

图神经网络导论

课程简介

授课教师：周晟

浙江大学 软件学院

2023.11.14



目录

- ① 课程介绍
- ② 从图数据到图神经网络
- ③ 图特征
- ④ 图数据挖掘任务



1 课程介绍

2 从图数据到图神经网络

3 图特征

4 图数据挖掘任务



课程信息

- ① 上课时间：冬学期每周二下午
- ② 上课地点：N306
- ③ 课程教师：周晟
- ④ 课程主页：<https://zhoushengisnoob.github.io/courses/index.html?course=gnn-2023>

课程考核

- ① 随堂点名
- ② 课后作业：1 ~2 次 (bonus)
- ③ 期末作业：课程报告（论文）



准备知识

课程准备知识：

- ① 线性代数（矩阵运算，矩阵性质，特征值特征向量）
- ② 机器学习基础（有监督，半监督，无监督，训练，测试）
- ③ 概率论与数理统计（条件概率，常见随机分布，贝叶斯网络）
- ④ 图论基础（连通性，中心性，拉普拉斯变换）

代码准备知识：

- ① Sklearn
- ② Pytorch
- ③ NetworkX
- ④ DGL¹ or PyG

¹<https://www.dgl.ai/>



NetworkX

NetworkX²:

Contact

[Mailing list](#)

[Issue tracker](#)

[Source](#)

Releases

Stable ([notes](#))

2.6.2 — July 2021

[download](#) | [doc](#) | [pdf](#)

Latest ([notes](#))

2.7 development

[github](#) | [doc](#) | [pdf](#)

Archive



NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.



Software for complex networks

- Data structures for graphs, digraphs, and multigraphs
- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be "anything" (e.g., text, images, XML records)
- Edges can hold arbitrary data (e.g., weights, time-series)
- Open source [3-clause BSD license](#)
- Well tested with over 90% code coverage
- Additional benefits from Python include fast prototyping, easy to teach, and multi-platform

©2014-2021, NetworkX developers.

²<https://networkx.org/>

DGL

DGL³:

The screenshot shows the official website for DGL (Deep Graph Library). The header features the DGL logo and navigation links for About, Get Started, Tutorials, Blogs, Docs, Forum, and GitHub. The main banner has a blue background with a network graph pattern. It displays the text "DEEP GRAPH LIBRARY" and "Easy Deep Learning on Graphs". Below the banner are two buttons: "Install" (green) and "GitHub" (white). The page content is organized into three columns: "Framework Agnostic" (with PyTorch, TensorFlow, and Apache MXNet logos), "Efficient And Scalable" (describing message passing primitives and multi-GPU acceleration), and "Diverse Ecosystem" (mentioning domain-specific projects like DGL-KE and DGL-LifeSci). A section titled "Find An Example To Get Started" includes a search bar and a table showing a paper example.

DEEP GRAPH LIBRARY

Easy Deep Learning on Graphs

Install GitHub

Framework Agnostic

Build your models with PyTorch, TensorFlow or Apache MXNet.

m

Efficient And Scalable

Fast and memory-efficient message passing primitives for training Graph Neural Networks. Scale to giant graphs via multi-GPU acceleration and distributed training infrastructure.

Diverse Ecosystem

DGL empowers a variety of domain-specific projects including [DGL-KE](#) for learning large-scale knowledge graph embeddings, [DGL-LifeSci](#) for bioinformatics and cheminformatics, and many others.

Select tags: heterogeneous graph, ...

Paper Title	Tags
A graph-convolutional neural network model for the prediction of chemical reactivity	molecules, reaction prediction

Find An Example To Get Started

³<https://www.dgl.ai/>

课程参考资料

- ① Graph Representation Learning⁴ William Hamilton
- ② Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges⁵ Michael M. Bronstein
- ③ 学术论文 (ICLR, NIPS, ICML, KDD, WWW)
- ④ 开源框架文档



WILLIAM L. HAMILTON
McGill University
2020

Geometric Deep Learning
Grids, Groups, Graphs,
Geodesics, and Gauges

Michael M. Bronstein¹, Joan Bruna², Taco Cohen³, Petar Veličković⁴

May 4, 2021

PRE-PUBLICATION DRAFT OF A BOOK TO BE PUBLISHED BY
MORGAN & CLAPPOOL PUBLISHERS.
UNPUBLISHED MATERIAL IS THE PROPERTY OF THE AUTHOR.
ALL RELEVANT COPYRIGHTS REST WITH THE AUTHOR AND
PUBLISHER EXTEND TO THIS PRE-PUBLICATION DRAFT.

