

Elseide 艾思亿德

艾思产研数据平台系列数据库

——企业族群数据库

<http://idr.elseide.com>



目录

一、研发背景 1

二、应用场景 1

 （一）金融风控 1

 （二）投资决策 2

 （三）情报分析 2

 （四）供应链管理 2

三、识别方法 3

 （一）相关概念 3

 （二）算法解析 4

 （三）实施步骤 5

四、内容介绍 6

 （一）数据内容 6

 （二）数据样例 7

 （三）数据统计 12

五、数据优势 13

 （一）高精度权重建模 13

 （二）多维度信息保留 13

 （三）动态稳定性保障 13

 （四）灵活扩展性 14

 （五）业务导向输出 14

艾思企业族群数据库介绍

一、研发背景

企业图谱作为企业、投资人以及投资、任职关系的复杂网络，在企业数据分析领域应用广泛。但在传统企业图谱分析中，LPA 标签传播算法、Louvain 算法等主要依赖节点间的“有无关系”，而忽略了企业网络中投资、任职等关系的权重差异（如股权占比、职位高低），导致无法判断企业或个人之间的紧密程度。此外，传统方法将节点仅归于单一族群，导致大量关联信息丢失（例如同一企业可能属于多个商业集团）。

族群划分作为复杂网络的关键特征，为解决上述问题提供了新思路。艾思亿德采用企业族群探索方法，基于企业图谱节点间关系开展族群划分，该方法能够识别企业间潜在商业集团或合作伙伴，具有广泛应用场景与显著数据优势。借助权重化处理、多族群置信概率分析以及相似度分析，该技术在投资推荐、风险控制、商业合作以及情报分析等领域展现出重要的应用价值。

二、应用场景

（一）金融风控

艾思企业族群数据库可识别隐性关联企业，防范集团性信贷风险。一些企业通过复杂股权结构、关联担保或共同实际控制人隐藏真实债

务，导致金融机构面临多头授信、过度融资风险。该库通过分析企业图谱中的控股、任职及资金流向，提前发现潜在风险，避免因信息不对称造成的信贷损失。这种深度关联分析提升了金融机构的风险识别能力，为贷前、贷中、贷后管理提供全面数据支持，保障资金安全。

（二）投资决策

艾思企业族群数据库助力投资者精准决策。通过分析企业图谱中的控股、任职及合作网络，投资者可全面评估投资标的的市场地位和风险，发现潜在联盟或竞争对手。例如，某些企业可能通过复杂股权结构隐藏关联或与竞争对手共享实际控制人。这种深度分析为投资者提供全面视角，帮助筛选优质投资机会，避免因信息不足导致失误，提升投资决策的科学性和准确性。

（三）情报分析

艾思企业族群数据库可深入剖析企业间的控股、任职及合作网络，挖掘实际控制人、关联企业和利益输送路径。例如，通过股权穿透、任职关联和资金流向分析，可发现企业通过隐性控股或交叉任职控制多家公司，实现利益输送或规避监管。这种深度情报分析为企业战略决策提供支持，助力企业精准识别合作伙伴或竞争对手，优化商业布局，提升市场竞争力。

（四）供应链管理

艾思企业族群数据库助力优化供应链管理。通过分析企业图谱中的控股、任职和合作网络，企业可识别关键节点供应商和客户的潜在

风险，优化资源配置，提升供应链的稳定性和效率。这有助于企业在市场竞争中占据优势地位。

三、识别方法

（一）相关概念

1、企业图谱

企业图谱是一种关于企业、投资人以及投资、任职关系的复杂网络，如图 1 所示。在这个网络中，企业、投资人被抽象为图中的节点，而节点之间的投资、被投资与任职关系则被抽象为节点之间的边。这些节点和边共同组成了一个复杂且庞大的企业关系网络，即企业图谱。

