

# 黑灰产网络资产可视分析系统

孙德晟 高琳 刘子奥 月小琪 周艺璇 胡海波（指导老师）

重庆大学

# CONTENT

任务分析

数据处理

可视化与  
交互设计

案例分析

总结与展望

## 可分析任务

### 1. 赛题解析

#### ① 网络资产类型划分：

- **外围网络资产**：向网民直接公开的黑灰产业务网站域名（Domain节点）
- **核心网络资产**：关系到众多外围网络资产运行的网络资产（IP、Cert节点等）
- **普通网络资产**：介于其它两类网络资产之间

#### ② 节点及边分析

- 分析重点：**IP、Cert、Domain**节点和关联强度“**很强**”的边
- Whois节点及边：存在域名贩子，无法作为重要线索进行分析，辅助验证子图挖掘结果

任务分析

数据处理

可视化与  
交互设计

案例分析

总结与展望

## 可<sup>视</sup>分析任务

### 2. 网络资产子图挖掘

#### ① 社区传播算法：

- 难以考虑**节点类型和产业类型**的影响
- 对于**边的强度约束不够**，不能保证满足子图挖掘**业务规则**，如节点a91593a45b  
( Cert )

## 可分析任务

### 2. 网络资产子图挖掘

#### ① 社区传播算法：

- 难以考虑节点类型和产业类型的影响
- 对于边的强度约束不够，不能保证满足子图挖掘业务规则

#### ② 根据网络资产类型进行挖掘

- 以潜在的核心网络资产（IP、Cert节点）为基础进行挖掘
- 根据赛题给出的业务规则，挖掘IP、Cert节点四跳内的子图，并简化为IP-Cert链路
- 根据IP-Cert链路构建核心网络资产之间的关系图，以此为基础进行子图挖掘
- 所有的节点均在IP-Cert链路中

## 可分析任务

### 3. 团伙特征分析

#### ① 核心网络资产和关键链路识别

- 根据赛题中对应业务规则筛选符合条件的节点
- 设计相应算法识别核心网络资产
- 根据**边的强度和路径长度**，设计对应算法识别关键链路

#### ② 运作机制分析

- 统计团伙主要信息
- 以识别的**核心网络资产**和**关键链路**为基础进行分析
- 根据黑灰产类型和边类型识别主要**运营方式**和**子图特征**

# 黑灰产网络资产图谱数据集

## 1. 数据预处理

### ① 数据索引

- 将Node.csv中每一个节点添加一个唯一标识的数字，并修改Link.csv的对应字段

### ② 数据修正

- 修正错误的industry字段
- 将industry字段从[“A”，“B”]格式规范为“AB”

### ③ 数据清洗

- 检查边的一致性，并删除了41条不符合题目描述的边

```
[  
    'r_whois_name',  
    'Whois_Phone_d09d0994cef3553708537f9e83b1cb339347fb529a557d0be0ff6a7961bb561d',  
    'Domain_838cc7afce19af1d89d296d0dee22d23509a7f74723caf2472f5403fe7c64774'  
]
```

任务分析

数据处理

可视化与  
交互设计

案例分析

总结与展望

## 黑灰产网络资产图谱数据集

### 2. IP-Cert链路挖掘

#### ① 对每一个IP、Cert节点，遍历其四跳及以内的所有节点

- 遇到关联强度不是“很强”的边，停止遍历
- 遇到另一个IP或Cert节点，停止遍历

#### ② 统计每一个IP、Cert节点四跳内节点的相关信息，将IP、Cert节点分类。

- ①型和②型节点为潜在的核心网络资产节点

路径节点	含有IP/Cert	不含IP/Cert节点
涉及黑灰产业	①型， 38010个	②型， 24381个
不涉及黑灰产业	③型， 127589个	④型， 48991个

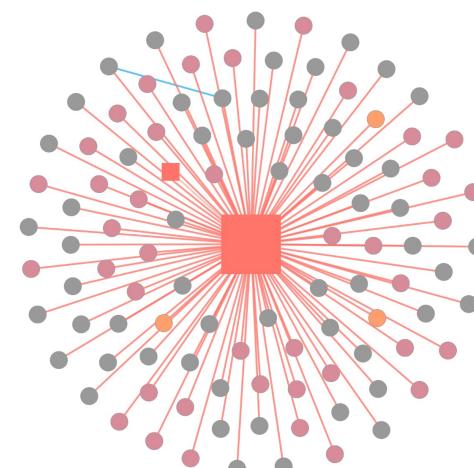
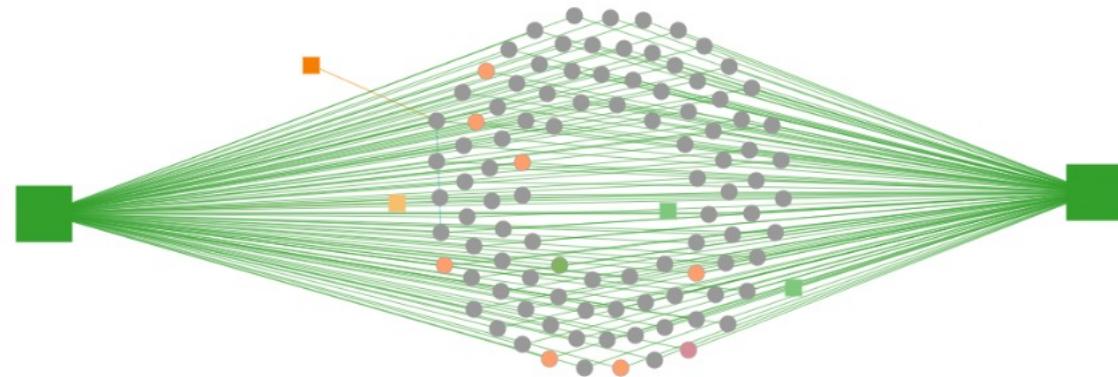
#### ③ 构成IP-Cert链路

- 由①型节点构成，起点和终点均为IP或Cert链路，统称IP-Cert链路，共计369402条
- ②型节点构成的链路成为IP-X链路或Cert-X链路，共计24381条

## 黑灰产网络资产图谱数据集

### 3. IP-Cert链路信息统计

- ① 计算每一个IP-Cert链路包含的节点和边
- ② 统计每一个IP-Cert链路信息
  - 包含的节点类型及数量
  - 包含的黑灰产类型及数量
- ③ 统计每一个节点所在的IP-Cert链路
- ④ 每一个IP、Cert节点直接相邻的Domain节点的信息
  - 节点数量
  - 涉及的黑灰产类型与数量
  - 黑灰产节点数量占比等



# 黑灰产网络资产图谱数据集

## 4. 核心资产识别

① 计算子图中各节点**介数中心性** $BC(N_i)$ :

$$BC(N_i) = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}}$$

- $n_{st}^i$  : 经过节点*i*的最短路径的边数量
- $g_{st}$ : 连接*s*和*t*的最短路径的数量

② 对节点集合按**介数中心性降序**排列。筛选满足  **$h_1$ -index** 条件的节点作为候选核心资产节点

$$BC(N_i) > (h_1 - 1) * s_1$$

- $h_1$  : 节点*N<sub>i</sub>*根据随机游走介中心性降序后的索引数
- $s_1$  : 放缩系数。

③ 计算候选核心资产**节点度中心性** $DC(N_i)$  :

$$DC(N_i) = \frac{K_i}{n - 1}$$

- $K_i$ : 该节点相连的边的数量
- $n - 1$  : 节点*N<sub>i</sub>*与其他节点都相连的边的数量

④ 对节点集合按照度**中心性降序**排列，满足以下  **$h_2$ -index** 筛选条件的作为候选核心资产节点:

$$DC(N_i) > (h_2 - 1) * s_2$$

- $h_2$  : 节点*N<sub>i</sub>*根据度中心性降序后的索引数
- $s_2$  : 放缩系数。

⑤ 筛选候选核心资产节点集合中类型为IP或Cert的节点，作为该子图**核心资产节点**

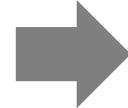
## 黑灰产网络资产图谱数据集

### 5. 关键链路识别

- ① IP-Cert链路中包含的边均为**关键链路**
- ② **关联链路重要性：边的强度、路径长度**
  - 给定两个核心资产节点，分别作为关键链路的start与end
  - 使用**深度优先遍历算法**，获取start与end节点之间的所有最短路
    - 两个核心网络资产间存在多条路径时，路径越短越重要
  - 两个核心资产间的**最短路**作为关键链路
  - 如果最短路超过四跳，则表明这两个核心资产之间没有关键路径

