2022 年第九届中国可视化与可视分析大会

数据可视化竞赛 赛道 1

(ChinaVis Data Challenge 2022 - mini challenge 1)

答　卷

参赛队名称：重庆大学-唐豪-挑战 1

团队成员： 唐豪，重庆大学，1963024305@qq.com，队长

姜润枫，重庆大学，401186943@qq.com

洪宇洋，重庆大学，hongyuyang@cqu.edu.cn

覃瑾，重庆大学，3047106923@qq.com

贺思嘉，重庆大学，2741482534@qq.com

胡海波，重庆大学，haibo.hu@cqu.edu.cn，指导老师

团队成员是否与报名表一致(是或否)：是

是否学生队(是或否)： 是

使用的分析工具或开发工具(如果使用了自己研发的软件或工具请具体说明)：D3，ECharts，3D Force Graph, Three.js，Neo4j，React，Egg.js，Node，LineUp.js，Ant Design

共计耗费时间(人天)： 60 人天

本次比赛结束后，我们是否可以在网络上公布该答卷与视频(是或否)：是

(灰色字为参赛信息填写模板，请参赛者在提交时参照模板填写)

挑战 1.1：请根据附录 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。

1.1.1 问题分析

根据题述，将黑灰团伙依据节点和边的数量分为小型、中型、大型团伙，黑灰团伙内部存在诸如涉黄、涉毒、诈骗、涉枪、黑客、非法交易、非法支付等黑灰产业，团伙所掌握的网络资产依靠一定的关键链路进行互通协作。因此，对黑灰团伙的挖掘需从异常节点出发，寻找到黑灰团伙的资产链路从而确定黑灰团伙所拥有的非法网络资产规模，获取对应的网络资产子图。

1.1.2 解决方案

本作品针对大规模网络结构数据采用社区发现算法(Modularity Optimization)进行子图划分，由于黑灰团伙之间很少存在交集，同时单个黑灰团伙内部业务相似度较高，因此采用社区发现算法能很有效地将网络分割为多个单一的社区。通过社区发现算法，每个节点包含了一个独立的社区信息 ID，因此可依据该 ID 进行相应的社区检索，发掘不同线索所在社区的联系，不断依据业务类型拓展社区，直至形成一个完整的网络资产子图。

表格 1 网络资产线索社区划分

黑灰产团伙	节点 ID	节点名称	节点类型	归属社区
团伙 1(小型)	Domain_c58c149eec59bb14b0c102a0f303d4c20366926b5c3206555d2937474124beb9	c58c149eec.com	Domain	1874669
	Domain_f3554b666038baffa5814c319d3053ee2c2eb30d31d0ef509a1a463386b69845	f3554b6660.com	Domain	1874677
团伙 2(中型)	IP_400c19e584976ff2a35950659d4d148a3d146f1b71692468132b849b0eb8702c	156.241.xxx.xxx	IP	1874856
	Domain_b10f98a9b53806ccd3a5ee45676c7c09366545c5b12aa96955cde3953e7ad058	b10f98a9b5.com	Domain	1767261
团伙 3(中型)	Domain_24acfd52f9ceb424d4a2643a832638ce1673b8689fa952d9010dd44949e6b1d9	24acfd52f9.com	Domain	1874775
	Domain_9c72287c3f9bb38cb0186acf37b7054442b75ac32324dfd245aed46a03026de1	9c72287c3f.com	Domain	1909994
	Domain_717aa5778731a1f4d6f0218dd3a27b114c839213b4af781427ac1e22dc9a7dea	717aa57787.com	Domain	1874769

	Domain_8748687a61811032f0ed1dcdb57e01efef9983a6d9c236b82997b07477e66177	8748687a61.com	Domain	1756336
	Whois_Phone_f4a84443fb72da27731660695dd00877e8ce25b264ec418504fface62cdcbbd7	+1.971xxxxx	Whois_Phone	
团伙 4(大型)	IP_7e730b193c2496fc908086e8c44fc2dbbf7766e599fabde86a4bcb6afdaad66e	23.82.xxx.xxx	IP	1874934
	Cert_6724539e5c0851f37dcf91b7ac85cb35fcd9f8ba4df0107332c308aa53d63bdb	6724539e5c	Cert	1910118
团伙 5(大型)	Whois_Phone_fd0a3f6712ff520edae7e554cb6dfb4bdd2af1e4a97a39ed9357b31b6888b4af	+86.400xxxxx	Whois_Phone	
	IP_21ce145cae6730a99300bf677b83bbe430cc0ec957047172e73659372f0031b8	3.234.xxx.xxx	IP	1874650
	Domain_7939d01c5b99c39d2a0f2b418f6060b917804e60c15309811ef4059257c0818a	7939d01c5b.com	Domain	1910296
	Domain_587da0bac152713947db682a5443ef639e35f77a3b59e246e8a07c5eccae67e5	587da0bac1.com	Domain	1874650

表格 2 节点样式说明表

节点		边	
	表示 Domain 类型节点		表示 r_cname 类型边
	表示 Cert 类型节点		表示 r_subdomain 类型边
	表示 IP 类型节点		表示 r_request_jump 类型边
	表示 Email、Phone、Name 类型节点		表示 r_cert 类型边
	表示选中当前节点		表示 r_dns_a 类型边

- **黑灰团伙 1（小型团伙）**

社区 ID [1874669, 1874677]

图 1 团伙 1 线索社区图

本作品利用社区检索功能，输入线索 1 所包含的两个社区 ID，将线索所在社区展示在主视图中。主视图中各节点样式说明见表 2。从图中可看出社区 1874677(左侧子图)与社区 1874669(右侧子图)通过中间的注册人信息节点连接在了一起，因此我们将我们的关注目标确定为该注册人信息的相关节点，通过查询，我们发现该用户拥有部分非法域名，于是我们将包含非法域名的社区添加至主图之中，见图 2。

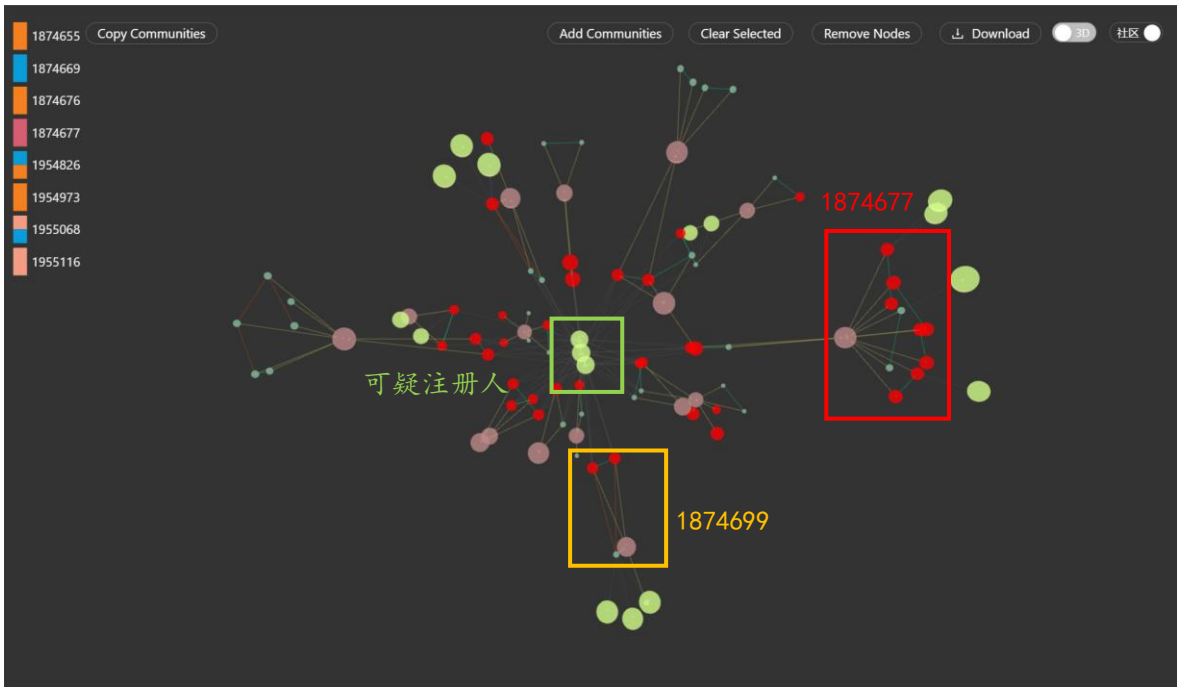

图 2 可疑注册人员联合社区图

图 3 邻居社区信息图

邻居社区图中不同高度的色块表示该社区内各种黑灰产业的数量占比，某种产业相对其他产业越多则该产业对应的色块高度越高。我们按具有相同类型产业同时异常节点数占比大于总节点数 1/2 的依据将符合条件的社区添加至主视图中，以此来获取团伙 1 的黑灰网络资产子图，最终结果见图 4。

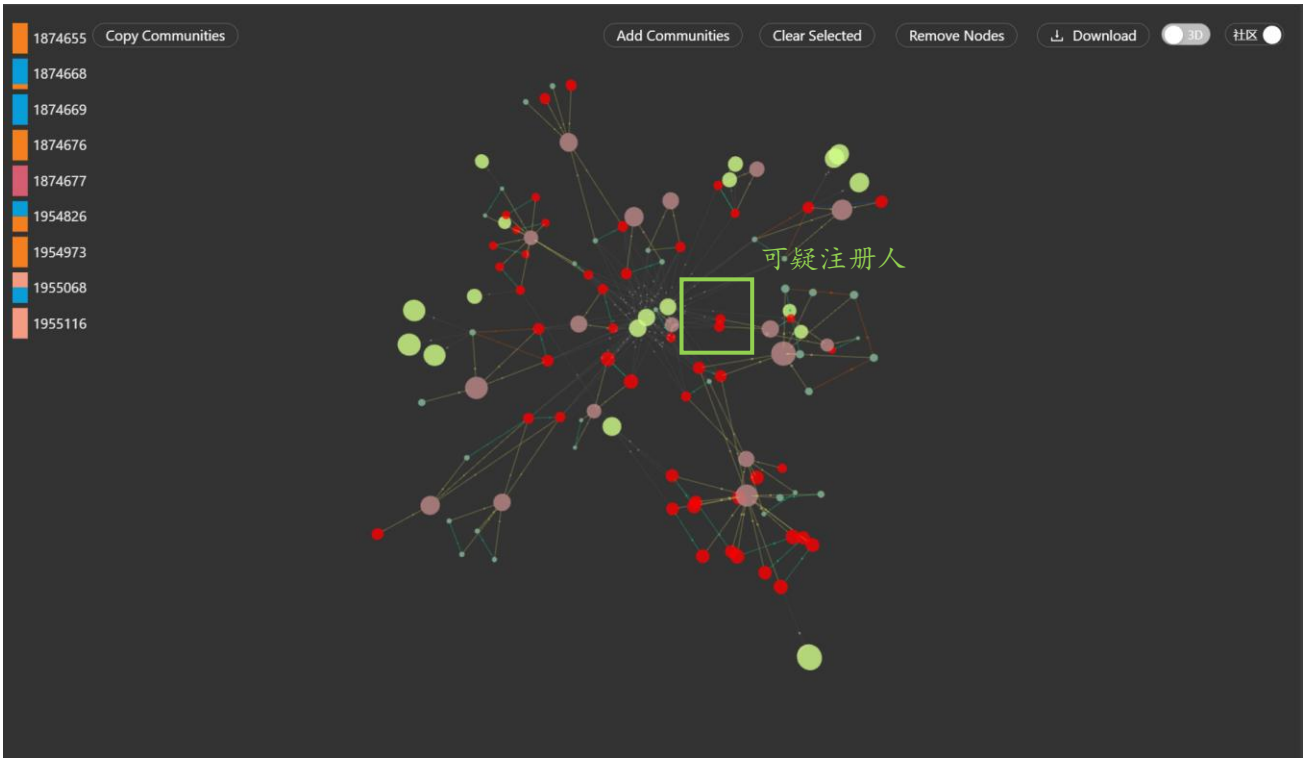

图 4 团伙 1 网络资产子图

通过团伙 1 网络资产子图，我们发现该团伙共拥有 119 个节点，其中异常节点有 50 个。其中主要的黑灰产业类型为赌博和非法交易，此外涉及少量的黄色产业和枪支。同时，我们发现该团伙主要的负责人为之前发现的可疑注册人，周围的异常节点都与之直接相关联。此外，在团伙外围也存在部分其他注册人的信息，他们均与某些异常节点相关联。因此，我们推断这些注册人应与中心的可疑注册人为同一团伙。

此外，我们发现该团伙周围存在利用内容分发的方式将其非法网站挂载在正常的网站上，以此来引诱用户进入其非法产业的现象。

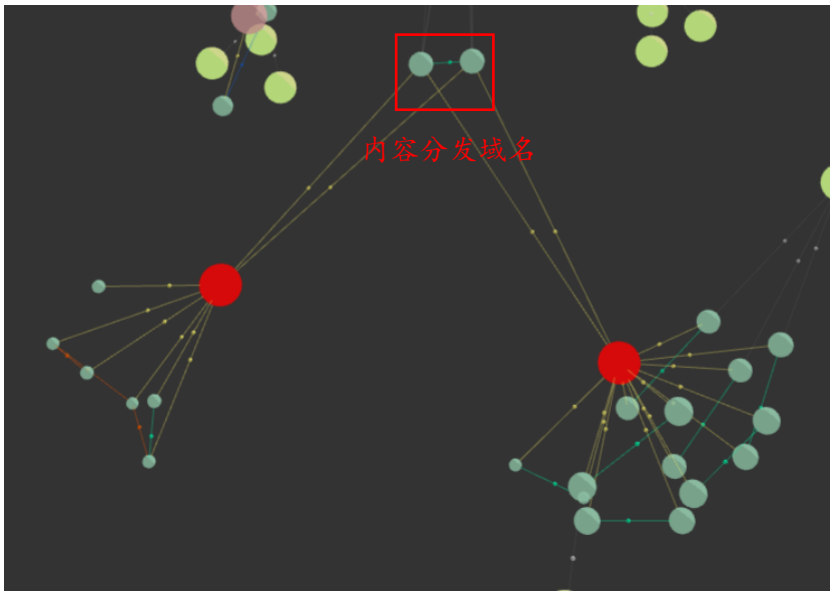

图 5 利用内容分发的方式推广非法网站

- **黑灰团伙 2（中型团伙）**

  社区 ID［1874856, 1767261］

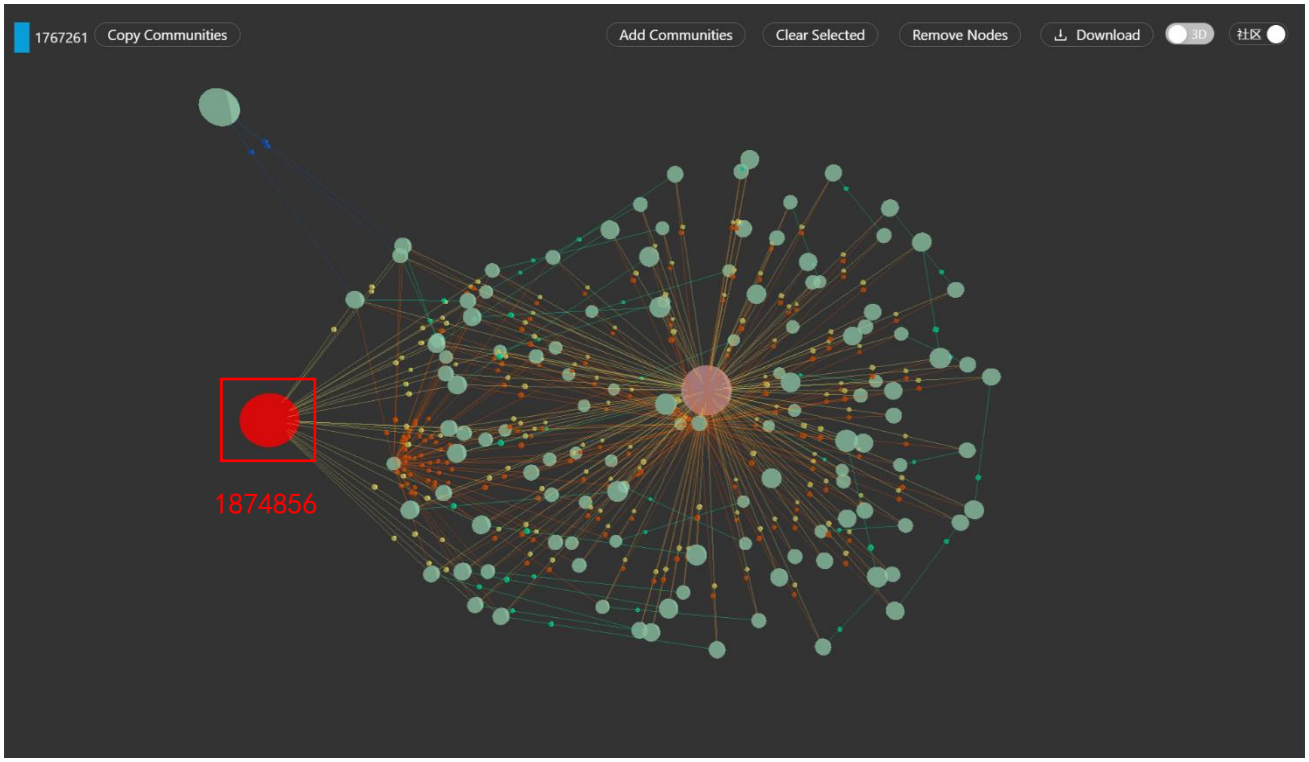

图 6 团伙 2 线索社区图

对团伙 2 的线索社区图进行分析发现，该线索社区图大部分节点都属于社区 1767261，仅线索节点 IP（156. 241. xxx. xxx）独立为一个社区 1874856，因此可以对独立的社区 1874856 进行拓展分析。

选中节点(156. 241. xxx. xxx)，节点邻居社区信息图展示该点四跳以内的社区信息。由该图可知，其周围的社区类型多为涉赌型社区，与线索类型保持一致。

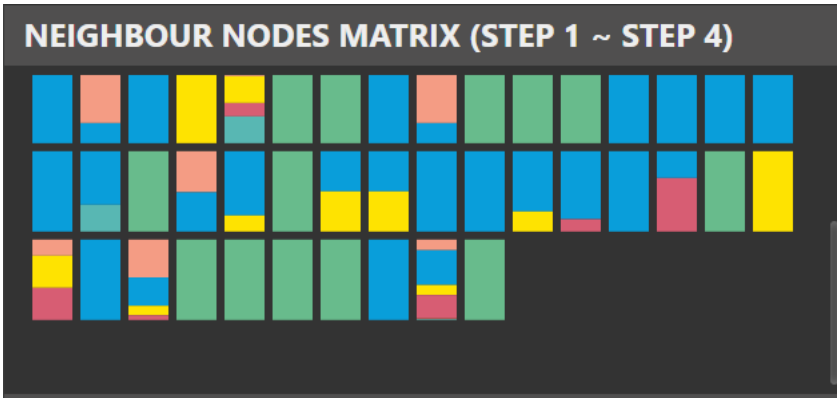

图 7 节点 156.241.xxx.xxx 邻居社区类型

我们将相同业务类型的社区添加至主视图中，最终形成黑灰团伙资产子图。异常节点在图中标红，该团伙共有 729 个节点，其中异常节点有 51 个，该团伙包含的产业类型主要为赌博，少量产业为涉黄和非法交易。

