ChinaVis Data Challenge 2022
赛道 1 评审指南

一、提交要求

作品提交材料要求：

（1）作品说明文档：要求参赛者根据组织方提供的文档推荐模板，用图文并茂的方式介绍作品，以 Word 或 PDF 格式提交；

（2）视频：要求参赛者制作带解说视频，围绕作品解释其可视分析流程，视频总长度不超过 5 分钟，视频数量 1 个，限 MP4 格式，视频大小建议在 50M 以内；

（3）作品代表性图片：要求参赛者请提供高清版本 1 张，限 JPG 格式，多图请拼接，图片大小不超过 20M。

二、背景介绍

网络黑灰产是指利用信息技术和网络技术，实施各类违法犯罪活动来谋取不正当利益的产业形态。目前，在互联网运行的内容秩序威胁型黑灰产是最常见网络黑灰产类型，它们以公开网站为载体来传播违法违规内容，开展网络诈骗、网络赌博、网络色情、违禁品交易等犯罪活动，严重侵害网络生态的健康发展，甚至威胁着网民生命财产安全。

网络黑灰产具有链条化、团伙化、资产化和跨域化等特点。链条化是指黑灰产形成了环环相扣的上、中、下游产业链，共同配合完成非法牟利。资产化是指黑灰产团伙掌握大量且关联复杂的多种网络资产，以支撑产业链的网络化运转，比如：上游信息盗取需要木马和钓鱼网站，中游业务网站运维需要域名和 IP 地址；下游支付需要安全证书。跨域化是指黑灰产团伙为躲避追查，将一部分网络资产和成员布置在境外。

分析黑灰产团伙掌握的网络资产是打击黑灰产的重要切入点。网络资产可以分为外围网络资产、普通网络资产和核心网络资产。外围网络资产主要是向网民直接公开的黑灰产业务网站域名。核心网络资产是关系到许多外围网络资产运行或关联多个业务线的网络资产，比如：同时支持多个网站域名运行的某 IP 地址，又比如：同一黑灰产团伙掌控的赌博业务网站和违禁品交易业务网站共同使用的

数字安全证书。核心网络资产信息一般不直接向网民公开，部分核心网络资产信息隐藏在多种公开数据源中。普通网络资产介于其它两类网络资产之间。

查证和封堵黑灰产团伙掌握的核心网络资产是目前打击黑灰产的主要手段之一。原因来自三个方面：一是封堵外围网络资产效率低且被动滞后，因为网站复本多，存活周期短，域名更换频繁。二是封堵核心网络资产可以让许多非法网站失效或陷入安全风险，造成高额恢复成本。三是深度分析核心网络资产有利于发现关联多资产或多业务的关键链路，还原多个业务线之间的联系，甚至发现真实世界中控制黑灰产的嫌疑人。

奇安信公司通过多种技术手段，在多个公开数据源中收集和整理黑灰产网络资产信息，形成了一个黑灰产网络资产图谱数据集（已脱敏）。该数据集以点边双异质有向图为数据结构，节点为网络资产，边为网络资产间关联关系。该数据集包含 8 类网络资产和 11 类资产关联关系，共有 237 万个节点，328 万条边。

假设你是网络黑灰产治理人员，请设计一套可视分析方案，从数据集中找到一些由同一黑灰产团伙掌握的网络资产子图，并识别子图中的核心资产与关键链路，将结果用图表形式呈现出来。

三、数据支持

本次挑战赛为参赛者提供了一份黑灰产网络资产图谱数据集（已脱敏），包括 Node.csv 和 Link.csv 两个数据文件，分别记录了节点和边的信息。

3.1 Node.csv 说明

Node.csv 数据文件大小为 229M，包括 237 万条数据记录，每一条数据记录一个节点，包括表 3.1 所示的 4 个字段。图 3.1 展示了 Node.csv 的数据样本。

表 3.1 Node.csv 数据文件—字段说明

字段	说明	类型	示例	说明
id	节点 id	String	Domain_0d9f06a82e90193f68e72e53acd55e23c74afb0e3589608627e423c64d19f6db	唯一标识节点
name	节点名称	String	0d9f06a82e.com	经过了 MD5 加密和无效化脱敏处理
type	节点类型	String	Domain	共 8 类，见表 3.2
industry	黑灰产业务类型(只对 Domain 类型节点有效)	String	['B']	共 10 类，见表 3.3

表 3.2 节点类型说明

字段	说明	数量	重要程度
Domain	网站域名	200 万	非常重要
IP	网站的 IP 地址	20 万	非常重要
Cert	网站用的 SSL 安全证书	13 万	非常重要
Whois_Name	网站域名的注册人姓名	1.8 万	重要
Whois_Phone	网站域名的注册人电话	0.2 万	重要
Whois_Email	网站域名的注册人邮箱	0.4 万	重要
IP_C	IP 的 C 段	0.6 万	一般
ASN	IP 的自治域	0.03 万	一般

表 3.3 黑灰产业务类型说明

industry 字段值	黑灰产业务类型	说明
A	涉黄	该域名的网站涉及色情传播
B	涉赌	该域名的网站涉及网络传播
C	诈骗	该域名的网站涉及网络诈骗，如仿冒著名网站
D	涉毒	该域名的网站涉及毒品交易
E	涉枪	该域名的网站涉及枪支交易
F	黑客	该域名的网站是嵌入恶意信息的黑客网站，如嵌入木马的钓鱼网站
G	非法交易平台	该域名的网站涉及非法交易，如个人信息买卖
H	非法支付平台	该域名的网站是非法支付平台
I	其他	其他黑灰产业务网站

```
Domain_0586b66338e82edf74a0a7d65d1e5835a86647b2e3781e5718c6330e0aca3617,0586b66338.com,Domain,['B']
Cert_fb7076fed16346aeb065c7d6f984ddff37b8dd4b35d2bd1a07f30ef7b819b03d,fb7076fed1,Cert,[]
IP_37f7ed5739b43757ff23c712ae4d60d16615c59c0818bf5f2c91514c9c695845,5.180.xxx.xxx,IP,[]
IP_44e642e648fa555970bfd01596dc1b67e65b357e469479b4105fed2758339462,156.245.xxx.xxx,IP,[]
Cert_5dd7cba66d526fbaaa23b4f2c375f2a10cf4cc9e927682e9602f423a9ae96d38,5dd7cba66d,Cert,[]
Whois_Name_da9834465d7bf75b26f00e78a2412c55a9bb160ab439ee4c0e7742c507a6ac78,lixxxxxxi,Whois_Name,[]
Whois_Email_e3ed53e22963da2784dc9aad7a83c123790617384f67d719fa31fa1c1872a417,sbiqqxxxxx@xxx.xxx,Whois_Email,[]
Whois_Phone_b9383e2d6af1ab1d9f4648f2b7bd348fb875f829124662f2ff4b510af4b66b89,+86.870xxxxx,Whois_Phone,[]
IP_C_80052b75991b23fad5ef78809203fc4e0f4af613c2414f51eba45772149a9625,156.245.xxx.0/24,IP_C,[]
Domain_a7eb1ab42b77f5806e61efe29fefa61bb58686f00f241c1753e7f399448e90f7,a7eb1ab42b.com,Domain,[]
ASN_894a39aa8f6405a82567c5c1832fd3a6b110552c2fe84eafa929a3e603fc4387,AS_894a39aa8f,ASN,[]
```