## 黑灰产网络资产图谱数据集

1. 数据预处理
2. IP-Cert链路挖掘
3. IP-Cert链路信息统计
4. 核心资产识别
5. 关键链路识别



### 1. 网络资产子图挖掘

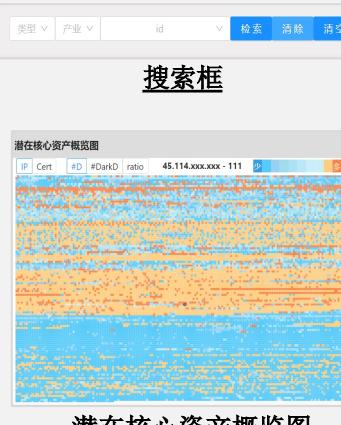
搜索框、潜在核心资产概览图、IP | Cert  
跳转图、IP-Cert链路图、黑灰产业对比  
图、黑灰产网络资产子图

### 2. 团伙特征分析

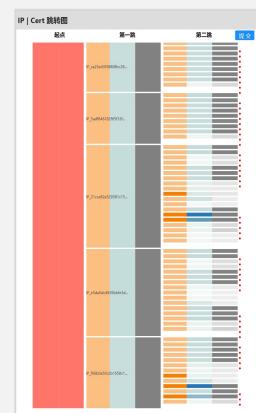
团伙基本信息、团伙网络信息、核心资产  
图、关键链路图、团伙分析结果

## 可视分析方案

### 发现线索

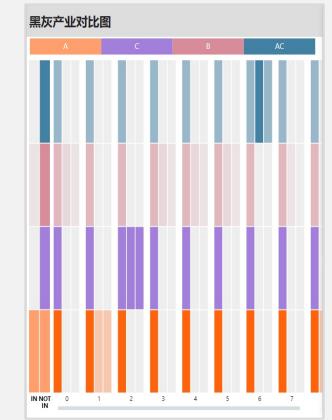


### 明确主要IP-Cert链路



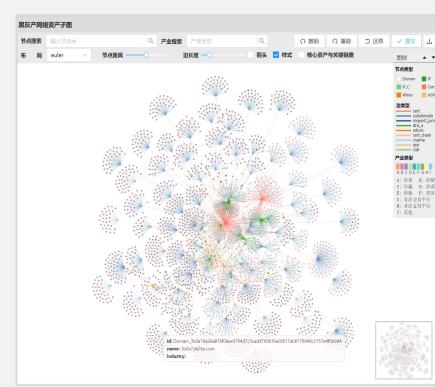
IP-Cert链路图

### 调整子图



黑灰产业对比图

### 确定子图



黑灰产网络资产子图

## 动态挖掘子图



## 分析团伙特征

### 统计分析



团伙基本信息



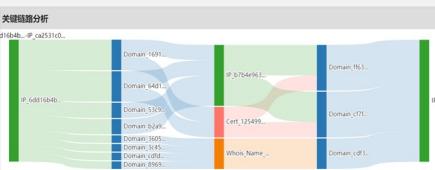
团伙分析结果

### 核心资产分析



核心资产图

### 关键链路分析



关键链路图

