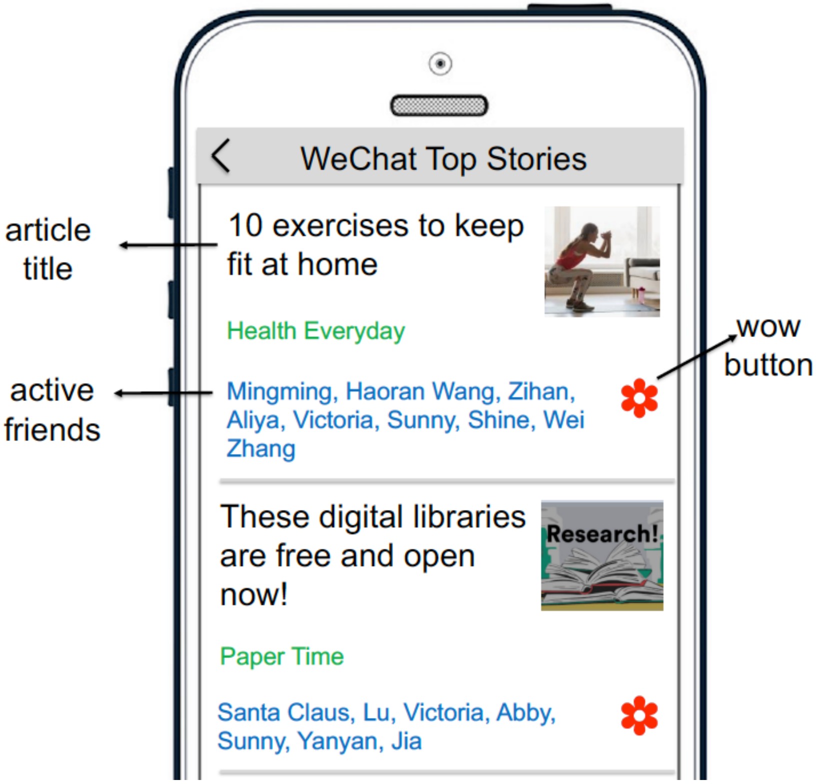
Understanding WeChat User Preferences and "Wow" Diffusion

Fanjin Zhang, Jie Tang, Xueyi Liu, Zhenyu Hou, Yuxiao Dong, Jing Zhang, Xiao Liu, Ruobing Xie, Kai Zhuang, Xu Zhang, Leyu Lin and Philip S. Yu

清华大学–腾讯联合实验室
TSINGHUA UNIVERSITY- TENCENT JOINT LABORATORY

WeChat Top Stories

- Top stories suggests contents based on users' personal preferences and their friends' sharing (referred to as "Wow").
- The user can choose to "Click" (view) the recommended article or follow to "Wow" the article.



The "Wow" feature acts as a diffusion process.

Motivation

- How can user attributes, the correlations between users, and the local network structure influence user behavior?
- What are the differences between various kinds of user feedback (such as “click”, “like”, and “share”) w.r.t. the above factors?
- To what extent users’ behavior can be predicted from their social connections and attributes.

Top Stories Dataset

- A social network $G = \{U, E\}$, where U is a set of users, and E represents edge set
- User attributes C including users' gender, age, regions, and so on
- The interaction between users and articles $L = \{(u, d, ts, is\_like, is\_click, af(u, d, ts)) | u \in U, d \in D\}$ where u is the ego user, d is a displayed article in article set D, $af(\cdot)$ refers to active "Wow"ed friends.
- 48,084,772 users, 61,252,317 articles, and 7,459,660,092 interactions.

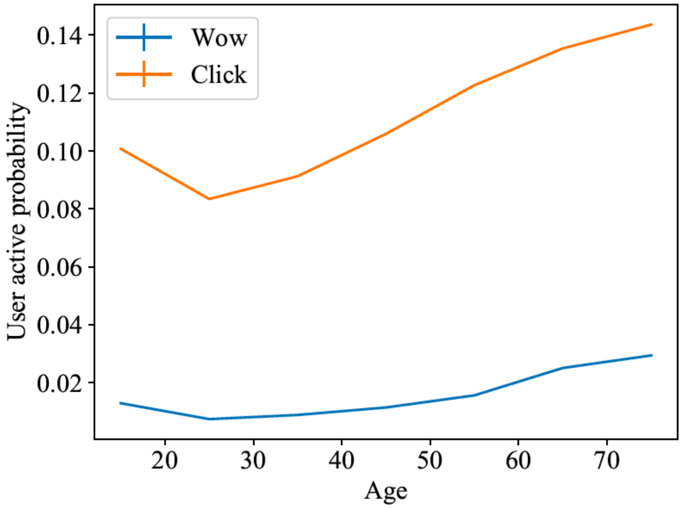
Analysis: User Demographics

- Males tend to consume content.
- Females are more active in social circles.