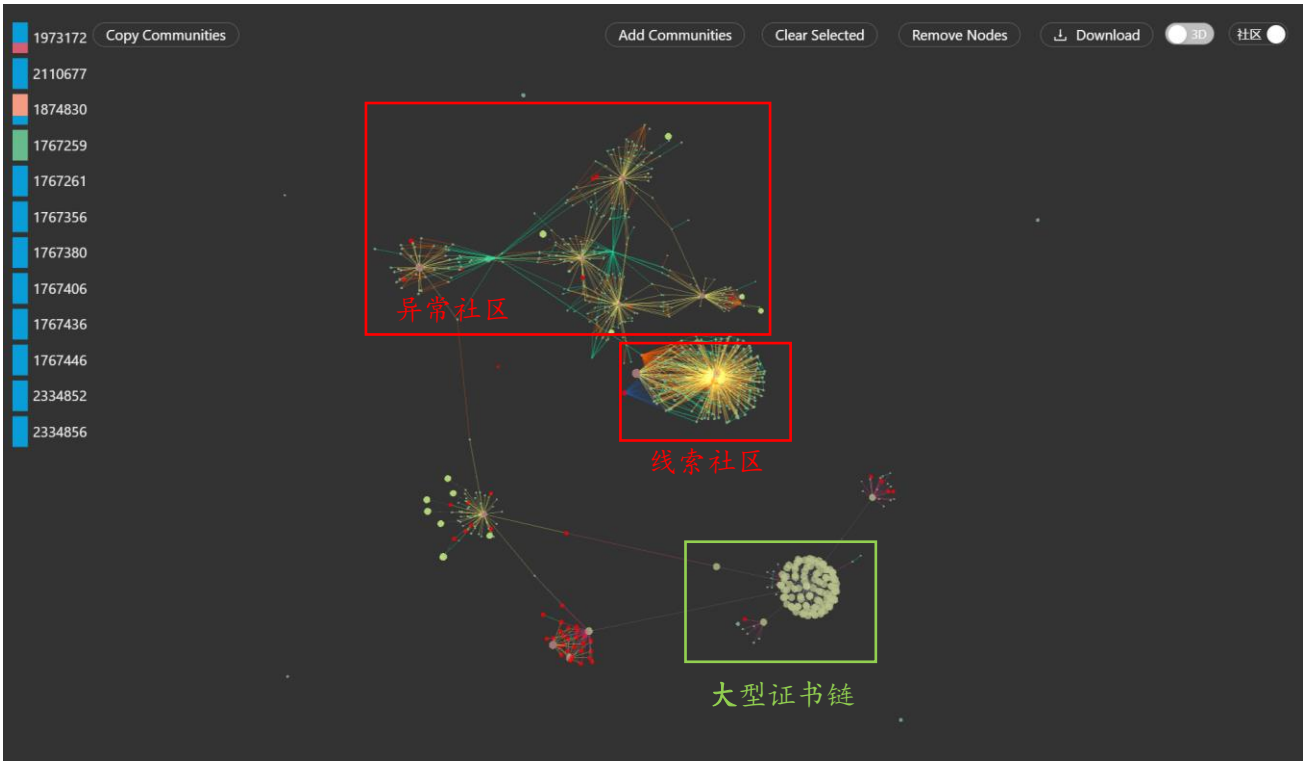

图 8 团伙 2 网络资产子图

对团伙网络资产子图进行分析可得，该团伙依靠一些正常网站作为其连接至其他非法资产的链路跳板，线索节点(b10f98a9b5. com)经由正常域名(e01c17f42e. com)跳转至一个大型的异常社区。

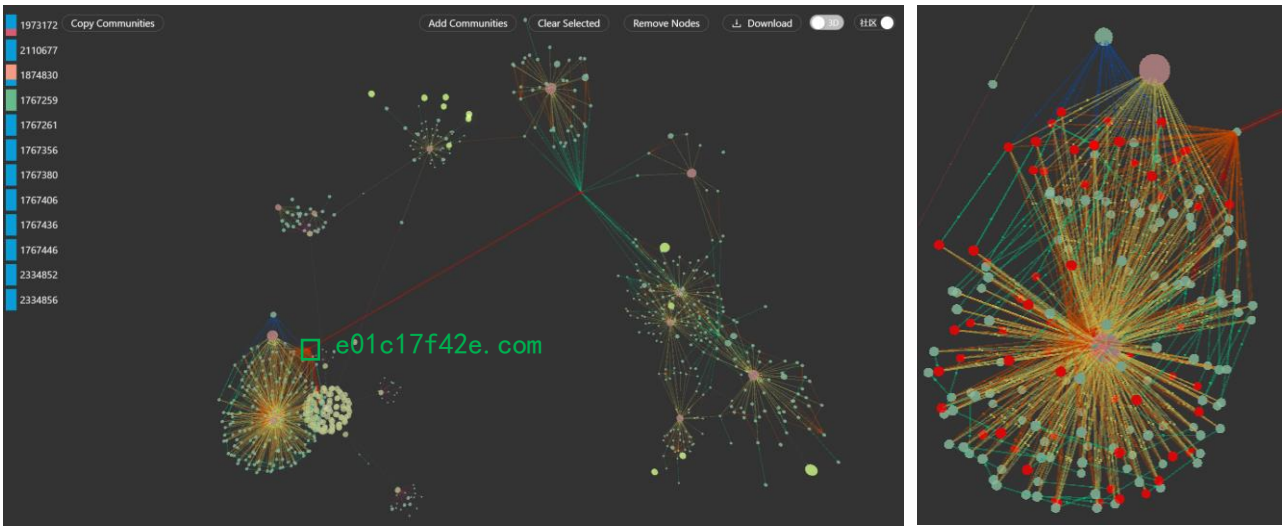

图 9 线索节点经由正常域名跳转至异常社区

图 9 右侧显示了该社区外层包裹了大量的正常域名，由图中标红所示。这些正常的域名之间互为子域名的关系，同时发现该社区还存在大量的内容分发域名，这些内容分发域名与一个正常域名(e01c17f42e.com)直接相连并依靠该域名通往其黑灰产业社区。此外我们通过不断拓展社区发现，该异常社区依赖一个大型的证书链，如图 8 所示，由于该证书链周围也聚集了大量的异常节点，因此推断，该证书链也为该团伙所拥有。

- **黑灰团伙 3（中型团伙）**

  社区 ID［1874775, 1909994, 1874769, 1756336］

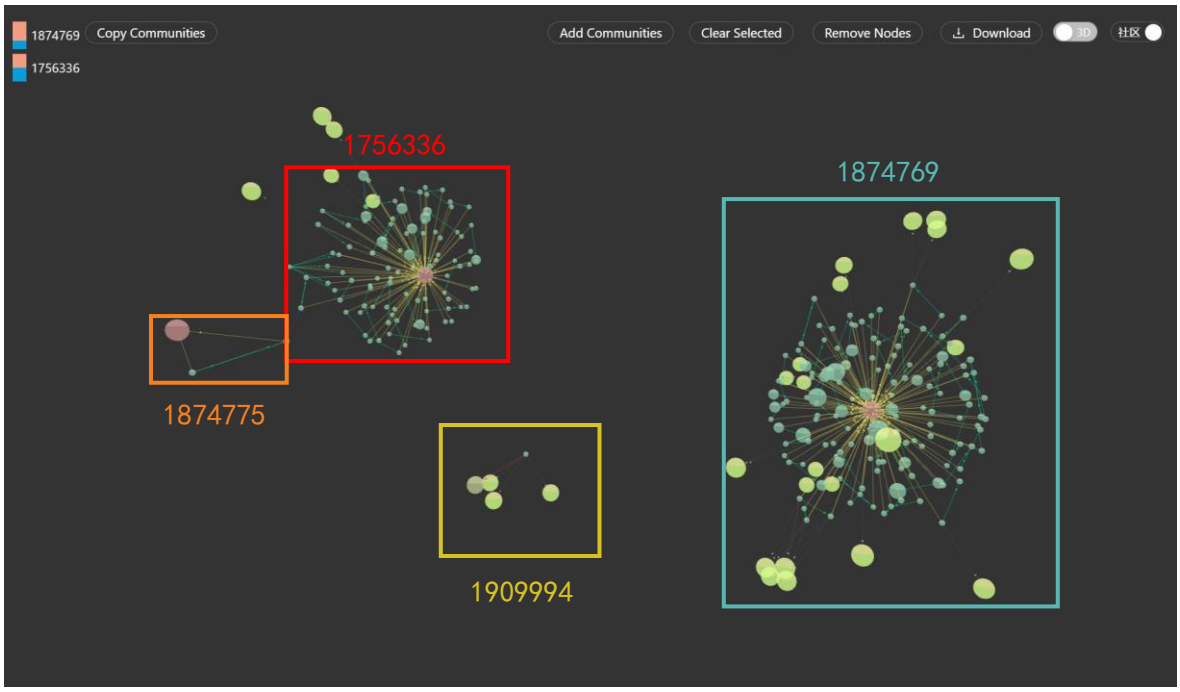

图 10 团伙 3 线索社区图

针对团伙 3，我们依据之前的划分方法，对分散的独立社区进行扩充和连接，最终形成图 11 所示的团伙 3 网络资产子图。

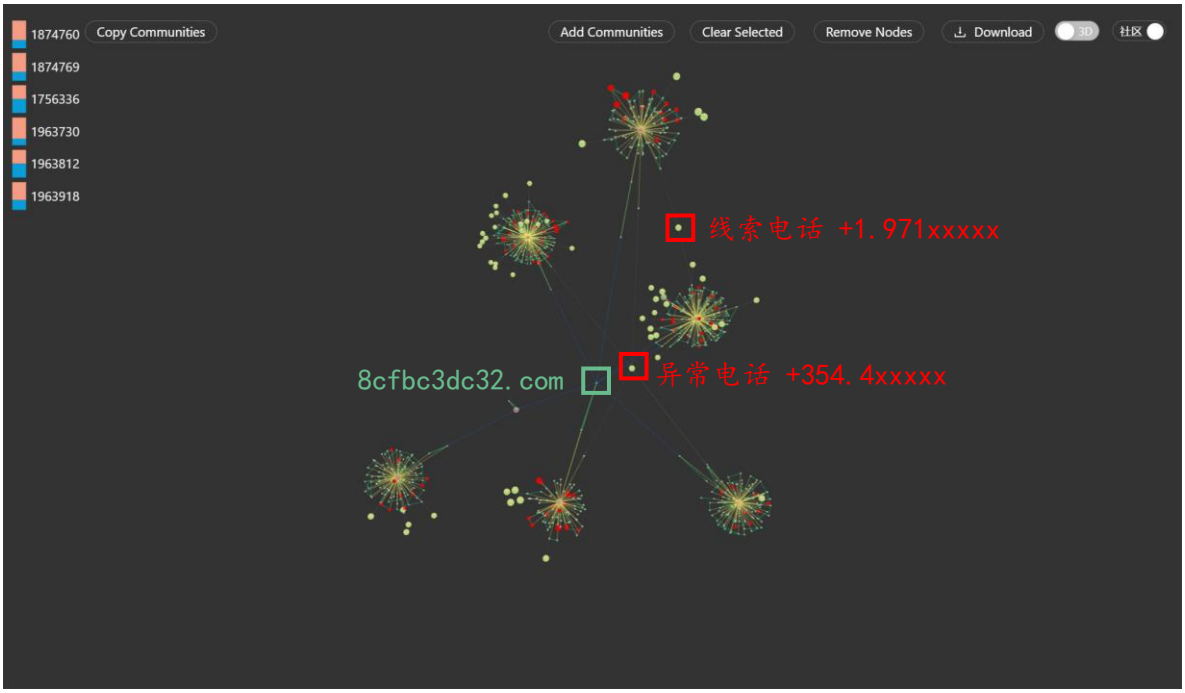

图 11 团伙 3 网络资产子图

从图中可以分析得到，该团伙与之前的团伙 2 有着相似的地方，都是采用一个正常的域名作为其跳板，诱导用户跳转至其黑灰产业之中。该团伙拥有 714 个节点，其中异常节点有 86 个，主要的黑灰产业类型为涉黄和赌博。在该网络资产子图中，我们发现多个存在异常问题的社区都经由中间的一个正常域名(8cfbc3dc32.com)进行的跳转，因此，该节点所关联的路径都可能是该团伙的关键链路，同时该域名没有同时关联两个 IP，因此排除其为内容分发网络的可能。

同时，我们发现该子图中也包含线索所提供的电话信息(+1.971xxxxx)，此电话信息同时关联了两个有异常节点的社区，该电话信息所关联的一个异常社区的同时也和另一个电话相关联(+354.4xxxxx)，该电话关联了四个不同的异常社区，因此怀疑此电话拥有者与线索提供的电话拥有者应属于同一黑灰团伙。

- **黑灰团伙 4（大型团伙）**

  社区 ID［1874934, 1910118］

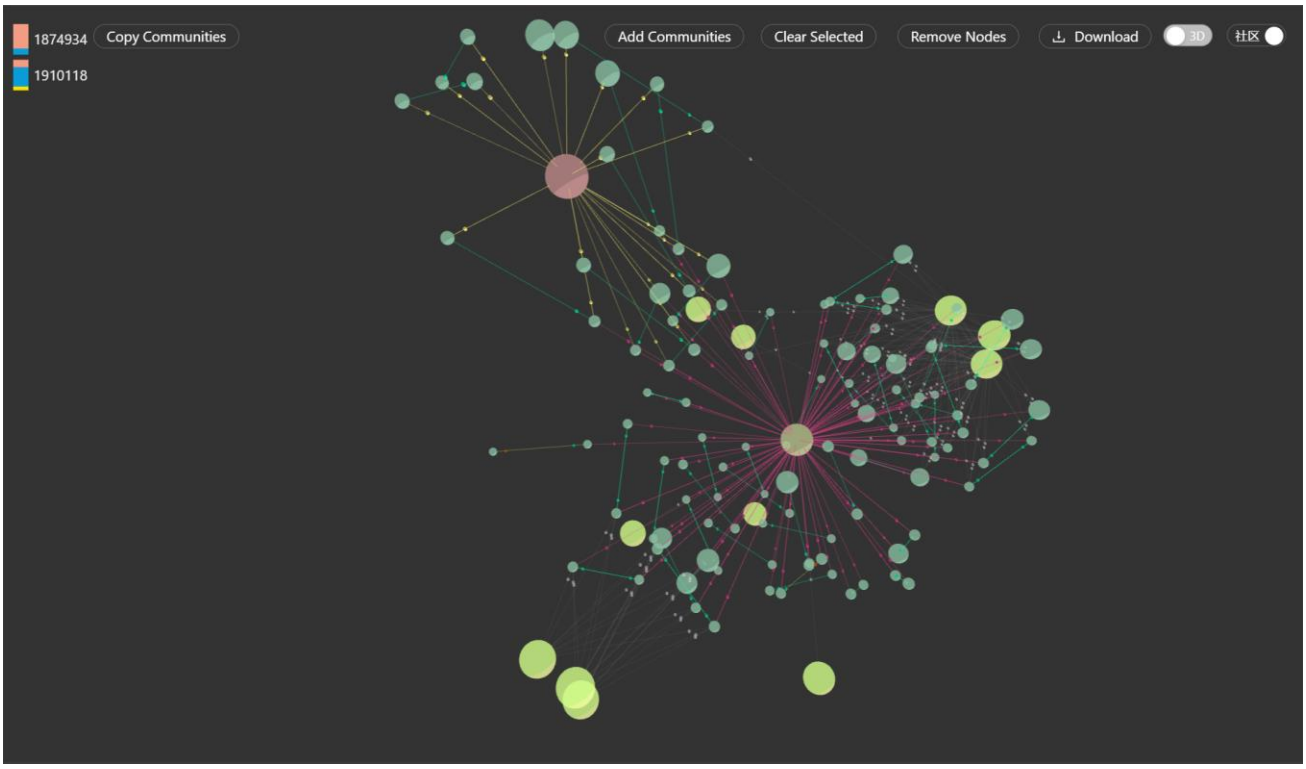

图 12 团伙 4 线索社区图

经过社区拓展，我们发现该社区与一个大型的异常社区直接相关联，于是我们将这个大型的异常社区添加至我们的视图之中。该大型社区共拥有 3737 个节点，其中异常节点高达 1469 个，主要黑灰产业类型为赌博和涉黄，其中还涉及少量的诈骗和涉枪。因此，我们推测该黑灰团伙应与这个大型社区有关。同时我们发现该社区虽然拥有大量的节点，但其中近 1000 个节点属于正常节点，不涉及黑灰产业。因此我们对这些正常的节点进行了剪枝操作，从而将联合社区的节点规模缩减到了 2079 个节点，最终的网络资产子图如图 12 所示，标红的节点为异常节点。

观察图 13 发现，线索社区经由线索 IP（23.82.xxx.xxx）连接至该大型异常社区。因此该线索证书（6724539e5c）到大型社区内部的核心证书的路径应为该团伙的一个关键链路。通过选中线索证书和大型异常社区的核心证书可以查询路径，关键链路图展示了查询所得的链路图和链路列表。