### 任务分析

### 数据处理

## 可视化与交互设计

### 案例分析

### 总结与展望

标题

黑灰产网络资产子图控制栏

团伙基本信息

搜索框

潜在核心资产  
概览图

黑灰产网络资产子图

团伙网络信息

筛选IC

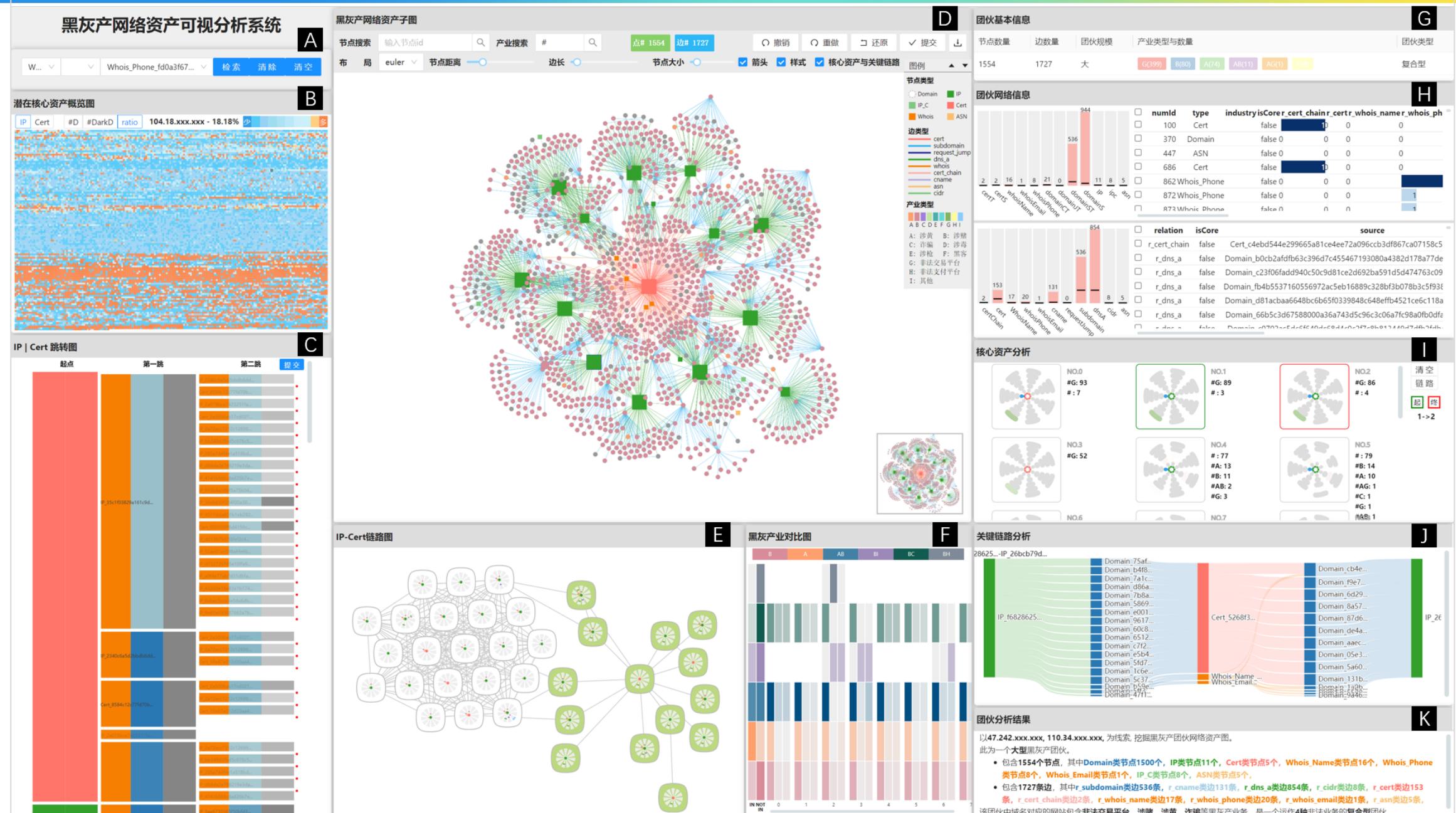
IP-Cert链路图

黑灰产业  
对比图

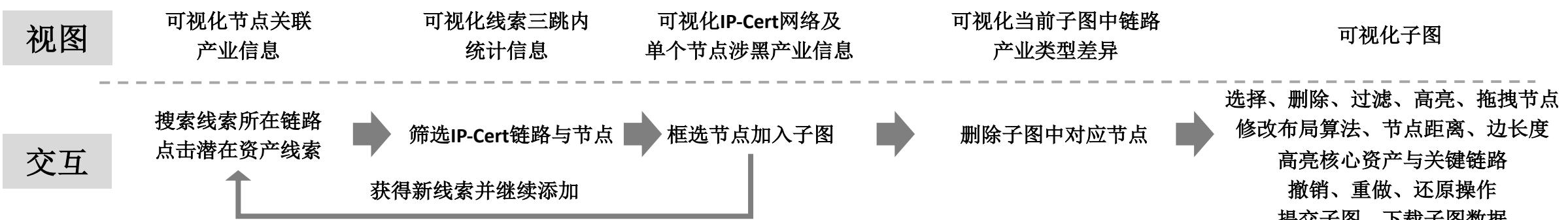
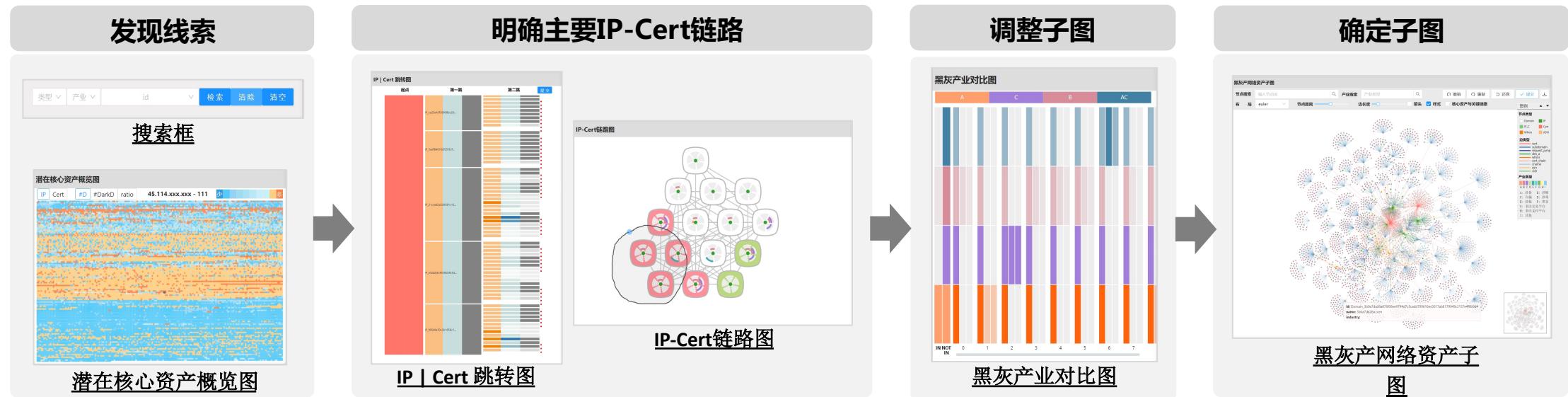
核心资产分析

关键链路分析

团伙分析结果



# 动态子图挖掘方法



## 挑战1.1-团伙2（中型团伙）：

### ① 以题目给的两个节点为线索，挖掘其所在的IP-Cert链路

- IP节点156.241.xxx.xxx，挖掘其所在的IP-Cert链路



任务分析

数据处理

可视化与  
交互设计

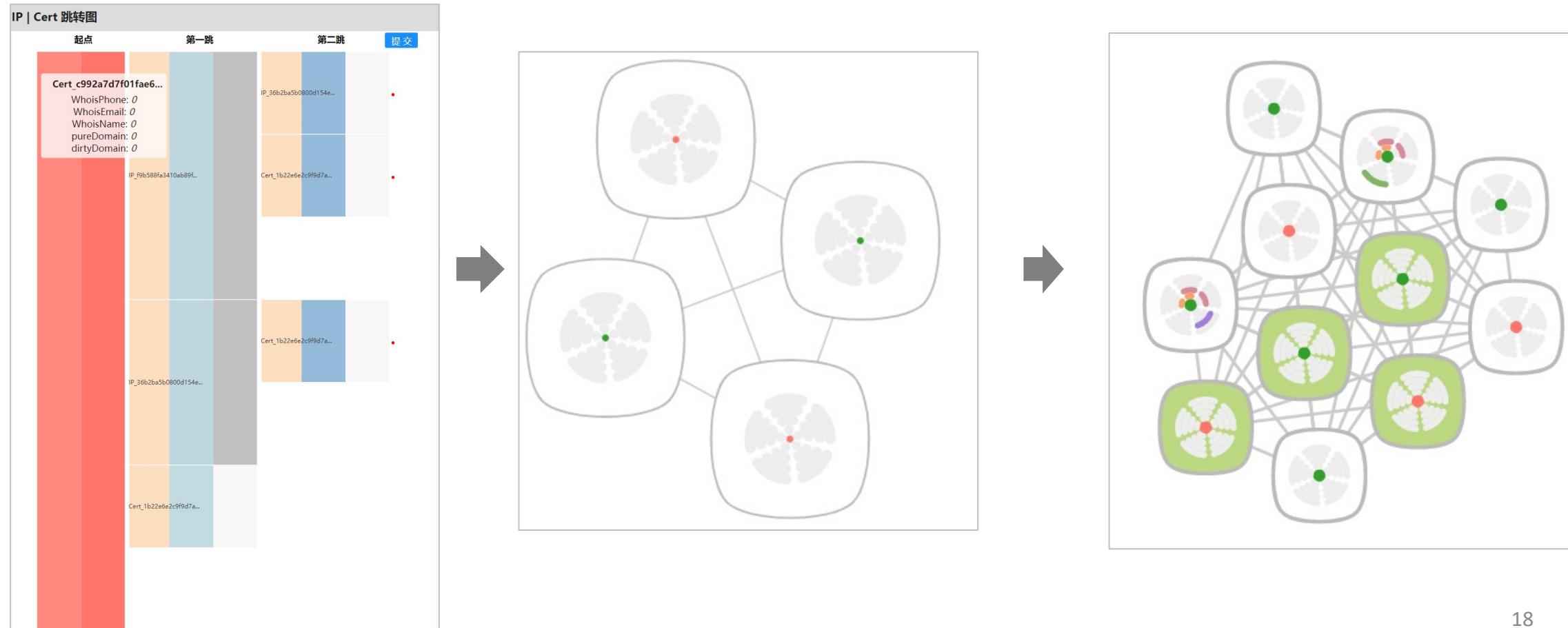
案例分析

总结与展望

## 挑战1.1-团伙2（中型团伙）：

### ① 以题目给的两个节点为线索，挖掘其所在的IP-Cert链路

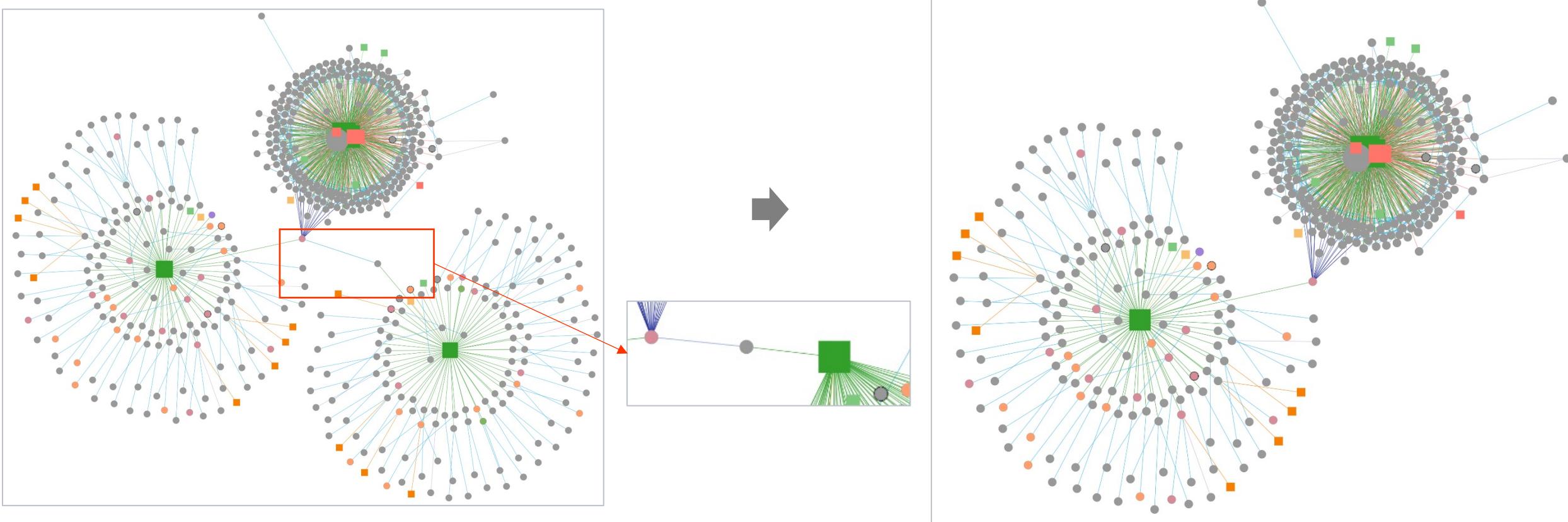
- Domain节点b10f98a9b5.com，挖掘其所在的IP-Cert链路



