S/C/N shares (%) and mean $D_B/C_B/S_B$.

Dataset	S/C/N share (%)	Similar $D/C/S$	Complement. $D/C/S$	Noise $D/C/S$
Youshu	92.71/5.97/1.32	0.414/0.822/0.499	0.521/0.829/0.567	0.429/0.369/0.472
NetEase	94.52/3.41/2.07	0.401/0.778/0.498	0.522/0.666/0.569	0.489/0.434/0.526
iFashion	53.88/11.56/34.56	0.375/0.834/0.518	0.585/0.835/0.612	0.468/0.172/0.554

(2) Soft bundle tendencies. Table 5 groups non-singleton bundles by $\arg\max \mathbf{p}_B$. Similar is largest for 92.71% of Youshu and 94.52% of NetEase bundles, Noise is largest for 34.56% of iFashion bundles. These are dominant-coordinate shares, not mean probabilities, ranking retains the full $\mathbf{p}_B$. Higher mean D_B for Complementary and lower C_B for Noise agree with Eq. (9), Complementary's higher S_B is observed, not implied by that rule.

At a frozen Youshu checkpoint, replacing member roles with one-hot intrinsic-primary roles changes $\arg\max \mathbf{p}_B$ for 42.4% of all $4{,}771$ catalog bundles, singletons included. This no-retraining probe shows role-to-state sensitivity, not improved ranking or external semantic validity.

(3) Interest selection. Interest switch is the share of users with at least two training bundles whose $\arg\max_k \omega_{uB}^k$ changes across those bundles, all candidates are checked. Router PPMI flip is the dominant-role change rate on PPMI-active repeated memberships after zeroing only the PPMI router source and renormalizing the rest at a frozen checkpoint. Fused win is the share of frozen validation pairs with $r_{uB+}^{\text{raw}} > r_{uB-}^{\text{raw}}$, this is the four-path fused score, not an individual path. Each positive has one seed-2024 negative unseen by that user in training, users without history are excluded. On $5{,}189/27{,}857/149{,}007$ pairs, Youshu/NetEase/iFashion win rates are $84.47\%/80.68\%/82.57\%$. None of these probes retrains the model.

Figure 2 gives two examples. The left side shows g_{iB} for Youshu item 14439 in three bundles. F3 is its intrinsic primary facet. Its weight is 0.91 in B2949, but B1482 and B3521 favor F0 (0.76) and F2 (0.48). Thus, 0.91 is contextual weight g_{iB}^3, not intrinsic salience. The right side shows ω_{uB} for NetEase user 2964. B4351 favors I1 (0.67), B148 and B10135 favor I3 $(0.46/0.49)$, and B20589 splits I2/I3 $(0.43$ each). Candidates change the mixture, not $\mathbf{h}_u^k$. These examples show mechanisms, not semantic labels, population rates, or ranking gains.

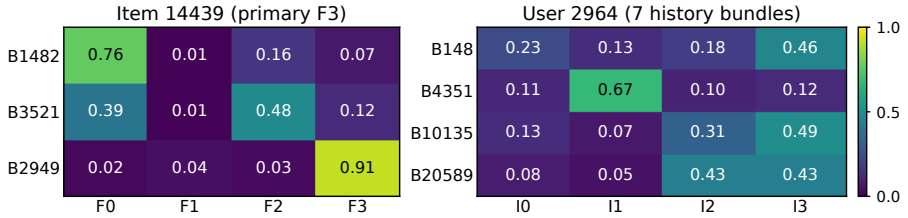


Figure 2: Contextual roles on Youshu and interest selection on NetEase. Indices are latent.

4.5 Hyper-parameters Analysis (RQ4)

Figure 3 studies d, K, and d_f on the seed-2024 Youshu validation set. Each point averages three $L_9(3^3)$ configurations at one factor level. For d, 64 to 128 raises both metrics because added capacity preserves richer collaborative signals, smaller gains at 256 show saturation. For K, 2 to 4 expands role and interest coverage, mainly boosting Recall@20, 4 to 8 adds finer facets that reduce top-rank interference, so NDCG@20 rises faster despite saturated coverage. For d_f, 32 to 64 adds role capacity, sharply improving Recall@20 but slightly lowering NDCG@20 as facets may overlap, 64 to 128 improves separation and top-rank order, so NDCG@20 rises faster than Recall@20.

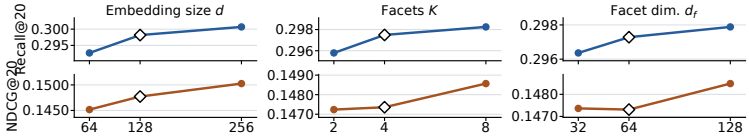


Figure 3: Impact of hyper-parameters d, K, d_f on Youshu.

5 Conclusion and Limitations

CFDGN keeps MultiCBR's collaborative score and adds a calibrated structural judgment based on contextual roles, bundle structure, and history interests. By correcting weak-margin pairs more accurately and more often favoring the positive bundle, the combined ranking improves over MultiCBR alone. Soft states lack labels, noise, dynamic membership, and cold-start settings remain open.

