蚂蚁集团 ANT GROUP | DataFun.

图机器学习在安全风控的应用

大安全—机器智能
朱 亮

目录 CONTENT

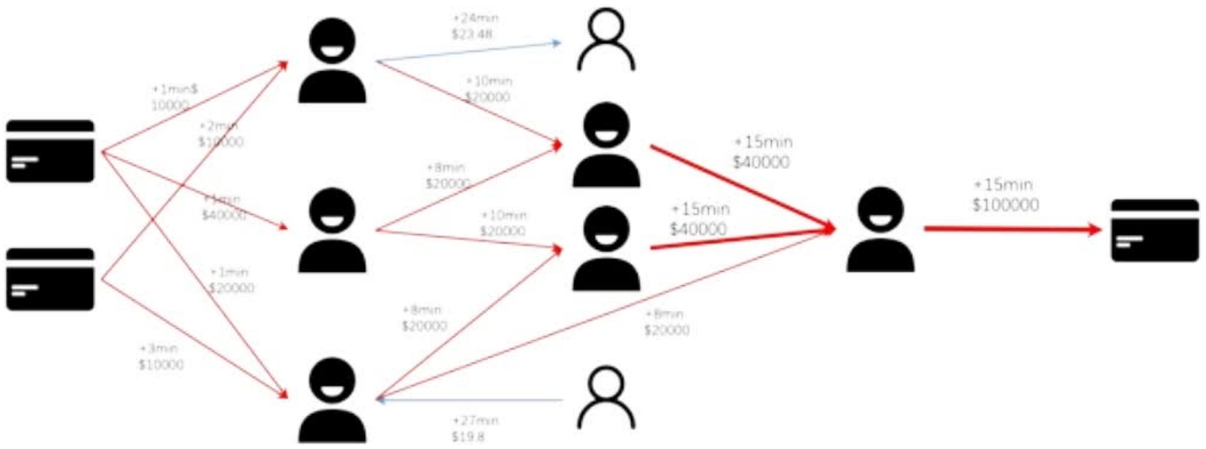
蚂蚁集团 ANT GROUP | DataFun.

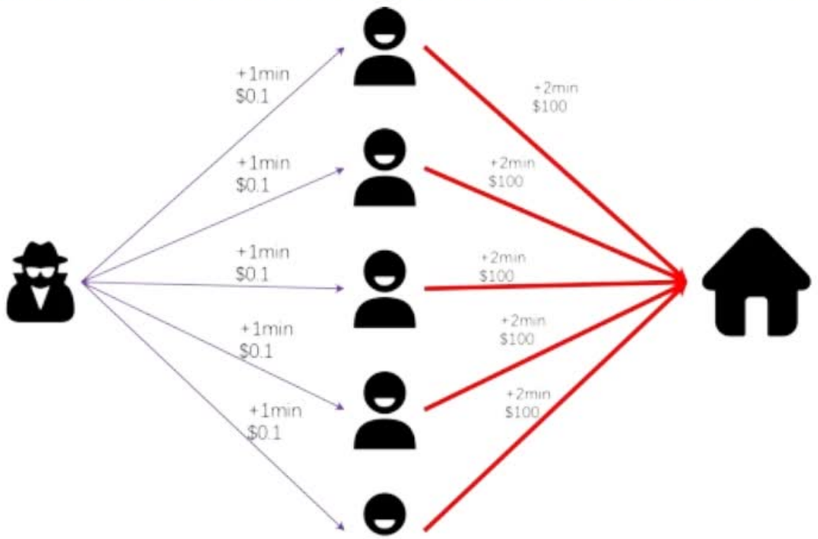
蚂蚁集团 ANT GROUP | DataFun.

