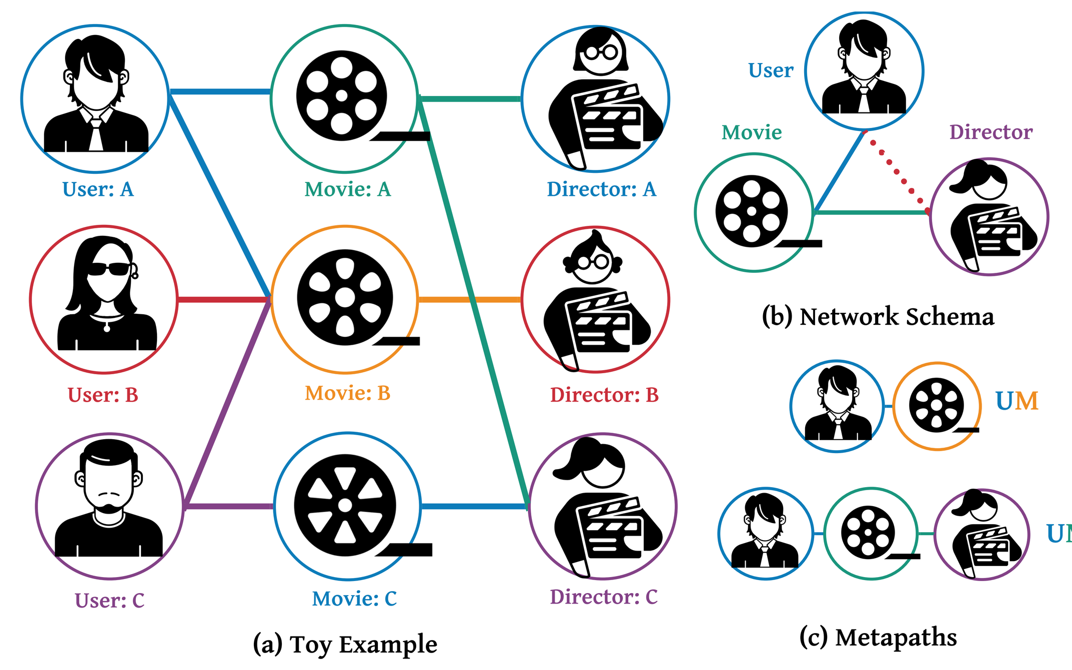
An Efficient Neighborhood-based Interaction Model for Recommendation on Heterogeneous Graph

Jiarui Jin[1], Jiarui Qin[1], Yuchen Fang[1], Kounianhua Du[1], Weinan Zhang[1], Yong Yu[1], Zheng Zhang[2], Alexander J. Smola[2]

