Financial Defaulter Detection on Online Credit Payment via Multi-view Attributed Heterogeneous Information Network

Qiwei Zhong[1]; Yang Liu[12]; Xiang Ao[2*];

Binbin Hu[1]; Jinghua Feng[1]; Jiayu Tang[1*]; Qing He[2]

1 Alibaba Group

2 *Institute of Computing Technology, Chinese Academy of Sciences*

* *denotes corresponding authors.*

Content

- Motivation
- Method
- Experiment
- Conclusion and Future Work
- Reference

Content

- Motivation
  - Background
  - Related Work
  - Challenges
- Method
- Experiment
- Conclusion and Future Work
- Reference

Background

- Payment
  - Cash, bank card, online money, etc.
  - What if no enough money on hand?
- Credit Payment
  - Promise to pay for it later.
  - What if you fail to repay the money in-time?

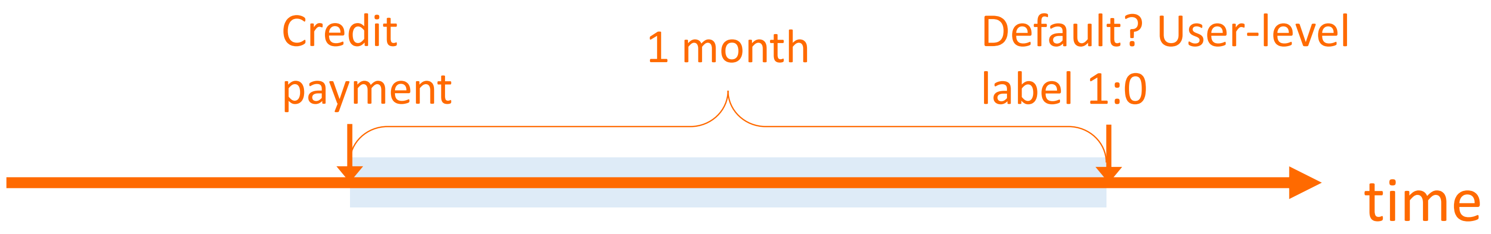
花呗
ANT CREDIT PAY

PayPal
CREDIT

Background

➢Defaulters

- Defaulters are those who could not pay the requirements within one month.



➢Task

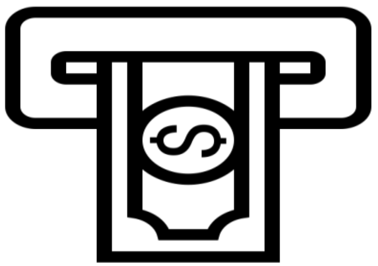
- Financial Defaulter detection
  - To predict whether a user will fail to make required payments in the next month.

➢Data

- User behaviors on credit-payment service platform
  - Payment transactions, log-in logs, etc.

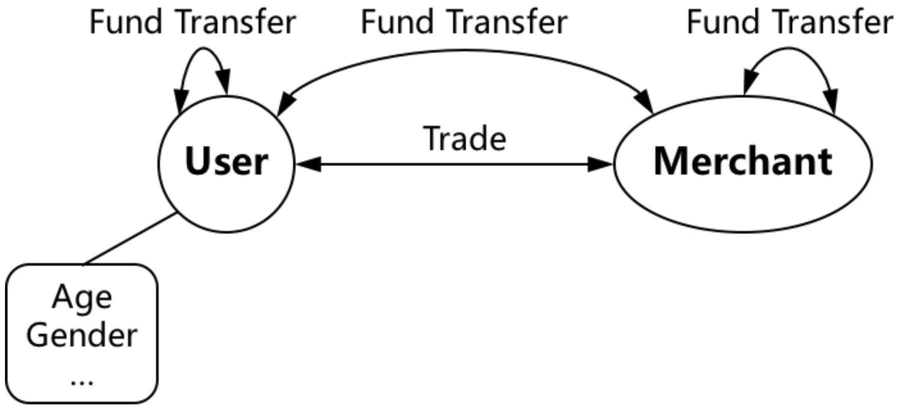
➢ Financial Defaulter Detection

- Fraud
- Cash-out
- Money Laundering



➢ Attributed Heterogeneous Information Network

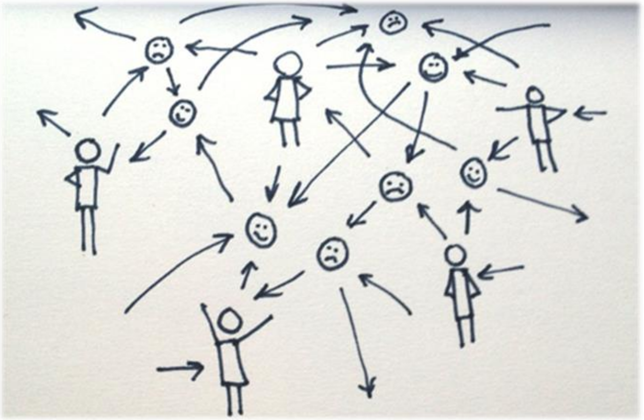
- Node
  - User, Merchant
- Link
  - Fund Transfer, Trade