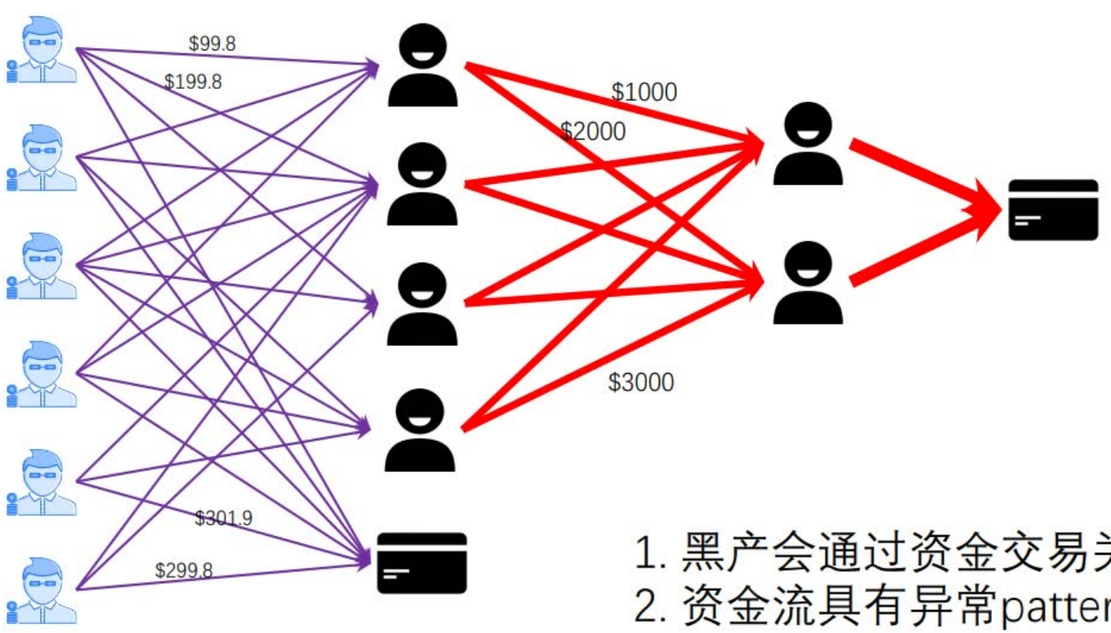
01

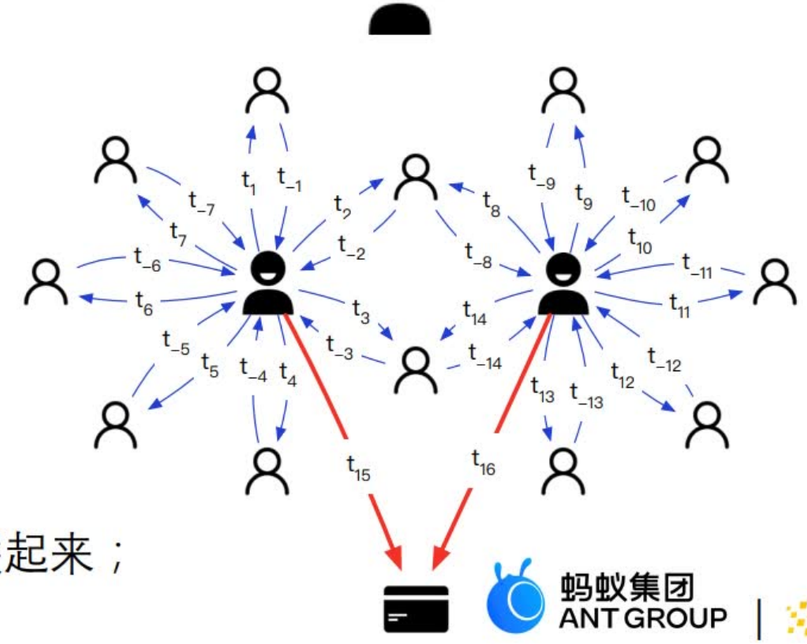
背景介绍

背景介绍









1. 黑产会通过资金交易关系关联起来；
2. 资金流具有异常pattern；

蚂蚁集团 ANT GROUP | DataFun.

蚂蚁集团 ANT GROUP | DataFun.

