

# 图网络在社会网络中的应用



中国科学院 信息工程研究所  
INSTITUTE OF INFORMATION ENGINEERING, CAS

宋传承  
2023年4月21日

# 目录

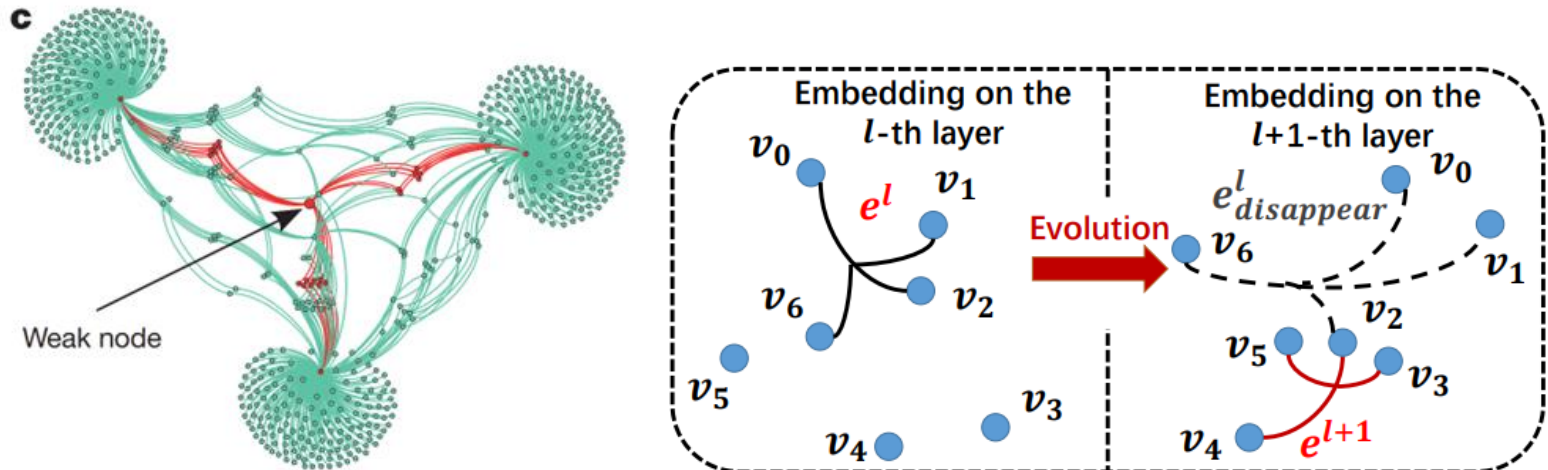
---

- 背景
- 方法论
- 中心问题
  - 孰强孰弱?
  - 孰亲孰远?
  - 孰新孰老?
  - 孰同孰异?
  - 孰胜孰衰?

# 背景

- 社会网络是一个由不同部分发挥各自作用而组成的复杂系统
- 社会网络革命 (social network revolution) 是与移动革命 (mobile revolution) 与互联网革命 (internet revolution) 并列为新时期影响人类社会的三大革命, 其中社会网络革命是最先出现, 却是最不为人关注的。
- “不在于你知道什么, 而在于你认识谁!” ——埃米尔·涂尔干
- “人的本质在其现实性上, 是一切社会关系的总和。” ——马克思

E.g.



# 理解复杂系统的解决问题的方法论

---

## 基本假设

- 社会/自然/技术“系统”由不同“部分”有机组成。
- 不同“部分”间存在某种“结构”。
- 各“部分”间通过“结构”进行交互保证系统的运转。
- 系统内部的个体之间相互依存，并且任何个体的行为结果潜在地依赖于其他个体的联合行为。

## 研究方法

- 确定研究问题
- 建模：构建图网络
- 分析：节点/边/子图/全图的变化
- 归纳：模式、原因、调控手段
- 应用：收集什么数据、进行什么分析、给出何种决策

# 中心问题

- **孰强孰弱**-个体能力分析：网络中哪些节点至关重要，哪些节点可以忽略？比如在某一社区进行广告投放，应该寻找哪个节点（意见领袖KOL发现）扩散广告信息可以获得最大信息传递和回报？
- **孰亲孰远**-群体关系分析：网络中是否存在关系紧密的群体？如果有，则如何发现它们？不同群体之间有何差异？比如在某科学领域中是否存在紧密合作的科学家群体，不同群体有何差异？
- **孰新孰老**-动态网络分析：动态网络是如何发展的？不同时期网络信息有何差异？比如随着红楼梦剧情的推进，哪些人消失了，哪些新角色登场了？演变分析可与个体/群体分析组合获得更多信息。
- **孰同孰异**-实体关系分析：网络之中哪些节点在某些方面相似或哪些节点之间更有可能产生新的连接关系（链路预测），以及同配异配问题。
- **孰胜孰衰**-综合情报决策：综合研判当前已收集的群体、个体、时间情报，概括和推断目标领域的分布格局和发展态势，可以帮助政策制定、产业布局、学术规划、人际交往、公司组织调整等，是网络分析的社会价值的体现。

# 孰强孰弱？个体能力分析





nature  
machine intelligence

ARTICLES

<https://doi.org/10.1038/s42256-020-0177-2>

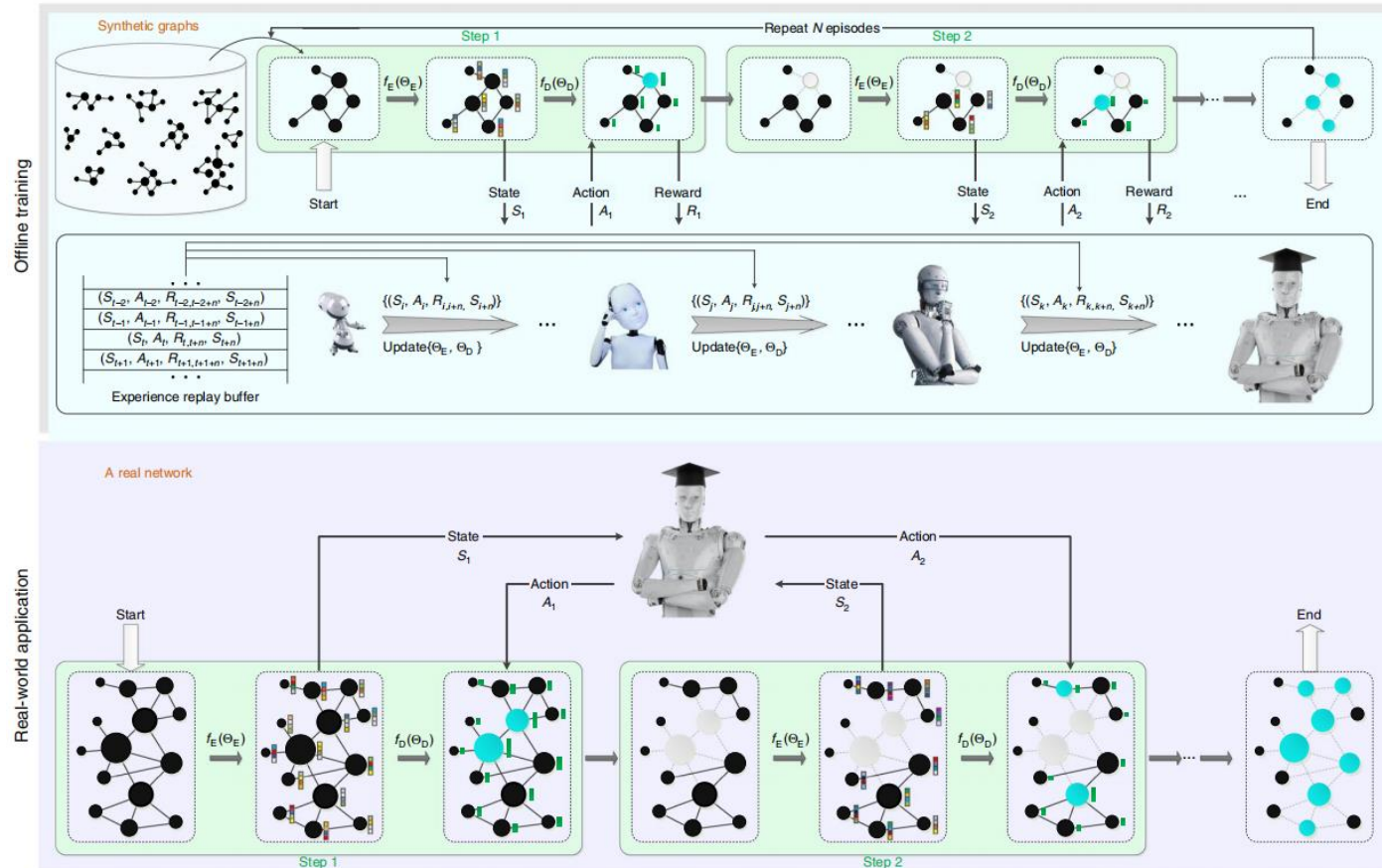


## Finding key players in complex networks through deep reinforcement learning

Changjun Fan<sup>1,2</sup>, Li Zeng<sup>1</sup>, Yizhou Sun<sup>2</sup>   and Yang-Yu Liu<sup>3,4</sup>  

