Machine Leanring on Temporal Graphs

——时序图上的机器学习

汇报人：刘猛（国防科技大学）

汇报时间：2022年10月28日

MindSpore

目 录

CONTENT

[M]S MindSpore

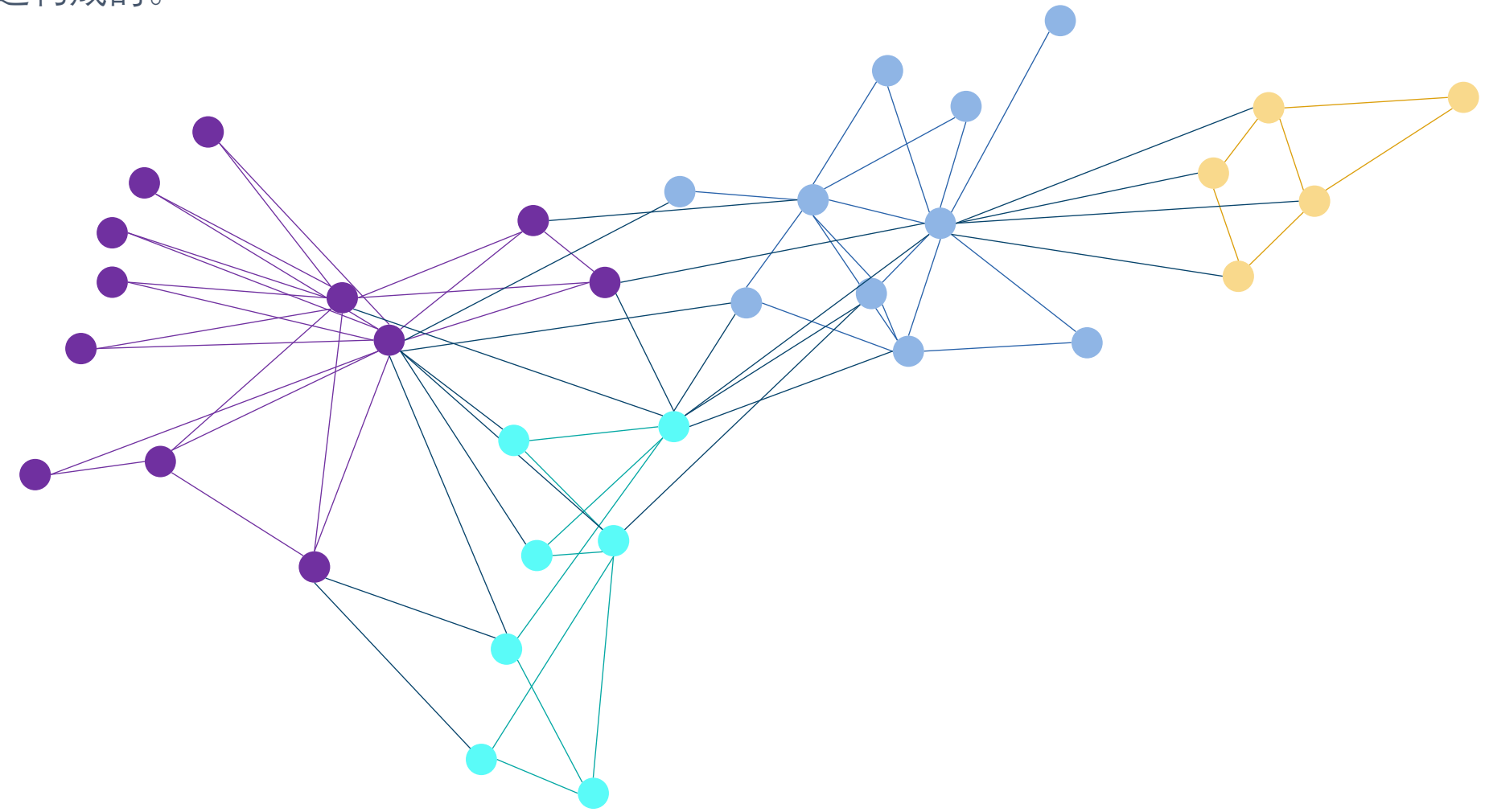
PART 01

01 图是什么

MindSpore

图是什么?

图是由节点和边构成的。



[M]S 昇思 MindSpore

为什么要研究图?

图 (Graph)， 又被称为网络 (Network)， 是一种现实世界广泛存在的数据形式。

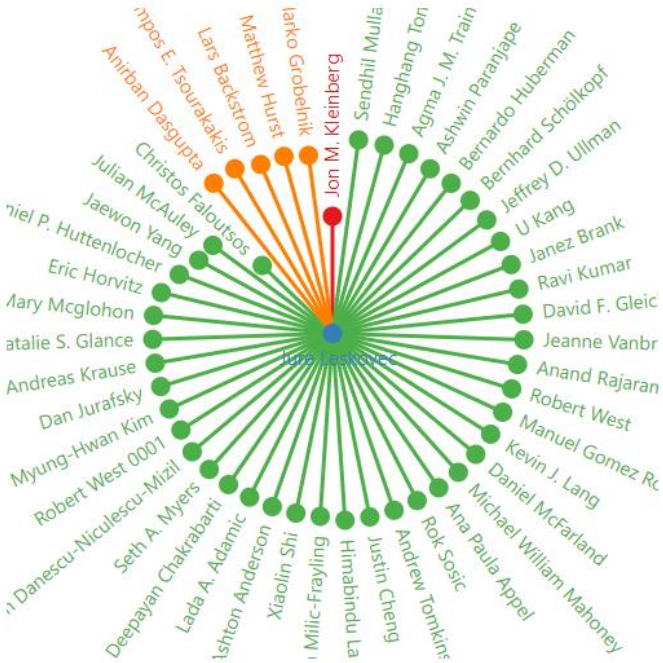
群体社交网络[1]

城市通信网络[2]

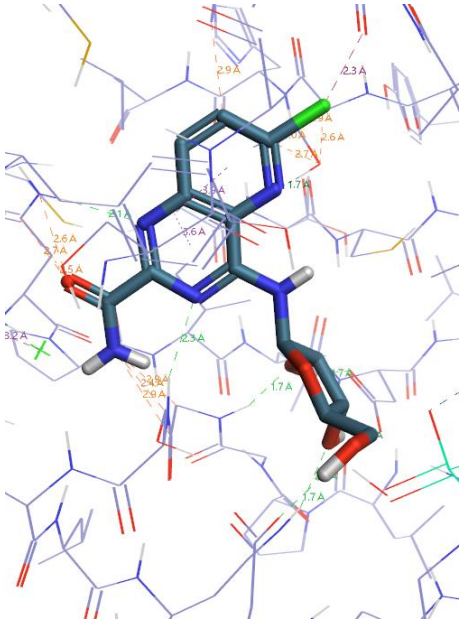
[M]S 昇思 MindSpore

现实世界有哪些图？

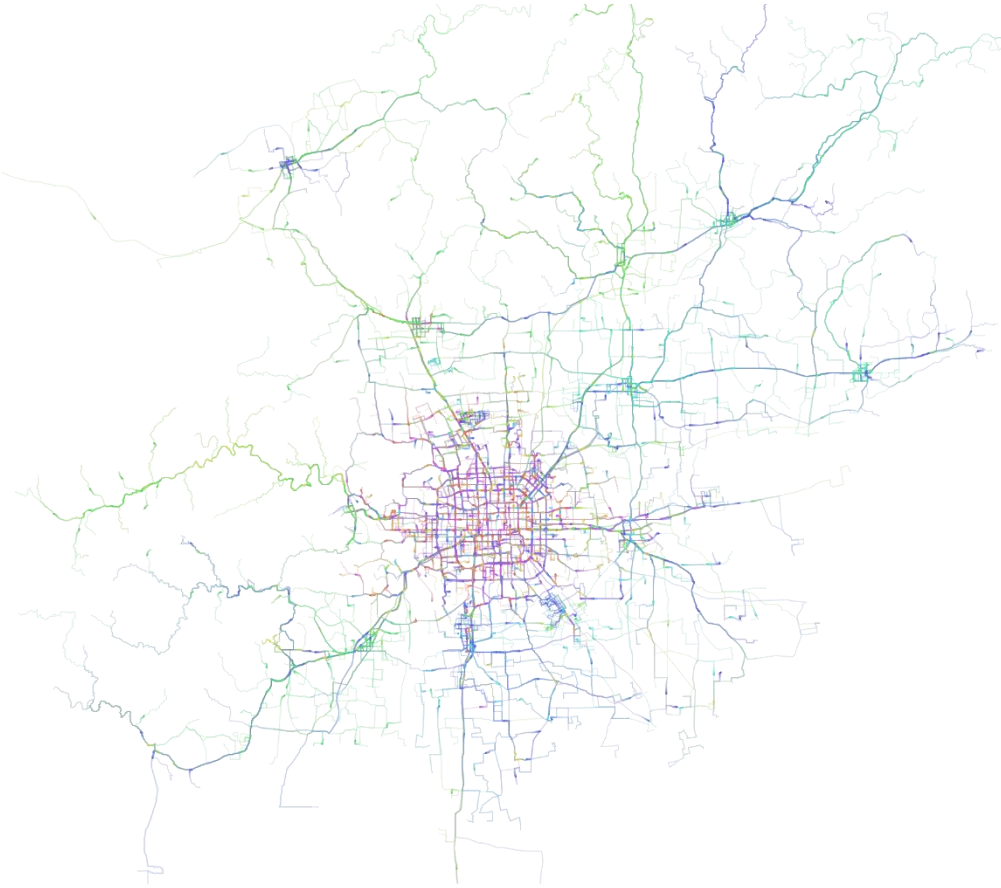
不仅于此，很多领域的数据都可以表现为图的形式。



学术合作[3]



分子结构[4]



城市交通[5]

[M]S 昇思 MindSpore

什么领域和图相关?