02

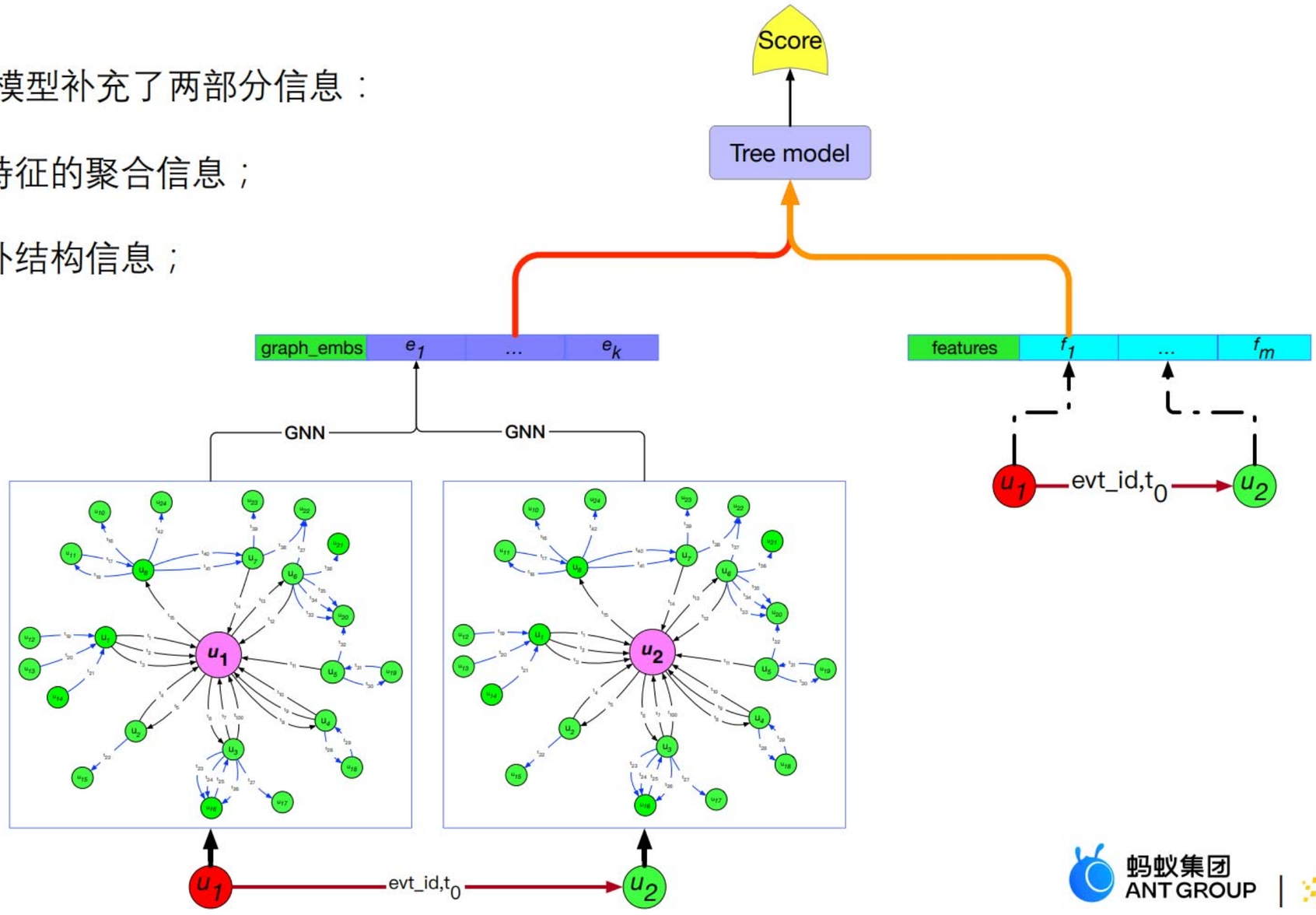
架构简介

算法架构简介

相较于树模型，图模型补充了两部分信息：

✓ k度子图内点边特征的聚合信息；

✓ k度子图内的拓扑结构信息；



蚂蚁集团 ANT GROUP | DataFun.

蚂蚁集团 ANT GROUP | DataFun.