图 3.1 Node.csv 数据样本示例

3.2 Link.csv 说明

Link.csv 数据文件大小为 493M，包括 328 万条数据记录，每一条数据记录对应一条边，包括表 3.4 所示的 3 个字段。图 3.2 展示了 Link.csv 的数据样本。

表 3.4 Link.csv 数据文件—字段说明

字段	说明	类型	示例	说明
relation	边类型	String	r_dns_a	共 11 类，见表 3.5
source	源节点	String	IP_37f7ed5739b43757ff23c712ae4d60d16615c59c0818bf5f2c91514c9c695845	源节点的 id 字段值
target	目标节点	String	Domain_2d3bbcec29453b6f56fb85ea28e8e5ea5fc5f5562e0f896b6b52b113a6cc1e44	目标节点的 id 字段值

表 3.5 边的名称说明

relation 字段	说明	数量	关联强度
r_cert	域名使用的安全证书	23 万	很强
r_subdomain	域名拥有的子域名	45 万	很强
r_request_jump	域名间跳转关系	0.06 万	很强
r_dns_a	域名对应的 IP 地址	205 万	很强
r_whois_name	域名的注册人姓名	10 万	较强
r_whois_email	域名的注册人邮箱	2.8 万	较强
r_whois_phone	域名的注册人电话	1.9 万	较强
r_cert_chain	证书的证书链关系	1.5 万	一般
r_cname	域名对应的别名	13 万	一般
r_asn	IP 所属的自治域	6.9 万	较弱
r_cidr	IP 所对应的 C 段	17 万	较弱

图 3.2 Link.csv 数据样本示例

3.3 黑灰产网络资产图谱模型

黑灰产网络资产图谱数据集中包括 8 种类型的节点和 11 种类型的边，图 3.3 给出了黑灰产网络资产图谱抽象模型，说明了各类型节点间的可能关联关系类型。

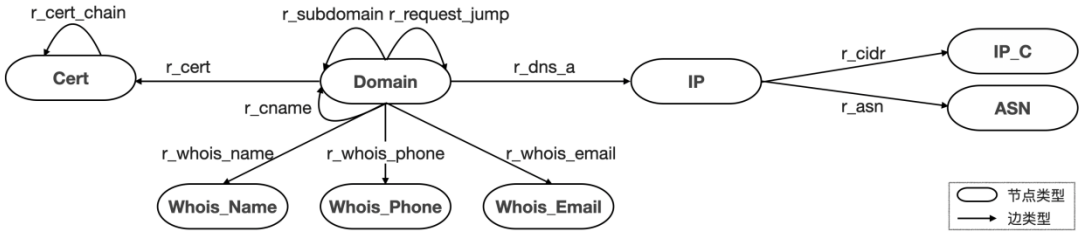


图 3.3 黑灰产网络资产图谱抽象模型

图 3.4 给出了以“5.180.xxx.xxx”IP 地址为线索（图 3.4 中红色节点），在黑灰产网络资产图谱数据集中挖掘到的小型黑灰产团伙的网络资产子图。图中的 N1、N3 是安全证书节点，绝大部分域名关联到这两个安全证书。N2 是 IP 节点，许多域名关联到这个 IP 地址。这些现象反映了许多域名（业务网站）共同使用了这两个安全证书，并且一部分网站部署在了同一个 IP 地址（服务器）上。另外，这些域名对应的网站大部分都是涉赌、涉黄、涉枪、游戏私服类网站。综上，该子图中的网络资产可能由同一个黑灰产团伙掌握，该黑灰产团伙同时开展了多项非法业务，其核心网络资产是 N1、N2 和 N3，这三个核心网络资产之间的通路是网络业务的关键链路。

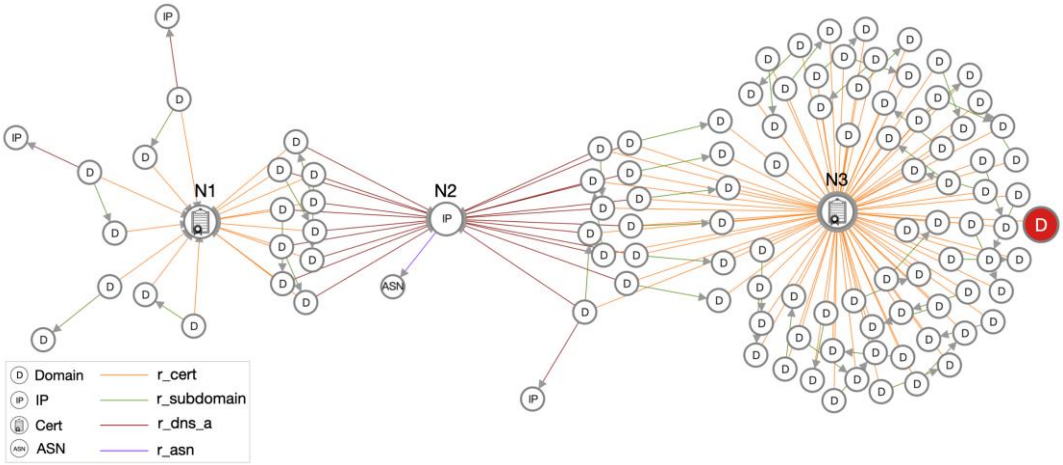


图 3.4 某小型黑灰产团伙掌握的网络资产子图示例

四、参考答案

评审总体说明：

1、要求回答问题准确且简明扼要；

2、要求以可视分析为主要技术路线探索问题答案；

3、要求用可视化的方式呈现与解释给出的答案；

4、鼓励给出参考答案以外的任何合理的新发现；

5、鼓励在分析过程中引入智能算法；

6、鼓励参赛队伍自行开发新颖的可视分析解决方案；

7、鼓励参赛队伍使用自己团队（公司）研发的分析工具。

4.1 挑战 1.1

请根据表 4.1 所示的五个黑灰产团伙的网络资产线索，在黑灰产网络资产图谱数据集中分别挖掘对应的网络资产子图（一个子图期望是由同一个黑灰产团伙掌握的网络资产及其关联关系）；识别每个子图中的核心网络资产和关键链路；用图表的形式呈现结果并简要分析每个黑灰产团伙网络运作机制。（请将答案尽量控制在 2000 字、10 张图片、10 个表格内）