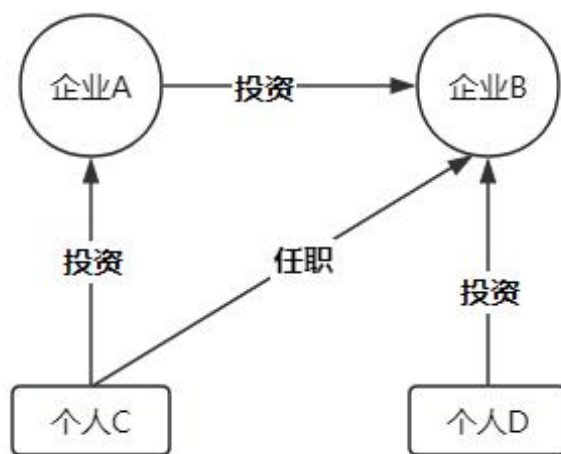


图 1 企业图谱示例

2、企业族群

企业族群则是企业关联图谱中通过关系权重（如投资关系、任职关系、相似度关系等）划分的紧密联系的子图，这些企业或个人节点在投资、任职或业务上存在较高的关联性，如图 2 所示。通过族群划分，可以更清晰地了解企业图谱中的结构特征，发现潜在的商业集团

或利益共同体。

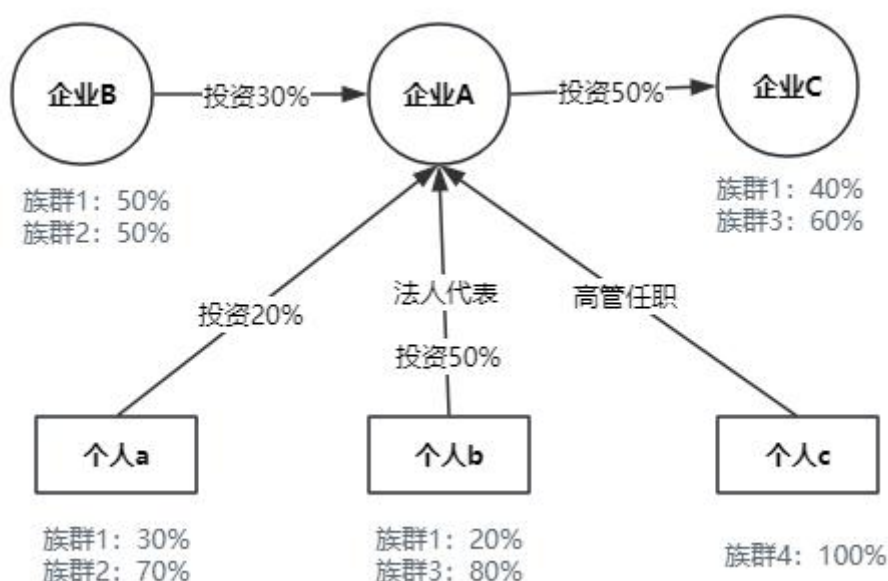


图 2 企业族群示例

如图 2 所示,在该企业图谱中,族群 1 的潜在成员包含企业 B(归属概率为 50%)、企业 C(归属概率 40%)、个人 a(归属概率 30%)以及个人 b(归属概率 20%)。

族群 2 的潜在成员包含企业 B(归属概率为 50%)、个人 a(归属概率 70%)。

族群 3 的潜在成员包含企业 C(归属概率为 60%)、个人 b(归属概率 80%)。

族群 4 的潜在成员包含个人 c(归属概率为 100%)。

(二) 算法解析

艾思企业族群数据库基于企业族群探索方法构建,企业族群识别依托于标签传播算法(Label Propagation Algorithm, LPA)。这是一

种基于图的半监督学习算法，凭借计算效率高、无需预先指定聚类数量等显著优势，在聚类、分类、社区发现等任务中广泛应用，涵盖社交网络分析、生物信息学、图像分割等诸多领域。