## 挑战1.1-团伙2（中型团伙）：

### ② 在主图中对链路进行修改边

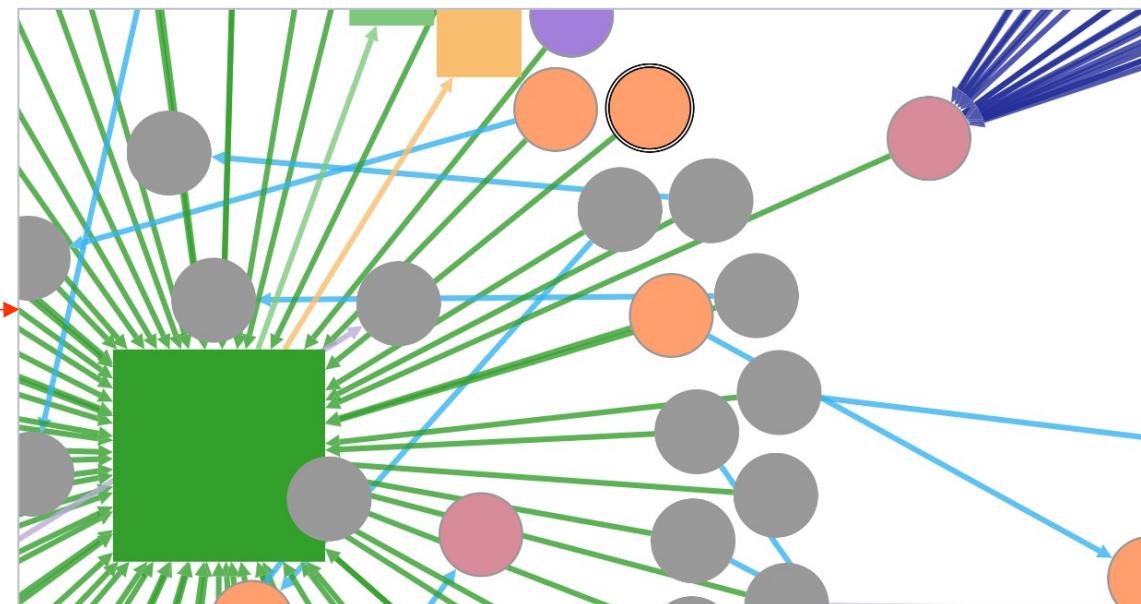
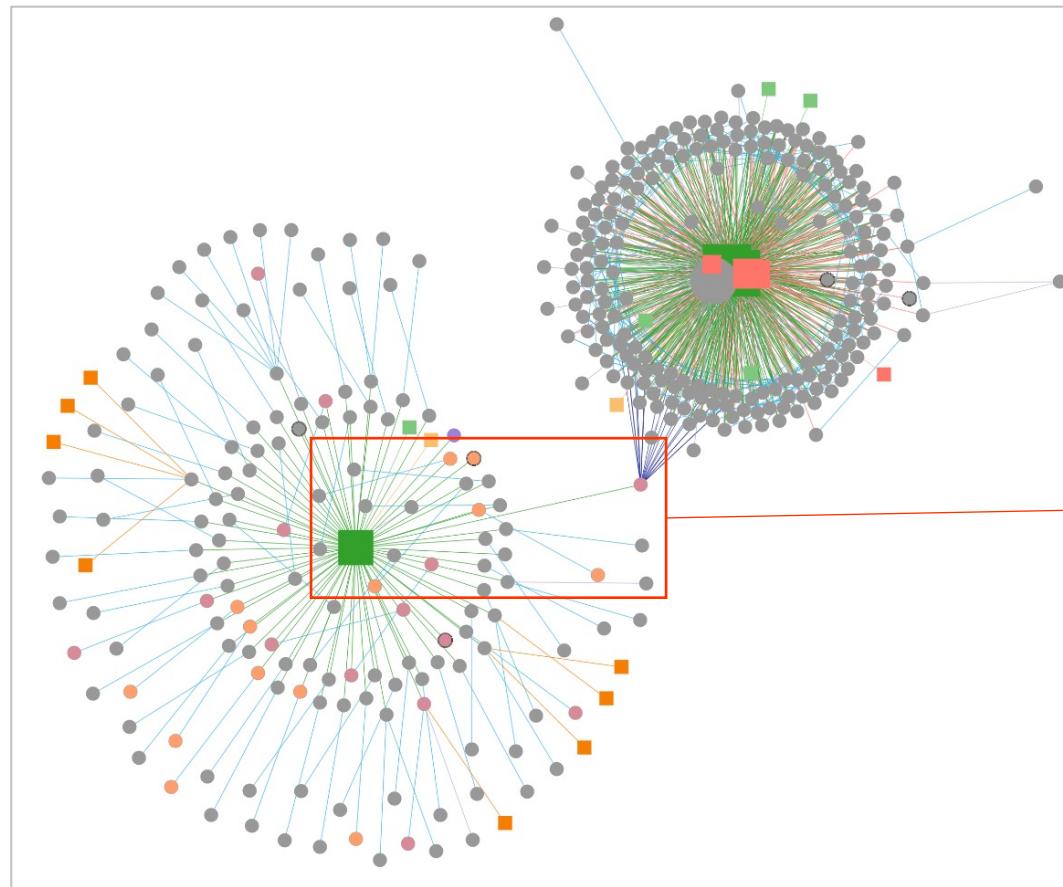
- 边强度较低，且只有一个节点进行关联，故进行删除



