

# 数据分析项目书（一）

## 项目背景

在工业界，根据业务方向不同数据分析又可以称为数据科学或商业分析，从业人员需要掌握扎实的数学功底、机器学习技术基础和编程能力。

人以群分、物以类聚。在校园里，不同的兴趣社团让不同兴趣的同学聚集在一起；不同的专业，又将同学们的社交网络作了稳定划分。与校园类似，在企业里，不同的部门和项目组，也会将企业员工划为不同的社群。那么，在社群中究竟谁是最有影响力的人，谁又是最关键的人物？

找出关键节点（KOL），用处有很多如广告营销、产品运营、战略投资。而社群划分，也可以让产品经理快速定位目标人群，从而实现精准的广告投放等等。

## 项目要求

- 学习图的基本概念，例如节点、度、边、有向图、无向图等。
- 学习并掌握图的基本算法，例如最短路径、pagerank、LPA、fastunfolding等。
- 学习图处理工具networkx，掌握图可视化方法。

## 数据介绍

本次项目提供某企业图数据，数据格式为：

1	1,2
2	1,3
3	5,9
4	9,12
5	...

## 作业提交

- 获取社交网络数据，并利用networkx、matplotlib等工具对企业图数据作可视化分析。
- 利用上述企业图数据，结合pagerank算法找出网络中的关键节点。
- 利用上述企业图数据，结合LPA等算法对网络做社群发现。

## 时间节点

20200116-20200122