Chapter: William L. Hamilton. [2021]. Graph Representation Learning:
Spherical Functions Applied to Signal Processing and Machine Learning, Vol. 34.
See: Page 1-18

¹Imperial College London / UK IDSA / Tunisie
²New York University
³Microsoft AI Research, Qualcomm AI Research is an initiative of Qualcomm
Technologies, Inc.
⁴DigitalOcean

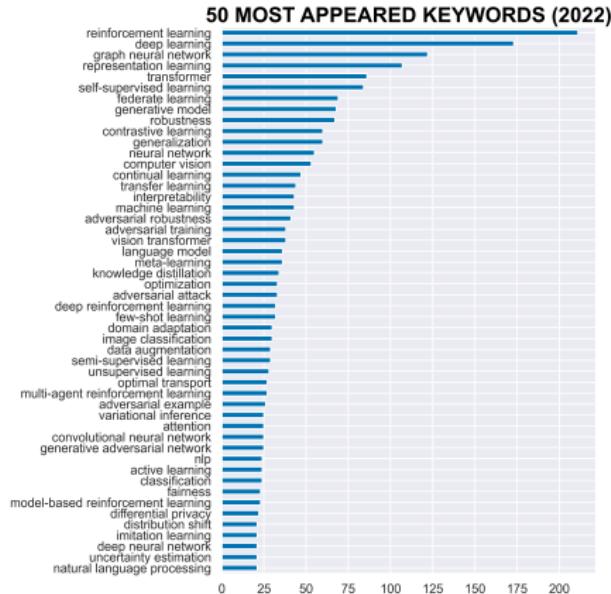
⁴https://www.cs.mcgill.ca/~wlh/grl_book/

⁵<https://arxiv.org/pdf/2104.13478.pdf>

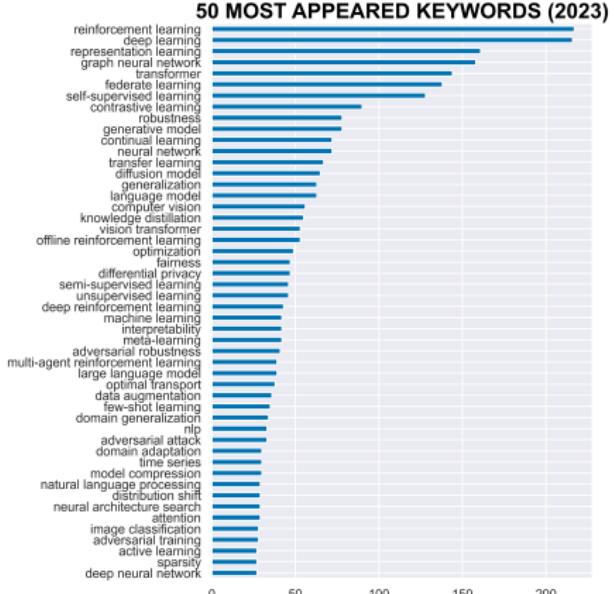


课程介绍

人工智能顶级会议 ICLR 上的图神经网络发展趋势



ICLR 2022



ICLR 2023

课程目标

本课程希望通过 8 次课的时间，帮助大家了解和掌握如下内容：

- ① 图神经网络的经典算法
- ② 图神经网络的实际应用
- ③ 图神经网络的理论基础
- ④ 图神经网络的**工业应用**
- ⑤ 动手图神经网络
- ⑥ 论文阅读能力 (Optimal)
- ⑦ 论文写作能力 (Optimal)