## 挑战1.1-团伙2（中型团伙）：

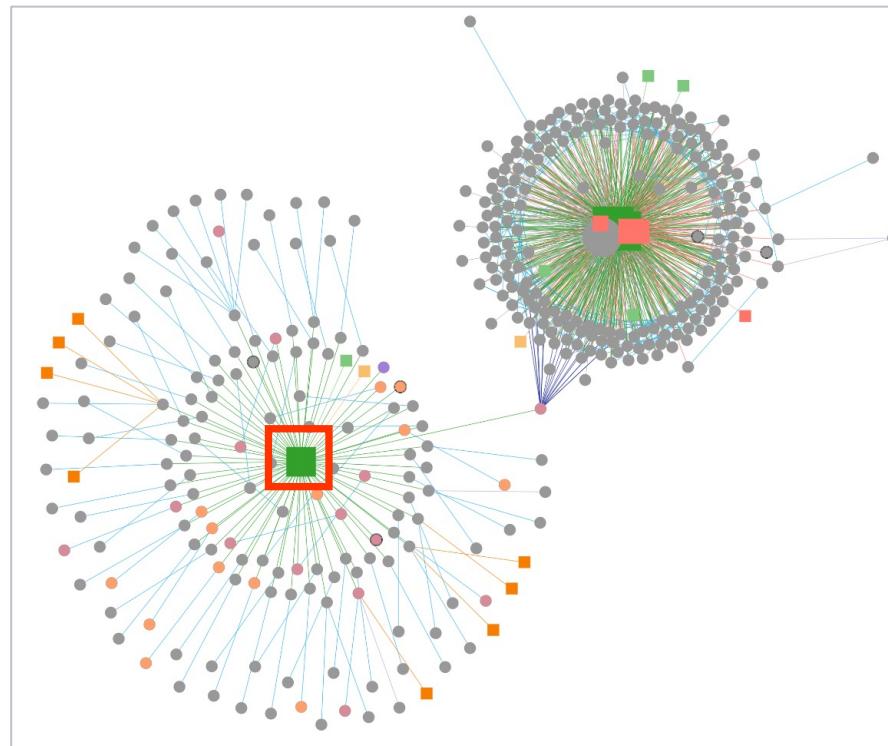
### ② 在主图中对链路进行修改边

- 大量节点通过r\_request\_jump与红色Domain节点关联，并与图中IP节点关联，故保留



## 挑战1.1-团伙2（中型团伙）：

- ③ 根据主图找到新的线索进行扩充（图中红色框标注的IP节点）



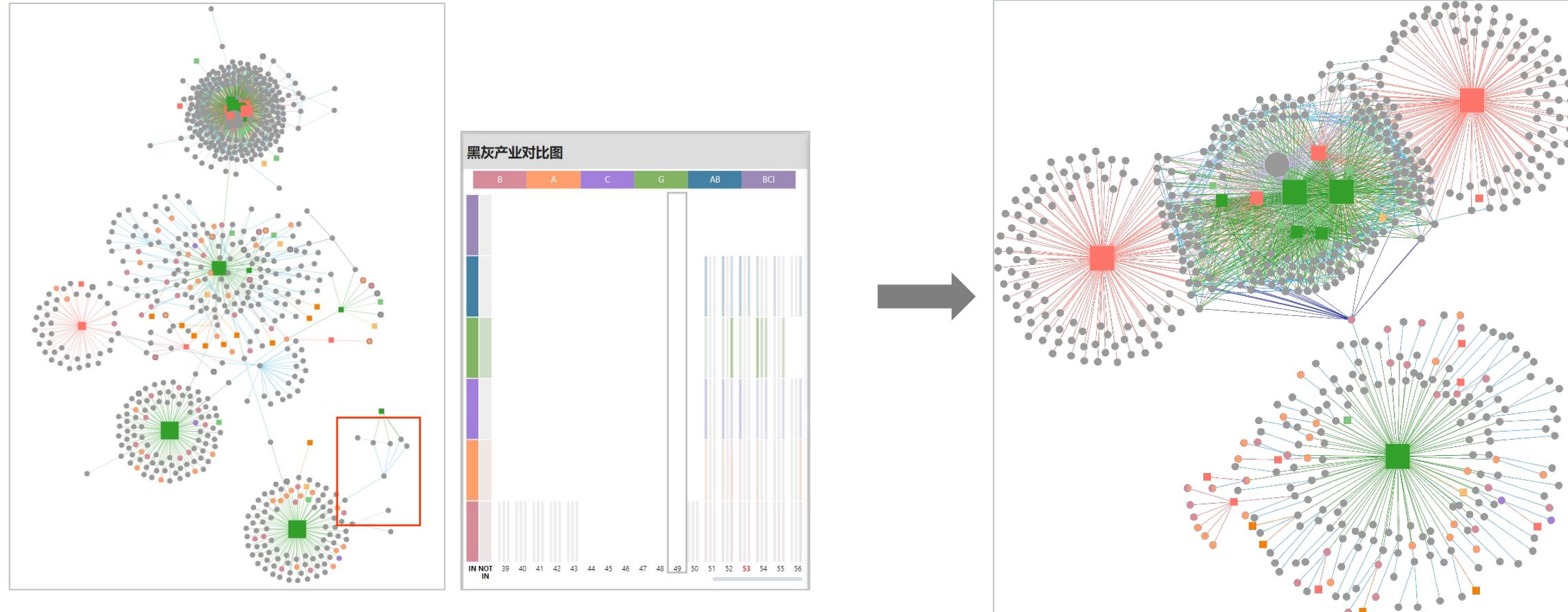
IP | Cert 跳转图

起点	第一跳	第二跳	提交
IP_12658a400c572d024...	IP_b8d2f87bb0a2f0a3...		
WhoisPhone: 0 WhoisEmail: 0 WhoisName: 0 pureDomain: 0 dirtyDomain: 0	IP_gaeBCa54dedad021...		
	IP_d7607115c11be77ad...		



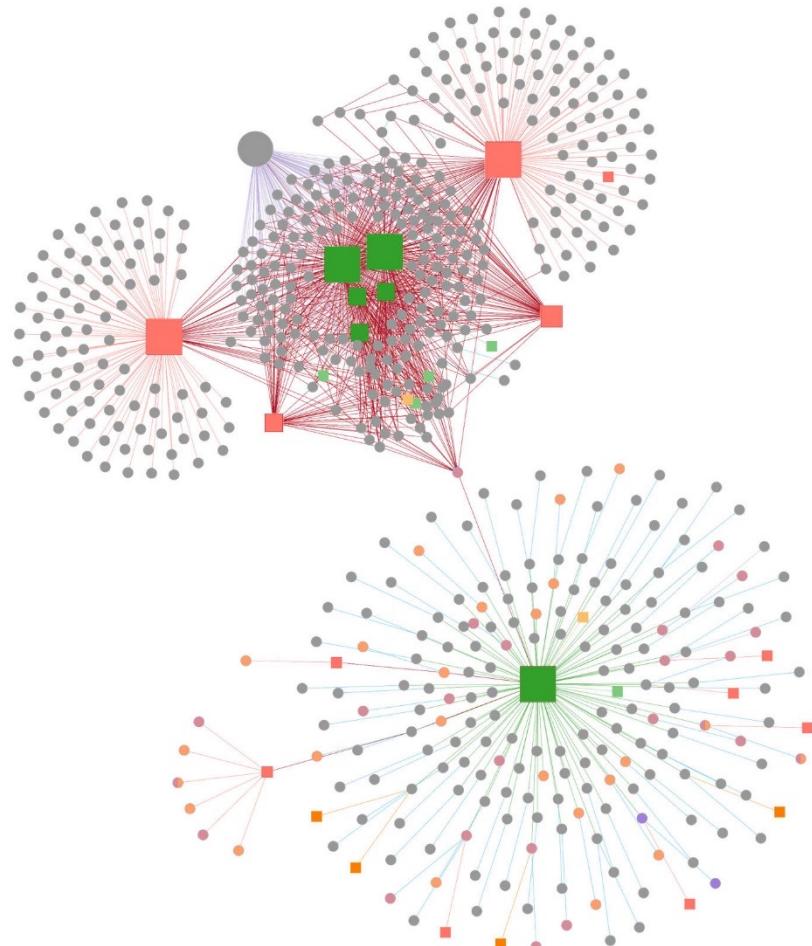
## 挑战1.1-团伙2（中型团伙）：

- ④ 删除主图中和线索节点距离过远、关联不密切的节点



## 挑战1.1-团伙2（中型团伙）：

### ⑤ 团伙统计



**节点：584个**

- ① Domain : 555
- ② IP : 6
- ③ Cert : 12
- ④ Whois\_Name : 4
- ⑤ IP\_C : 5
- ⑥ ASN : 2

**核心资产：**

- ① Cert : 6
- ② IP : 6

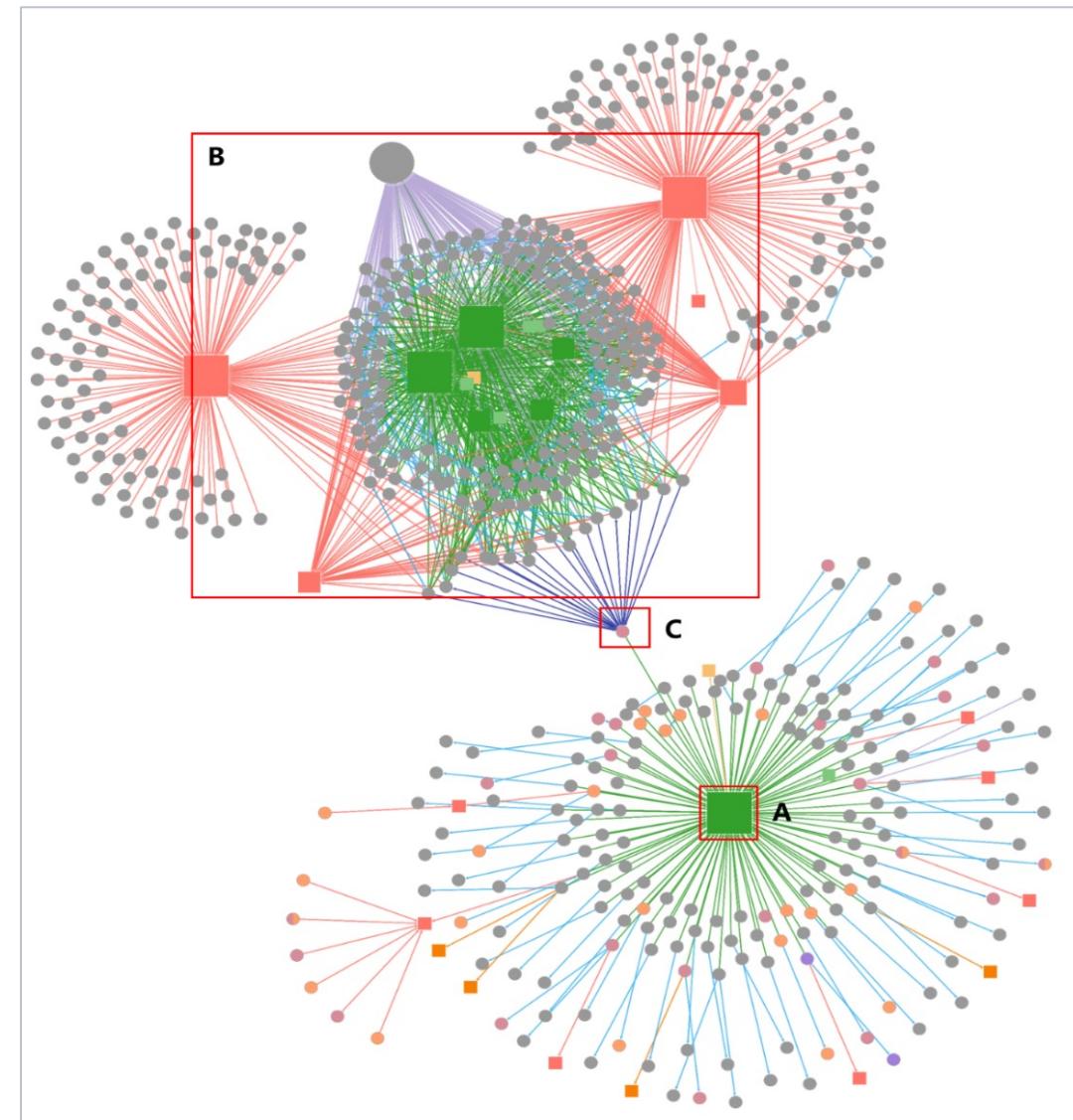
**边：1345条**

- ① r\_cert : 395
- ② r\_subdomain : 185
- ③ r\_request\_jump : 19
- ④ r\_dns\_a : 612
- ⑤ r\_whois\_name : 4
- ⑥ r\_cert\_chain : 1
- ⑦ r\_cname : 122
- ⑧ r\_cidr : 5
- ⑨ r\_asn : 2

