**图计算分析思路及任务分配**

一、常规分析

分析创业者数据中的学历分布、年龄分布、地域分布、性别分布等。哪些公司的人出去喜欢创业，哪些公司的人出去创业成功率最高。其他维度可再添加。这个主要由小甘来处理。

数据存放地址，teambition上的文件库/创投数据/160718\_创业者\_it桔子全.xlsx。

二、**图计算**。

图计算是我们本次项目的一个很重要的数据分析和挖掘任务，也是我们需要展示的最终成果。鉴于我个人时间和精力有限，同时各位具有计算机专业相关背景，对数据结构中的“图”都比较了解，因此这个任务交由小甘、陈总、杨总三位高手来完成，并由小甘牵头来完成这项任务。

**目标**：

分析和挖掘投资人、投资机构、创业公司、创业者、投融资项目之间的千丝万缕的关系，形成一个全面且深入的**社交关系图谱**。这样可以深入了解投融资圈的整体情况、行业热度、投资人偏好等。

工具：

Spark：图计算；

Echart：数据可视化；

其他工具。

**分析思路：**

1. 以投资机构为主线。看该投资机构投资了哪些公司。已启明创投为例，启明创投投资了拉勾网、逻辑思维、知乎等公司。已启明创投为中心节点，可得出启明创投与所投公司的一个关联图。（括号内的可先忽略。得出这个关联图后，我们可做其他维度的分析与展示，例如，展示所投公司现在所处的状态（天使、A轮、上市等）、所投公司的行业分布、所投公司的发展状况等等）。这样的话项目或者公司通过投资机构间接关联上。
2. 以项目或创业公司为主线。看该项目或创业公司获得了哪些投资机构的投资。这个可以按照轮次来区别不同的投资机构、也可按投资性质来分（财务投资、战略投资）、投资比例（领投、跟投等）等等。这样的话投资机构通过项目或公司简介关联上。
3. 以创业者为主线。看创业者的工作学习经历（派系）、创业经历、和投资人、投资机构的私人关系。找出创业者与创业公司、投资机构、投资人以及创业者之间的社交关系。
4. 以投资人为主线。看投资人的工作学习经历、创业经历等。找出投资人与投资机构、投资项目、投资人之间的社交关系。