图 13 团伙 4 关键链路图和链路列表

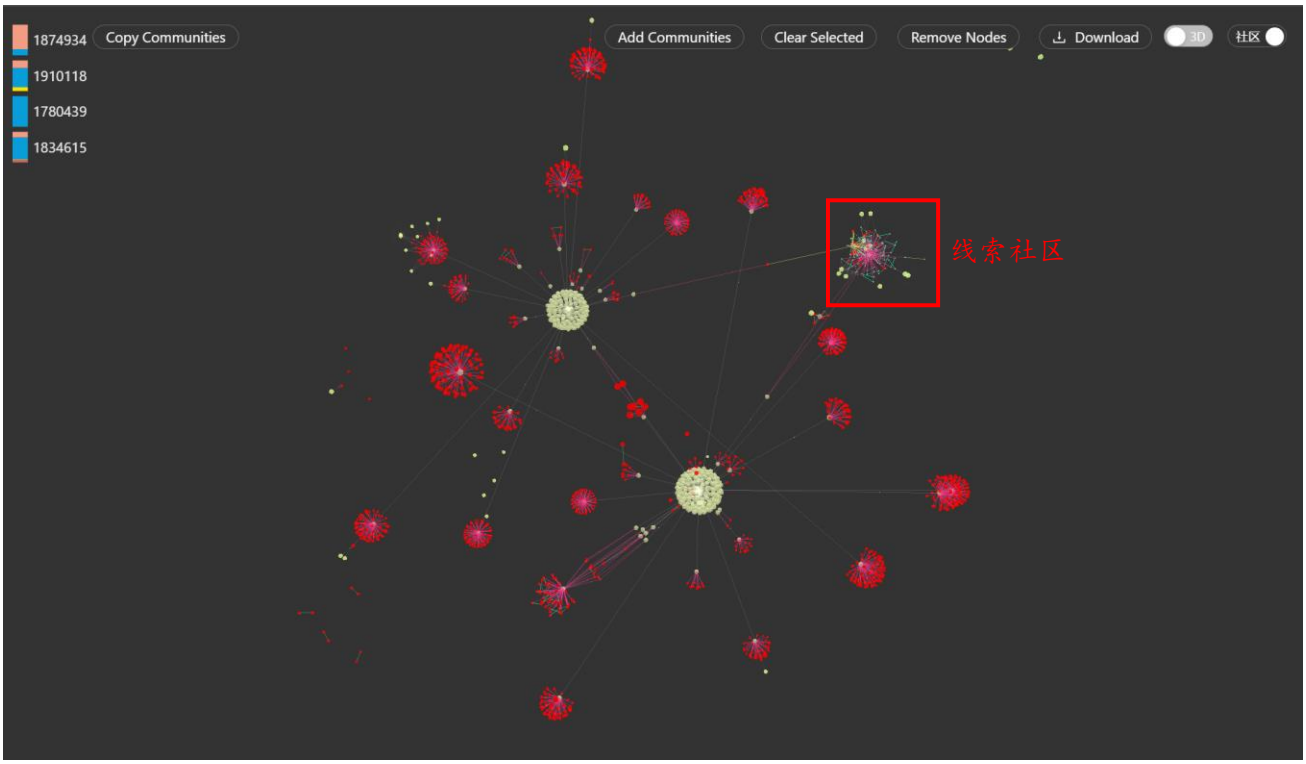

图 14 团伙 4 网络资产子图

- **黑灰团伙 5（大型团伙）**

  社区 ID [1874650, 1910296]

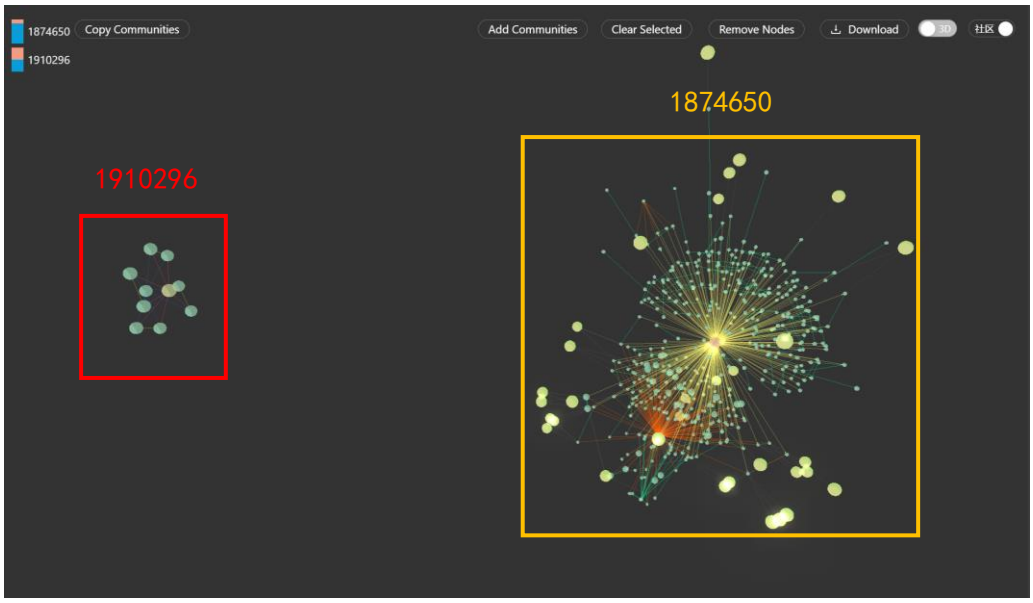

图 15 团伙 5 线索社区图

由线索社区图所示，线索节点所在社区之间并无连接。因此我们采用查询各社区核心资产间的路径来拓展社区，查询线索节点 IP(3.234.xxx.xxx)与线索 Domain 所关联的核心 Cert(050359ff6d)之间的路径，并将路径上的社区添加至主视图中，结果如图 16 所示。

图 16 团伙 5 联合社区图(左)及线索电话关联图(右)

为了更有效地拓展社区，我们查询了线索电话(+86.400xxxxx)所关联的节点图。经分析发现，该电话与 53 个 Domain 节点相关联，其中异常节点包含 62 个。通过将异常节点的社区添加至主图之中，我们发现它们与我们之前构建的团伙 5 联合社区具有直接连接，见图 17。其中线索电话已用红色高亮标记。我们发现该电话关联了多个 IP 地址，于是，我们对其关联的 IP 节点邻居社区进行了挖掘，并将业务类型一致的社区添加至主图中。

图 17 团伙 5 联合社区及线索电话关联异常社区图

同时我们发现，在该联合社区的周围存在一个异常的社区(图 18 右侧标红网络节点簇)与之直接相连，该异常社区规模为 794 个节点，拥有 532 个异常节点，主要产业类型为赌博。因此我们也将其加入了联合社区之中最终形成了我们的团伙 5 网络资产子图，如图 18 所示。

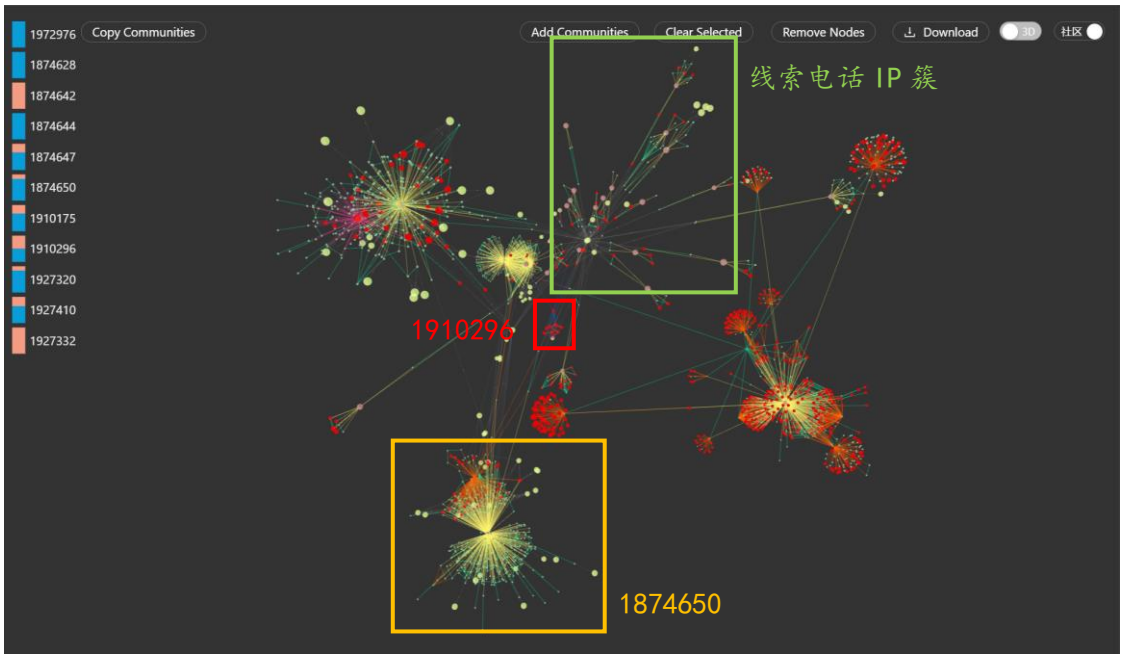

图 18 团伙 5 网络资产子图

该黑灰团伙共拥有 2180 个节点，其中异常节点为 650 个，团伙的主要黑灰产业为涉黄和赌博。我们从线索电话出发，分析该电话所拥有的 Domain(99df40f696.com)是否与大型的异常社区有关联。通过关键链路查询发现，该节点与异常社区存在两条关键路径，路径长度都为 4 跳，我们将其在主图中高亮，以此来分析其团伙运作方式。

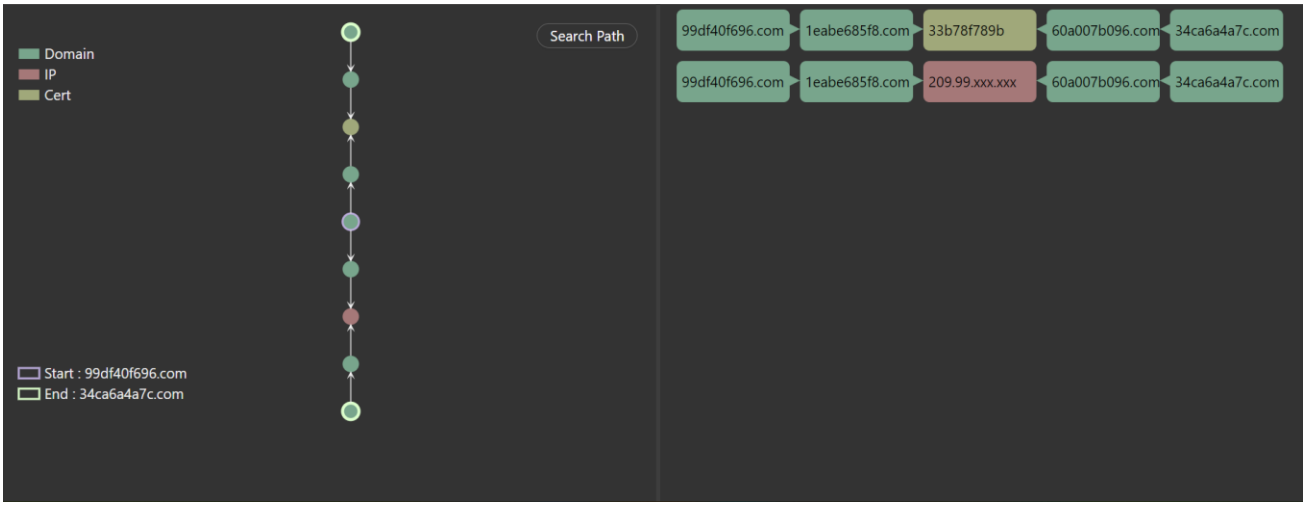

图 19 团伙 5 关键链路图及链路列表

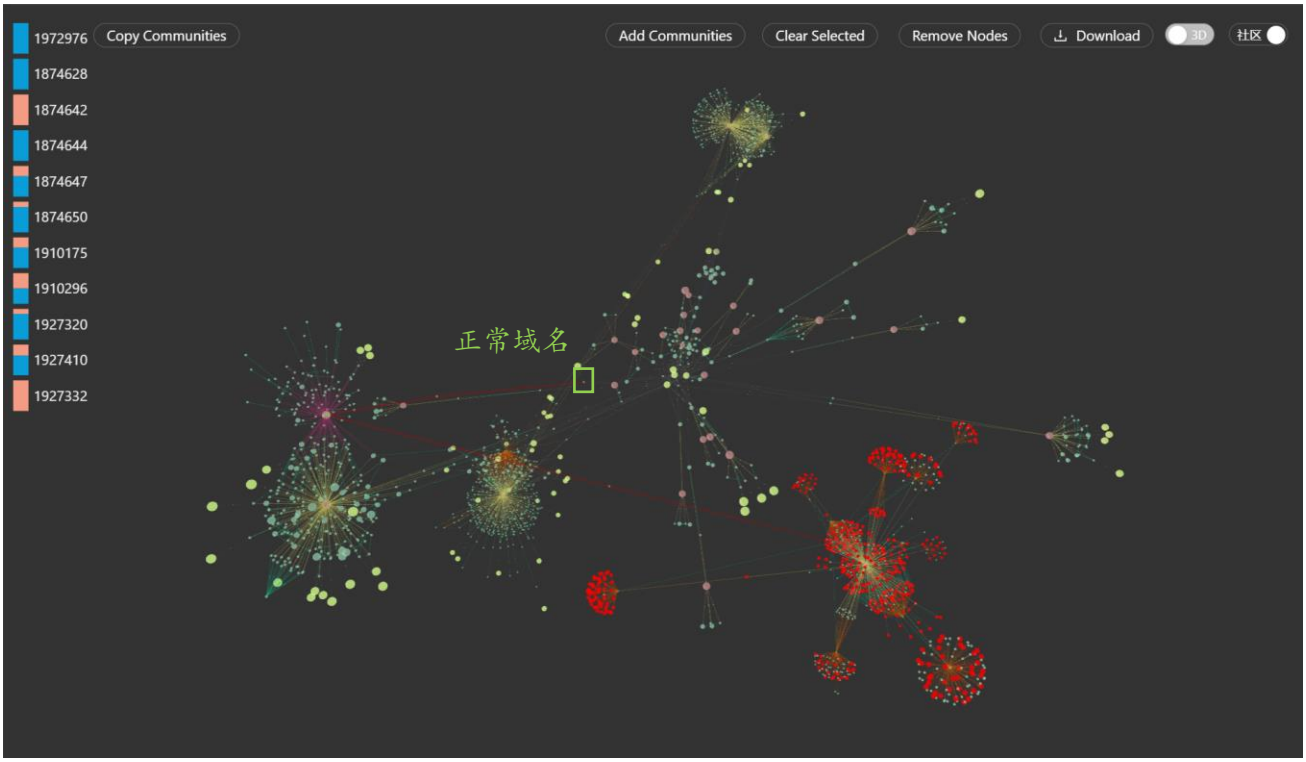

图 20 线索电话关键链路高亮展示

对高亮后的路径进行分析，发现该线索电话的拥有者利用多个正常的域名(99df40f696.com、b471916cb3.com、2d8104fd97.com)作为其连向其黑灰产业的入口，这些域名都指向一个存在问题的社区。该社区核心资产 IP(208.91.xxx.xxx)经由证书(33b78f789b)指向最终的大型异常社区，以此来构成其完整的黑灰产业链路。

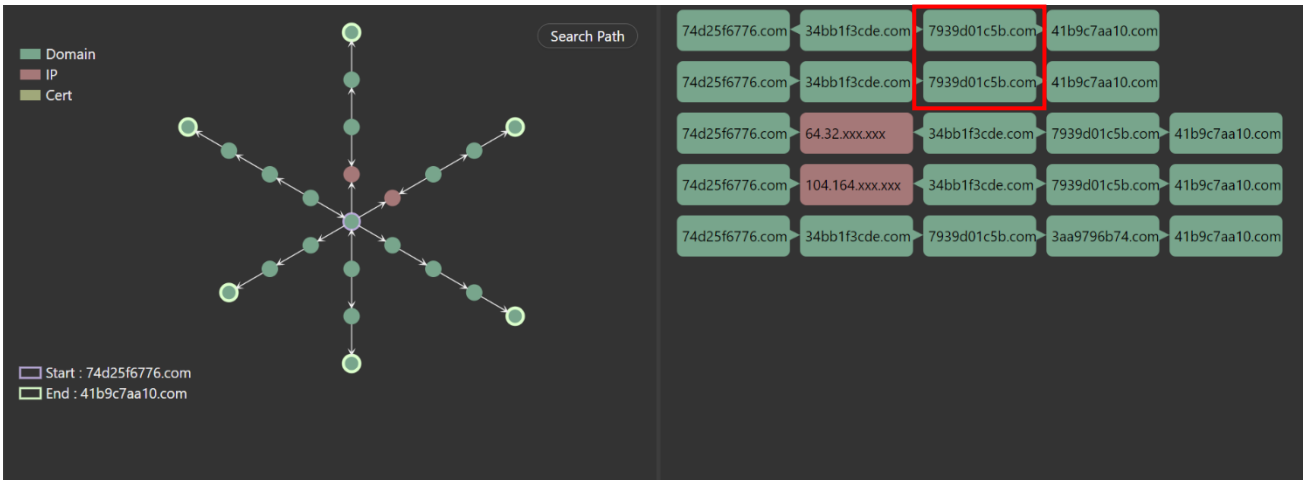

图 21 团伙 5 关键链路图及链路列表

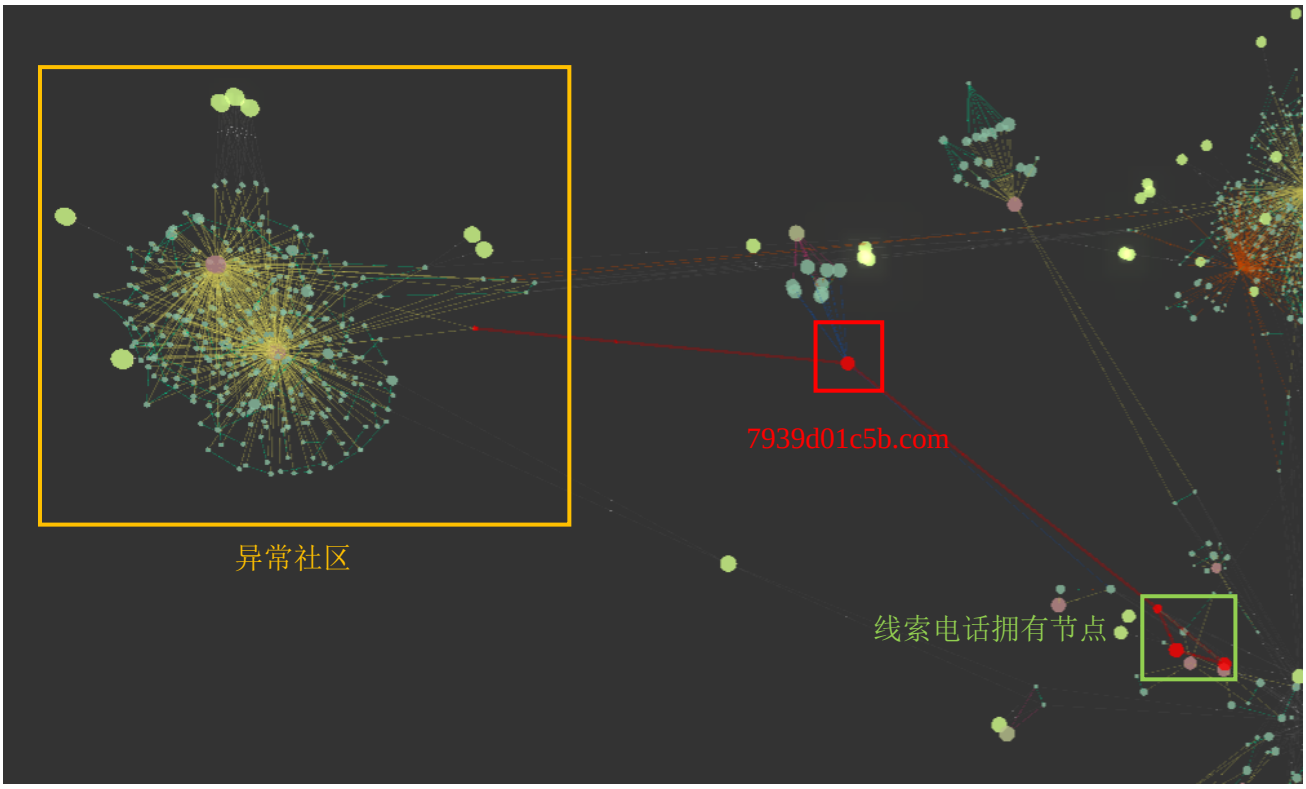

图 22 线索 Domain 所在关键链路

同时，我们注意到该团伙中，线索节点(7939d01c5b.com)还关联了另一条链路，该线索节点依靠该链路与线索电话(+86.400xxxxx)进行通信，此外该链路也与另一个异常的社区相连。因此推断该链路也与此线索电话拥有者相关联，应为该团伙的另一关键链路。