标签传播算法以图模型为基础，将数据集中的每个数据点视作图的节点，通过边的权重来体现节点间的相似性，从而构建出加权无向图。其核心思想在于，节点的标签会根据与之相连节点的标签信息不断传播和更新，直至达到稳定状态。此时，拥有相同标签的节点便被划分到同一类或社区中。

（三）实施步骤

1、数据预处理

首先，根据接收的企业数据建立企业图谱，企业图谱中包括节点和节点间的关系及权重，将每个节点定义为一个族群，并设置族群的置信概率为 100%。

其中，节点和节点间的关系及权重，包括：节点与节点间的投资关系、任职关系以及相似关系，[投资关系的权重由其股权占比确定](#)，[任职关系的权重由其职位确定](#)，[相似度关系的权重根据行业/地域匹配度确定](#)。越紧密的关系对应着越高的权重，紧密关系连接的双方也更应该属于同一个族群。当所述关联节点同时与所述目标节点存在多种关联关系时，其对应的关系权重通过各种关联关系对应的关系权重叠加得到。

2、选择目标节点

随机选择一个节点作为目标节点，开始族群划分过程。

3、确定目标节点族群

根据目标节点的关联节点和权重关系，确定目标节点的族群及其置信概率。

4、迭代处理剩余节点

将剩余节点依次作为目标节点，进行族群及其置信概率的确定。

5、判断族群稳定性

判断所有节点是否完成族群及其置信概率的确定，若完成则判断节点的族群及其置信概率是否稳定。如果稳定，则输出所述企业图谱中各节点的族群及其置信概率；如果不稳定，重置所有节点为未遍历状态，并迭代执行上述步骤，直到族群置信概率趋于稳定为止。

四、内容介绍

（一）数据内容

基于上述族群划分方法，艾思亿德构建了企业疑似商业集团/族群/个人合作伙伴库，共包括三张数据表，分别是个人族群关系表、企业族群关系表以及族群基础信息表，如表 1 所示。

表 1 艾思企业族群数据库

数据表	字段名称	字段释义	数据总量
个人族群关系	个人 UID	个人唯一标识符，用于唯一标识一个个人用户	10964 万+
	族群 id	族群标识符，用于标识某个企业族群	

	节点的度	节点连接边的条数	
企业族群关系	企业 UID	企业唯一标识符, 用于唯一标识一个企业用户	20433 万+
	族群 id	族群标识符, 用于标识某个企业族群	
	节点的度	节点连接边的条数	
	核心指数	与族群的关联程度, 指数越高, 关联度越高	
族群基础信息	族群 id	族群标识符, 用于标识某个企业族群	26587 万+
	总成员数	某个族群的总成员数量	
	企业成员数	属于某个族群的企业成员数量	
	个人成员数	属于某个族群的个人成员数量	

*注: 本数据库基于算法确定的企业族群成员, 以概率形式呈现, 并不意味着成员归属是必然发生的确定事件。

(二) 数据样例

随着《哪吒之魔童闹海》票房一路飙升, 截至 2025 年 2 月 15 日上午 10 点, 票房 (含预售) 突破 108.45 亿, 跻身全球影史票房榜前 13。作为背后主投上市公司, 北京光线传媒股份有限公司 2 月 15 日收盘斩获 20CM 五连板, 年初迄今股价最大涨幅 310.04%, 市值超 1000 亿元, 成为 A 股市值最高的影视院线上市公司。公司还计划构建“中国神话宇宙”体系, 动画储备项目众多。