TABLE 1
User activity w.r.t. gender.

Gender	"Wow" prob.	Click prob.
Male	1.17%	**10.62%**
Female	**1.19%**	9.86%

- The "wow" and click probability of the 20s is the lowest among all ages.



Analysis: User Demographics

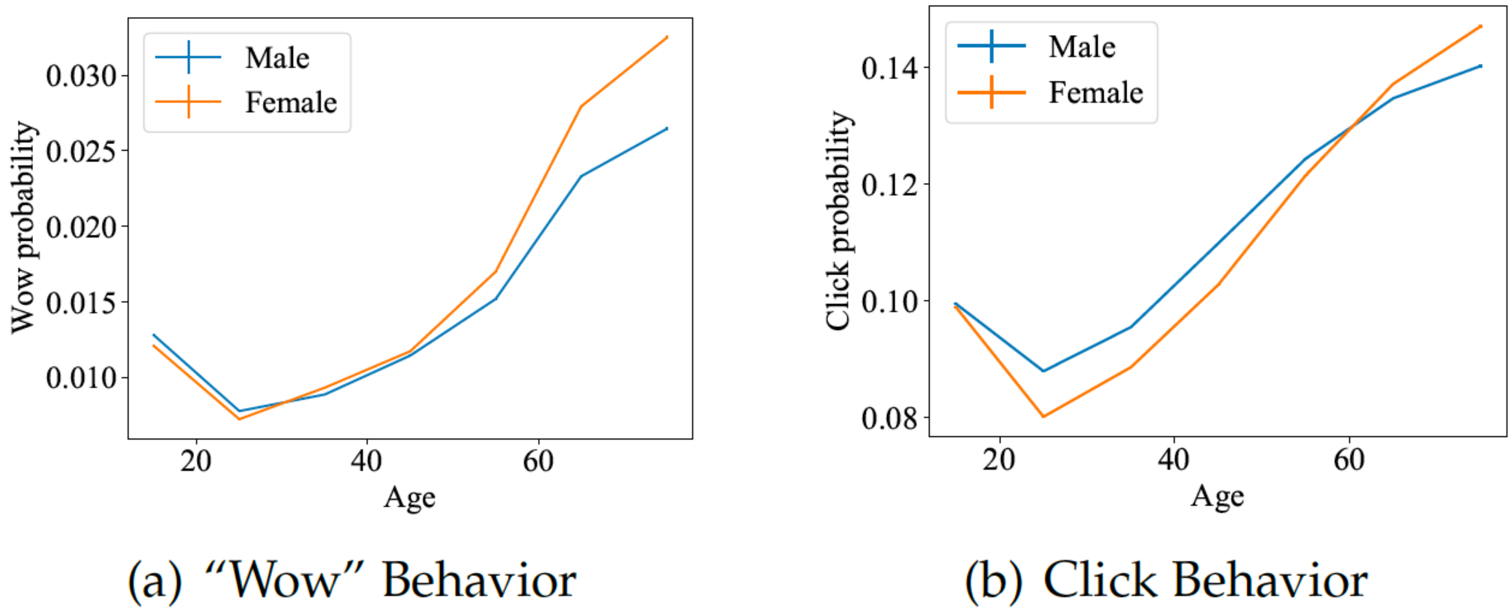


Fig. 3. User "wow" and click probability w.r.t. users' gender and age.

- The cross-attribute factor is more complicated.

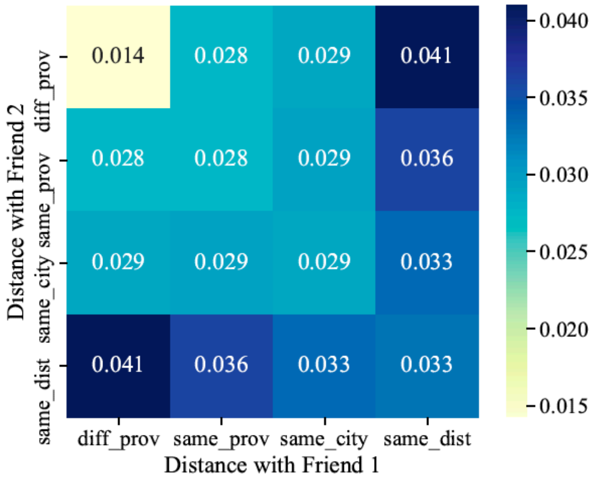
Analysis: Dyadic and Triadic Correlations

TABLE 3
Dyadic correlations w.r.t. the distance between the user and the friend.