[1]Shanghai Jiao Tong University, [2]Amazon Web Services

Background

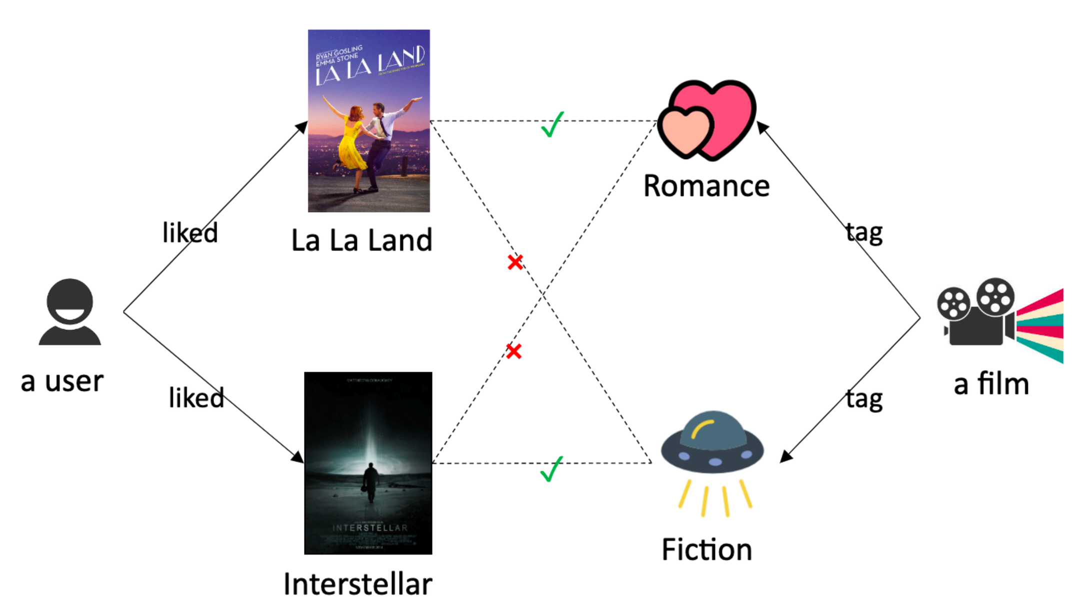
Heterogeneous Graph



(a) Toy Example (b) Network Schema (c) Metapaths

Challenges

- How to tackle the *early summarization issue* (to distinguish useful patterns from noise).



- How to design an end-to-end framework to capture and aggregate the interactive patterns between neighbors.
  - graph-based: low generalization formulation, unable to model high-order interactions.
  - metapath-based: metapaths are heterogeneous and often in different length.
- How to learn the whole system efficiently. (Extending interactions in heterogeneous information network, learning interactive patterns is always time-consuming)