- 提出一个深度强化学习统一框架FINDER。大量实验表明，FINDER在各方面的性能明显优于现有方法。
- 节点攻击策略
- 强化学习
- 编码-解码结构

# 孰强孰弱？个体能力分析

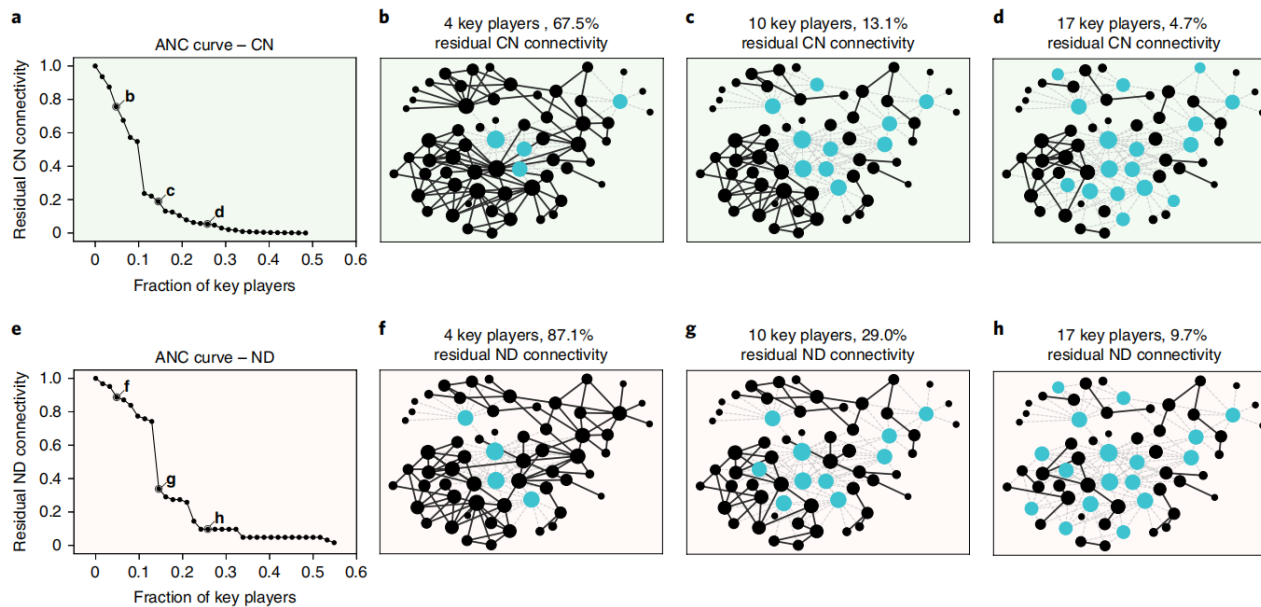


- 深度强化学习统一框架FINDER
- 节点攻击策略、强化学习原理、编码-解码结构

# 孰强孰弱？个体能力分析

## 节点攻击策略

- 不同于度中心性等静态方法，动态地评价节点重要性



- Accumulated Normalized Connectivity (ANC)

$$R(v_1, v_2, \dots, v_N) = \frac{1}{N} \sum_{k=1}^N \frac{\sigma(\mathcal{G} \setminus \{v_1, v_2, \dots, v_k\})}{\sigma(\mathcal{G})} \quad R_{\text{cost}}(v_1, v_2, \dots, v_N) = \sum_{k=1}^N \frac{\sigma(\mathcal{G} \setminus \{v_1, v_2, \dots, v_k\})}{\sigma(\mathcal{G})} c(v_k)$$

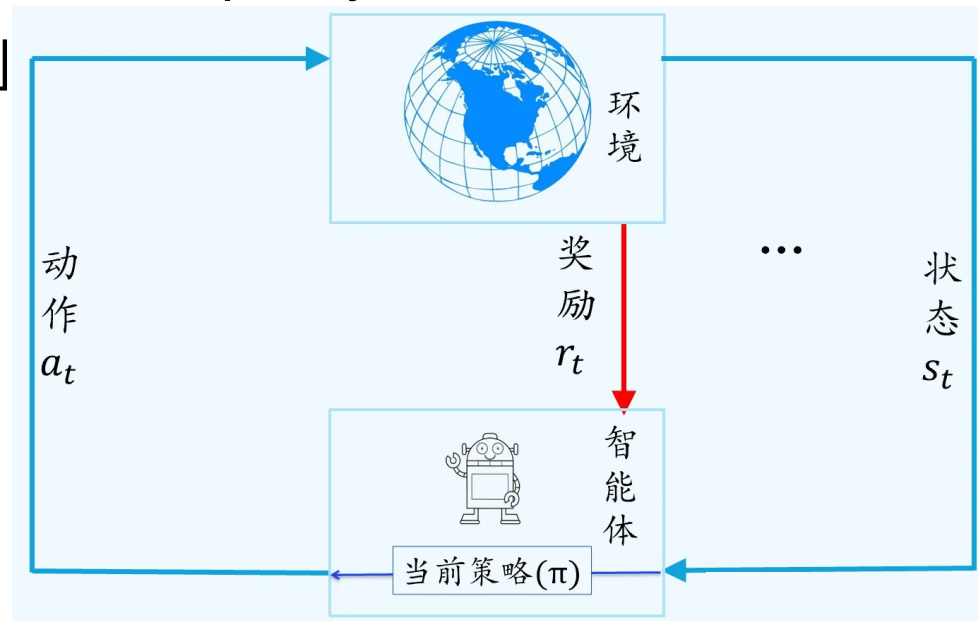


# 孰强孰弱？个体能力分析

## 强化学习原理

- 强化学习讨论的问题是一个「智能体(agent)」 怎么在一个「环境(environment)」 通过一定「动作(action)」 进行「状态(state)」 转移的过程中学习到一个「策略(policy)」，从而最大化它能获得的累计「奖励(Reward)」

- 环境：图网络
- 动作：攻击节点
- 状态：剩余网络
- 奖励：ANC值的降低
- 策略：怎么攻击节点



# 孰强孰弱？个体能力分析

## 强化学习原理

- 强化学习通过马尔科夫决策过程 (Markov Decision Process, MDP) 进行建模, 使用五元组  $(S, A, P, R, \gamma)$  表示。
- 在一个MDP过程中, 智能体的目标是学习到一个策略  $\pi$ , 用以指导在每一个状态  $s_t$  下, 采取动作  $a_t$ 。策略  $\pi$  是概率的集合或分布, 其元素  $\pi(a|s)$  为对过程中的某一状态  $s$  采取可能的行为  $a$  的概率:

$$\pi(a|s) = P\{a_t = a | s_t = s\}$$

- 一个策略完整定义了个体的行为方式, 也就是说定义了个体在各个状态下的各种可能的行为方式以及其概率的大小。
- 转移概率和奖励均仅和当前的状态有关, 与历史信息无关 (马尔科夫性)。

$$P\{s_{t+1} | s_0, a_0, \dots, s_t, a_t\} = P\{s_{t+1} | s_t, a_t\}$$

$$R\{s_{t+1} | s_0, a_0, \dots, s_t, a_t\} = R\{s_{t+1} | s_t, a_t\} = R\{s_t, a_t, s_{t+1}\} = r_t \in \mathbb{R}$$

# 孰强孰弱？个体能力分析

## 强化学习原理

- 假定在离散的时间域  $t \in \{0, 1, 2, \dots\}$  上，智能体从一个初始状态  $s_0$  出发，在每一个时间点  $t$  上，智能体观测到环境状态为  $s_t$ ，采取动作  $a_t$ ，得到即时奖励  $r_t$ 。将这个过程的记录为：

- 轨迹 (Trajectory)  $\tau = (s_0, a_0, r_0, s_1, a_1, r_1, \dots)$

$$\pi(a|s) = P\{a_t = a | s_t = s\}$$

$$R\{s_{t+1} | s_0, a_0, \dots, s_t, a_t\} = R\{s_{t+1} | s_t, a_t\} = R(s_t, a_t, s_{t+1}) = r_t \in \mathbb{R}$$