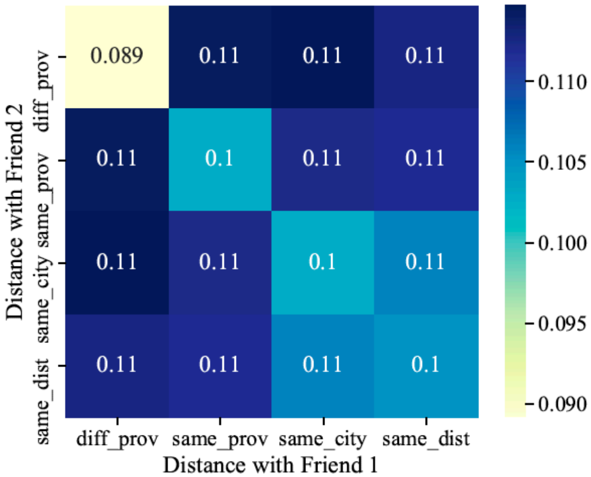
User	"Wow" prob.	Click prob.
All	1.01%	10.24%
Same province	1.05%	10.65%
Same city	1.08%	10.85%
Same district	**1.19%**	**11.27%**

- When the geographic distance between the ego user and the friend is closer, the "wow" probability and click probability of the ego user is higher.
- Interest homophily exists w.r.t. user region.

Analysis: Dyadic and Triadic Correlations



(a) "Wow" Behavior



(b) Click Behavior

- "Attribute diversity" may provide evidence that the "wow"ed articles are acknowledged by various users.

Analysis: Ego Network Properties

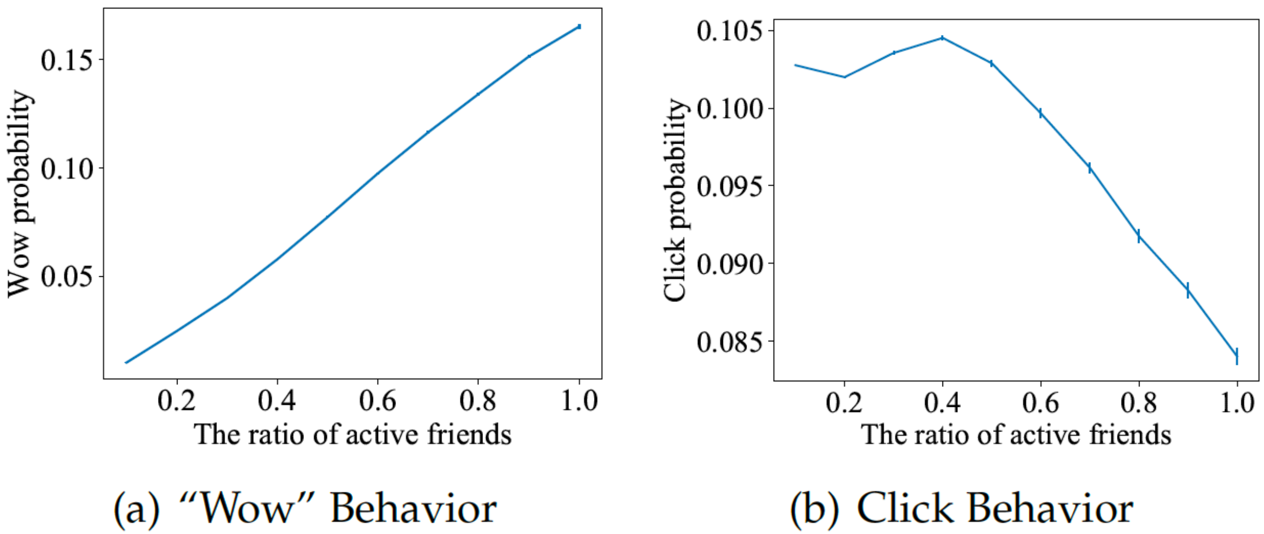
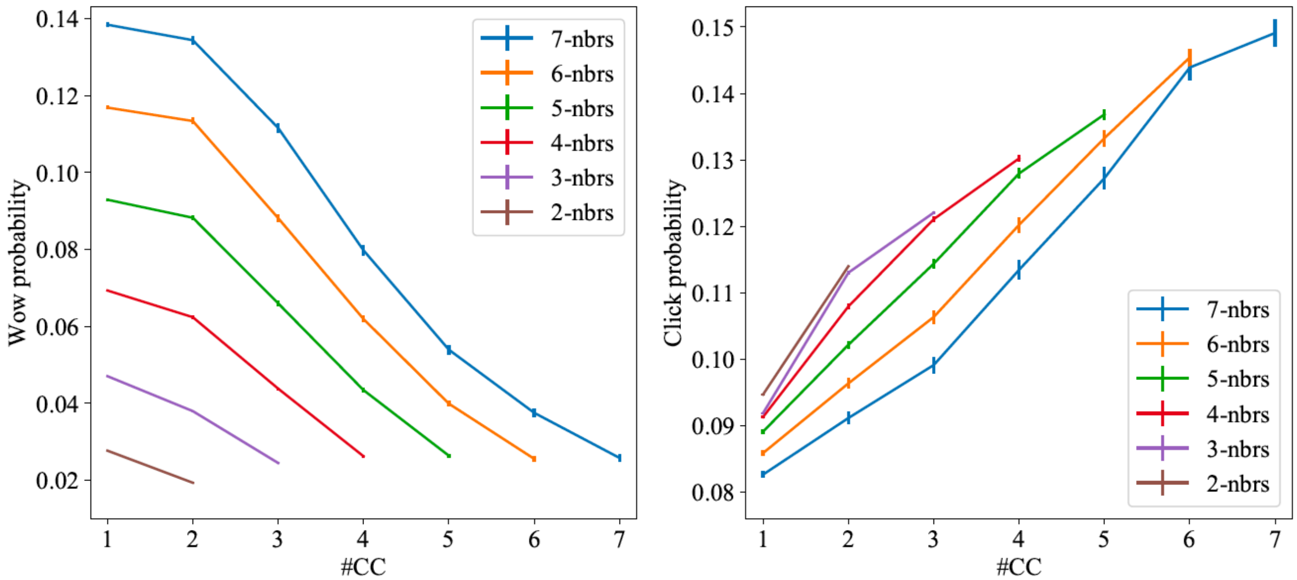