- 社会计算（社区发现/因果推理）
- 兴趣推荐（购物/视频/社交软件）
- 舆情分析（虚假新闻传播/信息影响力）
- 金融交易（风险评级/行为分析）
- 物联网（设备通信）
- 知识图谱（图结构数据库/关系推理）
- 智慧城市（交通预测）
- 生物医学（新药生成/靶向用药）

- 图像（像素间/图像间）
- 视频（不同帧间的关系变化）
- 文本（词间归属关系）
- 多模态数据（相似度矩阵）

[M]S 昇思 MindSpore

图和图像？图神经网络？

- 图 (Graph) 是一种抽象的数据结构，由节点和边构成，表达个体之间的关系。
- 图像 (Image) 是一种实际的数据表现，由像素构成，在二维空间中呈现事物。

- 图也经常被称为网络，都表达的是结构化的数据形式，但在图神经网络这个概念中，要对两者加以区分。
- Graph Neural Network (GNN) 中的 Graph 是图数据，而 Network 是由多组模拟人类神经元的函数组合而成的模型。

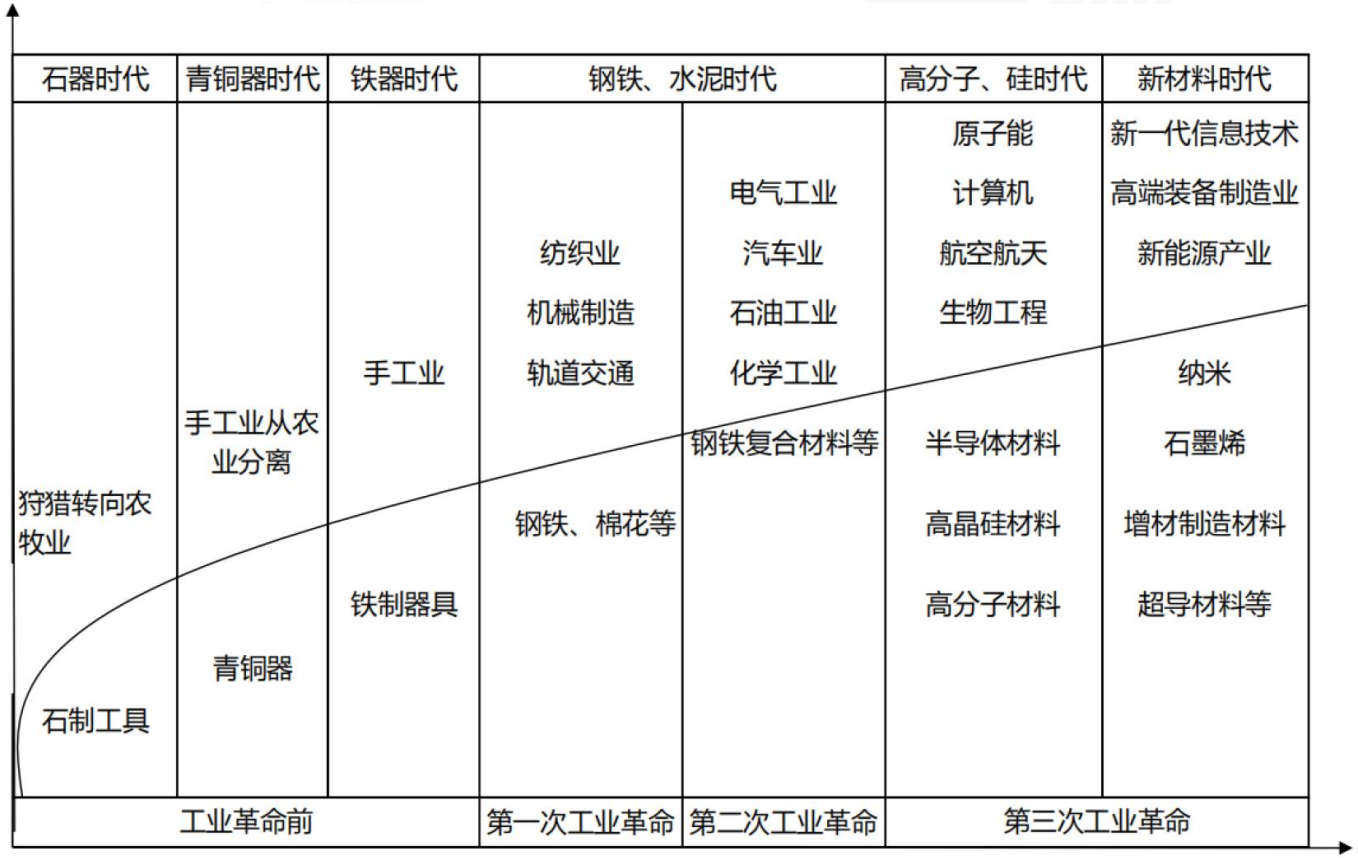
PART 02

02 图机器学习

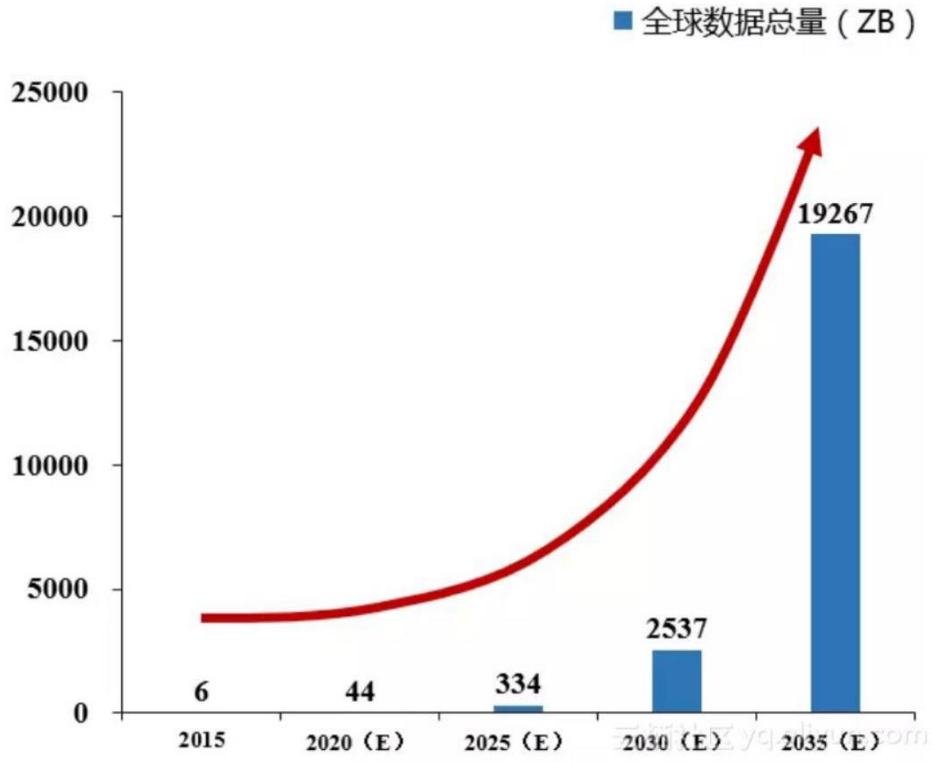
MindSpore

为什么用机器学习来挖掘图?

随着人类活动的日益频繁和复杂，传统的数据挖掘方法已不足以应对愈发庞大的图数据，因而研究者们将目光投向了新兴的机器学习方法。



人类生产生活的变化[6]



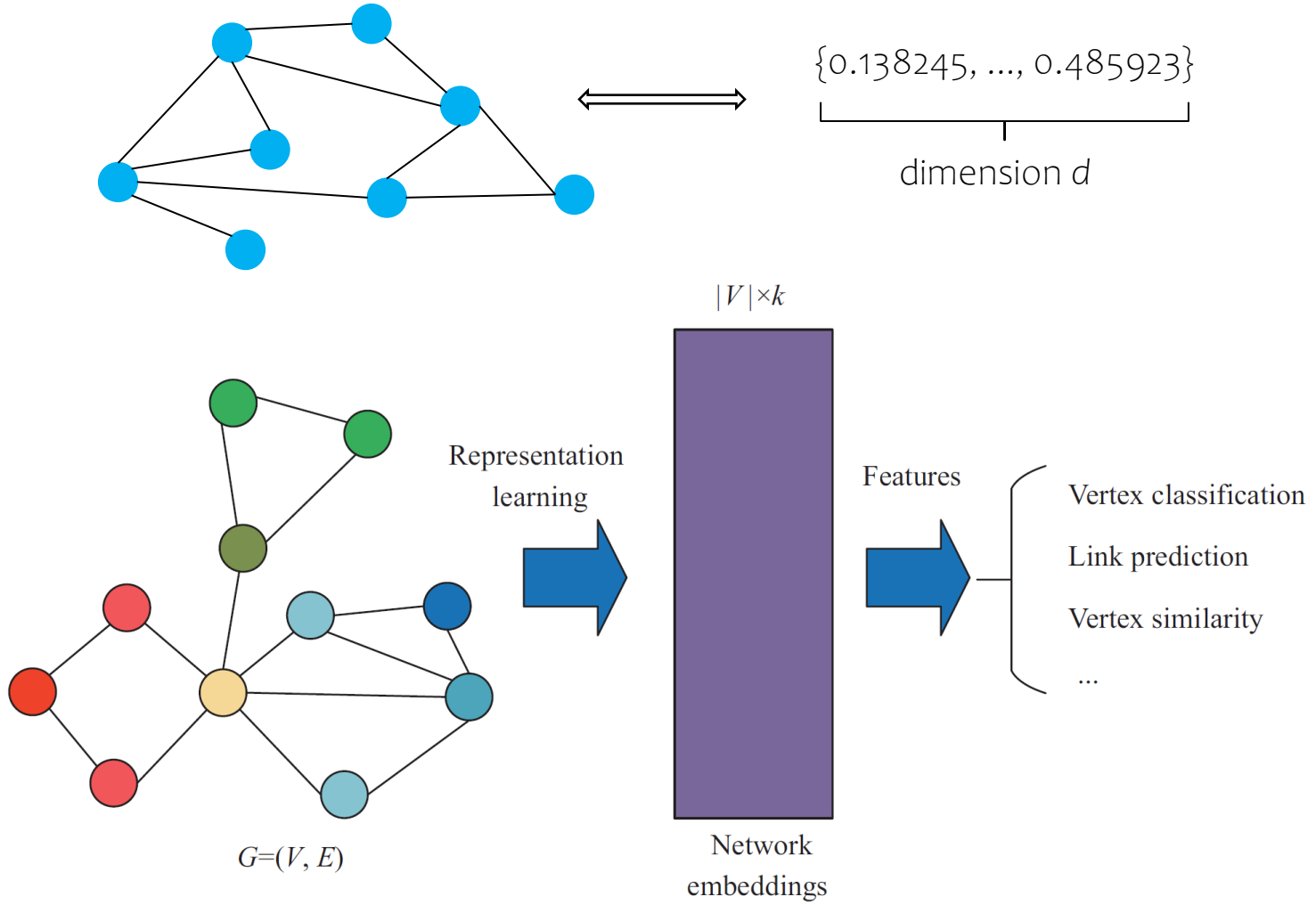
全球数据规模指数增长[7]

昇思 MindSpore

怎么学习图?