下文以北京光线传媒股份有限公司为例, 对其族群信息进行分析。

1、企业图谱

北京光线传媒股份有限公司的企业图谱如图 3 所示, 图谱包含如下信息: 1 个法定代表人, 45 个股东, 29 个历史股东, 13 个高管,

17 个历史高管，29 个对外投资企业，20 个历史对外投资企业，11 个分支机构。

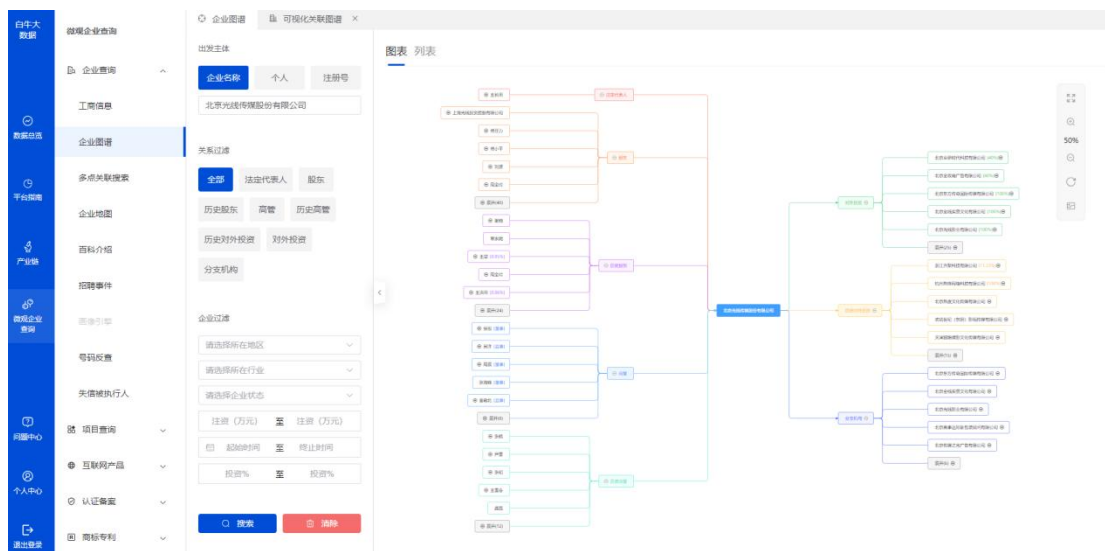


图 3 北京光线传媒股份有限公司企业图谱

点击“+”则可展开更多关联节点信息。在庞杂的企业关联网络中，寻找联系更为紧密的企业或个人合作关系，需要借助企业族群划分和知识图谱分析方法。

2、族群基础信息

依据艾思企业族群数据库，北京光线传媒股份有限公司隶属于两个潜在族群，族群 A 和族群 B。如表 2 所示，族群 A 的总成员数为 36 个，族群 B 的总成员数为 181 个。

表 2 北京光线传媒股份有限公司族群基础信息

族群 id	总成员数	企业成员数	个人成员数
111236983 (A)	36	25	11
114045699 (B)	181	99	82

3、企业族群关系

下文以族群 A 为例，对企业族群关系做分析。表 3 中的企业均属于族群 A（111236983），该族群共包含 25 家企业，这些企业之间通过控股、投资、业务协同等关系形成了紧密的关联网络，疑似为商业集团。族群 A 的企业覆盖影视制作、发行、广告、艺人经纪、科技等多领域，构建起完整产业链。

表 3 企业族群（A）关系-基于北京光线传媒股份有限公司

企业名称	族群 id	节点的度	核心指数
北京光线传媒股份有限公司	111236983	12	1351
北京光线影业有限公司	111236983	20	1308
光线控股有限公司	111236983	13	1177
霍尔果斯彩条屋影业有限公司	111236983	13	1175
霍尔果斯五光十色影业有限公司	111236983	23	1157
山南光线影业有限公司	111236983	36	1156
霍尔果斯青春光线影业有限公司	111236983	9	1156
北京彩条屋科技有限公司	111236983	12	1139
山南光启影视有限公司	111236983	42	1117
五光十色（东阳）文化传媒有限公司	111236983	49	898
鄂尔多斯市光线影业有限公司	111236983	41	766
霍尔果斯聚光影业有限公司	111236983	51	738
光线（北京）科技有限公司	111236983	17	712
北京全攻略广告有限公司	111236983	38	704
上海聚合广告有限公司	111236983	43	687
山南光线艺人经纪有限公司	111236983	43	687
天津夜线影业有限公司	111236983	13	686
山南光线电子商务有限公司	111236983	38	686
东阳东方传奇传媒有限公司	111236983	22	672