Pros of our model

- address early summarization issue (encode user/item neighbor's information into single embedding vectors before final prediction)
- efficient learning heterogeneous and high-order neighbors
- alleviate data sparsity and cold start problem

Algorithm

Algorithm 1 NIRec

INPUT: HIN $\mathcal{G} = (\mathcal{V}, \mathcal{E})$; node feature $\{e, i \in \mathcal{V}\}$; metapath set $\{\rho_0, \rho_1, \cdots, \rho_{P-1}\}$; source code n_s and target node n_t

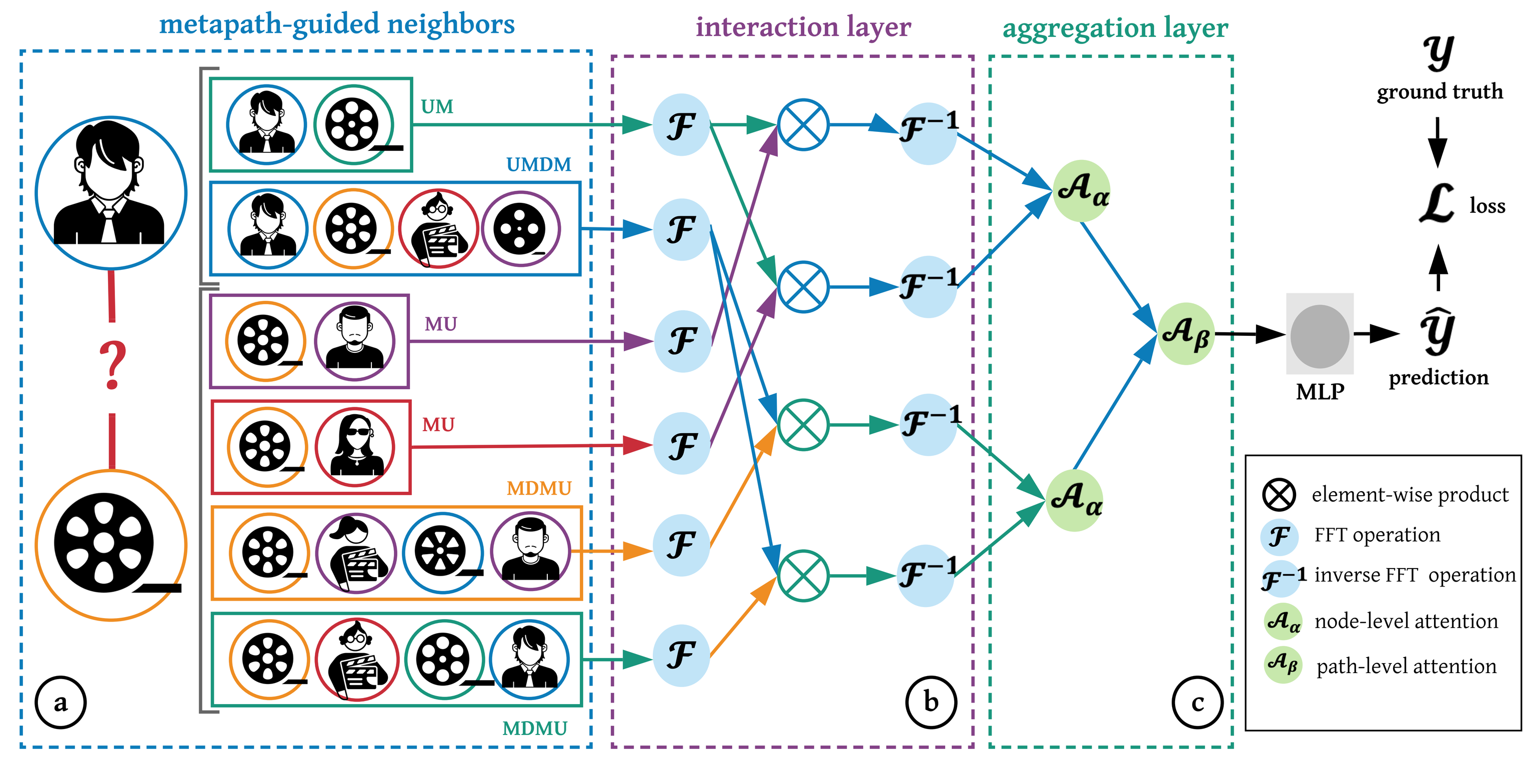
OUTPUT: final link prediction $\hat{Y}$ between n_s and n_t