图上的机器学习不再试图直接得到结果，而是通过建模图结构数据，为图中的节点生成表征向量，使这些向量能够代表每个节点的信息与属性。

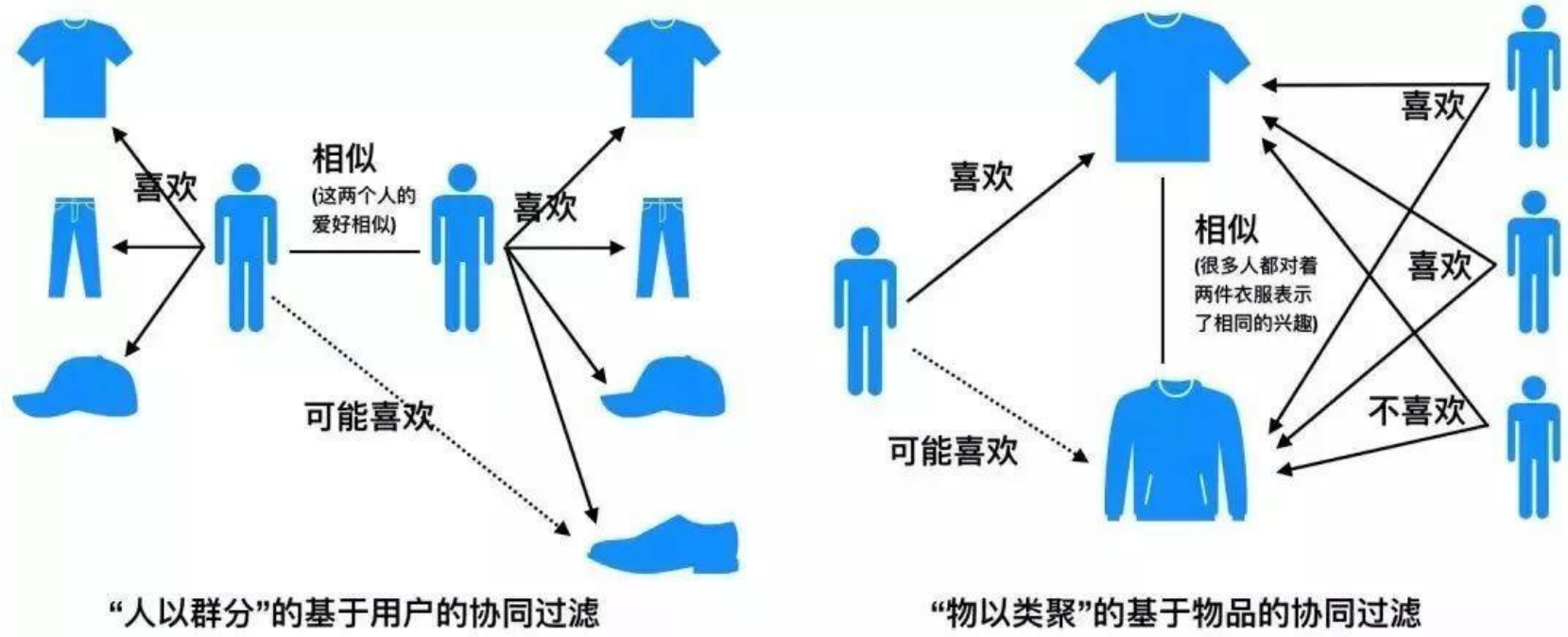
而后的下游任务都可基于节点表征计算，不需要频繁访问图结构。



图结构转换为节点表征[8]

节点表征能做什么？——以淘宝/抖音的兴趣推荐为例

通过为不同的用户、商品/视频学习表征向量，计算向量之间的相似度，就可以判断用户是否会对其感兴趣，从而进行推荐。



基于协同过滤的兴趣推荐[9]

[M]S 昇思 MindSpore

图学习的下游任务

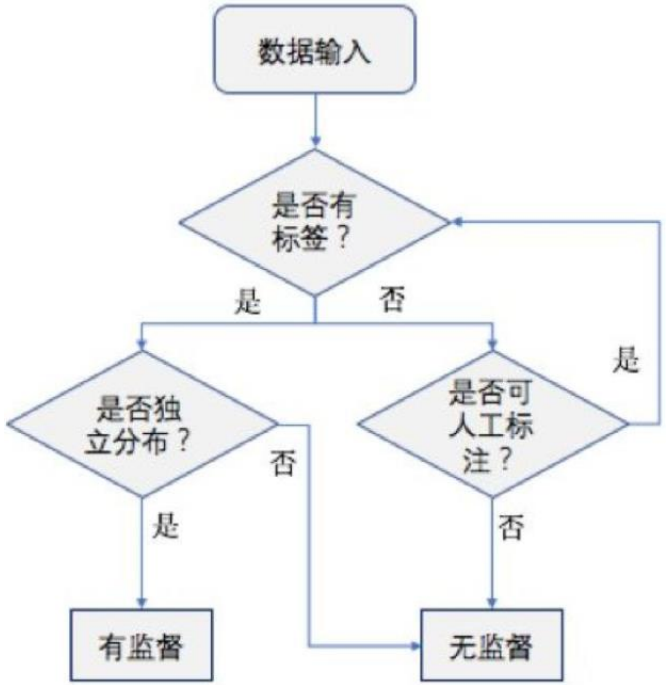
图学习难以清楚地界定为有监督学习或无监督学习，某些时候，它图学习同时具备有监督学习和无监督学习的特性[10]。事实上，依照其下游任务的不同，图学习的监督场景也是“薛定谔”式的。

有监督场景：节点分类、图分类

无监督场景：链路预测、节点聚类、图预测/聚类

另一方面，也可从不同的关注点出发，对下游任务进行划分。

- 节点级：针对节点类别属性
- 边级：针对节点对间的交互
- 图级：针对子图结构

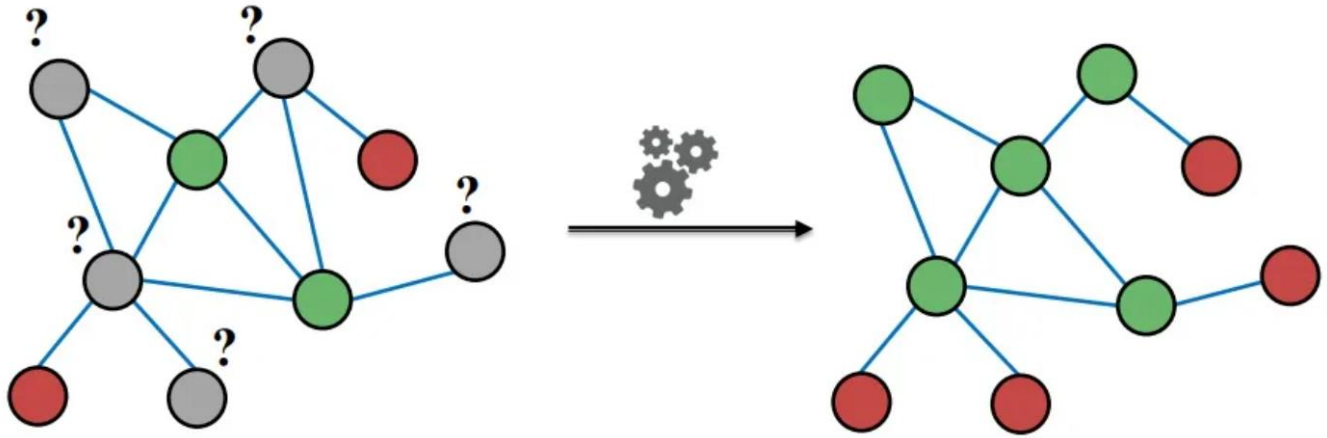


有监督和无监督 [10]

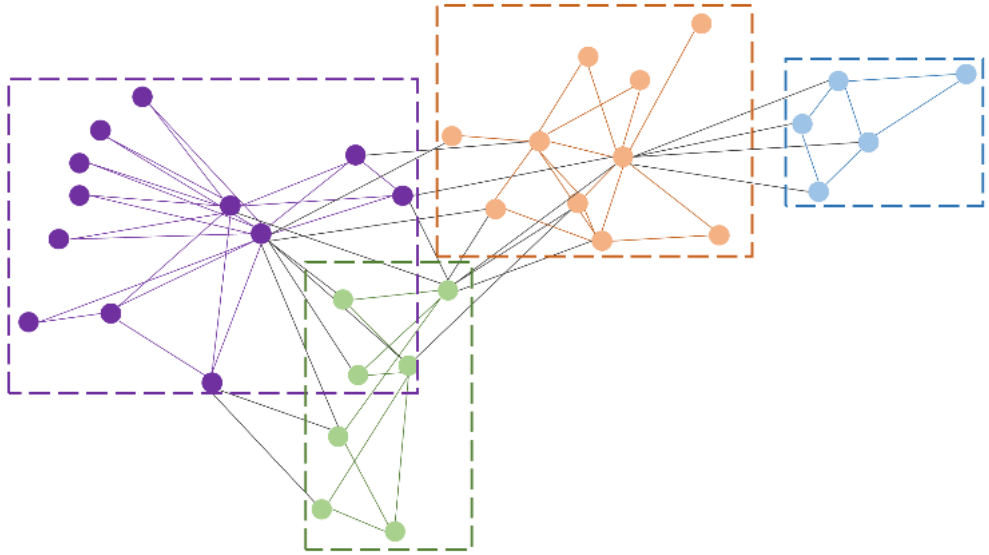
[M]S 昇思 MindSpore

节点级任务

节点级任务主要包括节点的分类和聚类，事实上，节点聚类也叫做社区检测 (community detection)。



节点分类[11]