## 挑战1.1-团伙2（中型团伙）：

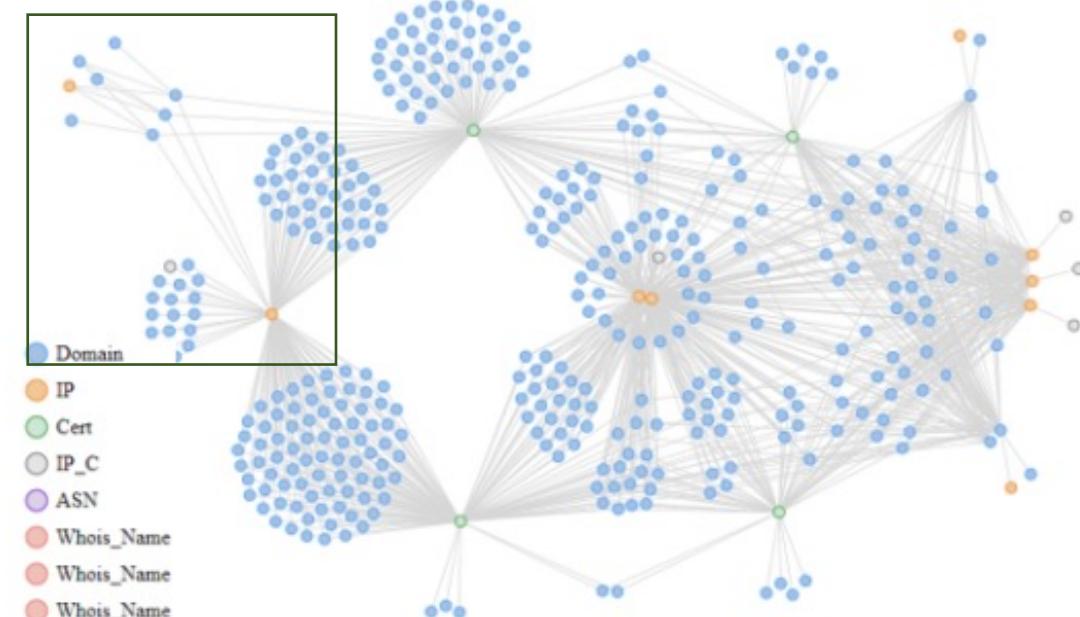
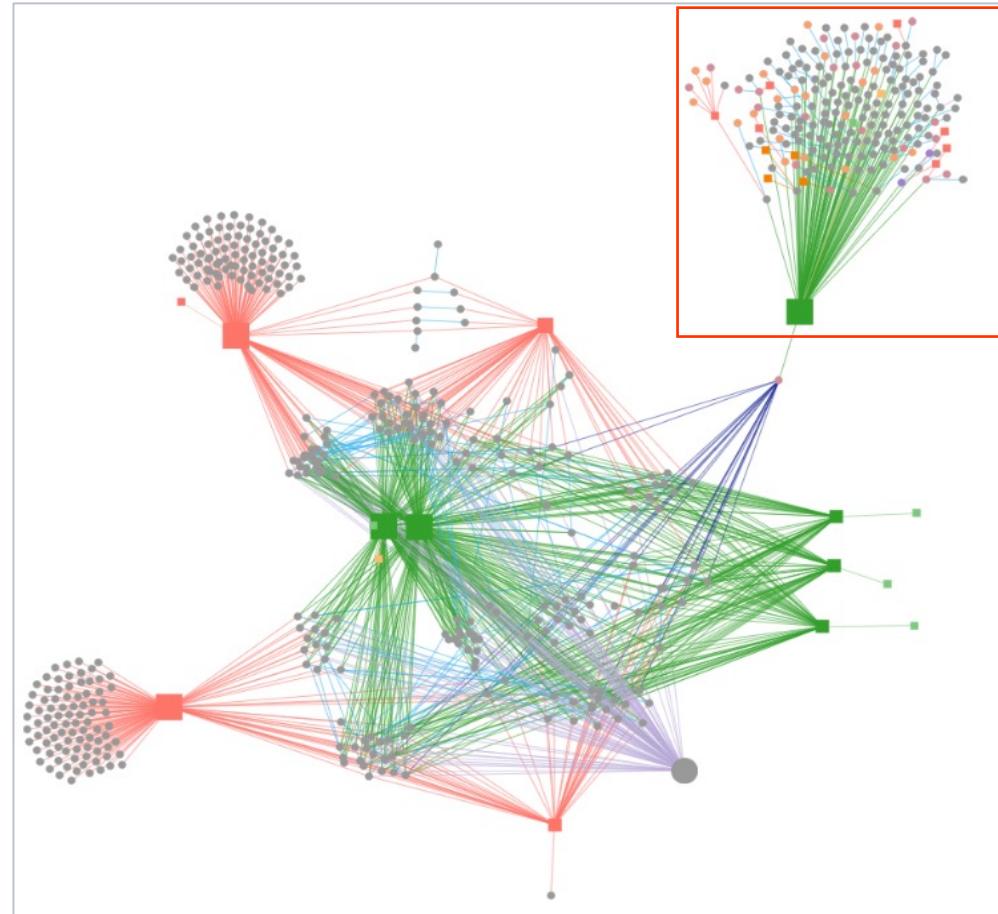
### ⑥ 运作机制分析及封堵策略

- 黑灰产业主要集中在**核心资产A**(IP 108.160.xxx.xxx节点)附近
- B部分使用了大量的**内容分发网络**，且只包含了少量的黑灰产业
- 大量的节点通过r\_request\_jump关联了C(Cert 8659e9de39.com)节点，C节点与核心资产A相邻。
- 封堵A：**有效打击**该团伙涉及的大量黑灰产业
- 封堵B：**有效减少**该黑灰团伙的运营范围
- 封堵C：**阻断**B部分的核心资产与核心资产A之间的**关联**，但是无法阻止该黑灰团伙其他产业的运营