AI USE STATEMENT

Generative AI tools were used to assist language editing. The authors reviewed and validated all AI-assisted technical content. The authors take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

All three datasets are public. We release our source code, processed datasets at the anonymous repository: `https://anonymous.4open.science/r/CFDGN-4C6A`.

REFERENCES

T. A. Brosh, A. Livne, O. Sar Shalom, et al. BRUCE: Bundle recommendation using contextualized item embeddings. In *Proceedings of RecSys*, pp. 237–245, 2022.

D. Cao, L. Nie, X. He, et al. Embedding factorization models for jointly recommending items and user generated lists. In *Proceedings of SIGIR*, pp. 585–594, 2017.

J. Chang, C. Gao, X. He, et al. Bundle recommendation with graph convolutional networks. In *Proceedings of SIGIR*, pp. 1673–1676, 2020.

J. Chang, C. Gao, X. He, et al. Bundle recommendation and generation with graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 35(3):2326–2340, 2023.

L. Chen, Y. Liu, X. He, et al. Matching user with item set: Collaborative bundle recommendation with deep attention network. In *Proceedings of IJCAI*, pp. 2095–2101, 2019.

Yankai Chen, Xinni Zhang, Henry Peng Zou, Bowei He, Yangning Li, Philip S. Yu, Irwin King, and Xue Liu. Adversarial encoding perturbation and synthesis for set representation auxiliary learning. In *Proceedings of ICLR*, 2026.

Q. Deng, K. Wang, M. Zhao, et al. Personalized bundle recommendation in online games. In *Proceedings of CIKM*, pp. 2381–2388, 2020.

X. Du, K. Qian, Y. Ma, et al. Enhancing item-level bundle representation for bundle recommendation. *ACM Transactions on Recommender Systems*, 2023.

Xiaoying Gao, Jianting Chen, Yunxiao Yang, Zaiyuan Di, and Yang Xiang. A diffusion-driven multi-view mixed contrastive learning framework for bundle recommendation. *Expert Systems with Applications*, 304:130690, 2026. doi: 10.1016/j.eswa.2025.130690.

H. Jeon, J. E. Lee, J. Yun, et al. Cold-start bundle recommendation via popularity-based coalescence and curriculum heating. In *Proceedings of the ACM Web Conference*, pp. 3277–3286, 2024.

Yuanzhong Jiang, Kun Yuan, Jennifer Shang, Xi Chen, and Yuanchun Jiang. From sparsity to precision: Enhancing bundle recommendations via large language model augmentation and multi-graph fusion. *Decision Support Systems*, 209:114732, 2026. doi: 10.1016/j.dss.2026.114732.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proceedings of ICLR*, 2015.

Jaeri Lee and U. Kang. Context-aware sequential bundle recommendation via user-specific representations. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pp. 1468–1476, 2025. doi: 10.1145/3746252.3761276.

C. Li, Y. Lu, W. Wang, et al. Package recommendation with intra- and inter-package attention networks. In *Proceedings of SIGIR*, pp. 595–604, 2021.

Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. Multi-interest network with dynamic routing for recommendation at tmall. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2615–2623, 2019. doi: 10.1145/3357384.3357814.

J. Li. Disentangled contrastive bundle recommendation with conditional diffusion. In *Proceedings of AAAI*, volume 39, pp. 12067–12075, 2025. doi: 10.1609/aaai.v39i11.33314.

Ming Li, Lin Li, Xiaohui Tao, Dong Zhang, and Jimmy Xiangji Huang. Divide-and-conquer: Cold-start bundle recommendation via mixture of diffusion experts. *ACM Transactions on Information Systems*, 2026a. doi: 10.1145/3799250.

X. Li, W. Xi, X. Xing, et al. Auto graph filtering for bundle recommendation. In *Proceedings of ICDM*, pp. 299–308, 2023.

X. Li, Y. Mu, and Y. Lin. MIMOSA: A unified multi-view multi-order contrastive learning framework for bundle recommendation. *Information Processing & Management*, 63(2):104446, 2026b. doi: 10.1016/j.ipm.2025.104446.

Yihang Li, Zhuo Liu, and Wei Wei. EpicCBR: Item-relation-enhanced dual-scenario contrastive learning for cold-start bundle recommendation. In *Proceedings of WSDM*, 2026c. doi: 10.1145/3773966.3777993.

J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991.

Y. Ma, Y. He, A. Zhang, X. Wang, and T.-S. Chua. CrossCBR: Cross-view contrastive learning for bundle recommendation. In *Proceedings of KDD*, pp. 1233–1241, 2022.

Y. Ma, Y. He, X. Wang, et al. MultiCBR: Multi-view contrastive learning for bundle recommendation. *ACM Transactions on Information Systems*, 42(4):1–23, 2024. Article 100.

H.-S. Nguyen, T.-N. Bui, L.-H. Nguyen, H. Hoang, C.-V. T. Nguyen, H.-Q. Le, and D.-T. Le. Bundle recommendation with item-level causation-enhanced multi-view learning. In *Machine Learning and Knowledge Discovery in Databases: ECML PKDD 2024*, volume 14948 of *LNCS*, pp. 324–341. Springer, 2024. doi: 10.1007/978-3-031-70371-3_19.