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。

为了能更好地获取不同社区的具体信息，我们构建了社区列表的可视化组件(图 23)，对我们的社区进行可视化展示。社区列表对每个社区的节点数、异常节点数、以及产业类型都进行了相应的统计；用可排序柱状图的方式来展示社区节点数和异常节点数；用热力图的方式来展示各种黑灰产业在该社区的占比。同时社区列表也会对该社区的邻居社区进行数据展示以帮助我们快速锁定独立的异常社区。

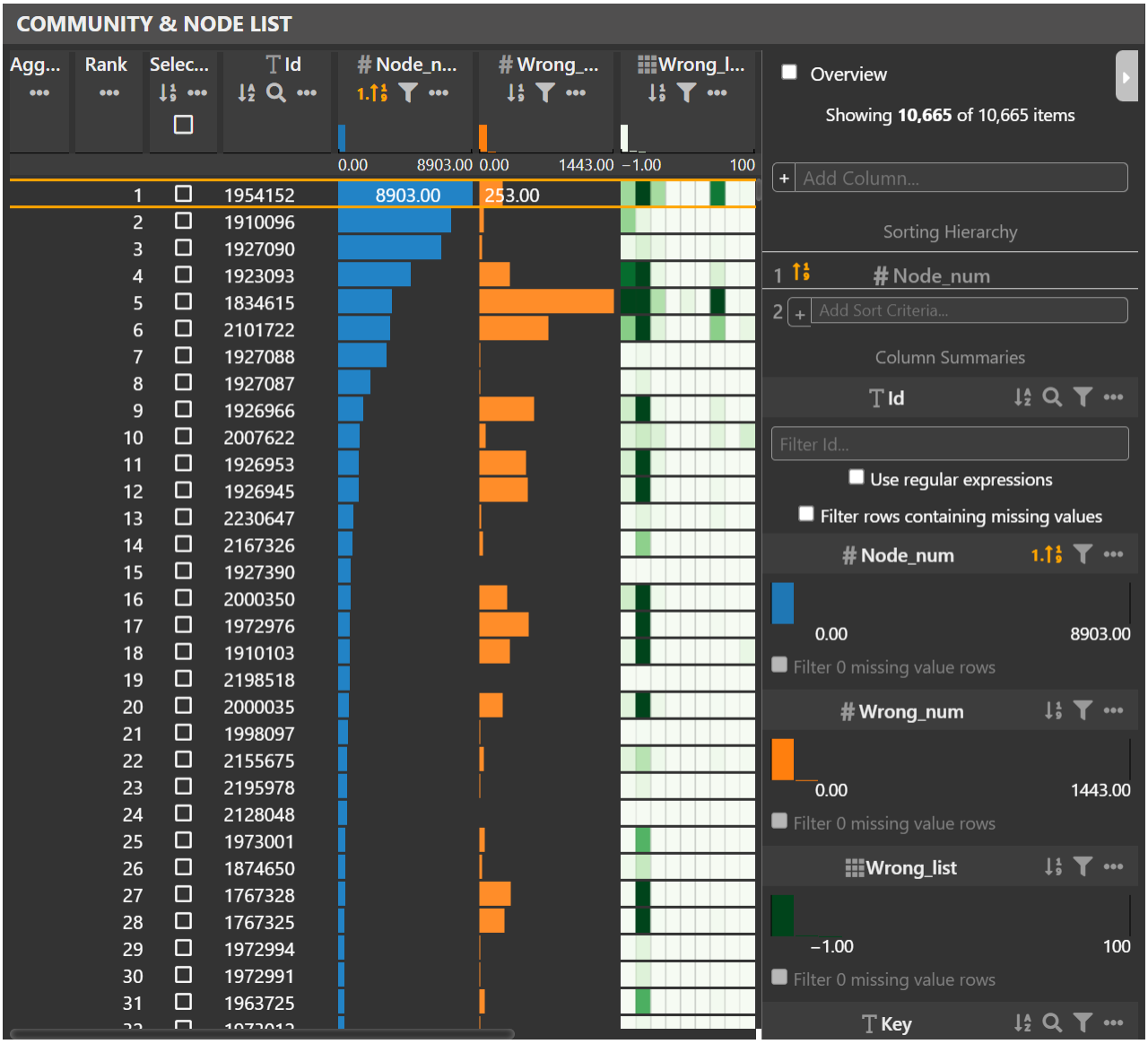

图 23 社区列表图

➢ 黑灰团伙 1（小型团伙）

我们在探索社区列表时发现一个小型社区拥有 62 个节点但其异常节点高达 51(图 24)，因此我们选择该社区作为了我们的观察对象。将该社区添加至主视图中，我们发现该社区主要由一个连接大量非法域名节点的 IP 资产构建而成。

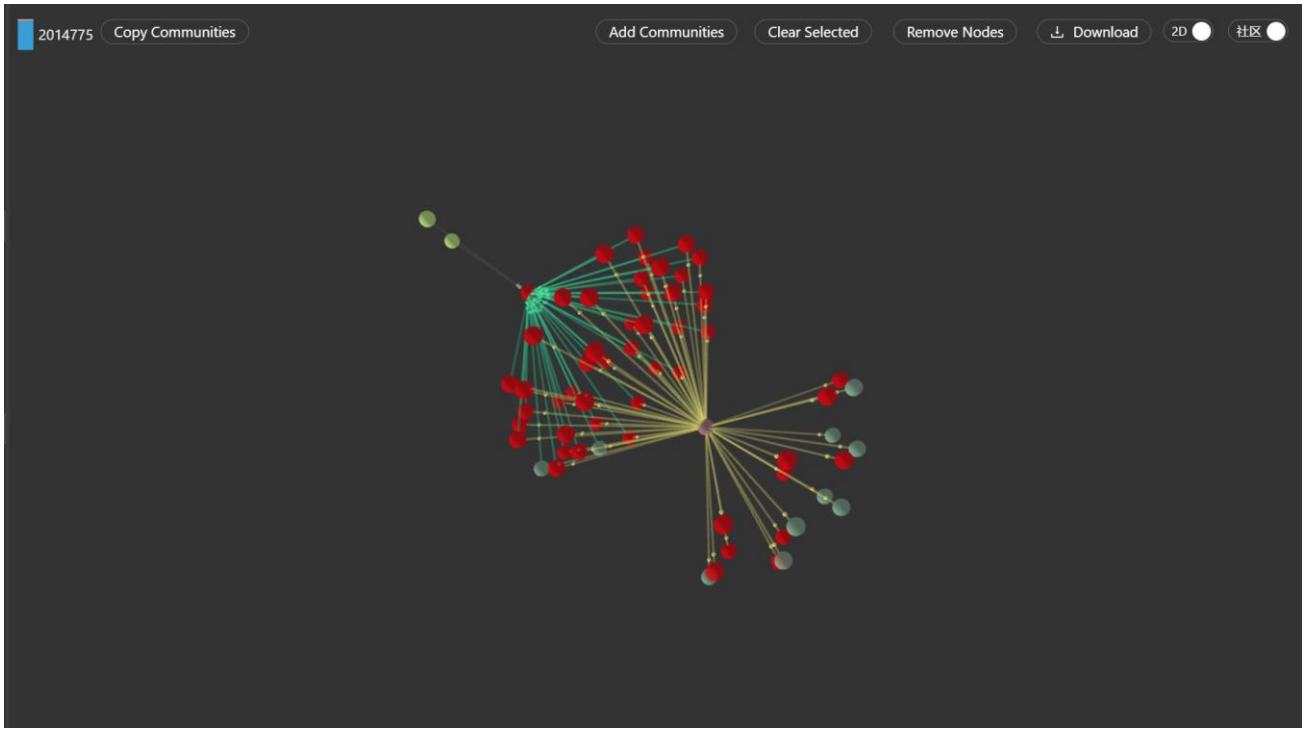

图 24 社区 2014775

在重复之前探索的过程中，我们发现该社区周围仅存在一个邻居社区(图 25)，且添加此邻居社区之后便不存在其他邻居社区。因此推测该团伙与外界相独立，很难通过常规方式查找到。

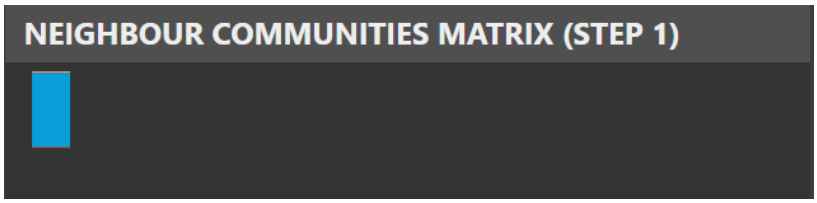

图 25 社区 2014775 邻居社区图

我们将其邻居社区加入当前主图中，获得了该小型黑灰团伙最终的网络资产子图(图 26)。分析得知，该黑灰团伙共有 162 个节点，其中非法节点高达 146 个，主要的产业类型为赌博。

为了探索该小型团伙的关键链路，我们将目光锁定到了该团伙内部的两个核心 IP 上。选中两个 IP(156. 235. xxx. xxx、156. 245. xxx. xxx)，查询两 IP 节点间的关键链路，并由关键链路图和链路列表来展示查询的结果。关键链路查询结果表示，该团伙依靠多条关键链路运作其黑灰产业，链路跳数主要为 3 跳和 4 跳，且所有的关键链路都会通过三个异常域名(fdbc1f1e07. com、8e6287744f. com、11e2201478. com)。因此如若对这三个域名进行封锁，能有效对该黑灰团伙的运作进行打击。此外，该团伙还包含了可能与此黑灰团伙相关系的注册人信息，其信息见表格 3。

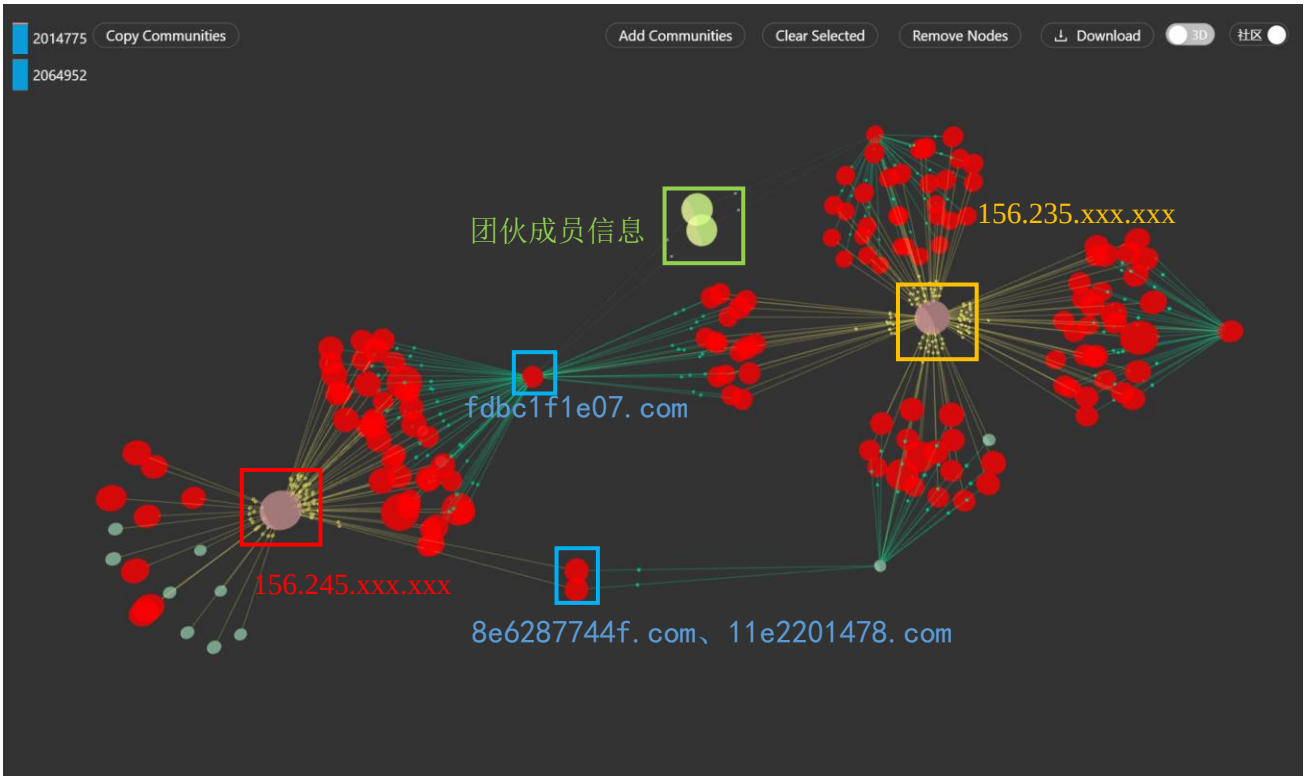

图 26 黑灰团伙 1 (小型)

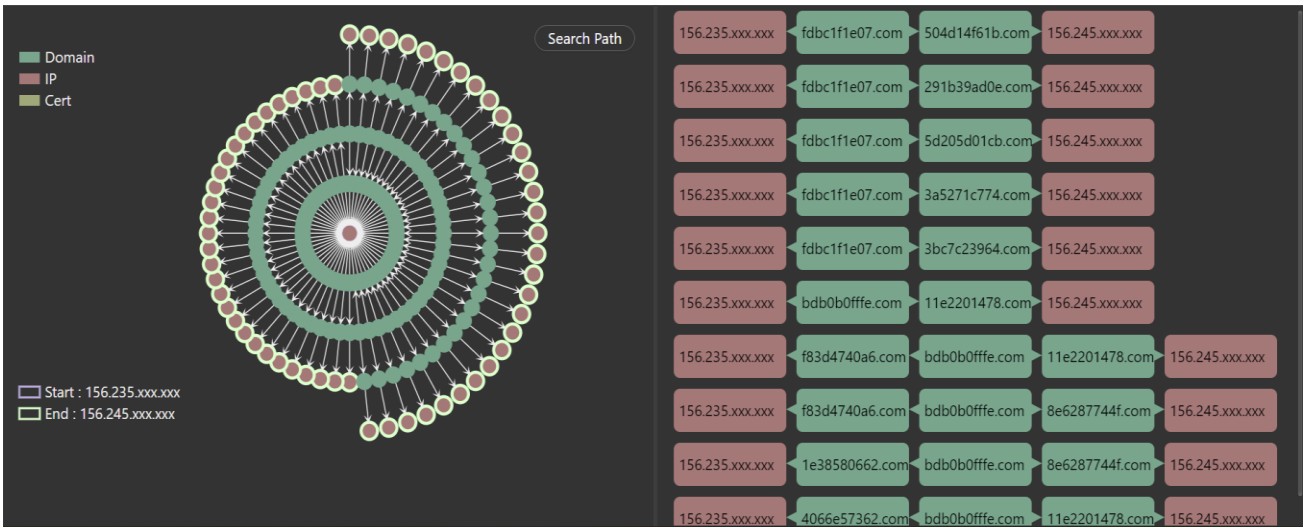

图 27 黑灰团伙 1 (小型) 关键链路图及链路列表

表格 3 黑灰团伙 1 (小型) 注册人信息表

节点 ID	节点 Name	节点类别
Whois_Email_d00bffb916dd17942e14e840058f853f78c3bc68cc1d52916cdee541e528f9e2	31857xxxxx@xxx.xxx	Whois_Email
Whois_Name_59757cfdc9e67c0beabce02094a43dc4c8138688efbb7ff724a1288c723047a2	张 xxxxx 勇	Whois_Name

➢ 黑灰团伙 2（小型团伙）

我们依据发现黑灰团伙 1 的方法继续探索社区列表，发现了一个与黑灰团伙 1 类似社区（图 28）。该社区节点数为 58，异常节点高达 57，仅有一个正常节点，因此我们将其作为了我们团伙 2 的探索目标。

