**西南科技大学本科毕业设计（论文）开题报告**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **学 院** | 计算机科学与技术学院 | **专业** | 软件工程 | **班级** | 软件卓越1901 |
| **姓 名** | 黄婧宇 | **学号** | 5120196267 | **指导教师** | 刘畅 |
| **设计（论文）题目** | 投融资数据的复杂网络建模与分析 | | | | |
| **一、选题背景（目的、意义）** | | | | | |
| **1.1选题背景**  复杂网络是一种理解现实世界复杂系统的抽象模型。它是由巨大的节点和节点之间错综复杂的关系共同构成的网络结构。它将复杂系统中的实体抽象成节点，将实体之间的关系抽象成连线。虽然数学中的图论也在研究网络，但是现实中的网络会有更多的随机特性。因此，复杂网络一般更加关注网络的统计特征。在网络理论的背景下，复杂网络是一个具有非平凡拓扑特征的图(网络)——这些特征不会出现在简单的网络中，如格或随机图，而是经常出现在代表实际系统的网络中。  复杂网络广泛存在于自然界和社会系统中，研究与分析这些复杂网络将极大的方便我们的生活。相关研究从上个世纪末至今，已经取得可人的成果，并以成为当前复杂性科学、系统科学等多个学科共同研究的热点。其研究内容涉及图论、统计物理、计算机、交通、社会学以及经济学等诸多领域。复杂网络理论无论是从网络的拓扑性质、形成机制，还是网络的传播动力学，即探究局部节点或者连边行为在网络上的扩散行为等，都能够给投融资数据下的融资行业网络分析以较大的启发。  **1.2选题目的**  利用数据计算获取每个城市在每个融资行业的融资企业数量，构建“城市-融资行业”二部分网络，探究城市在行业具有的显性比较优势及行业之间的相似度；  构建“融资行业网络”，根据“产品空间”来对未来可能的融资行业进行预测，同时探究经济发展水平对行业的潜在影响。  **1.3选题意义**  随着时代的发展，科学技术的快速进步，数据的探索与预测已成为当前时代的风口与核心之一。对投融资数据的挖掘也是行业分析相关的一大重要探索内容。从城市层面来说，对于未来行业的预测，是政府机构有效的资金分配工具，其帮助评估行业的战略价值，帮助管理人员估计成功的概率，从而预测行业的未来价值。而从企业的角度来说，能够预测并引导行业的未来发展趋势，判断行业投资价值，为投资者提供良好的决策参考。在预测的基础之上，结合城市的经济发展水平进行分析，能够有效探究经济对于行业的潜在影响，明确城市发展特征，从而为政府部门、投资者以及其他机构提供决策参考或投资参考，便于城市及企业对未来更好的进行一个规划，从而更好地制定行业融资策略和行业发展政策。 | | | | | |
| **二、国内外研究现状综述** | | | | | |
| 传统的经济发展理论无法解释各种产品类型在一个国家中经济的表现。传统的经济学观点认为，工业化工业化会对新产品产生“溢出”效应，从而促进后续增长。然而，这个想法并没有被纳入任何正式的经济模型。解释一个国家的经济，要么侧重于该国资本和其他生产要素的相对比例，要么侧重于技术能力的差异及其基础。这些理论都没有抓住产品之间固有的共性，因为这些产品之间的联系往往都会影响一个国家经济的增长。而在由Cesar a. Hidalgo等四人所攥写的文章《The Product Space Conditions the Development of Nations》[1]中提出了产品空间的概念，它是一个将全球经济中交易的产品之间相关性正式化的概念网络。文献通过显性比较优势RCA[2]来衡量国家在商品出口的竞争能力，并且使用一种称为邻近度[1]的度量方法来量化产品之间的相关性，得到我们的产品空间网络，它是在全球市场上交易的产品之间关系或接近程度所组成的网络。该网络表现出异质性和一个有核心的外围结构: 网络的核心是由金属产品、机械产品和化学品组成，而外围则由渔业、热带和谷物农业组成。产品所组成的集群与利默的产品分类方案有着惊人的相似之处。该网络对于经济政策有相当大的影响，其网络结构特征说明了部门国家具有稳定的经济增长而另一些国家停滞不前，无法继续向前发展的原因。  性质是事物所具有的特性和本质，反映了事物本身所具有的与其他事物不同的根本属性，捕捉事物的本质是研究其衍生行为与演化过程的基石，因而研究网络的拓扑性质对于刻画网络结构具有基本的重要性。而产品空间网络基于复杂网络与网络科学，欲探究产品空间网络的性质，则需对复杂网络进行探索。尽管近年来所兴起的网络科学与传统的图论关于基本概念是一致的，但两者在研究角度和研究方法上有着天壤之别。传统的图论往往着眼于某种规则结构、以及节点数较小的图，然而近年来复杂网络所关注的实际网络中往往有着数以万计的节点数目与连边。在物理学中，研究个体数量较少的系统和研究个体数量较大的系统往往采用不同的方法，前者能够采用类似于经典力学这样的精确方法，而后者往往需要采用类似于统计力学这样的统计学相关方法。