东阳光线影业有限公司	111236983	11	668
北京光线易视网络科技有限公司哈尔滨分公司	111236983	58	620
上海聚合广告有限公司北京办事处	111236983	58	610
上海光线电视传播有限公司浦东分公司	111236983	8	564
北京光线时代资讯有限公司	111236983	6	519
北京光线易视网络科技有限公司沈阳分公司	111236983	8	514

核心指数是衡量企业在族群中重要程度的关键指标。北京光线传媒股份有限公司核心指数为 1351，是族群的核心企业，全面布局影视、动漫、内容关联及产业投资等板块。北京光线影业有限公司(1308)、光线控股有限公司（1177）、霍尔果斯彩条屋影业有限公司（1175）以及霍尔果斯五光十色影业有限公司（1157）等企业，核心指数也位居前列，与族群 A 的关联程度更高。

节点的度反映了企业之间的连接数量。例如，霍尔果斯五光十色影业有限公司的节点度为 23，表明其在网络中与其他企业的连接较多，可能在业务协同、资源共享等方面更具优势。

4、个人族群关系

个人族群关系表揭示了企业背后的关键人物及其相互关系，可用于识别个人合作伙伴。与族群 A 相关联的个人，如表 4 所示，在光线传媒的生态中，王长田作为公司创始人和实际控制人，具有核心地位，其家族成员如王华也在公司股权结构中扮演重要角色。孙永焕作为五光十色影业的总裁，在影视制作和艺人经纪领域具有显著影响力。李晓萍作为非独立董事，参与公司治理和战略决策。

表 4 个人族群（A）关系-基于北京光线传媒股份有限公司

个人	族群 id	节点的度
王华	111236983	35
孙永焕	111236983	34
李慧	111236983	26
黎露	111236983	22
李晓萍	111236983	17
王鑫	111236983	12
罗霞	111236983	12
常索妮	111236983	12
其他	111236983	12
纪啸峥	111236983	11
王长田	111236983	9

通过对企业族群信息的深度分析，本数据库能够在复杂的企业关联网络中识别潜在的商业集团、企业族群及其核心成员，揭示关键的个人合作伙伴关系，在投资推荐、风险控制、商业合作以及情报分析等领域展现出重要的应用价值。

5、族群图谱

图 4 为集安市正发实业有限责任公司的族群图谱，呈现了该公司潜在族群的网络结构。如图所示，集安市正发实业有限责任公司存在若干潜在族群，这些族群主要通过集安市正发实业有限责任公司、集安市全域旅游服务有限公司以及刘云峰、刘正发等人相互关联。

