

doi:10.19920/j.cnki.jmsc.2022.04.003

# 金融股权知识大图的知识关联发现与风险分析<sup>①</sup>

洪亮<sup>1,2,3</sup>, 欧阳晓凤<sup>1,2,3</sup>

(1. 武汉大学信息资源研究中心, 武汉 430072; 2. 武汉大学信息管理学院, 武汉 430072;  
3. 武汉大学大数据研究院, 武汉 430072)

**摘要:** 金融机构的复杂股权网络是系统性金融风险的微观成因和传导渠道. 因此, 发现和分析股权网络中的关键风险特征结构对金融监管有着重要意义. 本文从知识关联理论出发, 基于大规模股权网络构建金融股权知识大图, 在此基础上提出了多层股权穿透 (MEP) 算法, 揭示以金融机构为中心的穿透式多层股权网络; 提出了关键股权路径发现 (CEP) 算法, 发现金融机构之间的关键股权路径. 本文结合案例对以上基于知识关联的股权风险结构进行了实证分析, 并对 MEP 和 CEP 算法进行了性能分析与对比. 基于真实数据的实证分析和性能分析表明, 与传统的股权结构分析方法相比, 本文所提出大数据算法能够更加高效和准确地发现和分析关键股权风险结构, 支撑系统性金融风险的穿透式监管.

**关键词:** 金融股权知识大图; 穿透式多层股权网络; 关键股权路径; 知识关联

**中图分类号:** F832   **文献标识码:** A   **文章编号:** 1007-9807(2022)04-0044-23

## 0 引言

大数据技术的广泛应用极大地促进了金融领域的发展, 改变了金融业态. 然而, 金融大数据多源异构、关联缺失等问题严重阻碍了其价值的分析与发现. 金融大数据的核心价值在于其蕴含广泛存在的知识关联. 其中, 金融股权数据的知识关联主要体现为金融机构通过股权网络建立的联系. 金融机构的股东来源众多且层层嵌套, 形成了结构复杂的金融股权网络, 使得实践中暴露出越来越多的股权治理问题. 金融股权网络已经成为系统性金融风险重要的微观成因<sup>[1,2]</sup>和传导渠道<sup>[3]</sup>. 现有的研究认识到了金融机构的股权结构对系统性金融风险的重要影响. Alexander<sup>[2]</sup>和Huang<sup>[1]</sup>研究了金融机构风险承担行为对系统性金融风险的影响. 如何穿透层层嵌套的股权网络, 识别实际控股股东为理解金融机构的风险承担行

为提供重要的信息. Iyer<sup>[3]</sup>等验证了银行间的联系会导致金融风险传染, 其中股权路径是一种重要的系统性风险传导渠道. 近期的研究表明系统性金融风险也可以通过非金融机构进行传导和扩散<sup>[4]</sup>. 因此, 本文聚焦于从包含海量金融机构和全量工商注册企业的大规模金融股权网络中发现和分析关键股权风险结构, 支撑系统性金融风险的穿透式监管.

2018年2月, 鉴于安邦保险集团存在隐瞒股权实控关系, 循环注资等严重违规违法行为, 保监会决定对其实施接管以进行风险处置. 如图1所示, 中国民生银行股份有限公司 (民生银行) 的直接股东中包含了安邦保险集团股份有限公司 (安邦保险), 持股比例为 4.49%, 为非控股股东. 然而, 穿透式多层股权网络可以发现民生银行的实际控股股东为安邦保险, 因为多个股权路径的叠加而得到的总持股比例为 29.88%. 一旦安邦保

<sup>①</sup> 收稿日期: 2019-09-08; 修订日期: 2021-01-28.

基金项目: 国家自然科学基金资助项目 (72074172); 湖北省重点研发计划资助项目 (2020BAB026).

作者简介: 洪亮 (1982—), 男, 江西抚州人, 博士, 教授, 博士生导师. Email: hong@whu.edu.cn

险发生风险,极有可能危及民生银行的经营秩序.进一步分析可得,安邦保险→安邦财险→安邦人寿→民生银行是持股比例最大的股权控制链路<sup>[5]</sup>,表明安邦保险通过多层股权形成的资金杠杆控制民生银行,该链路极易形成资本强化路径,是风险传导的重要路径.

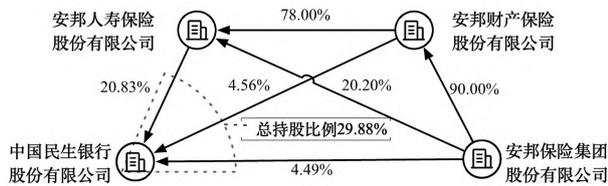


图1 民生银行案例分析

Fig.1 Minsheng bank case analysis

上述的案例表明,实际控制人和重要的股权控制链路是影响金融机构风险承担行为的关键风险特征.因此,本文发现并分析的关键股权风险结构包括穿透式多层股权网络和关键股权路径.其中,穿透式多层股权网络有利于实现穿透式监管,识别隐藏的实际控股股东.关键股权路径有利于识别股权控制链路,分析系统性风险传导路径.

大规模金融股权网络包含了众多的金融机构和非金融机构,总数可达数千万家,不仅具有复杂的股权关联,而且蕴含了丰富的语义信息.例如,一些金融机构的多层股权网络可达30层,结点数可达10万.已有的系统性金融风险研究<sup>[1-3]</sup>难以对大规模的金融股权网络进行穿透,并且缺乏从知识关联角度对金融风险结构进行发现和分析.目前的股权网络分析算法大部分基于复杂网络理论<sup>[1,6]</sup>,忽略了丰富的语义信息,难以直接应用于金融股权知识大图的查询与分析.如何利用金融知识大图中发现的多种知识关联,准确高效地发现并分析影响系统性金融风险的关键股权风险结构,支持系统性金融风险的“穿透式”监管,是本文亟需克服的重要挑战.

针对上述挑战,1)将银行、保险、期货、证券、资管、信托、公募基金共7类金融机构和全量的工商注册企业的股权数据进行融合,并基于知识关联将金融机构、非金融机构及其之间的股权关系统一组织和表示为一个金融股权知识大图;2)将金融股权网络中关键风险特征结构发现的问题转化为金融股权知识大图中的知识关联发现问题;

不同于现有的股权查询平台,提出了多层股权穿透算法,从复杂股权知识关联中发现穿透式多层股权网络,可以识别出隐藏在层层股权网络之后的实际控股股东;关键股权路径发现算法考虑持股和控股的不同优先级,从金融股权知识大图中高效地发现金融机构之间的关键持股路径;3)针对以上股权风险结构进行了特征分析和案例研究,发现了穿透式多层股权网络和关键股权路径的结构、语义和时序特征,及其在支撑系统性金融风险的穿透式监管中的基础性作用,为揭示股权结构与系统性金融风险之间的复杂关联提供新的视角与经验证据;4)基于亿级三元组的金融股权知识大图进行实验,结果证明了所提出的算法相比于现有同类算法的有效性和高效性,从知识关联视角为系统性金融风险识别和预警提供解决方案.

## 1 文献综述

### 1.1 股权结构与金融风险研究

股权结构作为股权治理的基础要素,近年来学术界对其进行了大量实证研究.股东结构决定了公司的治理机制与组织架构,从而影响到其战略、风险管理和资产负债行为.Barro等<sup>[7]</sup>实证了公司治理机制与股权结构的相关性与有效性.Alexander和Dhumale<sup>[2]</sup>研究内容聚焦于公司治理与风险的关系,提出制定基本的国际标准,以提高公司的稳健性.Wright等<sup>[8]</sup>从企业内部持股人与大股东、机构投资者的视角研究了股权结构对企业风险承担行为的影响.Barry等<sup>[9]</sup>利用欧洲商业银行的股权数据,分析了私有企业和上市银行的所有权结构和风险之间的关系,发现所有权结构在解释风险差异方面具有重要意义.Rotundo等<sup>[10]</sup>则研究了意大利股票市场股东关系的网络,并对股权网络相关特征进行分析.

为把握股权结构的重要特征,需要针对股权的集中程度进行细粒度的研究与分析.赵晶等<sup>[5]</sup>发现大股东表面上拥有着上市公司的控股权,但其只是实际控股股东控制链路中的一环,受实际控股股东的影响.贾明等<sup>[11]</sup>则研究了在公司内部治理中,控制性股东侵占投资者利益的同时使得

公司持续经营所面临的风险放大,并且这种风险还会通过市场机制传导到其他市场主体上,从而放大金融系统风险. Hooy 等<sup>[12]</sup>通过研究马来西亚公司所有权对公司绩效的影响,发现实际控股股东使用多个控制链会对公司绩效产生负面影响,并发现外来投资者最终持股比例与公司绩效呈非线性关系,这为解释最终持股股东特征与金融风险之间存在关联提供了证据.

目前已有相关文献对系统性金融风险进行了理论和实证分析. Iyer 等<sup>[3]</sup>利用印度大型银行倒闭造成的冲击以及银行间风险敞口的详细数据,检验了银行间联系导致的金融传导,并表明银行间关联是一种重要的系统性风险传导渠道. 网络分析技术的兴起使得基于金融网络进行系统性风险的研究成为可能. Haldane 等<sup>[13]</sup>从金融网络的角度探索了系统性金融风险的动态过程,从风险的形成,传导与扩散到提出政策性建议. Billio 等<sup>[14]</sup>提出了基于主成分分析和格兰杰因果网络的关联计量方法,并将其应用于美国对冲基金、银行、经济/交易商和保险的月度收益数据,研究发现银行在传递冲击方面发挥更大的作用. Elliott 等<sup>[15]</sup>利用欧洲债务交叉数据研究相互依存的金融机构网络的一系列风险级联,发现金融风险传导的可能性主要取决于整合和多元化. Elsing-er<sup>[16]</sup>和 Gouriéroux 等<sup>[17]</sup>在债务网络模型的基础上进行扩展,融入了股权数据. Huang<sup>[1]</sup>利用中国股市数据,研究了金融机构在金融网络中的局部拓扑结构对其系统性风险贡献的影响,裴茜等<sup>[18]</sup>则从基于金融资产价格变动的金融传染出发,研究我国股票市场的金融传导与渠道,并从银行和金融类的渠道发现近年来我国金融业的发展增大了我国股市的金融传染.

上述研究分析了股权结构与股权治理、股权的集中程度,以及金融风险之间的内在联系,并从金融机构复杂的网络结构出发,探讨网络结构对系统性风险的影响. 然而,目前的研究工作多以债权网络模型为基本网络,较少从股权网络视角研究系统性风险;其次,目前研究工作未将金融机构的股权网络置于完整金融体系的股权网络中进行分析,忽略了金融机构和非金融机构的股权关联;最后,股权知识大图组织和表示了股权网络中蕴

含的丰富语义信息与知识关联,可以支持金融风险结构精准发现与分析.

## 1.2 知识关联研究

知识关联是指构成知识的结点与结点之间存在的各种联系,知识本身和知识载体之间存在着各种关联,揭示和利用知识关联是知识组织、管理、发现和创造的起点<sup>[19]</sup>.

在金融领域,张俊<sup>[20]</sup>在证券信息服务分析与研究中就引入了知识关联的概念与思想,投资者通过判断证券承载的各种信息来判断其价值. 张瑞等<sup>[21]</sup>结合知识关联技术,探讨并实现金融数据可视化技术,为金融领域的关联数据分析助力. 李旭晖等<sup>[22]</sup>对金融等典型领域进行调研分析,发现大数据的价值主要源于其中蕴含的事物之间广泛存在的各种关联,对知识关联进行刻画、揭示和利用是大数据价值分析、发现和创造的核心问题. 唐旭丽等<sup>[23]</sup>分析了金融大数据领域的四种典型关联—分类、时空、统计、事件关联,为金融风险发现提供基础. 其中,分类关联用于描述金融概念间的层级结构,层次化、网络状的信息关联为金融大数据提供多维度的视角;统计关联是指使用统计方法分析金融风险事件主体之间在关联强度、关联结构和关联变化等方面的特征. Liang 等<sup>[24]</sup>则基于图网络模型将财务报告中的多领域知识进行整合,并通过知识关联推理以进一步挖掘企业关联的隐式知识,支持多视角和跨领域决策信息的智能融合.

O' Riain 等<sup>[25]</sup>利用关联数据技术与 XBRL(可扩展商业报告语言)结合起来,促进了 XBRL 与财务外部数据的关联. Li 等<sup>[26]</sup>基于关联数据建立了中国上市公司的投资开放数据,促进企业数据的商业使用. DeStefano 等<sup>[27]</sup>则使用基于资源描述框架(resource description framework, RDF)<sup>[28]</sup>和本体的关联数据提高管理企业数据资产的能力. 知识关联反映了知识从隐性关联到显性的过程,侧重于大数据价值分析、发现与创造;而关联数据则为海量异构数据中的知识关联提供了知识表示和共享的规范.

