arXiv:2601.01492v1 [cs.IR] 4 Jan 2026

数字森林中的面包屑：利用开源情报通过种子元数据追踪犯罪分子

Annelies de Jong[a], Giuseppe Cascavilla[a], Jessica De Pascale[a]

[a]蒂尔堡大学，荷兰蒂尔堡

摘要

本研究探讨种子元数据作为开源情报（OSINT）来源的潜力，重点关注用户画像和行为分析。尽管 BitTorrent 等点对点（P2P）网络在隐私和性能方面已得到充分研究，但其元数据很少被用于调查目的。本研究提出了一项概念验证，用于展示 Tracker 响应、种子索引数据和经过增强的 IP 元数据如何揭示与高风险行为相关的模式。

本研究遵循五步开源情报流程：来源识别、数据收集、数据增强、行为分析和结果呈现。研究数据来自海盗湾 和 UDP 跟踪器，形成了一个包含 206 个热门种子的、超过 60,000 个独立 IP 地址的数据集。研究通过地理定位、匿名化状态以及涉及儿童剥削材料（CEM）的标记对数据进行了增强。一项针对敏感电子书的案例研究表明，这类数据有助于发现对非法内容的潜在兴趣。

网络分析突出了对等节点聚类、共同下载模式，以及可疑用户对隐私工具的使用。研究表明，公开在线的种子元数据能够支持可扩展且自动化的 OS-INT 画像分析。

本研究提出了一种从噪声数据中提取有用信号的新方法，为数字取证领域作出了贡献，并可应用于执法、网络安全和威胁分析。

电子邮箱地址： anneliesdej2001@gmail.com (Annelies de Jong), g.cascavilla@tilburguniversity.edu (Giuseppe Cascavilla), D.DePascale@tilburguniversity.edu (Jessica DePascale)

关键词：开源情报、种子元数据、IP 元数据、用户画像、用户数据报协议抓取、匿名性检测、数字取证、网络分析

---

1. 引言

在全球互联和公开数据广泛可用所塑造的数字环境中，开源情报（OSINT）已成为收集和分析公开可访问信息的重要工具。从网络安全到执法和数字取证，开源情报通过追踪数字行为及其模式，为调查工作提供支持。然而，一些特定数据源仍未得到充分利用，其中包括点对点（P2P）文件共享活动所生成的元数据。

BitTorrent 是使用最广泛的点对点协议之一，它通过称为种子的小型元数据文件实现去中心化文件分发。这些文件包含跟踪器 URL、文件名和加密标识符等信息，但本身并不存储内容。尽管人们已对 BitTorrent 在盗版检测、网络优化和协议分析方面进行了广泛研究，[1, 2], 但在开源情报场景中，其元数据在用户画像和行为分析方面仍未得到充分探索 [3, 4, 5]。现有关于种子流量和 IP 情报的研究，主要聚焦于威胁检测和基础设施性能，而非用户画像或行为分析。

本研究通过评估种子元数据能否支持可扩展且结构化的基于开源情报的用户画像，弥补了这一空缺。本研究提出了一项多阶段概念验证，结合种子索引数据、Tracker 响应、IP 元数据和外部标记来源，以识别表明存在高风险活动的行为信号。

在本研究中，高风险行为 指可能表明犯罪或有害意图的潜在可能性上升的种子活动。例如，反复参与涉及已标记内容（如儿童剥削材料 - CEM）的群集、对敏感主题（如爆炸物手册）表现出浓厚兴趣，以及将匿名化技术与此类活动结合使用。尽管这些信号无法证明意图或违法行为，但可用于在开源情报工作流程中确定后续调查的优先级。

在本研究背景下，我们提出了一种结构化的多阶段概念验证，结合种子索引数据、Tracker 响应、IP 元数据，

以及外部标记来源。尽管目前尚未实现完全自动化，但该设计支持未来的可扩展性，旨在解决开源情报中常见的挑战，例如数据噪声以及跨来源关联的需求。

本研究的核心研究问题（RQ）如下：

> **RQ**：我们的概念验证在多大程度上能够展示利用种子元数据进行基于开源情报的用户画像的潜力？

我们通过三个子问题（sRQ）进一步探讨主要研究问题：

> **sRQ1**：有哪些公开可用的种子和 IP 元数据可用于检测高风险行为？

这个问题考察了可以从种子索引和追踪器等开放来源获取哪些类型的数据，并探讨这些数据的结构、可用性及其识别可疑活动的潜力。

> **sRQ2**：我们的概念验证利用公开可用的数据绘制种子用户网络的效果如何？

sRQ2探讨能否通过网络分析检测共同下载行为或内容聚类等模式。sRQ2还探讨这些模式如何有助于画像分析和行为洞察。

> **sRQ3**：种子和 IP 元数据在数据质量、结构和处理工作量方面，是否适合整合到开源情报工作流程中？

最后，sRQ3评估在现实世界的开源情报工作流程中使用此类数据的实际可行性。这包括考察数据质量、匿名化挑战和技术开销。它还对这类数据是否能够以及如何切实支持调查作出更全面的判断。

为回答这些问题，本研究采用五阶段开源情报工作流程：来源识别、数据收集、处理、分析和报告。sRQ1 涉及查找并收集有用的种子和 IP 数据。sRQ2 着重分析这些数据，以发现模式和用户行为。sRQ3 探讨这些数据在实际调查中的实用性和可靠性。这些问题共同引导研究逐步评估数据从获取到实际应用的过程。

数据采集自海盗湾和公开的用户数据报协议跟踪器，形成了一个数据集，包含 206 个热门种子中超过 60,000 个唯一 IP 地址。该数据集还补充了地理位置信息、匿名化标记以及外部指标（例如与儿童剥削材料的关联）。

这项针对敏感电子书的聚焦案例研究展示了元数据如何帮助揭示用户对高风险内容类别的兴趣。网络分析揭示了共同下载、隐私保护工具的使用以及对等节点聚类等模式，而这些都有助于行为画像。

通过这一方法，本研究作出了以下几项贡献：

- 本研究将**种子元数据引入为一种新颖且尚未充分利用的开源情报数据源**，用于用户画像。据我们所知，这是首次以结构化方式将种子数据用于这一目的。

- 本研究提出了一个**可复现的画像框架**，将种子记录、Tracker 响应、IP 情报和第三方标记整合到一个工作流程中。

- 本研究提升了**在噪声环境中的信号提取能力**，解决了从大量非结构化数据流中筛选相关情报这一常见难题。

- 本研究提供了**确定调查优先级**的工具，可识别暗示风险或意图较高的行为信号。

- 最后，尽管该概念验证目前需要手动执行，但它为**未来的自动化与可扩展性**奠定了基础，使其有望部署于实时开源情报或取证系统中。

总之，本研究凸显了种子元数据在开源情报工作流程中用于行为画像的潜力尚未得到充分挖掘。通过将非常规数据源整合到数字调查中，本研究扩展了威胁情报和执法界可用的分析能力。

2. 背景

开源情报（OSINT）是收集和分析公开可用的 Data，以生成可采取行动的知识 [6]。现代数字世界

产生了海量的《信息》 [3], ，其中大部分内容任何人都可以在任何地点、任何时间访问。开源情报利用这种可访问性来收集、处理和关联 Data，从而构建目标的更完整图景。其优势在于结合多个来源，揭示单一数据流无法提供的洞见。

开源情报随着技术的发展而变化。早期工作主要关注报纸、公共记录和政府出版物 [6]。社交媒体和在线论坛等数字平台的发展拓宽了其范围。近来，大数据、自动化和人工智能使大规模采集和实时分析成为可能 [4]。这些进步拓展了开源情报在众多领域的应用。在网络安全领域，开源情报有助于加强威胁检测和漏洞评估 [7, 8]。执法部门将其用于调查、反恐和数字取证 [9, 10]。它还通过核实来源和侦测泄密来支持国家安全 [11], ，也应用于保护生物学研究以及社会情绪或客户情绪研究 [12, 13, 14]。

多位研究人员提出了开源情报的结构化工作流程。尽管细节略有不同，但这些流程具有共同的结构。例如， [15] 概述了一套基本工作流程，包含以下阶段：

1. 根据情报目标确定信息来源 2. 从开放和公共来源收集数据 3. 处理数据，清理原始信息并将其结构化 4. 运用文本挖掘、社交网络分析和地理空间制图等技术分析数据 5. 汇报并传播调查结果

[4] 中的作者对 [16]的模型进行了拓展，并强调了规划和传播的作用，尤其是在研究发现能够推动方法和技术改进时。[17] 报告了相同的五个阶段，而 [18]则只强调最后四个阶段。 [3]聚焦于收集、分析和知识提取，展示了开源情报技术如何支持收集过程中的数据传输。这些步骤的产出为个人和组织分析提供了有用的见解。

尽管已有成熟的工作流程，仍然存在一些挑战。海量的噪声 Data 使筛选和提取变得困难 [4, 15]。自动化这些步骤能够提升 Data 的可用性 [19], ，但整合非开源情报 Data 需要仔细关注上下文和来源可靠性 [20, 21]。开源情报资源通常专注于特定地区，限制了其在全球范围内的应用 [3]。追踪