Initialize all parameters.
repeat
  for each metapath $\rho_k \in \{\rho_0, \rho_1, \cdots \rho_{P-1}\}$ **do**
    Find metapath-guided neighborhoods of n_s, n_t: $\mathcal{N}_{\rho_k}(n_s), \mathcal{N}_{\rho_k}(n_t)$.
    Obtain interaction result $\mathbf{H}[\mathcal{N}_{\rho_k}(n_s), \mathcal{N}_{\rho_k}(n_t)]$.
    Calculate node/element-level embedding z^{ρ_k}.
  end for
  Fuse path/matrix-level embedding Z.
  Obtain final predication $\hat{Y}$ via MLP.
  Calculate loss $\mathcal{L}(Y, \hat{Y})$, and Back propagation.
until convergence

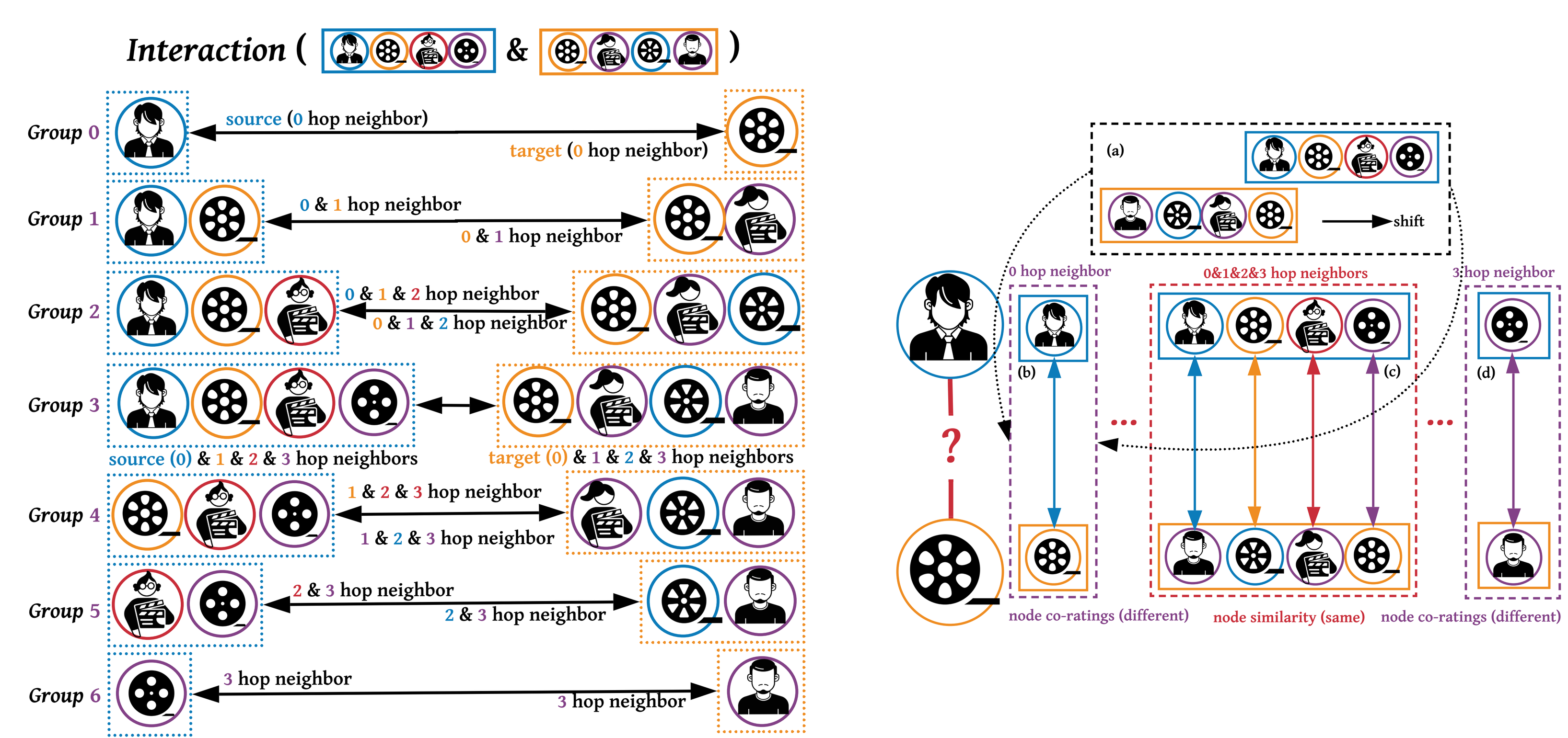
Table 1: The results of CTR prediction in terms of AUC, ACC. *Note*: "*" indicates the statistically significant improvements over the best baseline, with p-value smaller than 10^{-6} in two-sided t-test.

Model	Movielens		LastFM		AMiner		Amazon	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
NeuMF [7]	0.7890	0.7378	0.8900	0.8102	0.8130	0.7897	0.6841	0.6405
HAN [31]	0.8110	0.7530	0.9113	0.8289	0.8451	0.8284	0.7207	0.6831
HetGNN [35]	0.7830	0.7411	0.9020	0.8270	0.8202	0.7939	0.7061	0.6627
LGRec [8]	0.8030	0.7504	0.9127	0.8331	0.8308	0.8130	0.7058	0.6572
MCRec [9]	0.8161	0.7622	0.9274	0.8471	0.8512	0.8339	0.7274	0.6940
IPE [13]	0.8186	0.7693	0.9235	0.8440	0.8411	0.8209	0.7173	0.6789
$\text{NIRec}_{\text{CNN}}$	0.8342	0.7777	0.9353	0.8593	0.8734	0.8504	0.7397	0.7060
$\text{NIRec}_{\text{GCN}}$	0.8290	0.7630	0.9290	0.8571	0.8636	0.8475	0.7379	0.7042
NIRec	**0.8468***	**0.7896***	**0.9404***	**0.8665***	**0.8760***	**0.8562***	**0.7493***	**0.7110***

