

3月7日会议讨论纪要：

1. 研究新三板的公司的现状和原因、探索国内注册资本市场的可能性。需要先查看新三板数据的信息（格式、内容）同时把新三板公司与另一份数据对应起来（后一个数据等张千明）。
2. 下载并分析CrunchBase数据。
3. 在IT橘子上建立联合投资网络、通过pagerank算法来对VC排序。具体参考Scientific collaboration network。
4. 在IT橘子上做[15]上的实验，原文利用断点回归来分析天使投资对被投公司的影响。关于这个问题，我们IT橘子上的数据只有被投资公司，需要补全没有被投资公司的数据。

后续安排

- [deadline: 3-7] Tony查看IT橘子的数据和涛哥给的是否一致
- [deadline: 3-11] 陈志怀阅读Scientific collaboration network相关文献，并完成3。
- [deadline: 3-8] 陈志怀和Tony一起分析CrunchBase、新三板（名单，财务数据）获取的可行性分析。
- [deadline: 待定] 张千明从LiQian获取数据并抽空一起讨论数据的使用

OMP-PR算法及效果

建图

1. 二部图 $G=(L, R, E)$, $L = \{\text{公司}\}$, $R = \{\text{投资者}\}$, 两者有边当且仅当投资者投资过对应公司。
2. 二部图 $G=(L, R, E)$, $L = \{(\text{公司}, \text{融资轮次})\}$, $R = \{\text{投资者}\}$, 两边连边当且仅当投资者在对应轮次投资该公司

One Mode Projection

- 周涛的方式: $w'_{ij} = \frac{1}{d(j)} \sum_{l \in L} \frac{w_{li} w_{lj}}{d(l)}$ ¹
- Newman的方式: $w'_{ij} = \sum_{l \in L} \frac{w_{li} w_{lj}}{d(l) - 1}$ ²

对比数据

清科2016年度中国股权投资，包含早期投资机构前30和VC投资机构前50和PE投资前50。该数据通过对投资公司的管理资本量、新募集基金资本量、资案例的个数、投资资本量、退出案例个数、退出金额、回报水平等维度来评价。

对比方式，查看了这130家公司在我们PageRank的排名如何：

重名数量

前5%数量

前10%数量

早期投资公司	23/30	21/30	23/30
VC投资公司	38/50	28/50	31/50
PE投资公司	25/50	16/50	17/50

数据方面进度

- 确认之前@张千明 提供的IT桔子数据与@周涛 提供数据有差异，完成初步整理，待进一步分析，并与现有数据合并去重
- 得到@张千明 反馈，新三板部分企业缺详细工商数据。下周内取得该数据，并询问是否能取得现有IT桔子名单中企业的工商数据。

后续安排

- ☐ 寻早并对比其他数据排名
- ☐ 查找排名靠后的公司原因
- ☐ 本次建立二部图没有权重，改进之再看效果
- ☐ Crunchbase & 新三板报表数据可用性分析
- ☐ 两份IT桔子数据合并去重
- ☐ IT桔子数据的行业/时间线分布

References

1. Bipartite network projection and personal recommendation [↗](#)

2. [Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality](#), M. E. J. Newman, *Phys. Rev. E* **64**, 016132 (2001). [↗](#)