1 课程介绍

2 从图数据到图神经网络

3 图特征

4 图数据挖掘任务



认识图数据

图的定义

图是一种描述样本间关系的通用语言。许多复杂的模式往往需要通过数据之间的结构关系进行分析。

图的数学定义

图一般定义为： $G = \{V, E\}$ ，其中 V 是节点的集合， E 是边的集合。

图数据并不一定是显式构建或真实存在的，可以是用于建模没有显式关系的独立数据。

认识图数据



计算机网络



社交网络



交通网络



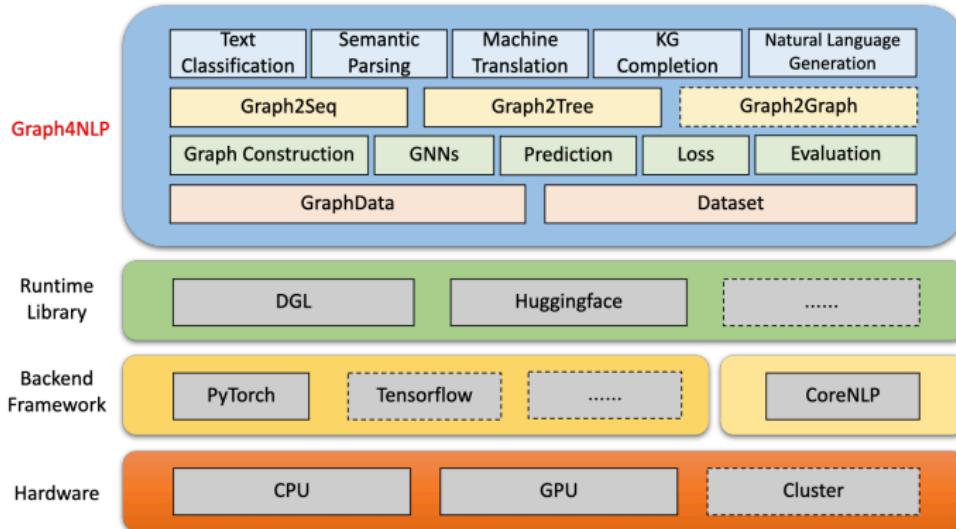
神经元网络



卫星网络



认识图数据



Graph4NLP⁶

⁶<https://github.com/graph4ai/graph4nlp>



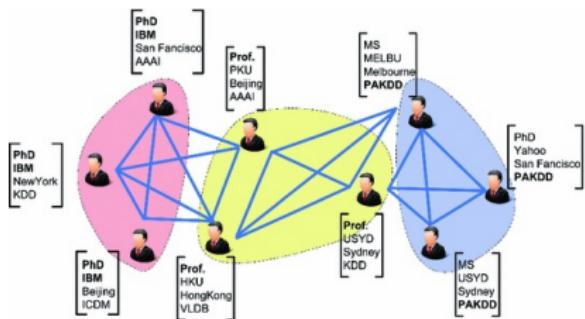
认识图数据

常见的复杂图：

- ① 属性图 (Attributed Graph)
- ② 有向图 (Directed Graph)
- ③ 有权图 (Weighted Graph)
- ④ 多关系图 (Multi-relational Graph)
 - ① 异构图 (Heterogeneous Graph)
 - ② 多样图 (Multiplex Graph)
- ⑤ 动态图 (Dynamic Graph)
- ⑥ ...



属性图 (Attributed Graph)



属性图的特点：

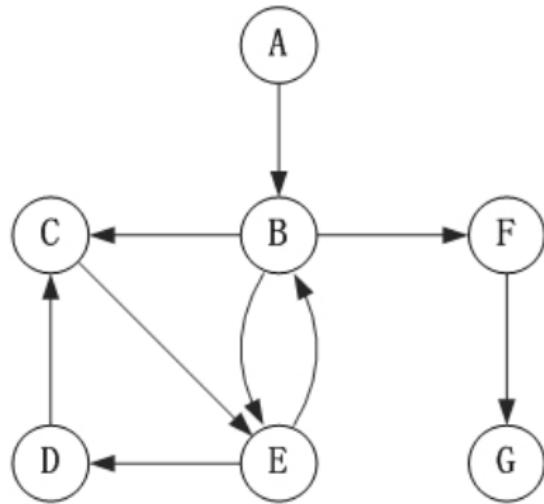
- ① 节点间关系度量多元
- ② Feature Homophily
- ③ 拓扑信息与属性信息融合
- ④ 拓扑信息与属性信息解耦

常见的属性图：

- ① 社交网络
- ② 生物信息网络
- ③ 学术网络



有向图 (Directed Graph)



有向图的特点：

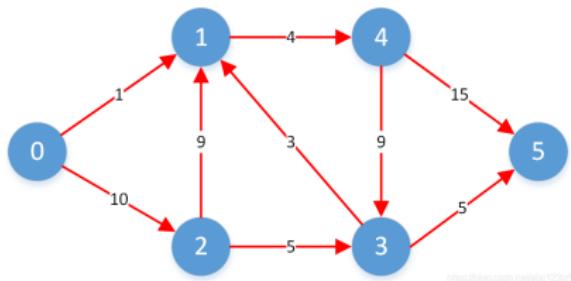
- ① 节点间关系非对称
- ② 单个节点属性多样化
- ③ 网络存在层次关系

常见的有向图：

- ① 金融网络交易关系
- ② 社交网络关注关系
- ③ 论文网络引用关系



有权图 (Weighted Graph)

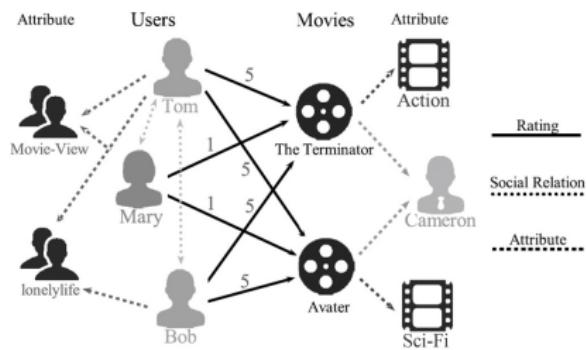


有权图的特点：

- ① 节点间关系非等价
 - ② 节点多样化邻居
 - ③ 局部结构多样化
 - ④ Signed Graph
-
- ① 金融网络交易关系
 - ② 交通网络流量关系
 - ③ 安防网络流量关系