或对个人进行去匿名化也面临困难 [22]。最后，技术的快速变化迫使开源情报方法持续适应 [23]。

2.1. 点对点网络

点对点（P2P）网络采用去中心化模型，每个节点既充当客户端，也充当服务器。与依赖中央服务器的客户端-服务器系统不同，P2P 网络允许对等节点直接共享资源。这种设计提升了可扩展性和容错能力 [24, 25]，，但也带来了拒绝服务攻击、蠕虫传播和内容投毒等风险 [25]。

P2P 网络可分为结构化网络和非结构化网络。结构化网络采用预定义拓扑和分布式哈希表（DHT），以高效发现资源 [26]。非结构化网络则自由连接，依靠泛洪来定位资源 [27]。尽管效率较低，非结构化网络更能适应频繁的节点更替，也适用于 BitTorrent 等文件共享系统 [28]。

BitTorrent 是使用最广泛的 P2P 文件共享协议，全球每月用户超过 1 亿 [29]。尽管它解决了大型文件分发问题，却也与版权侵权密切相关：研究表明，大多数共享内容未经授权或属于非法内容 [30, 31]。BitTorrent 使用引发的两个重大问题是版权侵权和儿童剥削材料（CEM）的传播 [31]。由于缺乏集中控制，监控十分困难，也加剧了法律争议 [32]。

另一个严重问题是儿童剥削材料（CEM）的传播。研究表明，许多犯罪者偏好使用 BitTorrent，而非 Gnutella 或 eMule 等其他平台 [33, 34]。P2P 网络的匿名性降低了被侦测和起诉的风险，使其成为犯罪用途的理想选择 [32, 35]。这些网络之所以吸引犯罪者，是因为它们免费、向所有人开放，且能在几乎不与服务提供者直接接触的情况下提供匿名性[35]。

3. 相关工作

点对点网络上的非法活动一直是研究人员持续关注的难题，相关研究主要聚焦于版权侵权和 CEM 分发。大多数研究追踪内容分发，少数研究则关注画像分析

个人用户。虽然种子元数据和流量有助于网络安全、执法和网络优化研究，但它们在开源情报中的应用仍然有限 [36, 7]。

关于种子数据的研究大致分为三个领域：（1）盗版与取证调查；（2）协议优化；（3）与情报相关的网络流量分析。有关盗版的研究通常绘制国家或区域尺度的活动分布图。尽管这些研究为了解 P2P 活动提供了宝贵见解，却在很大程度上忽视了种子数据在用户画像方面的潜力。

例如， [1] 按国家分析 BitTorrent 上的视频分发情况，并评估反盗版措施； [37]则表明，较小且较贫困的国家盗版率相对较高。尽管这些研究绘制了盗版趋势图并纳入地理数据，却并未关注个人用户。

种子数据也有助于取证调查，尤其是涉及 CEM 的案件。 [38] 估算了参与五个 P2P 网络的对等节点数量，并分析了内容分发趋势，但关注的是文件类型，而非犯罪者画像。 [33] 通过将行为证据分析（BEA）应用于 P2P 案件，在画像分析方面迈出了一步，展示了数字活动如何揭示犯罪者特征并支持证据重构。然而，他们的研究聚焦于个案，而非大规模画像分析，也未纳入 IP 元数据等开放数据，因此留下了研究空白。

大量研究致力于改进 BitTorrent 协议本身。这些研究优化文件传输、减少跨 ISP 流量，并设计更完善的路由和跟踪机制 [39, 40, 2, 41, 42, 43, 44, 45, 46]。其他研究则分析内容的受欢迎程度和网络使用情况，以改进带宽配置 [47, 48] 或推荐系统 [49]。尽管这些研究对网络性能和效率很有价值，但并未涉及开源情报收集。

更贴近开源情报的是，流量分析提供了实用的方法。在 [50], 中，作者构建了一个异常检测模型，根据行为和地理位置对 IP 地址进行评分；而 [51] 和 [52] 则出于网络安全目的，对加密流量和 VPN 使用进行分类。这些方法提升了网络安全水平，但并未用于用户画像分析。去匿名化和行为聚类提供了更相关的案例。在 [5], 中，研究表明可以从匿名化评分数据集中还原用户兴趣，这说明类似技术也可用于种子下载。 [53] 根据浏览特征，对共享同一 IP 地址的实体进行聚类，为行为归因提供了支持。这些发现与.....的目标一致

开源情报画像，因为种子下载可以揭示与个人用户相关的兴趣模式。 [54] 进一步凸显了利用证书和横幅等多源数据将用户与 IP 地址关联起来的难题。然而，他们的研究范围仍然有限，且仅限于特定数据集和地区。

尽管已有大量关于种子活动和网络行为的研究，文献中仍存在若干重要空白，本研究旨在填补这些空白。

- **缺乏用户级画像：** 大多数研究关注趋势、网络或协议效率，而非个人行为 [41, 2, 45, 46, 1,38]。

- **开源情报方法尚未得到充分运用：** 尽管开源情报方法常用于网络安全和社交媒体分析 [3, 4], ，却很少应用于 P2P数据。

- **交叉参照公开数据的做法有限：** 取证研究 [33] 会将行为与证据关联起来，但并未整合开放来源。

- **关注内容而非背景：** 研究往往追踪文件类别或受欢迎程度，而非用户在不同内容中的行为 [47, 49]。

为填补这些空白，我们提出一种概念验证，将开源情报方法应用于公开可访问的 P2P数据。该工作流程结合种子元数据、Tracker 响应和 IP 信息丰富化，以支持对参与高风险活动的用户进行行为画像分析。这种方法凸显了种子数据与结构化开源情报画像之间鲜少得到探索的交叉领域，为情报研究开辟了新的方向。

4. 方法论

我们提出的工作流程遵循 [15] 中概述的结构，并运用开源情报方法对 BitTorrent 用户进行画像和去匿名化处理。该流程采用一条数据管道，整合多个来源、处理收集到的信息，并通过系统化分析提取洞见。该处理流程包含五个阶段：来源识别、数据收集、数据处理、数据分析和报告编写（图1）。每个阶段都与下一阶段相衔接，从而支持全面且迭代式的调查。

以下各节将详细介绍每个阶段，重点说明所采用的数据源、技术和工具。

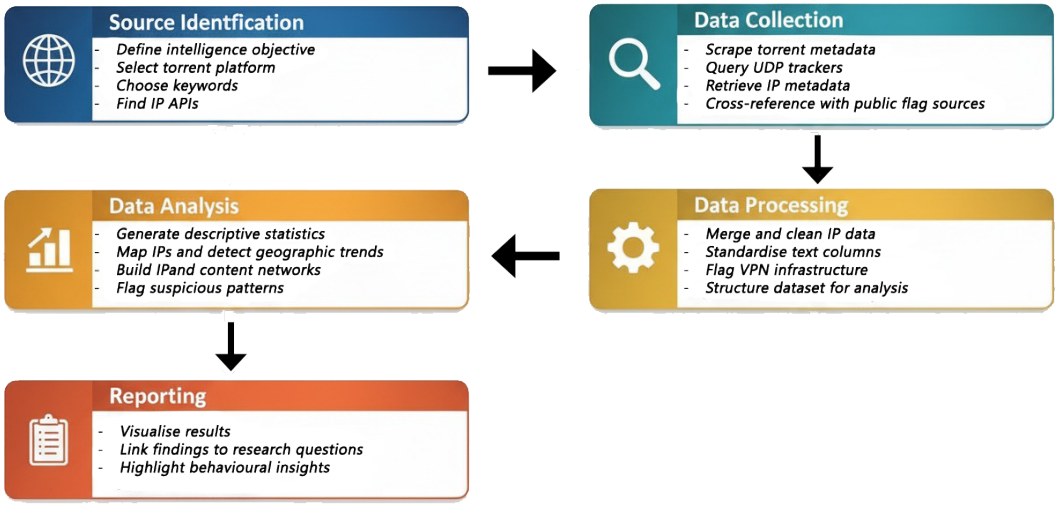


图1：数据管道设计

4.1. 来源识别

将不同的开放数据源与分析方法相结合，以持续构建目标对象的用户画像 [3]，开源情报便能发挥强大作用。基于这一方法，我们设计了一项研究，旨在为 BitTorrent 用户构建用户画像并对其进行去匿名化，尤其关注下载已标记内容的用户。我们重点研究了四类开源数据：(1) 用于获取文件和上传者元数据的种子索引网站；(2) 用于识别活跃节点的 Tracker 响应；(3) 来自公开地理定位和基础设施服务的 IP 元数据；(4) 用于标记高风险行为的种子监控数据库。这些数据集共同呈现了用户活动及其模式的多角度视图。

BitTorrent依赖追踪器和磁力链接。追踪器是一个中央服务器，通过共享用于交换特定文件的对等节点列表来协调群集（群集指整个对等节点网络） [55]。群集中的对等节点通常越多，下载速度就越快 [39, 47]。除了追踪器，分布式哈希表（DHT）还能让对等节点无需依赖中央服务器即可彼此发现，从而增强网络的韧性 [56]。磁力链接则通过用加密哈希值取代.torrent文件，进一步提升了网络韧性。这消除了对固定地理位置的需求，使对等节点能够通过DHT或公共追踪器直接连接 [57, 56]。