- 学习目标：找到最优的策略  $\pi^*$  将累计奖励最大化

$$r_0 + \gamma r_1 + \gamma^2 r_2 + \dots = \sum_{t \geq 0} \gamma^t r_t$$

# 孰强孰弱？个体能力分析

## 强化学习原理

- 学习目标：找到最优的策略将累计奖励最大化

$$r_0 + \gamma r_1 + \gamma^2 r_2 + \dots = \sum_{t \geq 0} \gamma^t r_t$$

- 策略价值函数：从状态 $s$ 开始，遵循当前策略 $\pi$ 所获得的收获的期望；或者说在执行当前策略 $\pi$ 时，衡量个体处在状态 $s$ 时的价值大小

$$V^\pi(s) = E_\pi \left[ \sum_{k \geq 0} \gamma^k r_{t+k} \mid s_t = s, \pi \right]$$

- 策略行为价值函数：在执行策略 $\pi$ 时，对当前状态 $s$ 执行某一具体行为 $a$ 所能的到的收获的期望；或者说在遵循当前策略 $\pi$ 时，衡量对当前状态执行行为 $a$ 的价值大小

$$Q^\pi(s, a) = E_\pi \left[ \sum_{k \geq 0} \gamma^k r_{t+k} \mid s_t = s, a_t = a \right]$$

# 孰强孰弱？个体能力分析

## 强化学习原理

- 通过迭代使得行为价值Q增量式更新：时序差分(TD)强化学习
- Q-learning：选择使Q最大的那个迭代方向

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[ r + \gamma \max_a Q(s', a) - Q(s, a) \right]$$

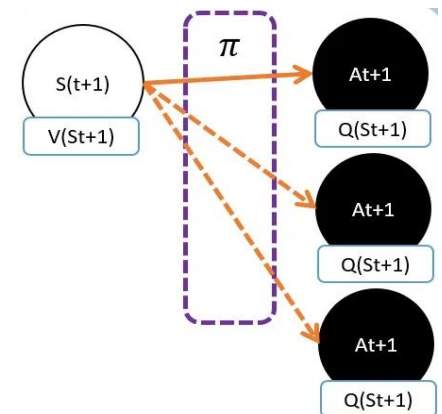
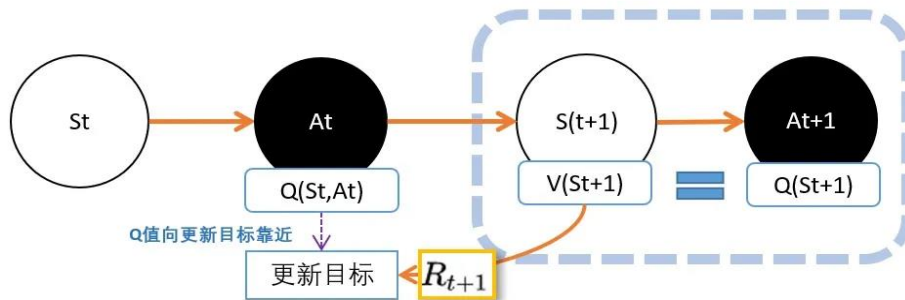
新估计值    旧估计值    当前步的奖励+下一步的价值    旧估计值  
(目标值/期望)

- 第k步迭代时:

$$Q_k^\pi(s_t, a_t) = Q_{k-1}^\pi(s_t, a_t) + \frac{1}{k} (R_k(s_t, a_t) - Q_{k-1}^\pi(s_t, a_t))$$

$$R_k(s_t, a_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots$$

$$= r_t + \gamma (r_{t+1} + \gamma r_{t+2} + \dots) = r_t + \gamma Q_{k-1}^\pi(s_{t+1}, a_{t+1})$$



# 孰强孰弱？个体能力分析

## 强化学习原理

- 通过迭代使得行为价值Q增量式更新：时序差分(TD)强化学习
- Q-learning：选择使Q最大的那个迭代方向

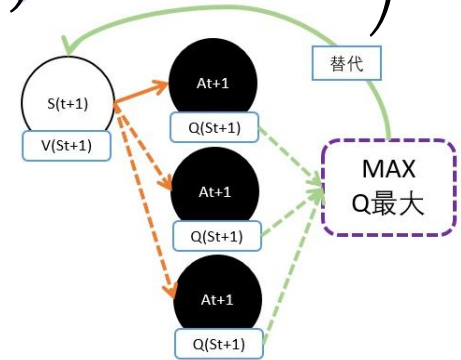
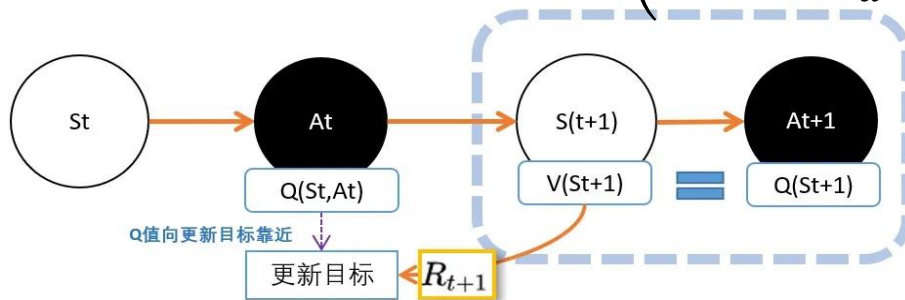
$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[ r + \gamma \max_a Q(s', a) - Q(s, a) \right]$$

新估计值    旧估计值    当前步的奖励+下一步的价值 (目标值/期望)    旧估计值

- 使用Q-learning的思想优化至max:

$$Q_k^\pi(s_t, a_t) = Q_{k-1}^\pi(s_t, a_t) + \frac{1}{k} \left( (r_t + \gamma Q_{k-1}^\pi(s_{t+1}, a_{t+1})) - Q_{k-1}^\pi(s_t, a_t) \right)$$

$$Q_k^*(s_t, a_t) = Q_{k-1}^*(s_t, a_t) + \frac{1}{k} \left( (r_t + \gamma \max_{a'} Q_{k-1}^*(s_{t+1}, a')) - Q_{k-1}^*(s_t, a_t) \right)$$



# 孰强孰弱？个体能力分析

## 强化学习原理

- 通过迭代使得行为价值Q增量式更新：时序差分(TD)强化学习
- Q-learning：选择使Q最大的那个迭代方向

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[ r + \gamma \max_a Q(s', a) - Q(s, a) \right]$$

新估计值    旧估计值    当前步的奖励+下一步的价值  
(目标值/期望)    旧估计值

- 当k迭代至无穷时， $Q_k^*(s_t, a_t)$ 收敛接近定值 $Q_{true}(s_t, a_t)$ ：

$$Q_k^*(s_t, a_t) = Q_{k-1}^*(s_t, a_t) + \frac{1}{k} \left( \left( r_t + \gamma \max_{a'} Q_{k-1}^*(s_{t+1}, a') \right) - Q_{k-1}^*(s_t, a_t) \right)$$

$$Q_k^*(s_t, a_t) - Q_{k-1}^*(s_t, a_t) = \frac{1}{k} \left( \left( r_t + \gamma \max_{a'} Q_{k-1}^*(s_{t+1}, a') \right) - Q_{k-1}^*(s_t, a_t) \right)$$

Loss

$$Q_{true}(s_t, a_t) - Q_{pred}(s_t, a_t) = \left( \left( r_t + \gamma \max_a \hat{Q}(s_{t+1}, a; \hat{\theta}) \right) - Q(s_t, a_t; \theta) \right)$$

# 孰强孰弱？个体能力分析

## 强化学习原理

- Loss函数

$$Q_{true}(s_t, a_t) - Q_{pred}(s_t, a_t) = \left( \left( r_t + \gamma \max_a \hat{Q}(s_{t+1}, a; \hat{\Theta}) \right) - Q(s_t, a_t; \Theta) \right)$$

- 延迟Loss函数：考察Q-learning

$$\left( r_{t,t+n} + \gamma \max_a \hat{Q}(s_{t+n}, a; \hat{\Theta}) \right) - Q(s_t, a_t; \Theta)$$