Fig. 8. User "wow" and click probability vs. the ratio of active friends

- Very different patterns
- "Wow": conformity; Click: information overload

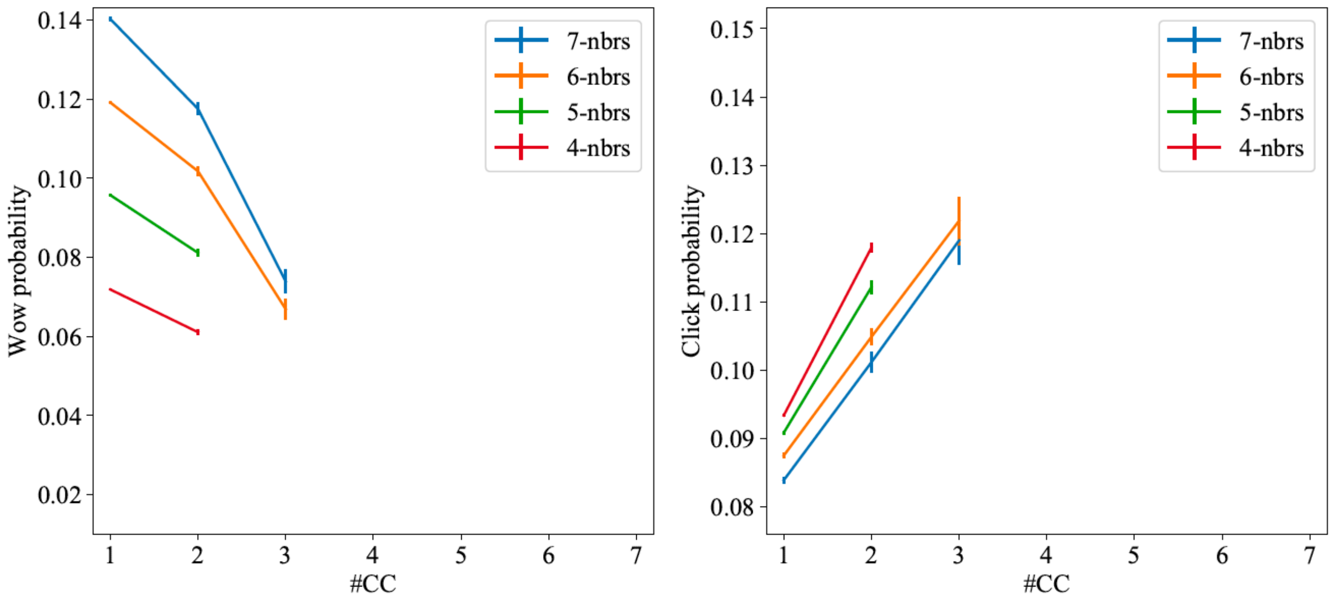
Analysis: Ego Network Properties



(a) Users' Active Rate v.s. #CC in the Ego network formed by active friends

- The number of connected components (#CC) — structural diversity

Analysis: Ego Network Properties



(b) Users' Active Rate v.s. #CC in the 1-core subgraph of the ego network formed by active friends