综上所述,如何基于知识关联进行金融风险发现与分析仍有待深入研究. 本文从知识关联的视角构建金融股权知识大图,对股权知识关联进

行多类型、多层次的表示,在此基础上对金融风险结构进行了准确的发现与深入的分析。

### 1.3 图路径算法研究

金融股权知识大图本质上为图结构,针对金融股权知识大图的价值分析与挖掘可以转为图路径算法研究。已有相关研究工作可以解决  $k$  简单最短路径,即无环的路径。Yen<sup>[29]</sup> 提出 Yen's 算法,求解从起点到终点的限定无环的  $k$  最短路径算法,是使用偏离路径方法解决  $k$  最短路径的最早算法。Martins 等<sup>[30]</sup> 对 Yen's 算法进行改进,提出了 MPS 算法;该算法引入边的缩小长度简化偏离路径长度的计算。Hershberger 等<sup>[31]</sup> 提出了一个偏离路径算法,将候选路径划分为多个等类价,使用替换路径算法求得每个类中的最短路径放入堆中,将堆顶最小的路径作为下一个最短路径。

$k$  一般最短路径,即存在环路的路径。Hart 等<sup>[32]</sup> 提出了一种启发式搜索的算法  $A^*$  算法,为启发式搜索奠定了基础。Eppstein<sup>[33]</sup> 提出 EA 算法,构建了一个路径图  $P(G)$  存放图中的所有非树边,并基于该路径图提出了时间和空间均较为高效的  $k$  最短路径算法,缺点在于构造图  $P(G)$  的时间成本较高。Jimenez 等<sup>[34]</sup> 针对 EA 算法的缺点,提出了 lazy 版的 EA 算法,简称为 LVEA 算法,在实际应用其运算速度要快于 EA 算法。Aljazzar 等<sup>[35]</sup> 受 EA 的启发提出了  $K^*$  算法,在图  $G$  上运用  $A^*$  算法进行向前搜索,比 EA 算法性能更好。

上述图路径算法<sup>[29-35]</sup> 没有考虑金融股权知识大图的语义与结构特征,因此不能直接应用于解决本文的问题。

## 2 研究思路

本文围绕金融股权知识大图风险结构发现和特征分析的问题,从知识关联的角度,提出层次化研究框架。如图 2 所示,基于海量多源异构的股权数据,通过融合和转换,构建金融股权知识大图;然后提出了多层股权穿透算法和关键股权路径发现算法,分别发现金融股权知识大图中的穿透式多层股权网络和关键股权路径;最后针对两类知识关联进行风险特征分析,结合风险案例支持系

统性金融风险的识别和监管。首先对金融股权知识大图进行了定义:

**定义 1** (金融股权知识大图)金融股权知识大图是一个加权有向图  $G(V, E)$ , 其中  $V$  为结点集合,  $E$  为边集合。  $\forall v_i \in V$ ,  $v_i$  为一个实体,表示金融机构、企业或者自然人,  $v_i$  具有一个属性集合  $\{a_1, a_2, \dots, a_n\}$ , 表示实体的不同属性,如企业名称、企业类型等;  $\forall e_{ij} \in E$ ,  $e_{ij}$  为实体  $v_i$  对于  $v_j$  的持股关系,  $e_{ij}$  具有属性  $\{w_{ij}, c_{ij}\}$ , 包括持股比例  $w_{ij}$  和持股类型  $c_{ij}$ 。

在知识大图构建层,原始的多源异构数据包括全国金融机构股权数据、工商注册企业股权数据和互联网数据。在工商注册企业股权数据中,金融机构股权数据是缺失与不准确的;互联网数据可以提供更多的金融语义信息。因此本文将基于共同的金融机构融合以上三类数据,建立金融机构之间的股权网络。然后,基于知识关联,通过语义转换将股权网络转换成语义丰富的金融股权知识大图。

在知识关联发现层,多层股权穿透算法穿透金融机构的层层股东,识别出该金融机构的实际控制人,并根据股权控制关系确定股东的层级,最终获取金融机构的穿透式多层股权网络。其中穿透式多层股权网络由中心金融机构及其多层级股东形成,隶属于不同类别(层级)的知识(股东信息)相互关联,属于知识关联中的分类关联;层次化、网络状的知识关联结构能够支持对于金融机构的穿透式监管。关键股权路径发现算法从金融股权知识大图中发现 Top- $k$  股权路径。关键股权路径使用统计的方法获得了关联强度、类型等方面的特征,因此属于统计知识关联。关键股权路径揭示了系统性金融风险产生和传导的结构和语义特征。

在风险分析识别层,将对以上两种风险结构特征,进行深入的量化分析。例如针对金融机构进行穿透式多层股权网络的最终股东结构类型统计分析。在复杂股权网络结构下,各类型股东穿透后的最终持股占比与直接股东的持股占比有很大的出入;特别是国有股权对金融机构的最终所有权占比,是国家金融安全、政治安全的基础判断依据。最后通过对典型金融事件的案例分析,验证算

法和风险特征分析的有效性,支持系统性金融风险的有效识别.

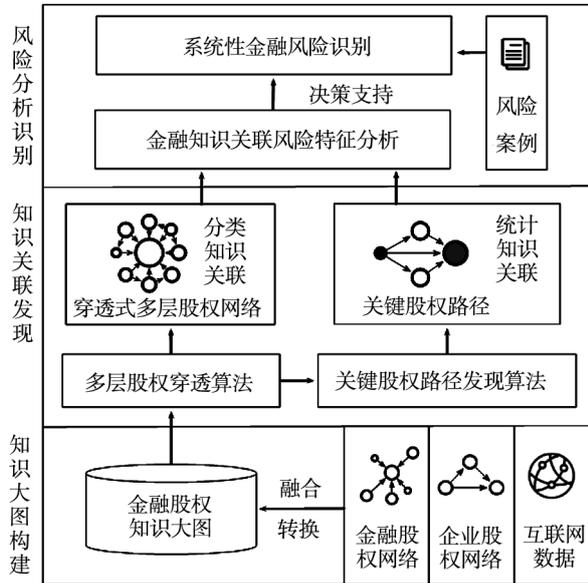


图2 研究框架  
Fig.2 Research framework

### 3 金融股权知识大图构建

#### 3.1 多源异构股权数据融合

本节将多源异构股权数据进行融合,构建大规模金融股权网络.其中,全量的工商注册企业数据库是结构化数据,时间跨度为2016年~2018年,包含了全国4200万家以上的工商注册企业的基本面信息和股东信息,其中包括了中国3867家上市公司的十大股东信息和十大流通股东信息,采用时间戳在知识大图上标注时变的股权关系,时间戳的单位为年.数据的基本情况如表1所示,结点和边的总数达到了亿级.

表1 工商注册企业股权网络结构信息

Table 1 Structural information of equity network of enterprises registered in state administration for industry and commerce

结点数	60 599 124
边数	103 330 303
容量	39.19GB
平均度	3.300
度最大值	72 173

图3是工商注册企业股权网络的度分布,结点的平均度为3.300,度的分布并不均匀.采用Clauset等提出的power-law拟合重尾分布,减少误差.可以看到图中出度和入度的拟合分布,其中出度的度概率分布具有较强的幂律特性,出度密度函数拟合方程为 $p(x) = 1.52x^{-2.520}$ ,拟合效果 $R^2 = 0.967$ ,入度密度函数拟合方程为: $p(x) = 1.785x^{-2.785}$ , $R^2 = 0.873$ ,服从幂律分布;对比图3(a)出度与图3(b)图入度概率密度的幂律拟合分布图,可以看到拟合的误差较小.该网络符合无标度网络的特征<sup>[36]</sup>,主要是由于大型金融机构有大量的金融网络连接,而小型金融机构只有较少的连接,也表明针对金融网络中度较大的节点冲击时的脆弱性.因此着重针对有代表性的金融机构进行了股权网络特征分析. Kogut

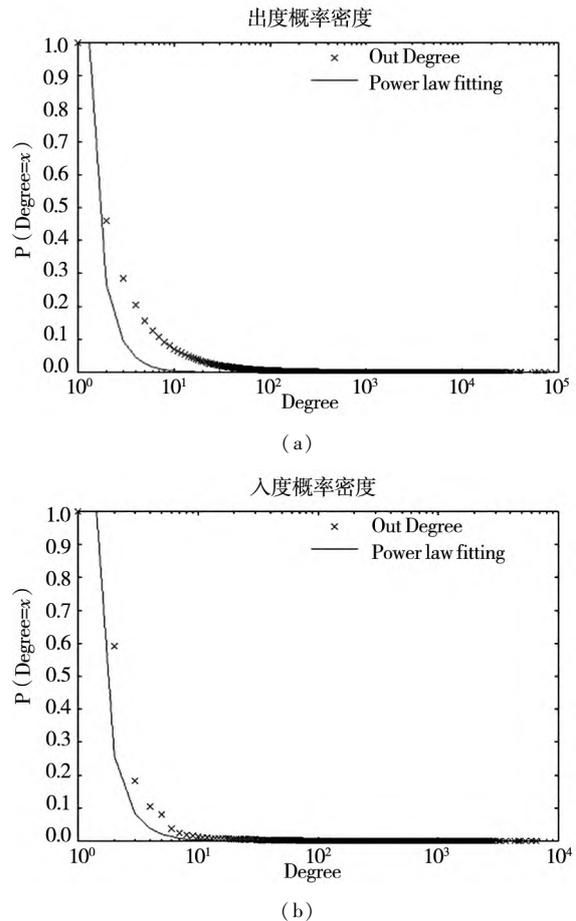


图3 工商注册企业股权网络结点出度和入度概率密度  
Fig.3 Probability density function of out-degree and in-degree of nodes in equity network of enterprises registered in state administration for industry and commerce

等<sup>[37]</sup>发现德国公司的所有权结构形成了一个“小世界”, Bertoni 等<sup>[38]</sup>则分析了1999年~2004年的意大利所有上市公司之间所有权与董事相互联系的复杂网络,结果显示该网络具有小世界的特征; Conyon 等<sup>[39]</sup>从图论的角度研究英国公司的所有权与控制权,发现企业所有权网络具有小世界的属性,且他们发现金融机构在企业所有权网络中起着潜在风险传导渠道的作用. 上述研究证明金融股权网络具有显著的小世界特征,说明小金融机构总是能通过有限的大金融机构形成连接,即金融风险发生时可以通过很短的路径传导到其他的结点. 因此,改变网络中有限的结点和边可以有效地防范和降低系统性风险.

金融股权网络的核心是金融机构的股权关联,而金融机构股权关联的核心是银行类金融机构的股权关联. 由于工商注册企业的股权数据中大量银行的股权信息缺失,同时已经有银行股权信息也不够准确. 因此,将采用互联网数据以及中国银行业数据库中的准确和最新的结构化股权数据,有效地补充缺失的银行股权数据,支持准确的实证分析. 中国银行业数据库中包括359家银行前十大股东,样本银行总资产占中资商业银行总资产的98.3%. 其中银行均为总行. 工商注册企业数据库中也包含了总行的分支机构,由于分支机构是总行的组成成分,所以不存在独立的股权结构. 为了保持工商注册信息的完备性,金融股权网络中将同时保留总行及其分支机构的结点.

银行股权数据和工商注册企业股权网络的融合方法是:判断银行B的股东A'和股权网络中的企业A是否为同一家金融实体;如果是,则将A'和A融合为一个股权网络结点A,则银行B和企业A建立起股权关联. 融合的关键问题是企业或者银行名称的共指消解,即将股权网络中同一实体的不同描述合并的过程. 比如,“中国工商银行股份有限公司”和“工行”描述的是同一家银行. 具体的技术方案是首先构建命名实体与实体指称的一对多映射关系,建立实体指称映射词典;其次对于中国银行数据库中的每个股东实体,查询实体指称映射词典,得到与该实体可能存在共指的

企业实体,生成候选共指实体对;最后采用改进的SimRank算法<sup>[40]</sup>来计算候选共指实体对的相似度. SimRank同时考虑图结构中实体的属性相似度和该实体邻居结点的属性相似度,当相似度大于阈值时,对于中国银行数据库中的某个股东实体,选择工商注册企业股权网络中与之相似度最高的实体进行链接,从而实现股权网络的融合. 当候选共指实体对的属性值或者邻居实体较为稀疏时,利用互联网上的企业基本面信息,百科数据等相关实体的属性和关联实体进行补充,通过多源异构数据的交叉验证,提高共指消解准确率. 针对股东重名的实体,仍采用SimRank计算重名实体之间的相似度,当相似度小于阈值时,则两个实体虽然同名但是并非同一个实体;反之,两个实体为同一个实体.

### 3.2 金融股权网络的转换

金融股权网络仅表示了股权结构信息,忽略了金融股权数据中蕴含的丰富的语义信息. 这些语义信息包含了金融机构的类别、属性,以及金融股权多维度、多层次的知识关联,是实现金融大数据价值分析与发现的基础. 例如,系统性金融风险识别需要考虑金融机构和企业的类型,以及金融机构间的控股或持股关系. 因此,将金融股权网络转换为金融股权知识大图. 表2是金融相关的概念与知识大图概念的对照表.