尽管效率很高，BitTorrent 仍存在隐私风险，因为其点对点结构会让群集中的任何人都能看到 IP 地址 [46]。与

客户端-服务器下载不同，后者只有服务器能看到请求，而 BitTor- rent 中的对等节点可以监视彼此的活动。监控团体、版权执法机构或恶意行为者都能轻易利用这一点。尽管有些用户会使用虚拟专用网络或代理服务器，但由于 BitTorrent 的设计，许多用户仍暴露在风险之中。

在数据收集方面，我们使用了四个来源。首先，我们从种子索引网站收集元数据。我们选择了海盗湾，这是一个热门的公共索引网站，拥有大量活跃种子、上传者和追踪器 [58, 59]。我们从其 Top 100 Torrents 页面提取了种子名称、类别、上传日期、文件大小、做种者和下载者数量、上传者信息以及磁力链接。做种者持有完整文件并将其分享，而下载者只下载（不分享）。做种者和下载者数量既能反映种子的受欢迎程度，也能反映其健康状况，因为做种者通常越多，下载速度就越快。该数据集展现了文件趋势，但与其他来源结合后，才能获得更深入的洞见。

为了研究种子数据如何揭示高风险或可疑行为，我们在海盗湾上进行了关键词搜索。我们搜索了“explosive(s)”、“IED”和“bomb(s)”，因为这些词常出现在媒体和执法部门关于简易爆炸装置的讨论中[1]。每次搜索都返回了同一位上传者**crwildman**发布的多个种子，他出现在所有搜索结果中。他以“IED”和“bomb(s)”为关键词上传的内容包括简易爆炸物、炸弹制造和自制化学化合物的指南。由于他反复出现，且在2013年上传了大量相关材料，我们选取了他上传的59本书作为代表性子集。借助该数据集，我们得以分析围绕一个明确界定的高风险内容群集所产生的群集活动和用户行为。

第二个数据源是追踪器响应数据。我们抓取了与种子关联的 UDP 跟踪器，以识别活跃节点。虽然 DHT 抓取能够覆盖范围更广、韧性更强的用户群体，但受技术限制，我们选择了中心化 UDP追踪器。我们使用来自海盗湾的磁力链接，获取文件哈希值并将其公布给追踪器。随后，追踪器返回已连接对等节点的 IP 地址和端口，使我们能够了解群集的实时分布及其组成。

第三个来源是 IP 元数据。我们使用公共 API（ip-api.co m

[1]虽然这些术语并非源自正式的关键词列表，但它们在公众、学术界和调查领域中都与简易爆炸装置广泛相关 [60, 61]。

和 ipinfo.io）为 IP 地址补充地理定位、互联网服务提供商、组织所有权和自治系统（AS）编号等信息。这些编号由单个组织管理的一组 IP前缀唯一标识，进一步揭示了基础设施所有权和托管模式。这些服务还提供了有关匿名性措施的指标。这些元数据帮助我们绘制地理分布图、检测托管服务提供商，并识别匿名性措施不足的用户。

最后，我们使用了一个外部监测数据库（iknowwhatyoudownload.com）。该数据库收集与 IP 地址相关的公开可见种子活动，并为其添加行为标签。数据库会标记此前下载过敏感或非法内容（包括 CEM）的 IP 地址。出于伦理和法律方面的限制，我们没有自行存储或处理此类材料，而只是将收集到的 IP 地址与数据库中被标记的条目进行交叉比对。这样，我们无需直接处理非法数据，也能完善风险画像。



图2：以下网站的截图*iknowwhatyoudownload.com*

4.2. 数据收集

我们结合使用了多种工具和库，从不同网页抓取数据。我们采用 Selenium WebDriver，这是一种用于自动化网页浏览器并模拟用户交互的开源框架。此方法使我们能够绕过自动抓取常见的限制并提取网站内容。为应对速率限制并减少被侦测的可能性，我们还在请求之间加入了延迟。我们使用 Beautiful- Soup 解析提取的 HTML，并将相关数据导出为 CSV 文件，以便进一步分析。

对于 UDP 抓取，我们通过改造 m2t（磁力链接转种子）项目 [62]，构建了一个基于套接字的定制抓取器。原始 m2t 项目专注于针对简单元数据的抓取请求，而我们此前已通过从海盗湾进行网页抓取收集了这些数据。因此，我们修改了代码，使其发送宣告请求，从而获取活跃种子群的实时对等节点列表。宣告请求会向跟踪器查询当前参与种子群的对等节点的实时信息，而原始实现并未获取这些信息。为此，我们重新构建了请求载荷，并重写了解析逻辑，以解读二进制跟踪器响应。

UDP 跟踪器使用由 BitTorrent 协议 [63] 定义的二进制字节缓冲区（数据包）。我们的流程首先向跟踪器发送连接请求，其中包含连接 ID 和事务 ID。随后，跟踪器会发放一个新的连接 ID，我们利用它发送宣告请求。这些请求不会下载种子，而是向跟踪器查询当前活跃的对等节点。跟踪器的响应以六个字节编码每个对等节点：前四个字节表示 IP 地址，后两个字节表示 TCP 端口。这种方法能够从 UDP 跟踪器高效提取数据，实时了解种子群组成和活跃参与者。

该处理流程既有效又富有成效。我们总共抓取了 206 个唯一种子的元数据，从 Tracker 响应中收集了 60,369 个唯一 IP 地址，并利用公共 IP 数据源中的元数据对其进行了补充。我们还从涉及敏感书籍的种子中收集了 42 个 IP 地址（见第 4.1 节），并将所有收集到的 IP 地址与公共标记来源进行了交叉比对，识别出 940 个此前与 CEM 分发相关联的地址。完整的描述性统计数据和数据集细分情况见第 5 节。

4.3. 数据处理

收集 IP地址并为其补充元数据后，我们构建了一个数据集，每个 IP地址对应一条记录。每条记录包括城市、国家、互联网服务提供商、自治系统编号、纬度和经度、解析后的主机名、匿名性和任播

指标，以及一个用于标记已知参与 CEM 分发情况的标志。任播是指由多个系统在多个地理位置共享的 IP 地址。我们合并并清理了这些记录，以确保一致性。

我们将批次采集的 CSV 文件合并为一个数据集，删除了重复项，并统一了文本字段格式。互联网服务提供商和组织名称均转换为小写，并更正了不一致的国家和城市条目。我们还将城市名称与区域数据进行交叉核对，以避免错误合并。例如，将“Dar Es Salaam”统一为“Dar es Salaam”，而名称相似但实际不同的地点，如“Lafayette”（路易斯安那州）和“LaFayette”（佐治亚州），则仍分别保留。这些更正提高了汇总的准确性，不过对于地理空间制图而言，仅凭纬度和经度就已足够。

该数据集还包含两个布尔字段：任播和隐私。由于没有任何 IP 地址被标记为任播，我们删除了这一字段。隐私字段表示某个 IP 地址是否与虚拟专用网络、代理或托管服务提供商相关联。隐私字段表示该 IP 地址是否与虚拟专用网络、代理或托管服务提供商相关联。为捕获未被$ipinfo.io$标记的缺失情况，我们扫描了互联网服务提供商字段和自治系统字段中的关键词，例如“云”“托管”“数据中心”“专用互联网”“专用网络”“代理”“TOR”和“虚拟专用网络”。如果发现这些关键词，我们就将隐私标记设为 True。

通过这一流程，我们创建了标准化且经过丰富处理的数据集，可用于按地理、互联网服务提供商或匿名性进行聚类等下游分析。这些准备工作为检测异常值以及识别与高风险种子活动相关的模式奠定了基础。

4.4. 数据分析

预处理完成后，我们对清理并整理好的数据集进行分析，以识别种子参与、地理分布、隐私措施和已标记活动方面的模式。此项分析结合统计方法与数据可视化，揭示关键趋势。

- 描述性统计：我们汇总种子类别、IP 地理分布、匿名性措施的使用情况以及内容受欢迎程度等关键属性，以识别用户和文件特征的一般模式。

- IP 网络分析：基于共享下载构建的 IP 间图揭示了用户之间的关系，并将种群级交互模式可视化。

- 内容网络分析：我们根据共同下载构建关联种子的图，以探索主题群组和跨主题兴趣。
- 异常值检测：将活动频繁或兴趣集中的 IP 地址标记为潜在异常值，并对其进行分析，以评估其与风险的相关性。
- 匿名化检测：我们评估虚拟专用网络、代理服务器和云托管IP的普及程度，并分析它们与敏感内容下载之间的关联，以评估匿名策略。

- 可视化与报告：通过图表、地图和统计图，帮助解读结果并有效传达关键发现。

4.5. 报告并传播研究结果

工作流程的最后阶段是对结果进行整理、解读和呈现，使其便于理解并能提供可采取行动的见解。我们结合网络可视化、地理空间制图、统计摘要和类别细分，突出用户行为、内容关联及风险指标中的模式。这些成果面向开源情报分析师、研究人员和调查人员。

我们使用三种主要的可视化类型：(1) 单部网络图，用于展示 IP 地址、内容和共同下载行为之间的联系；(2) 地理空间热力图和点状地图，用于显示种子参与者的全球分布以及敏感活动的聚类情况；(3) 条形图和表格，用于汇总种子类别、IP特征、匿名化使用情况和下载频率。每种可视化都附有说明文字，将结果与我们的研究问题直接联系起来。