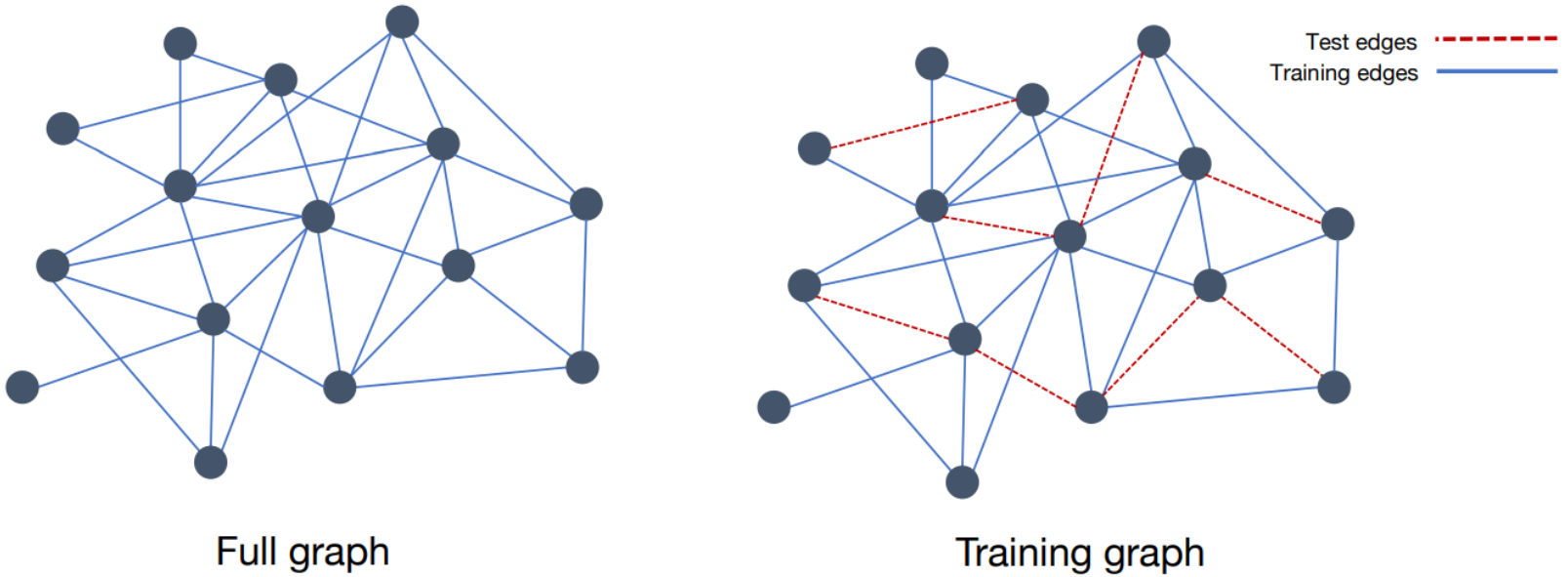


节点聚类

[M]S 昇思 MindSpore

边级任务

边级任务主要是链路预测/兴趣推荐，属于无监督场景下的任务。

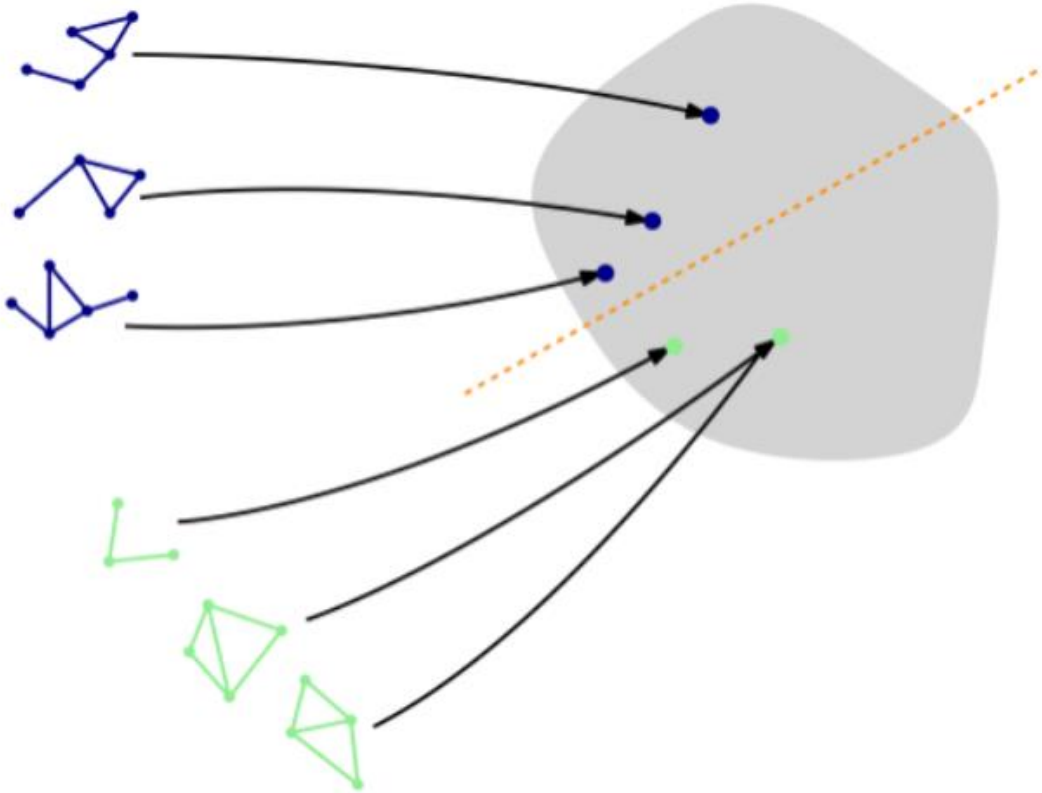


链路预测[12]

[M]S 昇思 MindSpore

图级任务

图级任务主要是图的分类和聚类。



图结构分类[13]

[M]S 昇思 MindSpore

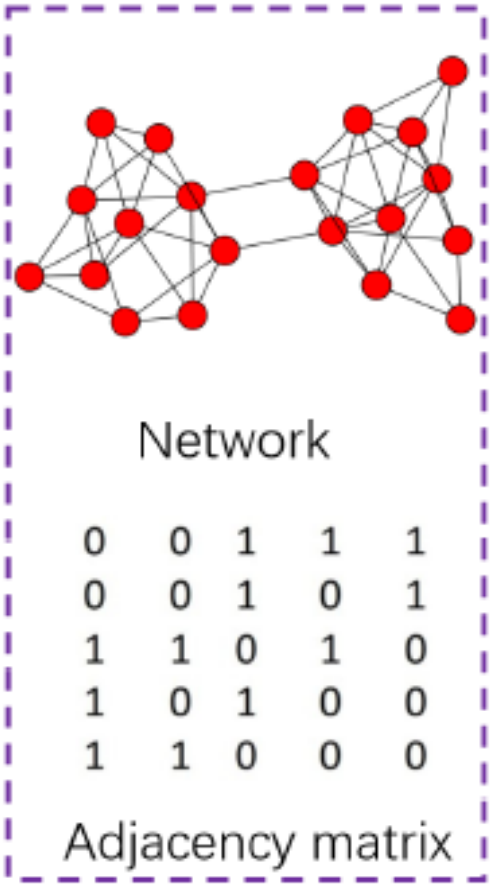
图数据的核心是什么？

一个基础的图结构 G = (V, E) 是由节点集 V 和边集 E 构成的，学习图中的信息，就是在学习节点与节点间的边。

不同图数据中，节点和边的含义也大不相同：

- 引文图中，论文为节点，引用关系为边；
- 购物图中，用户和商品为节点，购买关系为边；
- 分子图中，原子为节点，原子间的键为边……

这些节点和边的关系，可以通过邻接矩阵表现出来。

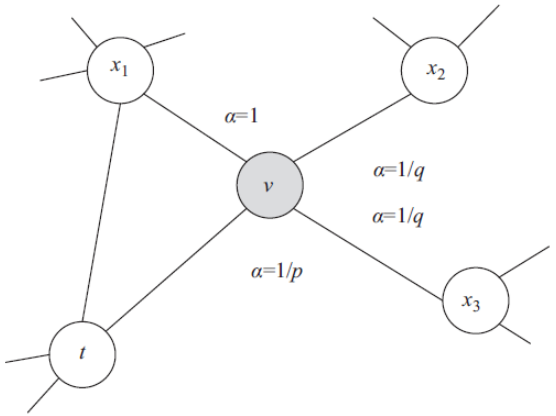


邻接矩阵[14]

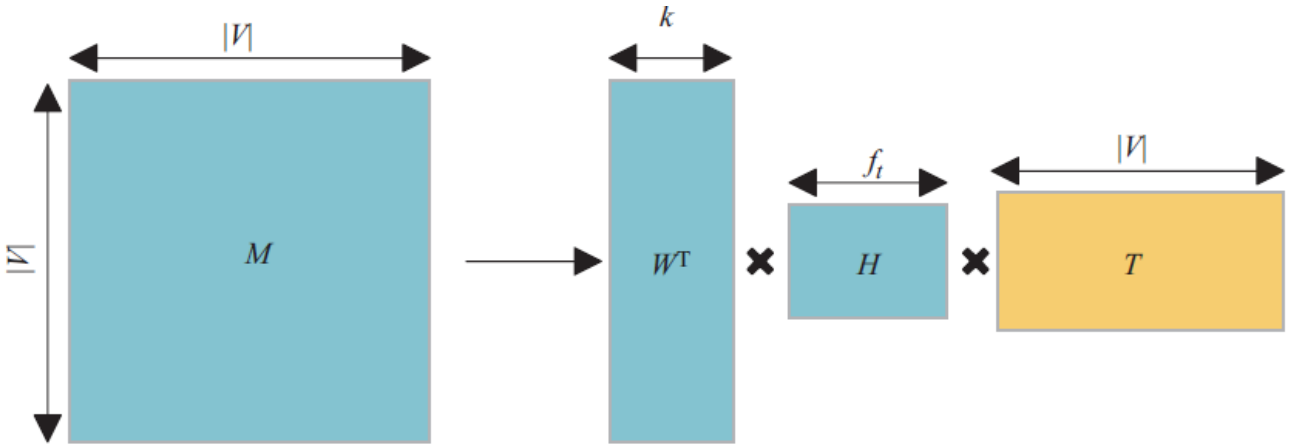
[M]S 昇思 MindSpore

学习图数据的方法

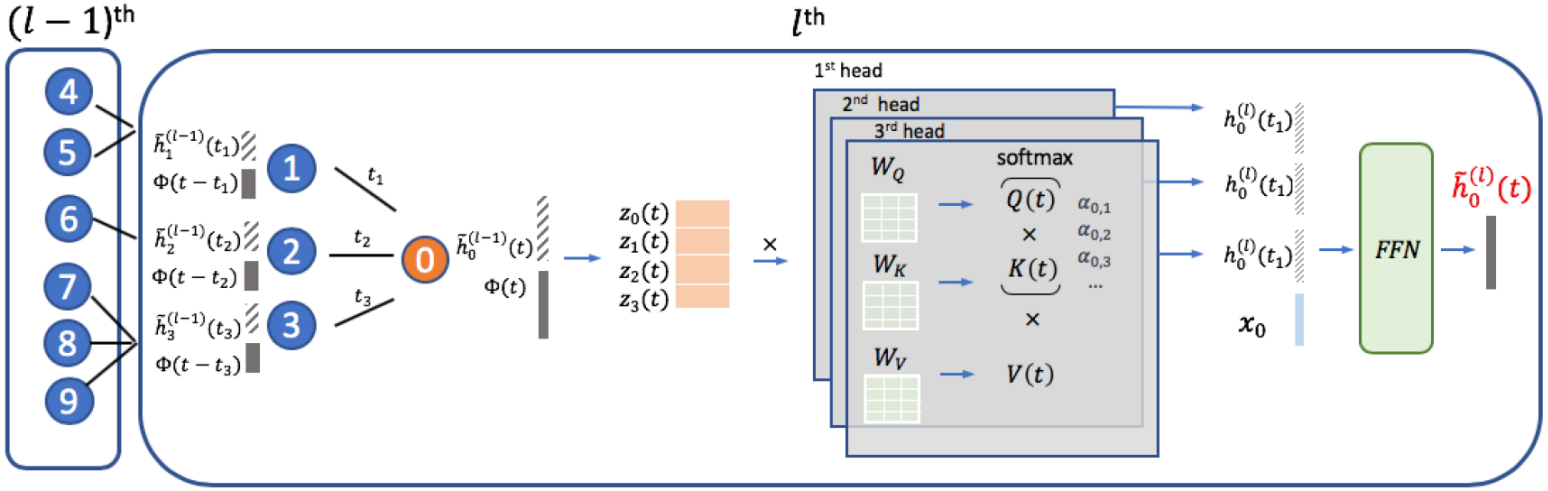
现有方法可以分为三类：随机游走、矩阵分解、神经网络。



随机游走 [15]



矩阵分解[15]



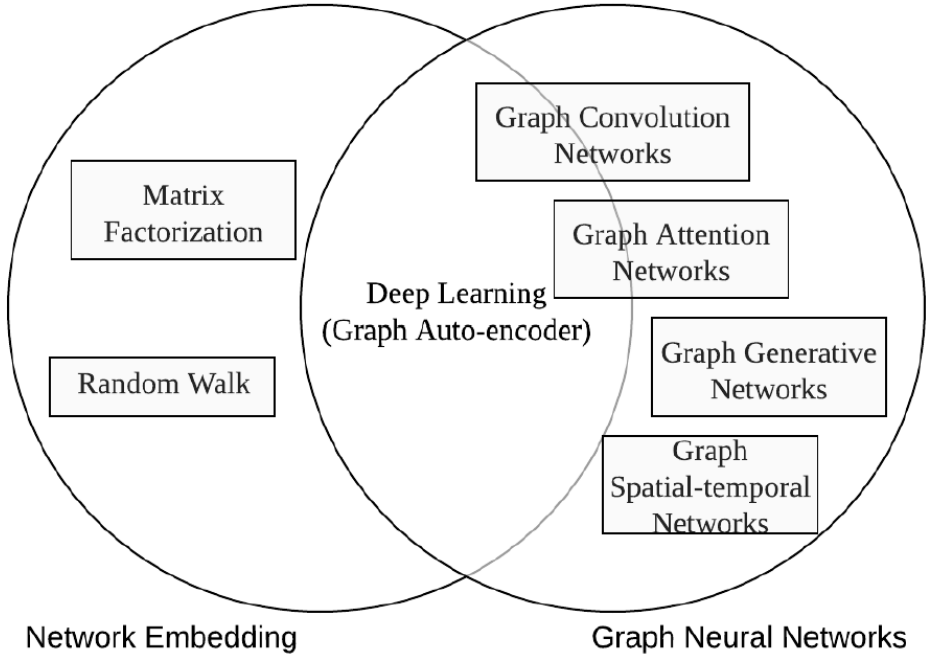
神经网络[16]

