

一. 构建半导体产业知识图谱需要以下技术方案：

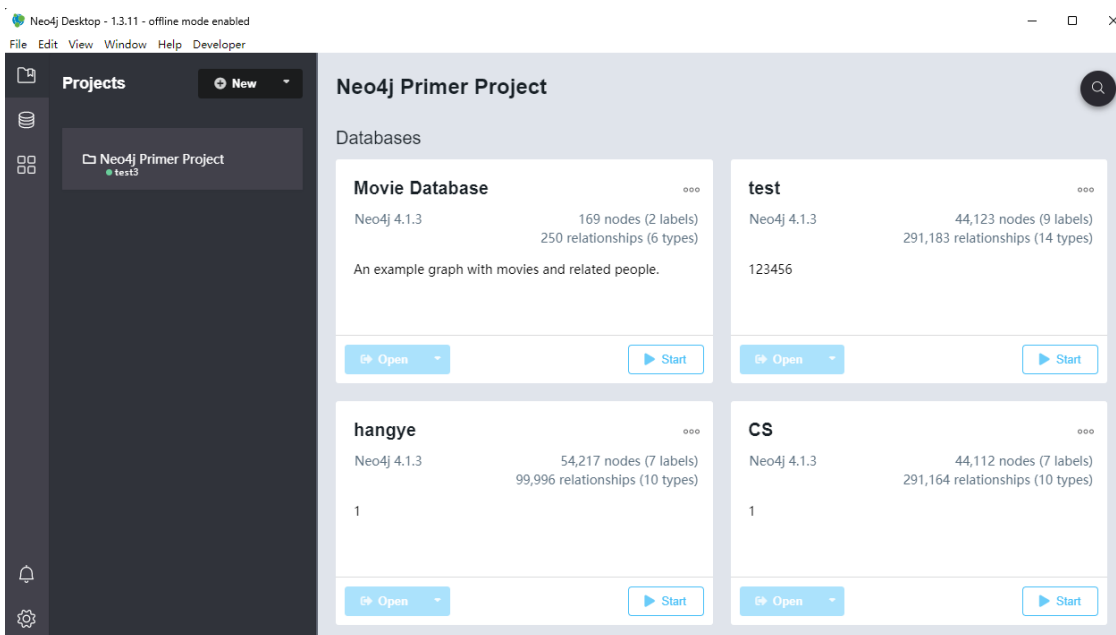
1. 语料库的构建：需要收集半导体产业相关的文本数据作为语料库，可以从互联网上的新闻、论文、专利等渠道进行收集。收集到的文本数据需要进行清洗和预处理，包括去除停用词、分词、词性标注等。
2. 实体识别：使用自然语言处理技术对语料库中的文本数据进行实体识别，识别出半导体产业相关的实体，如公司、人物、技术等。
3. 关系抽取：使用自然语言处理技术对语料库中的文本数据进行关系抽取，抽取出实体之间的关系，如公司之间的合作关系、人物与技术之间的关系等。
4. 数据存储：使用图数据库 Neo4j 存储实体和关系数据，Neo4j 是一种基于图的数据库，可以很好地存储实体和关系数据，并支持复杂的查询操作。
5. 数据可视化：使用 Python 编写程序，将从 Neo4j 中查询到的数据进行可视化展示，本项目借助了 GraphXR 这个开源可视化工具。