表 4.1 给出了 5 个黑灰产团伙的网络资产线索，根据这些线索可以分别找到 5 个网络资产子图，如图 4.1 所示。每个子图都对应一个潜在黑灰产团伙所控制的网络资产，每个子图中红色高亮并加大的节点对应表 4.1 中的网络资产线索。

表 4.1 黑灰产团伙的网络资产线索（赛题中提供）

团伙	节点 id	节点 name	节点类型
团伙 1	Domain_c58c149eec59bb14b0c102a0f303d4c20366926b5c3206555d2937474124beb9	c58c149eec.com	Domain
	Domain_f3554b666038baffa5814c319d3053ee2c2eb30d31d0ef509a1a463386b69845	f3554b6660.com	Domain
团伙 2	IP_400c19e584976ff2a35950659d4d148a3d146f1b71692468132b849b0eb8702c	156.241.xxx.xxx	IP
	Domain_b10f98a9b53806ccd3a5ee45676c7c09366545c5b12aa96955cde3953e7ad058	b10f98a9b5.com	Domain
团伙 3	Domain_24acfd52f9ceb424d4a2643a832638ce1673b8689fa952d9010dd44949e6b1d9	24acfd52f9.com	Domain
	Domain_9c72287c3f9bb38cb0186acf37b7054442b75ac32324dfd245aed46a03026de1	9c72287c3f.com	Domain

	Domain_717aa5778731a1f4d6f0218dd3a27b114c839213b4af781427ac1e22dc9a7dea	717aa57787.com	Domain
	Domain_8748687a61811032f0ed1dcdb57e01efef9983a6d9c236b82997b07477e66177	8748687a61.com	Domain
	Whois_Phone_f4a84443fb72da27731660695dd00877e8ce25b264ec418504fface62cdcbbd7	+1.971xxxxx	Whois_Phone
团伙 4	IP_7e730b193c2496fc908086e8c44fc2dbbf7766e599fabde86a4bcb6afdaad66e	23.82.xxx.xxx	IP
	Cert_6724539e5c0851f37dcf91b7ac85cb35fcd9f8ba4df0107332c308aa53d63bdb	6724539e5c	Cert
团伙 5	Whois_Phone_fd0a3f6712ff520edae7e554cb6dfb4bdd2af1e4a97a39ed9357b31b6888b4af	+86.400xxxxx	Whois_Phone
	IP_21ce145cae6730a99300bf677b83bbe430cc0ec957047172e73659372f0031b8	3.234.xxx.xxx	IP
	Domain_7939d01c5b99c39d2a0f2b418f6060b917804e60c15309811ef4059257c0818a	7939d01c5b.com	Domain
	Domain_587da0bac152713947db682a5443ef639e35f77a3b59e246e8a07c5eccae67e5	587da0bac1.com	Domain

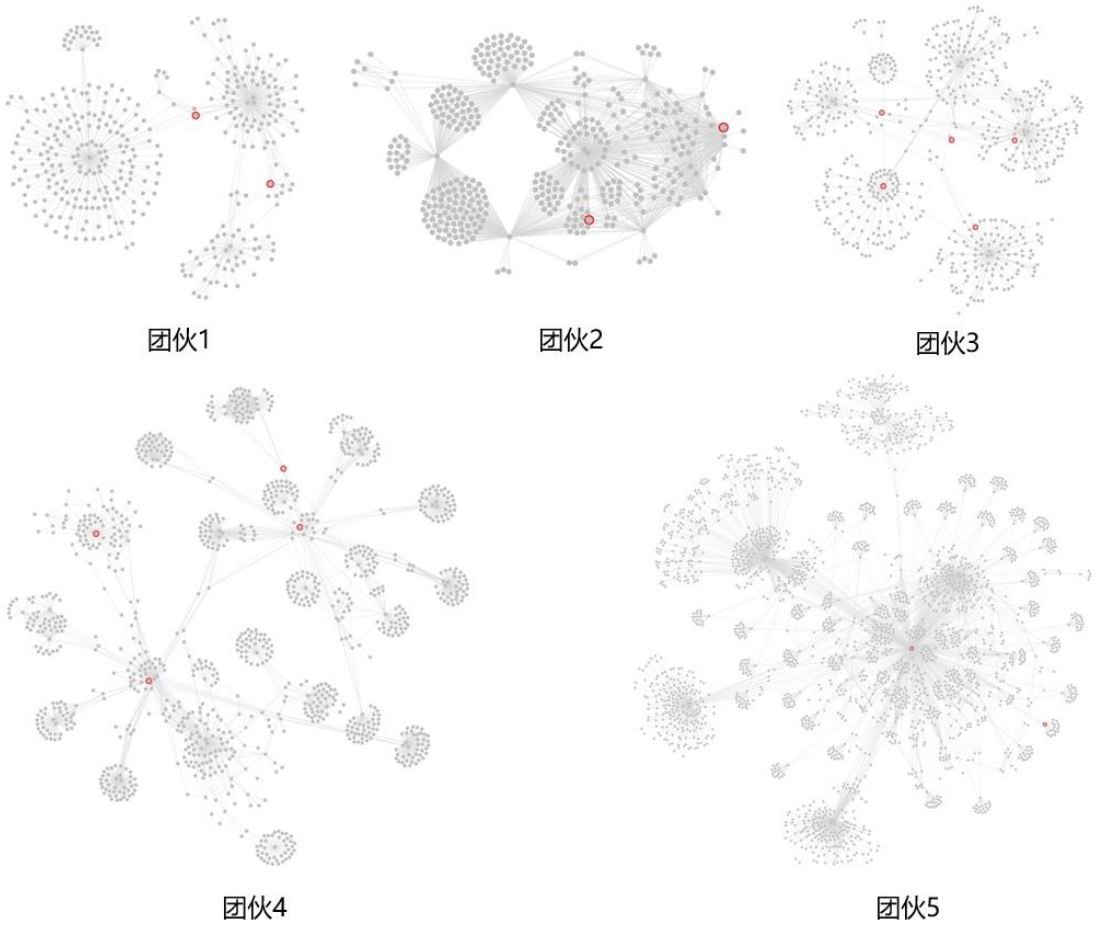


图 4.1 五个黑灰产团伙的网络资产子图概览。红色边框高亮并加大的节点对应子图的线索

接下来，我们介绍挑战 1.1 的参考答案。答案中的网络资产子图均由专家基于业务规则和领域知识，通过人工方式从黑灰产网络资产图谱数据集中挖掘得到的，并人工识别了子图中的核心网络资产和关键链路。