这种报告编写方法将原始元数据转化为情报信号，从而提升应用于种子生态系统的开源情报方法的可解释性。除了支持本研究之外，这套可视化与分析框架也进一步完善了更广泛的开源情报实践。通过在研究社群中分享这些结果，我们为改进点对点网络活动、在线匿名性及更广泛的数字调查分析技术作出了贡献。

5. 数据与描述性统计

我们使用包含三个相互关联表格的关系数据库来存储和分析收集到的 Data：ip_info、tor_info 和 hash_ip。这些表格系统地关联了 IP 地址、种子元数据和共享文件。

指标	热门种子	被标记书籍（案例研究）
种子数量	206	59
唯一 IP 地址	60.369	42
涉及国家	213	19
因 CEM 被标记	940 (1,56%)	3 (7,14%)
因隐私原因被标记（虚拟专用网络/代理）	12.392 (20,53%)	15 (35,71%)

表1：种子数据集汇总统计

ip_info 表以 IP 地址作为唯一标识符，包含地理位置、所有权和隐私属性。每条记录都包含城市、地区细分、国家、互联网服务提供商、组织和自治系统编号。我们还存储了纬度和经度坐标、解析后的主机名、隐私服务标记，以及标记与 CEM 分发相关联的 IP 地址的标记。

tor_info 表存储了种子元数据，并以文件哈希值作为唯一标识符。属性包括标题、类别、子类别、上传日期、文件大小、做种者和下载者数量、上传者用户名以及磁力链接。我们将**crwildman** 上传的书籍分为六个兴趣类别：（1）政府与军事行动；（2）暗杀、杀戮和战斗技巧；（3）自制爆炸物、武器和毒品；（4）生存与规避技巧；（5）撬锁、扒窃和保险箱破解；（6）食人。

Thehash_ip 表将种子与相应的 IP 地址关联起来，从而将用户与共享文件联系起来。它包含引用 tor_info 记录的哈希值，以及指向 ip_info 条目的 IP 地址。

我们通过每日抓取*Top 100 Torrents* 页面收集数据，时间为2024年2月8日至3月4日，共获得206个独立种子。该数据集包含141个电视和电影种子、63个色情作品标题、1个游戏和1本电子书。图3直观展示了数据集按类别和子类别划分的情况，清晰呈现各部分的相对占比和数据集规模。共有9名上传者贡献了这206个种子，其中3名用户各自仅上传了一个标题。

图4显示了分布情况，表明*Cristie65* 在我们的收集期间上传了所有色情内容。

表1总结了热门种子和被标记的书籍的关键统计数据。下文将对每项指标进行更详细的讨论。

热门种子数据集中的主题模式包括“性”，

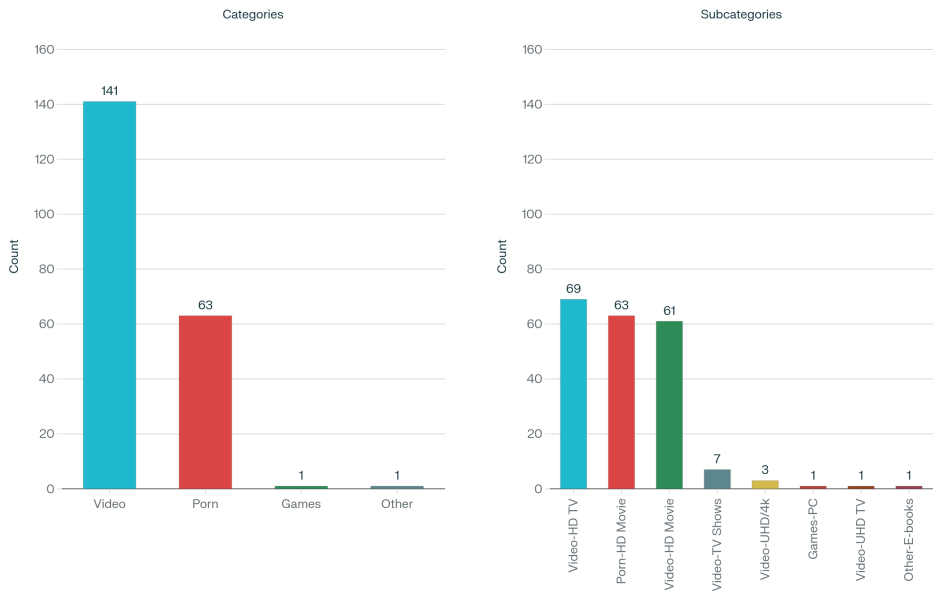


图3：各类别数量

“身体”、“姐妹 ”和“女性”，表明色情或成人主题内容占有很大比重。与此同时，“侦探”、“光环”、“降世神通”和“阿凡达 ”等词语反映出人们对犯罪、科幻和动作等主流娱乐类型持续感兴趣。这些发现与早前的定量结果（图3）相符，后者显示，大多数种子属于电影/电视或色情类别。

我们抓取了相关追踪器的数据，并从213个国家/地区获取了60,369个唯一IP 地址。每个 IP 地址平均关联 1.51 次热门种子下载，最多为 57 次。分析发现，940个 IP 地址（1.56%）因已知的 CEM 分发活动而被标记。此外，我们识别出12,392个 IP 地址（20.53%）极有可能是虚拟专用网络或代理服务器等隐私保护服务的用户。在其余 IP 地址中，有16,591个（27.48%）既没有隐私标记，也没有可识别主机名，是匿名化程度最低的群体。这一未匿名化子集中包含79个被标记为 CEM 相关的 IP 地址，占所有 CEM 相关地址的8.40%。相比之下，75.85%的 CEM标记地址使用了隐私保护技术，表明涉及此类活动的用户匿名化比例显著更高。这一模式表明，

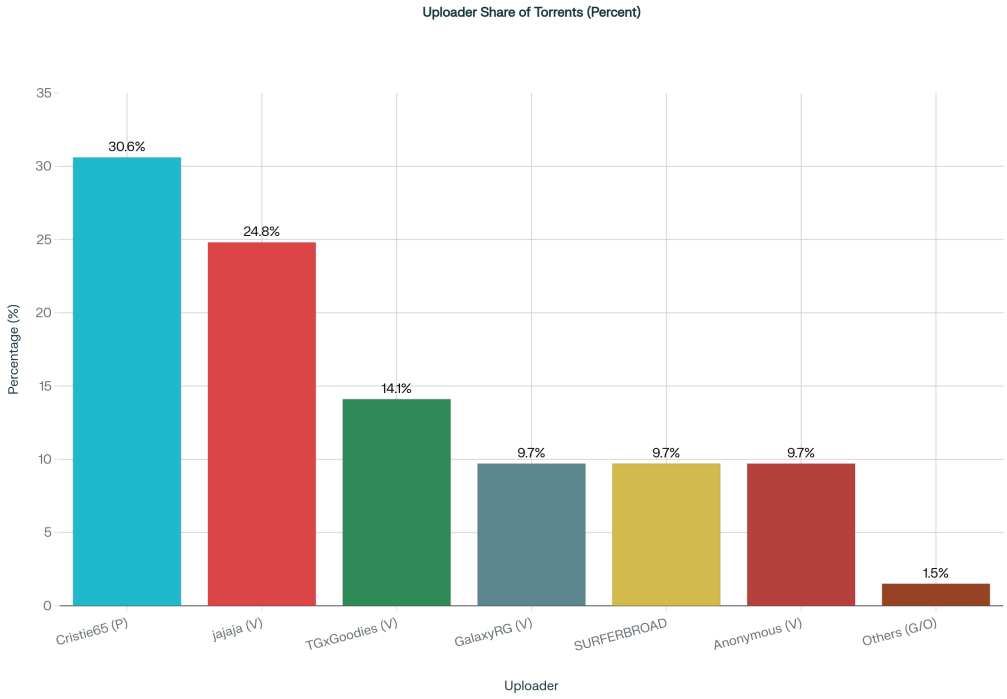


图4：热门种子的上传者及其占比（在 206 个种子中）

下载敏感内容的个人会主动使用隐私工具来隐藏身份。虽然大多数 IP 元数据完整，但许多条目缺少组织数据，限制了详细分析的可能性。不过，我们成功获取了所有 IP 地址的互联网服务提供商信息。表2列出了我们数据集中最常使用的十家互联网服务提供商。

最常见的互联网服务提供商是 *Datacamp Limited*，占捕获的 60,369 个 IP 地址的 2.46%。这些 IP 地址中近 99% 使用隐私保护技术，这与该公司作为 Surfshark 和 NordVPN 等商业 VPN 服务主要基础设施提供商的角色相符。

匿名化与较高风险信号密切相关。所有 IP 地址中有 1.56% 带有 CEM 标记，而与 *Datacamp Limited* 相关联的 IP 地址中，这一比例上升至 7.81%。这一模式证实了我们先前的发现：从事高风险活动的用户更有可能采用匿名化策略。