## 二. 详细过程

1. 整理半导体及相关产业的数据，包括上游企业、行业、产品及各产品小类的上游材料。将所有数据整合为 .json 形式，便于后续读取和使用。

```
1 [{"company_code": "300724.SZ", "product_name": "半导体掺杂沉积光伏设备", "rel_weight": 0.561175, "rel": "主营产品", "company_name": "捷佳伟创"}]
2 [{"company_code": "300037.SZ", "product_name": "半导体化学品", "rel_weight": 0.004953, "rel": "主营产品", "company_name": "新宙邦"}]
3 [{"company_code": "300429.SZ", "product_name": "半导体光刻胶光引发剂", "rel_weight": 0.022298, "rel": "主营产品", "company_name": "强力新材"}]
4 [{"company_code": "300429.SZ", "product_name": "半导体专用化学品", "rel_weight": 0.014181, "rel": "主营产品", "company_name": "强力新材"}]
5 [{"company_code": "600584.SH", "product_name": "半导体芯片凸块", "rel_weight": 0.040381, "rel": "主营产品", "company_name": "长电科技"}]
6 [{"company_code": "300690.SH", "product_name": "半导体设备", "rel_weight": 0.082785, "rel": "主营产品", "company_name": "至纯科技"}]
7 [{"company_code": "300408.SZ", "product_name": "半导体部件", "rel_weight": 0.169782, "rel": "主营产品", "company_name": "三环集团"}]
8 [{"company_code": "600520.SH", "product_name": "半导体封装模具", "rel_weight": 0.074211, "rel": "主营产品", "company_name": "文一科技"}]
9 [{"company_code": "600667.SH", "product_name": "半导体分部", "rel_weight": 0.420196, "rel": "主营产品", "company_name": "太极实业"}]
10 [{"company_code": "002409.SZ", "product_name": "半导体化学材料", "rel_weight": 0.176973, "rel": "主营产品", "company_name": "雅克科技"}]
11 [{"company_code": "002409.SZ", "product_name": "半导体前驱体材料", "rel_weight": 0.074646, "rel": "主营产品", "company_name": "雅克科技"}]
12 [{"company_code": "300373.SZ", "product_name": "半导体器件", "rel_weight": 0.869048, "rel": "主营产品", "company_name": "扬杰科技"}]
13 [{"company_code": "300373.SZ", "product_name": "半导体芯片", "rel_weight": 0.125734, "rel": "主营产品", "company_name": "扬杰科技"}]
14 [{"company_code": "300373.SZ", "product_name": "半导体硅片", "rel_weight": 0.056439, "rel": "主营产品", "company_name": "扬杰科技"}]
15 [{"company_code": "300656.SZ", "product_name": "功率半导体", "rel_weight": 0.066335, "rel": "主营产品", "company_name": "民德电子"}]
16 [{"company_code": "300757.SZ", "product_name": "半导体自动化设备", "rel_weight": 0.007348, "rel": "主营产品", "company_name": "罗博特科"}]
17 [{"company_code": "300757.SZ", "product_name": "半导体生产过程智能化设备", "rel_weight": 0.005088, "rel": "主营产品", "company_name": "罗博特科"}]
18 [{"company_code": "002185.SZ", "product_name": "半导体器件封装测试", "rel_weight": 0.979172, "rel": "主营产品", "company_name": "华天科技"}]
19 [{"company_code": "600363.SH", "product_name": "半导体激光系列", "rel_weight": 0.048109, "rel": "主营产品", "company_name": "联创光电"}]
20 [{"company_code": "300480.SZ", "product_name": "半导体封测装备类", "rel_weight": 0.175738, "rel": "主营产品", "company_name": "光力科技"}]
21 [{"company_code": "300480.SZ", "product_name": "半导体精密加工", "rel_weight": 0.001621, "rel": "主营产品", "company_name": "光力科技"}]
22 [{"company_code": "300123.SZ", "product_name": "半导体元器件", "rel_weight": 0.067943, "rel": "主营产品", "company_name": "亚光科技"}]
23 [{"company_code": "300123.SZ", "product_name": "半导体器件", "rel_weight": 0.023797, "rel": "主营产品", "company_name": "亚光科技"}]
24 [{"company_code": "300567.SZ", "product_name": "半导体", "rel_weight": 0.031147, "rel": "主营产品", "company_name": "精测电子"}]
```

## 2. 创建一个图数据库用于存储数据



### 3. 借助 Python 程序调用图数据库接口 API 进行图谱创建

#### 3.1 读取数据

```
def __init__(self):
    cur_dir = '/' + os.path.abspath(__file__).split('/')[-1]
    # 实体列表
    # company是上市公司，是我国重要的公司代表与行业标杆，选取上市公司作为基础实体
    # 从中选取与半导体及集成电路行业有关的上市公司
    # company包括企业名name, 企业全称fullname, 企业编号code, 上市地点及时间location, time
    self.company_path = os.path.join(cur_dir, 'data/company.json')

    # industry是行业分类，是承载企业、公司、产品的媒介
    # --行业指数和热点行业等指标
    # 对上市公司进行了行业归属
    # 选取申万发布了2021版的行业分类规范
    self.industry_path = os.path.join(cur_dir, 'data/industry.json')

    # 产品，代表公司的主营范围，用于定位公司性质
    # 其数据可以从公司的经营范围、年报等文本中进行提取得到。
    self.product_path = os.path.join(cur_dir, 'data/product.json')

    # 关系列表
    # 公司所属的行业--通过公开的上市公司行业分类表，可以得到上市公司所对应的行业分类数据。
    self.company_industry_path = os.path.join(cur_dir, 'data/company_industry.json')

    # 公司主营产品关系
    # 上市公司经营产品数据可以从两个方面来获得，
    # 一个是从公司简介中的经营范围中结合制定的规则进行提取，
    # 另一个是从公司每年发布的半年报、年报中进行提取。
```