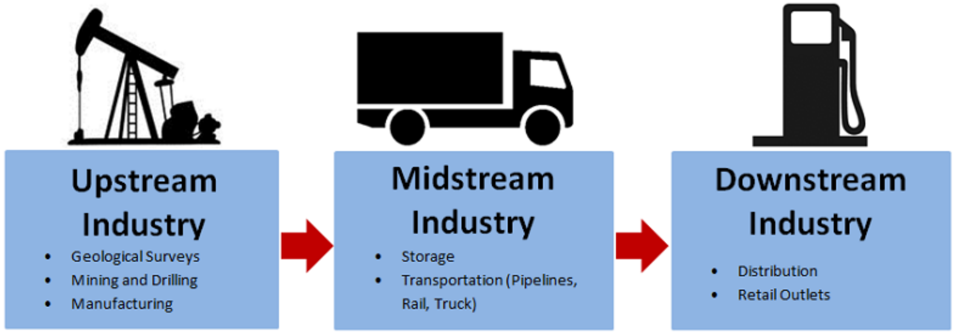
Please refer to [13, 19, 32] in our paper.

- Endogeny
  - Users could be subjectively reluctant to afford when they raise a debt.

- Adversary
  - The criminals may deliberately construct complex behaviors to avoid regulation.



- Accumulation
  - May be impacted by upstream or down-stream neighbor enterprises.



Solutions

- Endogeny
  - Users could be subjectively reluctant to afford when they raise a debt.
- Adversary
  - The criminals may deliberately construct complex behaviors to avoid regulation.
- Accumulation
  - May be impacted by upstream or down-stream neighbor enterprises.

Accurate user profiling from multi-view factors

+

Fine-grained behavioral patterns among users

+

Impacts of neighbors

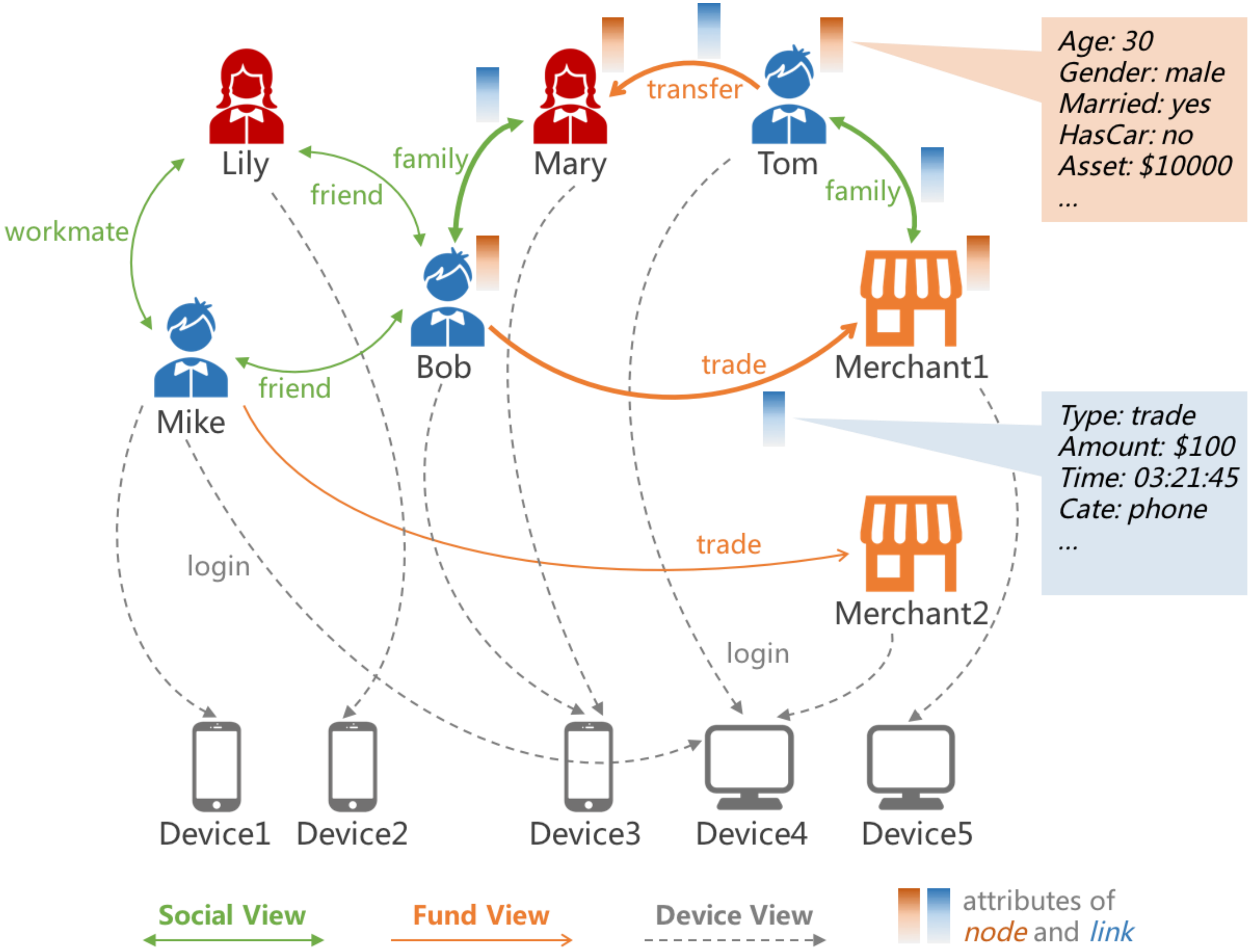
=

Multi-view Attributed Heterogeneous Information Network based financial DEfault useR detection