我们将该社区添加至主图之中，发现该黑灰团伙由一个被非法域名包裹的证书节点构成，其主要黑灰产业为赌博。

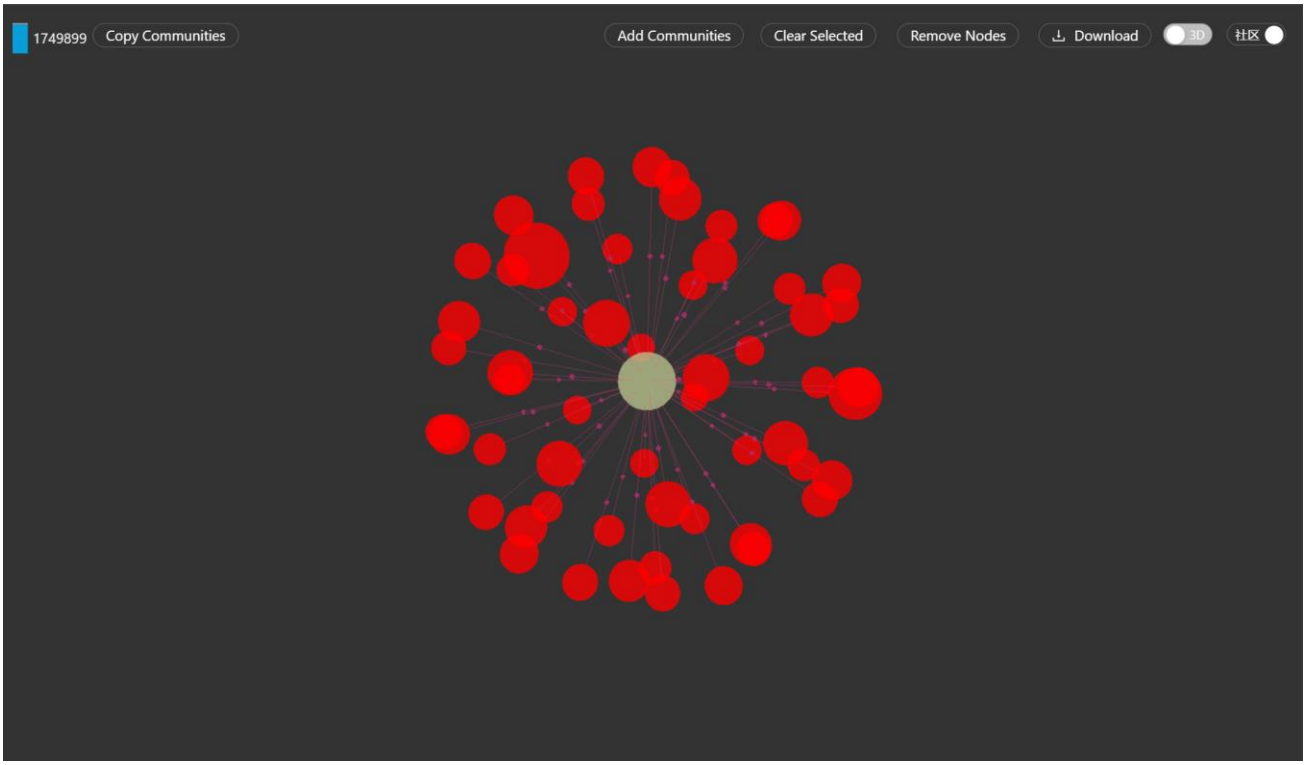

图 28 社区 1749899

通过观察该社区的邻居社区图发现，该社区周边都为与之类似的赌博型社区（图 29），于是我们将其邻居都添加至主图中，最终得到了该黑灰团伙网络资产子图（图 30）。

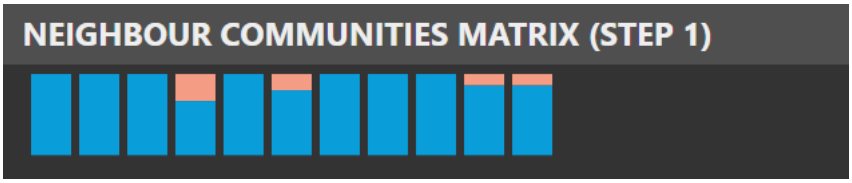

图 29 社区 1749899 邻居社区图

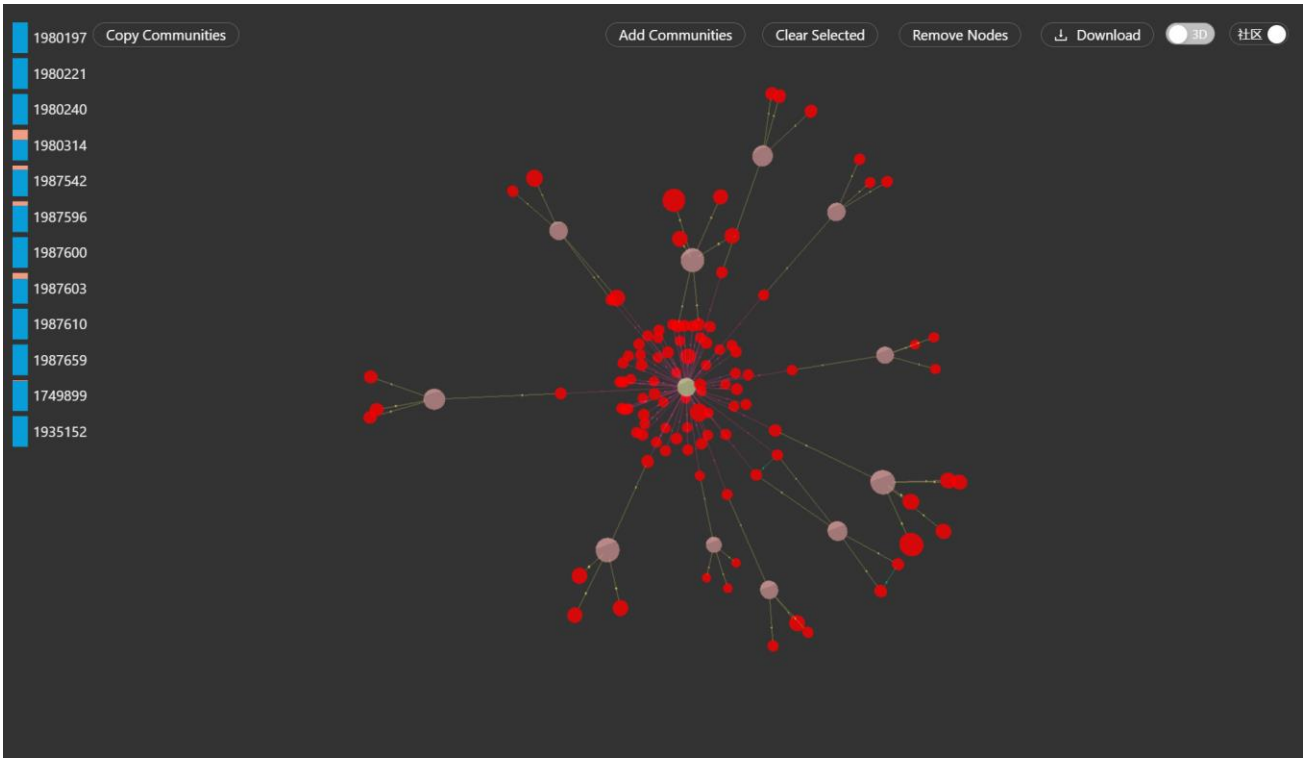

图 30 黑灰团伙 2 (小型)

经分析，该黑灰团伙共有 117 个节点，其中异常节点数目为 105 个，黑灰产业主要为赌博，涉及少量的黄色产业。由图中可明显看出，该黑灰团伙的运作方式为通过大量的边缘 IP 网络连接至中心证书。来引导用户从事赌博活动，因此其关键链路应与中心证书(cfb1f76dff)相关。由于图中每个外部 IP 节点至其中心证书节点跳数都在 4 跳以内，因此我们认为所有外部 IP 至其中心证书的路径都为该团伙的关键链路。

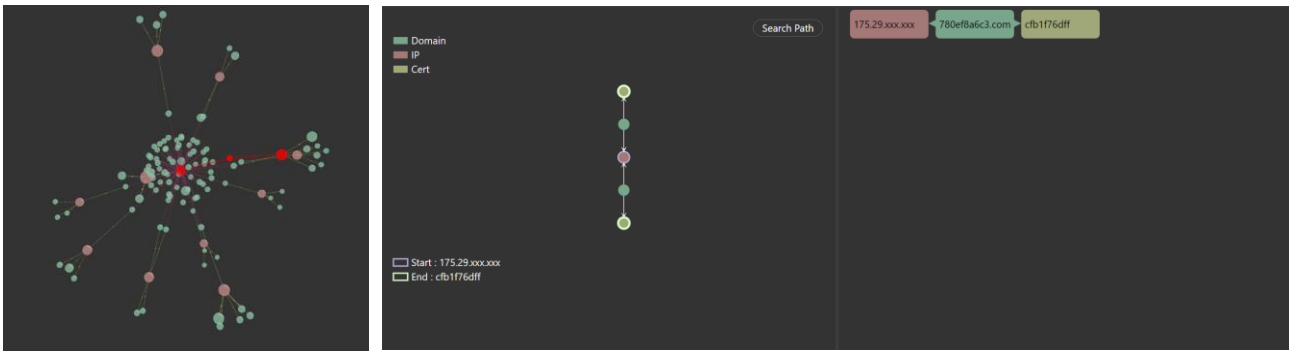

图 31 黑灰团伙 2 关键链路示意

此外，我们从降维视图(图 32)中发现该团伙构成社区呈现出高度的聚集态，这也间接说明了我们寻找的社区在一定程度上具有一定的相似性，符合我们的预期。

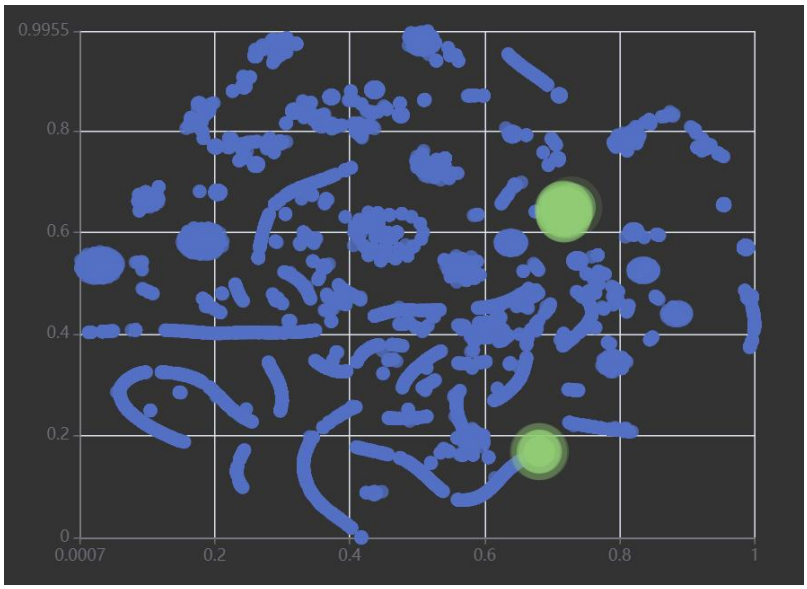

图 32 团伙 2 社区降维图

- **黑灰团伙 3（中型团伙）**

我们依据发掘团伙 1 和 2 的方法，在社区列表中选择了一个较小的社区(262678)进行探索，该社区共有 4 个节点，其中异常节点数为 2(图 33)。

图 33 社区 262678

依据之前的探索步骤，最终形成图所示的团伙 3 网络资产子图(图 34)。

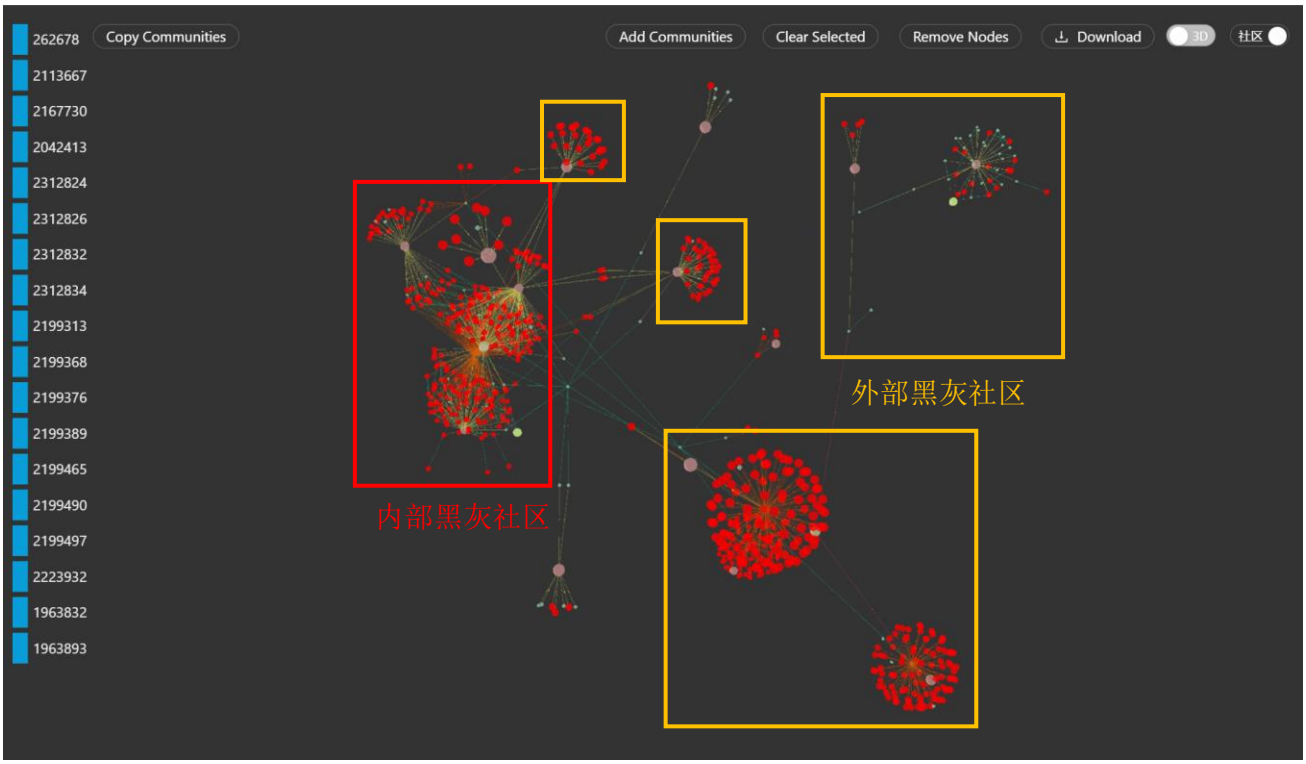

图 34 黑灰团伙 3 (中型)

如图可以分析出，该团伙与之前的团伙相似，均有正常的域名作为跳板。不过，与之前发现的团伙不同之处在于该团伙存在外部黑灰社区和内部黑灰社区。从图中可明显看出内部黑灰社区具有高度的聚集性，而外部的黑灰社区之间较为独立。此外这些外部社区都是通过多个正常域名跳转至内部社区之中。

因此，我们推测该黑灰团伙的运作方式可能是先通过外围的赌博社区吸引客户，再将客户引导至中心的内部黑灰社区之中。由于有正常的社区作为桥梁，即使外部的黑灰社区被查封，内部的黑灰社区也不会受到影响。该黑灰团伙共有 797 个节点，其中问题节点有 705 个，该黑灰团伙的主要营业类型为赌博行业。核心资产包含 15 个 IP，2 个证书以及大量的子域名。

由于外部社区需要通过正常域名跳转至内部的核心社区，因此我们将外部社区到内部社区的通路视为该团伙的关键链路。由于图中存在大量的关键链路，因此仅对部分链路进行了示意，但最终我们发现这些关键链路都会通过共同的几个域名。因此我们提供以上域名信息帮助相关人员对其进行分析，域名信息见下表。

表格 4 团伙 3 关键链路核心域名信息表

节点 ID	节点 Name	节点类别
Domain_c16a80b4e715b6c198383e1ab90b9292f5b00cce6c78d98528a1304264d4141b	c16a80b4e7.com	Domain
Domain_91918c2c1357bcbef57a3606562a26e797a9cde8a74c5f13f7e55d24e13bdf51	91918c2c13.com	Domain
Domain_1fe3f86041beeabbf806005d04a34c4cbec5a5dc11dbad99aa18e54296820d43	1fe3f86041.com	Domain
Domain_a831d31aa6dd7a936cc73dc9b789a8ad	a831d31aa6.com	Domain

9792e215a725bd5be5fa0a2cdb87330d		

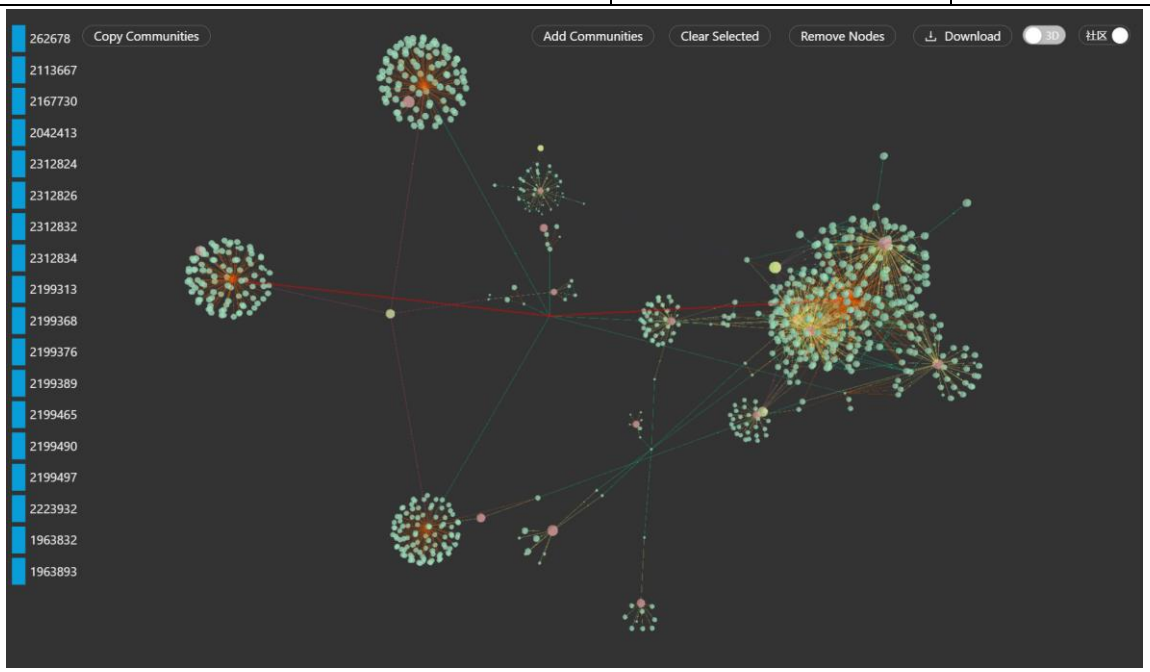

图 35 黑灰团伙 3 关键链路示意

➢ 黑灰团伙 4（大型团伙）

我们同样依据之前的方法来挖掘大型团伙。通过社区信息筛选，我们发现社区 1983184 中异常的节点数占总节点数的比重较高(图 36)。因此，我们选中此社区，并将其在主视图中显示。

通过观察邻居视图，我们发现该社区周围都以赌博和黄色产业的社区为主，但本社区的主要产业为赌博，因此我们需按照其产业类型扩充社区。

我们同样选择含有相同的业务类型的社区进行扩充。此外，扩充依据还应包含异常节点与正常节点的占比，当异常节点大于总节点数量一半且都是相同业务类型的社区才会被我们视作扩充的目标社区。依据该扩充准则，我们将当前社区的邻居社区不断添加至我们的主图中，直至邻居中已不存在可扩充的社区，则认为当前社区已饱和，不再继续扩充。最终我们得到了黑灰团伙 4 的网络资产子图(图 38)。