#### 3.2 创建节点

#### 3.3 加载数据

#### 3.4 创建实体

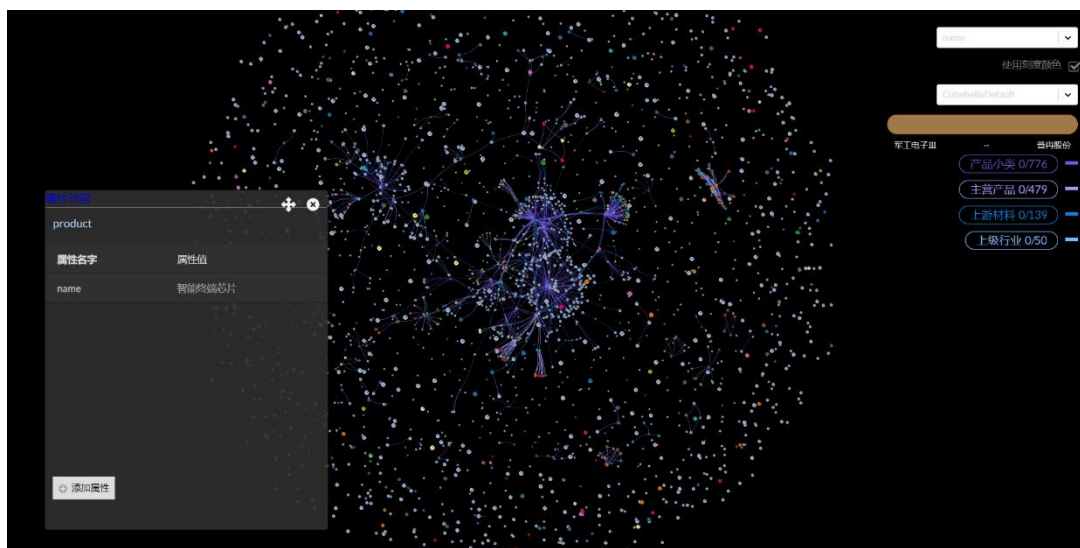
#### 3.5 创建实体关系边

#### 3.6 创建实体关联变

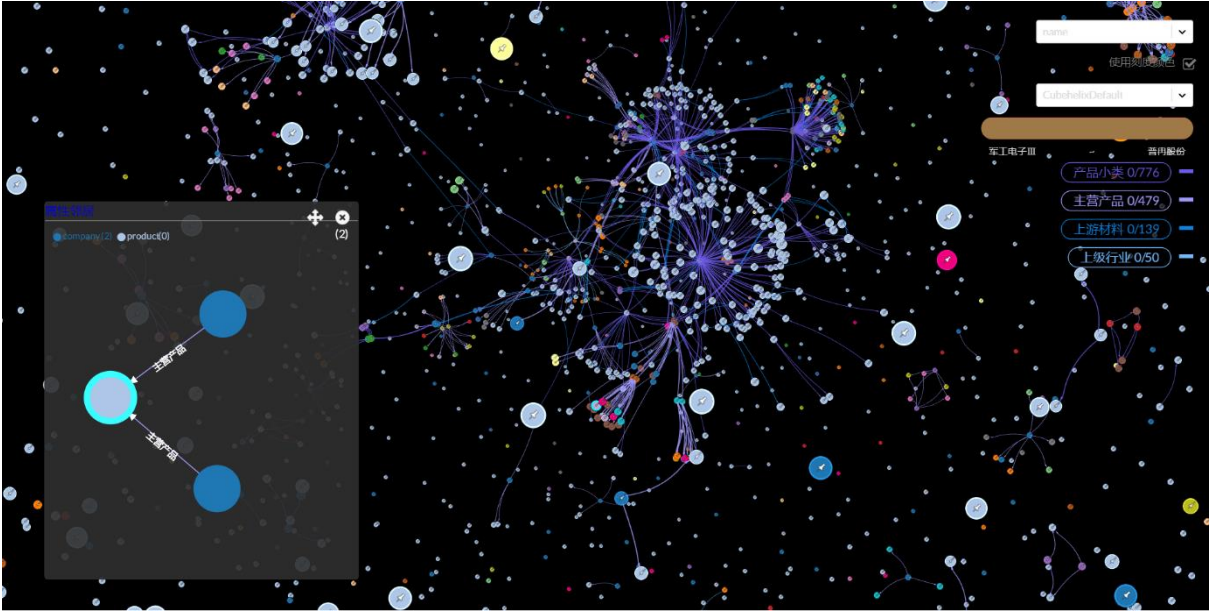
### 4. 借助 GraphXR 可视化工具进行可视化和结果分析

#### 4.1 整体全景图：

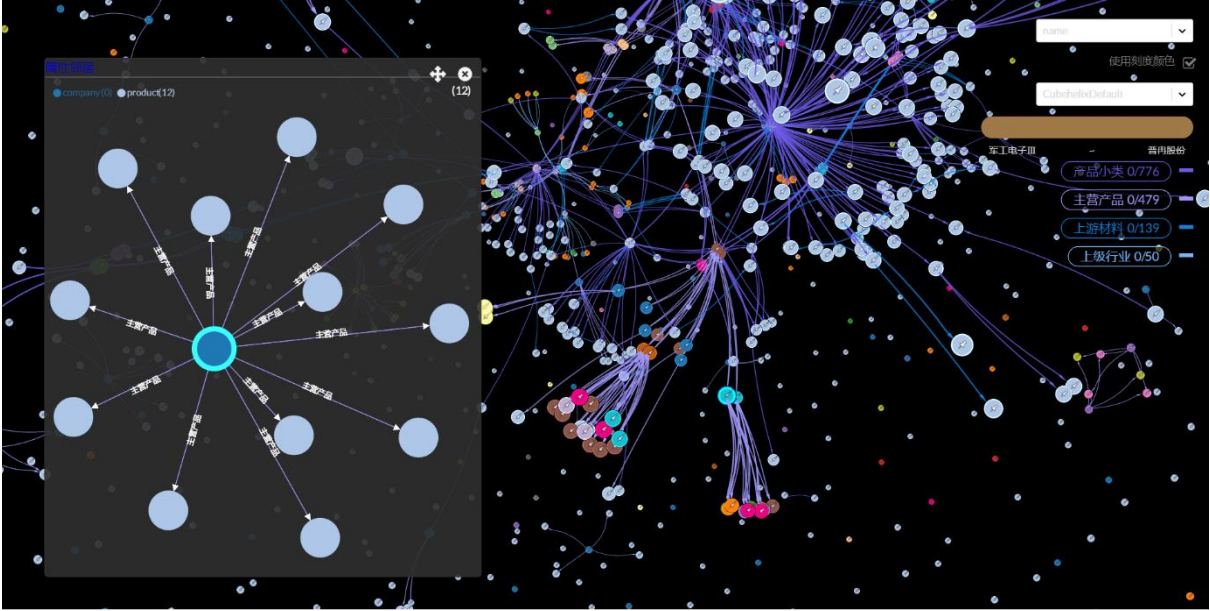
其中包含 2422 个产品实体、237 个企业实体、42 个行业实体。776 个产品小类、479 个主营产品、139 个上层材料、50 个上级行业