从哲学的角度来看，这就是量变到质变。因而不同于传统的图论研究，在面对庞大节点所组成的复杂网络时，我们需要采用截然不同的方式来探究其相关拓扑性质，如平均路径长度、聚类系数、度分布等相关内容[3]。  在探究了网络的拓扑性质之后，网络动力学也是网络研究的一大重点。在日常进行交友时，有这样一句话：“近朱者赤，近墨者黑。”其比喻接近好人可以使人变好，接近坏人会让人变坏，所结交的朋友通常会影响个人的言行举止。也就是说，社会网络的结构会影响个人的行为，这是网络动力学的一种体现。同样地，产品空间认为，一个国家生产一种产品的能力取决于它生产其他产品的能力，一个出口苹果的国家最有可能拥有适合出口梨的条件，因为这个国家已经有了合适的土壤、气候、包装设备、冷藏卡车、农艺学家、植物卫生法和工作贸易协定。所有这些都可以很容易地重新部署到运营梨的业务上。然而，如果该国转而选择生产铜线或家用电器等不同产品，这些投入将是徒劳的。基于这样的思路，文献[4]认为在当前国家具备优势的领域，其邻近领域在未来会更加容易产生优势，以此来对未来的研究产出进行了预测与评估。  常言道：“能者通吃”。城市经济发展水平在一定程度上影响着科技的多样性。文献[5]通过对科学领域分类，并对城市优势领域进行映射（如下图）后，发现城市经济水平影响着国家的科学多样性。低收入的国家，其研究领域会更加的倾向于自然科学类领域，如：环境科学，动物科学等相关领域。而高收入的国家，其研究领域更加靠近三角形的中心，代表其在各个领域中处于多样化发展的状态。  三角形  **图 2-1 城市收入水平与国家科学映射图** | | | | | |
| **三、研究目标与研究内容** | | | | | |
| 1. **学习网络科学基本方法与过程。**   通过阅读相关文献及资料，分析复杂网络的国内外研究现状，同时学习网络科学相关的基本方法与网络分析方式，明确研究所采用的思路与方法。   1. **构建融资行业网络。**   使用产品空间的构建方法构建融资行业网络，过滤掉融资企业地区、融资行业不完整的投资事件；之后，计算每个融资行业在每个城市的显性比较优势并对其进行过滤，以此来根据公式计算行业i和行业j之间的接近度。通过接近度矩阵来构建“融资行业”网络。   1. **分析利用网络预测城市未来可能的融资行业。**   分析网络的拓扑性质，包括度分布、聚类系数、平均路径长度等，了解网络基本结构；使用Louvain算法对社团进行划分，得到不同领域的社团，观察其分布特性；计算各个城市、行业与当前城市的行业水平能力的平均接近程度，对其进行排序得到未来可能的融资行业。   1. **结合城市经济水平，分析城市在社团下的位置，探究城市经济水平对于行业发展的潜在影响。**   计算各个城市在各个社团所具备的优势，通过Plotly库将其映射到三角形中，结合城市经济水平对其进行分析，探究城市行业多样性是否受城市经济水平(如中低水平)的影响。 | | | | | |
| **四、拟采用的研究思路（方法、技术路线、可行性论证等）** | | | | | |
| **4.1方法**  **数据：**选题数据来源于1990年至2020年投资事件数据共208312条，字段包括投资企业，融资企业、融资企业地区、投资事件、投资金额、轮次、行业等。其中包含涉及175个行业，344个地区。将数据的地域限制为中国的各个城市，并且移除数据中融资企业地区及融资行业不完整的投资事件。根据以上限制对数据进行过滤后得到有效数据，共计179936条数据。为了验证预测的准确性，我们根据投融资时间将数据划分为训练集及测试集两个部分，以1990~2015年的数据作为训练集，2016~2020年5年的数据作为测试集。  **显性比较优势（Revealed Comparative Advantage）.**城市c在行业i的RCA定义为：  其中表示城市在行业中的进行的投融资数目。表示城市c投融资的所有行业数目。表示所有城市在行业进行的投融资总数，表示所有投融资企业数量。  **行业相似度（Field proximity）.**行业i和行业j之间的相似度被定义为一个城市在一个行业中具有优势(RCA> 1)的同时在另一个行业中也具有优势的条件概率的最小值:  **社团划分（Identify the field clusters）.**融资行业网络是根据年份(1990年至2015年)的汇总数据得出的行业相似矩阵来构建的。使用Louvain算法来对得到的行业相似矩阵进行划分，算法流程如下图所示。社团划分  图 4-1 Louvain算法流程  **行业预测（Field prediction）**.表示一种新的潜在优势行业j与城市k当前行业水平的平均接近度的衡量标准。通过计算，并对其进行排序，得到行业预测的列表。  **分析不同城市在社团划分下的位置(Position analysis).**计算各城市在对应社团下的，代表当前城市在当前类别行业所具备的竞争能力。  通过Plotly库，将城市节点映射至三角形中，以此来探究不同城市在社团划分下的位置。  