

## 推动社会计算研究与社科理论结合

杨虎 郑路 蒋理慧 中央财经大学信息学院 清华大学社会学系

人工智能、大数据、复杂系统、动态网络等将成为社会科学新的理论增长点。同时，社会科学与自然科学的文理交叉、文理交融会对二者的发展起到促进作用。社会发展已然在技术变革的作用下悄然加速，直面社会发展所面临的真实问题（疫情应对、政府治理、大型公共活动、风投网、舆情、数据伦理、因果科学），可借助新的信息基础设施，探索建立新的社会计算研究方法，架构未来智能社会。近日，由中国人民大学信息学院孟小峰教授与清华大学社会学系罗家德教授召集的2020年第五届全国大数据与社会计算学术会议（China Conference on Big Data & Social Computing, BDSC2020），汇聚社会科学家、计算机学家、物理学家与统计学家于一堂。

### 大数据为社科理论提供证据支持

复杂理论和大数据方法，可用于讨论产业制度转变和治理结构的转变。这类转变表现为多层次和多时点演化，每个层次和时点的演化都是非线性的。产业发展实际上是宏观行动者与微观行动者频繁互动的结果，宏观行动者会通过多层网传递到每一个微观行动者；每一个微观行动者也会通过层层网络的演化，产生非线性涌现现象，以至于涌现出集体行动、规范制度，最后从微观个体层层传导到整个产业的宏观层面。

罗家德报告了关于产业制度转变和治理结构转变研究的架构以及其中已完成的部分。该研究发现，从个人行为到最终的系统制度与规范，实际上是层层变化、层层涌现新现象的过程。这一过程中的不确定性非常高，只能逐层进行分析。根据复杂社会系统演化理论建构的 long range correlation 方法，可将这一过程分为五个部分：关系、小团体、社区（或社群）、大型社会网、系统制度与规范。在风险投资产业的研究中，由少数风险投资机构（Venture Capital, 简称 VC）组成的小团体共同演化构成社群，一般为两层关系的组织，信息传递较为顺畅。而 VC 产业对应的是大型社会网，包含三层以上关系网。第一个是圈子，指一群人因为共同投资的关系联系在一起，圈子的带头人是领投者，在投资中起到非常重要的作用；第二个是产业中社区的划分，VC 圈子抱在一起形成社群，区分的因素一般是产权差异、产业差异等；第三个是产业结构的动态演化问题。

中央财经大学信息学院杨虎从系统的视角出发，讨论了 VC 圈子和 VC 的联合投资行为。这类研究关注复杂系统的模型与行为。将复杂性思维引入对金融问题的分析，最初源于跨学科研究，如金融物理学。包括风险投资在内的金融活动结果具有高度不确定性，其原因在于大量行为主体的互相作用：风险投资行业机构之间、投资项目之间都存在着密切联系，同时行为主体的预期也互相影响。因此，基于联合投资网络对 VC 投资行为和“圈子”进行研究，可更好地描述其投资行为和演化特征。若给定联合投资网络，评价投资机构的声望则等价于评价风险投资机构在联合投资网络中的位置。由于不同中心度指标反映了节点在网络中的不同结构特性，使用单一中心度指标来评价节点中心度具有一定局限性，因此综合多个中心度指标的评价方法被用来计算节点影响力，从而计算风险投资机构的排序并识别投资领袖，有较为准确和稳健的结果。

风险投资的投资绩效常和这些“产业领袖”的圈子相关。研究进一步发现，若以 IPO 和并购来衡量风险投资机构的投资绩效，风险投资机构在圈子网络中的位置越靠中心，就越容易成功退出投资项目，其投资绩效越高。

北京师范大学系统科学院樊瑛的研究又进一步从圈子扩大到了社区（community），考察了 VC 产业网层层递进的发展。社区结构可用于观察网络中顶点间的关系，它们可以被分成组，组

内顶点间的连接比较稠密，组间顶点的连接比较稀疏。社区结构是网络中观尺度的表现，是了解整个网络结构和功能的重要途径。已有的复杂网络社区探测算法大多是无监督的聚类方法，只考虑网络的拓扑结构，用于研究社区结构的信息比较局限。半监督的探测方法能够将先验信息引入社区发现过程中，提高社区划分的性能和社区发现的准确性。为更准确地描述 VC 产业的结构，可基于 VC 产业联合投资网络设计半监督的社区探测算法（LL 算法）进行社团划分。