Tefincom S.A. 占 IP 地址的 2.18%，并且与

ISP	捕获的 IP 数量	占总数百分比
Datacamp Limited	1486	2,46%
Tefincom S.A.	1315	2,18%
m247 europe srl	792	1,31%
philippine long distance telephone co.	785	1,30%
Chinanet	768	1,27%
claro nxt telecomunicacoes ltda	694	1,15%
cogent communications	615	1,02%
clouvider limited	489	0,81%
rcs & rds	473	0,78%
packethub s.a.	462	0,77%

表2：最常见的互联网服务提供商及其占总 IP 地址的比例（$n = 60.369$）

NordVPN 的开发存在历史渊源。Data 反映了这一联系：*Tefincom*的 IP 地址中有 100% 显示虚拟专用网络或代理标记。

排名前十的互联网服务提供商中，有九家的匿名化率都同样偏高，其 IP 地址中超过 95% 被标记为使用了保护隐私的技术。这一趋势凸显了种子生态系统中虚拟专用网络基础设施的广泛使用。唯一的例外是 *Claro NXT Telecomunicações Ltda*，这家巴西大型互联网服务提供商没有任何匿名化标记，表明其用户活动更加直接，未经过混淆处理。

我们将地理位置数据绘制成热力图，以进行空间分析。图5展示了按国家划分的 IP 地址全球分布情况，其中美国、英国、加拿大、巴西和南非所占比例最高。这一分布与 [37], 的发现相符：较富裕国家的盗版绝对数量更高。由于热力图着重呈现模式，而非精确数量，表3通过列出 IP 地址数量最多的十个国家及其在数据集中的占比，对可视化结果进行了补充。

为考察可疑的下载行为，我们分析了用户“*crwildman*”于2013年1月上传的59本书的书名。虽然其中23个种子目前没有活跃参与者，且只有15个有多个对等节点，但这些文件在十多年后仍吸引着用户下载。

图6展示了各内容类别中书名的分布。“自制爆炸物、武器和毒品”类别中的书籍数量最多，活跃下载量也占多数。活跃节点最多的三个种子都与爆炸物有关：军事 爆炸物化学 必备电子书、*A Guide to Field Manufactured Explosives.pdf* 和*Explosives and Weapons - Homebuilt Claymore Mines*。在测量期间，我们发现每个种子都有四个活跃

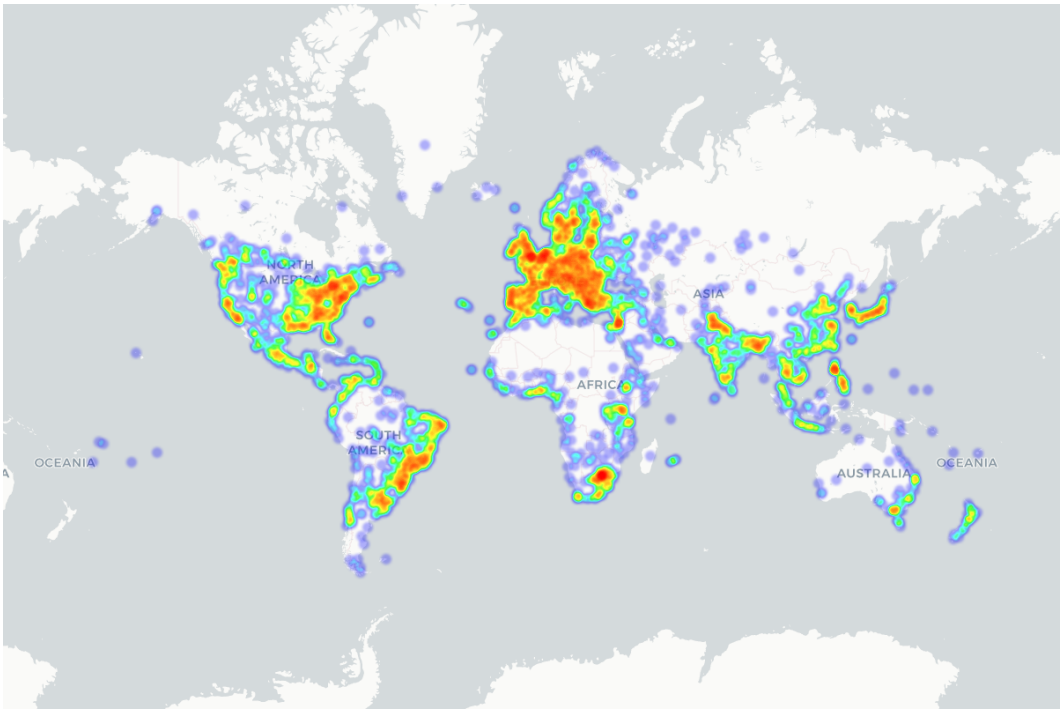


图5：热门种子活动的地理分布

国家	数量	占比
美利坚合众国	5.181	8,59%
英国	3.263	5,41%
加拿大	3.037	5,03%
巴西	2.835	4,70%
南非	2.733	4,53%
澳大利亚	2.198	3,64%
菲律宾	1.871	3,10%
荷兰	1.815	3,01%
中国	1.768	2,93%
葡萄牙	1.528	2,53%

表3：数据集中 IP 地址数量排名前 10 的国家（$n = 60.369$）

我们测量期间每个种子的对等节点。

用户“*crwildman*”上传的59本书中，出现频率最高的词包括“爆炸物”、“手册”、“生存”、“审讯”、“攻击”和“中央情报局”，反映出内容明显侧重于作战战术、简易武器以及情报相

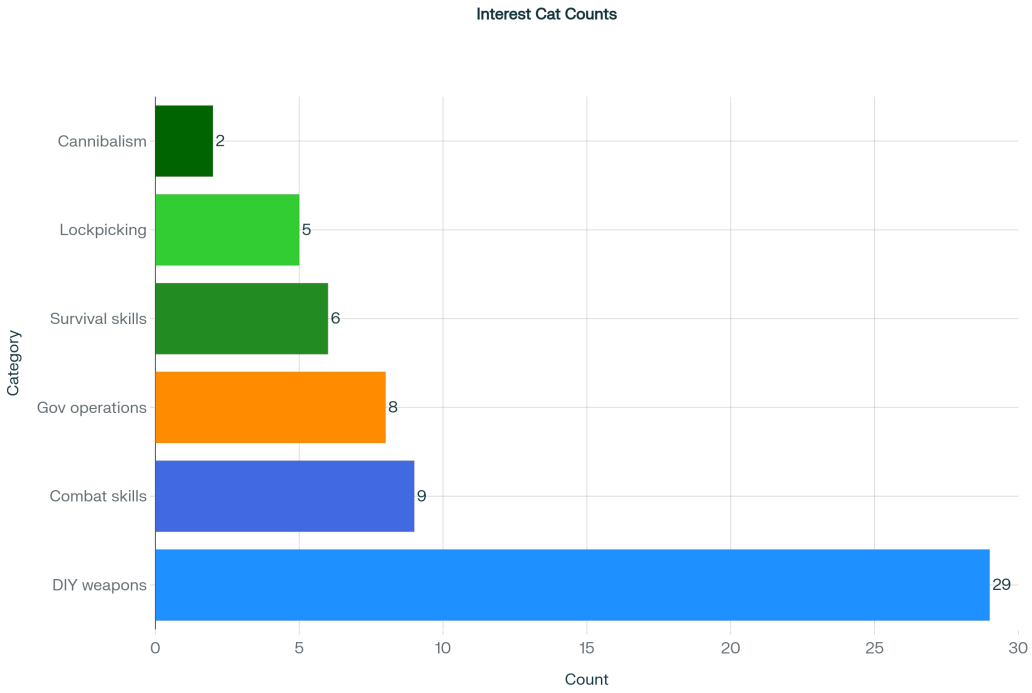


图6：各类别的书籍数量

关技术。其他词语，如“指南”、“催眠”、“毒药”和“游击战”，表明其中包含心理战和化学战内容，以及生存主义和颠覆性战术。

与一般种子数据集相比，这些材料中暴力或非法指导内容的比例明显更高。内容重点与我们的分类结果相符，支持将这一收藏视为开源情报画像分析中的一个重要风险因素（由于篇幅有限，我们在此不提供每个种子所对应的59个书名及其类别和对等节点数量的完整详情）。

我们抓取了电子书群，并识别出 42 个唯一 IP 地址，其中两个也出现在热门种子数据集中。这些 IP 地址来自 19 个国家：12 个来自美国，5 个来自加拿大，来自中国和荷兰的各有 3 个，其他国家各有 1—2 个。图7展示了这一地理位置分布，形状越大表示下载的书籍数量越多。圆圈表示最不可能使用匿名化的 IP 地址

