

7.1 基于微信数据的邓巴圈计算项目

7.1.1 项目时间

2017.05.15-2018.12.30（提出数据需求起算）

7.1.2 研究团队

罗家德老师团队；王晓冰老师团队；贺鹏老师团队

7.1.3 研究目的

1) 利用微信数据计算邓巴圈，根据微信互动数据将用户的微信好友按照不同关系强度（tie strength）进行分类；

2) 最终得到训练好的分类模型，将其运用于随机抽样得到的微信大数据中（北京上海广州 3 市每个城市 7,000 人样本）；

3) 算出不同人群的人脉情况，不同性别，不同年龄层，不同城市的人各会有拟似家人、熟人、认识之人以及微信中的间接连带各多少人？

4) 根据对微信大数据的分析结果结合中国本土人脉理论（费孝通，1998；黄光国，1988）对话邓巴理论，指出中国人脉结构的独特特征，及其对社会网络理论的贡献。

7.1.4 未来可能应用

1) 帮助微信使用者划出自己的人脉图谱；

2) 多期资料的分析可以帮助微信使用者看到自己人脉的变化情况；

3) 进一步从人脉算出社会资本，对个人升迁、找职、信用、借贷、心理健康、幸福感、社区参予以及公共事物态度都具有预测效果。

7.2 基于 QQ 大数据意见领袖识别项目

7.2.1 项目开始时间

2018 年 7 月

7.2.2 研究概述及研究目的

针对腾讯员工的绩效进行分析和预测，主要通过探讨员工人口属性变量、员工的社会资本、员工的培训记录、活水情况、受奖励情况以及员工在小组内部的位置（即是否是部门主管的班底、圈子或者是圈外人的关系）以及员工所在非正式圈子的社会资本的情况等七部分的变量与员工绩效之间的关系，从而对员工绩效进行预测。

其中本项目重点分析员工在小组领导圈子中的位置、社会资本、在非正式领导圈子中的社会资本以及其他人力资源相关指标对于员工绩效的影响，探究公司中的正式关系、非正式关系对于员工绩效所产生的影响。这部分研究内容不仅具有理论意义，同时兼具实际应用价值。

7.2.3 研究应用

1. 对于员工的绩效进行预测，探究员工产生不同绩效的原因，针对一些在组织、领导可控的因素上更好的引导和培训，以促进员工绩效的提高。

2. 刻画公司内部非正式关系图谱，探究组织中的非正式圈子的形成及规律，同时研究员工在非正式组织中的社会资本对于员工绩效的影响。

7.3 人脉圈层研究

当今社会处于一个信息爆炸，万物互联的状态，如何从错综复杂的人际关系中梳理出亲疏远近的人脉圈层，具有重大的理论与现实指导意义。人脉圈层的研究不仅能够帮助个人更好的管理关系、整合资源。帮助企业找寻可靠的生意伙伴，从而发现更好的投资及合作机会。同时中国人脉的圈层研究能够帮助我们比较中西方社交的异同点，从而挖掘出隐藏在中国社交背后的差序格局。在本研究中，我们采用了将社会科学与大数据结合的研究方法，利用理论指导数据挖掘，利用数据挖掘的结果重新认识梳理理论，在理论指导与数据挖掘不停迭代，反复对话的过程中，逐步提高了圈层分类算法的准确度，同时得出具有中国特色的人脉圈层理论。

研究结果及总结：

本次研究通过选取基于心理学及脑科学解释的邓巴圈理论和基于文化解释的差序格局理论，对于关系强度的输出问题进行更为细致的讨论，得出关系强度更合理的输出结构应该是从支持群体到部族的 5 分类变量或是家人-熟人-认识之人的 3 分类变量。最终研究通过分析有监督学习的模型建立过程，特别是基于同一组特征搭建的有监督学习模型的检验准确率，指出该检验准确率可以提供给研究者用于比较操作化理论方法与真实数据的拟合程度。

在接下去的部分，文章将介绍本次研究所使用的方法以及数据，并最终计算这两种理论与数据的拟合程度，即基于两组理论搭建的有监督学习模型的检验准确率。

7.4 中国风险资本（VC）联合投资网络研究

（1）按计划执行情况。

本研究按原先计划进行，执行情况良好，完成了英文期刊及书中章节论文四篇，英文会议论文六篇，以及投稿中在审稿论文两篇。中文期刊论文两篇，会议论文两篇以及投稿中在审稿论文三篇。同时发展了下一波研究方向，由风险投资产业联合投资网单一网络分析走向 VC 联合投资网，董事会联结网以及上市企业联合研发网的三层网分析，从而看到三层网内部及之间的互动可以对行动者造成什么样的影响。