异构图 (Heterogeneous Graph)



异构图的特点：

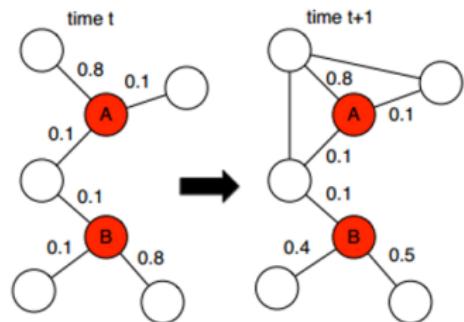
- ① 节点类型多样
- ② 边类型多样
- ③ 节点关系多样
- ④ 结构特征多样

常见的异构图：

- ① 金融网络
- ② 交通网络
- ③ 社交网络
- ④ 学术网络



动态图 (Dynamic Network)



动态图的特点：

- ① 动态形式多样
- ② 特征描述困难
- ③ 时序挖掘困难

常见的动态图：

- ① 金融网络
- ② 社交网络
- ③ 交通网络



图学习的困难

图学习面临的主要困难：

- ① 图数据数量巨大
- ② 图数据特征高维
- ③ 样本间关系复杂



传统图挖掘

图神经网络

图神经网络定义

图神经网络定义

图神经网络 (Graph Neural Networks) 是一种基于图结构的深度学习方法。

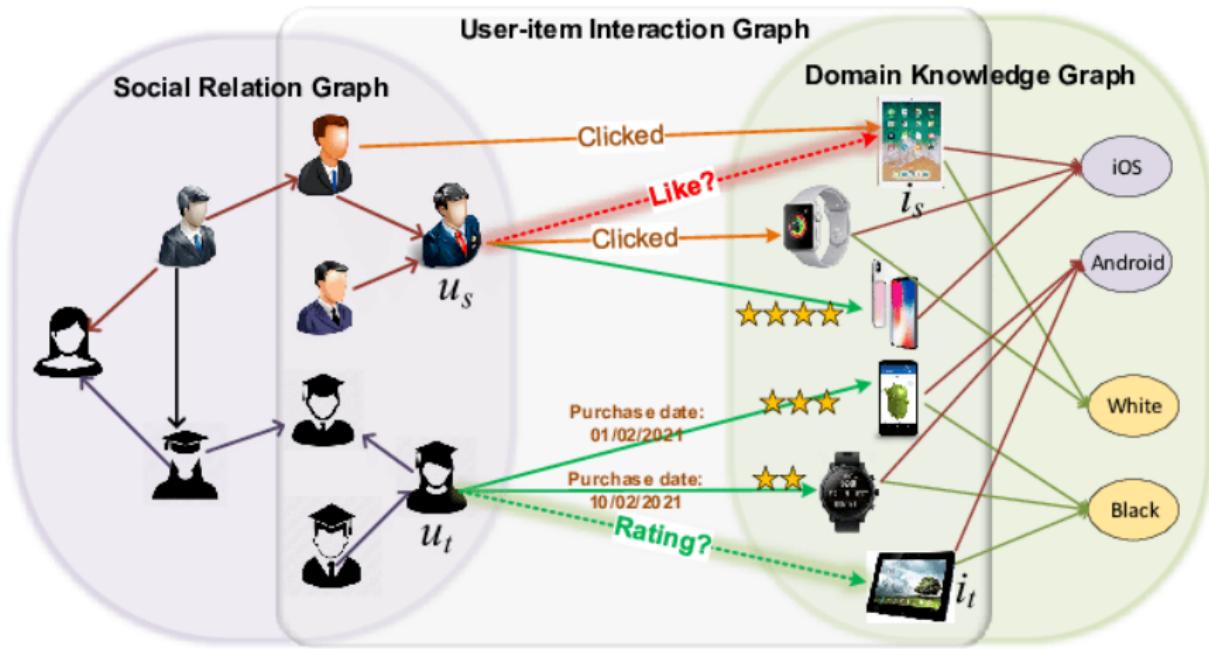
图结构是描述样本间关系的数据形式，在自然界广泛存在。

深度学习是特征提取，复杂关系建模，知识推理的有效工具。