表2 金融概念与知识图谱概念对照表

Table 2 Financial concepts and knowledge graph concepts

金融概念	知识图谱概念
金融机构、企业	实体
企业类型	类别
持股、管理关系等	关系
企业的属性	属性
金融机构及企业之间的持股链路	路径
股权网络	子图

如图4(a)(b)所示,语义转换将股权网络用RDF进行规范化的表示:金融机构或者企业的名称被转换为实体的属性;股权关系被转换为知识大图的边,并添加了语义标签,包括了边的类别(名字或股东等)和属性(持股比例,控股或持股关系等). 其中控股和持股关系需要通过持股比

例,结合股权网络的结构进行计算得到<sup>[41]</sup>.

对于不同的边或路径,控股关系并不意味着持股比例一定小于持股关系,因为控股关系的计算还需要考虑其他股东的数量和持股比例. 股权路径的持股比例等于该路径上所有边的权重(直接持股比例)的乘积,表示路径起点的股东对终点股东的间接所有权份额<sup>[42]</sup>. 如图 4(b)所示,控股关系路径  $(v_1, v_3, v_0)$  的持股比例也就是股东  $v_1$  对  $v_0$  的间接所有权份额为:  $p_{10}^0 = w_{13} \times w_{30} = 20.2\% \times 20.83\% = 4.207\%$ , 而持股关系路径  $(v_1, v_0)$  的直接持股比例即股东  $v_1$  对  $v_0$  的直接所有权份额  $p_{10}^1 = 4.49\% > p_{10}^0$ .

在金融意义上,控股关系的控股权要强于持股关系,不论其实际的持股比例是多少. 因此,在转换时要区分这两个关系的优先级;引入  $c_{mn}$  表示股权关联边  $e_{mn}$  类型,当  $e_{mn}$  为控股边时,表示  $v_m$  对  $v_n$  的直接控股权为 100%,  $c_{mn} = 1$ , 否则  $c_{mn} = 0$ .

金融知识大图用 RDF 三元组的形式存储到图数据库管理系统 gStore<sup>[43]</sup>中,以便对数据进行高效管理.

## 4 知识关联发现算法

### 4.1 多层股权穿透算法

#### 4.1.1 问题定义和总体思路

在金融股权知识大图中,某个金融机构或企业的直接股东为第一层股东,而每一个第一层股东又有第二层股东,以此类推,直到最终股东(定义 5). 以上所有的股东及其之间的股权关联组成了多层股权网络. 多层股权网络在穿透式监管中具有重要意义,有助于揭示层层股东对某个金融机构或企业的控股权,并找到最终持股比例和实际控制人.

**定义 2 (控股权)** 结点  $v_i$  对  $v_c$  的控股权  $u_{ic} = \max (\bigwedge_{e_{mn} \in E(p_{ic}), p_{ic} \in P_{ic}} c_{mn})$ . 其中  $P_{ic}$  表示  $v_i$  到  $v_c$  所有路径的集合;对于  $P_{ic}$  中的第  $j$  条路径  $p_{ic}^j (j \subseteq [0, M])$ ,  $M$  表示路径集合的数量;将  $p_{ic}^j$  路径中所有边的直接控股权进行逻辑与,计算该路

径的控股权.  $u_{ic}$  等于最大控股权路径  $p_{ic}$  的控股权.

**定义 3 (最终持股比例)** 最终持股比例  $\delta_i$  为  $v_i$  对  $v_c$  直接持股比例和间接持股比例的加权总和,即  $\delta_i = \sum_{p_{ic} \in P_{ic}} \delta(p_{ic})$ , 等于  $v_i$  到  $v_c$  所有路径(简单路径或非简单路径)的持股比例之和.

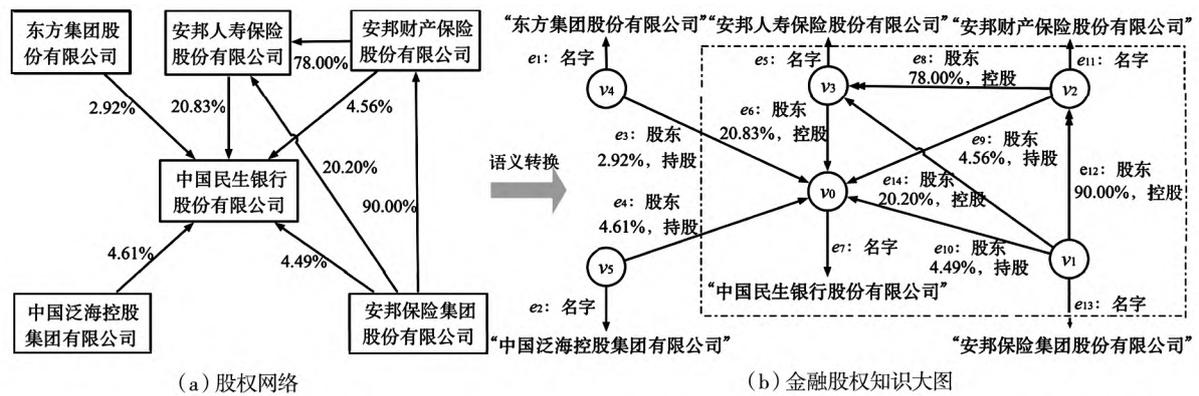
**定义 4 (穿透式多层股权网络)** 给定金融股权知识大图  $G(V, E)$ , 以及中心结点  $v_c$ , 穿透式多层股权网络  $S(v_c, V_s, E_s)$  是  $G$  的导出子图(Induced Graph).  $\forall v_i \in V_s, v_i$  具有层数  $l(v_i)$ 、控股权  $u_i$  和最终持股比例  $\delta_i$  属性. 在  $v_i$  到  $v_c$  控股权  $u_{ic}$  最大的路径中,如果路径  $p_{ic}$  持股比例  $\delta(p_{ic})$  最大,则  $v_i$  相对于  $v_c$  的层数  $l(v_i) = l(p_{ic})$ , 其中  $l(p_{ic})$  为路径  $p_{ic}$  的跳数.  $S$  中包含内部结点、最终股东(定义 5)和一个实际控制人  $v_{control}$  (定义 6).

**定义 5 (最终股东)** 在穿透式多层股权网络  $S(v_c, V_s, E_s)$  中,  $\forall v_i \in V_s, v_i \neq v_c$ , 存在  $v_i$  到中心结点  $v_c$  的路径  $p_{ic}$ , 如果不存在路径  $p_{jc}$ , 使得  $p_{ic}$  是  $p_{jc}$  的真子路径, 则  $v_i$  是  $S$  的最终股东, 否则  $v_i$  是  $S$  的内部结点. 最终股东类型包括“国有”、“私有”、“外资”、“广泛持股”, 因为这些类型的股东没有股东.

国有股东指的代表国家持有国有股权的政府机构或法人;私有是指单一自然人所持有的股权;外资是指港澳台及外国实质性法人、自然人实体所持有股权,不包括内资法人在境外设立并控制的法人实体所持股. 广泛持股是指中小投资者的持股.

**定义 6 (实际控制人)** 在穿透式多层股权网络  $S$  中,对于所有其他最终股东  $v_j$ , 如果最终股东  $v_i$  对中心结点  $v_c$  的控股权  $u_{ic} \geq u_{jc}$ , 且最终持股比例  $\delta_i \geq \delta_j$ , 则  $v_i$  为实际控制人  $v_{control}$ .

提出多层股权穿透(MEP)算法,给定中心结点  $v_c$ , 找出穿透式多层股权网络. MEP 算法需要解决的主要问题是确定穿透式多层股权网络中每个结点的层数,并计算其对中心结点的最终控股权和最终持股比例,确定实际控制人.



语义转换：  
 股权网络转换为金融股权知识大图

- ⊙ 实际控制人    0 持股
- 起点            1 控股
- 终点             $l$  层数
- 完全控制路径     $h$  代价函数

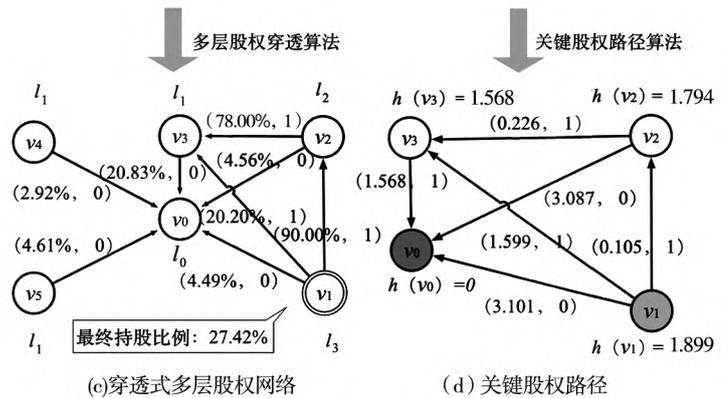


图 4 金融股权知识大图转换示例

Fig. 4 Conversion example of financial equity knowledge graph

#### 4.1.2 基于 BFS 的多层股权穿透

根据定义 4 可知,需要以  $v_c$  为中心结点,基于广度优先遍历(breadth-first search, BFS)往外扩展访问图中的邻居结点,然后找到每一层的股东  $v_i$  到  $v_c$  的控股权和持股比例最大的路径,从而确定股东结点的层数  $l(v_i)$ . 两点之间的持股比例最大的路径实际上是第 1 条关键股权路径,是 Top- $k$  关键股权路径的特例. 因此,调用关键股权路径发现算法(CEP 算法,详见 4.2 节)找到持股比例最大的路径. 以上过程以逐层迭代的方式进行,直至所有的邻居结点均为最终股东. 然后计算每个股东对于中心结点最终持股比例,在穿透式多层股权网络中,控股权最大条件下找到最终持股比例最大的结点,该结点为  $v_c$  的实际控股结点  $v_{control}$ . 注意到,CEP 算法仅需 1 次遍历,可以确定层级并计算最终持股比例.

Brioschi<sup>[44]</sup> 提出的 Integrated Ownership Share (IOS) 模型,通过稀疏矩阵运算,将持股比例向内

反推,确定任意一个股东对一个金融机构或企业的最终持股比例. IOS 模型可以在存在环路的情况下计算各股东结点对金融机构持股的最终比例;该算法需要对于股权网络中所有结点的两两关系进行全局的运算. 多层股权穿透仅找到与中心结点存在直接或间接股权关联的结点,这些结点数量远少于金融知识大图的全量亿级结点. 因此,只将穿透式多层网络的导出子图作为 IOS 模型的输入,并且在矩阵运算中,仅计算网络中其他结点到中心结点的最终持股比例. 这样可以较大地提高多层股权穿透的效率.

$V_s$  中的任意两点的股权关联表示为直接持股比例矩阵  $A$ ,其中  $v_i$  对  $v_j$  的直接持股比例表示为  $w_{ij}$ ,为  $A$  中第  $i$  行  $j$  列元素,  $\forall v_i \in V_s, w_{ii} = 0$ ,即公司对自身的直接持股比例为 0. 给定输入矩阵  $A$ ,IOS 模型通过直接持股比例的层层递推找到全部间接持股比例,即最终持股比例矩阵  $U = A \times (I - A)^{-1}$ ,其中  $I$  为  $|V_s| \times |V_s|$  的单位矩阵.

仅需要计算到  $V_s$  中所有结点到  $v_c$  的最终持股比例,可得

$$U_{*,c} = A \times (I - A)^{-1}_{*,c} \tag{1}$$

其中  $U_{*,c}$  表示  $U$  的第  $c$  列向量.