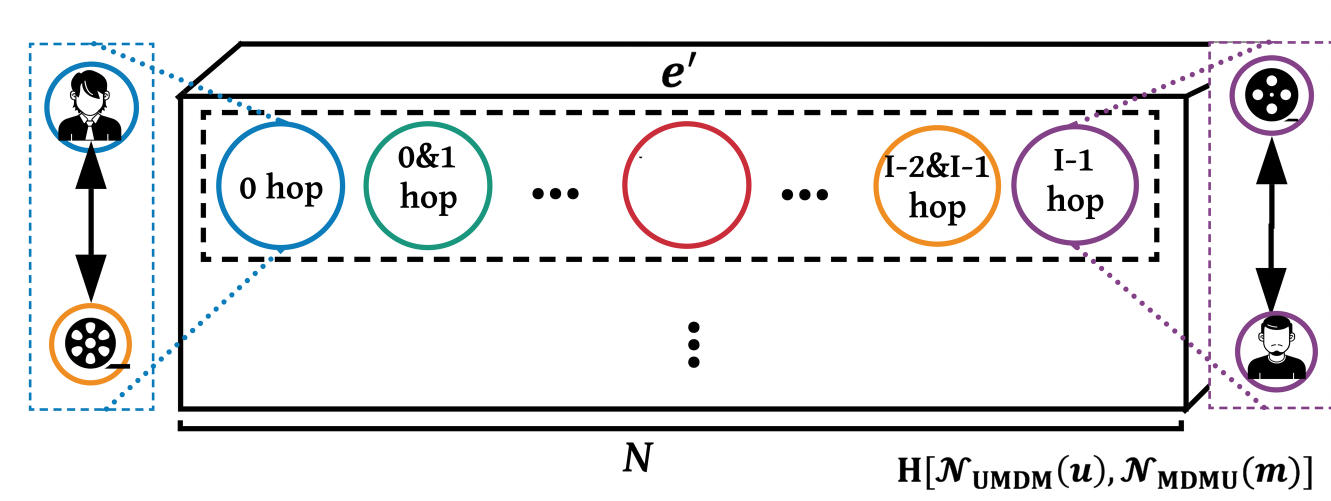
Our Model (NIRec)



- (a) first samples *metapath-guided neighborhood;*
  - $\mathcal{N}_\rho(o)$ is defined as the set of all visited objects when the object o walks along the given metapath ρ;
  - e.g., $\mathcal{N}_\rho^1(u_A) = \{(u_A, m_A), (u_A, m_B)\}$, $\mathcal{N}_\rho(u_A) = \mathcal{N}_\rho^2(u_A) = \{(u_A, m_A, d_A(u_A, m_A, d_C), (u_B, m_B, d_B)\}$ (metapath: UMD)



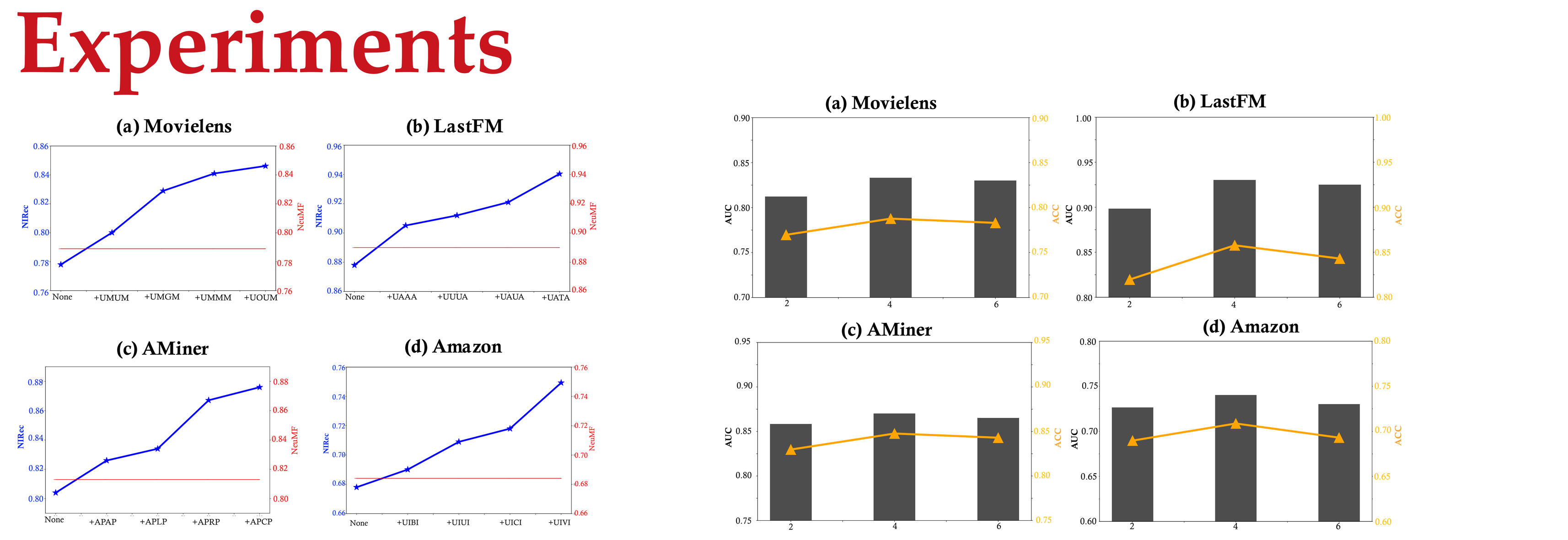
- (b) next constructs *interactive information via interaction layer;*
  - neighborhood is grouped according to the distance to the source/target node. Interaction is only employed between corresponding neighborhoods, which can be formulated as convolution operation in signal process field.
- (c) finally combines rich information through *aggregation layer*