- 图重构Loss函数：考察图结构

$$\alpha \sum_{i,j=1}^N s_{ij} \|y_i - y_j\|_2^2$$

*s<sub>ij</sub>* 的 embedding 向量  
示性和 *j* 是否邻接

- 总Loss函数

$$\text{Loss}(\Theta_Q, \Theta_{\text{Embed}}) = \underbrace{\mathbb{E}_{(s_t, a_t, r_{t,t+n}, s_{t+n}) \sim U(D)} [(r_{t,t+n} + \gamma \max_{a'} \hat{Q}(s_{t+n}, a'; \hat{\Theta}_Q) - Q(s_t, a_t; \Theta_Q))^2]}_{\text{Q-learning loss}}$$

$$\Theta_E = \{W_1, W_2, W_3\}$$

$$\Theta_D = \{W_4, W_5\}$$

$$+ \alpha \underbrace{\sum_{i,j=1}^N s_{ij} \|y_i - y_j; \Theta_{\text{Embed}}\|_2^2}_{\text{Graph reconstruction loss}}$$

$$\Theta_Q = \{\Theta_E, \Theta_D\}, \Theta_{\text{Embed}} = \Theta_E$$



# 孰强孰弱？个体能力分析

## 编码-解码结构

- 编码：类GraphSAGE模型

---

**Algorithm S2** Encoding process in FINDER

---

**Input:** Graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$ , input features  $\{X_v \in \mathbf{R}^{1 \times c}, \forall v \in \mathcal{V}\}$ , depth  $K$ , weight parameters  $W_1 \in \mathbf{R}^{c \times p}$ ,  $W_2 \in \mathbf{R}^{p \times (p/2)}$ ,  $W_3 \in \mathbf{R}^{p \times (p/2)}$

**Output:** Embedding vector  $z_v, \forall v \in \mathcal{V} \cup \{s\}$

1: Create a virtual node  $s$  which connects all nodes in the graph, denoted as the graph state

2: Initialize  $h_v^{(0)} \leftarrow \text{ReLU}(X_v \cdot W_1)$ ,  $h_v^{(0)} \leftarrow h_v^{(0)} / \|h_v^{(0)}\|_2, \forall v \in \mathcal{V} \cup \{s\}$

3: **for**  $l = 1$  to  $K$  **do**

4: **for**  $v \in \mathcal{V} \cup \{s\}$  **do**

5:  $h_{\mathcal{N}(v)}^{(l-1)} \leftarrow \sum_{j \in \mathcal{N}(v)} h_j^{(l-1)}$

6:  $h_v^{(l)} \leftarrow \text{ReLU}([W_2 \cdot h_v^{(l-1)}, W_3 \cdot h_{\mathcal{N}(v)}^{(l-1)}])$

7: **end for**

8:  $h_v^{(l)} \leftarrow h_v^{(l)} / \|h_v^{(l)}\|_2, \forall v \in \mathcal{V} \cup \{s\}$

9: **end for**

10:  $z_v \leftarrow h_v^{(K)}, \forall v \in \mathcal{V} \cup \{s\}$

---

# 孰强孰弱？个体能力分析

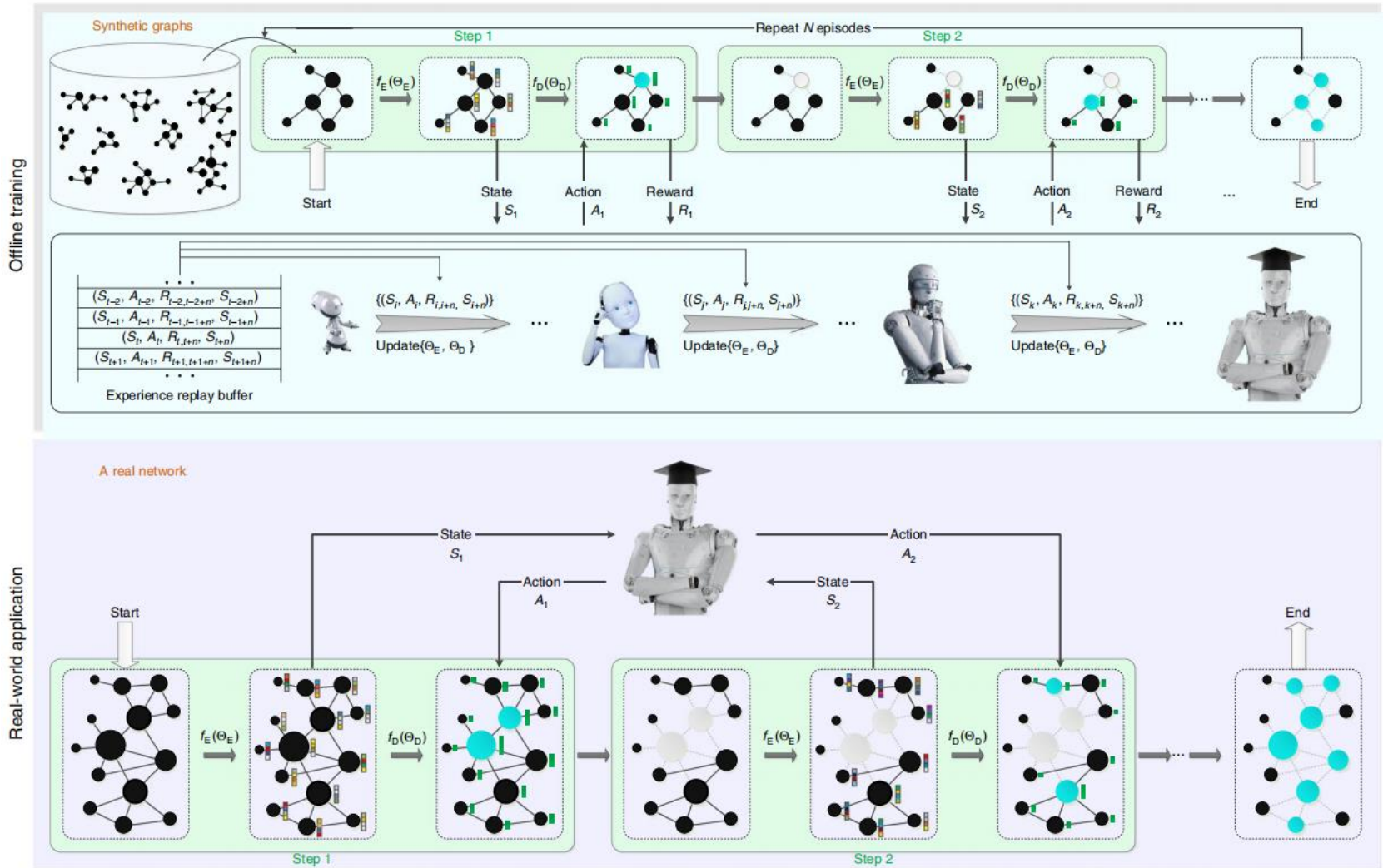
## 编码-解码结构

- 解码：学习一个映射使得state-action对 $(s, a)$ 映射到标量值 $Q(s, a)$ ，该Q值表示从状态 $s$ 选择动作 $a$ 以来最大的收益

$$Q(s, a) = \mathbf{W}_5^T \text{ReLU}(\mathbf{z}_a^T \cdot \mathbf{z}_s \cdot \mathbf{W}_4)$$

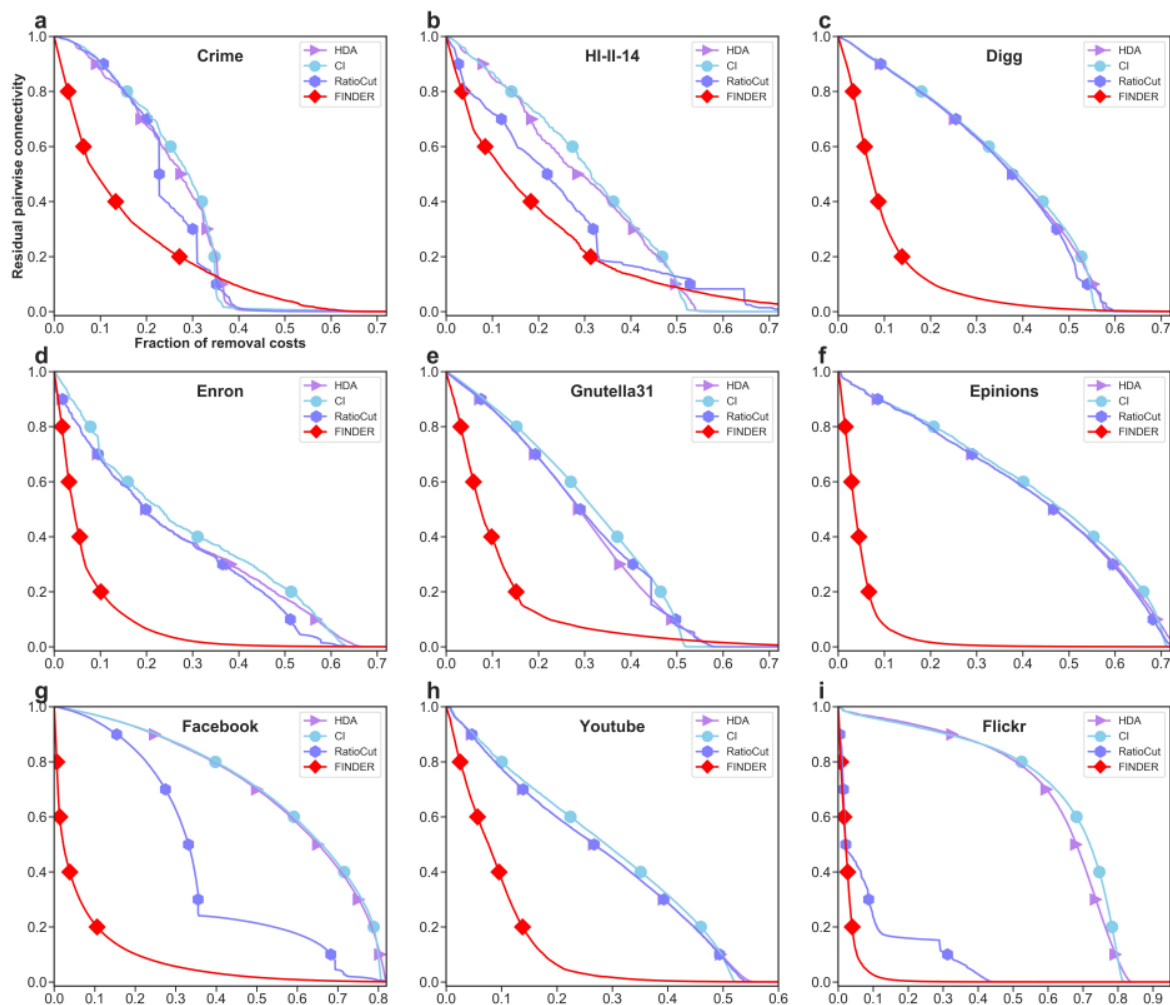
- 这里 $\mathbf{W}_4 \in \mathbb{R}^{p \times 1}$ ,  $\mathbf{W}_5 \in \mathbb{R}^{p \times 1}$ 是权重参数,  $\mathbf{z}_s \in \mathbb{R}^{1 \times p}$ ,  $\mathbf{z}_a \in \mathbb{R}^{1 \times p}$ 是编码的输出, for state and action respectively。