MEP 算法具体步骤如下:从查询中心点  $v_c$  开始,访问  $v_c$  在金融股权知识大图中的邻居结点,不断迭代,直到所有邻居结点均为最终股东,则迭代结束.在 BFS 过程中,每次访问一个新的结点  $v_i$ ,调用 CEP 算法计算  $v_i$  到中心  $v_c$  的第 1 条关键股权路径,并确定  $v_i$  的层数. BFS 结束后,使用公式(1)计算穿透式多层股权网络中所有结点到中心结点最终持股比例,根据定义 6 确定实际控股股东  $v_{control}$ .

```

算法 1 多层股权穿透算法 (MEP)
输入:金融股权知识大图  $G(V,E)$ ,中心结点  $v_c$ 
输出:穿透式多层股权网络  $S(v_c, V_s, E_s), v_{control}$ 
queueQ; //结点队列
 $V_s \leftarrow \{v_c\}$   $E_s \leftarrow \emptyset$ 
visited ( $v_c$ )  $\leftarrow 1$  //标识为已访问
Q.enqueue ( $v_c$ ) //中心结点入队列
While ! Q.empty ( ) do
     $v_i \leftarrow Q.dequeue$  ( )
    CEP ( $v_c, v_i, 1$ ) //找出 Top-1 关键股权路径
     $l(v_i) = l(p_{ic})$  //确定层级
    Foreach  $v_j$  in  $G.adjV(v_i)$ ,  $e_{ij}$  in  $G.adjE(v_i)$  do
        If ! visited ( $v_j$ ) //结点未被访问
            visited ( $v_j$ )  $\leftarrow 1$ 
             $V_s \leftarrow V_s \cup \{v_j\}$ 
             $E_s \leftarrow E_s \cup \{e_{ij}\}$ 
            Q.enqueue ( $v_j$ )
IOS ( $S(v_c, V_s, E_s)$ ) //计算到  $v_c$  的最终持股比例
If  $u_{ic}$  is maximal of all nodes and  $i = \text{argmax}(\delta_i)$ 
    //控股权和最终持股比例最大
     $v_{control} = v_i$ ;
Return  $S(v_c, V_s, E_s), v_{control}$ ;

```

如图 4(c) 所示,  $v_0$  为中心结点,通过 BFS,层层向外遍历,得到  $v_1, v_2, v_3, v_4, v_5$  组成了  $v_0$  的穿透式多层股权网络.注意到,  $v_1$  到  $v_0$  有 4 条路径,其中  $v_1 \rightarrow v_0$  的持股比例为 4.49%,控股权为 0;  $v_1 \rightarrow v_2 \rightarrow v_0$  的持股比例为  $4.56\% \times 90\% = 4.1\%$ ,控股权为 0,  $v_1 \rightarrow v_3 \rightarrow v_0$  的持股比例为  $20.2\% \times 20.83\% = 4.2\%$ ,控股权为 1,  $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_0$  的持股比例为  $20.83\% \times 78\% \times 90\% = 14.62\%$ ,控股权为 1.因此,控股权最大的

路径为  $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_0$ ,  $v_1$  为  $v_0$  的第 3 层股东,即层数为 3.将上述 4 条路径的持股比例相加,可得  $v_1$  对  $v_0$  的最终持股比例为 27.42%,为  $v_0$  的穿透式多层股权网络中所有结点最大的,因此实际控股股东  $v_{control}$  为  $v_1$ .

### 4.2 关键股权路径发现算法

#### 4.2.1 问题定义和总体思路

关键股权路径主要是指知识大图结点之间的持股链路.用户进行多层股权网络穿透后,可以更进一步查询两个结点之间的持股路径,确定两个企业之间的持股关系的强弱,甚至进一步判断控股和关联程度,计算路径的持股比例和控股权,发现关键股权路径.

**定义 7(关键股权路径)** 给定金融股权知识大图  $G(V,E)$ 、路径起点  $v_s$  (金融机构、企业或者自然人)和终点  $v_e$  (金融机构或者企业)、以及  $k$ ,  $k \geq 1$  且为整数,  $v_s$  到  $v_e$  的关键股权路径为两点之间 Top- $k$  条持股比例最大的路径,即按照持股比例递减排列的前  $k$  条路径.

在金融股权知识大图中发现关键股权路径并不是一个容易解决的问题.首先,由于知识大图规模巨大,穷举两点之间的所有路径的代价较大.其次,金融股权知识大图中路径权重具有以下特殊性:

股权路径  $p_{ij}$  的持股比例  $\delta(p_{ij})$ ,即  $v_i$  通过路径  $p_{ij}$  对  $v_j$  的持股比例,等于  $p_{ij}$  上每条边的持股比例的乘积

$$\delta(p_{ij}) = \prod_{e_{mn} \in E(p_{ij})} w_{mn} \tag{2}$$

其中  $e_{mn}$  为股权路径  $p_{ij}$  的边集  $E(p_{ij})$  中的边,  $w_{mn}$  为  $e_{mn}$  的权重,即持股比例.

如图 4(b) 所示,路径  $(v_2, v_3, v_0)$  的持股比例  $p_{20}$  为有向边  $(v_2, v_3)$  和  $(v_3, v_0)$  的持股比例  $w_{23}$  和  $w_{30}$  相乘,即  $p_{20} = w_{23} \times w_{30} = 78\% \times 20.83\% = 16.25\%$ .

现有的路径发现算法<sup>[30,31,35]</sup> 是基于权重相加的网络设计的,无法直接应用于发现金融股权知识大图中的关键股权路径.

最后,从关键股权路径的金融意义来看,根据定义 2,如果路径上所有的边均为控股关系,则该条路径为控股路径;反之,如果路径上至少有一条

边为持股关系,则该条路径为持股关系.对于同样的起点终点,控股路径的控股权要高于持股路径.因此,关键股权路径首先需要根据边的语义标签找到控股路径,在路径均为控股或者持股时再比较路径的持股比例,对股权路径按持股比例从大到小排序.

#### 4.2.2 基于股权优先级的 Top - k 最短路径发现

首先对知识大图的边权重进行转换.考虑到  $0 \leq p_{ij} \leq 1$ ,进行如式(3)的对数转换,就将式(2)中的权重乘积  $\delta(p_{ij})$  转换为式(4)中的权重之和  $\delta'(p_{ij})$ .

$$w'_{mn} = -\ln w_{mn}, e_{mn} \in E(p_{ij}) \quad (3)$$

$$\begin{aligned} \delta'(p_{ij}) &= \sum w'_{mn} = -\ln(\prod w_{mn}) \\ &= -\ln(\delta(p_{ij})) \end{aligned} \quad (4)$$

通过以上边权重的转换,关键股权路径发现问题转化为从带权有向图中查询两点之间的 Top - k 最短路径.同时,还要考虑股权路径的控股权优先级.两点间的最短路径问题等价于两点之间的 TSP(traveling salesman problem)问题,因此关键股权路径发现问题是一个 NP 难问题<sup>[45]</sup>.

提出一个基于优先级的启发式关键股权路径发现(CEP)算法. CEP 算法区分控股和持股关系的不同优先级,表达控股和持股的语义差异.现有的路径发现算法,如 A\* 算法等,在解决关键股权路径发现问题时,由于没有考虑股权关系的优先级,也忽略了金融股权知识大图中路径权重为各边权重乘积的特点,将无法输出准确的 Top - k 关键股权路径. CEP 算法维护了一个开始于起点的路径树,不断扩展树中的路径,一直到达终点  $v_e$ . 在每一次的迭代中, CEP 需要决定从哪条路径进行扩展. CEP 首先检查当前路径是否为控股路径,优先选择控股路径进行扩展. 如果控股路径数量小于  $k$ ,则选择持股路径. 在路径的控股权相同的情况下, CEP 算法基于当前路径的权重以及当前结点扩展到终点的估计权重进行路径选择.

具体地, CEP 算法中使用启发式评估函数  $h(v_j)$  表示当前访问结点  $v_j$  到终点  $v_e$  的路径权重估计值,  $g(v_j)$  表示从起点  $v_s$  到路径中当前访问结点  $v_j$  的实际路径权重. 因此从起点到终点经过结点  $v_j$  的路径长度  $f(v_j)$  可以表示为

$$f(v_j) = g(v_j) + h(v_j) \quad (5)$$

CEP 算法在优先级相同的情况下,逐个选择能够最小化公式(5)的路径.

CEP 算法的具体步骤如下:首先调用 Dijkstra 算法计算每个结点  $v_j$  到终点  $v_e$  的路径权重估计值  $h(v_j)$ ,具体方法为把  $v_e$  作为 Dijkstra 算法的起点,反向地计算路径的权重. CEP 算法设置两个队列  $open_p$ 、 $closed_p$ ; 其中  $open_p$  为一个优先级队列,保存待确定的结点.  $open_p = (v_j, g(v_j), f(v_j), c_{ij})$ , 其中  $c_{ij}$  为前一个访问结点  $v_i$  到结点  $v_j$  的持股类型. 路径拓展过程中,在当前路径为控股路径时,需要优先考虑  $c_{ij}$ , 在  $c_{ij}$  相同时再考虑  $f(v_j)$ , 最后在  $f(v_j)$  相同时才考虑  $g(v_j)$ , 即优先级  $c_{ij} > f(v_j) > g(v_j)$ .  $closed_p$  为一个普通队列,存储已确定的结点. CEP 算法从起点往终点逐步扩展,每次添加一个邻居结点  $v_j$ ,更新  $g(v_j)$  和  $f(v_j)$ , 以及  $c_{ij}$ . 注意到,如果当前路径为持股路径时,则所有的后续结点  $c_{ij}$  均赋值 0, 因为此时考虑新加入结点的优先级并无意义,整条路径仍然为持股路径. 当前确定的结点为终点时,则找到一条路径,从  $closed_p$  队列输出,如果输出的队列数量达到  $k$  条,则算法执行完毕.

#### 算法2 关键股权路径发现算法(CEP)

输入:金融股权知识大图  $G(V, E)$ ,  $v_s, v_e, k$

输出:Top - k 关键股权路径  $p_{se}^1, p_{se}^2, \dots, p_{se}^k$

count  $\leftarrow$  0

Dijkstra( $G(V, E), v_e, v_s$ ) // 终点  $v_e$  为 Dijkstra 算法的起点  
 $open_p \leftarrow (v_s, 0, f(v_s), 1)$  // 待确定的结点优先级队列,初始持股类型  $c = 1$ ;  $g(v_s) = 0$

**While** ! $open_p.empty()$  **do**

$(v_i, g, f, c) \leftarrow open_p.dequeue()$  // 按优先级顺序出队列

$closed_p \leftarrow v_i$  // 已确定的结点队列

**If**  $v_i = v_e$

count ++

**While** ! $closed_p.empty()$  **do**

$p_{se}^{count} \leftarrow closed_p.dequeue()$

**Return**  $p_{se}^{count}$

**If** count =  $k$  **break**

**Continue** // 开始新一轮循环找下一条路径

**Foreach**  $v_j$  in  $G.adjV(v_i)$  and ! visited( $v_j$ ) **do**

$g(v_j) = g(v_i) + w'_{ij}$

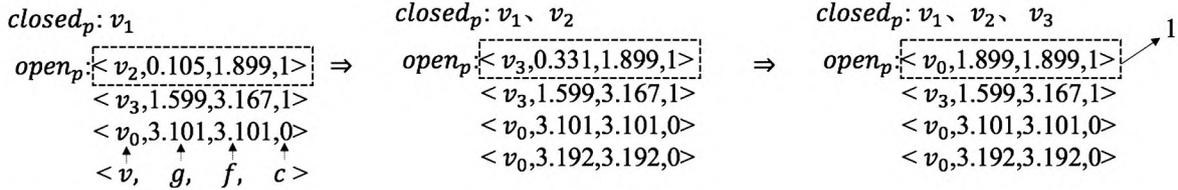
$f(v_j) = g(v_j) + h(v_j)$

**If** ( $u_{si} = 1$  and  $c_{ij} = 1$ ) // 如果前面不为 1 则优先级都为 0

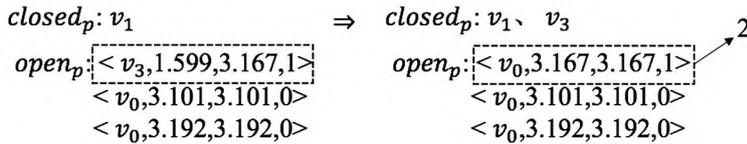
$open_p \leftarrow (v_j, g(v_j), f(v_j), 1)$

**Else**  $open_p \leftarrow (v_j, g(v_j), f(v_j), 0)$

第1条:  $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_0$



第2条:  $v_1 \rightarrow v_3 \rightarrow v_0$



第3条:  $v_1 \rightarrow v_0$

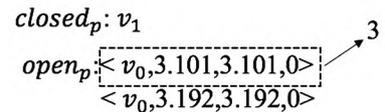


图5 Top-3 关键股权路径发现过程示例  
Fig. 5 Example of Top-3 critical equity path

如图4(d)图,其中  $v_1$  为起点,  $v_0$  为终点,首先进行权重转换,例如有向边  $(v_2, v_3)$  持股比例  $w_{23}$  为(78%, 控股)经过权重转换为(0.226, 1). 然后,求解启发式函数中的结点到终点的代价函数  $h(v)$ . 例如  $v_3$  到终点  $v_0$  的估计长度为 1.568, 则  $h(v_3) = 1.568$ . 如图5所示与起点  $v_1$  直接相连接的点有  $v_2, v_3, v_0$ , 放入  $open_p$  队列中, 优先级按照  $c_{ij} > f(v_j) > g(v_j)$ , 则  $\langle v_2, 0.105, 1.899, 1 \rangle$  位于队列顶部, 出  $open_p$  队列,  $v_2$  进  $closed_p$  队列; 将与  $v_2$  直接相连接的点及其  $g, f, c$  放入  $open_p$  队列, 可以看到有相同的结点  $v_3$ , 但是这两个结点经过的父结点不一样, 因此  $g, f, c$  也不一样. 弹出  $\langle v_3, 0.331, 1.899, 1 \rangle$ ,  $v_3$  加入  $closed_p$  队列; 以上出入队列的过程进行迭代, 直到弹出的点为  $v_0$ , 则得到第1条路径为  $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_0$ , 路径最短为 1.899, 股权比例最大, 为控股路径. 接下来求第2条路径, 目前的  $open_p$  序列里弹出  $\langle v_3, 1.599, 3.167, 1 \rangle$ , 与  $\langle v_3, 1.599, 3.167, 1 \rangle$  直接相连的  $v_0$  加入  $open_p$  队列, 根据优先级弹出  $\langle v_0, 3.167, 3.167, 1 \rangle$ , 得到第2条路径为  $v_1 \rightarrow v_3 \rightarrow v_0$ , 路径权重之和为 3.167, 也是控股路径. 依照上述的方法求的第3条路径为  $v_1 \rightarrow v_0$ , 路径长度为 3.101, 尽管比第2条更短, 但是为持股路径, 而  $v_1 \rightarrow v_3 \rightarrow v_0$  为完全控股路径, 因此第3条路径优先级低于第2条路径. 如果不考虑持股控股优先级的话, 仅考虑  $f(v_j) > g(v_j)$ , 计算的第2条路径会是  $v_1 \rightarrow v_0$ , 而不是 CEP 方法得到

的  $v_1 \rightarrow v_3 \rightarrow v_0$ , 因为  $v_1 \rightarrow v_0$  的路径上的持股比例要大于  $v_1 \rightarrow v_3 \rightarrow v_0$ , 出现结果偏差情况. 综上, 得到从起点  $v_1$  到终点  $v_0$  的 Top-3 路径.