昇思 MindSpore

图神经网络的分类

图神经网络可以大体分为五类[17]：图卷积、图注意力、图自编码、图生成、图时空。

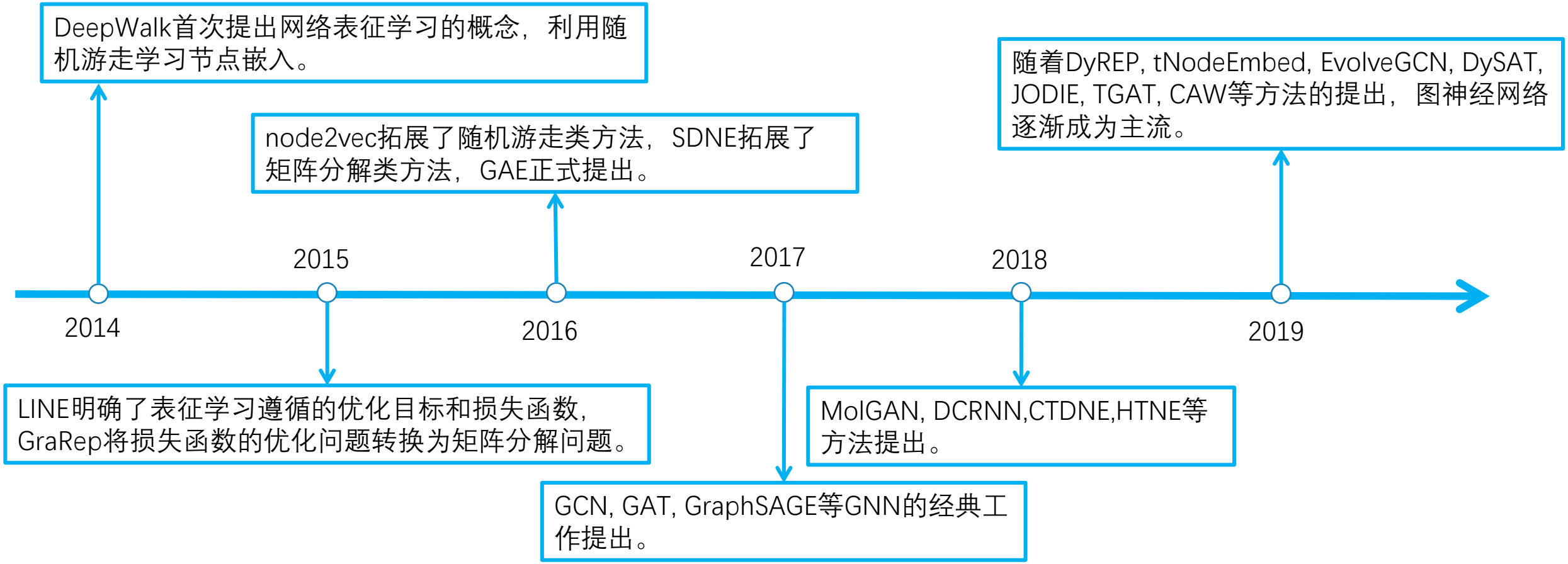
也可分为四类：循环GNN、卷积GNN、自编码GNN、时空GNN。



图神经网络[17]

昇思 MindSpore

图学习的演变



[M]S 昇思 MindSpore

谁在研究?

Jure Leskovec，斯坦福大学，Geogle引用量11万+，H指数129，提出node2vec和GraphSAGE等算法。

Steven Skiena，美国石溪大学杰出教授，Geogle引用量2万+，H指数65，《算法设计手册》作者，提出DeepWalk算法。

Thomas Kipf，Google Brain，Geogle引用量2万+，H指数22，GCN, GAE, R-GCN 等算法的第一作者。

昇思 MindSpore

谁在研究?

唐杰，清华大学，国家杰青，Geogle引用量2万+，H指数83，开发了Aminer平台和悟道大规模预训练模型。

崔鹏，清华大学， Geogle引用量1万+，H指数52，提出了SDNE和HOPE等算法。

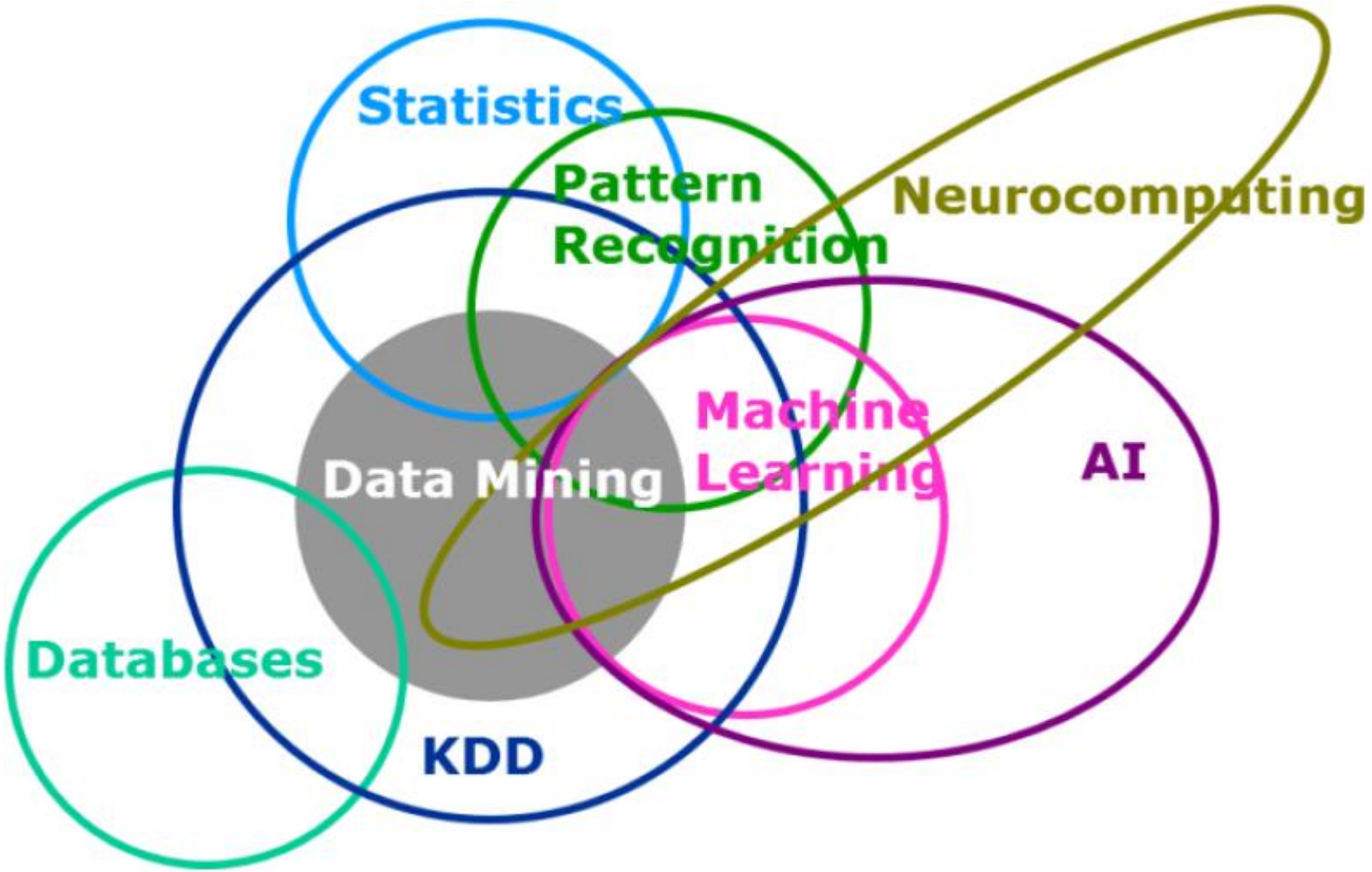
石川，北京邮电大学，Geogle引用量8千+，H指数42，专注于异质图和大规模图领域。

[M]S 昇思 MindSpore

图学习的定位

图学习本质上是用机器学习的方法来挖掘图数据，因此，它是一类同时衔接数据挖掘领域和机器学习领域的重要方向。

事实上，数据挖掘和机器学习也是很难区分的，当数据挖掘中寻找的知识等同于机器学习中寻找的函数时，两者基本也是等同的。



领域间关系 [18]