（2）研究目标完成情况。

研究期间，因为王俊程老师还在台湾，参予较少，所以建立模型的主要工作由新增伙伴，北京师范大学系统科学学院樊瑛教授及其团队，提供必要的建模工作。同时多名助理在此期间毕业或离职，又有多位助理加入，这将在人才培养中描述。

7.5 兴趣社交研究计划（2018 年拟申请）

项目起点：社群自组织理论

1 概念及应用价值

自组织过程理论的研究本身就是具有应用意义的，新媒体环境下兴趣社群形成过程的范式研究对于发起、培养以及扩大兴趣社群具有重要意义，这样大规模的社群组织长期依赖于产品进行互动以及完成共同目标，势必会为产品带来极大的用户粘性，而针对不同兴趣社群提供个性化的产品功能、营销服务也就可以提上日程。同时，兴趣社群自组织过程理论的架构为下一段利用大数据绘制兴趣社群的分化图谱和结合人脉圈子的研究打下基础，达到大数据和理论的结合，对用户粘性的要点进行深入探究。

2 学术意义

对于本次研究来说，自组织理论研究主要产生以下三个具有学术意义的研究成果：

- ◆ 根据扎根理论，对依托于微信、QQ 等新媒体社交平台下建立的兴趣社群进行实际的调研和访谈，构建兴趣社群的形成过程理论，对兴趣社群形成过程进行刻画，使兴趣社群过程理论对话传统社会学中自组织理论，为新媒体下兴趣社群的研究提供范式。
- ◆ 基于对兴趣社群形成过程动态的研究，可以对比新媒体社交平台兴趣社群和传统自组织特点及形成过程的异同，探究其异同背后的原因。
- ◆ 为下一步结合大数据进行兴趣社群的研究提供理论基础和科学依据，为大数据下模型的训练和验证、以及如何将社群弱关系、间接关系转化为强关系进行铺垫。

7.6 基于电信数据的社会资本挖掘项目

本次项目试图利用中国电信的后台数据，对电信用户的社会资本进行测量，并且发展出一种成熟的数据处理方式，使得社会资本指标可以在广泛的经济社会问题中得到应用。

值得说明的是，本次项目实际是一项应用导向的项目。经典的社会学理论从逻辑的角度说明了社会资本与人们在经济生活中的行为存在着深刻的联系，并且已有的经济模型往往忽略了这一指标可能产生的作用。本次项目希望能够通过通信数据，将社会资本指标引入传统的经济学模型，使得传统的经济学模型得到改善，并且基于这种改善实现实践性的价值。

7.7 新浪微博大数据研究

基于 CGSS（2008）的微博情感倾向抽样调查方法修正

通过数据挖掘的方法收集民众在日常上网中发表的对于时事、政策的言论是了解舆情的有效途径之一。与传统的问卷调查相比，该方法拥有更快速、更真实、更及时、样本量更大等诸多优点。由于样本量巨大，用人力对网络言论进行甄别、分类往往是不现实的，于是研究者们发展出了各种方法利用计算机对网络言论中的语义进行识别。

以微博为例，基于微博内容的语义分析可大致分为面向事实的文本挖掘与面向观点的文本挖掘两类。前者包括热点话题探测（杨亮、林原、林鸿飞，2012）、主题抽取、垃圾信息处理、自动摘要等几类，目的在于归纳微博内容的基本语音；而后者则是从用户发布的微博信息中挖掘出其讨论的主题与对该主题的情感偏好，发布者对于该话题的情绪是正面的还是负面的。本研究的主要目的在于找到更准确、方便的方式了解民众态度，因此主要关注面向观点的文本挖掘，即民众对于不同话题的情感偏好。

中国综合社会调查是中国第一个全国性的综合、连续、大型社会调查项目。调查涵盖从 2003 年起的全国 125 个县（区），500 个街道（乡、镇），1000 个居（村）委会，10000 户家庭中的个人。通过定期、系统的收集中国人与中国社会各方面的数据，总结社会变迁的长期趋势，探讨现实社会中的各项议题。目前，CGSS 与互联网相结合的研究多为以 CGSS 数据为基础，探讨其他变量与网络行为之间的相互影响。如陈云松（2013）在研究中发现互联网的日常使用对于城市居民的非制度化政治参与有促进作用。在本研究中，我们将按照 2008 年（下同）CGSS 数据中的若干人口学特征（如性别、年龄、教育程度等）比例来对微博用户数据进行筛选抽样，使筛选出的微博用户在这些特征的比例上与 CGSS 数据保持一致，再利用筛选出的用户的微博言论探究民众对于某一主题或政策的情感倾向。最后，我们会从 CGSS 问卷中挑选出较为类似的问题，以便对网络情感倾向进行校对，确定网络数据的判断是否准确。如果对于若干个话题，网络数据所得出的情感倾向与 CGSS 统计结果均较为一致，那么我们就找到了一种更为快捷的了解中国人与中国社会的方法。