，而红色标记表示与 CEM 分发相关联的 IP 地址。

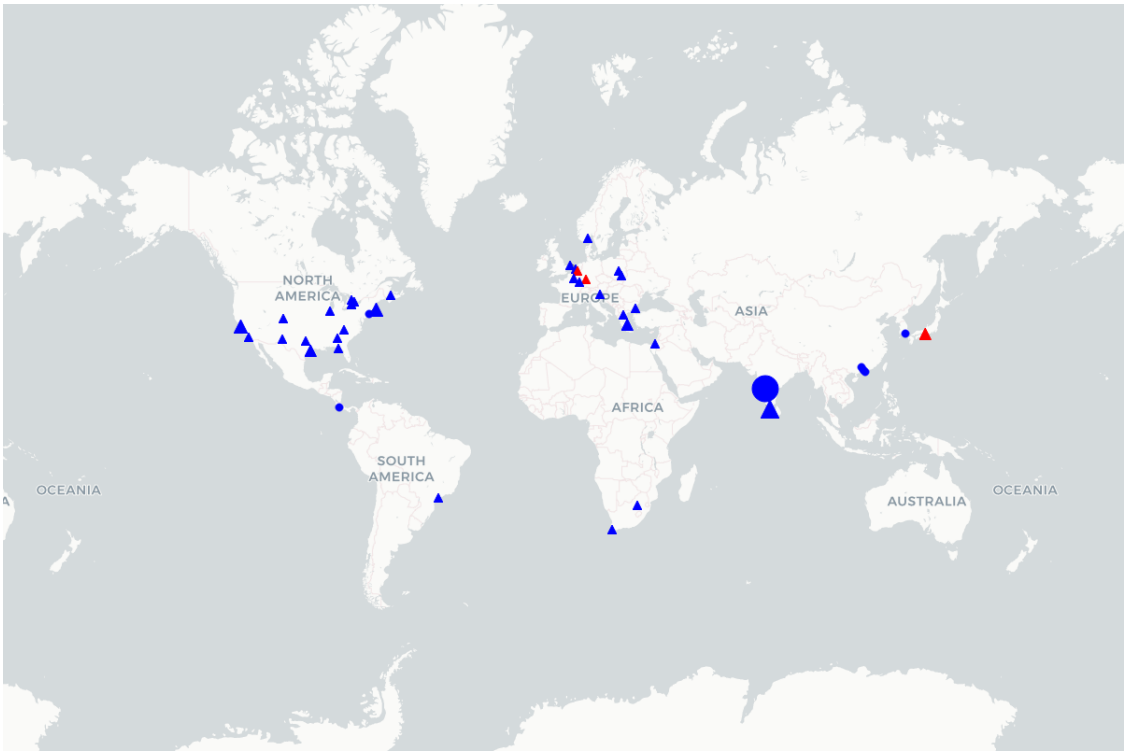


图7：下载敏感书籍的 IP 地址地理分布

我们发现，42 个 IP 地址中有 3 个带有 CEM 分布标记。其中两个访问了自制爆炸物材料：一个下载了一份野外制造的爆炸物指南，另一个获取了化学品供应商名单。第三个 IP 地址与一本包含格斗技巧内容的书籍有关。

约三分之一的 IP 地址可能通过虚拟专用网络或代理服务进行匿名化。在最不可能使用匿名化的 7 个地址中，大多数访问了自制爆炸物书籍。

表 4 展示了图7中的关键数据。对于拥有多个 IP 地址的国家，表中列出了下载总量、IP 地址数量、隐私服务的使用情况以及带有 CEM 标记的 IP 地址。地图直观地展示了空间分布，而表格则更清晰、易于理解地呈现了底层数据。

美国的下载量最高

国家	下载量 IP 数量	VPN/代理使用	CEM 标记的 IP	
美国	17	12	7	0
印度	14	2	0	0
加拿大	5	5	3	0
中国	3	3	0	0
荷兰	3	3	2	1
希腊	3	2	0	0
波兰	2	2	1	0
南非	2	2	0	0

表 4：被标记书籍下载量最多的国家

（17）且独立 IP 地址数量最多（12）。加拿大和荷兰的用户使用虚拟专用网络或代理的比例相对较高，表明他们的隐私意识较强。

一个值得注意的发现是与印度相关的活动：仅有两个 IP 地址就产生了 14 次下载。在图7中，这表现为两个大小明显偏大的圆点，其中最大的圆点代表一个单独的 IP 地址，该地址产生了 9 次下载。这种集中的活动表明人们对相关内容有针对性的兴趣，后续章节将对此进行更深入的考察。

6. 实验结果

为了测试我们的方法利用公开数据映射种子用户网络的效果，我们构建了一个包含两种节点类型的二部图：IP 地址和种子。随后，我们将其转换为单部IP间图，其中的边表示共同参与——即抓取时两个 IP 地址出现在同一群集中。

由于 IP 地址数量庞大，网络难以解读，因此我们按频率筛选，仅保留排名前 0.01% 的 IP间连接。这些边代表最强的行为关联——跨不同种子的重复共同下载，这可能表明用户行为模式稳定或使用共享网络。我们设定了较高的网络纳入活跃度阈值：IP 地址必须至少下载过七个热门种子。在符合条件的 IP 地址中，我们发现了明显的内容偏好：217 个仅下载电视和电影内容，140 个只关注色情内容，11 个同时访问这两个类别，另有 4 个下载了书籍和电视内容。

图 8 展示了这个高频网络，其中包含 372 个 IP 地址。每个节点都根据 IP 元数据进行地理定位，节点样式

同时表示匿名性和 CEM 标记状态:

- 圆形 =未标记为匿名化
- 三角形 = 匿名化 IP（检测到虚拟专用网络/代理）
- 红色 = 已标记为已知 CEM 分发
- 蓝色 =未标记

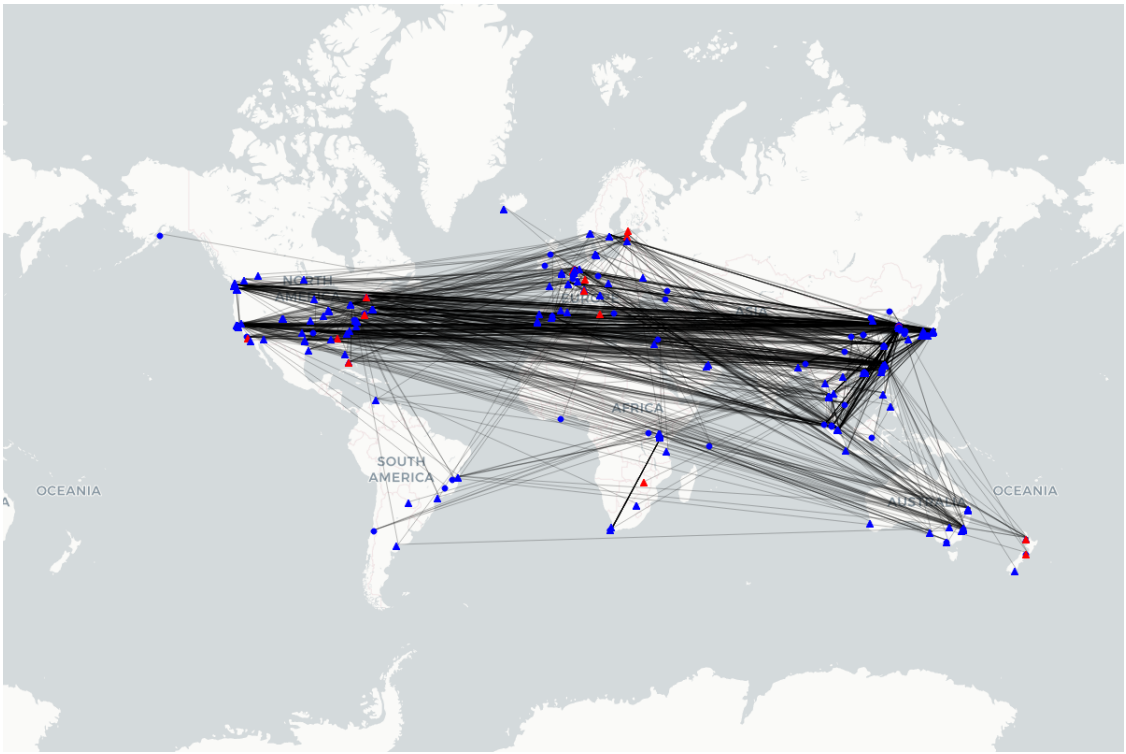


图 8：连接频率为 0.01% 的网络。

该网络展现了一个全球分布的用户群体，其中存在多个密集集群。地理聚类现象十分明显，许多高频连接发生在特定地区内。不过，洲际联系也很常见，这凸显了种子群的去中心化特性，以及 海盗湾内容的国际传播范围。

边的粗细表示共同下载频率，线条越粗，表明共同参与种子群的频率越高。不过，在当前筛选级别下，这些视觉差异

仍不太明显。为更清楚地展示最显著的关系，表5列出了共同下载频率最高的前十组 IP 对 IP 共同下载对。每一行代表一组独特的 IP 对，它们反复出现在同一批种子群中，这表明它们可能通过一致的下载时间、共同兴趣或共享基础设施形成了紧密的行为关联。

出于隐私和伦理方面的考虑，表中列出了每个 IP 地址的大致坐标和所在国家，而非原始地址。这些信息充当假名标识符，使我们能够在网络图和表格中追踪相关记录，同时不暴露私人信息。表中还显示了匿名化状态，以及任一地址是否带有 CEM 分布标记。该子集中所有高频交互都涉及色情内容。若干 IP 地址（以单元格颜色标示）出现在多个配对中，这表明它们可能是活跃度很高的用户，也可能是跨多个群充当持久对等节点的节点——这种系统性行为值得进一步调查。