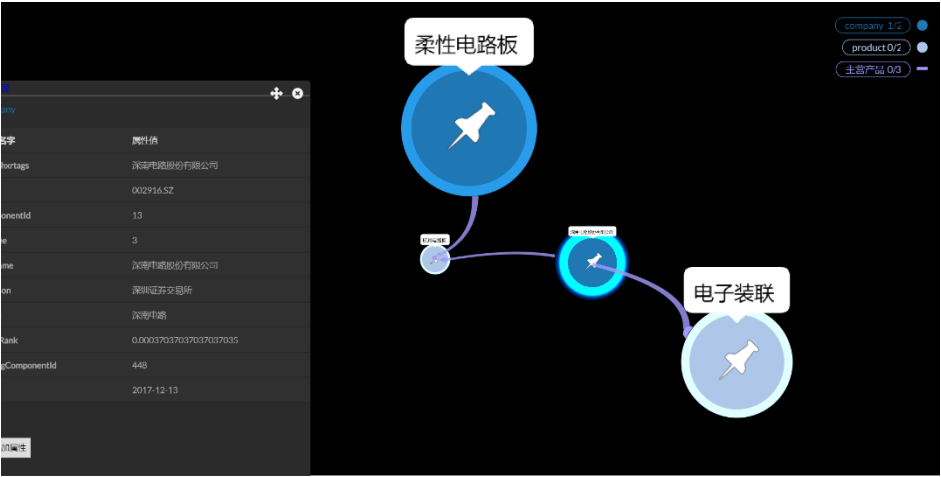
4.2 企业竞争关系



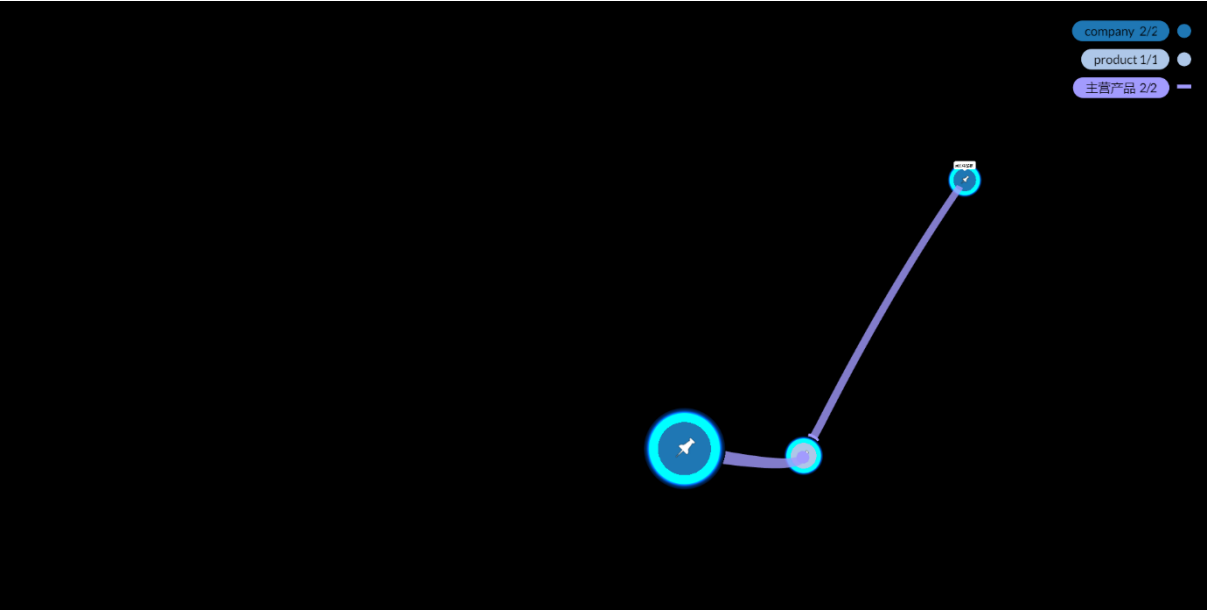
4.3 同一企业多种主营产品