# 孰强孰弱？个体能力分析





# 孰强孰弱？个体能力分析

## 实验结果



# 孰亲孰远？群体关系分析

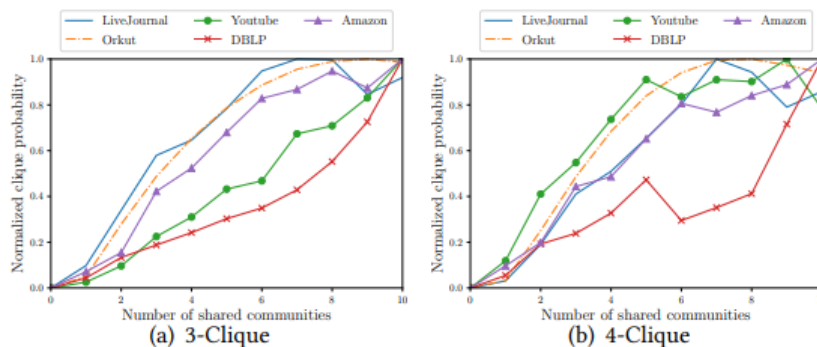
## CommunityGAN: Community Detection with Generative Adversarial Nets

Authors:  [Yuting Jia](#),  [Qinqin Zhang](#),  [Weinan Zhang](#),  [Xinbing Wang](#) [Authors Info & Claims](#)

- 通过真实数据集上的统计研究，验证了clique与community的强相关性，并基于GAN思路建模，结合AGM (Affiliation Graph Model) 来同步完成社区发现和节点表示学习任务——用clique扩展AGM模型，基于GAN进行社区发现。
- 当前缺少对重叠社区发现任务的建模；数据分析发现团和社区结构常同时出现。

Table 2: The occurrence probability of cliques for vertices sampled from all vertices or from one community. R: from all vertices. C: from one community.

Dataset	2-Clique		3-Clique		4-Clique	
	R	C	R	C	R	C
LiveJournal	4E-6	0.80	2E-11	0.18	0	0.08
Orkut	2E-5	0.82	8E-11	0.18	0	0.06
Youtube	4E-6	0.73	0	0.10	0	0.02
DBLP	1E-5	0.52	1E-10	0.31	0	0.23
Amazon	6E-6	0.53	0	0.18	0	0.05



- 表明节点集共享社区越多，越有可能形成团；此前AGM中节点对与共享社区关系的数据分析，可以理解为2-Clique，本文进一步研究3,4-Clique与共享社区的关系

# 孰亲孰远？群体关系分析

## 建模方法

- 基于GAN进行建模，生成器G和判别器D迭代学习（此处引入强化学习中常用的policy gradient来训练）。其中
- G用于生成最像clique的节点子集（也可理解为从图中抽取节点子集，并且该子集尽可能像clique）。
- D则判别上述节点子集是否在图中为clique。
- 上述目标形式化表示为：

$$\min_{\theta_G} \max_{\theta_D} V(G, D) = \sum_{c=1}^V \left( \mathbb{E}_{m \sim p_{true}(\cdot | v_c)} [\log D(m; \theta_D)] + \mathbb{E}_{s \sim G(s | v_c; \theta_G)} [\log(1 - D(s; \theta_D))] \right).$$

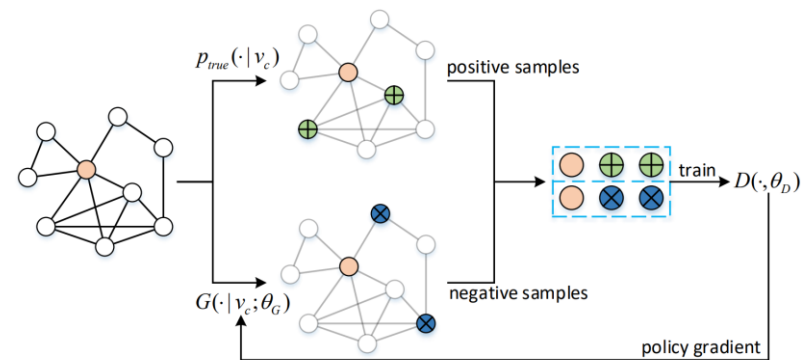


Figure 3: Framework of CommunityGAN

# 孰亲孰远？群体关系分析

## 生成器G和判别器D

- 判别器D：连乘经过sigmoid函数后的节点向量内积

$$D(s) = \prod_{(u,v) \in s, u \neq v} \sigma(\mathbf{d}_u^\top \cdot \mathbf{d}_v)$$

- 生成器G：通过从 $v_{s_1}$ 到 $v_{s_{m-1}}$ 的长序列建模来生成 $v_{s_m}$ ，最后得到生成的节点子集

$$\begin{aligned} G(s|v_c) \\ = G_v(v_{s_2} | v_{s_1}) G_v(v_{s_3} | v_{s_1}, v_{s_2}) \cdots G_v(v_{s_m} | v_{s_1}, \dots, v_{s_{m-1}}) \end{aligned}$$

- 最直接的方法，通过softmax来完成上述采样计算

$$G_v(v_{s_m} | v_{s_1}, \dots, v_{s_{m-1}}) = \frac{\exp(\sum_{i=1}^{m-1} \mathbf{g}_{v_{s_m}}^\top \mathbf{g}_{v_{s_i}})}{\sum_{v \notin (v_{s_1}, \dots, v_{s_{m-1}})} \exp(\sum_{i=1}^{m-1} \mathbf{g}_v^\top \mathbf{g}_{v_{s_i}})}$$

# 孰亲孰远？群体关系分析

## AGM思想

- GAN建模有一定缺陷：①还未能实现社区发现；②softmax计算开销大；③softmax忽略图结构信息。
- 用一个非负矩阵 $F$ 描述每个节点与社区的隶属关系，每行表示该行对应节点属于各个社区的概率。有了 $F$ 后，生成新图 $G'$ ，对于节点对 $(u, v)$ 间是否有边，依照如下公式计算概率：

$$p(u, v) = 1 - \exp(-F_u^\top \cdot F_v)$$

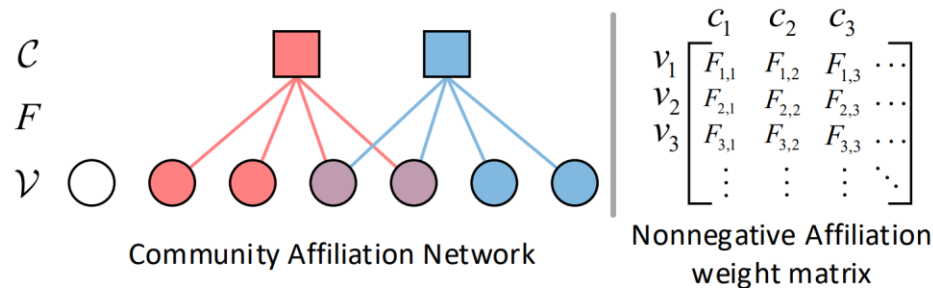


Figure 4: AGM framework.  $\mathcal{V}$ : vertices;  $\mathcal{C}$ : communities;  $F$ : affiliation of vertices to communities.