Huy-Son Nguyen, Quang-Huy Nguyen, Duc-Hoang Pham, Duc-Trong Le, Hoang-Quynh Le, Padipat Sitkrongwong, Atsuhiro Takasu, and Masoud Mansoury. RaMen: Multi-strategy multi-modal learning for bundle construction. In *Proceedings of ECAI*, pp. 1889–1896, 2025. doi: 10.3233/FAIA251022.

Y. Ren, H. Zhang, L. Fu, et al. Distillation-enhanced graph masked autoencoders for bundle recommendation. In *Proceedings of SIGIR*, pp. 1660–1669, 2023.

M. Sun, L. Li, M. Li, X. Tao, D. Zhang, Q. Xie, P. Wang, and J. X. Huang. A survey on bundle recommendation: Methods, applications, and challenges. *ACM Computing Surveys*, 58(11):1–42, 2026. doi: 10.1145/3802820. Article 292.

Z. Sun, J. Yang, K. Feng, et al. Revisiting bundle recommendation: Datasets, tasks, challenges and opportunities for intent-aware product bundling. In *Proceedings of SIGIR*, pp. 2900–2911, 2022.

Q. Tao, C. Liu, Y. Xia, Y. Xu, and L. Li. Adaptive multi-graph contrastive learning for bundle recommendation. *Neural Networks*, 181:106832, 2025. doi: 10.1016/j.neunet.2024.106832.

Teng Tu, Ai Li, Yunshan Ma, Shuo Xu, Xiaohao Liu, Haokai Ma, Liang Pang, and Tat-Seng Chua. Discrete diffusion for bundle construction. In *Proceedings of ICLR*, 2026.

Y. Wei, X. Liu, Y. Ma, et al. Strategy-aware bundle recommender system. In *Proceedings of SIGIR*, pp. 1198–1207, 2023.

C. Wu, H. Yuan, P. Zhao, et al. Dual-supervised contrastive learning for bundle recommendation. *IEEE Transactions on Computational Social Systems*, 11(3):3955–3965, 2024.

D. Zhang, L. Li, M. Li, A. Bhuiyan, M. Sun, X. Tao, and J. X. Huang. Modeling item-level dynamic variability with residual diffusion for bundle recommendation. In *Proceedings of AAAI*, volume 40, pp. 16253–16261, 2026a. doi: 10.1609/aaai.v40i19.38662.

Peng Zhang, Zhendong Niu, Ru Ma, Shunpan Liang, and Fuzhi Zhang. Cross-view contrastive representation learning on meta-path induced graphs with node features for bundle recommendation. *Neural Networks*, 199:108669, 2026b. doi: 10.1016/j.neunet.2026.108669.

S. Zhao, W. Wei, D. Zou, et al. Multi-view intent disentangle graph networks for bundle recommendation. In *Proceedings of AAAI*, volume 36, pp. 4379–4387, 2022.

Guorui Zhou, Xiaoqiang Zhu, Chengru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1059–1068, 2018. doi: 10.1145/3219819.3219823.

D. Zou, S. Zhao, W. Wei, et al. Towards hierarchical intent disentanglement for bundle recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 36(7):3556–3567, 2024.

A Complete Implementation and Reproducibility Details

A.1 Full experimental settings

All experiments use a batch size of 2,048, one negative sample per positive pair, Adam with a learning rate of 10^{-3}, and an ℓ_2 coefficient of 10^{-5}. The backbone has two propagation layers and embedding dimension $d = 128$. The structural branch uses $K = 4$ facets with $d_f = 64$. We apply noise augmentation during training.

Dataset-specific perturbation rates and view-fusion weights are listed in the order $(\mathrm{UB}, \mathrm{UI}, \mathrm{BI})$. For Youshu, the perturbation rates are $(0.20, 0.10, 0.10)$ and the fusion weights are $(0.35, 0.15, 0.50)$. It uses $\lambda_{\mathrm{CL}} = 0.04$, temperature 0.23, and 300 epochs. For NetEase, the perturbation rates are $(0.00, 0.00, 0.20)$ and the fusion weights are $(0.50, 0.20, 0.30)$. It uses $\lambda_{\mathrm{CL}} = 0.05$, temperature 0.20, and 100 epochs. For iFashion, the perturbation rates are $(0.05, 0.00, 0.15)$ and the fusion weights are $(0.10, 0.20, 0.70)$. It uses $\lambda_{\mathrm{CL}} = 0.10$, temperature 0.20, and 100 epochs.

Layer-0/1/2 aggregation coefficients are reported for the UB, UI, and BI views. For Youshu, they are UB $(0.70, 0.20, 0.10)$, UI $(0.60, 0.20, 0.20)$, and BI $(0.60, 0.10, 0.30)$. For NetEase, they are UB $(0.35, 0.15, 0.50)$, UI $(0.25, 0.65, 0.10)$, and BI $(0.40, 0.40, 0.20)$. For iFashion, they are UB $(0.50, 0.50, 0.00)$, UI $(0.30, 0.20, 0.50)$, and BI $(0.50, 0.50, 0.00)$.