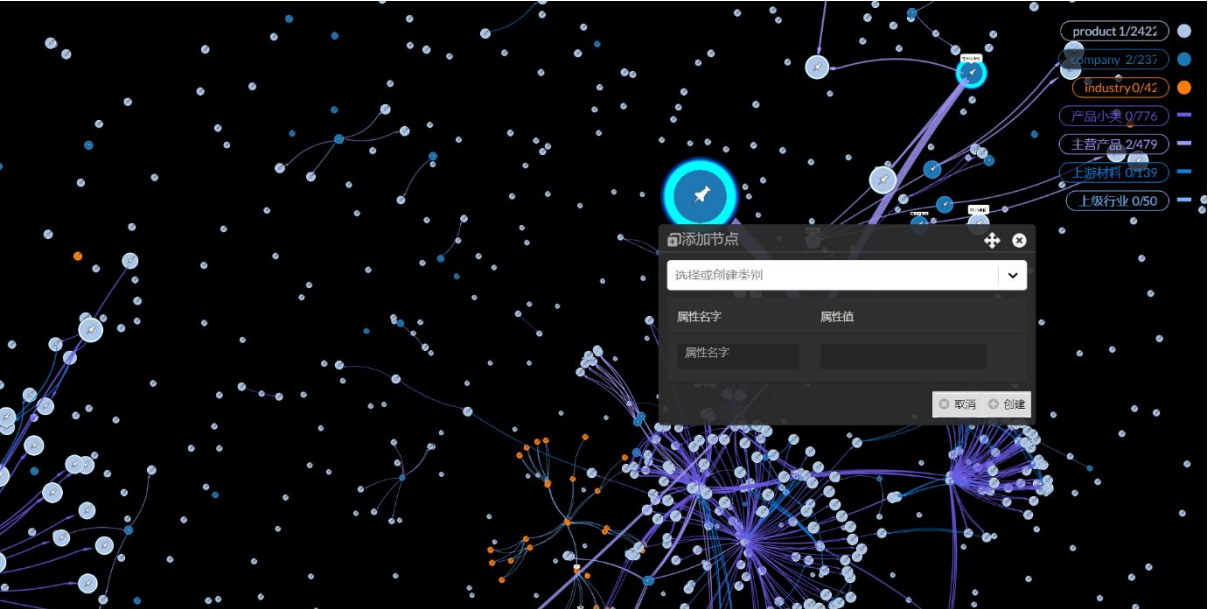
4.4 查询产品之间的关系



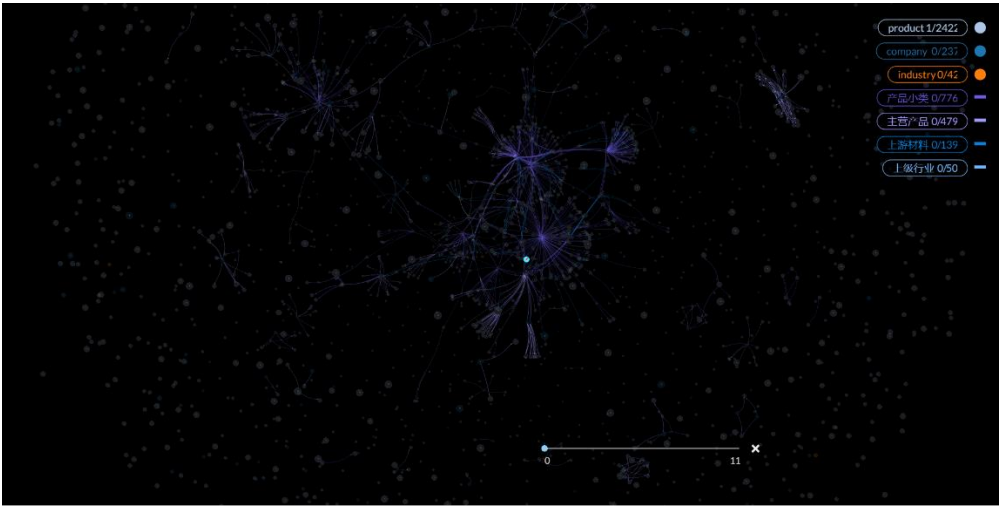
4.5 查询企业之间的关系