# 挑战1.1-团伙2（中型团伙）：

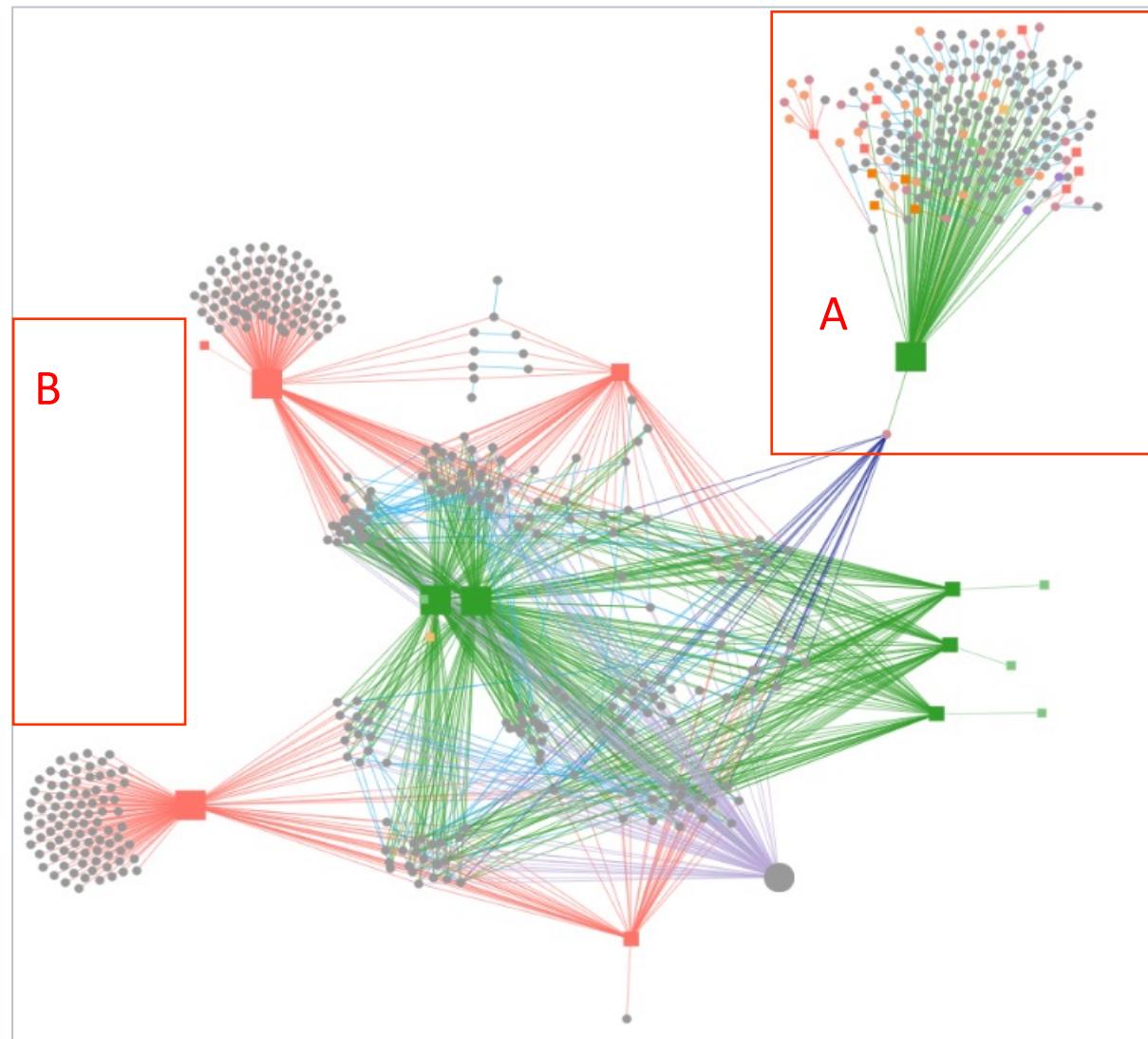
## ⑦ 结果对比



## 挑战1.1-团伙2（中型团伙）：

### ⑧ 原因分析：

- 在分析时，A部分的IP节点连接了大量的黑灰产业，我们对其进行了进一步的探索
- B部分的节点由于其所在IP、Cert节点与黑灰产业没有关联，B部分节点被筛掉



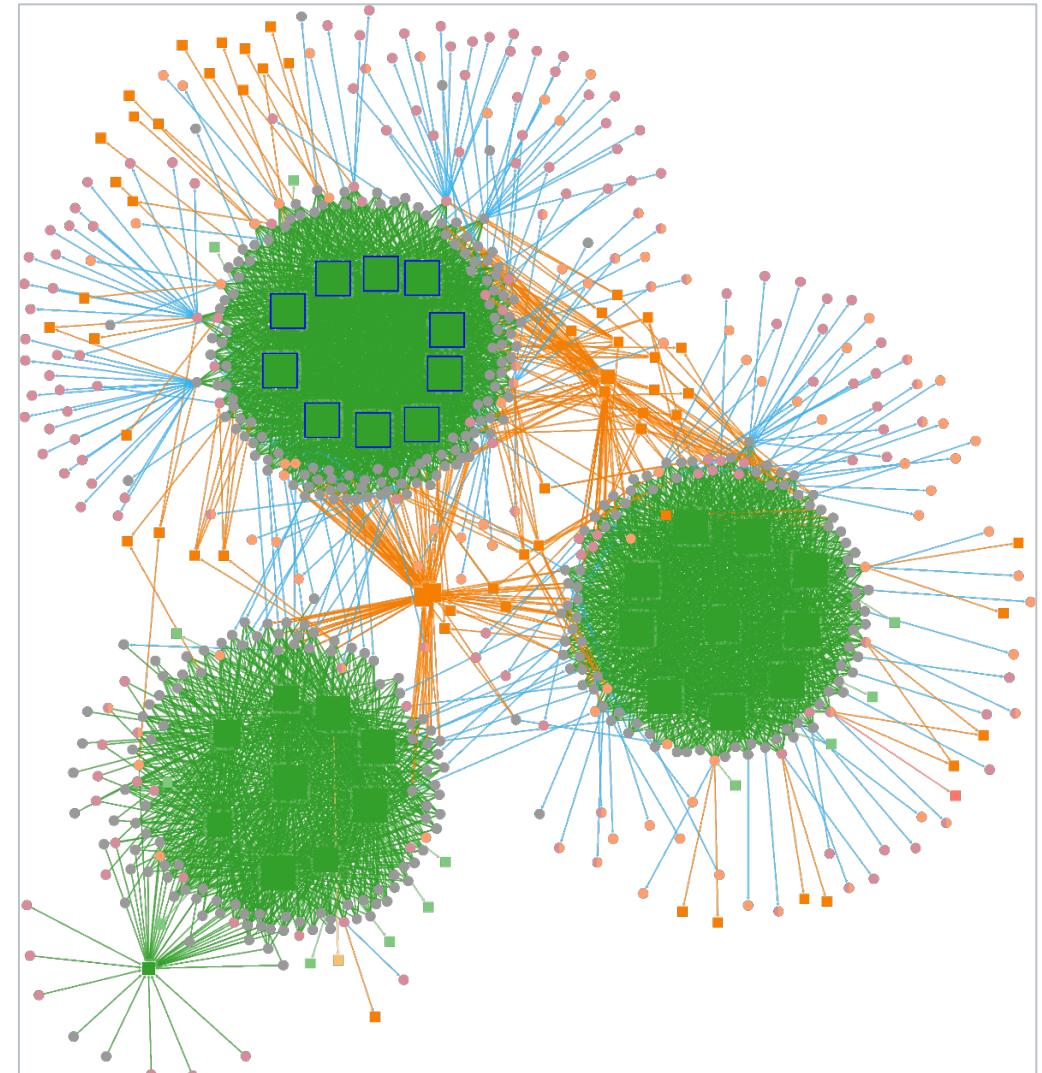
## 挑战1.2 团伙1：大量内容分发网络

### ① 团伙介绍

- 团伙共包含节点666个，边3979条，并包含30个使用内容分发网络的IP节点
- 团伙可分为**三个部分**，每个部分由多个IP共同组成
- 每个部分内的IP都使用了大量的**内容分发网络**，平均一个Domain节点关联超过**6个IP**
- 三个部分单独运作，通过**Whois、Domain**节点互相关联

### ② 打击方案

- 无法单独一个核心资产来打击该团伙，对同一个 Domain节点关联的所有IP节点同时封锁，才能够有效打击



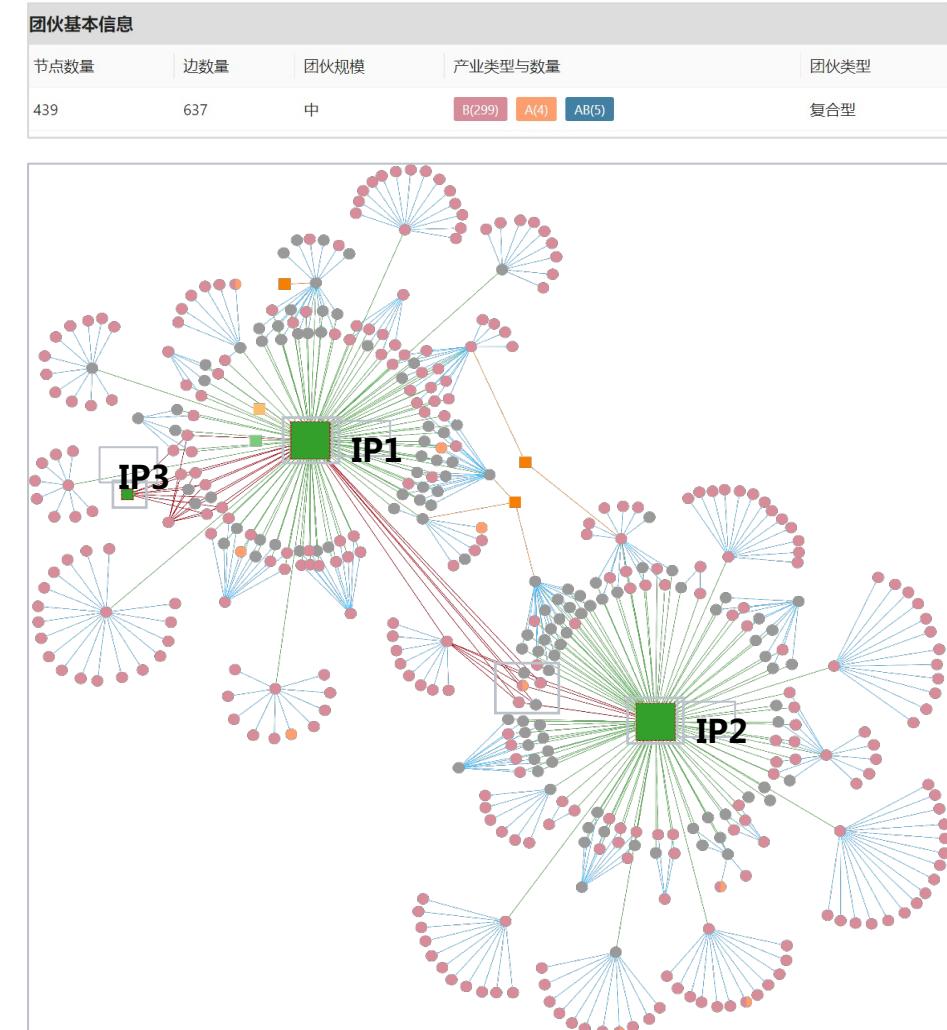
## 挑战1.2 团伙2：“哑铃”结构

### ① 团伙介绍

- 该团伙包含439个节点，637条边
- 三个IP均名为66.150.xxx.xxx（右图IP1、IP2、IP3），并以**IP1**、**IP2**为基础
- IP3节点关联的所有Domain节点均和IP1节点有关联，IP3不作为重要核心资产