# 孰亲孰远？群体关系分析

## AGM思想

$$p(u, v) = 1 - \exp(-F_u^\top \cdot F_v)$$

- 将上式扩展为“对于节点集  $\{v_1, v_2, \dots, v_m\}$ ，能形成clique的概率”：

$$\begin{aligned} p(v_1, v_2, \dots, v_m) &= 1 - \prod_c (1 - p_c(v_1, v_2, \dots, v_m)) \\ &= 1 - \exp(-\odot(F_{v_1}, F_{v_2}, \dots, F_{v_m})), \end{aligned}$$

- 自然而然，可以用上式作为判别器D，即

$$D(s) = 1 - \exp(-\odot(\mathbf{d}_{v_1}, \mathbf{d}_{v_2}, \dots, \mathbf{d}_{v_m}))$$

- 生成器G中每个节点的采样也可基于上式，使用softmax计算采样概率：

$$\begin{aligned} G_v(v_{s_m} | v_{s_1}, \dots, v_{s_{m-1}}) &= \frac{1 - \exp(-\odot(\mathbf{g}_{v_{s_1}}, \dots, \mathbf{g}_{v_{s_m}}))}{\sum_{v \notin (v_{s_1}, \dots, v_{s_{m-1}})} 1 - \exp(-\odot(\mathbf{g}_{v_{s_1}}, \dots, \mathbf{g}_{v_{s_{m-1}}}, \mathbf{g}_v))} \end{aligned}$$

# 孰亲孰远？群体关系分析

## 实验结果

F1-Score

Model	Amazon	Youtube	DBLP
MMSB	0.366	0.124	0.104
CPM	0.214	0.058	0.318
AGM	0.711	0.158	0.398
node2vec	0.550	0.288	0.265
LINE	0.532	0.170	0.208
GraphGAN	0.518	0.303	0.276
ComE	0.562	0.213	0.240
CommunityGAN	<b>0.860</b>	<b>0.327</b>	<b>0.456</b>




NMI

Model	Amazon	Youtube	DBLP
MMSB	0.068	0.031	0.000
CPM	0.027	0.000	0.066
AGM	0.635	0.025	0.059
node2vec	0.370	0.071	0.068
LINE	0.248	0.070	0.027
GraphGAN	0.417	0.049	0.083
ComE	0.413	<b>0.091</b>	0.059
CommunityGAN	<b>0.853</b>	<b>0.091</b>	<b>0.153</b>

- 与非重叠或稀疏重叠的社区检测的baseline基线相比，考虑密集社区重叠为CommunityGAN提供了更高的学习灵活性。
- 生成器和判别器之间的极大极小博弈驱动CommunityGAN在AGM的基础上获得了显著的性能改进。

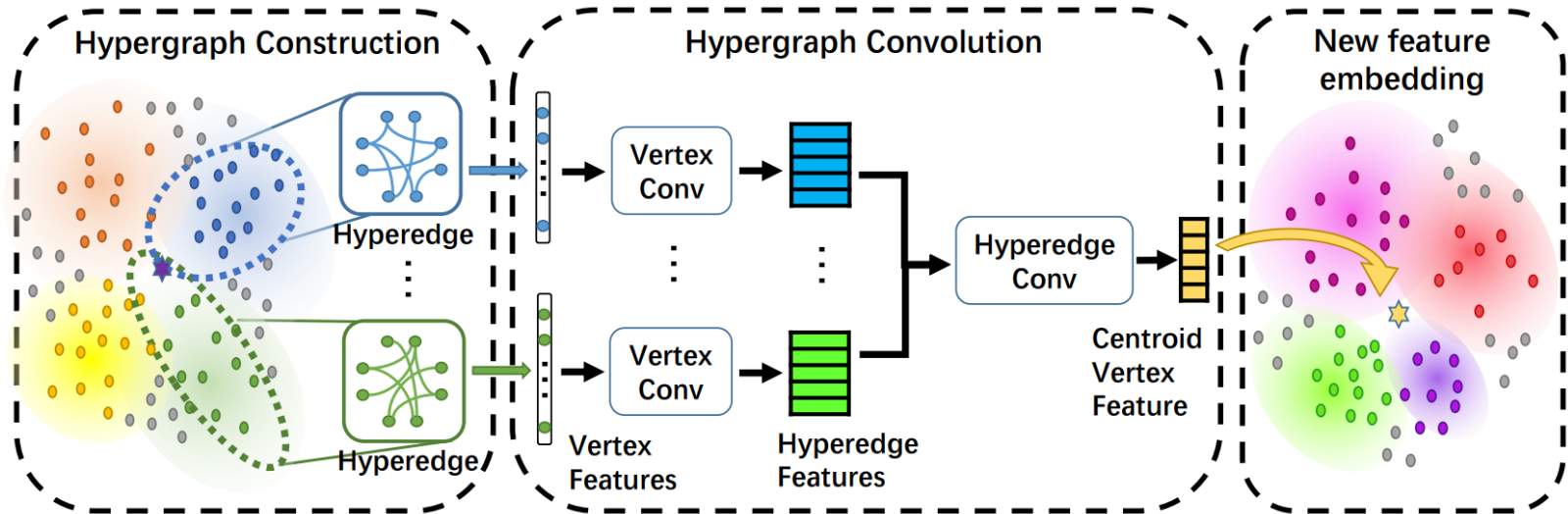
# 孰新孰老？动态网络分析

## Dynamic hypergraph neural networks

Authors:  [Jianwen Jiang](#),  [Yuxuan Wei](#),  [Yifan Feng](#),  [Jingxuan Cao](#),  [Yue Gao](#)

- 提出了一种动态超图构建(DHG)方法，该方法采用k-NN方法生成基本的超边，通过k - means聚类算法扩展相邻超边集。通过动态超图构造方法，可以提取数据局部关系和全局关系。
- 进行了基于引文网络分类和社交媒体情绪预测的实验。在基于网络的任务中，本文提出的方法优于最先进的方法，并且对不同的数据分布有更好的鲁棒性。在社交媒体情绪预测方面，本文提出的方法与最先进的方法相比，表现有所改进。

# 孰新孰老？动态网络分析



- 对于超图中的每一个顶点 $u$ ，先经过动态超图构建过程生成多条超边，然后对每一条超边分别进行顶点卷积，得到该超边的特征，然后对这些超边特征再进行超边卷积，最后得到中心顶点 $u$ 新的特征。整个过程就是一个更新顶点 $u$ 特征的过程，这样就可以根据顶点 $u$ 的性特征构建新的超图，不断循环。

# 孰新孰老? 动态演变分析

## 动态超图构建(DHG)

- 首先根据欧氏距离对每一层的整个特征图进行k-means算法。对每一个输入节点 $u$ , 通过knn算法找到和该节点最相似的 $k-1$ 个节点, 连同顶点 $u$ 形成一个超边; 同时根据k-means的聚类结果将离定点 $u$ 最近的 $S-1$ 簇被赋值为与该顶点相邻的超边。

---

### Algorithm 1 Hypergraph Construction

---

**Input:** Input embedding  $\mathbf{X}$ ; hyperedge size  $k$ ; adjacent hyperedge set size  $S$

**Output:** Hyperedge set  $\mathcal{G}$

**Function:** k-Means clustering  $kMeans$ ; k-nearest neighborhood selection  $knn$ ; distance function  $dis$ ;  $S - 1$  smallest distance index selection  $topK$

```
1:  $C = kMeans(\mathbf{X})$ 
2: for  $u$  in  $range(len(\mathbf{X}))$  do
3:    $e_b = knn(X[u], X, k)$ 
4:    $\mathcal{G}[u].insert(e_b)$ 
5:    $\mathbf{D} = dis(C.center, u)$ 
6:    $\mathbf{D} = sort(D)$ 
7:    $ind = topK(D, S - 1)$ 
8:   for  $i$  in  $ind$  do
9:      $\mathcal{G}[u].insert(C[i])$ 
10:  end for
11: end for
```