图神经网络已经成为当前图数据挖掘，复杂关系建模，知识推理的主流方法。

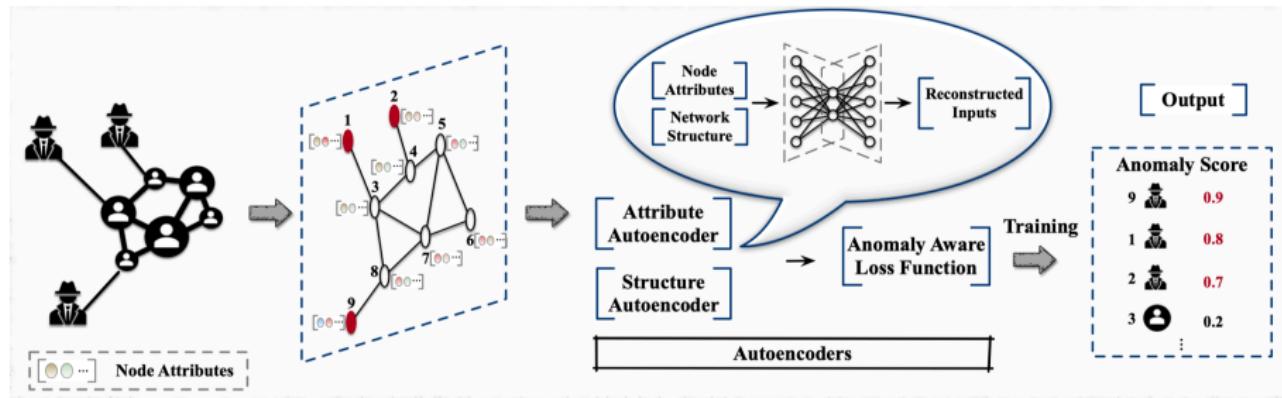
图神经网络的应用

推荐系统



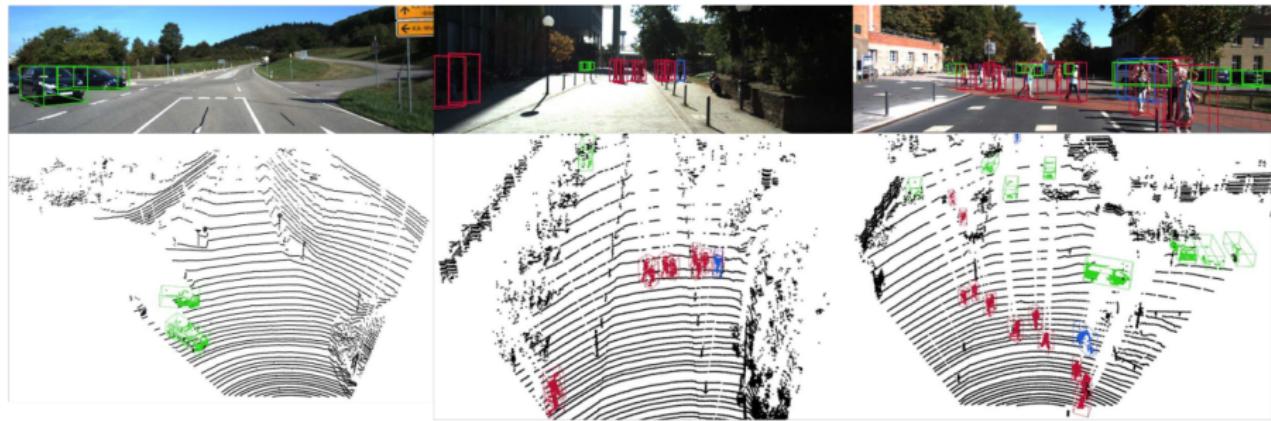
图神经网络的应用

异常检测



图神经网络的应用

计算机视觉建模



课程内容

本课程的主要内容包括：

- ① 经典图表征学习方法 (Graph Representation Learning)
- ② 卷积图神经网络 (Convolutional Graph Neural Network)
- ③ 循环图神经网络 (Recurrent Graph Neural Network)
- ④ 图结构学习的图神经网络 (Graph Neural Network with Structure Learning)
- ⑤ 富信息图神经网络 (Information Rich Graph Neural Network)
- ⑥ 图上的社区发现 (Community Detection on Graphs)
- ⑦ 图上的异常检测 (Graph Anomaly Detection)