表 4.2 简要介绍了 5 个黑灰产团伙的网络资产子图的统计信息，包括节点与边数量统计、核心网络资产数量统计、关键链路上的边数量统计、关键链路上的节点（下文简称桥节点）数量统计。注意：我们对原始挖掘得到的网络资产子图的规模进行了控制。在网络资产子图中存在一些大型的簇状结构，这些簇状结构中有很多结构同构的节点（节点与边类型相似、节点邻域连接结构相似）。因此，我们随机采样一些与簇中心连接的同构节点，采样率大致在 10~20%。在表 4.2 中，除了团伙 2，其它团伙的网络资产子图都应用了上述采样过程，所示的节点和边数量均为采样后数量。

表 4.2 黑灰产团伙的统计信息

团伙	节点数量	边数量	核心网络资产节点数量	核心网络资产节点间的关键链路上的边数量	核心网络资产节点间的关键链路上的桥节点数量
团伙 1（小型团伙）	368	617	6	22	7
团伙 2（中型团伙）	401	1099	8	732	271
团伙 3（中型团伙）	589	1057	2	1	0
团伙 4（大型团伙）	2354	5271	73	1261	357
团伙 5（大型团伙）	1079	2345	50	293	62

4.1.1 团伙 1

图 4.2 是团伙 1 的网络资产子图，表示该黑灰产团伙所控制的网络资产。该黑灰产团伙可能进行涉黄、涉赌、诈骗、涉枪等多种黑灰产业务。总体上，该黑灰产团伙掌握的网络资产呈现出以域名为主体，少量安全证书、IP、Whois 注册信息为核心网络资产，多域名与单一核心网络资产关联紧密，少量节点同时与多个核心网络资产间关联的分布模式。下面，我们分别介绍该子图的节点与边统计信息和核心网络资产与关键链路信息。

1、节点与边统计信息

该网络资产子图的节点数量 368 个，分布情况如图 4.2（a）所示。各种类型

节点的数量和比例如下：

- Domain 类型　　230 个（62.5%）；
- IP 类型　　107 个（29.1%）；
- Whois_Email 类型　　9 个（2.4%）；
- Whois_Phone 类型　　9 个（2.4%）；
- Whois_Name 类型　　6 个（1.6%）；
- Cert 类型　　5 个（1.4%）；
- IP_C 类型　　1 个（0.3%）；
- ASN 类型　　1 个（0.3%）。

该案例的边数量 617 条，分布情况如图 4.2（b）所示。各种类型边的数量和比例如下。

- r_dns_a 类型　　171 条（27.7%）；
- r_cert 类型　　147 条（23.8%）；
- r_whois_email 类型　　74 条（12.0%）；
- r_whois_name 类型　　70 条（11.3%）；
- r_whois_phone 类型　　70 条（11.3%）；
- r_subdomain 类型　　50 条（8.1%）；
- r_dns_cname 类型　　26 条（4.8%）；
- r_request_jump 类型　　4 条（0.6%）；
- r_cert_chain 类型　　3 条（0.5%）；
- r_asn 类型　　1 条（0.2%）；
- r_cidr 类型　　1 条（0.2%）。

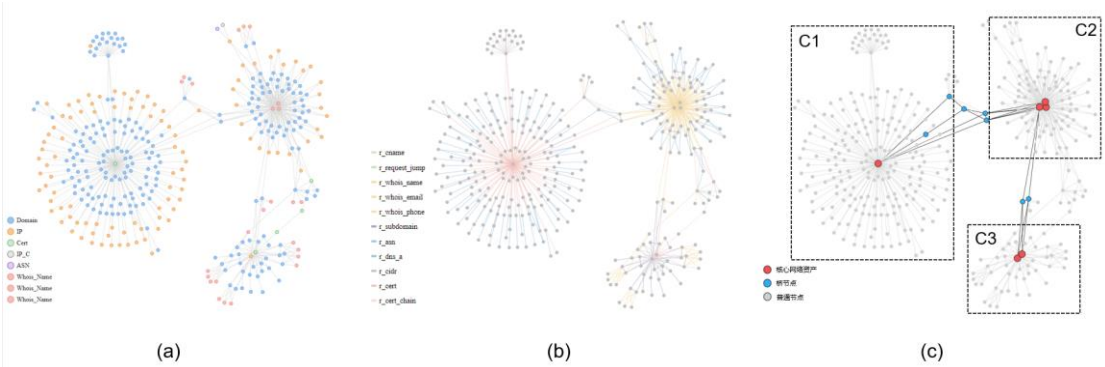


图 4.2 团伙 1 掌握的网络资产子图示例。（a）用颜色编码节点类型的网络资产子图；（b）用颜色编码边类型的网络资产子图；（c）核心网络资产与关键链路示意图。

2、核心网络资产与关键链路信息

如图 4.2（c）所示，该网络资产子图中有 6 个核心网络资产，分布在 C1、C2、C3 区域中。

结合图 4.2（a）、图 4.2（b）可知，C1 区域包含了很多蓝色的域名（Domain）节点，这些节点均通过 r_cert 类型的边连接到同一个绿色的安全证书（Cert）节点。同时，这些域名节点也各自通过 r_dns_a 类型的边连接到不同的黄色 IP 节点。以上信息说明了 C1 中域名的网站配置了相同的安全证书，但它们部署在不同的域名服务器上，域名的 IP 地址各不相同。因此，这些域名共用的安全证书，即，Cert 节点，就是该子图的一个核心网络资产。

C2 区域包含了很多蓝色的域名（Domain）节点，这些节点均通过 r_whois_name、r_whois_phone、r_whois_email 类型的连边连到相同的红色 Whois 注册信息（Whois_Name，Whois_Phone、Whois_Email）节点。同时，这些域名节点也各自通过 r_dns_a 类型的边连接到不同的黄色 IP 节点。以上信息说明了这些域名很可能是被同一个人或同一组织注册的，所以具有相同的 Whois 注册信息，但这些域名部署在不同的域名服务器上，所以域名的 IP 地址不同。因此，这些域名共同的注册信息，即 Whois_Name，Whois_Phone 和 Whois_Email 节点，就是该子图的核心网络资产。

C3 区域包含了一些蓝色的域名（Domain）节点，这些节点均同时通过 r_cert、r_dns_a 类型的边连接到相同的绿色安全证书（Cert）节点和黄色 IP 节点。同时，这些域名节点也通过 r_whois_name、r_whois_phone、r_whois_email 类型的连边连到各自不同的红色 Whois 注册信息（Whois_Name，Whois_Phone、Whois_Email）节点。以上信息说明了这些域名的网站配置了相同的安全证书，且部署在相同的域名服务器上，但很可能被不同的人或组织注册。这些域名共用的安全证书和域名服务器，即，Cert 节点和 IP 节点，就是该子图的核心网络资产。