- The structural topology of cleaned ego networks probably gives a better discriminative ability to predict ego users' activity.

Analysis Summary

- Males are more likely to click but less likely to “wow” articles than females. Counterintuitively, the young generations (people in their 20s and 30s) have the lowest active rate in Top Stories.
- For dyadic or triadic correlations,
  - there exists interest homophily between users and friends (such as about gender and region),
  - but attribute diversity (such as region) also positively correlates with users’ activity when there is more than one active friend.

Analysis Summary

- According to ego network topology, the patterns of “wow” and click behavior are very different.
  - For instance, when fixing the number of active friends, users’ “wow” probability is negatively correlated to #CC formed by active friends, but for click behavior, it is the opposite.
  - The patterns can be more significant when the ego network is cleaned.

User Behavior Prediction

- Input
  - $G_u^\tau = \{V_u^\tau, E_u^\tau\}$ is u's τ-ego network which is a subgraph induced by u and u's τ-degree friends. V_u^τ is the node set of subgraph G_u^τ and E_u^τ is the edge set of G_u^τ.
  - C_u^τ is the attribute matrix of users in V_u^τ.
  - $S_{(u,d,ts)} = \left\{s_{(v,d,ts)} \in \{0,1\} \,\middle|\, v \in V_u^\tau \backslash \{u\}\right\}$, where $s_{(v,d,ts)}$ is the action status of user v w.r.t. article d before timestamp ts.

User Behavior Prediction

- Goal
  - quantify the "wow" and click probability of ego user u after timestamp ts:

$$P(s_{(u,d,>ts)}|G_u^{\tau}, S_{(u,d,ts)}, C_u^{\tau})$$

- Learn independent models for predicting two behaviors
  - $P(is\_click_u = 1) \approx P(is\_click_u = 1 | is\_wow_u = 1)$

Ideas from Analysis

- To model cross-attribute factors for users' different attributes, we adopt the factorization machine technique to generate second-order features to model feature interactions for each individual.
- To remove noise in the ego networks, our model propagates initial user features in the modulated spectral domain, to generate user embeddings based on cleaned ego networks.

Ideas from Analysis

- To model dyadic correlations, we adopt a new graph attention mechanism to model feature interactions between neighbors.
- To model the connected components — the hierarchical structure of the ego networks, we generate hierarchical representations of ego networks by clustering nodes together and learning on the coarsened graphs iteratively.