Common structural settings are $\tau_{\mathrm{role}} = 0.5$, $\tau_{\mathrm{type}} = 0.25$, $\tau_{\mathrm{hist}} = 0.5$, $\tau_{\mathrm{match}} = 0.5$, and $\tau_{\mathrm{PCL}} = 0.2$. PCL label smoothing is 0.05. PPMI retains 20 neighbors with a minimum co-occurrence of two. The enrichment weight and temperature are 0.1 and 0.5, and the router weights are $0.40/0.45/0.15$ for intrinsic, plain-context, and PPMI-context sources. The relation dimension is 64, the relation bank size is 65,536, and the relation batch size is 2,048. We set $\lambda_1 = 0.05$, $\lambda_2 = 0.1$, $\alpha_{\mathrm{init}} = 0.1$, and $\alpha_{\max} = 0.5$. Item/Bundle/User PCL and Orth weights are 0.001 on NetEase and iFashion, and 6.25×10^{-5} on Youshu.

For each run, checkpoint selection maximizes

$$J_{\mathrm{val}} = \mathrm{Recall@20} + \mathrm{NDCG@20} + \mathrm{Recall@40} + \mathrm{NDCG@40} \tag{23}$$

on the validation split, the validation @10 sum breaks ties. The selected checkpoint is evaluated once on the test split. Main-table values average seeds 2024, 2025, and 2026. The ablation and hyperparameter results use seed 2024.

A.2 Sparse PPMI construction and relation supervision

We set $m_B^{\mathrm{tr}} = 1$ only when bundle B appears in the training UB graph. The co-occurrence matrix is

$$Z^{\mathrm{tr}} = \mathrm{Diag}(m^{\mathrm{tr}})Z, \qquad C = (Z^{\mathrm{tr}})^\top Z^{\mathrm{tr}}, \qquad C_{ii} = 0. \tag{24}$$

The positive pointwise mutual information prior is

$$\mathrm{PPMI}(i,j) = \max\left\{\log \frac{p(i,j)}{p(i)p(j)}, 0\right\}, \tag{25}$$

where $p(i,j)$ and $p(i)$ are empirical co-occurrence and item frequencies computed from C. We keep an item pair only when $C_{ij} \geq 2$ and its PPMI is positive. We keep the 20 largest values in each row, symmetrize by taking the elementwise maximum, and remove self-edges. This gives A^{PPMI} in Eq. (3). Local role routing uses only positive-PPMI neighbors already present in $B \setminus \{i\}$, it masks this source when no such co-member exists.

For a bundle with $n \geq 2$ members, let L_B be the number of unordered positive-PPMI member pairs and let $\mathcal{E}_B^P$ be their set. The PPMI contribution to compatibility is

$$\mathrm{Cov}_B^P = \frac{2L_B}{n(n-1)}, \quad \mathrm{Str}_B^P = \frac{1}{L_B} \sum_{(i,j)\in\mathcal{E}_B^P} (1 - e^{-\mathrm{PPMI}(i,j)}), \quad E_B^P = \mathrm{Cov}_B^P \, \mathrm{Str}_B^P, \tag{26}$$

We set $\mathrm{Str}_B^P = 0$ when $L_B = 0$.

For relation-vector supervision, we treat undirected PPMI edges as positive pairs and item pairs with zero raw co-occurrence as negatives. We do not label pairs observed only once as negative. We keep up to 65,536 pairs per class and use at most 2,048 pairs per class at each training step. The pair banks are equal in size and use deterministic seed 314159. Section 3.5 gives the exact loss.

A.3 SCORE-SCALE CALIBRATION

We denote the centered matrix of current main-branch bundle embeddings as $\mathbf{b}_c$, and set $\mathrm{Cov}_B = |\mathcal{B}|^{-1}\mathbf{b}_c^\top \mathbf{b}_c$. The full-catalog backbone score scale in Eq. (17) is

$$\sigma_{\mathrm{Main}} = \sqrt{\mathbb{E}_{u\in\mathcal{U}}[(\mathbf{e}_u^{\mathrm{main}})^\top \mathrm{Cov}_B\, \mathbf{e}_u^{\mathrm{main}}]}. \tag{27}$$

We refresh it from the non-augmented training graph at the start of each epoch and before validation or test evaluation. The calibration coefficient is $\alpha = \alpha_{\max}\,\mathrm{sigmoid}(\theta_\alpha)$, we initialize it at 0.1 with $\alpha_{\max} = 0.5$.