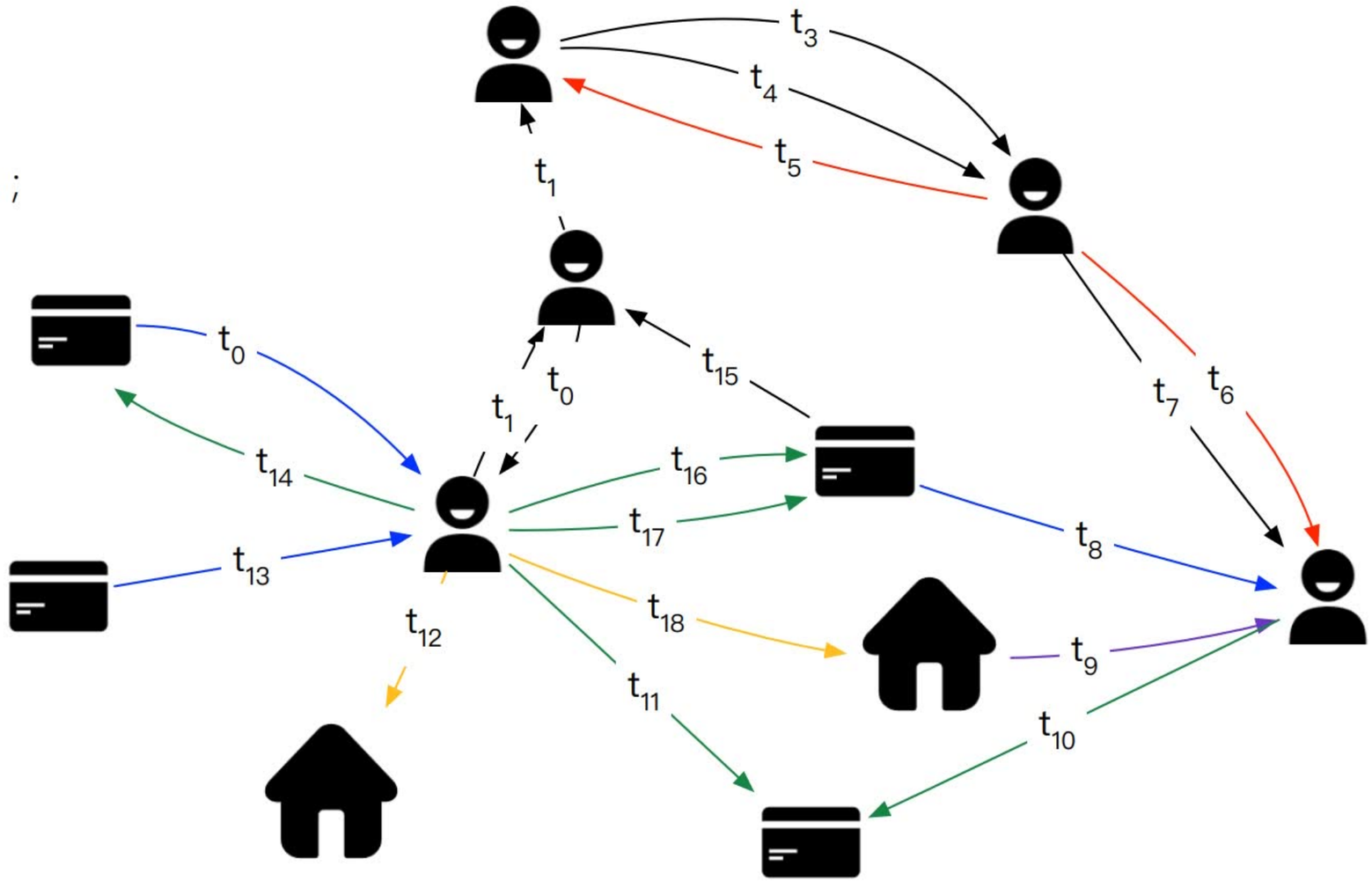
03

安全风控图模型

- 有向动态异质资金图
- 主网络介绍
- DDGCL

安全风控图模型——有向动态异质图

- 点：账号、卡号、商家等；
- 边：有方向，多种类型边关系；



蚂蚁集团 ANT GROUP | DataFun.