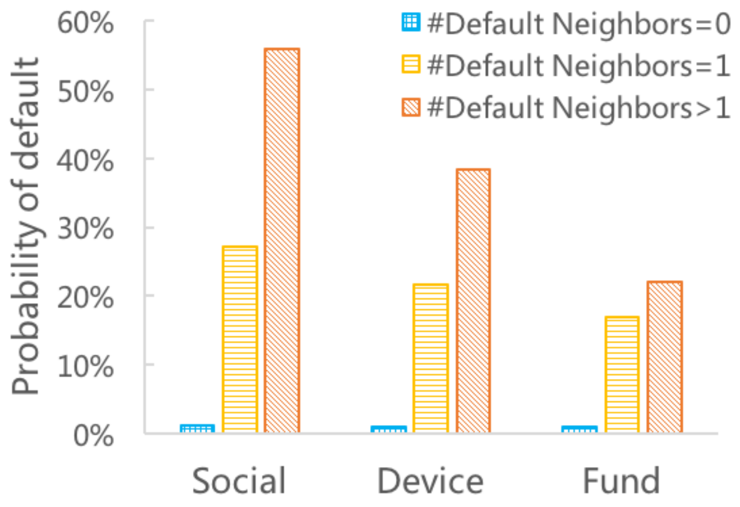
Content

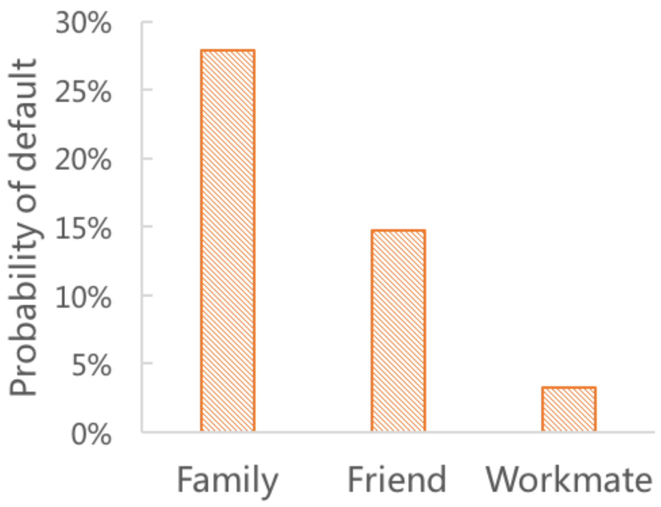
Multi-view Attributed Heterogeneous Information Network

- View
  - Social
  - Fund
  - Device
- Node
  - User
  - Merchant
- Link
  - Friend, family, workmate
  - Transfer, trade
  - Login



Statistical Analysis on Three Views





- Observation:
  - Users are more likely to be default when they have default neighbors.
  - Different views have different impacts on users.
  - Different relations have different impacts.