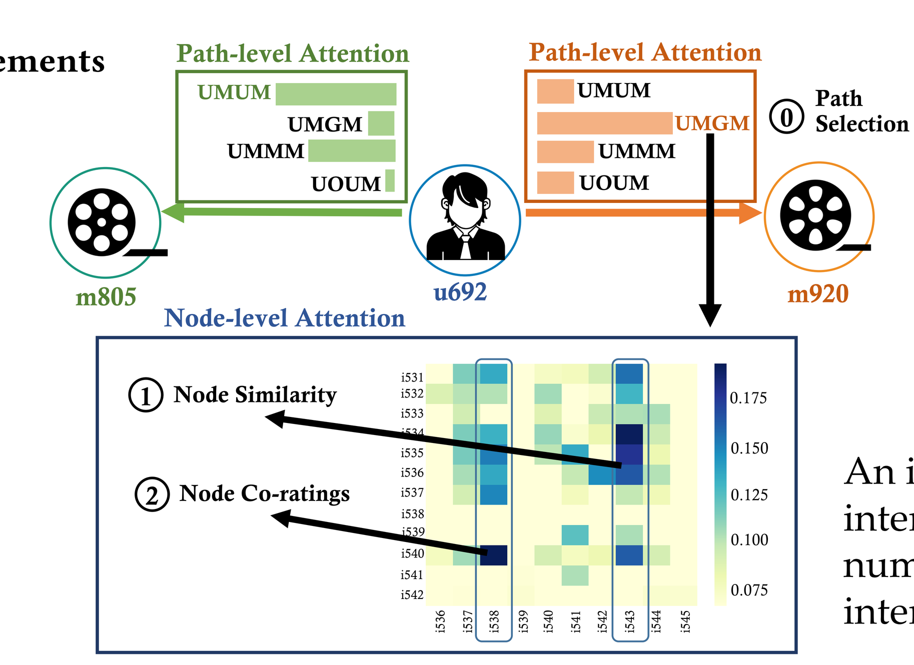


$$Z = \sum_{j=0}^{P-1} \beta^{\rho_j} \cdot Z[\mathcal{N}_{\rho_j}] \longleftarrow z_i^\rho = \sigma(W_q \cdot \left(\frac{1}{H}\sum_{n=1}^{H}\sum_{j \in \mathcal{N}_\rho} \alpha_{ijn}^\rho \left(h_j^\rho W_C^n\right)\right) + b_q)$$

path/matrix-level attention node/element-level attention

Experiments





Full Paper Link

Jiarui Jin's WeChat

An illustrated example of the interpretability of interaction-specific attention distributions. The number denotes the logic flow of interpretation.

If you have any question, please feel free to contact Jiarui Jin (jinjiarui97@gmail.com)