## 5 实证分析

### 5.1 风险特征分析

基于2018年的金融股权知识大图, 对以下风险特征进行分析: 1) 关键股权特征的统计分布与概率近似; 2) 基于股权结构特征的风险特征识别.

#### 5.1.1 穿透式多层股权网络

第一, 借助 MEP 算法从金融股权知识大图中抽取金融机构的穿透式多层股权网络图. 选取了金融机构中最具有代表性的7种类型: 银行、保险、期货、证券、资管、信托、公募基金, 其中资管研究对象仅为国有资产管理公司, 由于其资产规模和利润总额占据绝对优势, 且拥有主流金融牌照, 具有较强的代表性. 私募基金存在信息透明度低, 信息披露有限等问题, 将进一步收集和更新私募基金的准确股权数据. 表3表示的是整体金融机构样本中各类型的股权结构的统计, 可以看到银行的平均层数为13层, 平均股东达4134家, 平均最终股东占比66.20%, 而证券等金融机构的股东数与层数要高于其他几类金融机构, 原因在于大型证券公司股权结构较为复杂.

表3 金融机构股权结构特征分析

Table 3 Analysis of equity structure about financial institutions

类型	数量	平均总层数	平均股东总数	平均最终股东数	平均最终股东占比
银行	359	13	4 134	2 798	66.20%
保险	117	15	5 359	3 605	60.46%
期货	12	18	8 058	5 412	62.08%
证券	35	28	12 348	8 285	66.81%
资管	2	2	8	6	70.91%
信托	67	12	3 938	2 642	52.46%
公募基金	116	12.8	1 348	849	61.80%

表4 银行类金融机构股权结构特征分析

Table 4 Analysis of equity structure of banking financial institutions

类型	数量	平均总层数	平均股东总数	平均最终股东数	平均最终股东占比
国有银行	7	30	13 190	8 849	65.44%
股份制银行	11	28	12 603	8 454	64.85%
城商行	128	15	5 137	3 466	65.14%
农商行	172	11	3 013	2 058	69.77%
外资银行	37	4	1 666	1 117	52.97%
民营银行	4	12	3 940	2 656	74.27%

表5 银行类金融机构的最终持股结构

Table 5 The final shareholding structure of a financial institution's bank

	国有/%	私有/%	外资/%	广泛持股/%
国有银行	72.22	0.04	1.86	25.88
股份制银行	32.30	8.34	11.92	47.44
城商行	27.03	15.97	5.74	51.26
农商行	17.86	20.68	1.27	60.19
民营银行	1.14	61.63	10.06	27.17
外资银行	0.05	0.02	99.89	0.04
总计	53.29	5.29	5.76	35.66

表4表示的是银行类金融机构的股权结构统计,可以看到国有银行与股份制银行平均总层数分别为30、28层,股东总数超过1万,大型国有商业银行总资产占银行业金融机构比例高达35.4%;股份制银行相对于国有银行股权较为分散,股东数量较多,其规模和资产占银行业总资产和负债的层数的特征相符合。

进一步分析银行类金融机构平均总层数与股东总数、最终股东占比的关系,如图6(a)所示,总层数与股东总数具有明显的分层结构,且中间出现层数间隙,在0层~20层内,股东总数在0~4 000内浮动,而在30层~40层,股东总数则跃升

到了15 000左右。结合表4可以分析的得出,城商行、农商行、外资银行、民营银行的股东层数在0层~20层之间,而国有银行和股份制银行股东层数较多,股权结构更为复杂。图6(b)展示了最终股东占比的分布规律,同样在0层~20层时,最终股东占比均高于20%,大部分集中于40%~85%区间,而到了30层~40层之后最终股东占比稳定于66%上下,可见银行的股东总数与最终股东占比与总层数有一定的相关性,同样结合表4可知国有银行和股份制银行的最终股东占比很稳定,与这两类银行的总资产占比比例紧密相关。

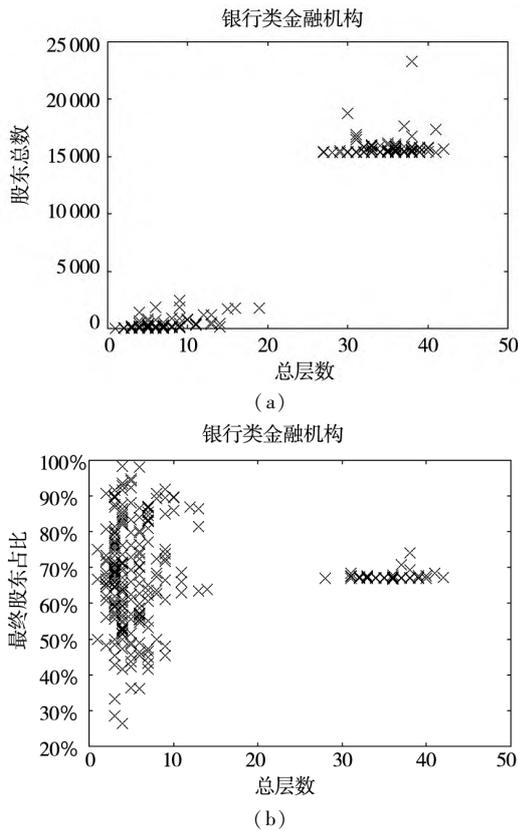


图 6 银行类金融机构股权结构特征分析图  
 Fig. 6 Analysis of equity structure of banking institution

第二,在系统性金融风险的识别分析中,另一

个重要的分析指标是分类型的最终持股比例.分析银行的最终股东,包括国有、私有、外资、广泛持股 4 个类型.使用 MEP 算法找到 6 大类型银行的穿透式多层股权网络,并统计银行最终股东的类型和占比.如表 5 所示,银行的国有最终持股资产加权比例高达 53.29%,私有为 5.29%,外资为 5.76%,广泛持股为 36.66%.李涛等<sup>[46]</sup>发现目前中国企业的制度环境中,国有比例越高,就越能发挥其监督作用,防止内部人控制,因此对业绩也有正向的促进作用.徐二明等<sup>[47]</sup>发现在企业中,国权股权占比越大,国有股在企业战略创新中的作用就越大,因此,过程创新和绩效创新获得绩效就好.且国有股权比例越高,企业能享受政府提供的关键资源的优惠待遇,对企业和金融机构都有正向的促进作用.上述理论与实证研究证实:国有股占比的提高会提升企业和金融机构的风险承担水平;从公司内部内控水平来看,银行国有最终持股占比越大,其风险承担水平越高.基于以上理论,可以计算每一类型银行的国有最终持股比例的平均值,判断某银行国有最终持股比例是否低于其所属类型的平均值,给系统性金融风险监管提供参考.

表 6 银行类金融机构 Top - k 关键股权路径统计表

Table 6 Statistical table of Top - k critical equity paths of banks

起点类型	终点类型	路径数量	第 1 条		第 2 条		第 3 条		第 4 条	
			长度	比例/%	长度	比例/%	长度	比例/%	长度	比例/%
城商	城商	21	3	8.523 8	9	0.000 0	0	0.000 0	0	0.000 0
城商	农商	6	5	3.413 0	4	0.009 3	9	0.002 1	9	0.000 4
股份制	城商	16	9	0.626 6	11	0.000 0	11	0.000 0	11	0.000 0
股份制	农商	18	9	0.333 4	10	0.000 0	10	0.000 0	10	0.000 0
股份制	外资	6	8	0.000 0	10	0.000 0	10	0.000 0	11	0.000 0
国有	城商	181	8	0.294 0	9	0.023 1	9	0.004 7	10	0.000 0
国有	股份制	16	5	0.162 4	7	0.000 1	7	0.000 0	11	0.000 0
国有	国有	12	4	0.006 9	6	0.000 0	6	0.000 0	6	0.000 0
国有	民营	8	7	0.026 9	8	0.000 2	8	0.000 0	6	0.000 0
国有	农商	115	8	0.254 8	9	0.000 0	10	0.000 0	10	0.000 0
国有	外资	16	8	0.598 2	8	0.000 0	8	0.000 0	9	0.000 0
农商	城商	10	5	2.173 2	8	0.000 0	8	0.000 0	9	0.000 0
农商	农商	65	2	5.914 8	5	0.085 1	6	0.000 6	7	0.000 0
外资	城商	51	9	0.443 8	10	0.000 0	11	0.000 0	12	0.000 0
外资	股份制	4	5	0.000 2	0	0.000 0	0	0.000 0	0	0.000 0
外资	国有	6	6	6.212 5	0	0.000 0	0	0.000 0	0	0.000 0
外资	民营	4	8	0.000 1	0	0.000 0	0	0.000 0	0	0.000 0
外资	农商	22	6	0.204 6	10	0.000 0	12	0.000 0	10	0.000 0
外资	外资	4	10	0.000 0	10	0.000 0	11	0.000 0	11	0.000 0
合计		581	6.76	1.344 4	8.73	0.012 3	9.17	0.001 7	9.65	0.000 0

### 5.1.2 关键股权路径

使用 CEP 算法查询银行类金融机构之间的关键股权路径,并进行相关的特征分析.路径的起点和终点分别枚举所有银行类机构,共有  $359 \times 359 = 128\ 881$  种组合方式,其中仅有 581 种起点和终点的组合之间存在连通路程.连通路程的数目远小于可能存在连通路程的数量,原因在于股权关系较为稀疏,使得两个银行之间存在连通路程的概率较小.对于任意一个存在连通的起点和终点组合,即路径长度大于 0,依次设定  $k = 1, 2, 3, \dots$ ,进行 Top -  $k$  查询.注意到第  $k$  条路径的持股比例随着  $k$  增加不断较少,对于持股比例接近于 0 的路径,在金融上缺少实际意义.因此,仅保留持股比例的小数点后 4 位.

第一,对起点和终点的银行类金融机构所属的不同类型进行统计分析.如表 6 所示,Top - 4 关键股权路径中的第 4 条路径的持股比例大部分为 0.因此,仅列出 Top - 4 路径.1)从纵向来看,起点银行与终点银行类型组合一共有 19 种,其中有银行与城商行,国有银行与农商行之间的路径数量最多,说明它们之间的股权关系较为稠密.同时,发现银行起点和终点类型为国有银行和股份制银行的股权路径数量比例最大,与穿透式多层股权网络的分析一致.国有银行和股份制银行规模和资产占比占据银行业的 53%,拥有数量庞大的股东,在银行与银行的股权关联网络中,起着重要的连接作用.其中,起点为国有银行,终点为城商行和农商行的路径数量最大.城商行与农商行统称为区域性商业银行.关键股权路径数量说明区域性银行受到国有银行与股份制银行等大型银行的股权控制,我国银行业龙头集中度进一步提升.在第 1 条路径中,城商行 - 城商行的平均路径权重达到 8.523 8%,平均长度为 3,相比其他类型的第 1 条路径持股比例较高,路径长度更短.这些关键股权路径包含大量的资产规模千亿级以上的大型城商行,如北京银行、南京银行、上海银行等.这与城商行依赖高杠杆盈利的结论<sup>[48]</sup>一

致.关键股权路径持股比例的计算实质是终点对起点的最终收益权份额,随着路径长度的累加,杠杆也会逐级累加.起点银行(股东)如果资金实力不够、大幅依赖杠杆融资,则最终收益权越大,对终点银行的稳健经营与资金充足率的冲击也就越大.因此,关键股权路径也反映了系统性金融风险的微观成因.