1 课程介绍

2 从图数据到图神经网络

3 图特征

4 图数据挖掘任务



图特征

图特征

图特征是传统图机器学习/图数据挖掘的输入，通常依赖人类的先验知识。

- ① 节点的度
- ② 节点的中心度
- ③ 介数中心性
- ④ 聚类系数

图神经网络的发展离不开传统图特征！



节点的度

节点的度

节点的度是指连接到节点的边的个数：

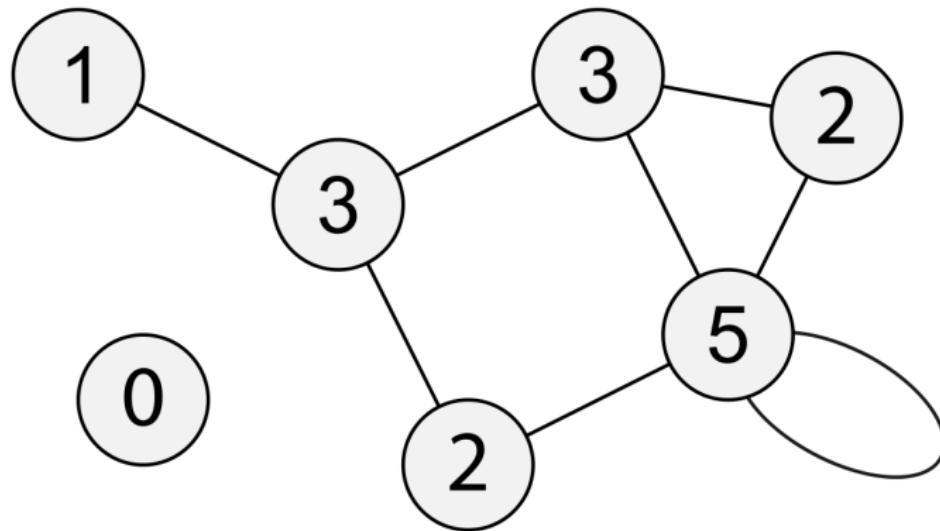
$$d_u = \sum_{v \in \mathcal{V}} A_{u,v}$$

节点的度是图数据挖掘/图机器学习中最重要的特征之一

节点的入度和出度

在有向图中，节点的度又分为入度（In-degree）和出度（Out-Degree）。

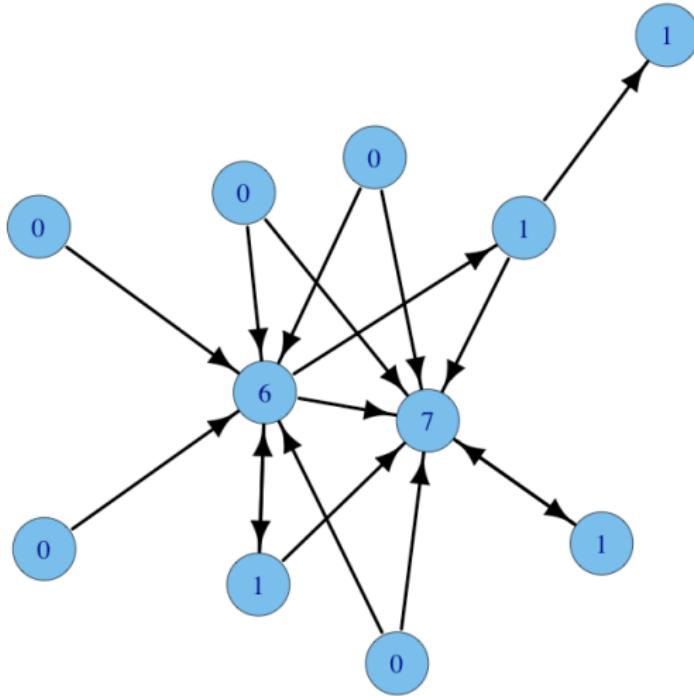
节点的度



无向图的度



节点的度



节点的中心度

中心度 (Node Centrality)

节点的度仅考虑节点边的个数多少，而不考虑节点的重要性（仅用节点的度难以衡量节点的重要性）。

特征向量度 (Eigenvector Centrality)

节点的度仅能度量节点邻居的个数，节点的特征向量度则在度的基础上进一步考虑节点邻居的重要性：

$$e_u = \frac{1}{\lambda} \sum_{v \in \mathcal{V}} A_{u,v} e_v$$

等价为：

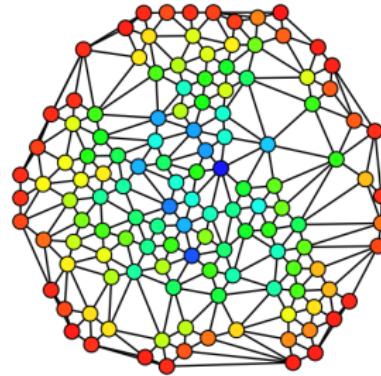
$$\lambda \mathbf{e} = \mathbf{A}\mathbf{e}$$

介数中心性

介数中心性 (Betweenness centrality)

介数中心性 (Betweenness Centrality) 描述了节点属于其他节点的最短路径的概率。

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$



聚类系数

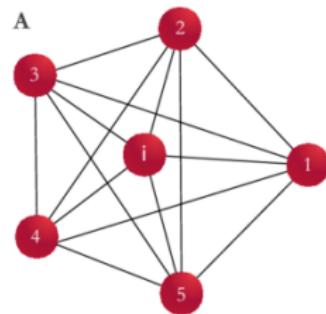
聚类系数 (Clustering Coefficient)