Model Framework — DiffuseGNN

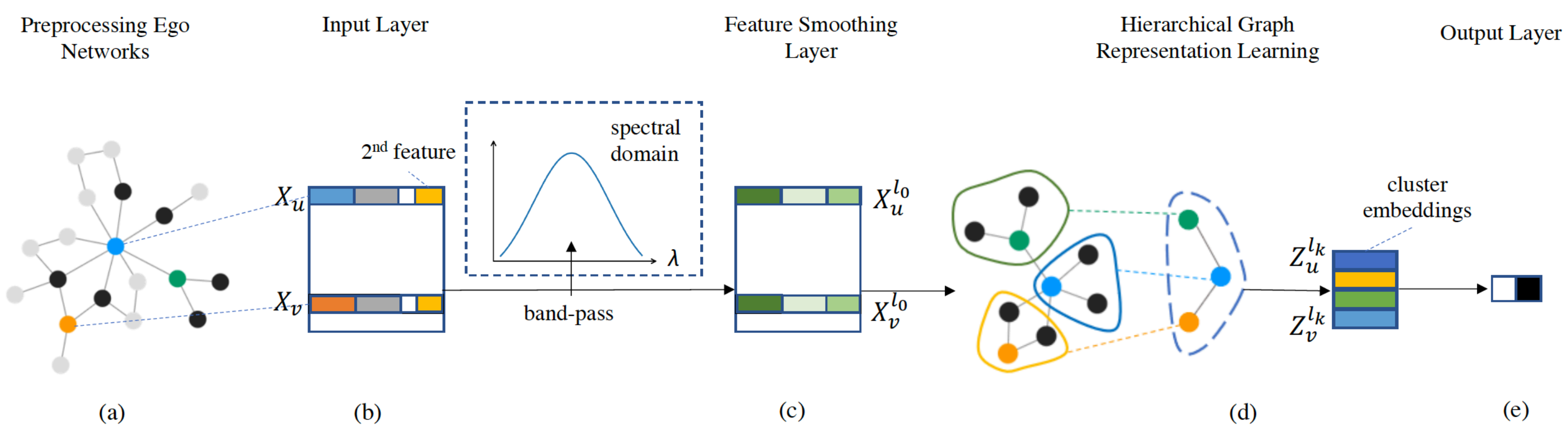


Fig. 10. Model Framework: (a) A sketch map of the processed ego network; (b) The input layer, in which each user's pretrained embedding is concatenated with handcrafted features and her influence feature, which indicates her active status and whether she is an ego user; (c) The feature smoothing layer, in which each user's concatenated input features are filtered by a band-pass filter in the spectral domain; (d) Each ego network's features are passed into a hierarchical graph representation learning model; (e) The output layer, where MLP layers are employed to predict user behaviors based on graph representations.

(a) Preprocessing Ego Networks

- Sampling a subset of users from one’s ego-network
  - Breadth-First Search (BFS)

(b) Input Layer

- Input features X
  - demographic and social role features
  - two-dimensional contextual features (indicating each user's active status and positions in the ego network)
  - pre-trained user embeddings via ProNE
  - factorization machine (FM) technique to model feature interactions

$$X^{\text{2nd}} = \frac{1}{2}((\sum_{i=1}^{F} W_i x^{(i)})^2 - \sum_{i=1}^{F}(W_i x^{(i)})^2)$$

(c) Feature Smoothing Layer

- Random walk normalized graph Laplacian
$$\mathcal{L} = I_m - D^{-1}A$$
- where A is the adjacency matrix of the ego network, m is the size of each ego network, I_m is the identity matrix, and $D = \sum_j A_{ij}$.
- $\mathcal{L}$ can be decomposed as
$$\mathcal{L} = U\Lambda U^T$$
- where $\Lambda = \mathrm{diag}[\lambda_1, \ldots, \lambda_m]$.

(c) Feature Smoothing Layer

- Small (large) eigenvalues in a graph Laplacian control the network's global clustering (local smoothing) effect.

$$\widetilde{\mathcal{L}} = U\mathrm{diag}([g(\lambda_1), g(\lambda_2), ..., g(\lambda_m)])U^\top$$

- where $\tilde{\mathcal{L}}$ is the modulated Laplacian and g is the spectral modulator.

$$g(\lambda) = e^{-\frac{1}{2}[(\lambda-\mu)^2-1]\theta}$$

$$X^{l_0} = D^{-1}A(I_m - \widetilde{\mathcal{L}})X$$

(d) Hierarchical Graph Representation Learning

- Idea:
  - design a hierarchical representation learning method to encode the substructures of ego networks
  - cluster similar nodes iteratively to encode these substructures
  - employ graph neural networks as basic modules

(d) Hierarchical Graph Representation Learning

- We first generate node embeddings of the entire ego networks via a GNN

$$Z^{l_1} = \mathrm{GNN}_{0,\mathrm{embed}}(A^{l_0}, X^{l_0})$$

- Following DIFFPOOL, we learn an assignment matrix $B^{l_{k+1}}$ to map nodes to high-order clusters.

$$B^{l_{k+1}} = \mathrm{softmax}(\mathrm{GNN}_{k,\mathrm{pool}}(A^{l_k}, X^{l_k}))$$

- where $B^{l_{k+1}} \in R^{m_k \times m_{k+1}} (m_{k+1} < m_k,\ m_0 = m)$ and $b_{ij}^{l_{k+1}}$ represents the probability of assigning node i to j-th cluster in $(k+1)$-th assignment layer