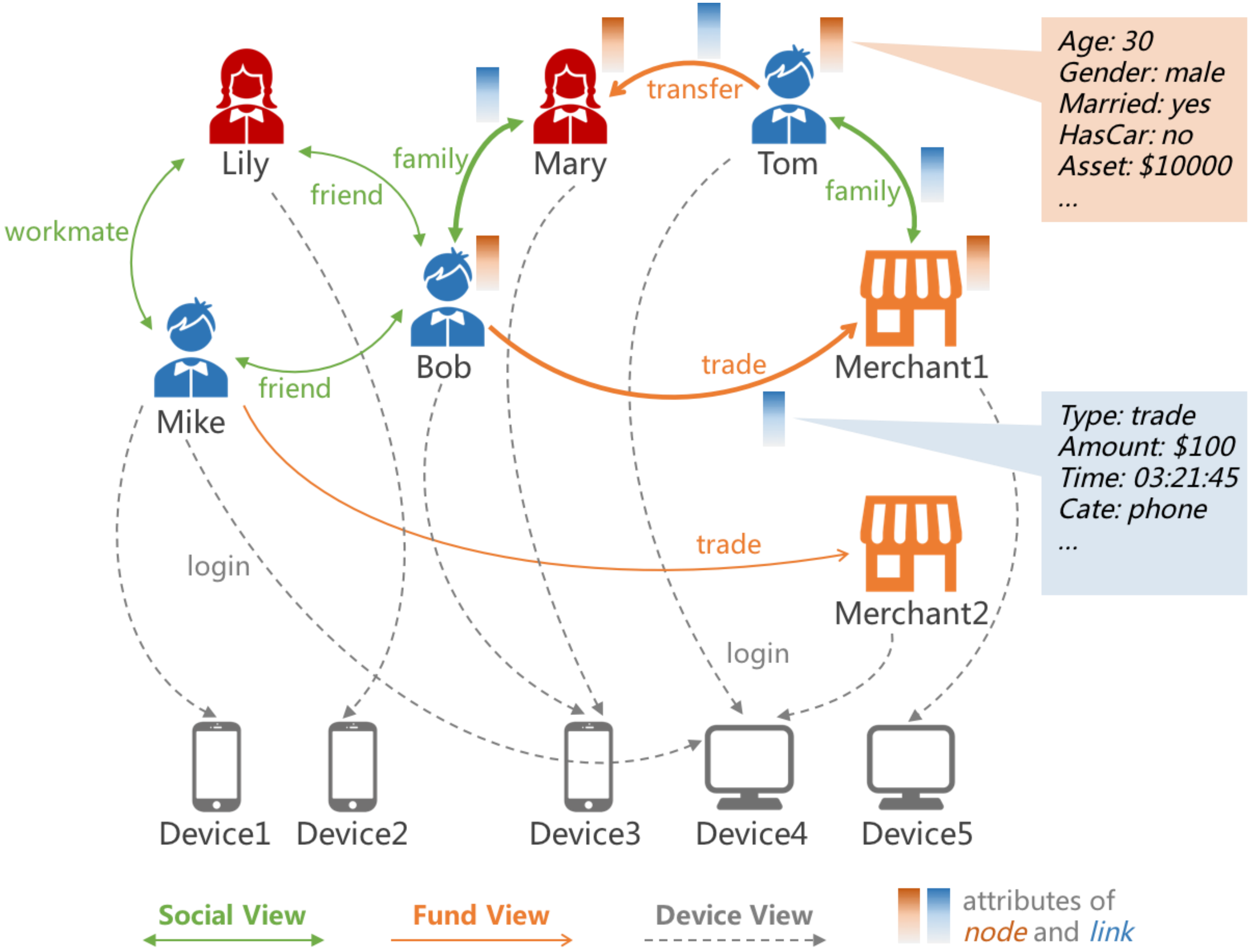
Meta-path on MAHIN

➢Intra-view meta-path

- UsU: $User \xrightarrow{social} User$
- UdU: $User \xrightarrow{device} User$
- UfU: $User \xrightarrow{fund} User$
- UsUsU: $User \xrightarrow{social} User \xrightarrow{social} User$
- UfUfU: $User \xrightarrow{fund} User \xrightarrow{fund} User$

➢Cross-view meta-path

- UdUsU: $User \xrightarrow{device} User \xrightarrow{social} User$
- UfUsU: $User \xrightarrow{fund} User \xrightarrow{social} User$
- UfUsUfU: $User \xrightarrow{fund} User \xrightarrow{social} User \xrightarrow{fund} User$



Meta-path based Path Encoder

$$z_i^o = \tanh(\boldsymbol{W}_a \boldsymbol{h}_i^o)$$

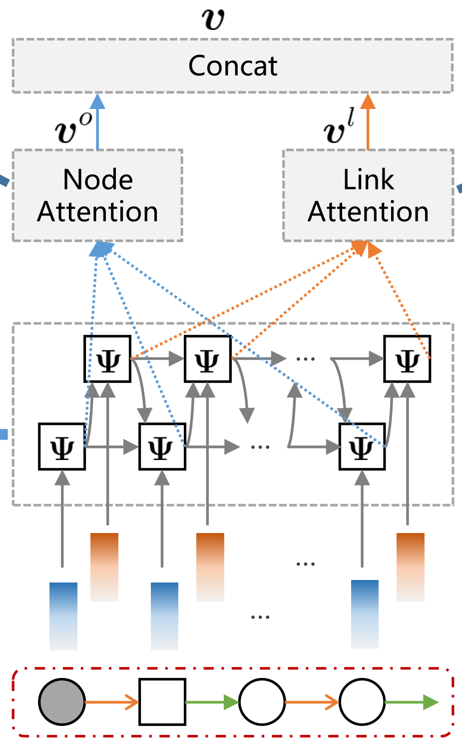
$$a_i^o = \frac{\exp(\boldsymbol{q}_a^{\mathrm{T}} \boldsymbol{z}_i^o)}{\sum_{j=1}^{L} \exp(\boldsymbol{q}_a^{\mathrm{T}} \boldsymbol{z}_j^o)}$$

$$\boldsymbol{v}^o = \sum_{i=1}^{L} a_i^o \boldsymbol{h}_i^o$$

$$z_i^l = \tanh(\boldsymbol{W}_a \boldsymbol{h}_i^l)$$

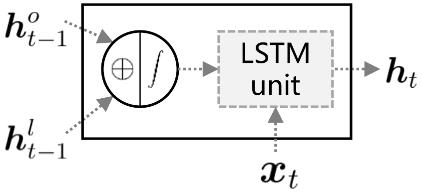
$$a_i^l = \frac{\exp(\boldsymbol{q}_a^{\mathrm{T}} \boldsymbol{z}_i^l)}{\sum_{j=1}^{L} \exp(\boldsymbol{q}_a^{\mathrm{T}} \boldsymbol{z}_j^l)}$$

$$\boldsymbol{v}^l = \sum_{i=1}^{L} a_i^l \boldsymbol{h}_i^l$$



Meta-path based Path Encoder

Neural Unit Ψ



$$\boldsymbol{h}'_{t-1} = [\alpha \boldsymbol{h}_{t-1}^o; (1-\alpha)\boldsymbol{h}_{t-1}^l]$$

$$\boldsymbol{h}_{t-1} = \sigma(\boldsymbol{W}_\phi \boldsymbol{h}'_{t-1} + \boldsymbol{b}_\phi)$$