图算法和风控的相遇

“作案”有团伙特性

“作案”有相似性

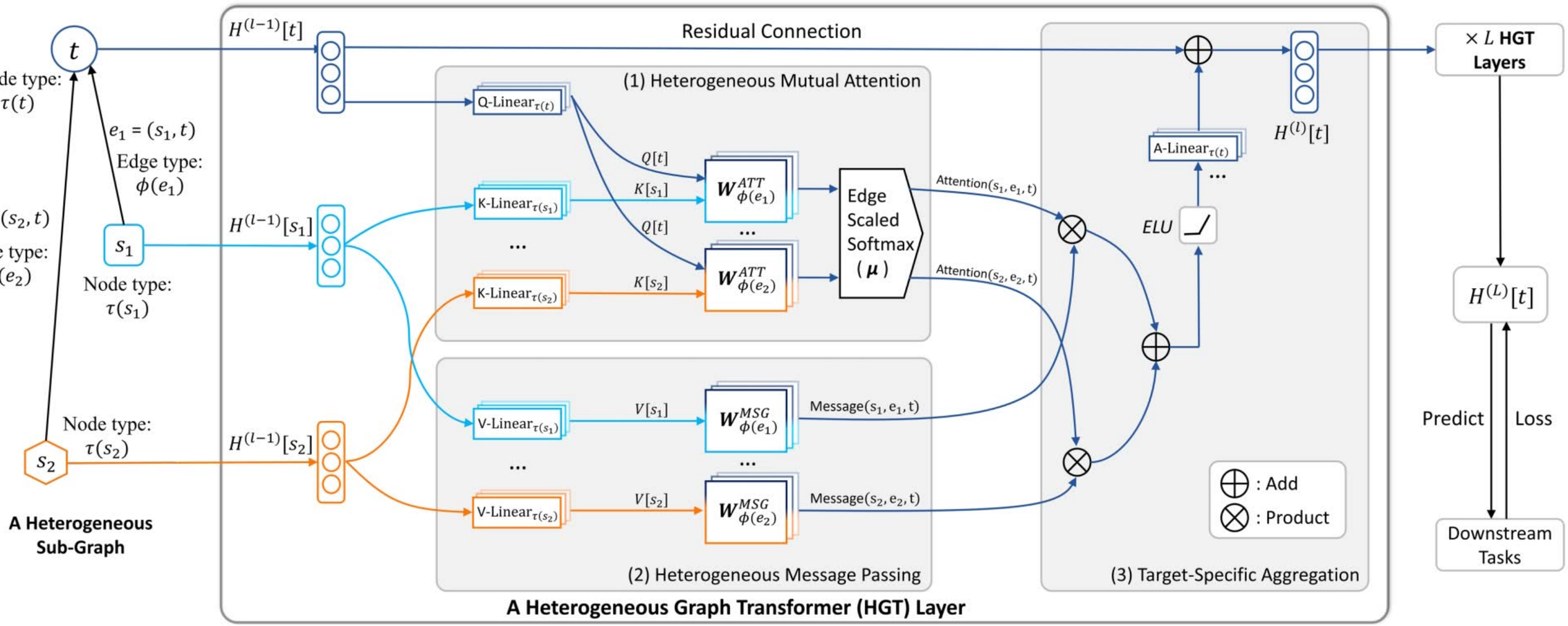
“作案”需要大量账号和设备资源配合

“作案”具有成本因素

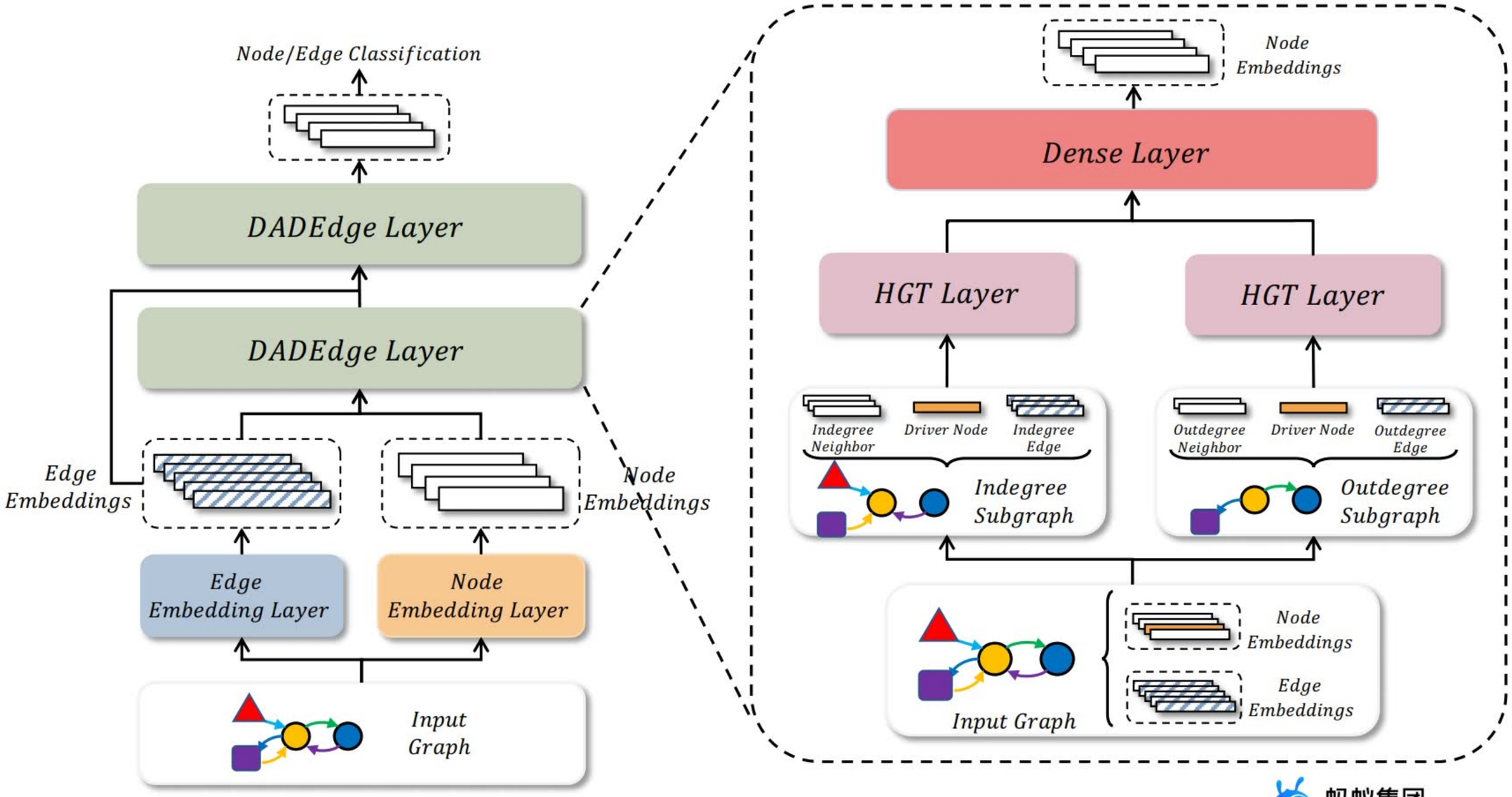
物以类聚，人以群分

安全风控图模型——HGT+DADEdge(Directional Attention Dual Edge Embedding)