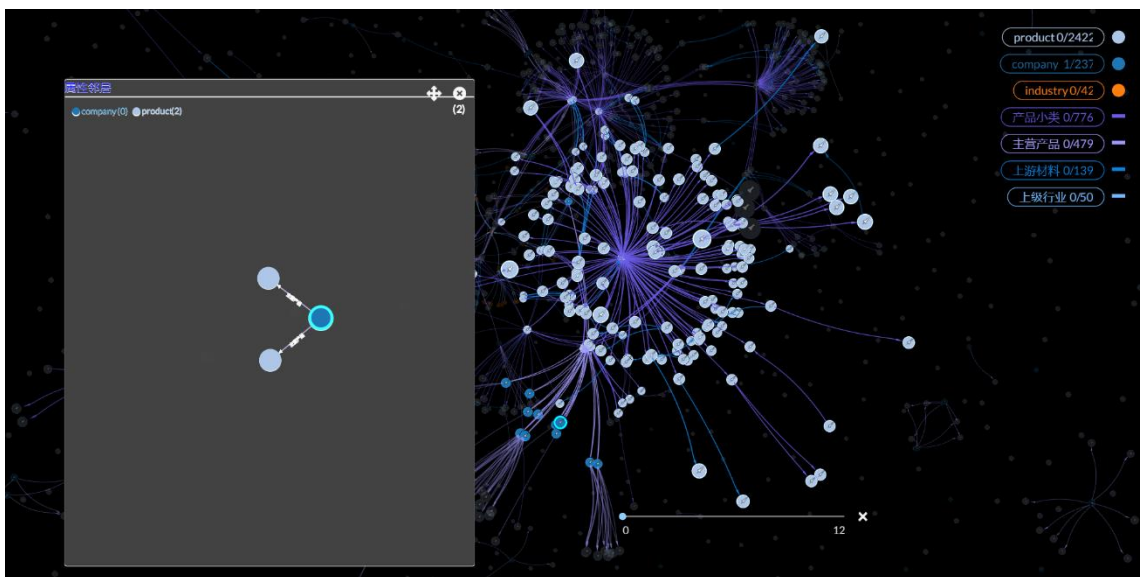
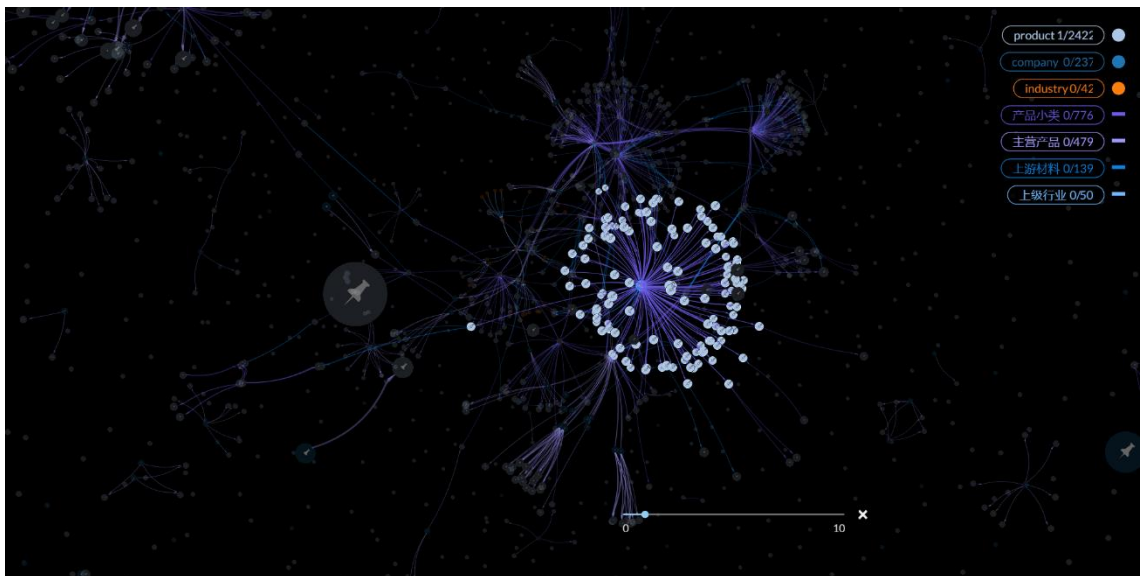
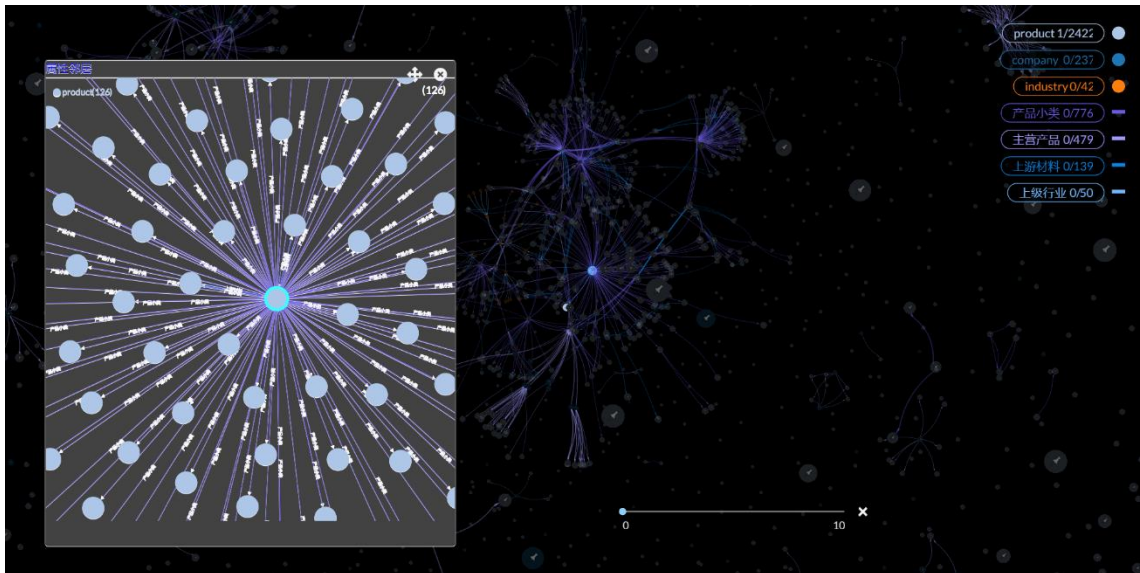
4.6 添加删除节点