C1 和 C2 区域之间、C2 和 C3 区域之间，均通过少量与各区域内核心网络资产关联的域名节点产生关联。比如，C1 和 C2 区域之间，有少量域名即通过 r_cert 边与 C1 中的安全证书节点相连，又通过 r_whois_name、r_whois_phone、r_whois_email 类型的连边与 C2 的 Whois 注册信息节点相连，说明这几个域名的网站即配置了 C1 的安全证书又有 C2 的注册信息。C2 和 C3 区域与之类似，有少量域名的即通过 r_cert 边、r_dns_a 边分别与 C3 中的安全证书节点、IP 节点相连，又通过 r_Whois_Name、r_Whois_Phone、r_Whois_Email 类型的连边与 C2 的 Whois 注册信息节点相连，说明这几个域名的网站即配置了 C3 的安全证书又部署了 C3 的域名服务器，还具有 C2 的注册信息。以上关联不同区域的边的关联程度均为很强或较强，反映出这些网络资产关联紧密，很可能是同一个黑灰产团伙掌握的网络资产。这些关联不同区域间的节点就是桥节点，经过核心网络资产

与桥节点之间的路径就是该子图的关键链路。

4.1.2 团伙 2

图 4.3 是团伙 2 的网络资产子图，表示该黑灰产团伙所控制的网络资产。该黑灰产团伙可能进行涉赌的黑灰产业务。总体上，该黑灰产团伙掌握的网络资产呈现出以域名为主体，少量安全证书、IP、域名为核心网络资产，多域名同时与多核心网络资产关联，使得多核心网络资产间关联紧密的分布模式。下面，我们分别介绍该子图的节点与边统计信息和核心网络资产与关键链路信息。

1、节点与边统计信息

该网络资产子图的节点数量 401 个，分布情况如图 4.3（a）所示。各种类型节点的数量和比例如下：

- Domain 类型　　383（95.5%）；
- IP 类型　　9 个（2.2%）；
- Cert 类型　　4 个（1.0%）
- IP_C 类型　　5 个（1.2%）。

该案例的边数量 1099 条，分布情况如图 4.3（b）所示。各种类型边的数量和比例如下。

- r_dns_a 类型　　595 条（54.1%）；
- r_cert 类型　　377 条（34.3%）；
- r_request_jump 类型　　19 条（1.7%）；
- r_subdomain 类型　　26 条（2.4%）；
- r_ cname 类型　　76 条（6.9%）；
- r_cidr 类型　　6 条（0.55%）。

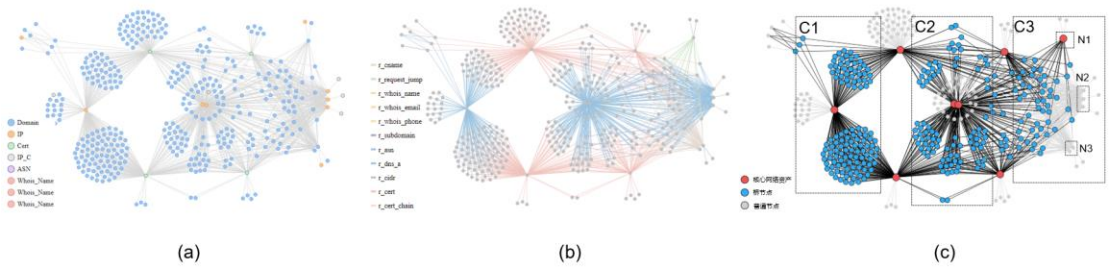


图 4.3 团伙 2 掌握的网络资产子图示例。（a）用颜色编码节点类型的网络资产子图；（b）用颜色编码边类型的网络资产子图；（c）核心网络资产与关键链路示意图。

2、核心网络资产和关键链路信息

该网络资产子图中有 8 个核心网络资产，可以把该子图整体划分为 C1、C2、C3 区域，如图 4.3（c）所示。

结合图 4.3（a）、图 4.3（b）可知，C1 区域包含了很多蓝色的域名（Domain）节点，这些节点均通过 r_dns_a 类型的边连接到相同的黄色 IP 节点，说明了这些域名的网站部署在同一个域名服务器上。因此，这些域名共用的 IP 地址就是该子图的一个核心网络资产。

C2 区域与 C1 区域类似，包含很多蓝色的域名（Domain）节点通过 r_dns_a 类型的边连接到两个相同的黄色 IP 节点，说明了这些域名的网站部署在两个相同的域名服务器上。因此，这些域名共用的两个 IP 地址就是该子图的核心网络资产。

C3 区域包含了很多蓝色的域名（Domain）节点，这些节点通过 r_request_jump 类型的边连接到同一个域名节点（N1），并通过 r_dns_a 类型的边共同连接到三个 IP 节点（N2），还通过 r_cname 边共同连接到 2 个域名（N3）。以上信息说明了访问这些域名的网站会自动跳转到同一个其他网站，且这些域名的网站部署在相同的三个域名服务器上，并配置了相同的两个域名别名。根据业务规则可知，当大量域名同时有三个及以上的 IP 地址时，大概率使用了内容分发服务，所以这些 IP 地址就不是该团伙控制的核心网络资产。此时，域名与其别名间也就失去了直接关联，两个 CNAME 域名（域名的别名）也不是核心网络资产。因此，这些域名共同跳转到的 N1 域名是该区域的核心网络资产。

C1 区域和 C2 区域之间，以及 C2 和 C3 区域之间，均由于各区域内大多数域名节点通过 r_cert 边指向相同的证书节点而产生了紧密关联，说明位于不同区域内的域名的网站配置了相同的安全证书。证书类型的边的关联程度很强，反映出这些网络资产关联紧密，很可能是同一个黑灰产团伙掌握的网络资产。所以，使不同区域间产生关联的四个证书就是核心网络资产，使核心网络资产间产生关联的边和节点就是关键链路和桥节点。

4.1.3 团伙 3

图 4.4 是团伙 3 的网络资产子图，表示该黑灰产团伙所控制的网络资产。该黑灰产团伙可能进行涉黄、涉赌的黑灰产业务。总体上，该黑灰产团伙掌握的网络资产呈现出以域名为主体，多域名与少量 IP 或少量域名网络资产关联紧密，且少量 IP 与少量域名网络资产间直接关联或通过一些域名间接产生关联的分布模式。下面，我们分别介绍该子图的节点与边统计信息和核心网络资产与关键链