将边方向作为一个边上一个属性



安全风控图模型——HGT+DADEdge(Directional Attention Dual Edge Embedding)



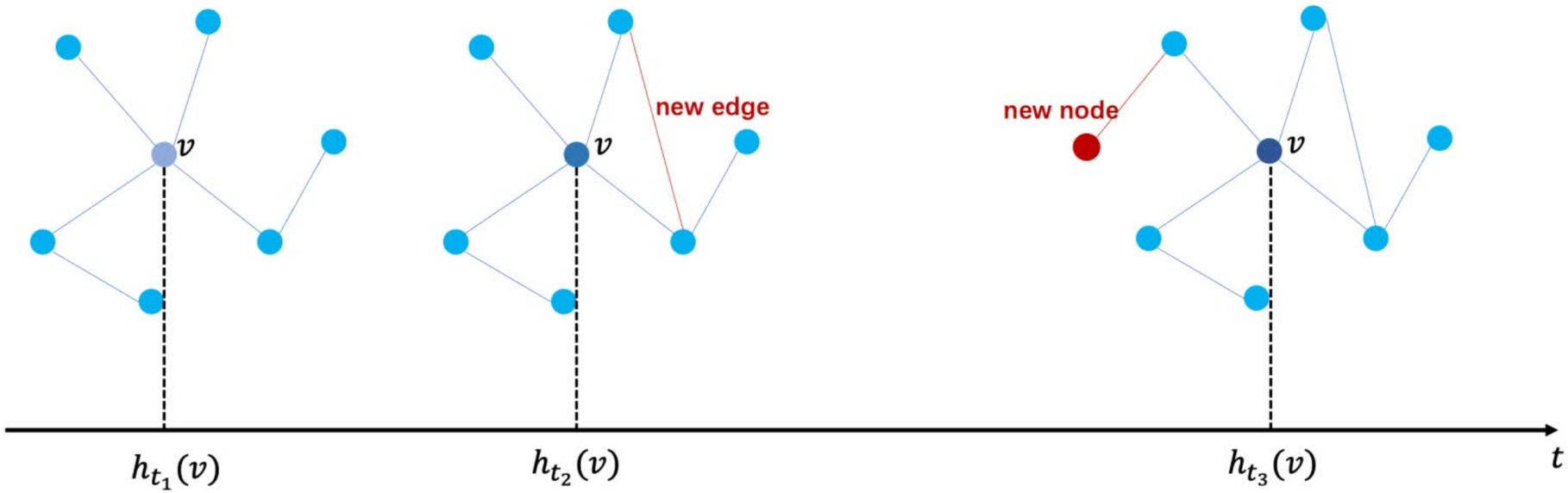
蚂蚁集团 ANT GROUP | DataFun.

安全风控图模型

以某一个**场景为例，各模型效果对比如下：

	AUC 提升	千一打扰下 召回提升
TGAT	0	0
HGT	+0.0327	+5.634%
HGT+DADEdge	**+0.0335**	**+10.856%**

安全风控图模型——DDGCL(Debiased Dynamic Graph Contrastive Learning, CIKM2021)



- Many real-world graphs are dynamic in the sense that they evolve over time.
- Node v's representation will depend on its structural and compositional information, as well as the temporal information, and its representation shall be time dependent.
- Some methods derived from static graph scenarios are not directly applicable and may even lead to a questionable inference on these dynamic data.

蚂蚁集团 ANT GROUP | DataFun.

安全风控图模型——DDGCL

假设：大多数节点k度邻域子图在短时间内变化具有一致性

$$\widehat{\mathrm{sim}}(x,y) = \sum_{1\le i,j\le d} \mathrm{RELU}\left(\langle \omega_{ij}, \mathrm{TE}(|t_x - t_y|)\rangle\right) x_i y_j$$