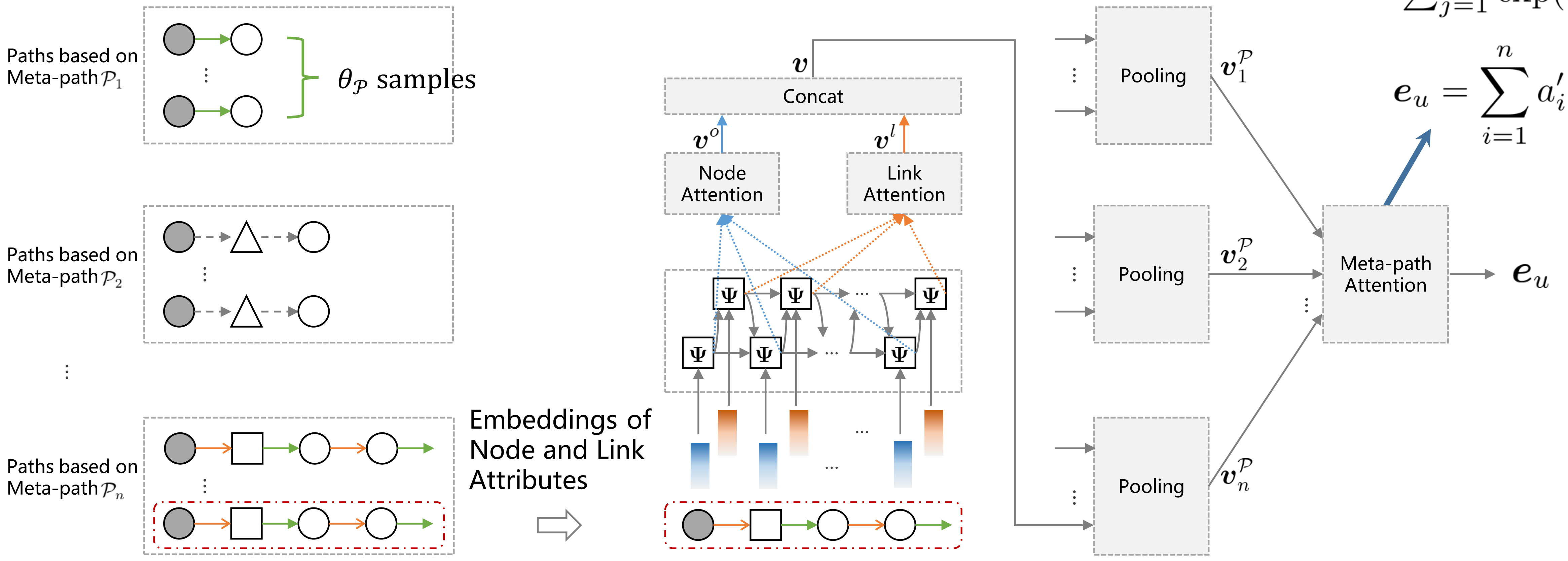
$$\boldsymbol{h}_t = \mathbf{LSTM}(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t, \boldsymbol{c}_{t-1})$$

Importance of Views

$$z_i' = \tanh(\boldsymbol{W}_a' \boldsymbol{v}_i^{\mathcal{P}})$$

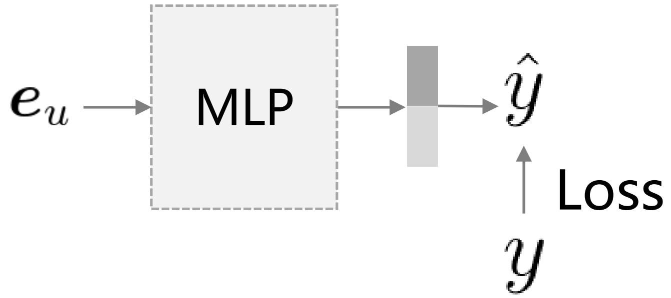
$$a_i' = \frac{\exp({\boldsymbol{q}_a'}^{\mathrm{T}} {\boldsymbol{z}_i}')}{\sum_{j=1}^{n} \exp({\boldsymbol{q}_a'}^{\mathrm{T}} {\boldsymbol{z}_j}')}$$

$$\boldsymbol{e}_u = \sum_{i=1}^{n} a_i' \boldsymbol{v}_i^{\mathcal{P}}$$



Meta-path on MAHIN **Meta-path based Path Encoder** **Modeling Importance of Views**