表5：出现频率最高的前10组 IP 地址（坐标）配对

注： 单元格颜色根据坐标标示重复出现的 IP 地址。相同颜色表示不同配对中的 IP 地址相同。这些颜色与表7中所示的 IP 地址颜色一致

IP地址1坐标	IP地址2坐标	频次	Cat.	国家	VPN	CEM 标记
(37.2980, 127.0777)	(31.2222, 121.4581)	34	Porn	韩国/中国	否/否	否/否
(37.5245, 127.0354)	(37.2980, 127.0777)	30	Porn	韩国/韩国	否/否	否/否
(37.2980, 127.0777)	(24.1440, 120.6844)	25	Porn	韩国/台湾	否/否	否/否
(24.1440, 120.6844)	(37.4483, 127.0681)	24	Porn	台湾/韩国	否/否	否/否
(24.1440, 120.6844)	(31.2222, 121.4581)	23	Porn	台湾/中国	否/否	否/否
(37.2980, 127.0777)	(37.4483, 127.0681)	23	Porn	韩国/韩国	否/否	否/否
(35.6893, 139.6899)	(37.2980, 127.0777)	23	Porn	日本/韩国	是/否	否/否
(37.2980, 127.0777)	(1.3521, 103.8200)	22	Porn	韩国/新加坡	否/是	否/否
(48.4471, -123.3018)	(24.1440, 120.6844)	22	Porn	加拿大/台湾	否/否	否/否
(37.5245, 127.0354)	(31.2222, 121.4581)	22	Porn	韩国/中国	否/否	否/否

图 8 展示了全球高频对等节点交互，并体现了我们的方法在识别和分析目标子网方面的能力。通过使用公开元数据和灵活的筛选标准，我们可以创建反映特定行为模式或风险指标的专业网络视图。

一个例子是检测有已知 CEM 分发历史的 IP 地址。我们必须澄清，**本研究不包含或处理任何含有已知 CEM 内容的种子**。相反，一些出现在主流种子群中的 IP 地址带有此前曾参与 CEM 分发活动的独立标记。它们的出现并不意味着它们在我们收集的样本中参与了

非法活动，但它们在不同种子群中的行为重合提供了重要的风险信号。这一发现表明，即使不涉及明确的 CEM 场景，种子元数据也有助于识别高风险用户模式。

图9展示了一个由70个被标记为儿童性虐待材料相关的IP地址构成的聚焦子网络，这些地址仅参与了色情种子的下载。边表示共同参与同一群集。与更广泛的网络相比，该子网络明显更稀疏，仅有272条边，但围绕若干地理集群呈现出清晰的结构。东亚、欧洲部分地区和美国的活动密度较高。这一模式表明，被标记为儿童性虐待材料相关的用户会聚集在特定内容细分领域中，支持了这样一种假设：内容类型有助于缩小开源情报应用中的行为画像分析范围。

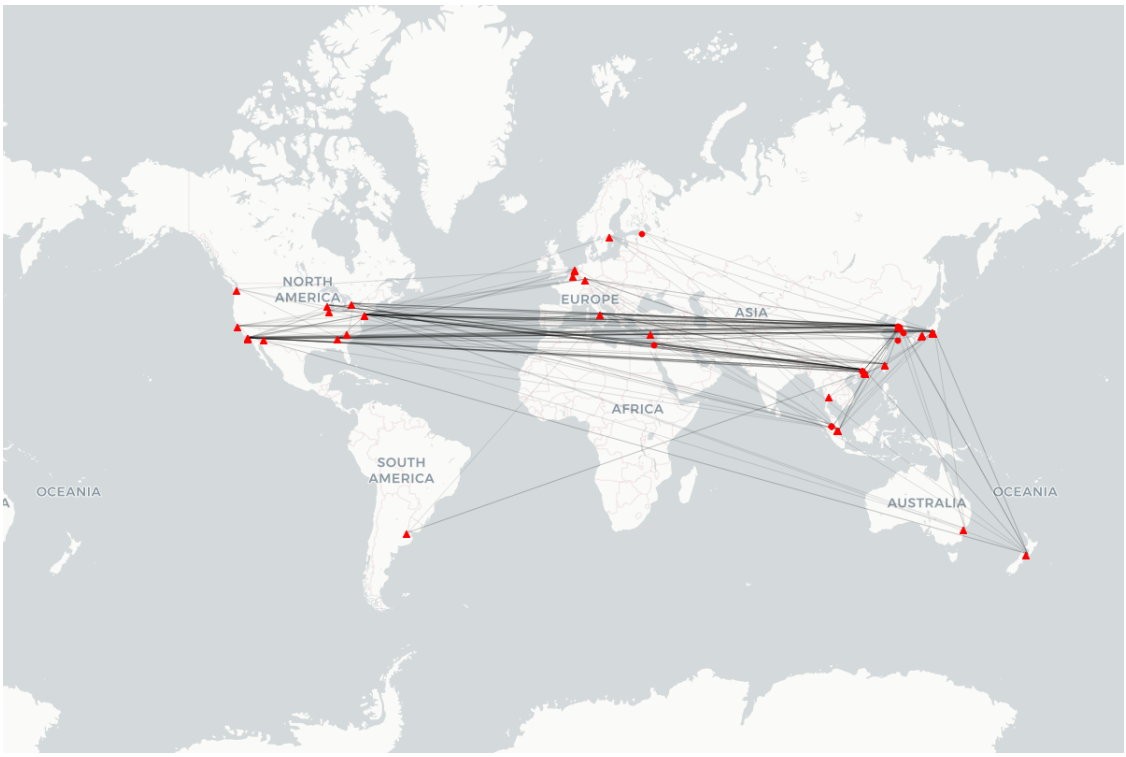


图9：数据集中仅下载色情内容的CEM标记IP地址网络地图。

我们通过介数中心性分

析识别出若干具有高度影响力的节点，它们在网络中充当结构性桥梁。这些节点连接了原本联系较弱的 IP 集群，表明它们可能在促进仅访问色情内容的用户之间的行为重叠或反复群集活动方面发挥着重要作用。

虽然该网络遍布全球，但中心性最高的节点集中在北美、东亚和欧洲部分地区。这种模式表明，活动枢纽扎根于各个区域，同时保持着强大的国际联系。识别主要枢纽和规模较小的参与者，有助于确定调查工作的优先顺序并监测升级风险。

表 6 展示了 CEM 标记的色情网络中介中心性得分最高的十个 IP 地址，并按得分排序。我们选择这一指标，是因为它能识别出充当原本互不相连的网络部分之间桥梁的节点，从而可能揭示那些能够助推高风险内容更广泛传播的参与者。每行包含大致坐标、国家、中心性得分、虚拟专用网络使用情况和下载次数。我们添加了 CEM 列，以提醒读者注意标记状态。

表 6：按中介中心性排名前 10 的 IP 地址（坐标）（CEM标记）

坐标	国家	介数	度	虚拟专用网络	下载量	CEM
（40.71230, -74.0068）	美国	0.403	45	True	23	Yes
(40.94680, 14.3015)	意大利	0.170	24	False	13	Yes
(43.89780, -79.2617)	加拿大	0.101	14	假	5	Yes
(33.67500, -118.0027)	美国	0.095	28	假	11	Yes
(35.68930, 139.6899)	日本	0.087	6	True	3	Yes
(37.53320, 126.9692)	韩国	0.086	9	否	4	Yes
(37.52450, 127.0354)	韩国	0.080	26	否	10	Yes
(33.74850, -84.3871)	美国	0.066	20	True	9	Yes
(1.28967, 103.8500)	新加坡	0.053	10	True	3	Yes
(37.53600, 126.9710)	韩国	0.052	11	否	5	Yes

这些指标揭示了每个节点的位置和行为特征。与早期分析不同，东亚和美国在此处表现最为突出，其中韩国和日本贡献了多个中心性较高的 IP 地址。排名最高的 IP 地址兼具较高的群体参与度和匿名化特征，这可能表明其行为具有系统性。相比之下，其他几个中心性较高的 IP 地址并未使用匿名化，因此可能更有调查价值，因为与之相关的元数据或许更可靠。

图10通过纳入标记 IP 地址与未标记对等节点之间的直接连接，扩展了标记IP网络。这个扩展视图仅显示与标记 IP 地址相连的边；为便于阅读，我们排除了未标记 IP 地址之间的边。

完整网络包含 3,354 个 IP 地址和 374,327 条连接。图9仅聚焦于标记节点，而这个更新后的视图还纳入了它们的直接邻居；这些邻居也只参与了热门色情种子。这样的筛选使我们能够考察被标记用户与主流成人内容参与者之间的行为重叠。