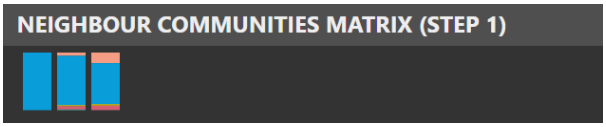

图 36 社区 1983184 邻居社区信息图

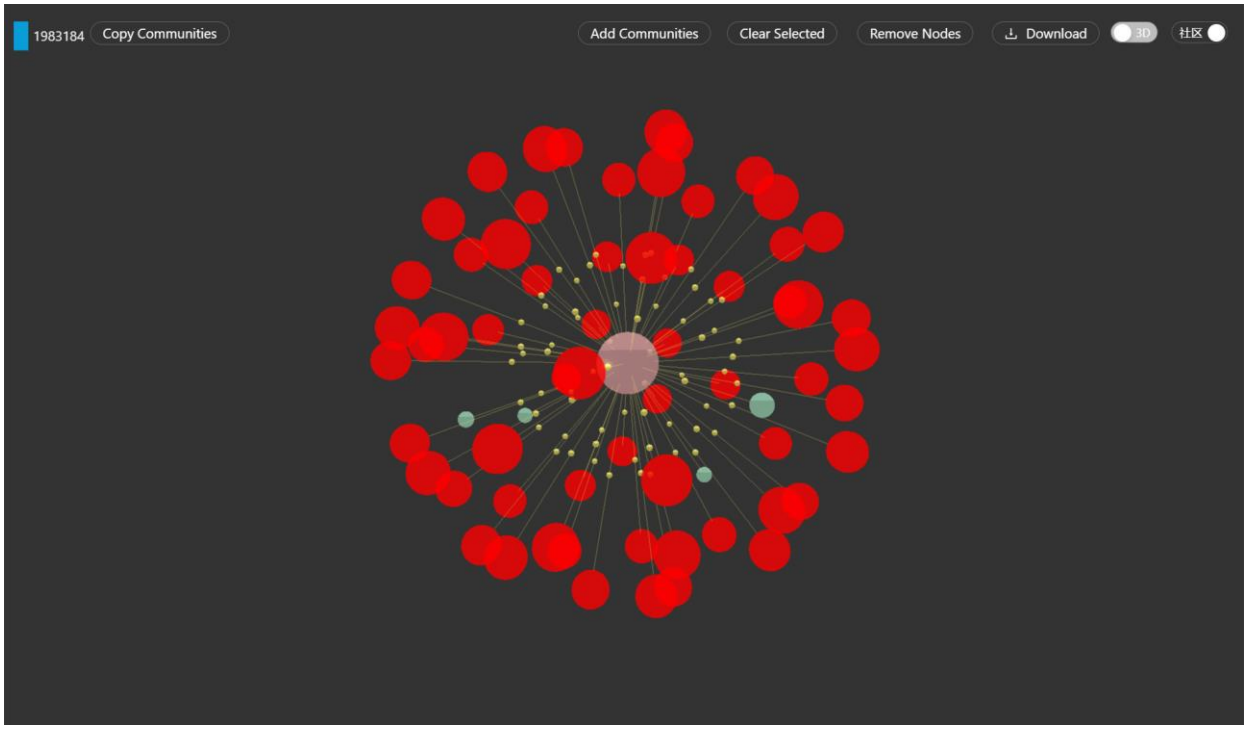

图 37 社区 1983184

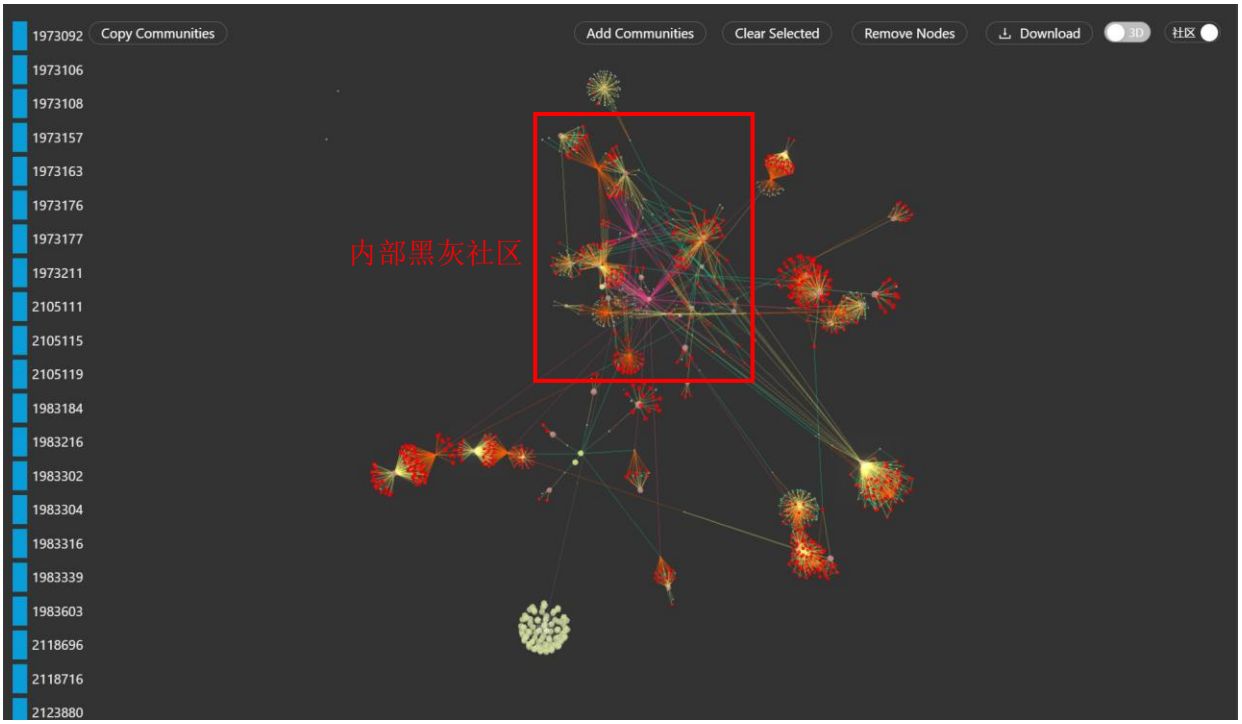

图 38 黑灰团伙 4 (大型)

该大型黑灰团伙包含 1695 个节点，其中异常节点有 983 个，该团伙的黑灰产业类型主要为赌博。由图分析可知，该团伙和之前的中型团伙类似，也具有外部黑灰社区和内部黑灰社区之分。不同的是内部的黑灰社区主要依靠两个证书运行，而非直接相连。此外，外部的黑灰社区也不是形成一个个单

一的域名集群，而是多个域名集群之间相互连接的模式。此模式与黑灰团伙的基本模式类似，这一点与我们之前发现的团伙 3 有着很大的不同。

我们推测，大型黑灰团伙除了会利用正常的域名做为周转，在外部异常节点之间的连接模式会参照自身的连接模式做改造。当相关人员按照以往黑灰产业的基本模式进行网络匹配时，很可能在完成外部社区审查后便终止审查，从而不会对内部的社区造成影响。

与之前的团伙 3 类似，我们依然考虑将外部黑灰社区通往内部黑灰社区的路径作为该团伙的关键链路(图 39)。

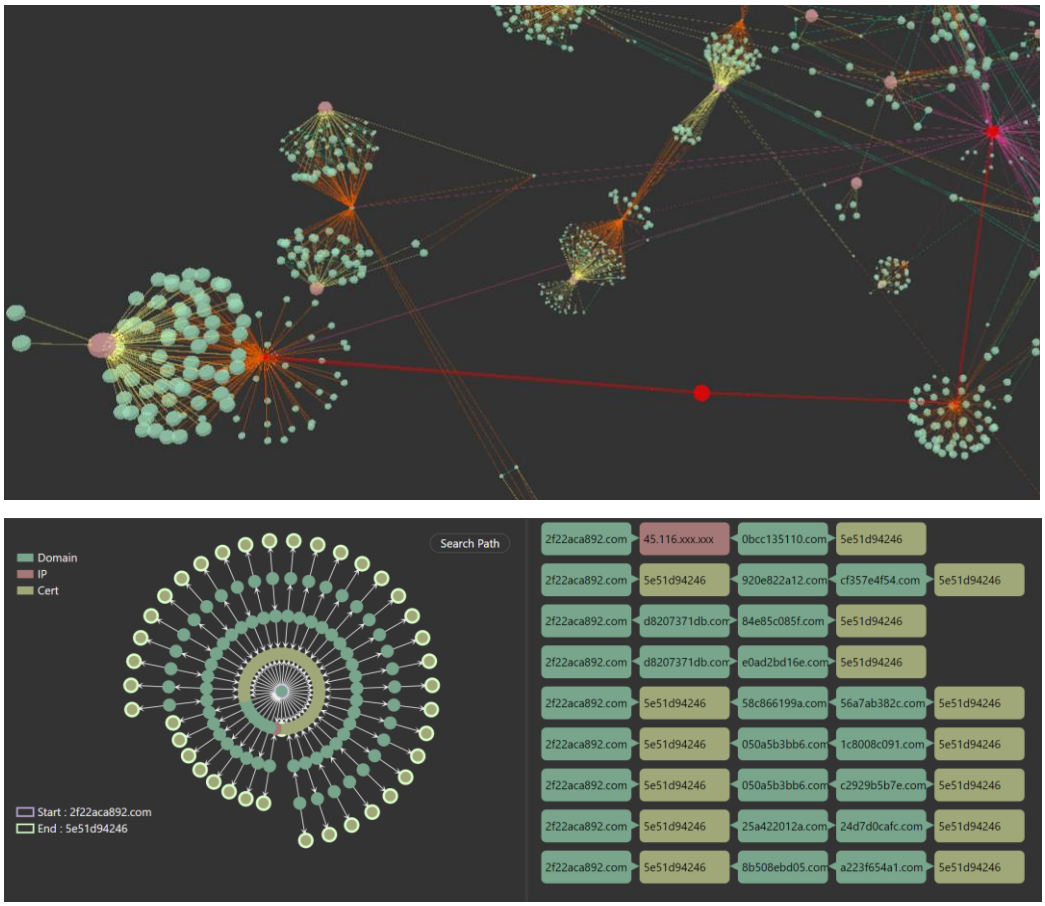

图 39 黑灰团伙 4 关键链路示意

表格 5 团伙 4 关键链路核心节点

节点 ID	节点 Name	节点类别
Cert_5e51d942460c7a85a67e99b3e38c05dd193222bddfafc3753514acd2923da8a8	5e51d94246	Cert
Cert_3a0454ff13c800ec019c888620cf3df359771c37c0fae1fb2c2f044841b249b8	3a0454ff13	Cert
IP_fbbba09f972cbbefcc14a33e87ec12e29cfecf1879b0dd847ba14d2038ef8e21	103. 253. xxx. xxx	IP
IP_f6c0c25ac33609674ec4994dcd88c645c08b54fa489c6435886264611536ba9d	103. 253. xxx. xxx	IP
IP_8216b7456870cb01b7d7b68d45a1b186	45. 116. xxx. xxx	IP

dab72164de84dd35f083452a0a33a111		
IP_d6a235eecdc8228c386cf15836797e90 2f0bd5d83143d9461f95992da305f304	45. 116. xxx. xxx	IP

➢ 黑灰团伙 5（大型团伙）

通过同样的方法，我们在社区中挖掘出第二个大型黑灰团伙。该黑灰团伙被发现的起始点为社区 1972945（图 40）。

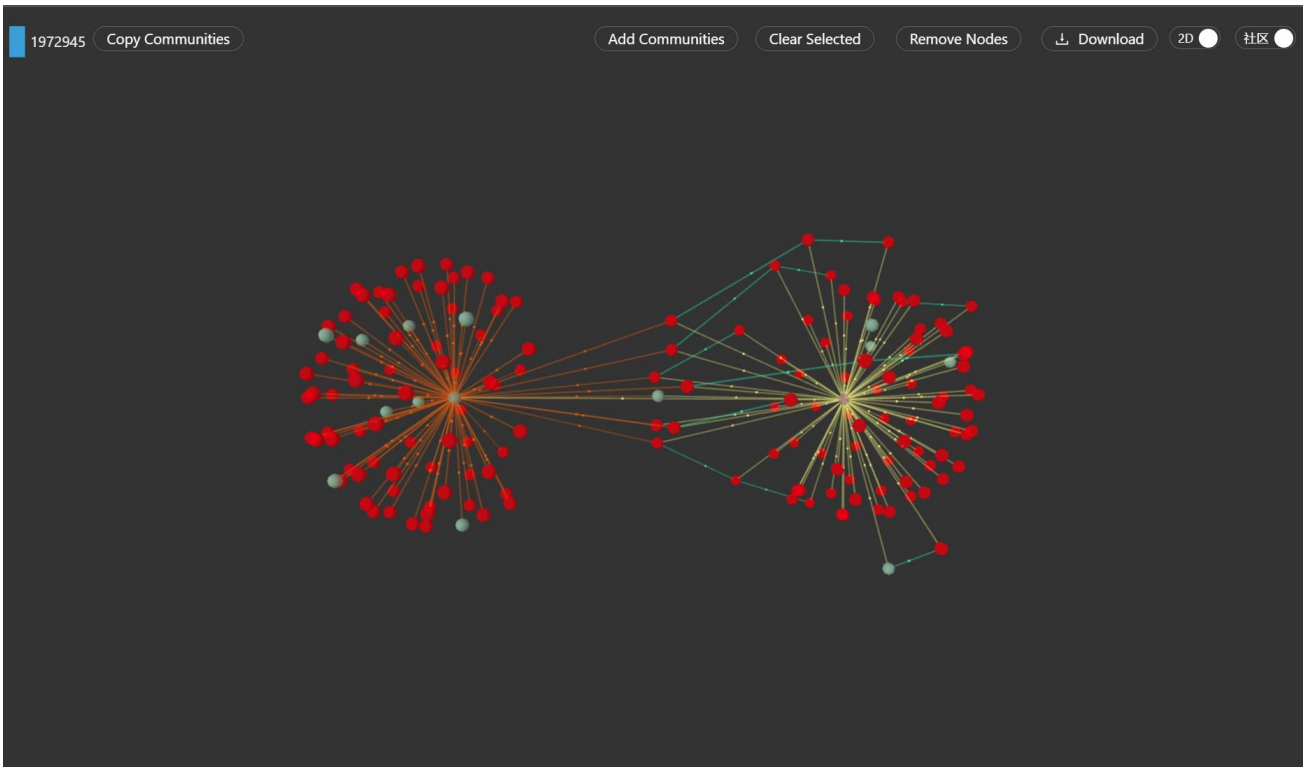

图 40 社区 1972945

通过该社区的邻居社区图（图 41），我们发现，其邻近社区大多数为赌博型的社区，当前社区类型相同，且每个邻居社区中异常节点的数量都达到了当前总节点数的一半以上。因此，这些邻居社区都应被添加到当前的社区之中。

图 41 社区 1972945

将相同业务类型的社区按照扩充规则添加至主视图，最终形成的黑灰团伙资产子图如图 42 所示。异常节点在图中标红所示，该团伙共有 2965 个节点，其中异常节点为 2141 个，黑灰产业类型主要为赌博以及少量的色情产业。

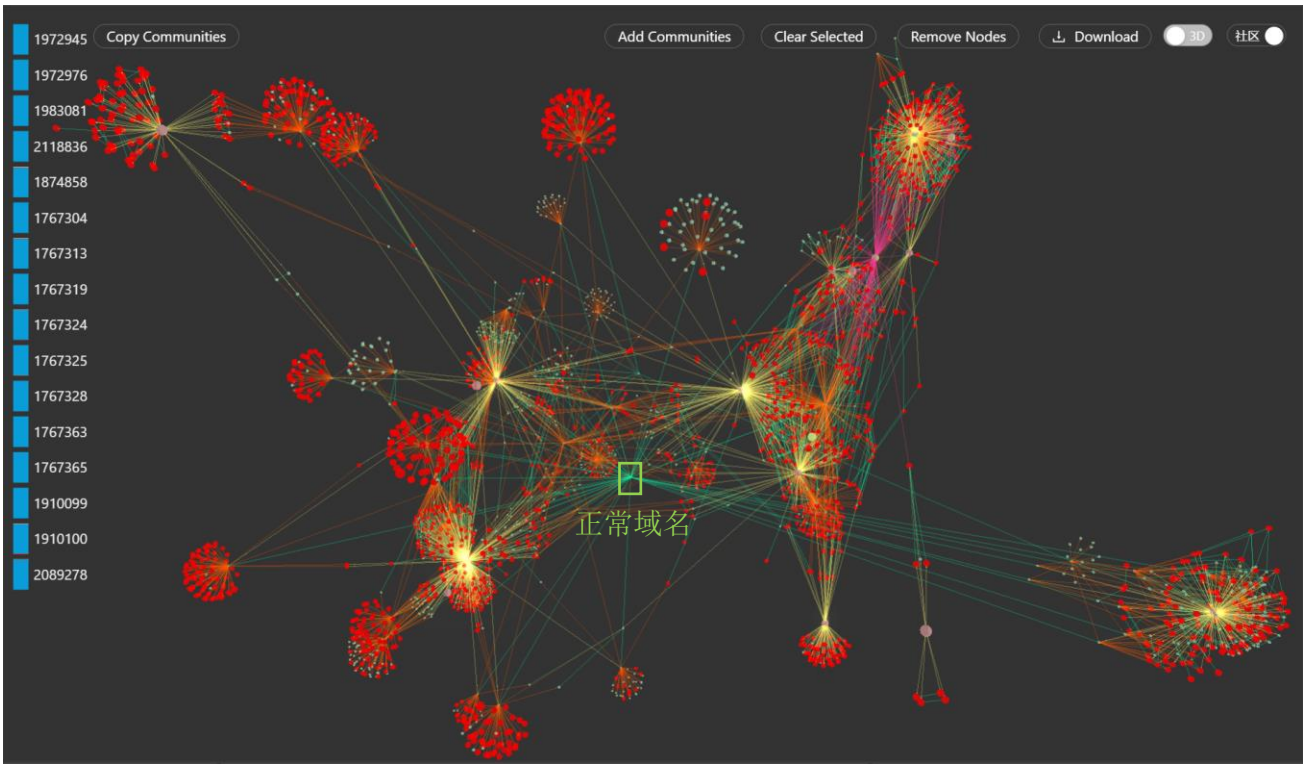

图 42 黑灰团伙 5 (大型)

对团伙 5 网络资产子图进行分析发现，该团伙依靠大量正常的域名为跳板，与团伙 3 和 4 类似。同时该团伙内部存在 4 个大型的黑灰社区，且这 4 个社区之间紧密相连，在该团伙的外部有着和团伙 4 及团伙 3 类似的异常域名集群。

因此，我们推测在大中型团伙中，利用正常域名作为周转和在核心资产外部构建与之类似域名集群应为黑灰团伙共有的一种模式。

表格 6 团伙 5 关键链路核心节点

节点 ID	节点 Name	节点类别
Domain_34ca6a4a7c099f5727c64c60df902fa58d144c5d727c0ff96e44669cb8c8d5e2	34ca6a4a7c.com	Domain
IP_aabca6424f0c2f75959003efa57136233d2dafb8e0e607fe2eaaee813ed303be	45.200.xxx.xxx	IP
IP_aa326da0ce3091617ac26d4ca1a1e4bf492d1218c38d2a9c73551ea48ff67ffe	45.200.xxx.xxx	IP
IP_500b7fa52c05f27c0a17b635e1b2fca4960a70632947ba8e0801fae1eea077a3	45.200.xxx.xxx	IP
IP_41fcfc3c79898b784f30db8631bf8734ca9cefde9971cd8e4fb38a5362bdecba	182.160.xxx.xxx	IP

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。

1.3.1 子图挖掘方法

本作品对子图挖掘的算法采用了 Newman 提出的 Modularity Optimization 算法。

对于挑战一所给定的网络结构，Modularity Optimization 方法被用于将该网络划分成不重叠的多个社区。如图 43 所示，我们用光球来表示划分的社区，并用异常节点的数量映射了光球的大小。