图学习的相关会议和期刊

《中国计算机学会推荐国际学术会议和期刊目录》

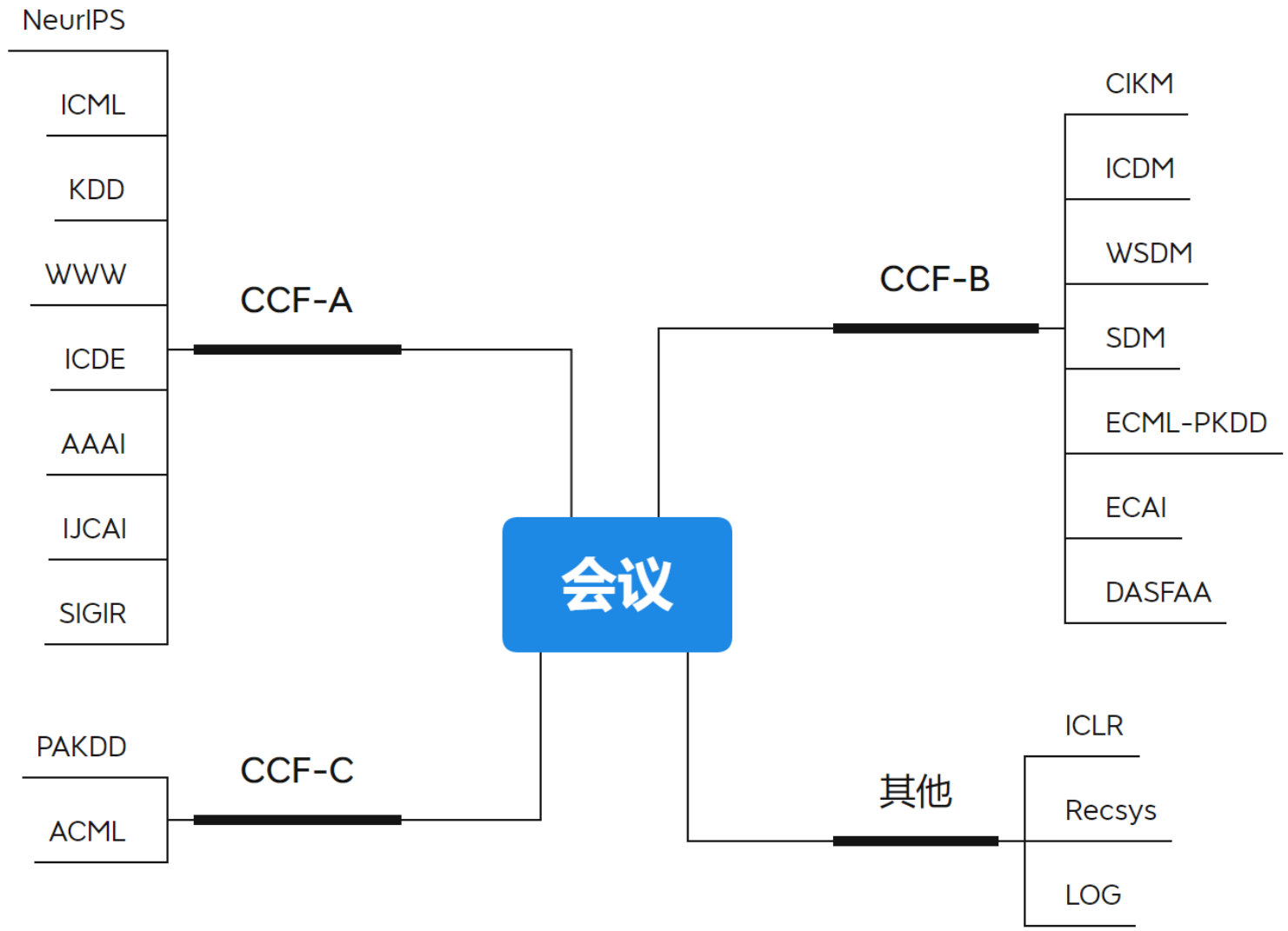
是CCF制定的用来评价计算机会议与期刊级别的推荐列表，分为A, B, C三个等级。

值得注意的是，CCF评级并不能完全代表会议或期刊水平，一些排名较低或未在排名中的会议期刊同样有着较高的行业认可度。（注：下文列出的内容并未包含所有相关会议期刊，部分水平很高的因相关度有限暂未列出，如有疏漏敬请谅解。）

此外还有CCF推荐的几本较为重要的中文期刊：计算机学报、软件学报、中国科学：信息科学、计算机研究与发展等。

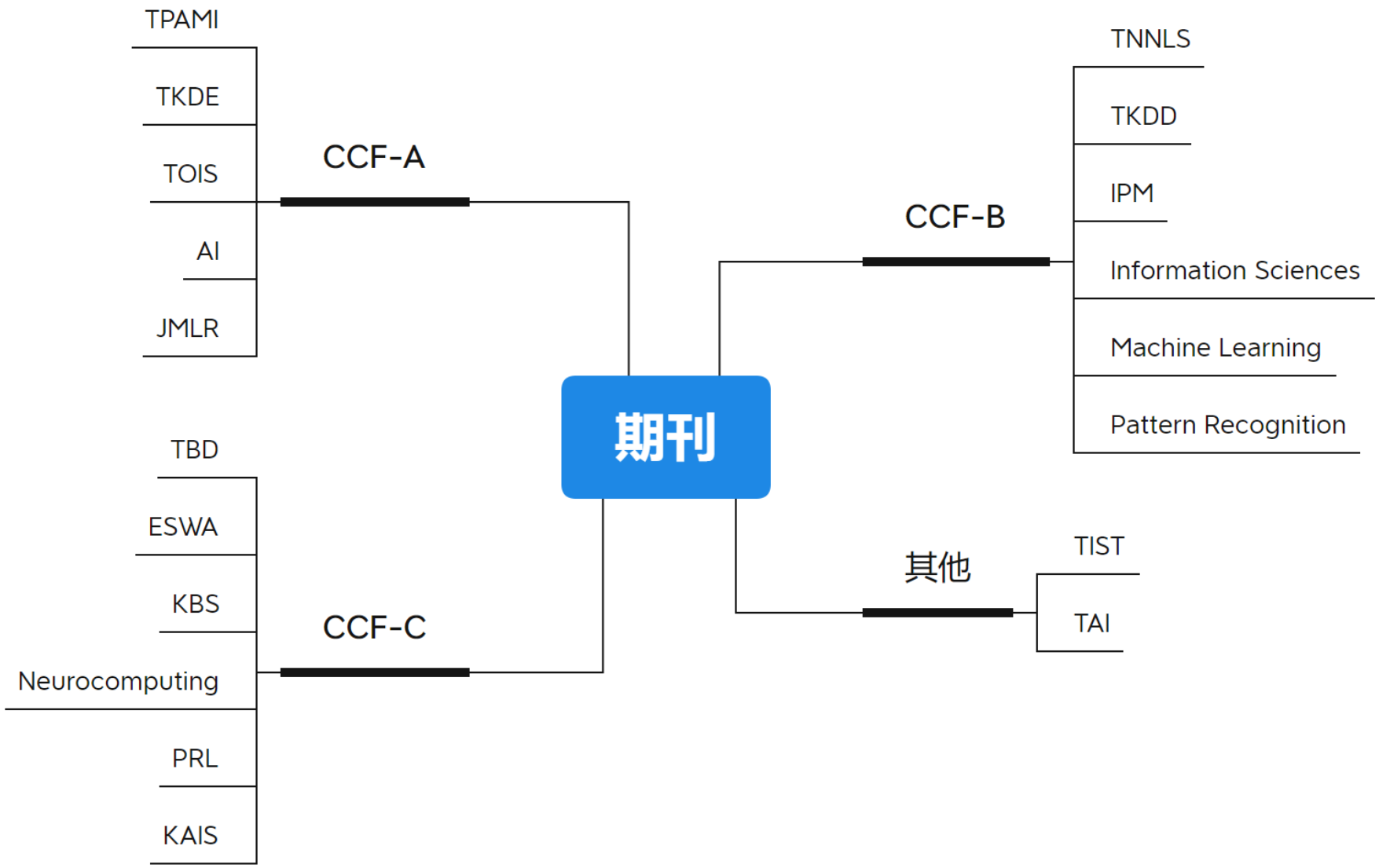
[M]S 昇思 MindSpore

图学习相关会议



昇思 MindSpore

图学习相关期刊



[M]S 昇思 MindSpore

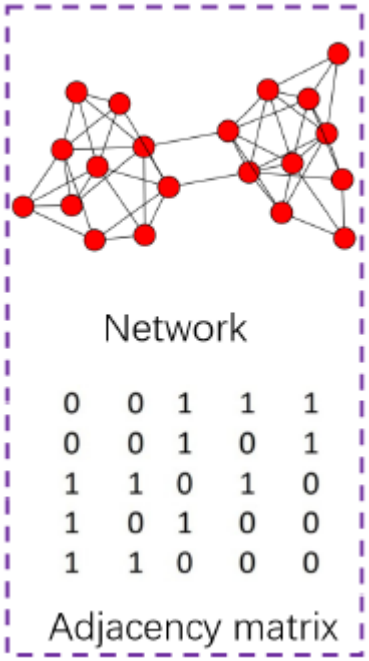
PART 03

03 时序图学习

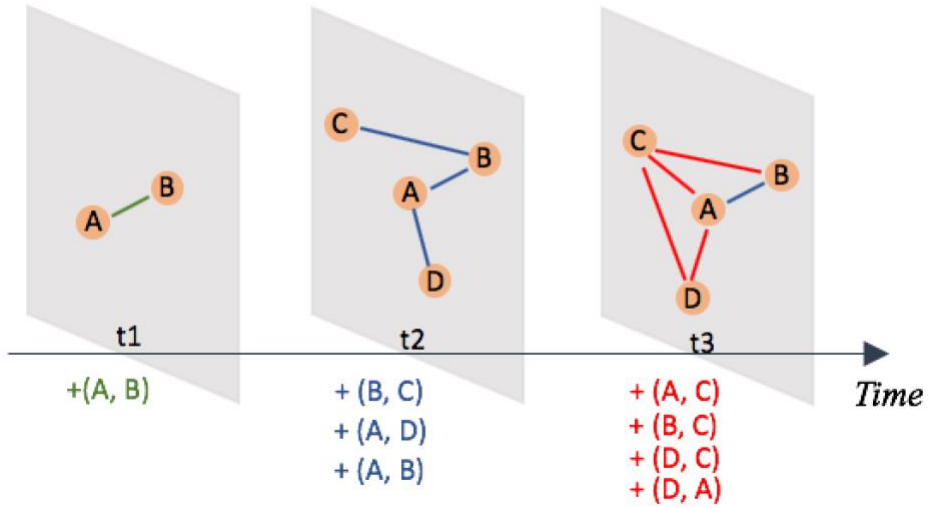
MindSpore

图数据如何分类?

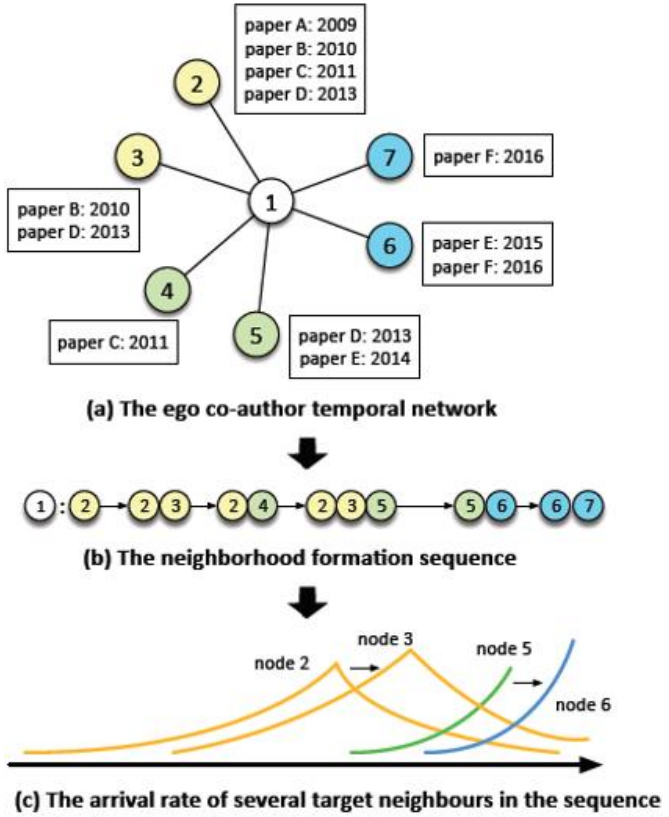
图数据可以根据其是否含有节点交互的动态信息，分为静态图 (static graph) 和动态图 (dynamic graph)，动态图又可细分为离散图 (discrete graph/static snapshot)和时序图 (temporal graph)。



静态图



离散图[16]



时序图[19]

时序图是什么样的?