北京化工大学信息科学与技术学院谷伟伟则讨论了整个产业网，并以模型拟合了实际联合投资网结构，从而探讨了一种“精英俱乐部式小世界网络”涌现的过程。研究 VC 网络的形成和演化有助于把握网络形成的微观机制，找到这些机制对宏观网络的影响，总结出 VC 产业的演化规则。在风险投资行业，实力较弱的 VC 一般会围绕在较强的 VC 周围进行联合投资，最后形成以较强的风险投资公司（即投资领袖为枢纽）以联合投资为纽带的合作投资形式，也就是精英团体的小事件网络。从数据中验证是否存在精英团体的小事件网络，可发现 elite（较强的 VC）比 follower（较弱的 VC）联合投资的次数更多，所形成网络的密度更大。同时，通过文献调研法和访谈法可得到网络演化的两种机制——关系嵌入机制和结构嵌入机制，同时可发现：结构嵌入模型反映了真实网络演化的机制，展现出“精英俱乐部式小世界网络”涌现的过程。

### 社会计算推动社会发展

社会计算与社会发展之间存在互融共生的关系。人工智能时代为社会科学带来数据、方法、方法论上的深刻变革，社会科学对大数据、人工智能的深入研究也有助于引导科技向善。

近些年，伴随人工智能在各领域的广泛应用，对于技术的反思也逐渐进入了人们的视野。清华大学社会与金融研究中心郑路以当下最为流行的两款短视频 App——抖音和快手为例，围绕二者在推荐算法上的差异，考察其在内容风格、平台治理等方面的异同，及其所产生的社会和经济后果。抖音致力于“记录美好生活”，算法设计关注头部爆品；快手声称“看见每一种生活”，算法设计注重扶持长尾的草根 KOL，降低不同视频之间关注度的“基尼系数”。相应地，抖音的用户网络中，度的分布更符合幂律分布，即“赢家通吃”；快手的用户网络中，度的分布则介于正态分布和幂律分布之间，表现为长尾分布。不同的算法设计，不仅影响企业的长远发展，而且将产生深远的社会影响。

社会计算的发展方兴未艾，中国社会学定量研究也正在快速崛起。南京大学社会学系陈云松认为，大数据时代的到来纾解了社会学定量研究“微观旨趣”的学术困境，大数据提供了以往问卷调查无法提供的宏观数据，能够重启宏观定量社会分析。通过从海量数据中利用计算社会方法提取构建出宏观社会指标，继而用传统计量回归模型特别是面板模型和时间序列模型，可实现宏观层次的关联分析和因果推断。在方法方面，机器学习在大数据时代的社会学研究中大有可为。其中，无监督学习可以应用于海量文本的社会学分析；基于监督学习的社会预测可推动社会科学量化研究进行高精度预测。

面对新冠肺炎疫情，中国是世界上首个大规模应用人工智能数据、算法、算力应对公共卫生事件的国家。中国社会科学院社会学研究所吕鹏及其合作者深入分析了疫情防控中数字平台企业发挥的作用。他们认为，国家、数字平台企业、其他企业、社会共同形塑了“平台治理场域”，其中，数字平台企业既是治理者，又是被治理对象，是“国家—市场—社会”的中介。在疫情前，我国就已形成国家和数字平台企业“非对称相互依存”的“平台治理场域”；在疫情防控中，数字平台企业积极通过多种类型的自我治理和合作治理参与防控，体现了数字技术发展带来的治理效能，也为企业发展带来了机遇。但从长远看，平台治理的外部压力不会因疫情中获得的合法性而被完全消除。正视平台企业的“自主性”，才能呈现和理解平台治理的完整图景。社会学在此过程中也应当以更具建设性的姿态参与平台治理，共建良好的治理生态。

3S（GIS、RS、GNSS）方法技术可以从时空观角度研究计算社会科学的时空特性，是计算社会科学的重要发展方向。西安交通大学实证社会科学研究所贺力结合自己团队的两项重要研究展示了如何应用 GIS 空间计量方法，实现城市犯罪预防。在预防城市暴力犯罪的研究中，贺力团

队以谷歌街景图像为数据源，利用虚拟环境审计方法采集微观建成环境数据，更为准确地揭示出城市微观建成环境与暴力犯罪的关系。在城市盗窃犯罪的研究中，贺力团队借助高时空分辨率客观数据如手机大数据、夜间灯光大数据等，精准刻画 24 小时城市空间中的人口流动，绘制城市人口的微环境，识别城市盗窃类刑事案件的时空分布，分析城市局部空间环境及场所功能对犯罪机会性的影响。

通过多学科交叉融合，以社会计算为方法论，以人工智能、大数据等信息技术为科学工具，可构建“社会计算试验场”。这有助于深刻剖析社会计算与社会智能的内在机制，实现对新型社会现象的发现与机理揭示，促进社会计算与社会智能的发展。会议还有美国芝加哥大学 James Evans、IEEE Fellow 王海勋、美国亚利桑那州立大学刘欢和德国哥廷根大学傅晓明作为嘉宾演讲，另有“社会计算与 AI”“社会计算与数据伦理”“社会计算与疫情应对”“社会计算与舆情治理”“社会计算与大型公共活动治理”“社会计算与数字空间政府治理”等分论坛。文理交融在大数据研究上崛起的趋势逐渐显现。