三角  图 4-2 三角形绘制示例图  **4.2技术路线：**  a) 过滤掉融资企业地区、融资行业不完整的投资事件。利用数据计算获取每个城市在每个融资行业的融资企业数量，构建“城市-融资行业”二部分网络，如下图所示（图中省份在这里应为城市，图中连边权重就是计算得到的企业数量）。  图片1  **图 4-3 数据处理结果及二分图示例**  b)计算每个融资行业在每个城市的显性比较优势（Revealed Comparative Advantage，RCA）（如下图所示，图中省份在这里应为城市）。如果融资行业i在城市c的RCA值大于等于1，则将连边保留，并认为该行业在城市c中具备显性比较优势，其余RCA小于1的连边删除。  图片2  **图 5 RCA的计算**  c)计算行业i和行业j之间的接近度，计算公式为:  图片3  **图 6 相似度计算**  d)构建“融资行业”网络，其中节点为融资行业，连边权重为融资行业之间的接近度。若两个行业之间接近度为0，则不建立连边。  e)分析网络的拓扑性质，包括度分布、聚类系数、平均路径长度等。  f)对网络进行社团划分[6]。  g)利用该网络预测城市可能未来的融资行业。  h)分析不同城市在社团划分下的位置。联系城市经济发展水平进行分析。 | | | | | |
| **五、研究工作进度安排** | | | | | |
| **第一阶段**（2023年1月13日- 2023年3月10日）：查阅国内外相关文献相关资料，明确课题研究的意义和目的，对课题研究内容进行技术可行性和需求分析。完成开题报告及开题答辩。  **第二阶段**（2023年3月11日- 2023年3月22日）：进行流程设计与整理，编码数据预处理及网络构建与分析部分。  **第三阶段**（2023年3月23日- 2023年4月14日）：编码完成社团划分与网络预测部分。  **第四阶段**（2023年4月15日- 2023年4月27日）：结合城市经济水平，分析不同城市在社团划分下的位置，进行中期检查。  **第五阶段**（2023年4月28日- 2023年5月26日）：根据任务流程以及网络分析，完成毕业论文目录以及初稿的撰写。  **第六阶段**（2023年5月27日- 2023年6月2日）：进行论文评阅，并根据撰写规范和修改意见修改论文，准备毕业答辩。  **第七阶段**（2023年6月5日- 2023年6月底）：完成毕业答辩以及毕业设计（论文）等相关材料的提交。 | | | | | |
| **六、参考文献** | | | | | |
| 1. Hidalgo C A, Klinger B, Barabási A L, et al. The product space conditions the development of nations[J]. Science, 2007, 317(5837): 482-487. 2. B. Balassa, The Review of Economics and Statistics 68, 315 (1986). 3. 汪小帆,李翔,陈关荣,2006.网络科学导论,87-105. 4. Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., & Hidalgo, C. A. (2016). The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. Scientometrics, 109(3), 1695-1709. 5. Miao L, Murray D, Jung W S, et al. The latent structure of global scientific development[J]. Nature Human Behaviour, 2022: 1-12. 6. Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre.Fast unfolding of communities in large networks[J].Statistical Mechanics: Theory and Experiment 2008 (10), P1000. | | | | | |
| **指导教师意见** | **指导教师（签名）**  **年 月 日** | | | | |
| **答辩小组意见** | **□通过 □不通过**  **答辩组成员（签名）\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  **答辩组组长（签名）\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  **年 月 日** | | | | |
| **学院审核意见** | **分管教学院领导签字（公章）\_\_\_\_\_\_\_\_\_\_\_**  **年 月 日** | | | | |