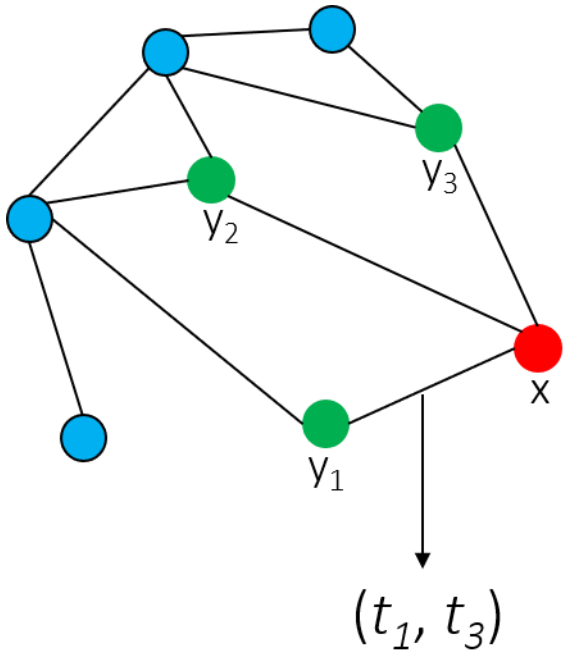
输入 G = (V, E, T)，其中 V 是节点集， E 是边集，T 是伴随时间信息的节点交互集。

数据形式： (node id, node id, timestamp)

(1, 2, 0.0000), (1, 3, 0.0504), (4, 3, 0.1200), (64, 2, 0.2500)

……

(3, 5, 0.3800), (12, 9, 0.4613), (1, 2, 0.6983), (8, 2, 0.9674)



在对时序图建模的初始阶段，图结构是不存在的，每有一对节点发生交互，抽象概念上的图才会生成对应的节点和边。随着时间的推进，时序图结构就被不断完善。

[M]S 昇思 MindSpore

为什么时序图不用邻接矩阵？会带来哪些问题？

时序图数据是不同于传统静态图或离散动态图的，这给时序图学习的方法既带来了好处，也存在着不便。

优势:

可以处理大规模图，通过分批次送入模型，避免了邻接矩阵过大从而读取困难的情况。

劣势:

传统的GNN方法无法适用，因为邻接矩阵在此不适用。

为什么邻接矩阵不适用?

两个节点间可能会存在着多次交互，一条边难以表达多个时间戳。（这个难不仅在于原始数据的存储难，也在于数据读取之后的存储难。）

[M]S 昇思 MindSpore

时序图学习方法

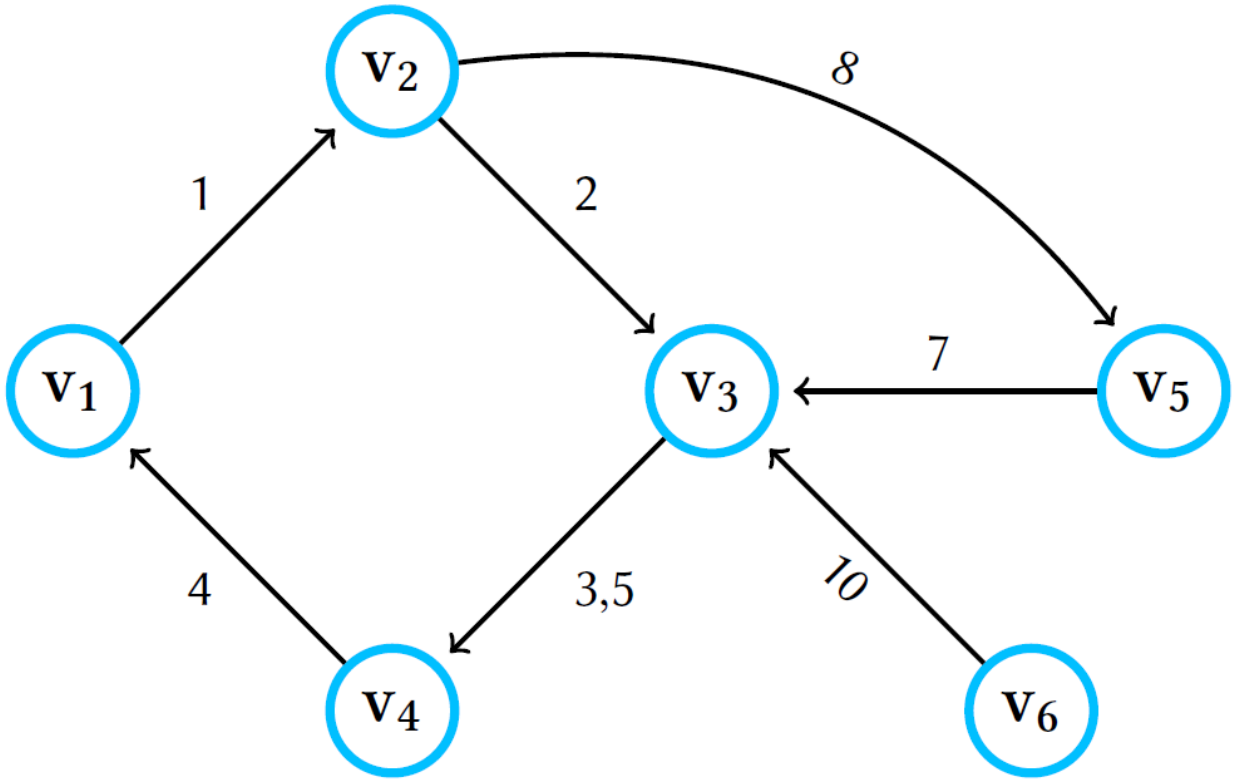
在此，我们简单介绍几种时序图学习的经典或最新方法：

- CTDNE：利用随机游走建模时序图
- HTNE：利用霍克斯过程建模时序图
- JODIE：提出了预测节点未来表征的思想
- CAW：提出了因果匿名随机游走的思想
- TREND：用GNN建模时序图

[M]S 昇思 MindSpore

CTDNE

CTDNE通过执行有时间先后的随机游走，对时序图进行建模。

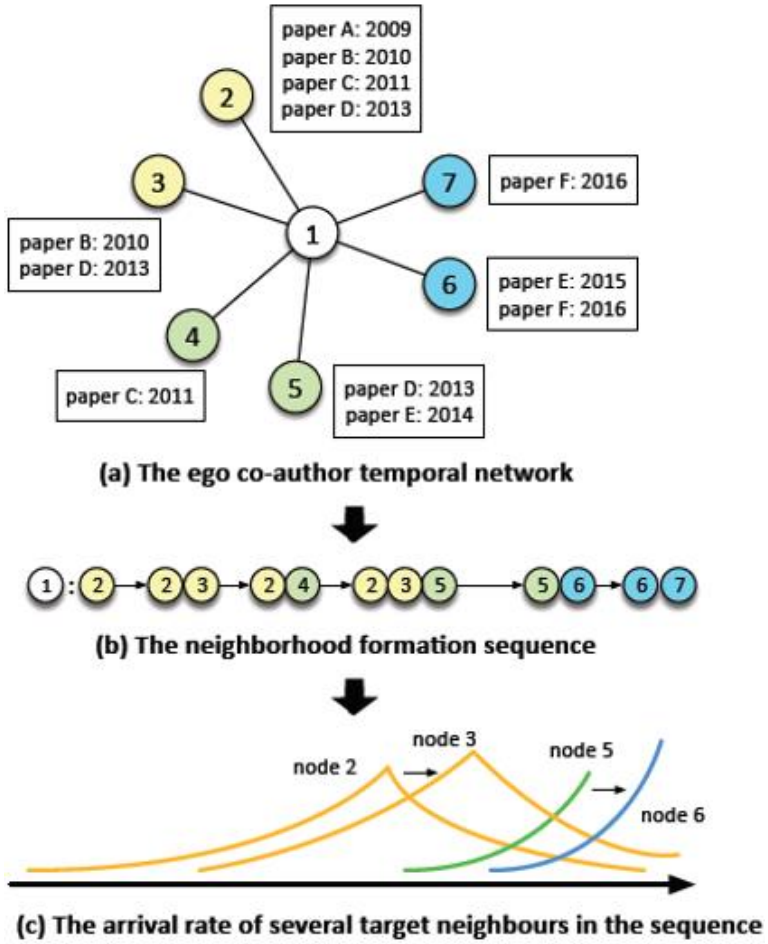


$$\Pr(w) = \frac{\exp\left[\tau(w) - \tau(v)\right]}{\sum_{w' \in \Gamma_t(v)} \exp\left[\tau(w') - \tau(v)\right]}$$

时序随机游走[20]

HTNE

HTNE引入霍克斯过程，认为两个节点的交互不仅与它们自身相关，还会受到历史交互邻居的影响。



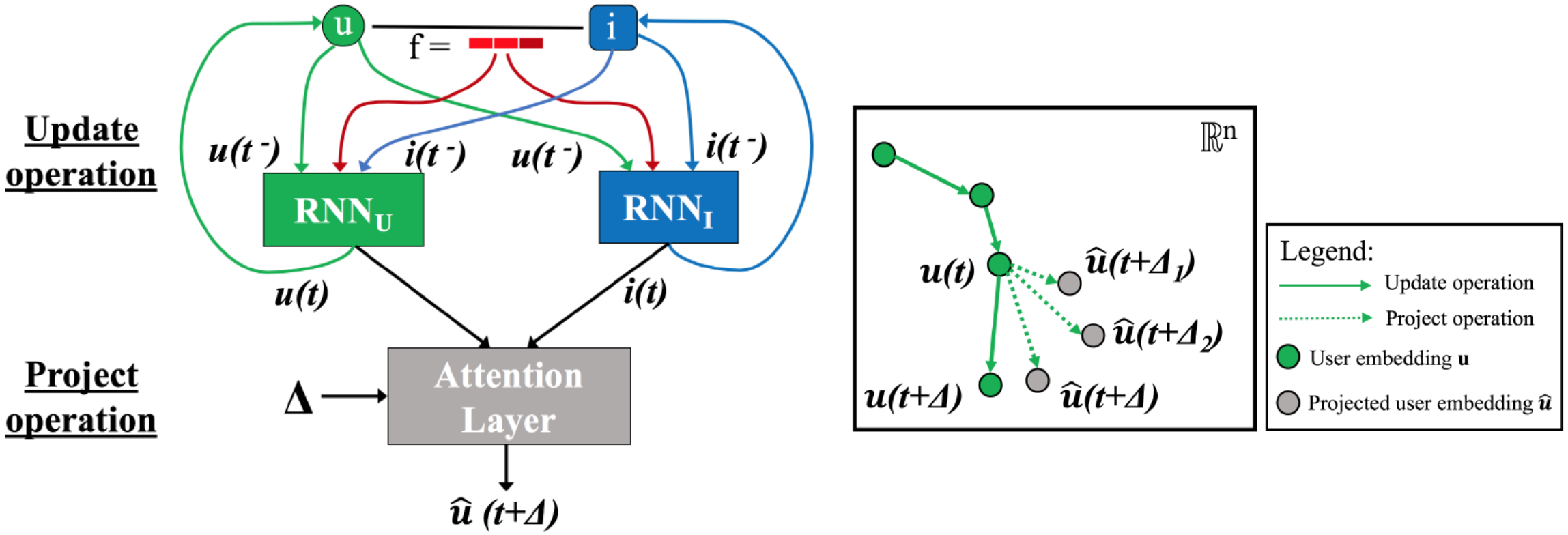
$$\tilde{\lambda}_{y|x}(t) = \mu_{x,y} + \sum_{t_h<t} \alpha_{h,y}\kappa(t-t_h)$$

$$\alpha_{h,y} \cdot k(t-t_h) = -||z_h - z_y||^2 \cdot \frac{e^{(-||z_x - z_h||^2)}}{\sum_{h'} e^{(-||z_x - z_{h'}||^2)}} \cdot e^{(-\delta(t_c - t_h))}$$