We define a difficulty weight from the detached normalized main-branch margin:

$$m_{\mathrm{main}} = \frac{\mathrm{sg}(s_{uB^+}^{\mathrm{main}} - s_{uB^-}^{\mathrm{main}})}{\sigma_{\mathrm{Main}}}, \qquad n = \mathrm{sigmoid}(-m_{\mathrm{main}}), \qquad w_{\mathrm{rank}} = 0.1 + 0.9n. \tag{28}$$

We optimize the structural representations and the calibration coefficient through separate gradient paths. Let $m_{\mathrm{raw}} = r_{uB^+}^{\mathrm{raw}} - r_{uB^-}^{\mathrm{raw}}$, $\delta_{uB} = \alpha\,\mathrm{sg}(r_{uB}^{\mathrm{raw}})$, and $m_\delta = \delta_{uB^+} - \delta_{uB^-}$. We define $\mathcal{L}_{\mathrm{raw}} = \mathbb{E}[\mathrm{sg}(w_{\mathrm{rank}})\,\mathrm{softplus}(-m_{\mathrm{raw}})]$, $\mathcal{L}_{\mathrm{cal}} = \mathbb{E}[w_{\mathrm{rank}}\,\mathrm{softplus}(-m_\delta)]$, and $\mathcal{L}_{\mathrm{preserve}} = \mathbb{E}[(1 - w_{\mathrm{rank}})(\mathrm{mean}(\delta^2) + m_\delta^2)/2]$. The residual objective is

$$\mathcal{L}_{\mathrm{res}} = \mathcal{L}_{\mathrm{raw}} + \mathcal{L}_{\mathrm{cal}} + \mathcal{L}_{\mathrm{preserve}}. \tag{29}$$

In the calibration terms, we detach the raw residual. This prevents the scale parameter from changing the structural representations just to suppress the residual.

A.4 LEAVE-ONE-OUT HISTORY AND EVALUATION

For a positive training bundle $B^+ \in \mathcal{H}_u$, CFDGN removes B^+ from the attention statistics used for the history representation. For example,

$$\mathbf{h}_{u,-}^k = \frac{\sum_{B\in\mathcal{H}_u} e^{\ell_{uB}^k}\mathbf{h}_B^k - e^{\ell_{uB^+}^k}\mathbf{h}_{B^+}^k}{\sum_{B\in\mathcal{H}_u} e^{\ell_{uB}^k} - e^{\ell_{uB^+}^k}}. \tag{30}$$

The type and composition values are computed in the same way. If no history remains, we skip the structural ranking update for that example, but backbone learning continues. At inference, we mask training-observed UB pairs and compute Eq. (17) over the full bundle catalog in chunks of 512. Chunking changes memory use but not candidate scores.

We obtain the fixed anchors by QR orthogonalization of K random vectors with component seed 12345. We store them as non-trainable buffers. We keep the component-specific random states separate from the random streams used for data loading and graph augmentation. These deterministic component seeds are distinct from experiment seeds 2024-2026.

B AUXILIARY STRUCTURAL DETAILS

The main text defines the role router, bundle vector, history values, selector, and four matching paths. This section gives auxiliary normalization, history, and facet-constraint details. The role diversity in Eq. (7) uses generalized Jensen-Shannon dispersion normalized by $\log\min(|B|, K)$.

B.1 FACET AND MEMBER PROJECTIONS

The item-facet norms and stabilized directions used by the role router are

$$a_i^k = \|\mathbf{r}_i^k\|_2, \qquad \mathbf{z}_i^k = \frac{\mathbf{r}_i^k}{a_i^k + \varepsilon}, \tag{31}$$

where $\varepsilon > 0$ prevents division by zero and a_i collects the K norms.

The member weights in Eq. (6) and the semantic pair vectors in Eq. (8) are

$$\alpha_{iB}^k = \frac{g_{iB}^k}{\sum_{j\in B} g_{jB}^k + \varepsilon}, \qquad \mathbf{v}_i = \mathrm{Norm}(W_v\mathbf{t}_i), \tag{32}$$

where W_v is a learned projection and the stabilized denominator applies only within candidate bundle B.

B.2 PPMI FACET MESSAGE AND GATE

Equation (2) defines the same-index PPMI message and similarity gate from the row-normalized prior $\bar{A}$. The two normalizers used in Section 3.2 are $\mathrm{Norm}(x) = x/(\|x\|_2 + \varepsilon)$ and $\mathcal{N}(x) = (x - \mathrm{mean}(x))/\sqrt{\mathrm{mean}[(x - \mathrm{mean}(x))^2] + \varepsilon}$, with componentwise means. Global PPMI neighbors can influence a member's enriched facets, but local routing and bundle aggregation still use only actual candidate members.

The compatibility scale in Eq. (8) is $\gamma = \mathrm{softplus}(\gamma_{\mathrm{raw}}) + 10^{-4}$, which keeps the semantic score monotonic in pairwise similarity. For a singleton, the masked pair calculation yields an unused provisional $C_B^{\mathrm{sem}} = \mathrm{sigmoid}(b)$ before the final $C_B = 0.5$ is assigned.

B.3 HISTORY SUMMARIES AND RELIABILITY

The candidate-independent query and the candidate-conditioned mixture vector are

$$\mathbf{q}_u^k = \mathrm{Norm}(W_q^k \,\mathrm{sg}(\mathbf{e}_u^{UI})), \qquad \mathbf{z}_{uB} = \sum_k \omega_{uB}^k \,\mathrm{Norm}(\mathbf{h}_u^k), \tag{33}$$

where W_q^k is a learned interest-specific projection, $\mathbf{e}_u^{UI}$ is the UI-view user embedding, and only the mixture weights ω_{uB}^k depend on candidate B.