### ② 打击方案

- 切断“哑铃”两端IP交流的枢纽，切断产业之间依赖
- 可以对**IP节点进行封堵**，使依赖此IP的大量网站瘫痪



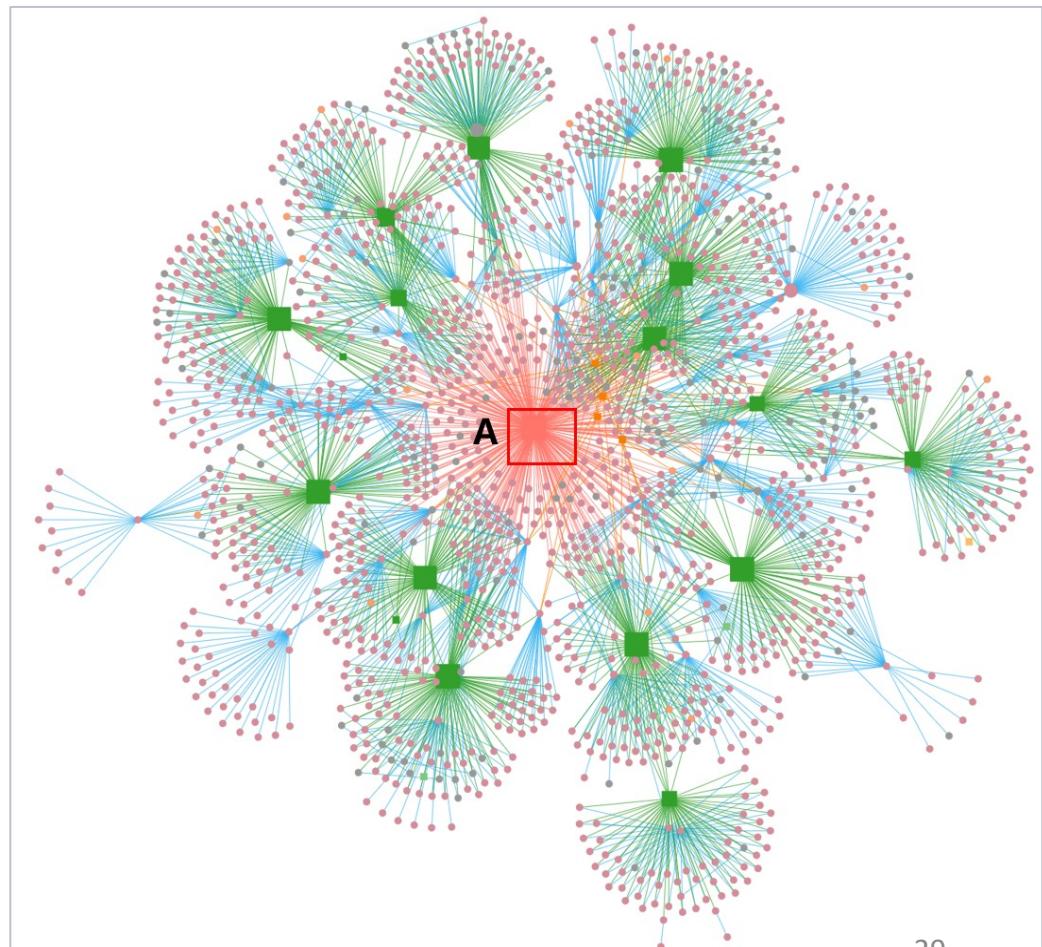
## 挑战1.2 团伙3：多种产业类型；中心性

### ① 团伙介绍

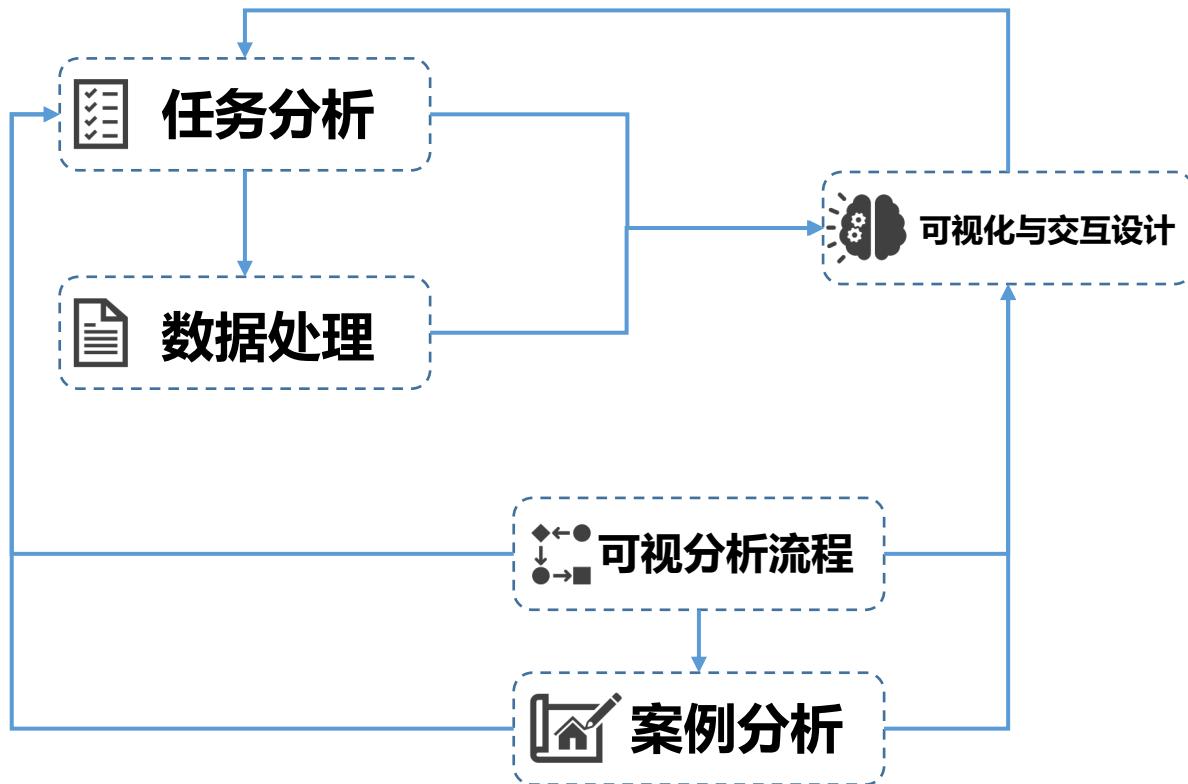
- 该团伙包含1601个节点，3138条边，其中1462个节点涉及黑灰产业，占总结点数量约90%，
- 以5268f305a6 ( Cert ) 节点为**中心**进行**扩散**，关联到众多IP节点
- 其外围的IP节点之间几乎没有**直接关联**，它们都只和中心的Cert节点关联。

### ② 打击方案

- **封堵中心Cert节点**封堵可以有效的打击团伙
- **封堵外围的IP节点**可以有效减少该团伙运作范围
- **从外围的IP节点进行逐步封锁，直至核心的Cert节点**



## 解题思路



## 讨论与改进

- ① 视图颜色过多，上手难度较大
- ② 视图过多，部分辅助视图使用较少

任务分析

数据处理

可视化与  
交互设计

案例分析

总结与展望

# 感谢各位聆听 敬请批评指正

---

黑灰产网络资产可视分析系统

孙德晟 高琳 刘子奥 月小琪 周艺璇 胡海波（指导老师）

重庆大学