---

# 孰新孰老？动态网络分析

## 超图卷积(HGC)

- 超图卷积由两个子模块组成：顶点卷积子模块和超边卷积子模块。顶点卷积将顶点特征集合到超边上，然后超边卷积将相邻的上边缘特征集合到形心顶点。

---

### Algorithm 2 Hypergraph Convolution

---

**Input:** Input sample  $\mathbf{x}_u$ ; hypergraph structure  $\mathcal{G}$

**Output:** Output sample  $\mathbf{y}_u$

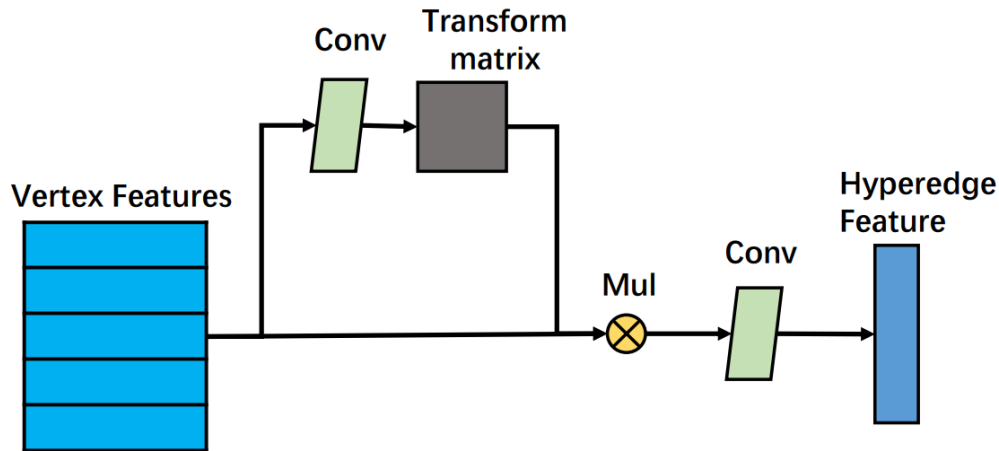
```
1:  $\mathbf{xlist} = \Phi$ 
2: for  $e$  in  $Adj(u)$  do
3:    $\mathbf{X}_v = VertexSample(\mathbf{X}, \mathcal{G})$ 
4:    $\mathbf{x}_e = VertexConv(\mathbf{X}_v)$ 
5:    $\mathbf{xlist.insert}(\mathbf{x}_e)$ 
6: end for
7:  $\mathbf{X}_e = stack(\mathbf{xlist})$ 
8:  $\mathbf{x}_u = EdgeConv(\mathbf{X}_e)$ 
9:  $\mathbf{y}_u = \sigma(\mathbf{x}_u \mathbf{W} + \mathbf{b})$ 
```

---

# 孰新孰老？动态网络分析

## 顶点卷积

- 节点卷积时通过构建一个  $k * k$  的 transform matrix, 来将节点的特征维度压缩到  $k$  维大小。通过 transform matrix 和节点特征相乘来对同一个 hyperedge 内节点的相互关系进行建模和表示。最后经过一个卷积操作进行维度的压缩, 得到和包含这些节点的边的特征。



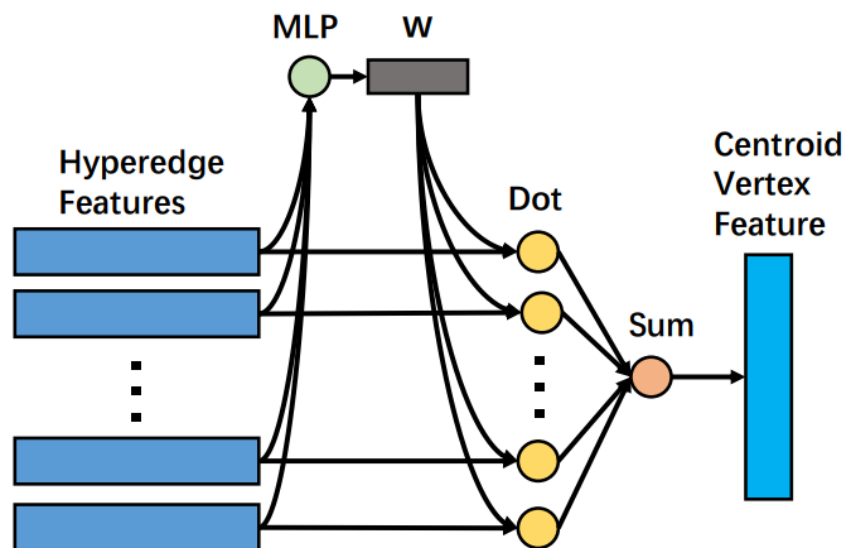
$$\mathbf{T} = MLP(\mathbf{X}_u)$$

$$\mathbf{x}_e = conv(\mathbf{T} \cdot MLP(\mathbf{X}_u))$$

# 孰新孰老？动态网络分析

## 超边卷积

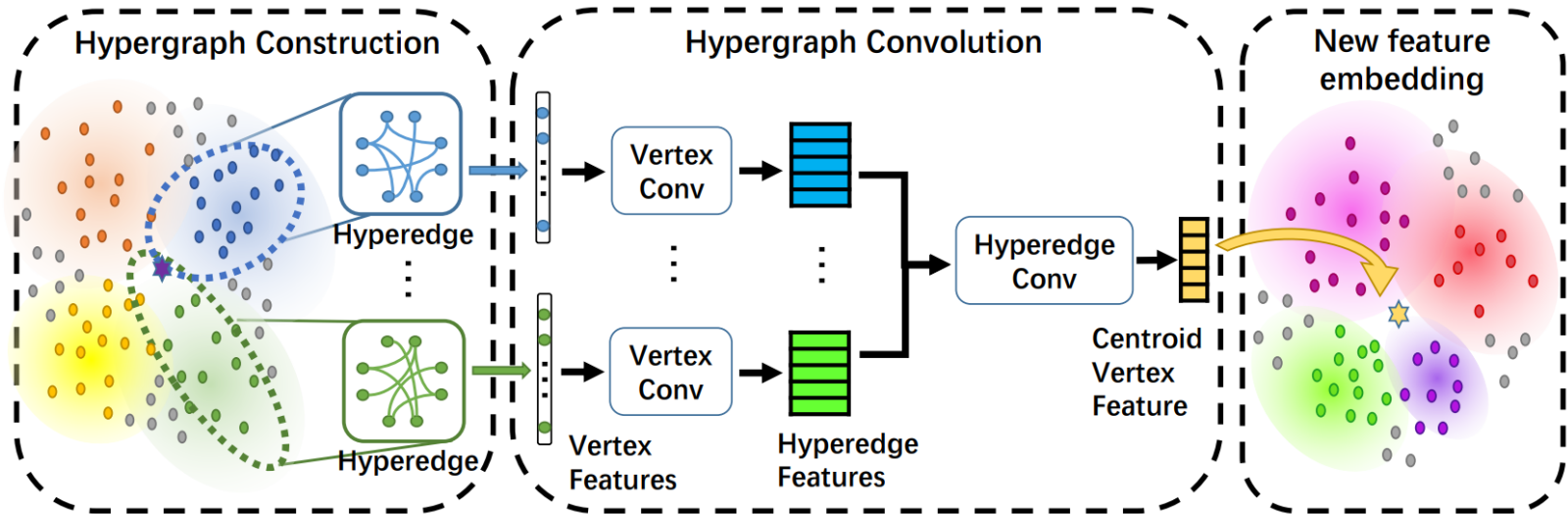
- 将超边特征通过卷积进行聚合得到新的形心顶点特征。超边卷积首先利用MLP生成每个超边的权值，然后再根据得到的权重和超边特征进行加权求和，得到输出的形心顶点特征。



$$w = \text{softmax}(\mathbf{x}_e \mathbf{W} + \mathbf{b})$$
$$\mathbf{x}_u = \sum_{i=0}^{|\text{Adj}(u)|} w^i \mathbf{x}_e^i$$



# 孰新孰老？动态网络分析



- 对于超图中的每一个顶点 $u$ ，先经过动态超图构建过程生成多条超边，然后对每一条超边分别进行顶点卷积，得到该超边的特征，然后对这些超边特征再进行超边卷积，最后得到中心顶点 $u$ 新的特征。整个过程就是一个更新顶点 $u$ 特征的过程，这样就可以根据顶点 $u$ 的性特征构建新的超图，不断循环。

# 孰新孰老？动态网络分析

## 实验结果：引文网络实验

- 对公共基准引文网络数据集Cora进行了一系列的实验。Cora 数据集具有固定的图结构，每个节点代表一篇学术文章，关系表示文章之间的相互引用，每个顶点都有一个词包特征向量和一个类别标签，表明论文所属的主题。通过随机抽取不同比例的数据作为训练集来验证方法的有效性，实验结果如下：

lr	#train	GCN	HGNN	GAT	DHGNN
std	140	81.5%	81.6%	<b>83.0%</b>	82.5%
2%	54	69.6%	75.4%	74.8%	<b>76.9%</b>
5.2%	140	77.8%	79.7%	79.4%	<b>80.2%</b>
10%	270	79.9%	80.0%	81.5%	<b>81.6%</b>
20%	540	81.4%	80.1%	83.5%	<b>83.6%</b>
30%	812	81.9%	82.0%	84.5%	<b>85.0%</b>
44%	1200	82.0%	81.9%	85.2%	<b>85.6%</b>