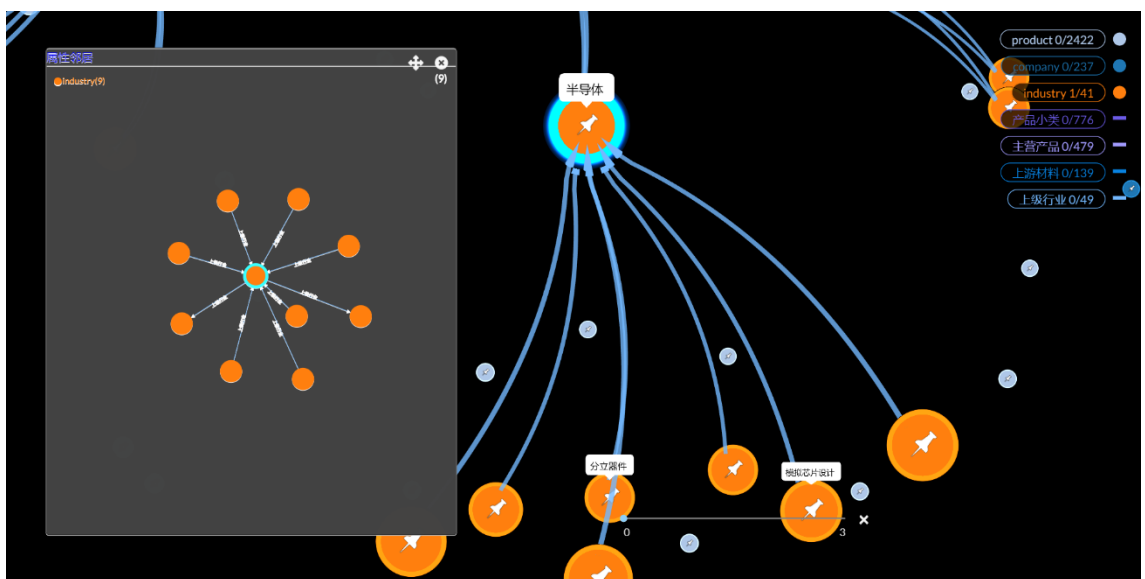
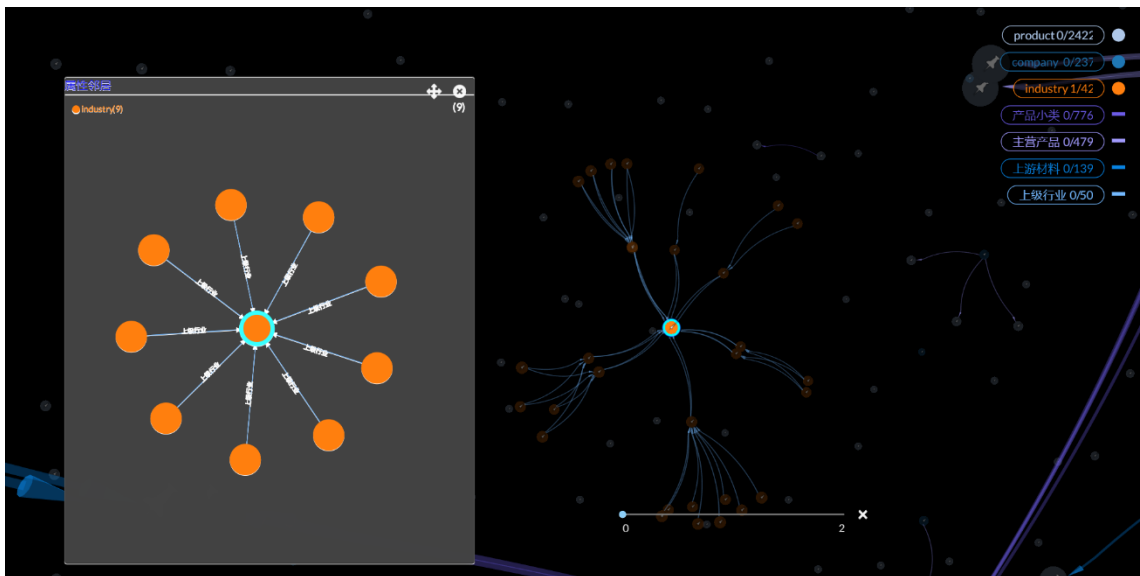
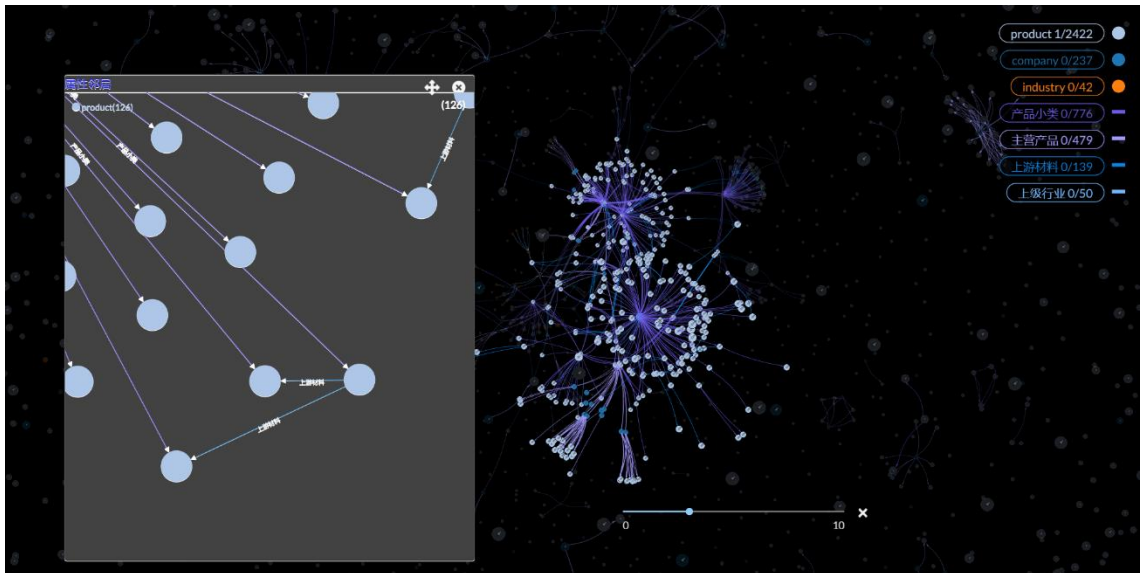