For each interest k, the attention weights in Eq. (11) also aggregate historical states and role compositions:

$$\boldsymbol{\pi}_{u,\mathrm{type}}^k = \sum_{B_h \in \mathcal{H}_u} \beta_{uB_h}^k \mathbf{p}_{B_h}, \qquad \boldsymbol{\pi}_{u,\rho}^k = \sum_{B_h \in \mathcal{H}_u} \beta_{uB_h}^k \rho_{B_h}. \tag{34}$$

For each interest, the effective history size is $N_{\mathrm{eff},u}^k = 1/\sum_{B_h \in \mathcal{H}_u} (\beta_{uB_h}^k)^2$, and its reliability is $R_u^k = N_{\mathrm{eff},u}^k/(N_{\mathrm{eff},u}^k + K)$. Given the candidate match $c_{uB}^k = \cos(\boldsymbol{\pi}_{u,\rho}^k, \rho_B)$, we shrink it toward 0.5:

$$\widetilde{c}_{uB}^k = 0.5 + R_u^k (c_{uB}^k - 0.5). \tag{35}$$

B.4 FACET ALIGNMENT AND ORTHOGONALITY

Fixed-anchor prototype contrastive learning (PCL) aligns facet index k with the same orthogonal prototype $\mathbf{p}_k$ at item, bundle, and user levels. This prototype is non-trainable:

$$\mathcal{L}_{\mathrm{PCL}} = \mathrm{CE}\left(\left\{ \frac{\cos(\mathbf{z}^k, \mathbf{p}_q)}{\tau_{\mathrm{PCL}}} \right\}_{q=1}^{K}, \text{target } k \right). \tag{36}$$

Let Q stack an entity's K unit facets and let $G = QQ^\top$. Orthogonality regularization penalizes redundant facets within each entity:

$$\mathcal{L}_{\mathrm{Orth}} = \frac{\sum_{k \neq l} G_{kl}^2}{K(K-1)}. \tag{37}$$

We apply PCL and Orth at the item, bundle, and user levels. Their weights are reported in Appendix A.1.

C COMPLETE STRUCTURAL MECHANISM ANALYSIS

C.1 CONTEXTUAL ITEM ROLES AND SECONDARY FACETS

If the router behaved as a static item classifier, the same item would retain a nearly unchanged g_{iB} across bundles. We therefore measure several quantities for items appearing in multiple bundles: dominant-role switches, pairwise switches, secondary-facet promotion, attention lift, and cross-context total variation (TV).

Table 6: Item-role diagnostics (lift: ratio; others: proportions).

Dataset	Repeated coverage	Context switch	Pair switch	Secondary active	Secondary promoted	Attn. lift	Context TV
Youshu	0.4631	0.4012	0.2081	0.3569	0.4353	2.7237	0.1110
NetEase	0.9999	0.3436	0.1330	0.4236	0.4811	2.7406	0.0844
iFashion	0.3759	0.4304	0.2859	0.2762	0.4516	2.3408	0.1542

Context-switch rates are 40.12%, 34.36%, and 43.04% on Youshu, NetEase, and iFashion. Secondary-activation rates are 35.69%, 42.36%, and 27.62%. Attention lift for promoted secondary facets exceeds 2.3 on each dataset. Figure 2(a) shows representative role distributions. Facets 0-3 are latent coordinates rather than human-annotated semantic categories.

C.2 PPMI OPPORTUNITY AND USE

The four PPMI injection points share the same prior, which is computed from bundles observed during training. Only the local role source requires a positive-PPMI co-member already present in the candidate. Table 7 reports raw repeated-co-occurrence support and compares it with the activation and role effects inside CFDGN.

Table 7: Repeated-relation opportunities and PPMI use.

Dataset	Repeated-item coverage	Pair ≥ 2 support	PPMI-active context	PPMI role TV	Dominant-role flip
Youshu	0.4631	0.3702	0.6973	0.0546	0.0567
NetEase	0.9999	0.5767	0.9457	0.0579	0.0572
iFashion	0.3759	0.0537	0.1611	0.0543	0.0524

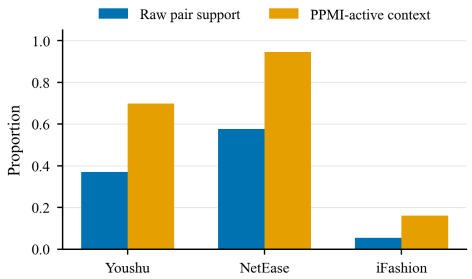


Figure 4: Repeated-pair support and PPMI-active contexts.

Raw support and model activation differ numerically because of positive PMI, top-20 pruning, symmetrization, and local availability. Both quantities nevertheless rank the datasets in the same order: iFashion $<$ Youshu $<$ NetEase. PPMI changes the dominant contextual role for about five percent of audited cases on each dataset. The ranking ablation is nearly tied. Together, these results suggest a limited but measurable structural effect, rather than a stable direct accuracy gain.

C.3 MEMBER-DERIVED BUNDLE STRUCTURE

Table 5 in RQ3 reports dominant soft-state shares and mean structural features. Across all datasets, Complementary has higher mean diversity than Similar. Noise has lower mean compatibility than the two compatible states. These relations follow the monotonic construction in Eq. (9), they are not classification accuracy against external labels. Noise energy also contains D_B. As a result, low-diversity and low-compatibility bundles have low energy for all three states and tend toward a less decisive distribution. We treat this region as ambiguous, rather than relabeling earlier results under a changed formula.