路信息。

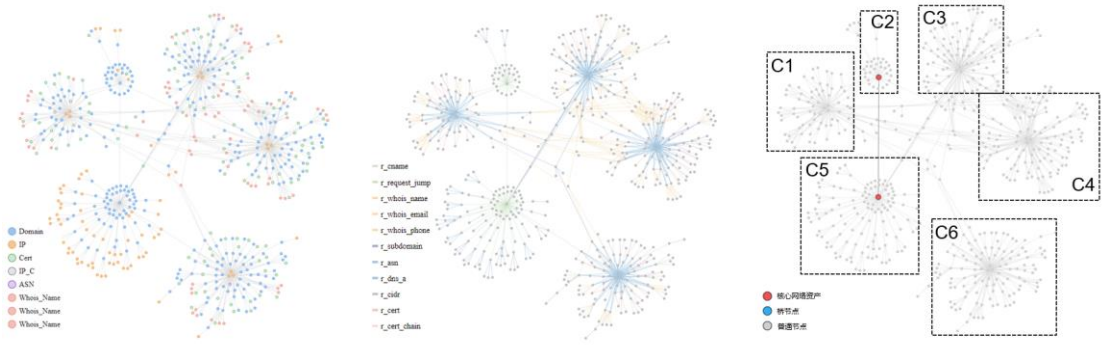
1、节点与边统计信息

该网络资产子图的节点数量 589 个，分布情况如图 4.4（a）所示。各种类型节点的数量和比例如下：

- Domain 类型　　317 个（53.8%）；
- Cert 类型　　106 个（18%）；
- IP 类型　　83 个（14%）；
- Whois_Email 类型　　38 个（6.5%）；
- Whois_Phone 类型　　24 个（4%）；
- Whois_Name 类型　　17 个（3%）；
- IP_C 类型　　3 个（0.5%）；
- ASN 类型　　1 个（0.2%）。

该案例的边数量 1057 条，分布情况如图 4.4（b）所示。各种类型边的数量和比例如下。

- r_dns_a 类型　　548 条（51.8%）；
- r_cert 类型　　130 条（12.3%）；
- r_request_jump 类型　　121 条（11.4%）；
- r_subdomain 类型　　75 条（7.1%）；
- r_whois_phone 类型　　61 条（5.8%）；
- r_whois_emai 类型　　56 条（5.3%）；
- r_whois_name 类型　　52 条（4.9%）；
- r_ cname 类型　　8 条（0.7%）；
- r_asn 类型　　3 条（0.3%）；
- r_cidr 类型　　3 条（0.3%）。



(a)　　(b)　　(c)

图 4.4 团伙 3 掌握的网络资产子图示例。（a）用颜色编码节点类型的网络资产子图；（b）用颜色编码边类型的网络资产子图；（c）核心网络资产与关键链路示意图。

2、核心网络资产和关键链路信息

该网络资产子图整体上包括六个区域（C1-C6）的网络资产，包括两个核心网络资产，如图 4.4（c）所示。

结合图 4.4（a）、图 4.4（b）可知，C1、C3、C4、C6 区域内的网络资产分布情况类似，均包含了很多蓝色的域名（Domain）节点，这些节点通过 r_dns_a 类型的边共同连接到三个 IP 节点，且通过 r_cert 类型的边连接到各自不同的安全证书节点，或通过 r_whois_name、r_whois_phone、r_whois_email 类型的连边连到各自不同的红色 Whois 注册信息（Whois_Name，Whois_Phone、Whois_Email）节点。以上信息说明这些域名部署在相同的三个域名服务器上，但是各自的证书、注册信息等均不相同。根据业务规则可知，当域名有三个及以上的 IP 地址时，说明这些域名大概率使用了内容分发服务，所以域名共同指向的这些 IP 地址就不是核心网络资产。因此，这几个区域内没有核心网络资产。

C4、C5 区域内大多数都是蓝色的域名（Domain）节点，且绝大多数域名节点均通过 r_request_jump 边连接到同一个域名，且 C5 区域内的域名节点还通过 r_dns_a 类型的边连接到各自不同的一个或多个 IP 节点。以上信息说明了访问这些域名的网站时会自动跳转到同一个其他域名的网站，但是这些域名部署在不同的域名服务器上。因此，大部分域名共同跳转到的域名节点就是 C4、C5 区域的核心网络资产。

C1、C4、C5 区域间，由于各区域内均存在少量域名节点通过 r_request_jump 边连接到 C5 的核心网络资产节点而产生关联，说明访问这些域名的网站时均会自动跳转到 C5 区域的核心网络资产所对应的域名的网站。C1 与 C3 区域间、C1 与 C4 区域间、C1 与 C6 区域间、C3 与 C5 区域间，均因为区域内有少量域名节点通过 r_subdomain 或 r_dns_a 类型边关联到同一个节点而产生关联，说明不同区域内的域名存在子域名关系，或者部署在同一个域名服务器上。这六个区域间连边的关联程度很强，反映出这些网络资产关联紧密，很可能是同一个黑灰产团伙掌握的网络资产。C2、C5 的核心网络资产之间的路径就是关键链路。

4.1.4 团伙 4

图 4.5 是团伙 4 的网络资产子图，表示该黑灰产团伙所控制的网络资产。该黑灰产团伙可能进行涉黄、涉赌的黑灰产业务。总体上，该黑灰产团伙掌握的网络资产呈现出以域名，少量 IP、Whois 注册信息、安全证书为核心网络资产，多

域名与单一核心网络资产关联紧密，且核心网络资产之间通过多节点产生关联的分布模式。下面，我们分别介绍该子图的节点与边统计信息和核心网络资产与关键链路信息。

1、节点与边统计信息

该网络资产子图的节点数量 2354 个，分布情况如图 4.5（a）所示。各种类型节点的数量和比例如下：

- Domain 类型　　1923 个（81.7%）；
- IP 类型　　326 个（13.8%）；
- Cert 类型　　29 个（1.2%）；
- Whois_Email 类型　　21 个（0.9%）；
- Whois_Phone 类型　　21 个（0.9%）；
- IP_C 类型　　19 个（0.8%）；
- Whois_Name 类型　　11 个（0.5%）；
- ASN 类型　　4 个（0.2%）。

该案例的边数量 5271 条，分布情况如图 4.5（b）所示。各种类型边的数量和比例如下。

- r_dns_a 类型　　1837 条（34.9%）；
- r_whois_phone 类型　　910 条（17.3%）；
- r_cert 类型　　637 条（11.9%）；
- r_whois_name 类型　　599 条（11.4%）；
- r_whois_email 类型　　553 条（10.5%）；
- r_subdomain 类型　　522 条（9.9%）；
- r_cidr 类型　　105 条（2.0%）；
- r_asn 类型　　79 条（1.5%）；
- r_ cname 类型　　18 条（0.3%）；
- r_cert_chain 类型　　9 条（0.17%）；
- r_request_jump 类型　　2 条（0.04%）。