2)从横向来看,总体上从第 1 条到第 4 条的平均长度是递增的趋势(6.76 - 9.65),因为长度增加使得路径的持股比例不断衰减,从而在 Top -  $k$  路径中的持股比例排名下降.一个例外是城商 - 农商股权路径,第 1 条的平均长度为 5,而第 2 条路径的平均长度为 4,原因在于在发现关键股权路径时,不仅考虑了路径的持股比例,也考虑了路径的控股或持股的优先级.一般而言,长度较长的股权路径的持股比例较低,如果出现反常现象需引起监管部门的注意,警惕路径起点的金融机构运用杠杆.如果是为躲避监管而设计较长的股权路径,其背后动机非常复杂,可能目的之一是从金融机构套取资源.注意到起点为外资银行,终点为国有、股份制或民营银行的股权路径仅存在 Top - 1 的路径;而起点为外资银行,终点为农商行或外资银行的路径则存在 Top - 4 的路径;这与银行的性质有很大的关系,因为国有、股份制等类型的银行对外资持股占比进行了严格把控,而农商行正处于发展时期,亟需积极引进战略股东,所以外资银行股东占股较多.

第二,发现关键股权路径是识别不正当的关联交易的基础.股东对银行资源的侵蚀,本身就是系统性金融风险的重要来源.这种风险一旦爆发会沿着股权路径进行传导和扩散.例如,由上述分析可知,城商行与城商行之间的股权路径一般较短,最终收益权也高于其他类型的银行,且城商行的高杠杆会随着路径累加,造成被控制银行的经营风险,需要监管部门的重点关注.由此可见,发现关键股权路径并考虑银行的类别可以进一步帮助监管部门细化风险防控措施和体系.

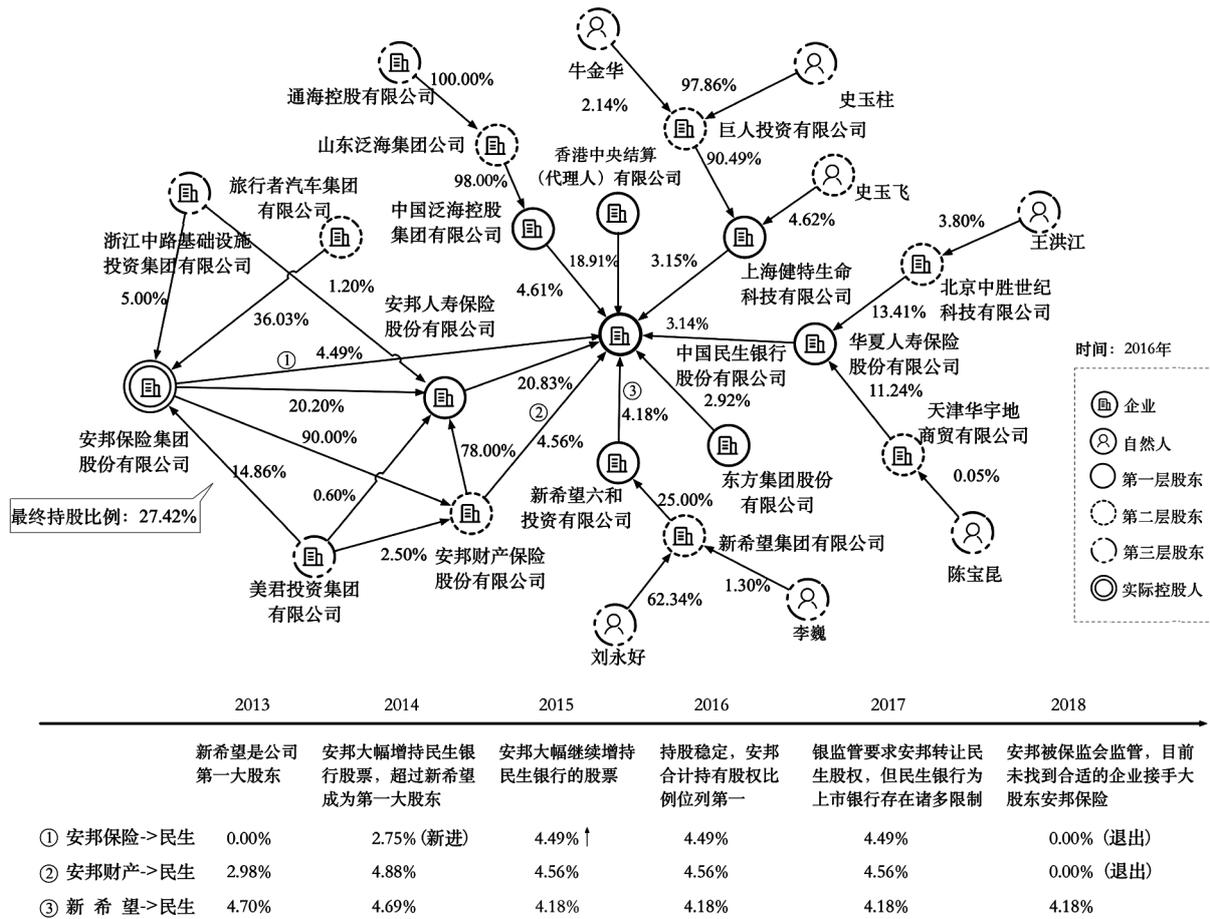


图7 民生银行穿透式多层股权网络

Fig. 7 Minsheng Bank penetrating multi-layer equity network

### 5.2 风险案例分析

安邦保险集团利用多层次复杂股权网络隐藏实际控股股东,并通过多层次股东循环持股、虚假注资形成600多亿的资本金.安邦保险集团最终被中国保监会托管,以防止安邦事件可能带来的风险传导效应,实质上防范了系统性金融风险.围绕“安邦事件”分析单个金融机构的穿透式多层股权结构的特征.通过调用MEP算法,在真实的金融股权知识大图中进行查询,穿透安邦系被接管之前的多层股权网络,评估其风险.图7展示了以“民生银行”为中心的穿透式多层股权网络(仅展示部分4层股东结构),并以时间轴的形式说明关键股权结构随时间的变化:从2013年到2018年,安邦从增持代替新希望成为民生银行的控股股东,到被接管之后相继退出民生银行的大股东之列.关键时间点包括了安邦财产保险股份有限公司从2013年的2.98%增持到2014年的

4.88%,再到2016年的4.56%,最后2018年的0%,表明股权的结构变化与金融风险的关联.如果通过MEP和CEP算法能够预先发现重要股权结构的时序变化,如股权增持引发的实际控股股东改变,则可以帮助监管部门预警可能的金融风险.

所有最终股东的持股比例可以通过公式(1)计算.如表7所示,按照最终持股比例排序得到的前10大股东中,安邦保险集团有限公司(简称安邦保险)为排名第一的最终股东,最终持股比例高达27.42%,可以认定为民生银行的实际控股股东.香港中央结算(代理人)有限公司是直接持股比例最高的股东,其持股比例为股东账户的股份总和,而非实质持股;这些股份的权益通过多层股权网络归属最终股东,实际控股股东被隐藏在层层股东之外.这样的高杠杆股权结构本身具有很大的风险.当实际控股股东出现风险时,风险有

较大概率传导给被控股公司.事实也证明,当安邦保险发生风险事件时,民生银行受到了风险冲击,股价大跌.

表7 民生银行最终股东持股比例

Table 7 Minsheng Bank's final shareholding ratio

最终股东	最终持股比例	排名
安邦保险集团股份有限公司	0.274 243 200	1
香港中央结算(代理人)有限公司	0.189 100 000	2
旅行者汽车集团有限公司	0.112 993 680	3
上海德悦股权投资管理有限公司	0.071 786 768	4
联通租赁集团有限公司	0.071 786 768	5
深圳祥隆股权投资管理有限公司	0.051 798 854	6
通海控股有限公司	0.046 100 000	7
卢志强	0.035 561 540	8
上海标准基础设施投资集团有限公司	0.028 408 142	9
史玉柱	0.027 894 357	10

表8 安邦保险到民生银行的关键股权路径

Table 8 Critical equity paths from Ampang Insurance to Minsheng Bank

路径	路径企业	控股权	长度	权重
第1条	安邦保险集团股份有限公司 → 安邦财产保险股份有限公司 → 安邦人寿保险股份有限公司 → 中国民生银行股份有限公司	1	3	0.146
第2条	安邦保险集团股份有限公司 → 安邦人寿保险股份有限公司 → 中国民生银行股份有限公司	1	2	0.042
第3条	安邦保险集团股份有限公司 → 安邦财产保险股份有限公司 → 中国民生银行股份有限公司	0	2	0.045

在得到民生银行的穿透式多层股权网络之后,进一步分析网络中两个公司之间的持股链路,找出关键的股权关联路径,帮助监管部门确定重要的股权关系,及时阻断风险.调用 CEP 算法,查询起点安邦保险与查询终点民生银行之间的 Top-3 关键股权路径.

查询结果如表 8 所示,第 1 条、第 2 条关键股权路径的控股权为 1,是完全控股路径;这两条路径具有较高的杠杆率,极易发生风险,从侧面印证了穿透式多层股权网络中的风险识别结果.纵观

Top-3 的股权路径,第 3 条路径的路径权重比例大于第 2 条路径,这是因为杠杆的运用使得较少的持股比例实现了控股权.这也印证了 CEP 算法需要区分持股、控股优先级的原因.关键股权路径说明了安邦保险注资资金流很多是通过第 1 条股权路径流通的,需要监管部门重点关注,当发生金融风险时,应及时进行阻断,防止风险的传染和扩散.

使用 MEP 和 CEP 算法对近年来典型金融风险事件进行了风险结构的发现和分析,如表 9 所示,此类事件共同点是股权结构成为引发风险的主要原因.金融复杂股权网络可以通过以下渠道对风险产生影响.其一是股东资本的真实性与合规性及多层股东带来的杠杆叠加机制,例如安邦事件和包商银行事件;其二是大股东关联交易带来的风险,及复杂股权网络下实际控股股东的高隐蔽性带来风险扩大,例如抚顺银行事件和中美天元事件.如表 9 所示,事前、事中、事后的实际控股股东发生了变化,安邦保险集团和辽宁永大集团通过前期股本大幅增持取代之前的实际控股股东的位置,再利用实际控制股东的权利操纵银行.包商银行事件和抚顺银行事件在风险发生之后,股权进行重组,共同点是重组之后的实控权由的民营企业转移至地方政府,最大限度的弥补损失,避免更多的风险.结合从实际控股股东到金融机构的关键股权路径的变化分析,可以看到具有较高风险隐患的实际控制股东的持股链路较为复杂,实际控股股东隐藏在层层股东之后,具有较高的隐蔽性.交叉持股即存在两家公司直接或间接地互为股东的情况.发现民生和包商银行的股权网络中存在交叉持股情况,在复杂的股权网络的情况下,容易造成虚假出资的潜在风险.

从以上案例分析中可以看出算法的重要性:在层层嵌套的金融机构股权网络之下,部分股东特别是实际控股股东可能不当干预金融机构的经营管理,将金融机构作为“取款机”,使得金融风险增加.而且实际控股股东到金融机构的关键股权路径也是风险传导扩散的重要路径.因此通过

算法分析穿透式多层股权网络和关键股权路径为理解金融机构的风险承担行为提供重要性的参考信息. 其中, MEP 算法可以洞察隐藏在多层股权网络中的实际控股股东, 实现穿透式监管, 为系统性金融风险提供决策支持. 关键股权路径的发现可以揭示金融机构之间的控股权和杠杆, 当出现持股比例与路径长度正相关的异常情况时, 则应警惕高杠杆的应用. 与此同时, 监管部门在分析和阻断风险传染的路径时, 应是重点考虑关键股权路径, 因此 CEP 算法可为系统性风险的监管提供决策支撑.

基于以上结论, 提出了 4 点政策建议: 1) 对金融机构进行穿透式股权监管, 基于时序股权数

据持续监控和识别实际控股股东, 与公开披露的控股股东进行比对验证, 如不同应确认该金融机构的实际控股股东, 避免实际控股股东处于监管盲区. 2) 计算重点金融机构的最终股东的国有股权占比, 如果占比低于该机构所属类别的国有股权占比平均水平, 则需要监管部门进行关注. 3) 建议对实际控股股东与金融机构的关键股权路径进行发现和核查, 基于股权高杠杆、交叉持股等风险特征的计算和分析对可能的风险进行预警. 4) 应将关键股权路径上的金融和非金融机构纳入监管范围, 当风险发生时, 对以上机构进行重点关注并拟定针对性的监管措施, 防范风险沿路径传导.