Model Training



$$z_u = ReLU(W_L \cdots ReLU(W_1 e_u + b_1) + b_L)$$

$$p_u = \sigma(w_p^T z_u + b_p)$$

$$\mathcal{L}(\Theta) = \sum_{\langle u, y_u \rangle \in \mathcal{D}} - (y_u \log(p_u) + (1 - y_u) \log(1 - p_u)) + \lambda \|\Theta\|_2^2$$

Content

- Motivation
- Method
- Experiment
  - Dataset
  - Compared Methods
  - Evaluation Metrics
  - Main Results and Analysis
- Conclusion and Future Work
- Reference

➢Data

Dataset	#Positive	#Negative	#Total	#Positive Rate
Training	6,950	1,374,355	1,381,305	0.503%
Testing	2,522	511,116	513,638	0.491%

➢MAHIN

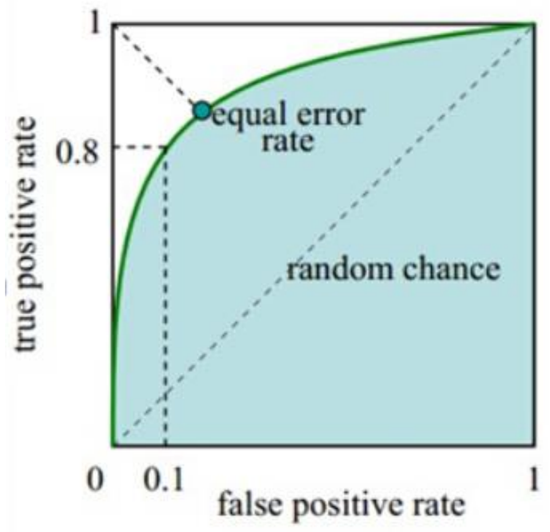
Dataset	Type		Attribute		Total
	Number	Examples	Number	Examples	
Node	4	User/ Merchant/ Phone/ Computer	100	NodeType/ [*User Profiles*]: Age/Gender/Married/IsVIP/... [*Credit Information*]: CreditScore/IsInBlacklist/... [*Purchase Behaviors*]: PurchaseAmountAYear/... [*Asset Information*]: Asset/HasCar/HasFactory/...	14,984,670
Link	6	Family/Friend/Workmate/ Trade/Transfer/ Login	45	LinkType/ [*Social*]: FirstRelatedTime/... [*Fund*]: TradeCategory/TransferAmount/... [*Device*]: LoginTime/StayMinute/...	168,864,052

Compared Methods

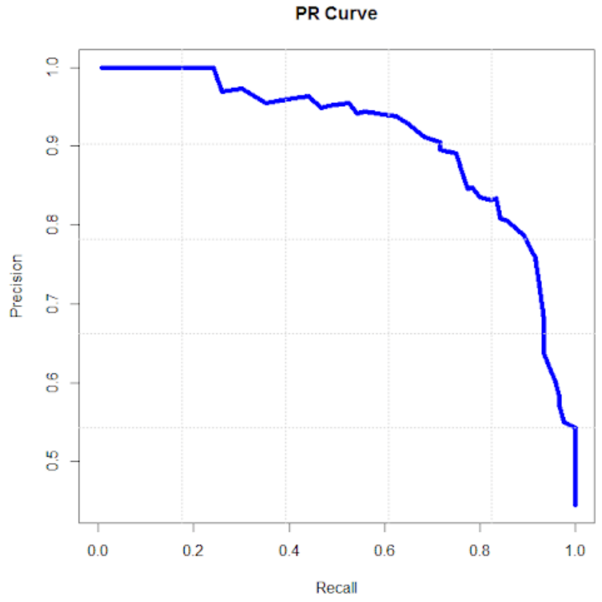
- GBDT[7]
  - A scalable tree-based model for feature learning and classification task.
- DeepForest[39, 42]
  - A deep model based on decision trees.
- HAN[33]
  - A graph neural network with node-level and semantic-level attention.
  - HAN_{S2} extracts interactive features of a target user following the meta-paths defined in our paper.
- HACUD[13]
  - A cash-out user detection method based on attributed heterogeneous information network.
  - $HACUD_{S2}$ extracts interactive features of a target user following the meta-paths defined in our paper.

Evaluation Metrics

- AUC
  - The area under the ROC curve



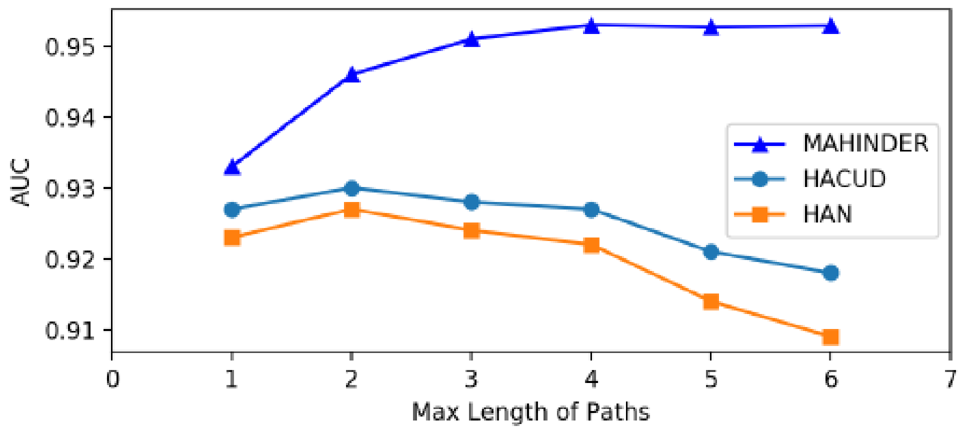
- R@P$_N$
  - The Recall when Precision equals N