聚类系数 (Clustering Coefficient) 是衡量节点的局部邻居中闭环三角关系 (closed triangle) 比例的指标。

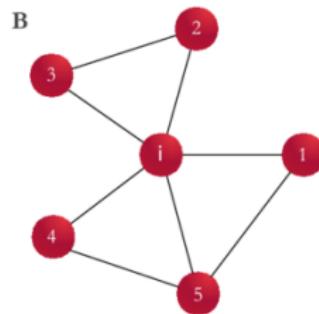
$$c_u = \frac{|(v_1, v_2) \in \mathcal{E} : v_1, v_2 \in \mathcal{N}(u)|}{\binom{d_u}{2}}$$

该指标等价于计算节点的邻居之间边的个数。客观反映了节点邻居的稠密程度。

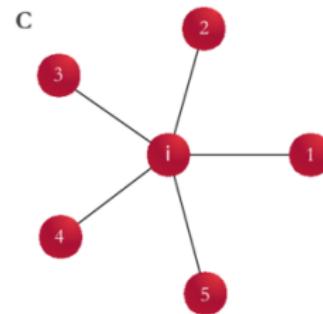
聚类系数



$$CC(i)=1$$



$$CC(i)=0.5$$



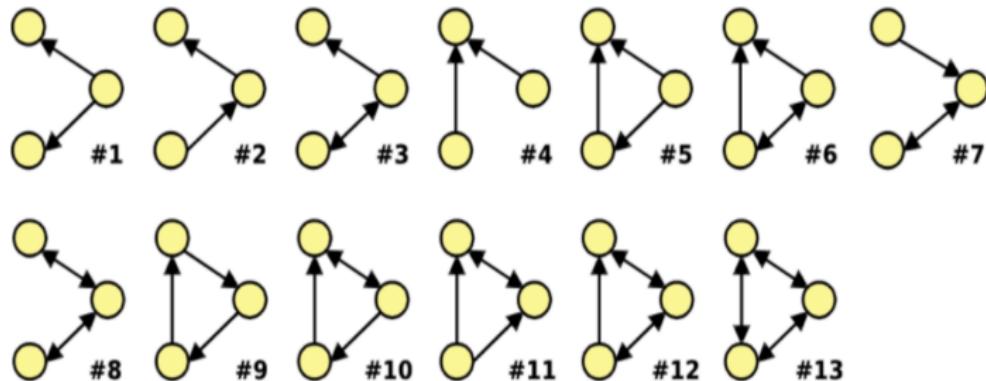
$$CC(i)=0$$



局部拓扑特征

子图 (Subgraph)

子图是节点，节点的邻居以及邻居之间边的集合，它是构成网络的基础单元。



节点数为 3 的有向子图的所有可能情况

图模板 (Motif)

图模板 (Motif)

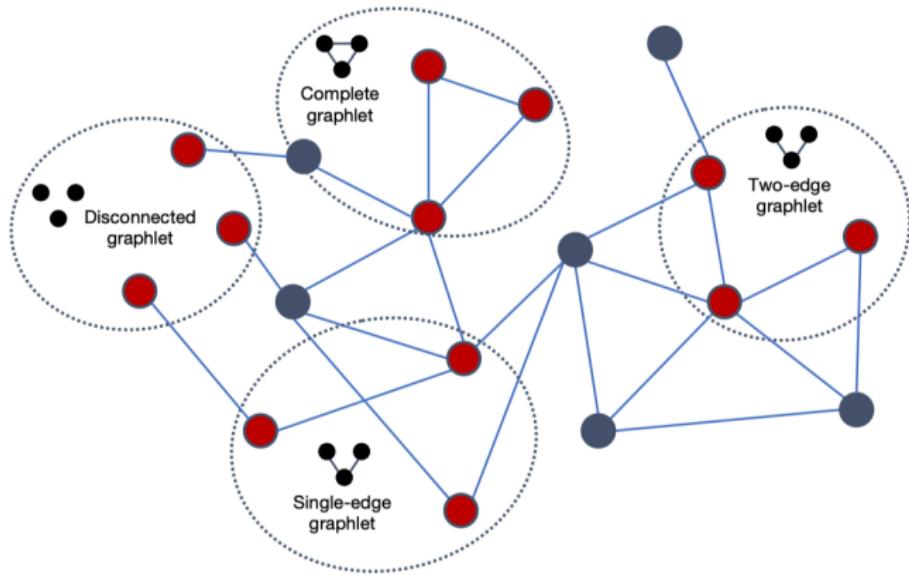
图模板 (Motif) 是指图中一类高频出现 (recurring) 的小的诱导子图 (Small induced subgraph) 的总称，它通常具有如下的特点：

- ① 高频出现：在网络中普遍存在
- ② 诱导子图：节点之间的连接都包含在内

图元 (Graphlet)

图元 (Graphlet) 是图模版的一类扩充。相比于图模板用于描述网络的局部特征，图元通常通过统计的方式描述图的全局分布特性。

局部拓扑特征



图中的图元 (Graphlet)