表 9 典型金融风险事件的股权结构分析

Table 9 Analysis on the equity structure of typical financial risk events

事件案例	事件说明	金融机构	实际控股股东	关键股权路径	交叉持股
安邦事件	2018 年 2 月安邦集团通过层层股权循环出资放大资本, 被保监会监管	中国民生银行股份有限公司	新希望六和投资有限公司(2013) → 安邦保险集团股份有限公司(2016)	Top-1(2013): 新希望六和投资有限公司 → 中国民生银行股份有限公司 Top-1(2016): 安邦保险集团股份有限公司 → 安邦财产保险股份有限公司 → 安邦人寿保险股份有限公司 → 中国民生银行股份有限公司	有
包商银行事件	2019 年 5 月包商银行出现大股东占用资金出现信用风险被接管	包商银行股份有限公司	明天控股有限公司(2016) → 中国人民银行(2019)	Top-1(2016): 明天控股有限公司 → 包头市北普实业有限公司 → 包头市聚锋贸易有限责任公司 → 包头市百川投资有限公司 → 包商银行股份有限公司 Top-1(2019) 中国人民银行 → 存款保险基金管理有限责任公司 → 包商银行股份有限公司	有
抚顺银行事件	2014 年永大集团大股东借抚顺银行炒作永大股票, 然后大幅减持套现	抚顺银行股份有限公司	建龙钢铁控股有限公司(2005) → 辽宁永大集团(2013)	Top-1(2005): 建龙钢铁控股有限公司 → 抚顺新钢铁有限责任公司 → 抚顺银行股份有限公司 Top-1(2013): 辽宁永大集团 → 抚顺银行股份有限公司	无
中美天元事件	2014 年吴东集团从柳州银行骗贷金额达 313 亿的贷款诈骗案浮出水面	柳州银行股份有限公司	吴东(2014) → 柳州市人民政府国有资产监督管理委员会(2017)	Top-1(2014): 吴东 → 广西中美天元融资性担保集团有限公司 → 广西海生药业有限公司 → 柳州银行股份有限公司 Top-1(2017): 柳州市人民政府国有资产监督管理委员会 → 柳州市龙建投资发展有限责任公司 → 柳州银行股份有限公司	无

5.3 算法性能分析

实验使用 python 开发环境实现算法. 机器的配置是 CentOS 7.4, 8 核 CPU, 32GB 内存. 随机进

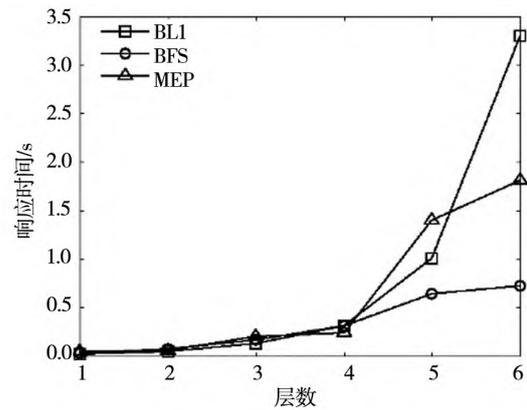
行 1 000 次查询, 分别比较 MEP、CEP 算法与对比算法的查询性能. 查询性能的度量基准为查询的响应时间.

MEP 算法的对比算法包括：(a) BFS 算法，即广度优先遍历算法，首先通过遍历确定整个多层股权结构图，再由外往内分层确定层级，需要对多层股权网络进行多次遍历。(b) BaseLine1 (BL1) 算法，运用 SPARQL 查询语言查询图数据库以确定股权的层级。BL1 算法调用 Dijkstra 算法重新计算带权最短路径。实验分别评测平均查询响应时间(简称查询时间)对于层级以及图结点数量的变化。

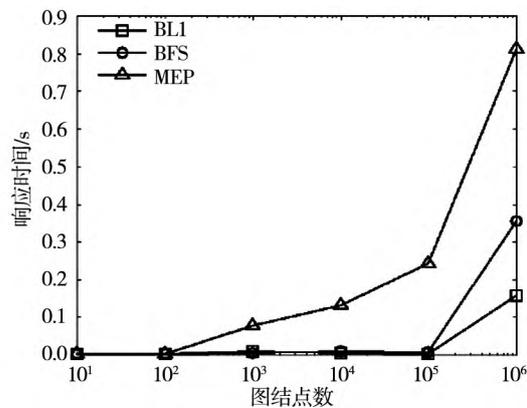
如图 8(a) 所示，查询时间与层级数量成正比，当平均层级较低时三种算法的查询时间相差不大，然而随着层级的增加，MEP 算法的查询时间相对于 BFS 算法和 BL1 算法来说，涨幅最小，其中 BFS 算法的查询响应时间增加最快。这是因为 BFS 算法需要多次遍历，而 MEP 算法仅需一次遍历。当查询穿透式多层股权层级不变时，如果 8(b) 所示，各算法的查询时间随着图结点数量增加而随之增加，可以看到 BL1 算法增长速度明显高于其他两种算法。因此，MEP 算法的查询性能要优于 BFS 算法和 BL1 算法。

CEP 算法的对比算法包括：(a) 对于求解有向图中包含环路的 KSP 问题而言， $K^*$  是目前最好的 KSP 算法<sup>[35]</sup>，因此选择经典算法  $K^*$  进行对比。 $K^*$  算法首先用  $A^*$  算法对图进行搜索直到目标结点被找到，并根据搜索部分构造  $p(G)$ ，再调用 Dijkstra 算法计算最短路径，如果找到 Top- $k$  路径则停止，否则继续启用  $A^*$  算法搜索重复以上过程，直到找到所有路径。由于  $K^*$  算法并未考虑边的优先级，因此所找到的 Top- $k$  路径并不一定是正确结果，如 4.2.2 节偏差情况中  $v_1 \rightarrow v_3 \rightarrow v_0$  路径虽然持股比例小于  $v_1 \rightarrow v_0$ ，但是  $v_1 \rightarrow v_3 \rightarrow v_0$  是控股路径， $v_1 \rightarrow v_0$  是持股路径。因此， $v_1 \rightarrow v_3 \rightarrow v_0$  的优先级更高。为了能够和 CEP 算法进行公平的比较，将  $K^*$  算法修改为  $K^* - m$  算法，使得  $K^* - m$  算法能解决多优先级的 KSP 问题。 $K^* - m$  算法步骤如下：首先运行  $K^*$  算法找到 Top- $k'$  路径 ( $k = k'$ )，然后将结果与 CEP 查询所得的 Top- $k$  路径进行比较，如果相同，则输出结果；否则  $k' = k' + n$ ，其中  $n$  为金融股权知识大图中结点的

平均度，根据表 1，本实验  $n = 3$ ；以上过程不断迭代，直到  $K^*$  算法找到的 Top- $k'$  中包含 CEP 查询所得的 Top- $k$  路径，则算法结束，输出结果，注意此时  $k' > k$ 。(b) BaseLine2 (BL2) 算法，运用 SPARQL 查询语言查询图数据库中两个结点之间的股权路径，只能输出确定跳数内的股权路径，且需要计算出所有路径的权重，以输出 Top- $k$  关键股权路径。实验分别评测查询时间与平均长度、 $k$  值、图结点数的变化关系。



(a)



(b)

图 8 穿透式多层股权算法查询响应时间

Fig. 8 Query response time of penetrating multi-level equity algorithm

如图 9(a) 所示。当图结点数、 $k$  值固定时，CEP 算法查询性能随着路径平均长度的增加优于  $K^* - m$  算法和 BL2 算法，主要原因是  $K^* - m$  算法需要进行多次迭代，而且由于设定的  $k$  值较小， $K^* - m$  算法首先需要建立一个复杂的路径图，使得效率降低。如图 9(b) 所示，各个算法的查询时间与  $k$  值成正比，其中 CEP 算法的增长速度最慢，具有较好的可扩展性。BL2 算法随着  $k$  的增

幅最大,值得注意的是在  $k = 2$  之前出现了一个交叉点,主要是因为所以在  $k$  较小时,  $K^* - m$  的整体代价高于 BL2. 如图 9(c) 所示, CEP 算法的查询时间最短, 当到达  $10^4$  结点数量之后, BL2 算法呈指数增加, 而 CEP 算法与  $K^* - m$  算法的增长较为平缓. 但是在此之后  $K^* - m$  算法响应时间高于 CEP 算法, 主要是因为  $K^* - m$  需要迭代多次得到结果. 相比于对比算法, CEP 算法查询效率较高, 且在大规模的数据集上具有良好的可扩展性. 因为 CEP 算法可以根据优先级和路径的权重(即持股比例)过滤不在 Top -  $k$  中的路径, 不需要遍历起点和终点之间所有的边, 可以节约更多的时间, 适用于大数据环境下金融股权知识大图的查询和分析.

MEP 算法的时间复杂度为  $O(|V| + |E|)$ , 其中  $|V|$  为大图中结点的数量,  $|E|$  是边的数量, 因为最坏情况图中的每个点和边都会被访问到; 空间复杂度为  $O|V|$ , 最坏情况需要存储图中所有结点. 当图的规模过大而无法存储时, 则可以根据 MEP 算法穿透的最大层数  $l$  (详见定义 4) 来描述复杂度, 此时 MEP 算法的时间复杂度和空间复杂度均可表示为  $O(b^l + 1)$ , 其中  $b$  为知识大图结点的平均出度. 同理, CEP 算法的时间复杂度为  $O(b^l)$ , 其中  $l$  是关键股权路径的最大长度(注意到 MEP 在确定结点层级时调用了 CEP), CEP 算法的空间复杂度为  $O(b^l)$ , 因为 CEP 在内存中存储了算法过程中所有产生的结点.

BFS 算法的时间复杂度为  $O(|V| + |E|)$ , 空间复杂度为  $O|V|$ ; BL1 和 BL2 由于在调用 SPAR-QL 语句过程中均涉及遍历图中的邻接结点和边, 因此在最坏情况下需要访问图中所有的点和边, BL1 和 BL2 的时间和空间复杂度均为  $O(|V| + |E|)$ ;  $K^* - m$  算法的时间和空间复杂度为  $O(|E| + |V| \ln |V| + k)$  [35].

由以上分析可知, MEP 算法的时间复杂度等于 BFS 和 BL1, 空间复杂度等于 BFS, 小于 BL1, 但是由于 MEP 算法采用了优化技术仅需一次遍历, 因此查询时间仍然优于 BFS 和 BL1. 因为  $b^l \leq |V|$ , 所以 CEP 算法的时间和空间复杂度均优于  $K^* - m$  和 BL2.

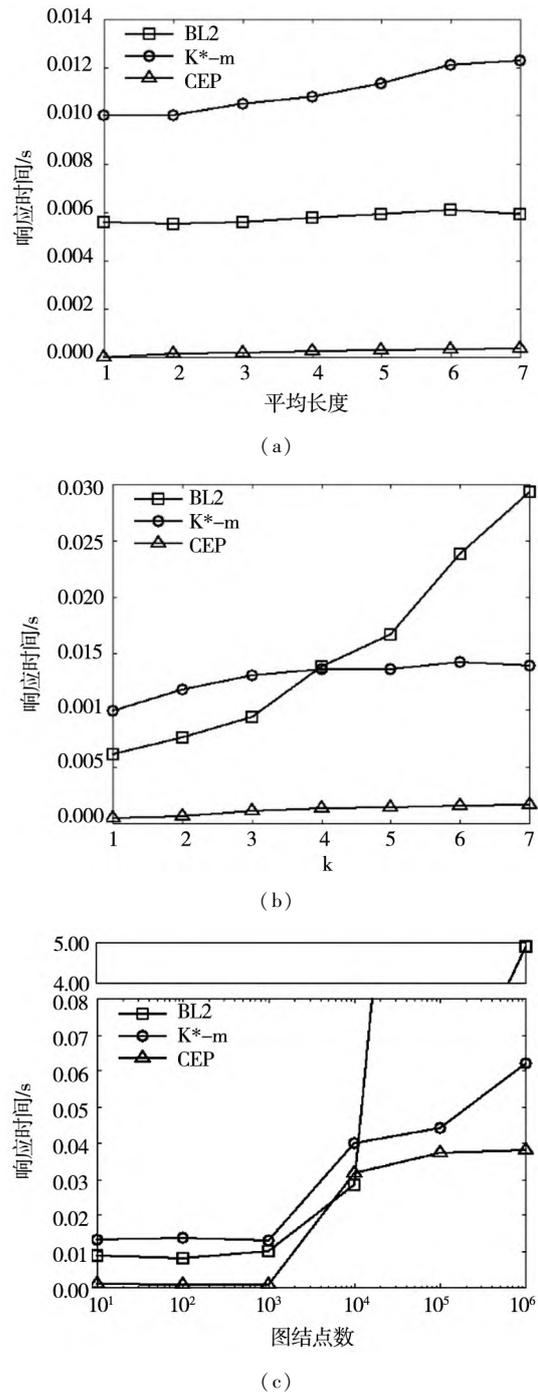


图9 关键股权路径发现算法查询响应时间

Fig.9 Query response time of critical equity path algorithm

## 6 结束语

基于知识关联理论, 构建了亿级结点的金融股权知识大图, 将金融股权网络中的风险结构发现与分析问题转化为金融股权知识大图中的知识

关联发现与风险分析. 提出了多层股权穿透(MEP)算法和关键股权路径发现(CEP)算法,能够从金融股权知识大图中准确高效地发现穿透式多层股权网络和关键股权路径.同时,对发现的两个知识关联结构进行了深度分析,并通过对案例的分析进一步验证了股权知识关联与系统性金融风险的关系.