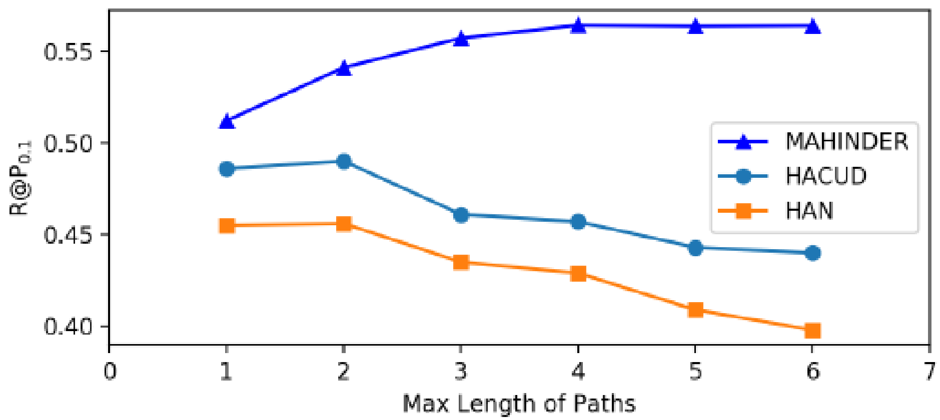


Main Results

Table 1 Performances of different methods on the dataset. The subscripts indicate the increasing value compared to GBDT.

Metric	GBDT	DeepForest	HAN	HACUD	HAN_{S2}	$HACUD_{S2}$	MAHINDER
AUC	$0.891_{/0.000}$	$0.914_{/0.023}$	$0.920_{/0.029}$	$0.925_{/0.034}$	$0.927_{/0.036}$	$0.930_{/0.039}$	$\mathbf{0.953}_{/\mathbf{0.062}}$
$R@P_{0.1}$	$0.403_{/0.000}$	$0.411_{/0.008}$	$0.424_{/0.021}$	$0.433_{/0.030}$	$0.456_{/0.053}$	$0.490_{/0.087}$	$\mathbf{0.564}_{/\mathbf{0.161}}$



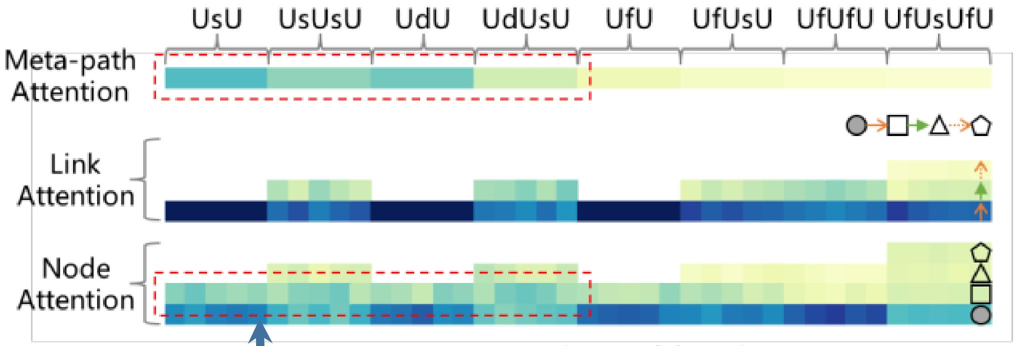


Ablation Test

Metric	MAHINDER$_{\setminus S}$	MAHINDER$_{\setminus D}$	MAHINDER$_{\setminus F}$	MAHINDER$_{\setminus L}$	MAHINDER$_{\setminus EnAtt}$	MAHINDER$_{\setminus MpAtt}$	MAHINDER
AUC	$0.929_{/-0.024}$	$0.934_{/-0.019}$	$0.938_{/-0.015}$	$0.936_{/-0.017}$	$0.945_{/-0.008}$	$0.942_{/-0.011}$	$\mathbf{0.953}_{/\mathbf{0.000}}$
R@P$_{0.1}$	$0.487_{/-0.077}$	$0.510_{/-0.054}$	$0.521_{/-0.043}$	$0.525_{/-0.039}$	$0.543_{/-0.021}$	$0.536_{/-0.028}$	$\mathbf{0.564}_{/\mathbf{0.000}}$

- MAHINDER$_{\setminus S}$ removes social view and its corresponding meta-paths
- MAHINDER$_{\setminus D}$ removes device view and its corresponding meta-paths
- MAHINDER$_{\setminus F}$ removes fund view and its corresponding meta-paths
- MAHINDER$_{\setminus L}$ removes link information and its corresponding attention module
- MAHINDER$_{\setminus EnAtt}$ removes node and link attention mechanisms in path encoder
- MAHINDER$_{\setminus MpAtt}$ removes attention mechanism modeling importance of views

