

6-5-cn

1 本节我们学习社交网络分析相关内容

2 数据处理系统提供大数据计算处理能力和应用开发平台。从计算架构的角度，将数据处理系统分为数据算法层、计算模型层、计算平台层、计算引擎层等。

与大数据相关的计算算法包括机器学习算法和数据挖掘算法。计算模型是指不同类型的大数据在不同场景下的处理方式，包括批处理、流计算、结构化数据

的大规模并发处理(MPP)模型、内存计算模型和数据流图模型。

计算平台和引擎**提供各种开发套件和操作环境**，我们选取 Spark Mllib 和 TensorFlow 为例

典型应用我们选择了推荐系统和社交网络分析，本节我们讨论社交网络分析

3 在本节中，我们将学习社交网络、社交网络分析、社交网络分析的度量标准和力导向图绘制  
最后我们将做社交网络分析的实验

#### 4 什么是社交网络

社会网络是由社会结构组成的

- 一组社会参与者(如个人或组织)，
- 二元关系的集合，
- 以及参与者之间的其他社会互动。

它通过节点(网络中的个体参与者、人员或事物)和连接它们的纽带、边缘或链接(关系或交互)来描述网络结构。

这些结构的研究使用社会网络分析来识别本地和全球模式，定位有影响力的实体，并检查网络动态。

让我们看一个关于社交网络概述的视频

5 社交网络可以用图表来表示。

节点指的是社会行动者，边指的是关系。

一条边的权重衡量关系的密切程度。

社交网络的平均距离较低。

如果我们画一张世界上每个人的社交图，将会有超过 70 亿个节点。

但根据六度分离，两个节点的最小距离很可能小于或等于 6。

由于互联网的发展，两个节点的平均距离越来越低。

社会网络分析(SNA)是通过使用网络和图论来调查社会结构的过程。

社会网络及其分析本质上是一个跨学科的学术领域，它起源于社会心理学、社会学、统计学和图论。

让我们看一个关于什么是社交网络分析的视频

6 社交网络分析的一些指标，

节点和边缘的度量包括平均聚类系数;特征向量中心;平均路径长度等。

网络的度量包括平均度、平均加权度、网络直径、图密度、最短路径、点击率、模块化、PageRank、连接组件等。

#### 7 Force-directed graph 力导向图绘制

可视化是分析社交网络的一种重要方法。

力向图是可视化社交网络的好方法。

力向图绘制是社交网络可视化的一项重要技术。

力向图绘制算法是一类考虑审美的绘制图的算法。

目前力向图绘制算法有 Force atlas 算法、Fruchterman-Reingold 算法、Yifan Hu 算法等。  
基本理论是在物理模型中假设节点之间存在力和拟合。  
目标是使整个系统稳定或者能量最低。

8 Fruchterman-Reingold 算法是最常用的算法之一。

该算法利用电磁学原理，生成圆形图。

斥力  $f_a(d)$  等于  $d$  的平方除以  $k$ ；引力  $f_r(d)$  等于负  $k$  的平方除以  $d$

最终状态是使力平衡

左图为原始图，右图为 Fruchterman-Reingold 风格，更有美感。

## 9 社交网络分析 SNA 实验

我们将在《权力的游戏》社交网络数据集的基础上，使用 NetworkX 和 Gephi 进行实验，

NetworkX 是一个 Python 包，用于创建、操作和研究复杂网络的结构、动态和功能。

Gephi 是一个开源和免费的可视化和探索软件，用于各种图形和网络。

实验 1 包括加载数据并生成图表、网络分析、使用 Networkx 进行可视化以及使用 datashader 进行边缘绑定。

实验 2 包括负载数据分析、Fruchterman-Reingold 算法简化图、可视化美化

实验 3 包括社交网络推荐环境设置、有偏随机游走、节点表示学习和推荐。

平台上提供了所有的实验材料，包括手册和代码，可以帮助大家完成实验。

10 本节我们学习了社交网络分析，今天的内容就到这里，谢谢大家