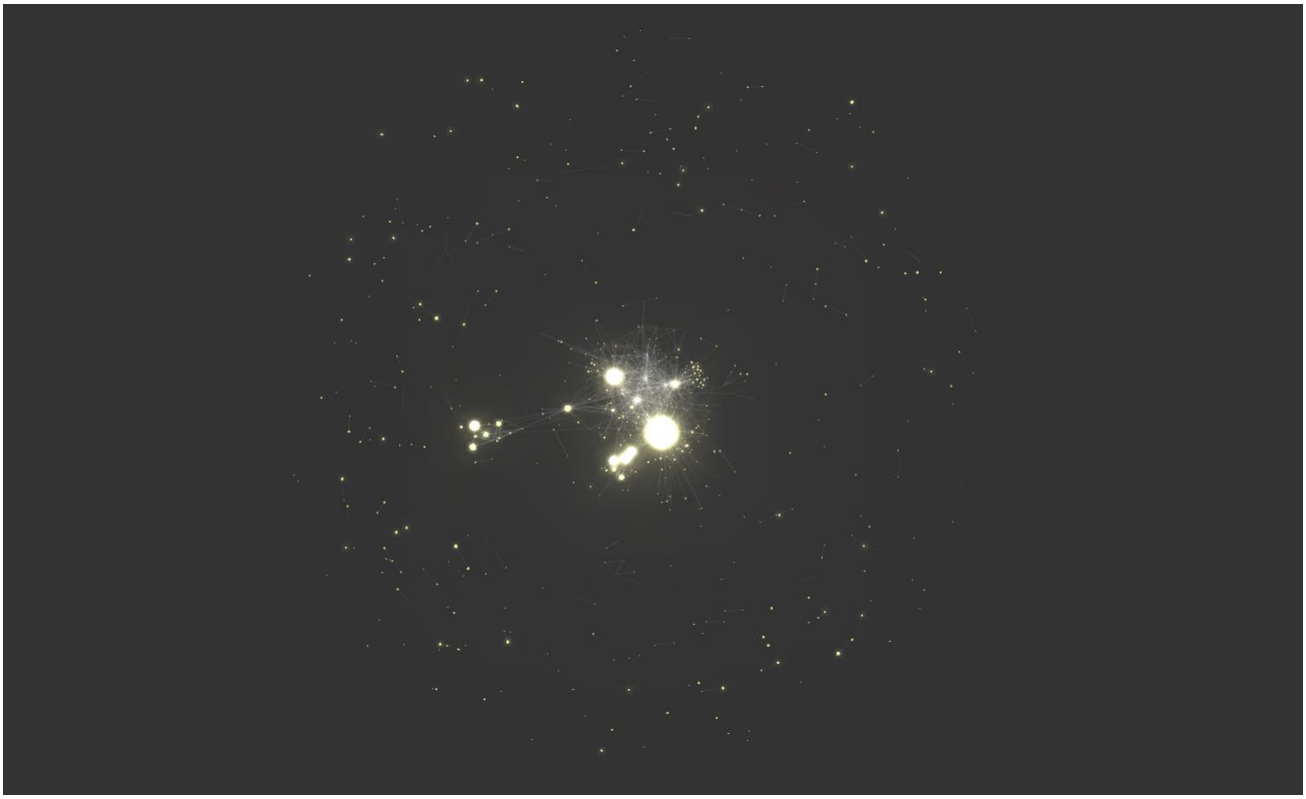

图 43 由社区算法自然划分出的社区连接图 (共 10655 个社区)

Modularity Optimization 方法首先定义模块度(Modularity)作为衡量一个社区的划分是否为相对较好的指标。若一个划分结果在社区内部的节点相似度较高，而在社区外部节点的相似度较低，则该社区应为一个相对较好的划分结果。

由于本题中所给数据集为一个完整的网络结构化数据，因此利用社区发现算法对网络划分能有效地获取各个黑灰团伙子图信息。在本题中，我们选用全局模块度Q作为网络划分的评价指标。全局模块度Q定义为社区内部的总边数和网络中总边数的比值减去一个期望值。该期望值是当网络设定为随机网络时，同样的社区划分所形成的社区内部的总边数和网络中总边数比值的大小。全局模块度Q的计算公式为:

$$Q = \frac{1}{2m}\sum_{vw}\left[A_{vw} - \frac{k_v k_w}{2m}\right]\delta(C_v, C_w) \tag{1}$$

其中，k_v表示节点v的度，A_{vw}为网络邻接矩阵的一个元素，其值定义为:

$$A_{vw} = \begin{cases} 1 & \text{节点 v 与 w 相连} \\ 0 & \text{其他情况} \end{cases} \tag{2}$$

函数$\delta(C_v, C_w)$的取值定义为：C_v为节点v所在的社区，C_w为节点w所在的社区。若v和w在一个社区，则$C_v = C_w$，函数值为 1，否则为 0。m为网络中边的总数。若v和w分别位于两个社区，则社区内部的边数和网络中总边数的比例为：

$$\frac{\sum_{vw} A_{vw}\delta(C_v, C_w)}{\sum_{vw} A_{vw}} = \frac{1}{2m}\sum_{vw} A_{vw}\,\delta(C_v, C_w) \tag{3}$$

因此，在进行每次的网络划分的时候需要计算Q值，Q取值最大的时候即为此网络的比较理想的划分。Q值的范围在 0 至 1 之间，Q值越大则说明网络划分的社区结构的准确度越高。在实际的网络划分中，Q值最高点一般出现在 0.3 至 0.7 之间。

在定义了全局模块度Q之后，我们使用 GN 算法对整体网络进行划分。其主要思想为：在一个网络之中，通过社区内部的边的最短路径相对较少，而通过社区之间的边的最短路径的数目则相对较多,其关键为计算网络中的边介数，边介数即网络中任意两个节点通过当前边的最短路径的数目。算法每次选择边介数较高的边进行删除，直至网络中的任一顶点作为一个社区为止。

因此，算法共分为以下 5 步：

(1) 计算网络的每一条边的边介数；

(2) 将边介数除以对应边的权重得到边的边权比；

(3) 找到边权比最高的边，并将其删除，计算网络的全局模块性Q。在计算中当边权比最高的边有多条时，同时移除这些边，并将此时移除的边和Q值进行存储；

(4) 重复步骤(1)(2)，直到网络中所有的边均被移除；

(5) GN 算法划分结束后，取出Q值最大时的序号，在原始矩阵中依次去除该次划分之前的边，得出最终连通矩阵；

1.3.2 核心网络资产识别方法

在说明核心网络资产识别方法之前，我们将对我们的节点划分方法进行简要的介绍。由于原始数据节点规模较大，难以在该规模的数据集上进行社区划分，因此在处理数据的过程中我们对节点进行数据约简。我们注意到，在挑战赛文档说明中有专门对于不同类型的节点的重要程度进行分类，此外在附录 5 中也给出了黑灰产网络资产图谱抽象模型，因此我们依据该抽象模型，对节点进行了重构。

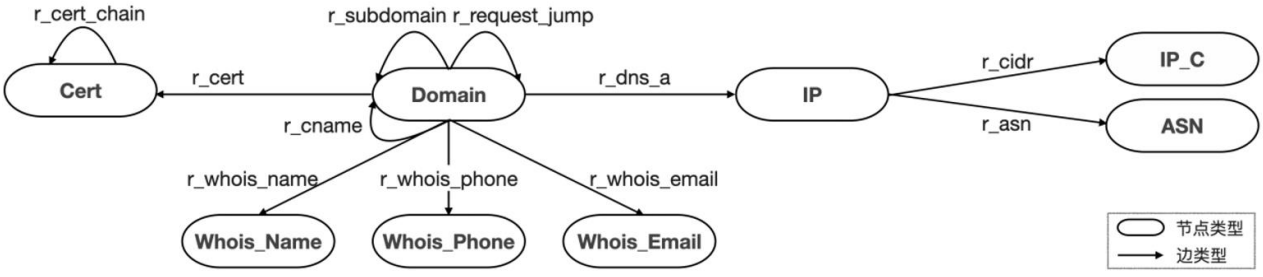

图 44 黑灰产网络资产图谱抽象模型

我们将 Whois_Name，Whois_Phone，Whois_Email 三个节点信息归类到 Domain 节点上，因为该类型节点只与 Domain 节点相关联，因此我们认为让其作为 Domain 节点属性更为合理。同理，对于只和 IP 相关联的 IP_C 节点和 ASN 节点，我们也将其归类到了 IP 节点的属性中，这样大幅地简化了数据规模，同时也没有丢失数据信息。值得一提的是，在后续的可视化探索阶段，我们发现图中丢失注册人信

息节点可能会对我们判断社区之间是否存在关联造成误判。因此我们在系统中将原先丢失的注册人节点信息进行了复原，但该复原仅仅停留在可视化系统之中，数据库中并不会包含这些节点的真实信息，因此不会对我们的后台的数据查询造成影响，同时也增强了我们系统可视化分析的能力。

对于核心网络资产的识别方法，我们根据挑战赛说明文档中对于核心网络资产的定义来进行核心网络资产识别，即：

(1) 如果某个网络资产一半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；

(2) 如果某个 Domain 关联了多个 IP 地址，则其关联的 IP 地址同样不被认为是核心网络资产；

针对规则 1，由于我们系统对节点进行了重构，原先关联强度较弱的 r_asn 和 r_cidr 类型的边在我们的数据库中已被清除，因此该规则我们的数据自然满足。我们的核心网络资产的判别方法主要依据规则 2，我们的判别方法通过构造与规则 2 类似语义的图数据库查询语句(neo4j-cypher)对每个节点进行判别，以此来区分核心网络资产和普通网络资产，这一点会在我们后续对关键链路查询的介绍中会再次提到。

1.3.3 关键链路识别方法

对于关键链路的识别方法，我们同样依据挑战书说明文档中对于关键链路的定义来识别的。

(1) 两个核心资产间长度大于 4 跳的路径不被认为是关键链路；

(2) 两个核心网络资产间存在多条路径时，路径越短越重要；

(3) 两个核心网络资产间路径的关联强度越强则越重要；

依据规则 1，我们识别关键链路从节点出发，因为关键链路路径长度应在 4 跳以内，因此我们通过确定起点和终点，查询起点终点间 4 跳以内路径的方式来返回关键链路，同时，我们注意到关键链路节点应为核心网络资产，因此我们在查询 4 跳以内节点之前，对于节点是否是核心网络资产进行了额外的判别，如果选中的节点不符合核心网络资产条件，则以该节点为起点或终点的关键链路应不存在。对于规则 2，我们采用可视化的方法对 4 跳以内的关键链路进行了展示，我们将关键链路构造成一个圆环，这样用户能很快确定选中节点之间关键链路的跳数分别是多少，也能更快地关注到那行跳数更短的路径。对于规则 3，我们选择采用链路列表结合主图高亮的方式来展示链路，通过链路列表，用户可很快确定该链路通过了哪些节点，与哪些类型的边相关联，因此可以由此来确定该链路的关联强度，同时在主图中高亮路径也有助于帮助用户更好地分析与关键链路相关联的其他节点。

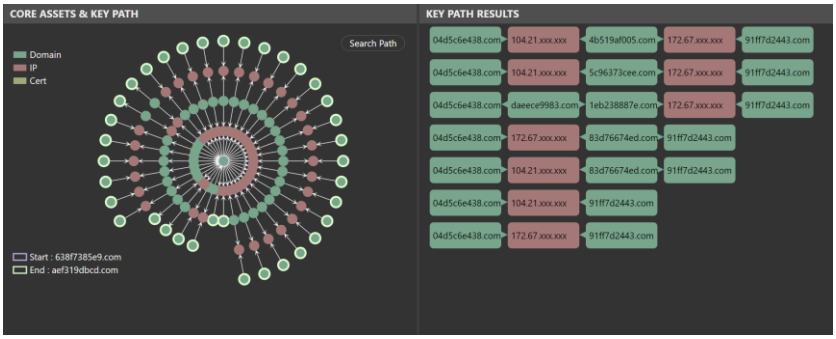

图 45 关键链路图及链路列表

1.3.4 图可视化方法与交互分析方法

本作品系统设计并实现了多种视图以及增强人机互动性的控制系统(图 46)，辅以合理的视觉编码，同时通过控制系统支持了更加灵活直观的交互设计，支持用户开展对黑灰产团伙网络资产子图挖掘、子图的核心网络资产分析、子图的关键链路分析以及网络社区信息概览等分析任务。

本可视分析系统主要采用深色为底，并根据实际展示效果对系统进行配色；通过对犯罪类型、社区等元素的颜色编码进行区分，并辅以深灰色边框区分系统中的不同视图。

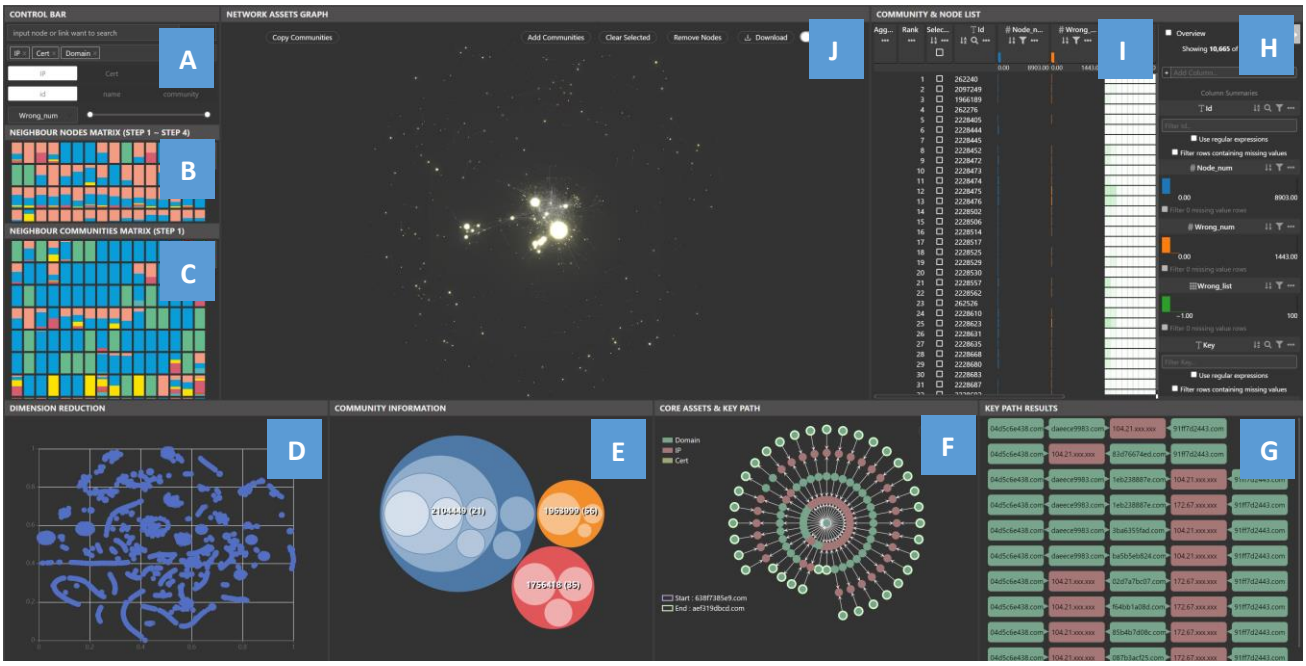

图 46 系统概览图

本系统页面中主要包括 2 个控制区域和 8 个视图区域。

控制区域包括主视图自定义控制台(图 46A)、信息总览列表控制台(图 46H)。以上两个控制台分别对主视图(图 46J)和社区&节点信息总览列表(图 46I)进行用户自定义设置，实现内容过滤、图例设置、信息搜索等功能。

视图区域包括邻近节点矩阵(图 46B)、邻近社区矩阵(图 46C)、网络社区类型散点图(图 46D)、社区信息气泡树图(图 46E)、核心资产&关键链路图(图 45F)、关键路径详细信息链接图(图 46G)、社区&节点详细信息列表(图 46I)以及网络资产力导向图(图 46J)。

本系统主要基于网络资产数据，对黑灰产团伙的网络资产子图、核心网络资产、关键路径以及团伙关联信息、黑灰产团伙网络运作机制等分析任务提供支持。

1.3.4.1 主视图自定义控制台

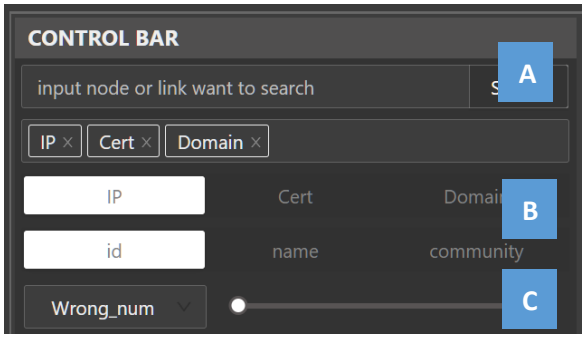

图 47 主视图自定义控制台

（1）视图设计

主视图自定义控制台对节点&社区&关系搜索信息(图 47A)、各类节点具体图例信息(图 47B)使用文本展示；对主图点标记数目使用滚动控制条展示(图 47C)；同时包含各类设置选择按钮。

（2）交互设计

该控制台主要是对主视图(网络资产力导向图)进行点标记筛选及信息检索。各方面具体交互设计如下：

a) 信息检索条(图 47A)：该设计通过以下三种自创的搜索语句进行搜索：
   - Domain?name:xxx(通过 name 属性搜索符合条件的 Domain 节点，并在主视图中高亮)；
   - nodeID(直接输入节点 ID，并在主视图高亮)；
   - communityID1,communityID2,communityID3(同时检索多个社区并在主视图中高亮)；
   - nodeID>nodeID(检索两个节点间的关系，并在主视图中高亮)

b) 主视图图例设置(图 47B)：该设计通过分别选择设置的对象(IP/Cert/Domain)和展示的图例信息(id/name/community)，实现对主视图图例信息(图 48)的自定义设置。

c) 主视图点标记过滤：该设计通过滚动控制条设置条件，从三个维度(Node_num/Wrong_num/Neighbour_num)筛选主视图显示的点标记。

图 48 主视图图例信息

1.3.4.2 邻近节点矩阵

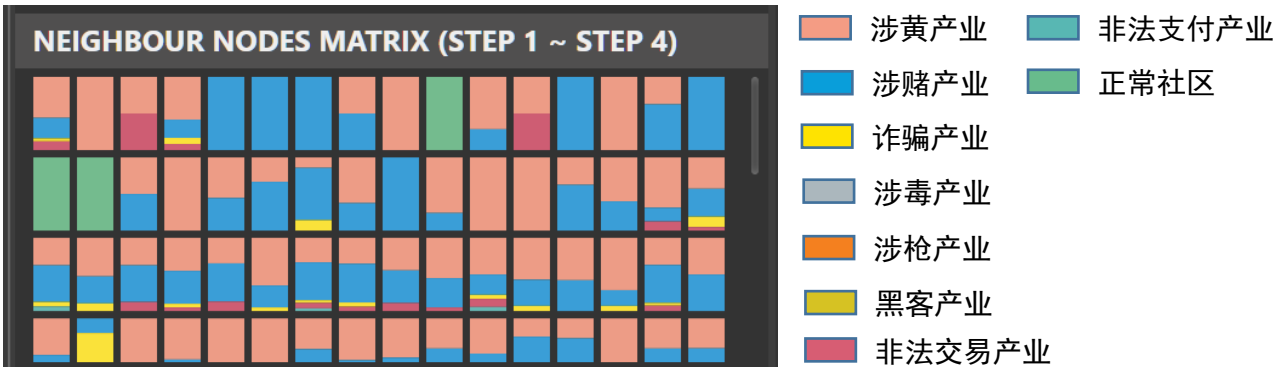

图 49 邻近节点矩阵(左) 颜色编码(右)