The frozen-forward counterfactuals disable local context, PPMI, or secondary facets without retraining. Their Youshu bundle-state flip rates are 6.4%, 2.1%, and 42.4%, respectively (Figure 5). They inspect the learned mechanism and are distinct from the retrained ablations in Table 3.

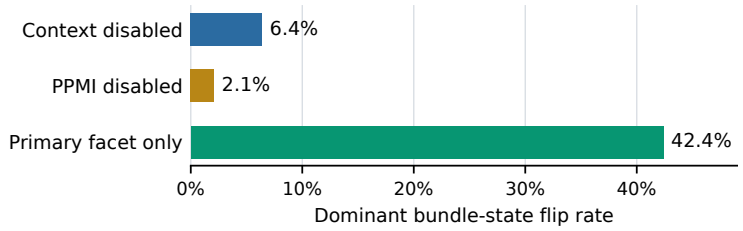


Figure 5: Youshu bundle-state flips under frozen-model interventions (no retraining).

C.4 Candidate-conditioned interest selection

Table 8 reports five diagnostics: dominant-interest switching, pairwise switching, divergence, selector concentration, and history reliability. Candidate switch is the fraction of users with at least two training bundles that show any dominant-interest change across all such bundles. Pairwise switch first evaluates every unordered bundle pair for each eligible user and then averages the within-user rates.

Table 8: Candidate-conditioned interest-selection diagnostics.

(a) Switching and divergence

Dataset	Eligible users	Candidate switch	Pairwise switch	JSD	TV
Youshu	2,573	0.1869	0.0356	0.0099	0.0576
NetEase	18,475	0.8735	0.4647	0.0344	0.1119
iFashion	53,897	0.9436	0.4054	0.0260	0.1007

(b) Concentration and reliability

Dataset	Selector entropy	Top-1 weight	History reliability
Youshu	0.8388	0.5077	0.5742
NetEase	0.8875	0.4226	0.6565
iFashion	0.9249	0.4022	0.8189

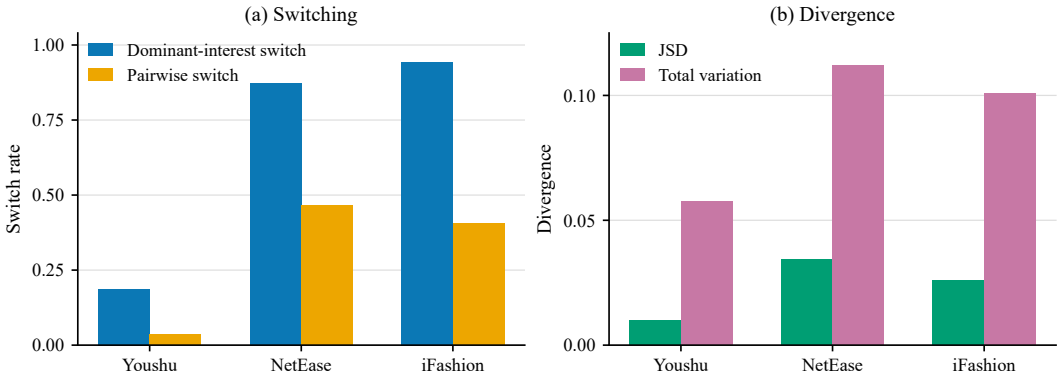


Figure 6: Interest selection: switch rates (left) and distribution distances (right).

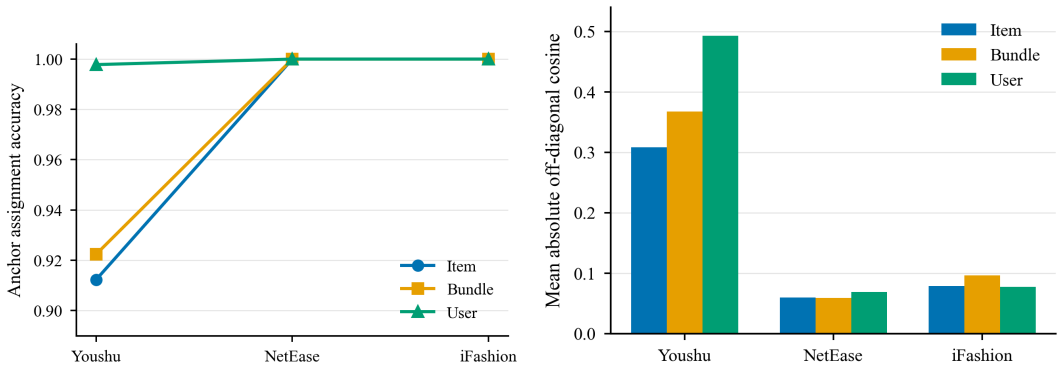
Dominant-interest switch rates are 18.69% on Youshu, 87.35% on NetEase, and 94.36% on iFashion. Youshu also has fewer users with at least two training bundles: 2,573, compared with 18,475 on NetEase and 53,897 on iFashion. Figure 2(b) illustrates how latent-interest selection changes across candidates. It does not assign external semantics to individual interest coordinates.