霍克斯建模[19]

JODIE

JODIE通过RNN学习节点表征，并预测节点未来的表征变化。

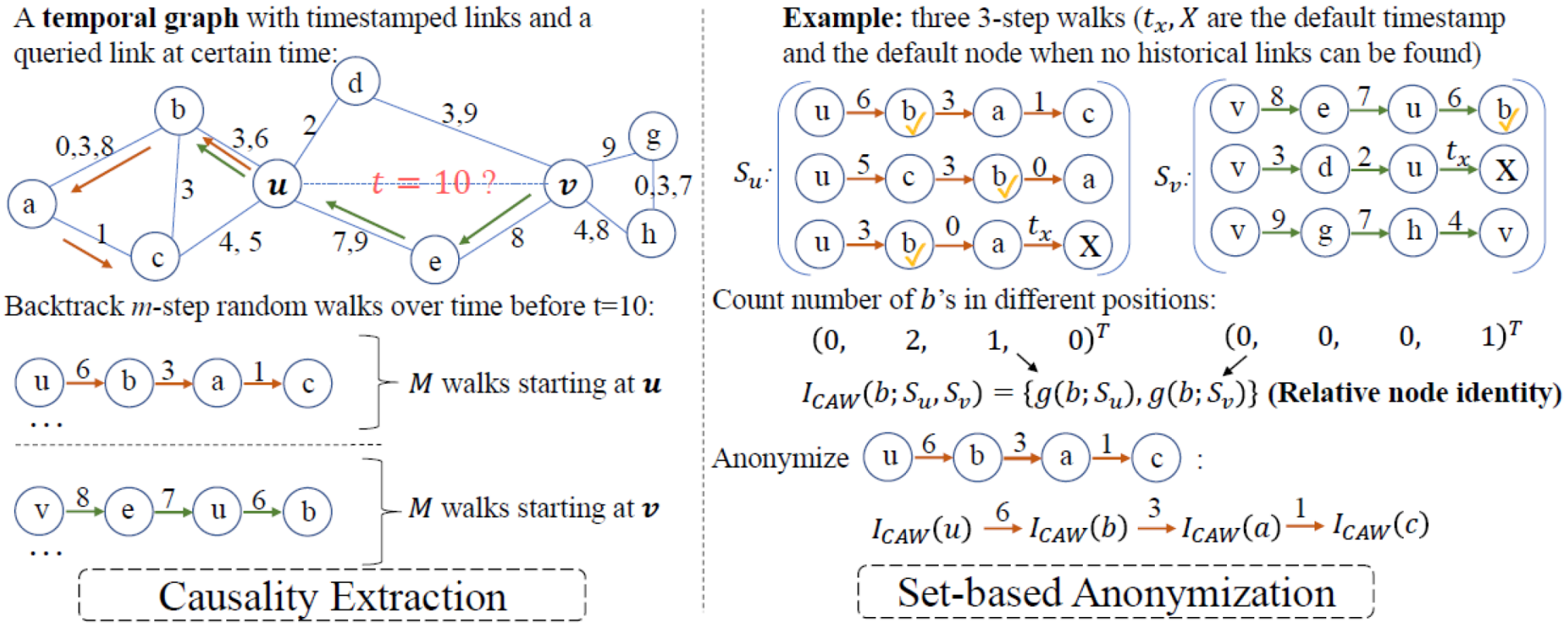


JODIE[21]

[M]S 昇思 MindSpore

CAW

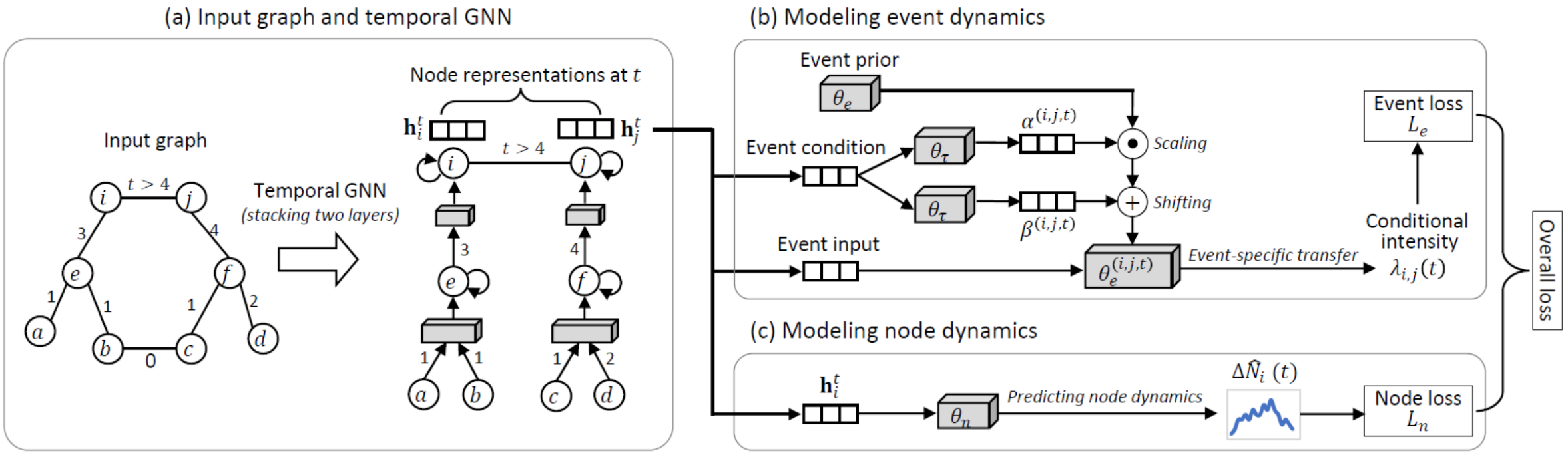
提出了因果匿名随机游走方法，对节点的多次时序游走进行匿名化后做因果抽取，计算边表征。



因果匿名随机游走 [22]

TREND

TREND用GNN替换霍克斯过程对节点的高阶邻域进行建模。



$$\lambda_{y|x}(t) = \sigma(-||z_x - z_y||^2 \odot \theta_e)$$

$$z_x^{t,l} = \sigma(z_i^{t,l-1} W_{self}^l + \sum_{h \in N_x} z_h^{t_h,l-1} W_{hist}^l k(t - t_h))$$

TREND[23]

时序图学习面临什么挑战?

- 邻居序列长度：时序图要获取节点的邻居序列，但过长会影响计算效率，过短又导致信息缺失。
- 时序图上的GNN：高效的GNN在时序图上还不能很好地适用，这受到邻接矩阵的限制。
- 对时序信息的进一步利用：当前算法多数是简单地计算时间差。

[M]S 昇思 MindSpore

PART 04

04 总结与展望

MindSpore

汇报内容

- 图是表达个体间关系的抽象数据结构。
- 社会中很多场景都可以表现为图的形式，并利用图学习来解决问题。
- 图学习本质是用机器学习的方法对图数据做数据挖掘。
- 通过研究节点间的关系，图学习将抽象的图数据提取为具体的表征向量。
- 图学习的下游任务可以分为有监督和无监督，也可分为节点级、边级、图级。
- 图学习可分为静态图和动态图，动态图又可分为离散图和时序图。
- 时序图学习与传统静态图不同，不再使用邻接矩阵的数据形式，而是表示为交互序列。

图学习可能的发展方向

昇思 MindSpore

有什么资源帮助学习?

- 斯坦福大学CS224W课程：Machine Learning with Graphs（图机器学习），主讲人 Jure Leskovec
- 《Graph Representation Learning》，作者 William L. Hamilton
- 《Deep Learning on Graphs》，作者 马耀，汤继良
- **https://github.com/MGitHubL/Chinese-Reading-Notes-of-Graph-Learning**

Graph Learning Notes (In Chinese)

在此整理了一些个人的文献阅读笔记，主要是图学习领域的，希望大家多多指正。

Survey

- Graph self-supervised learning: A survey (**TKDE 2022**) [paper][note]
- 面向社会计算的网络表示学习 (**清华博士论文** 2018) [paper][note]
- A Survey on Network Embedding (**AAAI 2017**) [paper][note]
- 网络表示学习专题 (**CCF 2017**) [note]

Paper

2022

- SAIL: Self Augmented Graph Contrastive Learning (**AAAI**) [paper][note]
- TREND: TempoRal Event and Node Dynamics for Graph Representation Learning (**WWW**) [paper][code][note]
- CGC: Contrastive Graph Clustering for Community Detection and Tracking (**WWW**) [paper][note]
- Pre-Training on Dynamic Graph Neural Networks (**Neurocomputing**) [paper][note]

2021

- Do Transformers Really Perform Bad for Graph Representation (**NeurIPS**) [paper][note]

[M]S 昇思 MindSpore

参考文献

[1] https://cloud.tencent.com/developer/news/46613

[2] https://699pic.com/tupian-500793249.html

[3] https://www.aminer.cn/

[4] Yuanxun Wang et al. Replacement of Protein Binding-Site Waters Contributes to Favorable Halogen Bond Interactions. J. Chem. Inf. Model. 2019.

[5] https://echarts.apache.org/

[6] 2020新材料行业研究报告，睿兽分析

[7] IDC，中国电子学会

[8] Cunchao Tu et al. “CANE: Context-Aware Network Embedding for Relation Modeling”. In: ACL. 2017.

[9] http://sykv.cn/cat/depth/20436.html

[10] https://blog.csdn.net/u010420283/article/details/83758378

[11] https://zhuanlan.zhihu.com/p/451082389

[12] William L. Hamilton et al. “Graph Representation Learning”.

[13] https://zhuanlan.zhihu.com/p/435945714

参考文献

[14] Peng Cui et al. “A Survey on Network Embedding”. In: TKDE (2019).

[15] 涂存超 等. 网络表示学习综述. 中国科学：信息科学

[16] Da Xu et al. “Inductive representation learning on temporal graphs”. In: ICLR. 2020.

[17] Zonghan Wu et al. "A comprehensive survey on graph neural networks." In: TNNLS (2020).

[18] 清华大学-中国工程院知识智能联合研究中心. 2019人工智能发展报告.

[19] Yuan Zuo et al. “Embedding Temporal Network via Neighborhood Formation”. In: KDD. 2018.

[20] Giang Hoang Nguyen et al. “Continuous-Time Dynamic Network Embeddings”. In WWW. 2018.

[21] Srijan Kumar et al. “Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks”. In KDD. 2019.

[22] Yanbang Wang et al. “Inductive Representation Learning in Temporal Networks via Causal Anonymous Walks”. In: ICLR. 2021.

[23] Zhihao Wen and Yuan Fang. “TREND: TempoRal Event and Node Dynamics for Graph Representation Learning”. In: WWW. 2022.

敬请批评指导！

汇报人：刘猛（国防科技大学）

汇报时间：2022年10月28日

MindSpore