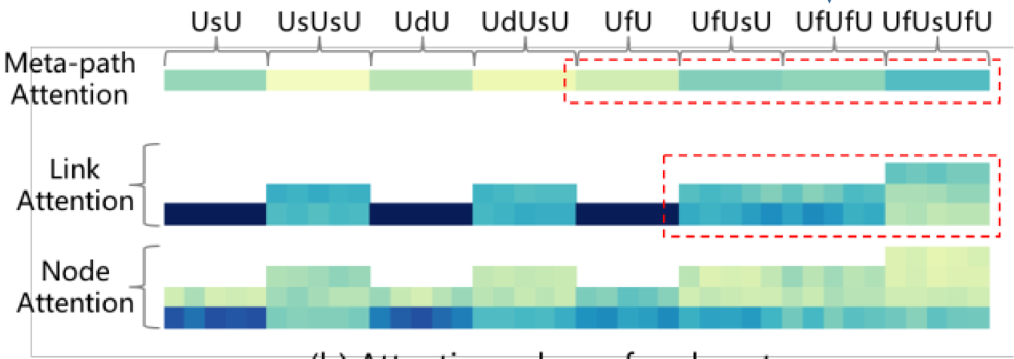
Visualization



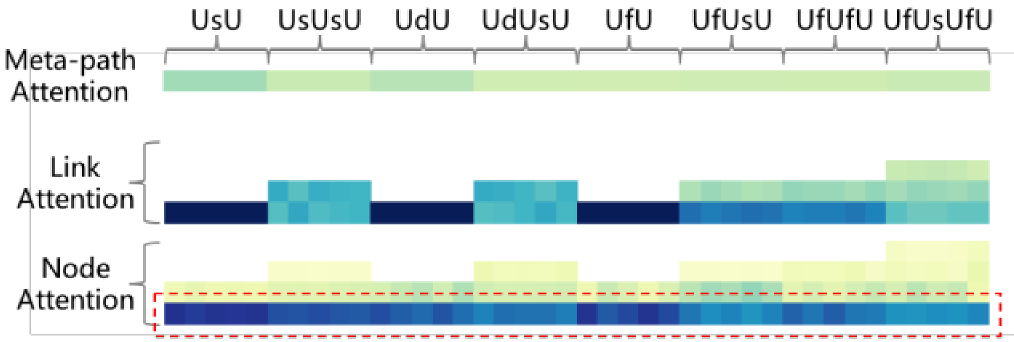
(a) Attention values of fraud user

The fraud users have higher attention values on social and device views (e.g., UsU, UdU) and first-order neighbors.

The cash-out users have higher attention values on fund and social views (e.g., UfUsU, UfUsUfU) and high-order links.



(b) Attention values of cash-out user



(c) Attention values of unintentional defaulter

The unintentional defaulters have higher attention value on themselves.

Content

- Motivation
- Method
- Experiment
- Conclusion and Future Work
- Reference

Conclusion and Future work

➢Conclusion

- We construct a multi-view attributed heterogeneous information network for better user profiling.
- We propose a novel model named MAHINDER which is effective in financial defaulter detection.

➢Future Work

- End-to-end model without pre-defined meta-paths
- Interpretability

Reference

[7] Jerome H Friedman. 2001. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics (2001), 1189–1232

[13] Binbin Hu, Zhiqiang Zhang, Chuan Shi, Jun Zhou, Xiaolong Li, and Yuan Qi. 2019. Cash-out User Detection based on Attributed Heterogeneous Information Network with a Hierarchical Attention Mechanism. In AAAI. 946–953

[19] Ziqi Liu, Chaochao Chen, Xinxing Yang, Jun Zhou, Xiaolong Li, and Le Song. 2018. Heterogeneous Graph Neural Networks for Malicious Account Detection. In CIKM. 2077–2085.

[32] Daixin Wang, Jianbin Lin, Peng Cui, Quanhui Jia, Zhen Wang, Yanming Fang, Quan Yu, Jun Zhou, Shuang Yang, and Yuan Qi. 2019. A Semi-supervised Graph Attentive Network for Financial Fraud Detection. In ICDM.

[33] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous Graph Attention Network. In WWW. 2022–2032.

[39] YaLin Zhang, Xiaolong Li, Yuan Qi, ZhiHua Zhou, Jun Zhou, Wenhao Zheng, Ji Feng, Longfei Li, Ziqi Liu, Ming Li, and et al. 2019. Distributed Deep Forest and its Application to Automatic Detection of Cash-Out Fraud. ACM Transactions on Intelligent Systems and Technology 10, 5 (2019).

[42] Zhihua Zhou and Ji Feng. 2017. Deep Forest: Towards An Alternative to Deep Neural Networks. In IJCAI. 3533–3539.

Acknowledgment

Thanks for listening!

If you have any question, feel free to contact us at

yunwei.zqw@alibaba-inc.com
liuyang17z@ict.ac.cn
aoxiang@ict.ac.cn