（1）视图设计

本视图的目的在于显示主视图中与选中节点的关系在四跳以内的社区信息(图 49 左)。该视图中使用矩形块代表社区，使用不同的颜色映射社区包含的不同的犯罪类型(图 49 右)，使用色块的大小映射各犯罪类型的节点数量；并且将矩形块根据其与选中节点的跳数排序，跳数越低即与之距离越近的社区对应的矩形块排在越前面。

（2）交互设计

该视图的交互主要是对主视图中显示的节点进行添加工作。具体的交互设计如下：

主视图节点添加：点击不同的矩形块，页面中添加相应的社区包含的节点和关系边。同时相应的矩形块会消失。

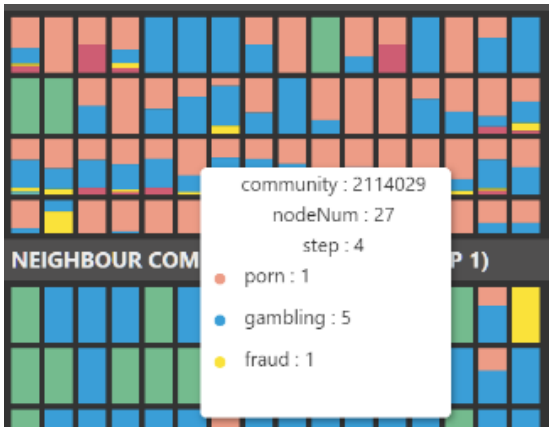

图 50 邻近节点矩阵图图例信息

矩阵图例显示：鼠标悬浮时，展示选中的社区的 ID、包含的节点数目、与主视图中选中的节点的关系跳数、包含的犯罪类型及该犯罪类型节点数目(图 50)。

1.3.4.3 邻近社区矩阵

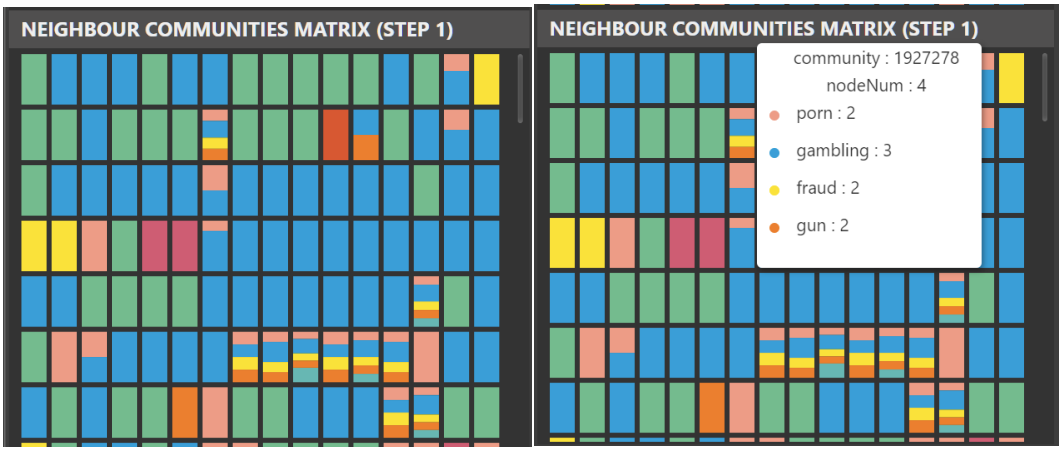

图 51 邻近社区矩阵(左) 视图图例信息(右)

（1）视图设计

本视图的目的在于显示与主视图中显示的社区的关系在一跳以内(直接相连)的社区信息(图51左)。该视图中使用矩形块代表社区，使用不同的颜色映射社区包含的不同的犯罪类型(图50右)，使用色块的大小映射各犯罪类型的节点数量。

（2）交互设计

该视图的交互主要是对主视图中显示的社区进行添加工作。具体的交互设计如下：

主视图节点添加：点击不同的矩形块，页面中添加相应的社区包含的节点和关系边，相应的矩形块会消失。与此同时，还会添加与新添加的社区直接相连的社区。除了在主视图中添加节点，还会在主视图左上角添加社区信息。

矩阵图例显示：鼠标悬浮时，展示选中的社区的 ID、包含的节点数目、包含的犯罪类型及该犯罪类型节点数目(图51右)。

1.3.4.4 网络社区类型散点图

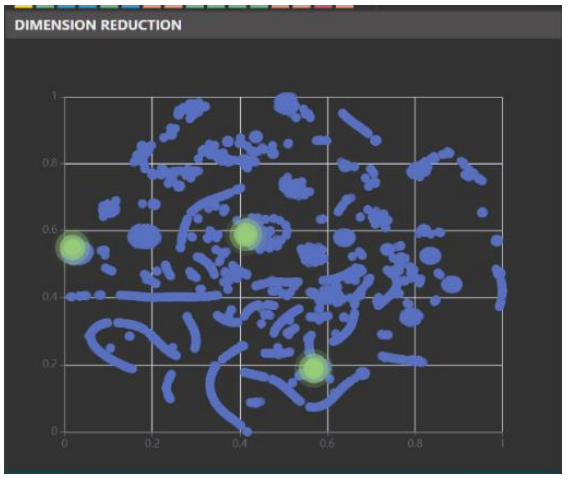

图 52 网络社区类型散点图

（1）视图设计

视图的目的在于显示对社区属性信息(具体犯罪类型节点数、正常节点数等)降维后的二维空间位置映射(图 52)。该视图利用位置信息映射社区的属性，位置越接近的点对应的社区相似度越高；而高亮的绿色点标记是目前在主视图中显示的社区的散点映射，用户可以直观的看到目前在主视图中的社区属性信息。

（2）交互设计

该视图的交互主要是与主视图中的显示信息进行同步。具体的交互设计如下：

- 信息同步：主视图中显示的社区都会在该视图中以绿色高亮点的形式显示。
- 散点图图例显示：鼠标悬浮时，展示选中的社区的 ID。

图 53 网络社区类型散点图图例信息

1.3.4.5 社区信息气泡树图

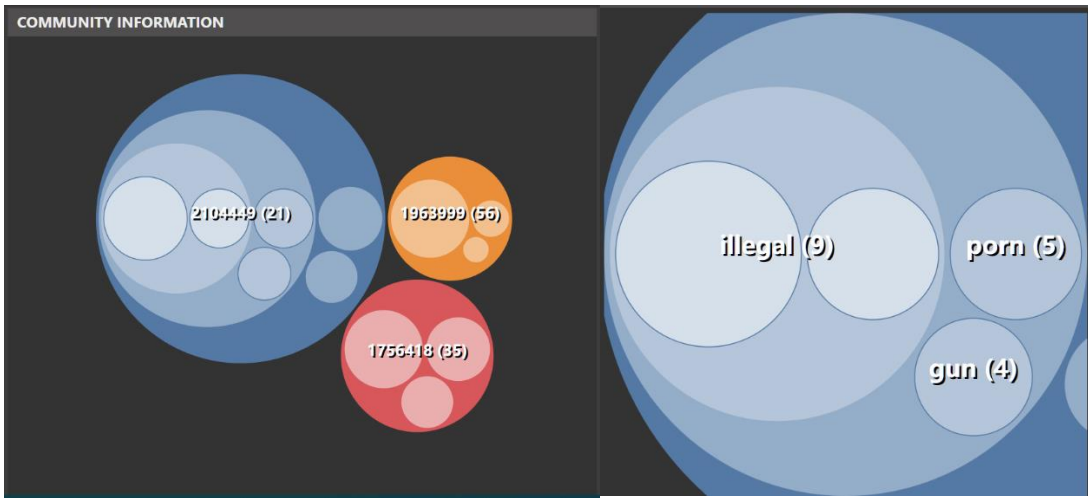

图 54 社区信息气泡树图(左) 视图详细信息(右)

（1）视图设计

视图的目的在于显示主视图中的显示的社区信息(图 54 左)。该视图利用颜色映射不同的社区、利用透明度映射不同的节点类型、用树图的包含关系映射社区与节点类型及犯罪类型的层级关系、用气泡的大小映射社区中各类节点的数目。每个气泡中包含以下信息：Domain 节点数目、Domain 节点包含的犯罪类型数量、IP 节点数目以及 Cert 节点数目。

（2）交互设计

该视图的交互主要是与主视图中的显示信息进行同步。具体的交互设计如下：

- 信息同步：主视图中显示的社区都会在该视图中以与其他气泡不同颜色的气泡的形式显示。
- 详细信息显示：鼠标点击相应的气泡时，该气泡会放大并显示更为详细的信息(图 54 右)。

1.3.4.6 核心资产&关键链路图与关键路径详细信息链接图

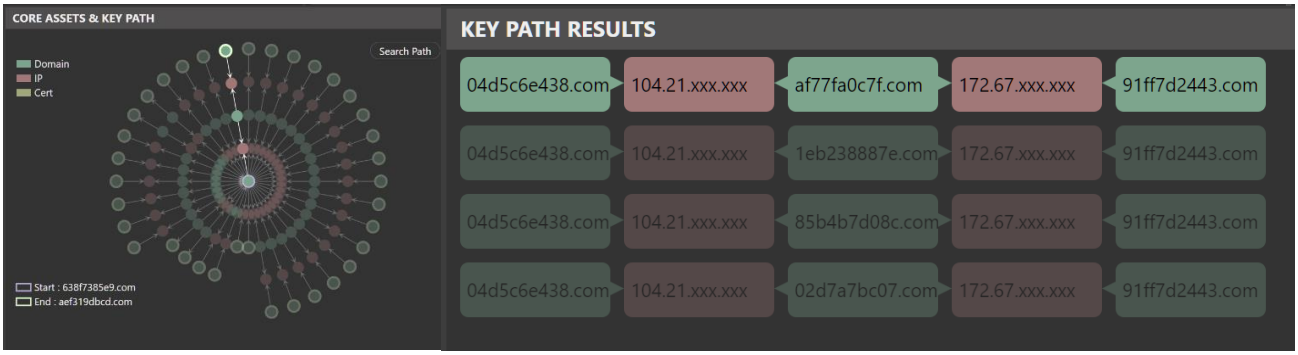

图 55 核心资产&关键链路图(左) 关键路径详细信息链接图(右)

（1）视图设计

以上两种视图呈现十分紧密的捆绑关系。核心资产&关键链路视图(图 55 左)用圆点映射不同的节点，用颜色映射不同的节点类型，用箭头映射关系的方向。同时关键路径详细信息链路图(图 55 右)以块映射不同的节点，用颜色映射不同的节点类型，以文字的方式直接显示节点的 name，并用三角形的方向映射不同的关系走向。

（2）交互设计

以上两个视图的交互主要是显示节点间的关键路径及详细信息。具体的交互设计如下：

- 关键路径检索：用户可以在主视图中选中两个点，再点击核心资产&关键链路视图中的“Search Path”，核心资产&关键链路图以径向布局的方式显示两个点之间的所有关键路径。
- 关键路径信息同步：鼠标点击核心资产&关键链路视图中的路径时，相应路径会高亮，左下角会显示该路径的起始节点的 name 和终止节点的 name。同时关键路径详细信息链接图中会添加一条详细的关键路径信息。

1.3.4.7 社区&节点详细信息列表及控制台

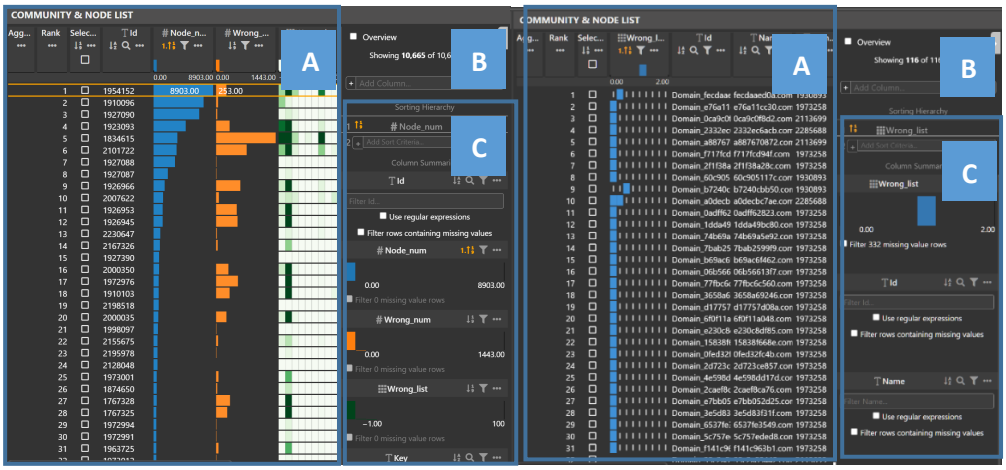

图 56 社区详细信息视图(左) 节点详细信息视图(右)

（1）视图设计

该详细信息列表分为两个视图。社区详细信息视图(图 55 左)对主视图中所显示的所有社区信息(图 55 左 A)、列表显示信息总览(显示与选中的信息总条数)(图 55 左 B)使用文本展示；使用不同颜色的条形图的长度映射此社区的包含的各种节点数目；用不同的颜色饱和度映射各类犯罪类型的节点数占比。节点详细信息视图(图 55 右)对主视图中所显示的所有节点信息(图 55 右 A)、列表显示信息总览(显示与选中的信息总条数)(图 55 右 B)使用文本展示；使用蓝色矩形块的位置映射其包含的不同的犯罪类型。

（2）交互设计

该视图与控制台主要是对主视图(网络资产力导向图)中的信息进行详细展示并过滤主视图中的数据。各方面具体交互设计如下：

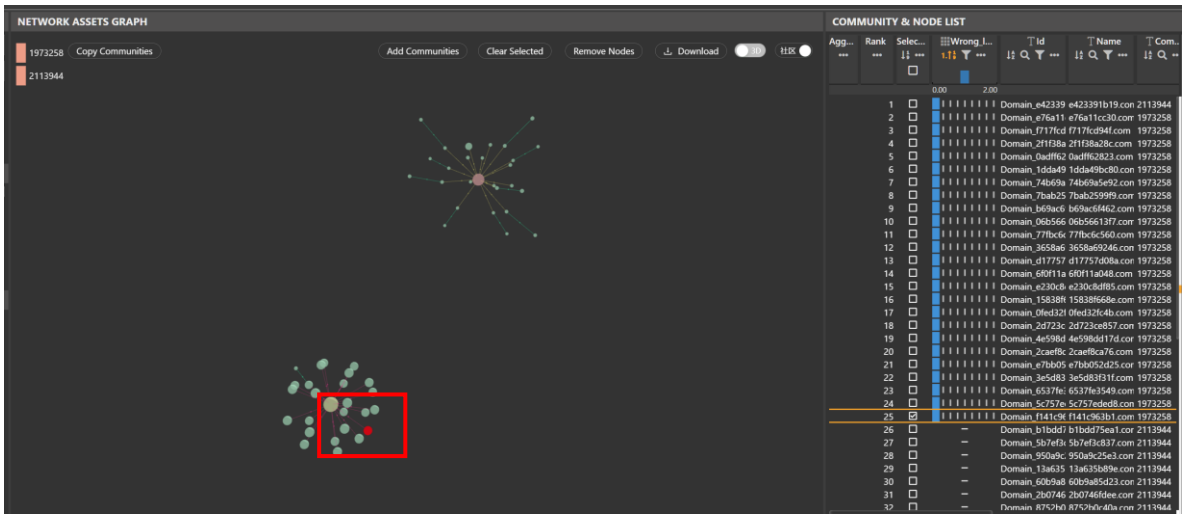

图 57 勾选主视图高亮显示

a) 主视图信息展示与筛选：详细信息列表(图 55 左 A、图 55 右 A)中将显示主视图中的所有社区或节点的详细信息。通过勾选列表中的信息，在主视图中进行相应的高亮显示(图 56)。

b) 所示信息总览(图 55 左 B、图 55 右 B)：对列表中的信息完成统计，包括此时显示的信息总条数和勾选的信息总条数。

c) 社区详细信息过滤(图 55 左 C)：该功能主要通过控制台实现。从上到下依次实现以下功能：分别以节点数目、不正常节点数目、犯罪类型列表为过滤条件，通过拉取控制条实现过滤；通过选择或输入社区中心节点 id 以及邻近社区 id 实现社区信息检索。

d) 节点详细信息过滤(图 55 右 C)：该功能主要通过控制台实现。从上到下依次实现以下功能：以犯罪类型数目为过滤条件，通过拉取控制条实现过滤；通过选择或输入节点 id、节点 name、节点所属社区 id、域名的注册人邮箱或域名的注册人电话实现社区信息检索。

1.3.4.8 网络资产力导向图

图 58 网络资产力导向图

（1）视图设计

该视图为主视图，以社区为单位显示了各资产(节点)及资产之间的关系(边)(图 57A)。该视图以不同的颜色映射不同的节点类型和边类型；以线标记体现点与点之间的关系；力导向的布局也使得视图的布局更加美观和稳定；以矩形色块的形式显示主视图中添加的社区(图 57B)。该视图以极其直观的方式表现了社区网络的分布。

（2）交互设计

该视图的交互主要体现在对力导向图中的节点进行筛选过滤。各方面具体交互设计如下：

a) 社区筛选(图 57C)：点击相应社区的矩形块可以消除该社区的色块和社区对应的力导向网络。

b) 节点添加及过滤(图 57B)：该设计以多个按钮的形式实现。点击“Add Communities”在主视图中显示的路径基础上添加沿路节点所在的社区；点击“Clear Selected”清除主视图中的高亮标记；点击“Remove Nodes”清除主视图中与社区&节点详细信息列表选中的信息所对应的点标记；点击“Download”下载主视图中所有的点信息(id/name/犯罪类型)和边类型(起始点id/终止点 id)；点击“社区”切换按钮可以切换此时主视图展示的是添加的社区还是初始的所有点总览。

c) 点击悬浮事件：鼠标点击相应的点可以进行高亮；悬浮于点或边上方可以显示相应的图例信息。