1 课程介绍

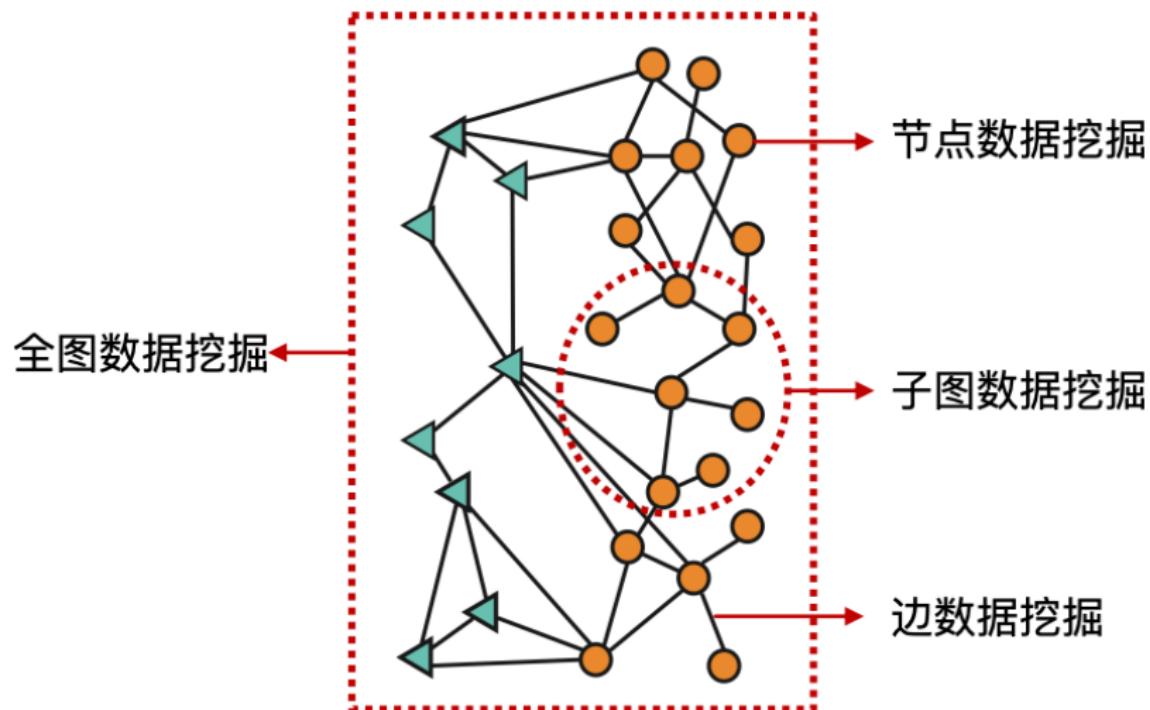
2 从图数据到图神经网络

3 图特征

4 图数据挖掘任务



图机器学习/图数据挖掘主要任务



节点分类

节点分类任务

给定网络中部分节点的标签 y_{train} 以及网络信息 $G = \{V, E, X\}$, 训练分类器预测未知节点的标签。

节点分类任务是图神经网络最常见的任务之一! [Ou et al., 2016]

节点分类任务的应用:

- ① 作弊团伙发现
- ② 用户兴趣定位
- ③ 网络身份识别



节点分类的基本假设

同质偏好 (Homophily)

图上的节点会和它们的邻居有相似的属性/标签，许多节点分类算法假设图上相近的节点具有相似的标签。

- ① 社交网络中，好友之间往往有相似的兴趣偏好。
- ② 论文引用网络中，相互引用的论文之间会有相似的内容
- ③ 作弊行为中，作弊者往往团伙作案且有相似的特性。



边预测

边预测 Relation/Link Prediction

给定一个已观测到的网络 $G = \{V, E, X\}$, 预测未观测到的边。

边预测任务大致可以分为:

- ① 静态图边预测
- ② 动态图边预测

边预测的应用:

- ① 推荐系统
- ② 网络结构优化
- ③ 异常检测



边预测的基本假设

边预测的基本假设有：

- ① 有相似属性的节点之间应该存在边 KNN
- ② 有相似标签的节点之间应该存在边
- ③ 有相似结构的节点之间应该存在边
- ④ ...

边预测的主要挑战是样本间关系难以直接刻画（通常用低维向量表示）



图聚类与社区发现

图聚类与社区发现

图聚类/社区发现是指发现图中的群聚节点（社区）。在同一个社区内的节点，有更大的概率生成边。

社区发现的应用：

- ① 蛋白质分子预测
- ② 犯罪团伙识别
- ③ 商品推荐



课程内容 QA



References I

-  Ou, M., Cui, P., Pei, J., Zhang, Z., and Zhu, W. (2016). Asymmetric transitivity preserving graph embedding. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1105–1114.