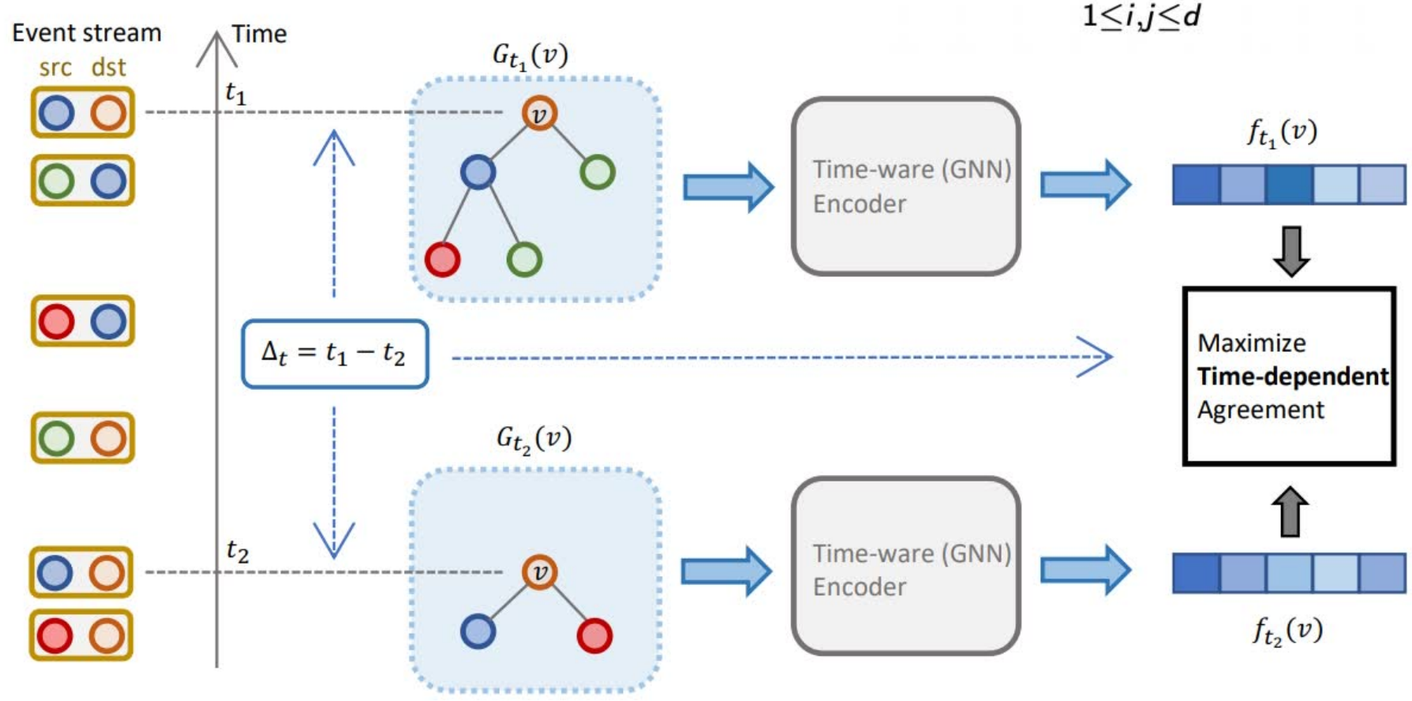


Figure: Building blocks of DDGCL: As a contrastive-based self-supervised framework, there're three ingredients of DDGCL: a Time-Aware GNN encoder, a positive sample reconstruction method, and a novel contrastive loss function.

蚂蚁集团 ANT GROUP | DataFun.

安全风控图模型——DDGCL

DGI(Deep Graph Infomax) Loss:

$$-\mathbb{E}_x\left[\mathbb{E}_{x^+\sim p^+}\log\frac{1}{1+e^{-\mathrm{sim}(f(x^+),f(x))}}+\mathbb{E}_{x^-\sim p^-}\log\frac{1}{1+e^{\mathrm{sim}(f(x^-),f(x))}}\right]$$

DDGCL Loss:

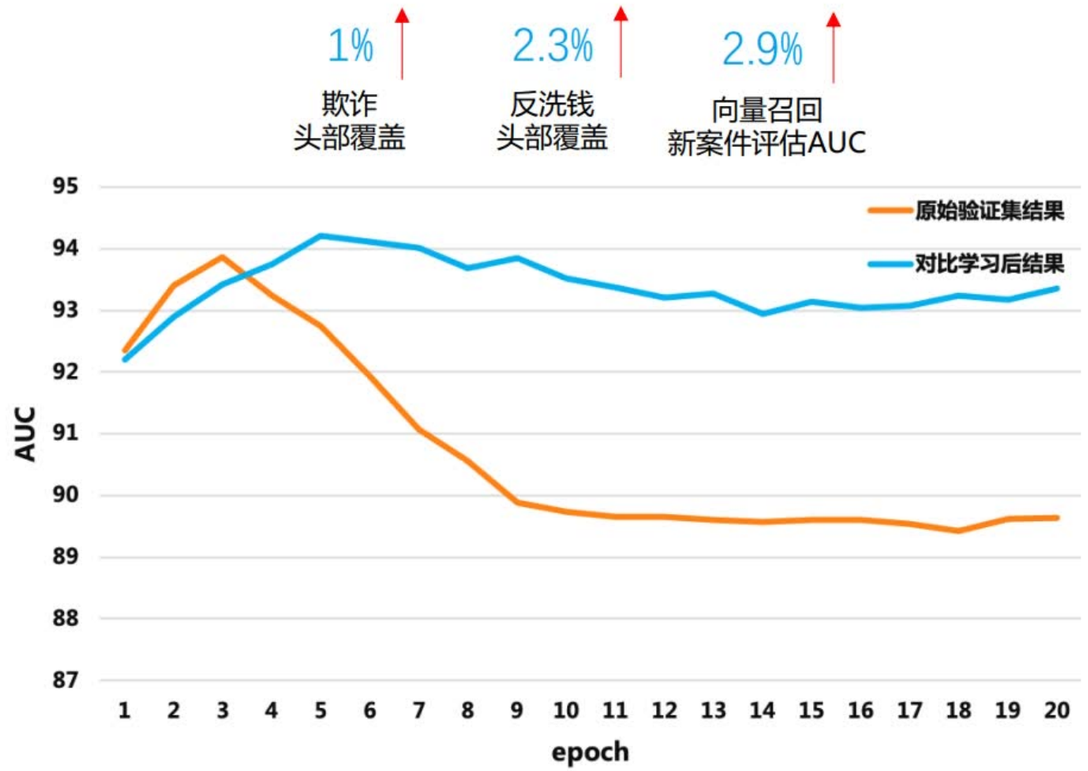
$$\mathcal{L}_{\mathrm{DDGCL}}=-\frac{1}{N}\sum_{x\in\mathbf{X}}\sum_{l=1}^{N_{\mathrm{pos}}}\frac{1}{N_{\mathrm{pos}}}\log\frac{1}{1+e^{-\mathrm{sim}(f(x_l^+),f(x))}}$$
$$-\frac{1}{N(1-\tau^+)}\sum_{x\in\mathbf{X}}\sum_{i=1}^{N_{\mathrm{neg}}}\left(\frac{e^{\beta\mathrm{sim}(f(x_i^-),f(x))}}{\sum_j e^{\beta\mathrm{sim}(f(x_j^-),f(x))}}\right)\log\frac{1}{1+e^{\mathrm{sim}(f(x_i^-),f(x))}}$$
$$+\frac{\tau^+}{N(1-\tau^+)}\sum_{x\in\mathbf{X}}\sum_{i=1}^{N_{\mathrm{pos}}}\left(\frac{e^{\beta\mathrm{sim}(f(x_i^+),f(x))}}{\sum_j e^{\beta\mathrm{sim}(f(x_j^+),f(x))}}\right)\log\frac{1}{1+e^{\mathrm{sim}(f(x_i^+),f(x))}}$$

蚂蚁集团 ANT GROUP | DataFun.

安全风控图模型——DDGCL

Table: Comparison of DDGCL with other previous approaches on graph contrastive learning on the dynamic node classification task

Multi task learning			
	Wikipedia	Reddit	MOOC
Link prediction	84.45 ± 0.8	65.41 ± 0.7	73.38 ± 0.4
GraphCL(on)	87.67 ± 0.9	64.72 ± 0.6	73.48 ± 0.5
GraphCL(nn)	87.86 ± 0.6	65.69 ± 0.7	73.81 ± 0.4
GraphCL(nl)	87.52 ± 0.9	66.13 ± 0.7	73.73 ± 0.4
GraphCL(nm)	**87.89** ± 0.8	**66.16** ± 0.7	73.64 ± 0.6
GraphCL(ns)	87.10 ± 0.6	64.40 ± 0.7	**73.89** ± 0.4
GCC	**88.46** ± 0.6	**69.83** ± 0.8	**73.94** ± 0.3
DDGCL	**89.32** ± 0.5	**71.13** ± 0.8	**74.54** ± 0.2



稳定模型训练，验证集**过拟合缓解**

蚂蚁集团 ANT GROUP | DataFun.

蚂蚁集团 ANT GROUP | DataFun.

04 展望

展望

- 资金链表征进一步学习：MaskGAE、AdaPath；
- 图模型鲁棒性的进一步提升：
  I. 子图去噪、子图预计算；
  II. 子图对抗攻击防御；
  III. DRO；
- 图结构的进一步挖掘；

非常感谢您的观看

蚂蚁集团 ANT GROUP | DataFun.