



中國人民大學  
RENMIN UNIVERSITY OF CHINA

# 计算传播理论与实务

2019-2020秋季学期

## 课程总结

授课教师：范举副教授、塔娜讲师

时间：2019年12月23日

# 计算传播概论

- 计算社会科学
- 大数据带来了
  - 新数据
  - 新方法
  - 新挑战
- 数据科学
  - 基本概念
  - 工作流程
  - 能够回答的基本问题
- 数据科学模型和社会科学模型的异同

# 统计思维

- 描述统计
  - 辛普森悖论
  - 探索数据分析
  - 数据可视化常用图表
- 推论统计
  - 总体与样本
  - 参数估计

# 使用机器学习

- 分类问题的训练与预测
- 分类问题方法设计的基本步骤
  - 提出问题、准备数据、选择特征、学习算法、评价模型
- 分类问题的评测指标
  - 精度Accuracy
  - 准确率、召回率、F值
  - 灵敏度、特异度

# 文本分析

- 文本分类
  - 文本分类的定义
  - 文本分类的常见方法
  - 朴素贝叶斯分类方法
  - 支持向量机分类方法
  - 文本分类的Python实现
- 数据采集
  - 网页抓取
  - 网页解析
  - 数据采集的Scrapy实现Scrapy

# 文本分析

- 文本分词
- 文本检索
  - 文本检索的定义
  - 倒排索引
  - 排序模型
- 情感分析
  - 情感分析的主要任务
  - TF-IDF模型
  - 基于朴素贝叶斯的情感分析
- 文本分析Python包的使用

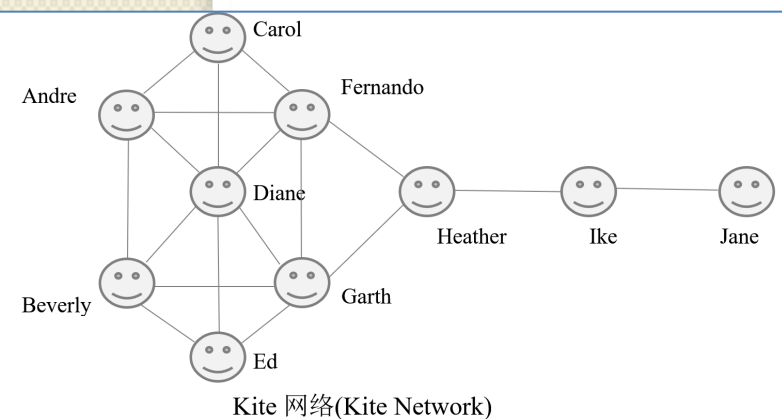
# 网络分析

- 图论的基本概念
  - 图的形式化定义
  - 有向图/无向图
  - 边的权重
  - 路径
  - 连通图/连通分支
  - 最短路径与距离

# 网络分析

- 节点中心度分析
  - 度中心度
  - 相近中心度
  - 中介中心度

正规化  $C(2, n - 1)$



网络各个节点的中心性

Degree		Closeness		Betweenness	
Diane	0.667	Fernando	0.643	Heather	0.389
Fernando	0.556	Garth	0.643	Fernando	0.231
Garth	0.556	Diane	0.600	Garth	0.231
Andre	0.444	Heather	0.600	Ike	0.222
Beverly	0.444	Andre	0.529	Diane	0.102
Carol	0.333	Beverly	0.529	Andre	0.023
Ed	0.333	Carol	0.500	Beverly	0.023
Heather	0.333	Ed	0.500	Carol	0.000
Ike	0.222	Ike	0.429	Ed	0.000
Jane	0.111	Jane	0.310	Jane	0.000



# 网络分析

- PageRank分析
  - PageRank的计算
  - PageRank的应用
- 社区检测
  - 三元闭包
  - 强关系、弱关系
  - Louvain算法的优化目标
  - 模块度的计算

# 大数据与伦理

- 信息茧房
- 算法歧视
- 算法推荐

# 成绩评定

- 期末考试 – 30%
- 三次作业 – 60%
- 课堂参与 – 10%



计算技术和传播理论

相向而行

Let us meet in the middle!