- DHGNN算法对比其他方法有一些提升，并且在训练集较小时提升效果比较明显。

# 孰新孰老？动态网络分析

## 实验结果：微博情感预测

- 微博数据集是没有固定图结构的数据集，含有 5550 条推文，其中 4196 条为正向情感，1345 条为负向情感。每个推文都包含文字、图片、表情三个模态，实验中分别为三个模态构建了三个超边集，并将这些超边集合并为一个多模态超边集。

Method	Acc	Train Time
CBM-NB [Wang <i>et al.</i> , 2014]	71.6%	-
CBM-LR [Wang <i>et al.</i> , 2014]	79.9%	-
CBM-SVM [Wang <i>et al.</i> , 2014]	81.6%	-
HGNN [Feng <i>et al.</i> , 2018]	86.8%	2m11s
MHG_noW [Chen <i>et al.</i> , 2015]	87.3%	-
MHG [Chen <i>et al.</i> , 2015]	88.6%	-
MMHG [Chen <i>et al.</i> , 2015]	88.7%	-
Bi-MHG [Ji <i>et al.</i> , 2019]	90.0%	58.5h
DHGNN (our method)	<b>91.8%</b>	1m32s

- 与目前最先进的的方法相比，DHGNN在多模式情绪预测任务中、获得了1.8%的准确性提升，同时大大缩短了训练时间。

# 孰胜孰衰？综合情报决策

## 战场情报

- 国防科技大学：战场目标实体类型识别的鲁棒图神经网络方法
- 动态时间规整(Dynamic Time Warping, DTW)算法，来计算不同实体间轨迹的相关度。DTW 算法早期被用于判别两段语音相似度，其优势在于可以对时间序列的延展或压缩具有一定的适应能力。DTW 方法同样可以应用到战场作战实体轨迹的相似度衡量。

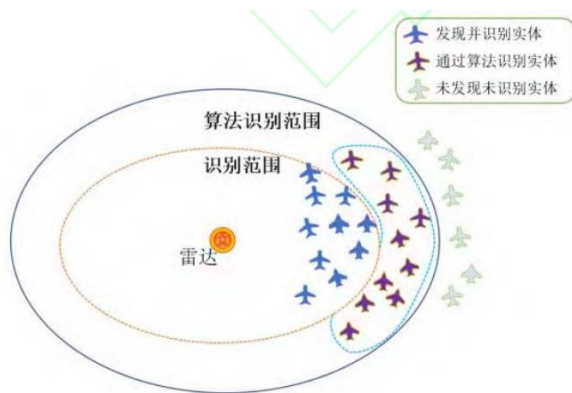


表 1 某时刻  $t_i$  获取到的战场态势数据  $K_i$

Table 1 The situation data of battlefield  $K_i$  in  $t_i$

编号 $i$	位置 $L_i$	速度	航向	类型
1	(-3151,-12556,100)	166	187	歼击机
2	(-4751,-2816,743)	187	216	歼击机
3	(3869,-2111,673)	193	189	未知
4	(-3151,-5367,118)	80	156	未知
5	(4651,-3576,23)	68	142	轰炸机
...	...	...	...	...

### 算法 1 DTW 图模型抽取算法

输入: 态势数据  $\{K_1, K_2, \dots, K_m\}$

参数:  $\alpha$ : 参数用来调节连接边数量的阈值

输出: 图结构数据  $G = (V, E, X)$

- 1: 从态势数据中提取当前时刻所有节点的轨迹时间序列  $\text{Tr}(n)$ ,  $n = 1, 2, \dots, N$ , 以及特征矩阵  $X$
- 2: 借助 DTW 算法, 计算所有节点之间的 DTW 距离:  $\text{DTW}(i, j)$ ,  $i, j = 1, 2, \dots, N$ , 记 DTW 距离构成集合为  $\mathcal{D}$
- 3: 对所有节点的 DTW 距离进行 Min-Max 归一化  $\text{dtw}(i, j) = \frac{\text{DTW}(i, j) - \min \mathcal{D}}{\max \mathcal{D} - \min \mathcal{D}}$
- 4: 连接节点归一化距离小于  $\alpha$  的节点, 即令  $E(i, j) = \mathbf{1}_{\text{dtw}(i, j) < \alpha}$ , 输出图数据  $G = (V, E, X)$ 。

# 孰胜孰衰？综合情报决策

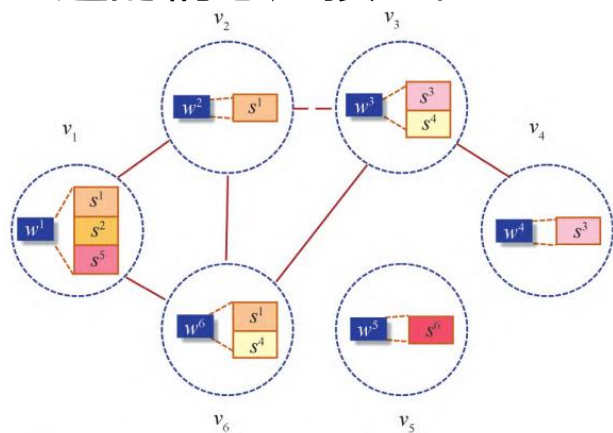
## 开源情报

- 美国战略情报局柏林站站长阿·杜勒(A·Dulles)：“最终情报产品80%的内容应来自开源信息资源的开发和利用”。
- 《美国陆军战场手册 2-22.9 公开源情报》手册中列举的开源情报五大类渠道：学术机构、政府组织/政府间组织/非政府组织、商业/公共信息服务机构、图书馆/研究中心、个人/团体。现今，这五大类渠道无一例外都成为了社交网络的用户，而且数量巨大。
- 全球最大的社会化媒体专业传播公司 WeAreSocial 发布的《2022年数字、社交网络及移动报告》数据显示，全球有51.6亿互联网用户，47.6亿社交媒体用户，占全球总人口的59.4%，全球社交媒体用户同比增长3.0%。
- 以全球最大的社交网站Facebook为例，截止2023年1月活跃用户已达29.63亿，语言多达75种。政府、媒体、企业和学术机构等专业内容生产者大量入驻极大提高了社交网络开源情报的质量。

# 孰胜孰衰？综合情报决策

## 开源情报

- 四川大学网安学院：基于图注意力网络的舆情演变预测研究
- 情感交互图中的顶点是在该时间步长内出现的情感词，若评论文本中存在情感词  $w$ ，则将该条评论  $s$  归属到对应的顶点  $v$ 。一个顶点可以包含多条评论，一条评论也可以分配给多个顶点。该映射关系也间接反映出两个情感词关联的公共评论文本越多，所表达的情感越接近。



— 两个顶点之间存在公共评论句  
- - - 两个顶点之间不存在公共评论句，但其文本相似度大于阈值0.3

图2 情感交互图

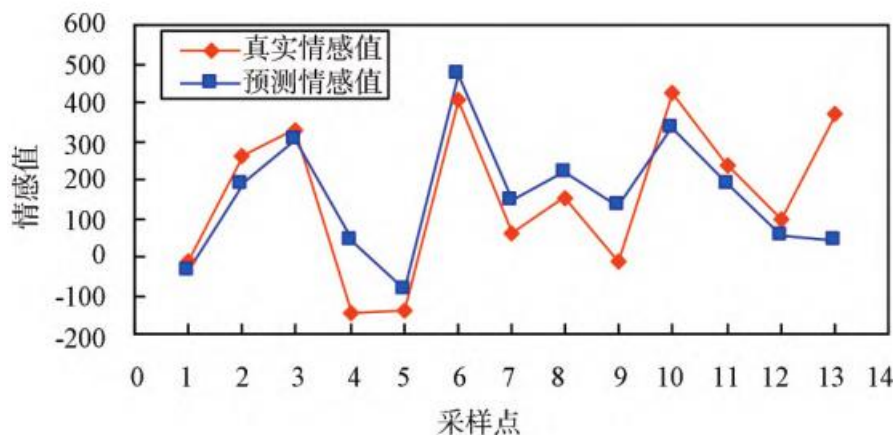



图3 预测情感值与真实情感值对比图

# 后续讨论班介绍

---

- 第一周：图神经网络在金融领域的应用-石逢钊
  - 第二周：图神经网络赋能的知识图谱研究与应用-刘 瑜
  - 第三周：图网络在生物化学领域的应用-周玉晨
  - 第四周：图网络在社会网络中的应用-宋传承
- 
- 第五周：图神经网络在推荐系统中的应用-吴咏萱
  - 欢迎访问课题组官网<https://ascii-iie.github.io/>