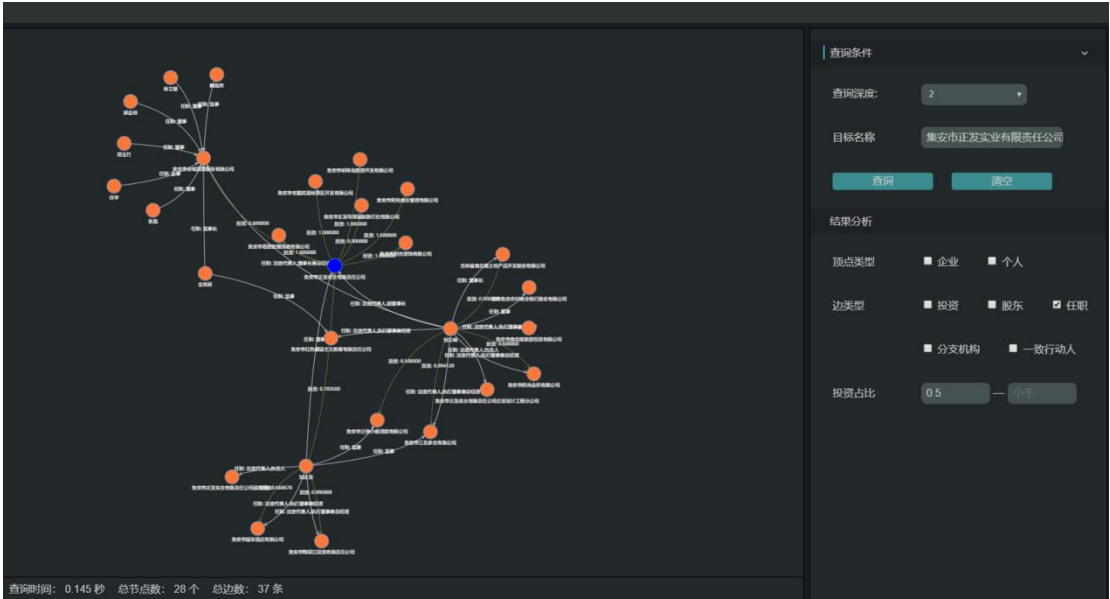


图 4 族群图谱实例

(三) 数据统计

企业族群划分所使用的数据包括企业基本信息、股权结构数据、高管及人员信息、投资关系数据、分支机构、担保与合作关系、联系方式关联以及外部数据等多维度数据,其节点和关系数据总量见表 5。这些数据来源包括企业内部数据库、工商信息平台,通过整合和分析这些数据,能够挖掘企业之间的深层关联关系,如实际控制人和所属集团等。

表 5 艾思企业族群数据库节点和关系数据总量

维度	内容	总量
企业与企业关系	投资	1100 万+
	历史投资	490 万+
个人与企业关系	法定代表人	9600 万+
	投资	1.5 亿+
	任职	2.19 亿+

	历史投资	3100 万+
	历史任职	4300 万+
节点数量	企业	1.13 亿+
	个人	3.25 亿+

五、数据优势

（一）高精度权重建模

区别于传统的简单处理方式，本数据库对股权、职位等企业关系关键参数进行量化分析。依据股权占比判定投资方话语权，按职位高低区分任职影响力，高股权占比和关键职位赋予更高权重。通过引入关系权重，精确呈现企业间强弱关系，有效提升族群划分准确性，为企业战略决策提供关键依据，助力企业明确自身在商业生态中的定位。

（二）多维度信息保留

传统分析将节点局限于单一族群，致使大量信息遗漏。本数据库允许节点归属多个族群，并留存多族群置信概率。以科技企业为例，其在技术研发和硬件生产领域分属不同族群。通过保存各族群置信概率，可清晰展现企业在不同关系中的参与程度，全方位呈现复杂商业网络，帮助企业精准把握全局，做出明智决策。

（三）动态稳定性保障

分析时随机遍历顺序会干扰结果，本数据库利用迭代收敛机制解决此问题。每次迭代，算法综合考量节点关联强度、周边节点归属和关系权重等因素，动态调整节点族群及置信概率。多次迭代后，结果

趋于稳定，无论初始顺序如何，都能得出可靠结论，为企业决策提供坚实的数据支撑。

（四）灵活扩展性

商业环境多变，新关系类型不断涌现。本数据库可轻松兼容供应链合作、专利关联等新关系。以供应链合作关系为例，依据合作深度等因素赋予权重，分析企业在产业链的位置和协同关系。对专利关联关系也同理，这使数据库能紧跟商业发展，为企业提供实用的商业洞察，保持竞争优势。

（五）业务导向输出

本数据库以服务业务为导向，提供置信概率排名和关键路径分析等功能。置信概率排名帮助企业快速筛选潜在合作伙伴，关键路径分析帮助企业识别商业网络关键节点，制定针对性策略。相似度分析还能提升族群划分准确性，在跨行业合作和投资推荐中发挥重要作用，助力企业把握机遇，高效决策。