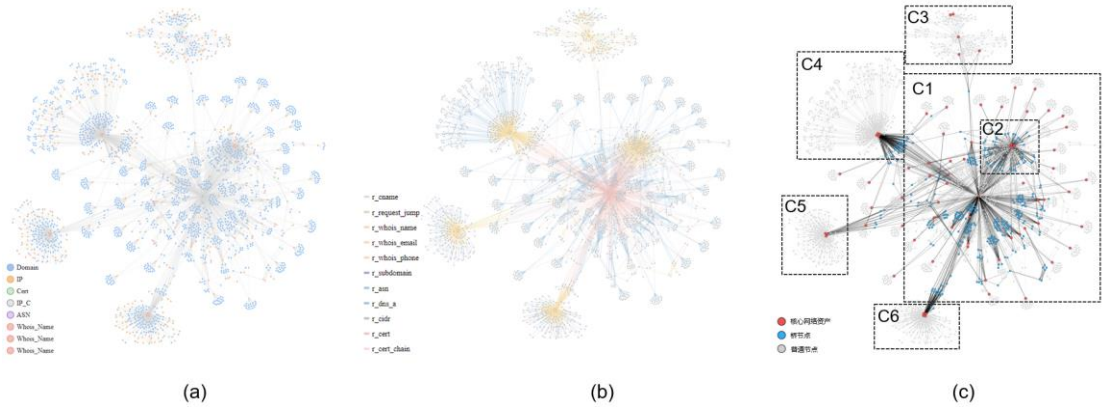


图 4.5 团伙 4 掌握的网络资产子图示例。（a）用颜色编码节点类型的网络资产子图；（b）用颜色编码边类型的网络资产子图；（c）核心网络资产与关键链路示意图。

2、核心网络资产和关键链路信息

该网络资产子图中有 73 个核心网络资产，分布在 C1-C6 区域，如图 4.5（c）所示。

结合图 4.5（a）、图 4.5（b）可知，C1 区域中，绝大部分节点是蓝色的域名（Domain）节点，这些节点均通过 r_cert 类型的边连接到同一个绿色的安全证书（Cert）节点，这些节点中，约 20-50 个域名节点为一组，通过 r_dns_a 类型的边连接到同一个 IP 节点。以上信息说明该区域的域名的网站都配置了同一个安全证书，且这些域名整体部署在不同的域名服务器上。因此，大部分域名共同关联到的证书节点就是该区域的核心网络资产。

在 C2-C6 区域中，均存在大量域名节点即通过 r_cert 类型的边连接到 C1 区域的证书，还通过 r_whois_name、r_whois_phone、r_whois_email 类型的连边连到相同的红色 Whois 注册信息（Whois_Name，Whois_Phone、Whois_Email）节点。说明这些域名的网站配置了相同的安全证书，且域名被同一个人或组织注册。因此，各区域内域名共同的注册信息，即 Whois_Name，Whois_Phone 和 Whois_Email 节点，就是核心网络资产。

C1 区域与 C2-C6 区域间，均有一些域名节点即通过 r_cert 边指向 C1 区域的证书节点（核心网络资产），又通过 r_whois_name、r_whois_phone、r_whois_email 类型边指向 C2-C6 区域内的 Whois 注册信息节点（核心网络资产），说明了区域间由于域名的网站配置了同一个安全证书且域名被同一人/组织注册而产生了紧密关联，反映出网络资产间关联紧密，很可能是同一个黑灰产团伙掌握的网络资产。因此，关联不同区域的节点就是桥节点，经过核心网络资产和桥节点的路径就是关键链路。

4.1.5 团伙 5

图 4.6 是团伙 5 的网络资产子图，表示该黑灰产团伙所控制的网络资产。该黑灰产团伙可能进行涉黄、涉赌、诈骗、非法交易平台等多种黑灰产业务。总体上，该黑灰产团伙掌握的网络资产呈现出以域名为主体，少量 Whois 注册信息、IP、域名为核心网络资产，多域名与单一核心网络资产关联紧密，且大部分核心网络资产间通过少量节点与某个核心网络资产产生关联的分布模式。下面，我们分别介绍该子图的节点与边统计信息和核心网络资产与关键链路信息。

1、节点与边统计信息

该网络资产子图的节点数量 1079 个，分布情况如图 4.6（a）所示。各种类型节点的数量和比例如下：

- Domain 类型　　970 个（89.9%）；
- IP 类型　　40 个（3.7%）；
- Whois_Phone 类型　　17 个（1.6%）；
- Whois_Name 类型　　14 个（1.3%）；
- Whois_Email 类型　　14 个（1.3%）；
- IP_C 类型　　13 个（1.2%）；
- Cert 类型　　9 个（0.8%）；
- ASN 类型　　2 个（0.2%）。

该案例的边数量 2345 条，分布情况如图 4.6（b）所示。各种类型边的数量和比例如下。

- r_whois_name 类型　　635 条（27.1%）；
- r_whois_phone 类型　　612 条（26.1%）；
- r_whois_email 类型　　592 条（25.2%）；
- r_cname 类型　　135 条（5.8%）；
- r_dns_a 类型　　116 条（5.0%）；
- r_subdomain 类型　　102 条（4.3%）；
- r_request_jump 类型　　60 条（2.6%）；
- r_cert 类型　　57 条（2.4%）；
- r_asn 类型　　16 条（0.7%）；
- r_cidr 类型　　16 条（0.7%）；
- r_cert_chain 类型　　4 条（0.2%）。