(d) Hierarchical Graph Representation Learning

- An ego network can be transformed into a smaller graph iteratively

$$X^{l_k} = {B^{l_k}}^\top Z^{l_k} \in \mathbb{R}^{m_k \times h_k}$$

$$A^{l_k} = {B^{l_k}}^\top A^{l_{k-1}} B^{l_k} \in \mathbb{R}^{m_k \times m_k}$$

- Based on the coarsened graph, the coarse-level node embeddings can be generated by

$$Z^{l_{k+1}} = \mathrm{GNN}_{k,\mathrm{embed}}(A^{l_k}, X^{l_k})$$

$$Z^{\mathrm{graph}} = \big\Vert_{k=1}^{L} \sigma(Z^{l_k})$$

(d) Hierarchical Graph Representation Learning

- Basic GNN modules — Graph Attention Networks (GAT)
  - additive attention

$$\alpha_{ij}^{\mathrm{AA}} = \frac{\exp(\mathrm{act}(a_{\mathrm{src}}^{\top} W_p x_i + a_{\mathrm{dst}}^{\top} W_p x_j))}{\sum_{t \in \mathcal{N}_i} \exp(\mathrm{act}(a_{\mathrm{src}}^{\top} W_p x_i + a_{\mathrm{dst}}^{\top} W_p x_t))}$$

  - dot attention (proposed)

$$\alpha_{ij}^{\mathrm{DA}} = \frac{\exp(\mathrm{act}((a_{\mathrm{src}}^{\top} W_p x_i + b_{\mathrm{src}}) \cdot (a_{\mathrm{dst}}^{\top} W_p x_j + b_{\mathrm{dst}})))}{\sum_{t \in \mathcal{N}_i} \exp(\mathrm{act}((a_{\mathrm{src}}^{\top} W_p x_i + b_{\mathrm{src}}) \cdot (a_{\mathrm{dst}}^{\top} W_p x_t + b_{\mathrm{dst}})))}$$

- Denoted as $\mathrm{DiffuseGNN}_{\mathrm{AA}}$ and $\mathrm{DiffuseGNN}_{\mathrm{DA}}$

(e) Output Layer

- Pass $\mathrm{Z}^{\mathrm{graph}}$ into fully connected layers to generate prediction scores
- Use cross-entropy loss function

Prediction Results

Method	WeChat "Wow"				WeChat Click				Weibo			
	Prec	Rec	F1	AUC	Prec	Rec	F1	AUC	Prec	Rec	F1	AUC
Random	47.84	50.06	48.92	50.05	28.64	50.13	36.45	50.02	25.12	50.48	33.55	50.15
LR	68.08	70.08	69.06	76.73	41.71	67.01	51.41	70.07	42.97	71.37	53.64	76.38
RF [22]	69.20	65.17	67.12	76.69	39.52	74.69	51.69	70.12	40.03	73.66	51.87	75.14
xDeepFM [21]	66.23	80.96	72.85	78.25	40.88	75.09	52.94	71.61	30.20	73.90	42.88	64.38
DeepInf [31]	70.28	**81.46**	75.46	83.06	43.88	76.03	55.65	74.50	48.09	71.67	57.56	81.46
Wang et al. [44]	69.76	79.40	74.27	81.91	41.91	75.07	53.79	72.31	45.58	74.63	56.59	80.26
SAGPool [19]	81.74	75.43	78.46	86.18	46.58	79.19	58.66	77.37	43.79	73.81	54.97	78.89
ASAP [32]	71.13	79.81	75.22	83.28	44.92	76.57	56.62	75.48	46.55	70.64	56.12	79.87
StructPool [48]	67.56	79.21	72.92	79.46	40.20	78.55	53.19	71.54	30.47	72.87	42.98	61.83
DiffuseGNN_{AA}	84.95	76.81	**80.67**	87.64	**46.63**	82.01	59.46	78.05	**50.09**	72.87	**59.37**	**83.08**
DiffuseGNN_{DA}	**85.46**	76.30	80.62	**87.69**	46.45	**82.81**	**59.52**	**78.27**	48.70	**74.88**	59.01	82.76
w/o pre-train	74.96	78.42	76.65	84.91	45.68	75.77	57.00	76.09	47.33	74.15	57.78	81.51
w/o node feature	85.01	75.69	80.08	87.10	45.64	82.26	58.71	77.64	46.34	74.46	57.13	81.10
w/o 2nd feature	86.40	76.16	80.16	87.50	46.66	81.66	59.39	78.16	46.12	75.02	57.12	81.03
w/o smoothing	79.23	77.57	78.39	86.04	46.26	78.38	58.18	76.89	48.95	72.20	58.35	82.13

- DiffuseGNN outperforms traditional classifiers, SOTA social influence prediction models and hierarchical graph representation learning methods.