研究取得了诸多具有理论意义与实践指导意义的结论,其中包括:1)多层股权网络穿透算法可以发现隐藏在层层股东之中的实际控股股东,支持穿透式监管.2)关键股权路径发现算法可以识别股权控制链路,从而为系统性金融风险传导路径提供参考信息,特别是从金融机构向非金融机构风险传导的路径.3)金融机构的

最终股东的国有股权最终持股占比,是国家金融安全、政治安全的基础判断依据.本文的研究结论表明,股权网络风险特征分析有助于识别和分析系统性风险,为穿透式监管提供必要的决策支持.

未来的工作包括:首先,进一步优化所提出的算法,提高算法的有效性和性能.其次,是在股权知识大图上实现对我国金融机构的整体股权结构的穿透式解析与关键特征分析,并建立系统性金融风险的度量标准.最后,将基于金融股权知识大图总结出系统性风险的形成机理,结合提出的知识关联发现算法,构建系统性风险的识别与传导模型,实现时序动态分析,为进一步防范系统性金融风险提供辅助决策.

#### 参考文献:

- [1] Huang W Q, Zhuang X T, Yao S, et al. A financial network perspective of financial institutions' systemic risk contributions [J]. *Physica A: Statistical Mechanics and its Applications*, 2016, 456: 183 - 196.
- [2] Alexander K, Dhumale R. *Enhancing Corporate Governance for Financial Institutions: The Role of International Standards* [M]. Cambridge: ESRC Centre for Business Research, University of Cambridge, 2001.
- [3] Iyer R, Peydro J L. Interbank contagion at work: Evidence from a natural experiment [J]. *The Review of Financial Studies*, 2011, 24(4): 1337 - 1377.
- [4] Silva T C, Da Silva Alexandre M, Tabak B M. Bank lending and systemic risk: A financial-real sector network approach with feedback [J]. *Journal of Financial Stability*, 2018, 38: 98 - 118.
- [5] 赵晶, 郭海. 公司实际控制权, 社会资本控制链与制度环境 [J]. *管理世界*, 2014, 9: 160 - 171.  
Zhao Jing, Guo Hai. Company actual control rights, social capital control chain and institutional environment [J]. *Management World*, 2014, 9: 160 - 171. (in Chinese)
- [6] 陈富赞, 李敏强. 基于邻接网络的频繁项目集发现算法 [J]. *管理科学学报*, 2006, 9(3): 54 - 61.  
Chen Fuzan, Li Minqiang. Frequent itemsets discovery algorithm based on adjacency network [J]. *Journal of Management Sciences in China*, 2006, 9(3): 54 - 61. (in Chinese)
- [7] Barro J R, Barro R J. Pay, performance, and turnover of bank CEOs [J]. *Journal of Labor Economics*, 1990, 8(4): 448 - 481.
- [8] Wright P, Ferris S P, Sarin A, et al. Impact of corporate insider, blockholder, and institutional equity ownership on firm risk taking [J]. *Academy of Management Journal*, 1996, 39(2): 441 - 458.
- [9] Barry T A, Lepetit L, Tarazi A. Ownership structure and risk in publicly held and privately owned banks [J]. *Journal of Banking & Finance*, 2011, 35(5): 1327 - 1340.
- [10] Rotundo G, D'arcangelis A M. Ownership and control in shareholding networks [J]. *Journal of Economic Interaction and Coordination*, 2010, 5(2): 191 - 219.
- [11] 贾明, 张喆, 万迪昉. 控制性股东侵占行为及其负外部性研究 [J]. *管理科学学报*, 2010(1): 70 - 77.

- Jia Ming, Zhang Zhe, Wan Difang. Research on controlling shareholder embezzlement and its negative externalities[J]. *Journal of Management Sciences in China*, 2010(1): 70–77. (in Chinese)
- [12] Hooy G K, Hooy C W, Chee H K. Ultimate ownership, control mechanism, and firm performance: Evidence from Malaysian firms[J]. *Emerging Markets Finance and Trade*, 2020, 56(15): 3805–3828.
- [13] Haldane A G, May R M. Systemic risk in banking ecosystems[J]. *Nature*, 2011, 469(7330): 351–355.
- [14] Billio M, Getmansky M, Lo A W, et al. Econometric measures of connectedness and systemic risk in the finance and insurance sectors[J]. *Journal of financial economics*, 2012, 104(3): 535–559.
- [15] Elliott M, Golub B, Jackson M O. Financial networks and contagion[J]. *American Economic Review*, 2014, 104(10): 3115–53.
- [16] Elsinger H. *Financial Networks, Cross Holdings, and Limited Liability*[M]. Vienna: Oesterreichische Nationalbank Austria, 2009.
- [17] Gouriéroux C, Heam J-C, Monfort A. Liquidation equilibrium with seniority and hidden CDO[J]. *Journal of Banking & Finance*, 2013, 37(12): 5261–5274.
- [18] 裴 茜, 朱书尚. 中国股票市场金融传染及渠道——基于行业数据的实证研究[J]. *管理科学学报*, 2019(3): 90–112.
- Pei Qian, Zhu Shushang. Financial contagion and channels in China's stock market: An empirical study based on industry data[J]. *Journal of Management Sciences in China*, 2019(3): 90–112. (in Chinese)
- [19] 阮光册, 夏 磊. 基于词共现关系的检索结果知识关联研究[J]. *情报学报*, 2017, 36(12): 1247–1254
- Ruan Guangce, Xia Lei. Research on knowledge association of retrieval results based on word co-occurrence relationship [J]. *Journal of Information*, 2017, 36(12): 1247–1254. (in Chinese).
- [20] 张 俊. 商业化证券信息服务及其实现策略[D]. 武汉: 武汉大学, 2005.
- Zhang Jun. *Commercialized Securities Information Service and Its Realization Strategy*[D]. Wuhan: Wuhan University, 2005. (in Chinese)
- [21] 张 瑞, 唐旭丽, 王定峰, 等. 基于知识关联的金融数据可视化分析[J]. *情报理论与实践*, 2018, 41(10): 131.
- Zhang Rui, Tang Xuli, Wang Dingfeng, et al. Visual analysis of financial data based on knowledge association[J]. *Information Theory and Practice*, 2018, 41(10): 131. (in Chinese)
- [22] 李旭晖, 凡美慧. 大数据中的知识关联[J]. *情报理论与实践*, 2019, 42(2): 72–77.
- Li Xuhui, Fan Meihui. Knowledge association in big data[J]. *Information Theory and Practice*, 2019, 42(2): 72–77. (in Chinese)
- [23] 唐旭丽, 马费成, 傅维刚, 等. 知识关联视角下的金融知识表示及风险识别[J]. *情报学报*, 2019, 38(3): 286–298.
- Tang Xuli, Ma Feicheng, Fu Weigang, et al. Financial knowledge representation and risk identification from the perspective of knowledge association[J]. *Journal of Information*, 2019, 38(3): 286–298. (in Chinese)
- [24] Liang Z, Pan D, Deng Y. Research on the knowledge association reasoning of financial reports based on a graph network [J]. *Sustainability*, 2020, 12(7): 2795.
- [25] O'riain S, Curry E, Harth A. XBRL and open data for global financial ecosystems: A linked data approach[J]. *International Journal of Accounting Information Systems*, 2012, 13(2): 141–162.
- [26] Li H, Zhai J. *Constructing Investment Open Data of Chinese Listed Companies Based on Linked Data*[C]. New York: International Digital Government Research Conference on Digital Government Research, 2016.
- [27] Destefano R, Tao L, Gai K. Improving Data Governance in Large Rrganizations through Ontology and Linked Data[C].

- New York;2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud), 2016: 279 – 284.
- [28] Broekstra J, Kampman A, Van Harmelen F. Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema[C]. Berlin: International Semantic Web Conference, 2002: 54 – 68.
- [29] Yen J Y. Finding the k shortest loopless paths in a network[J]. *Management Science*, 1971, 17(11): 712 – 716.
- [30] Martins E Q, Pascoal M M. A new implementation of Yen’s ranking loopless paths algorithm[J]. *Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, 2003, 1(2): 121 – 133.
- [31] Hershberger J, Maxel M, Suri S. Finding the k shortest simple paths: A new algorithm and its implementation[J]. *ACM Transactions on Algorithms (TALG)*, 2007, 3(4): 45 – es.
- [32] Hart P E, Nilsson N J, Raphael B. A formal basis for the heuristic determination of minimum cost paths[J]. *IEEE Transactions on Systems Science and Cybernetics*, 1968, 4(2): 100 – 107.
- [33] Eppstein D. Finding the K Shortest Paths[C]. Santa Fe: Symposium on Foundations of Computer Science, 1994.
- [34] Jiménez V M, Marzal A. A Lazy Version of Eppstein’s K Shortest Paths Algorithm[C]. Berlin: International Workshop on Experimental and Efficient Algorithms, 2003: 179 – 191.
- [35] Aljazzar H, Leue S. K\*: A heuristic search algorithm for finding the K shortest paths[J]. *Artificial Intelligence*, 2011, 175(18): 2129 – 2154.
- [36] Barabási A-L, Albert R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509 – 512.
- [37] Kogut B, Walker G. The small world of Germany and the durability of national networks[J]. *American Sociological Review*, 2001: 317 – 335.
- [38] Bertoni F, Randone P A. The small-world of Italian finance: Ownership interconnections and board interlocks amongst Italian listed companies[J]. Available at SSRN 917587, 2006.
- [39] Conyon M J, Muldoon M R. Ownership and control: A small-world analysis[J]. *Advances in Strategic Management*, 2008, 25: 31 – 65.
- [40] Kusumoto M, Maehara T, Kawarabayashi K-I. Scalable Similarity Search for SimRank[C]. New York: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, 2014: 325 – 336.
- [41] Levy M. Control in Pyramidal structures[J]. *Corporate Governance An International Review*, 2010, 17(1): 77 – 89.
- [42] Chapelle A, Szafarz A. Controlling firms through the majority voting rule[J]. *Physica A Statistical Mechanics & Its Applications*, 2005, 355(2): 509 – 529.
- [43] Zou L, Mo J, Chen L, et al. gStore: Answering SPARQL queries via subgraph matching[J]. *Proceedings of the VLDB Endowment*, 2011, 4(8): 482 – 493.
- [44] Brioschi F, Buzzacchi L, Colombo M G. Risk capital financing and the separation of ownership and control in business groups[J]. *Journal of Banking & Finance*, 1989, 13(4 – 5): 747 – 772.
- [45] Cheeseman P C, Kanefsky B, Taylor W M. Where the Really Hard Problems Are[C]. San Francisco: IJCAI, 1991: 331 – 337.
- [46] 李涛. 混合所有制公司中的国有股权[J]. *经济研究*, 2002, 8: 19 – 27.  
Li Tao. State-owned equity in a mixed-ownership company[J]. *Economic Research*, 2002, 8: 19 – 27. (in Chinese)
- [47] 徐二明, 张晗. 中国上市公司国有股权对创新战略选择和绩效的影响研究[J]. *管理学报*, 2011, 8(2): 206.  
Xu Erming, Zhang Han. Research on the influence of state-owned equity of Chinese listed companies on innovation strategy choice and performance[J]. *Chinese Journal of Management*, 2011, 8(2): 206. (in Chinese)
- [48] 步艳红, 赵晓敏, 杨帆. 我国商业银行同业业务高杠杆化的模式, 影响和监管研究[J]. *金融监管研究*, 2014, 2: 33 – 46.

Bu Yanhong, Zhao Xiaomin, Yang Fan. Research on the mode, impact and supervision of high leverage in my country's commercial bank interbank business[J]. Financial Supervision Research, 2014, 2: 33–46. (in Chinese)

## Knowledge association discovery and risk analysis based on financial equity knowledge graph

*HONG Liang*<sup>1, 2, 3</sup>, *OUYANG Xiao-feng*<sup>1, 2, 3</sup>

1. Center for the Studies of Information Resources, Wuhan University, Wuhan 430072, China;
2. School of Information Management, Wuhan University, Wuhan 430072, China;
3. Big Data Institute, Wuhan University, Wuhan 430072, China

**Abstract:** The complex equity network of financial institutions is the micro cause and transmission channel of systemic financial risk. Therefore, it is of great significance for financial supervision to find and analyze the key risk characteristic structures in the equity network. According to the theory of knowledge association, a financial equity knowledge graph is constructed based on large-scale equity networks. A multi-layer equity penetration (MEP) algorithm is proposed to reveal the penetrating multi-layer equity network centered on a financial institution. A critical equity path (CEP) algorithm is proposed to discover the critical equity path between financial institutions. This paper analyzes empirically the above equity risk structure based on knowledge association and compares the performance of MEP algorithm and CEP algorithm. The experimental results indicate that the proposed algorithms are more efficient and accurate in discovering and analyzing the key risk characteristic structures in equity networks, compared with the traditional equity structure analysis methods. Furthermore, the study profits the penetrating supervision of systemic financial risk.

**Key words:** financial equity knowledge graph; penetrating multi-layer equity network; critical equity path; knowledge association