4.7 从一个产品辐射到整个半导体产业链的过程









### 三. 项目总结

为了提高我国产业链风险全过程管理水平，我们需要依靠知识图谱等技术来开发制造业产业链全景信息系统，实现对制造业产业链的完整描述和动态表达。这将支持产业链风险的动态监控、精准识别、高效评估和提前预测，为产业链的稳定运行和强链补链提供支持，提高我国产业链的韧性和国际竞争力。我们使用 Python 和 Neo4j 图数据库构建了一个半导体行业全景图，旨在帮助用户更好地了解这个行业的生态系统。

这个半导体行业全景图项目可以从不同的数据源中提取数据，并将其存储在 Neo4j 图数据库中。然后，我们使用 Neo4j 的可视化工具呈现这些数据，以帮助用户更好地理解半导体行业的生态系统。用户可以看到半导体行业中的各种组织、公司和技术，以及它们之间的关系。此外，用户还可以看到半导体行业中的各种市场趋势和技术发展，以帮助他们更好地了解行业的未来发展趋势。我们相信，这个半导体行业全景图项目可以为半导体行业的从业者、投资者和研究人员提供有价值的信息和洞察力，帮助他们更好地了解这个复杂而庞大的行业。

在这个全景图中，用户可以看到半导体行业中的各种公司和组织，以及它们之间的关系。例如，用户可以看到哪些公司是半导体生产商，哪些公司是芯片设计公司，哪些公司是设备供应商，等等。用户还可以看到不同公司之间的合作关系，例如哪些公司在联合开发新产品，哪些公司在共同投资研发，等等。

此外，全景图还包括了半导体行业中的各种技术和市场趋势。例如，用户可以看到哪些技术正在成为主流，哪些技术正在被淘汰，以及哪些市场正在快速增长。这些信息可以帮助用户更好地了解半导体行业的未来发展趋势，以及哪些公司和技术可能会成为行业的领导者。这个半导体行业全景图提供了一个全面的视角，帮助用户更好地了解这个复杂而庞大的行业。

### 四. 功能简介

这个项目的主要功能是构建一个半导体行业全景图，以帮助用户更好地了解这个行业的生态系统。具体来说，这个项目可以实现以下功能：

1. 数据提取：使用 Python 编写脚本，从不同的数据源中提取半导体行业相关的数据，例如公司和组织信息、技术和市场趋势等。
2. 数据存储：将提取的数据存储在 Neo4j 图数据库中，以便后续的分析 and 可视化。
3. 数据分析：使用 Neo4j 的查询语言 Cypher 对数据进行分析，例如查找公司之间的关系、查找市场趋势等。
4. 数据可视化：使用 Neo4j 的可视化工具呈现数据，以使用户更好地理解半导体行业的生态系统。用户可以查看公司之间的关系、市场趋势、技术发展等信息。
5. 用户交互：用户可以通过可视化工具与全景图进行交互，例如缩放、选择节点、查看详细信息等。

简单来说，这个项目的主要功能是帮助用户更好地了解半导体行业的生态系统，通过数据提取、存储、分析和可视化，用户可以深入了解行业中的各种组织、公司和技术，以及它们之间的关系。这个项目可以为半导体行业的从业者、投资者和研究人员提供有价值的信息和洞察力。