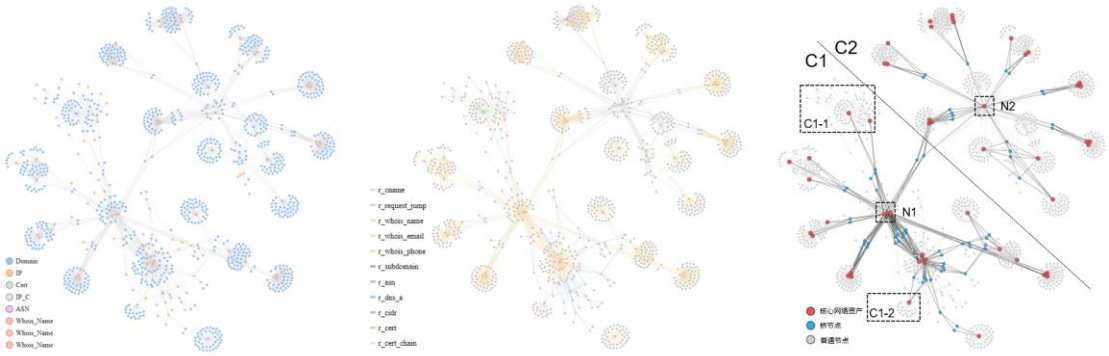


图 4.6 团伙 5 掌握的网络资产子图示例。（a）用颜色编码节点类型的网络资产子图；（b）用颜色编码边类型的网络资产子图；（c）核心网络资产与关键链路示意图。

2、核心网络资产和关键链路信息

该网络资产子图整体上可以划分为 C1、C2 两个区域，包括 50 个核心网络资产，如图 4.6（c）所示。

结合图 4.6（a）、图 4.6（b）可知，C1、C2 区域中，均存在很多域名节点通过 r_whois_name、r_whois_phone、r_whois_email 类型的连边连到相同的红色 Whois 注册信息（Whois_Name，Whois_Phone、Whois_Email）节点，说明这些域名被同一人注册。因此，这些 Whois 注册信息节点就是核心网络资产。C1 区域中，有一些域名节点通过 r_request_jump 的边与同一个蓝色域名（Domain）节点相连，还有一些域名节点通过 r_cert 的边同一个绿色安全证书（Cert）节点相连，分别如区域 C1-1 和 C1-2 所示。分别说明访问这些域名的网站时会跳转到同一个其他的网站，以及这些域名的网站部署了相同的安全证书。因此，这些域名共用的安全证书节点和共同跳转的域名节点也是核心网络资产。

C1 区域的核心网络资产之间，均存在与核心网络资产关联的少量域名节点通过 r_whois_name、r_whois_phone、r_whois_email 类型的连边与该区域的 N1 核心网络资产（Whois 注册信息节点）连接，说明这些域名使用过多个信息进行过注册。因此使得各核心资产区域之间产生紧密关联，很可能是同一个黑灰产团伙掌握的网络资产。

C2 区域的核心网络资产之间，均存在与核心网络资产关联的少量域名节点通过 r_dns_a 类型的连边与 C2 区域的 N2 的核心网络资产（IP 节点）连接，说明这些域名部署在相同的域名服务器上，使各核心资产之间关联紧密，很可能是同一个黑灰产团伙掌握的网络资产。

C1 区域和 C2 区域之间，有 2 个域名节点即通过 r_whois_name、r_whois_phone、r_whois_email 类型的连边连到 N1 的 Whois 注册信息节点，又通过 r_dns_a 类型

的边连接到 N2 的 IP 节点。说明这些域名的网站即具有 Whois 注册信息，又部署在 N2 的域名服务器上，使 C1 区域和 C2 区域产生紧密关联。关联各区域内和各区域间核心网络资产间的节点就是桥节点，经过核心网络资产和桥节点的路径就是关键链路。

4.2 挑战 1.2

请在黑灰产网络资产图谱数据集中挖掘不少于五个网络资产子图（与挑战 1.1 不同的子图）；识别每个子图中的核心网络资产和关键链路；用图表的形式呈现结果并简要分析每个子图对应的黑灰产团伙的网络运作机制。（请将答案尽量控制在 2000 字、10 张图片、10 个表格内）

该题目主要考察参赛者采用的可视分析方法挖掘子图、识别核心网络资产、识别关键链路的性能，只要答案合理、论述清晰、理由充分即可。建议专家结合以下几个方面进行评判。

（1）参赛者挖掘子图的规模。建议小型黑灰产团伙的网络资产子图规模在 400 个节点、800 条边以内；中型黑灰产团伙的网络资产子图规模在 800 个节点、1600 条边以内；大型黑灰产团伙的网络资产子图规模在 3000 个节点、6000 条边以内。

（2）参赛者挖掘子图的丰富程度。这种多样性主要体现在三个方面：1）参赛者挖掘到的多个网络资产子图的拓扑结构类型是否丰富；2）网络资产子图中包含的局部拓扑结构是否丰富；3）网络资产子图包含的节点与边类型是否丰富。

（3）参赛者识别核心网络资产的结果。主要关注参赛者识别的核心网络资产是否是使多网络资产间产生关联的节点；核心网络资产识别参考附录 1 提供的业务规则；参赛者是否还额外参考了其他业务规则或领域知识。

（4）参赛者识别关键链路的结果。主要关注参赛者识别的关键链路是否使核心网络资产间产生了关联；是否考虑了附录 2 提供的业务规则；参赛者是否还额外参考了其他业务规则或领域知识。

4.3 挑战 1.3

请简述采用的可视分析方法，比如：子图挖掘方法、核心网络资产识别方法、关键链路识别方法、图可视化方法、图交互分析方法等。（请将答案尽量控制在

2000 字、5 张图片、3 个表格内）

该题目是开放性问题，只要参赛者描述清晰、逻辑通顺即可。

附录 1：子图挖掘业务规则

我们提供了一些识别核心网络资产的基础业务规则以供参考，我们推荐参赛选手在实际应用过程中引入更多的业务规则及相关领域知识。

业务规则 1：建议主要在起始节点 3 跳关联内挖掘网络资产子图。

业务规则 2：建议参考边的关联强度挖掘网络资产子图。对于关联强度较弱的边指向的目标节点，不建议挖掘其 1 跳以外的节点；对于关联强度一般的边指向的目标节点，不建议挖掘其 2 跳以外节点。

业务规则 3：建议根根据实际场景丰富或过滤网络资产子图，比如：对于某些节点，允许加入与之关联的 3 跳外节点；又比如：当一个节点有成百上千个同类型邻居节点（关联关系类型也一样）时，可以适当过滤其邻居节点，以减少子图总体规模。

业务规则 4：我们推荐挖掘的网络资产子图规模如下：（1）小型黑灰产团伙的网络资产子图规模在 400 个节点、800 条边以内；（2）中型黑灰产团伙的网络资产子图规模在 800 个节点、1600 条边以内；（3）大型黑灰产团伙的网络资产子图规模在 3000 个节点、6000 条边以内。

附录 2：核心网络资产识别业务规则

我们提供了一些识别核心网络资产的基础业务规则以供参考，我们推荐参赛选手在实际应用过程中引入更多的业务规则及相关领域知识。

业务规则 1：如果某个网络资产 50%以上的邻边关联强度较弱，则该资产不被认为是核心网络资产。

业务规则 2：同时关联 2 个以上 IP 地址的 Domain 网络资产很大概率使用了内容分发网络。因此，Domain 网络资产所关联的多个 IP 地址不被认为是核心网络资产。

附录 3：关键链路识别业务规则

我们提供了一些识别关键链路的基础业务规则以供参考，我们推荐参赛选手在实际应用过程中引入更多的业务规则及相关领域知识。

业务规则 1：两个核心网络资产间长度大于 4 跳的路径不被认为是关键链路。

业务规则 2：两个核心网络资产间存在多条路径时，路径越短越重要。

业务规则 3：两个核心网络资产间路径的关联强度越强则越重要。