Parameter Analysis

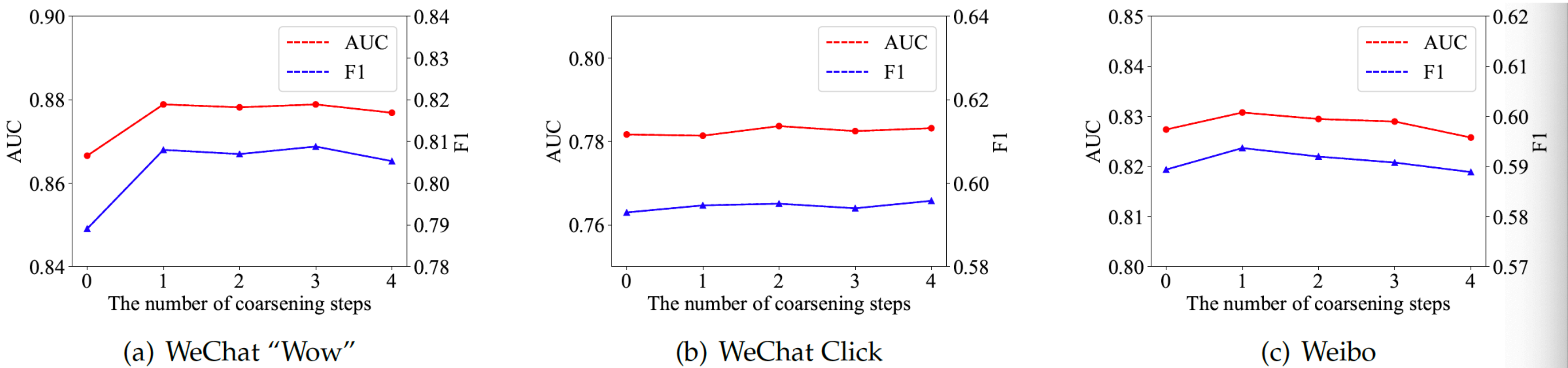
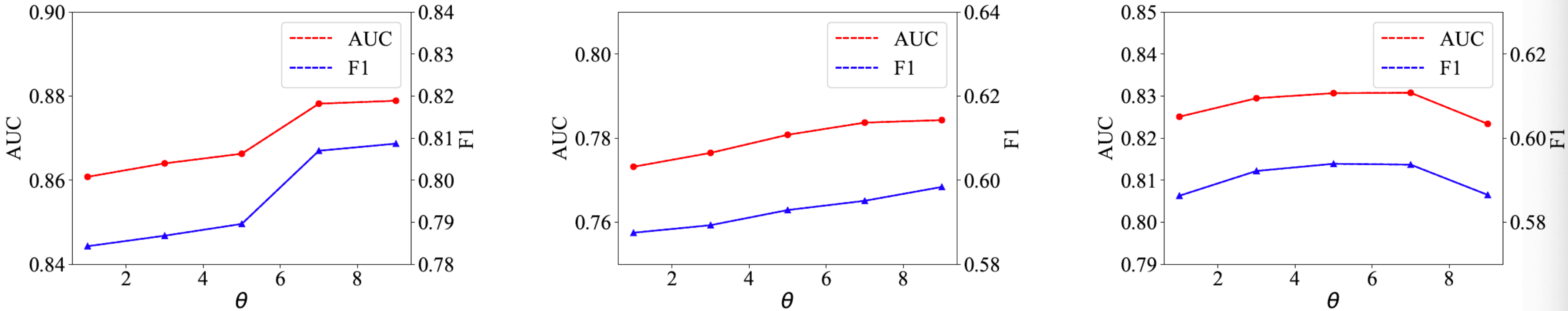


Fig. 11. "Wow" and click performance (AUC) on test dataset w.r.t. the number of pooling layers in hierarchical graph representation learning

- Hierarchical graph representation is clearly better than "flat" representation.
- 1 coarsening step is best for WeChat "Wow" and Weibo, while 2 coarsening step is best for WeChat Click.

Parameter Analysis



(a) WeChat "Wow" (b) WeChat Click (c) Weibo

Fig. 12. Prediction performance (AUC and F1) on test dataset w.r.t. θ in the graph filter of the feature smoothing step.

- θ in g affects the peak value of modulated eigenvalue.
- Larger θ is often better ($1 \leq \theta \leq 9$)

$$g(\lambda) = e^{-\frac{1}{2}[(\lambda-\mu)^2-1]\theta}$$

Conclusion

- Our study reveals several interesting patterns about the correlations between users' "Wow" and click behavior and factors of different granularities.
- Based on our discoveries, we propose a new model framework DiffuseGNN to predict user behaviors.
- Experiments show that DiffuseGNN consistently outperforms state-of-the-art baseline methods on WeChat and Weibo dataset.

Thank You

ArXiv: https://arxiv.org/abs/2103.02930
Code: https://github.com/zfjsail/wechat-wow-analysis