C.5 FACET GEOMETRY

PCL aligns facet index k with the same fixed anchor across item, bundle, and user levels. Orth reduces similarity among facets within one entity. Table 9 reports both measurements.

Table 9: Fixed-anchor alignment and within-entity orthogonality.

Dataset	Level	Anchor acc.	Anchor margin	Abs. off-diag.	Sq. off-diag.
Youshu	Item	0.9123	0.2915	0.3083	0.1697
	Bundle	0.9224	0.3180	0.3679	0.2050
	User	0.9978	0.3835	0.4929	0.3140
NetEase	Item	1.0000	0.6645	0.0598	0.0059
	Bundle	1.0000	0.7321	0.0586	0.0056
	User	1.0000	0.8137	0.0687	0.0092
iFashion	Item	1.0000	0.6317	0.0783	0.0096
	Bundle	1.0000	0.6644	0.0960	0.0152
	User	1.0000	0.8066	0.0769	0.0124



(a) Fixed-anchor assignment **(b)** Off-diagonal facet cosine

Figure 7: Facet geometry: anchor assignment (left) and off-diagonal cosine (right).

NetEase and iFashion approach 100% anchor assignment accuracy at all three levels. On Youshu, item and bundle accuracy exceed 91%, and user accuracy is 99.78%. Mean absolute off-diagonal cosine is below 0.10 on NetEase and iFashion but higher on Youshu. These statistics support facet separation and index alignment. However, they do not show that every latent coordinate has an external meaning.

C.6 SCORE PATHS

Table 10 reports the positive-minus-negative score gap and pairwise win rate for each structural path and their fusion.

Table 10: Discrimination by structural path and fixed fusion.

Dataset	Metric	Facet	Bundle	Type	Composition	Fused
Youshu	Mean gap	0.4226	0.7908	-0.0008	0.0463	0.3147
	Win rate	0.8250	0.8239	0.4862	0.6784	0.8447
NetEase	Mean gap	0.1852	0.6735	0.0096	0.0995	0.2419
	Win rate	0.7857	0.7854	0.5725	0.6731	0.8068
iFashion	Mean gap	0.1866	0.7754	0.0429	0.1389	0.2860
	Win rate	0.7796	0.8144	0.6876	0.6809	0.8257

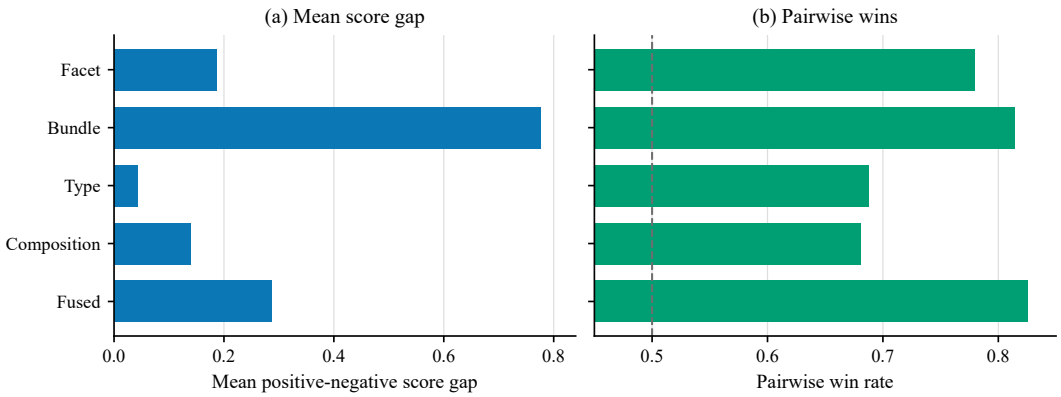


Figure 8: iFashion path gaps (left) and win rates (right); dashed line indicates chance.

Facet, Bundle, and Composition have positive mean gaps on all datasets. Fused win rates are 84.47%, 80.68%, and 82.57% on Youshu, NetEase, and iFashion. The Type path is weakly positive on NetEase and iFashion but slightly negative on Youshu (-0.0008, win rate 48.62%). This shows that a structural diagnostic need not be a strong ranking path on every dataset.

C.7 Independent training-data statistics

Table 11 compares model-based diagnostics with statistics computed directly from training data. Repeated-item coverage matches the membership surface used in the role audit, the count of users with at least two training bundles bounds where candidate switching is assessable, and raw repeated-pair support has the same dataset ordering as local PPMI activation. These statistics contextualize the mechanisms but do not supervise the latent roles or states.

Table 11: Training-data opportunities and mechanism diagnostics.

Dataset	Raw repeated-item coverage	Audit repeated-item coverage	Eligible users	Candidate-interest switch	Pair count at least 2	PPMI-active context
Youshu	0.4631	0.4631	2,573	0.1869	0.3702	0.6973
NetEase	0.9999	0.9999	18,475	0.8735	0.5767	0.9457
iFashion	0.3759	0.3759	53,897	0.9436	0.0537	0.1611