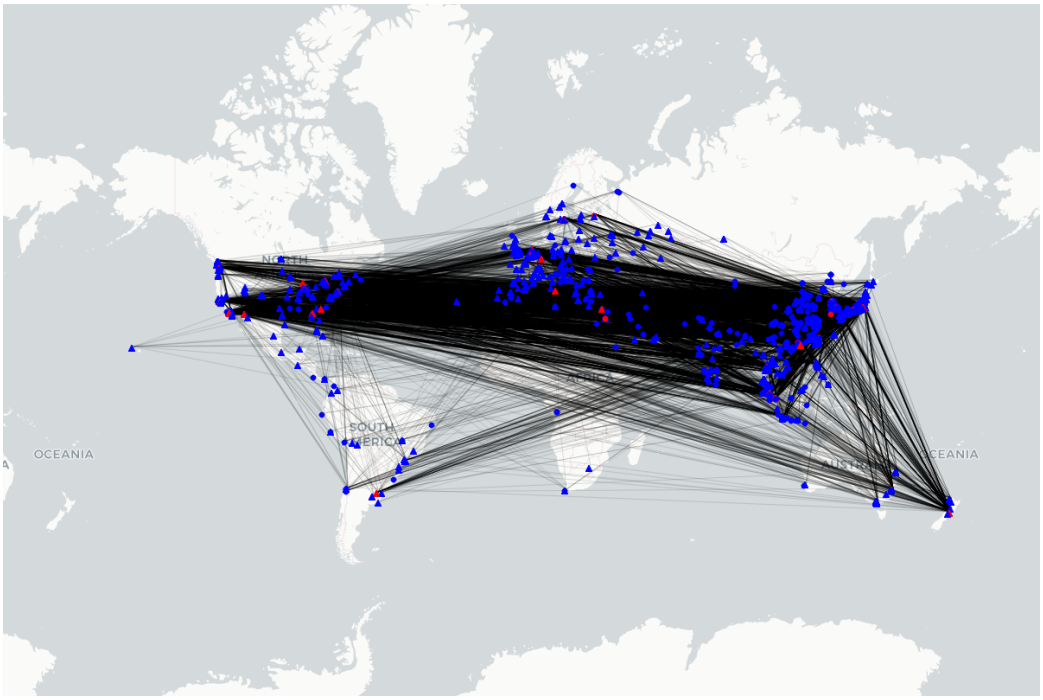


图10：分发 CEM 的 IP 地址扩展网络地图（仅显示涉及已标记 IP 地址的边）

为支持这一可视化，表7列出了扩展网络更新后的中心性指标。图10展示的是经过筛选的边（仅包含涉及已标记 IP 地址的边），而表7则基于完整网络进行计算，其中也包括非已标记 IP 地址之间的边。

与仅展示 CEM 标记的 IP 交互的表 6 相比，扩展表揭示了网络影响力的显著变化。现在，按介数中心性排名最高的节点均为未标记用户，几乎全部位于东亚，并在色情内容群集中充当密集的本地枢纽。然而，在这个更密集的网络中，最大介数中心性显著下降。

表7：按中介中心性排序的前10个坐标（CEM标记，扩展网络）

注： 单元格颜色表示根据坐标识别的重复 IP 地址。相同颜色表示该 IP 地址在本表和表5中相同

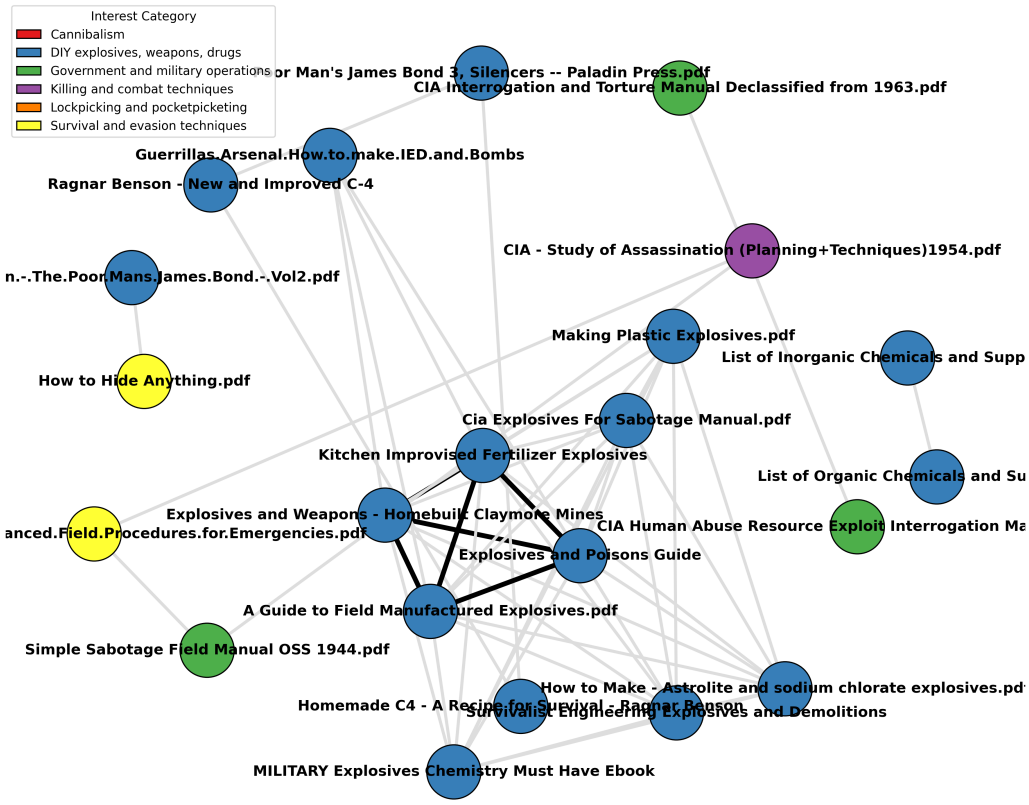
坐标	国家	介数	度	VPN下载量	CEM		
(37.2980, 127.0777)	韩国	0.067	2819	假	51	No	
（31.2222，121.4581）	中国	0.047	2499	假	41	No	
（37.5245，127.0354）	韩国	0.028	2171	否	34	No	
(35.6893, 139.6899)	日本	0.027	1930	True	32	No	
(37.4483, 127.0681)	韩国	0.022	2127	否	29	No	
(47.6034, -122.3414)	美国	0.021	1710	True	17	No	
(24.1440，120.6844)	台湾	0.018	2105	否	32	No	
(31.2222, 121.4581)	中国	0.016	1744	否	22	No	
(1.3321, 103.8940)	新加坡	0.015	1493	True	18	No	
(37.4897, 127.0639)	韩国	0.013	1756	True	24	No	

许多具有高中心性的 IP 地址也出现在表5中，该表记录了共同下载量最多的 IP 对。它们反复出现（通过一致的单元格着色标示），表明这些 IP 地址既处于结构中心位置，在行为上也十分突出。它们位于连接交汇处，并反复与其他用户进行共享下载，这表明它们可能是各自群集中的核心参与者或行为锚点。

图10展示了 CEM标记用户如何持续与未标记用户共同参与种子群集。两组用户都参与了纯色情内容种子的传播，表明他们的内容偏好一致。这种模式引发了关于种子生态系统中内容传播路径和用户聚类机制的重要问题。全球连接分布如今更明显地延伸至南美洲和西亚，展现了敏感内容如何通过共享的群集活动，在彼此分离的用户群体间传播。

尽管 CEM 分发清晰地标示了高风险活动，但它并非可疑行为的唯一指标。我们的数据集还包括被标记为含有敏感或有害材料的种子，尤其是由“*crwildman*”上传的书籍，这构成了情报框架中风险评估的另一个重要方面。

为了探索用户交互之外的模式，我们将关注点转向二部图中的内容侧。我们构建了一个内容-内容网络，其中节点代表单个种子，边表示共享下载者，从而基于用户行为为文件关系提供互补的见解。



F图 11：用户“*crwildman*”上传书籍的内容到内容网络

图11展示了生成的内容-内容网络。每个节点代表一个可能包含高风险内容的种子，边则表示共同下载关系。我们按主题类别为节点着色，而边的颜色和粗细则表示共同下载频率。

该网络显示，爆炸物相关内容周围形成了密集的集群，表明这些用户之间存在紧密的主题关联。较小的群体则关注生存指南、暗杀手册和化学品供应清单，这说明部分用户的兴趣跨越多个主题。度中心性和介数中心性揭示了网络中的四个核心节点。表8汇总了中心性最高的标题，并列出每个标题对应的唯一对等节点数量、度和介数得分。您可以将关联标题与表 9 对照，以获取更多背景信息。

表8：敏感书籍网络指标与对等节点发现

标题	发现的对等节点	度	介数	关联标题 (#)
《野外制造爆炸物指南》sives.pdf	4	0.429	0.006	2、3、4、5、6、7、8、9、10
《爆炸物与毒物指南》	2	0.429	0.006	1、2、4、5、6、7、8、9、10
Explosives and Weapons - 自制阔剑地雷 自制阔剑地雷	4	0.429	0.006	1、2、3、5、6、7、8、9、10
厨房自制肥料炸药	2	0.429	0.006	1、2、3、4、5、6、8、9、10
Cia 爆炸物 用于破坏活动 Manual.pdf	3	0.381	0.000	1、3、4、6、7、8、9、10

表 9：按编号排列的关联标题参考列表

#	关联标题
1	A Guide to Field Manufactured Explosives.pdf
2	CIA 炸药破坏手册.pdf
3	《爆炸物与毒物指南》
4	Explosives and Weapons - 自制阔剑地雷
5	Guerrillas.Arsenal.How.to.make.IED.and.Bombs
6	如何制作——阿斯特罗莱特炸药和氯酸钠爆炸物.pdf
7	厨房自制肥料炸药
8	制作塑胶炸药.pdf
9	MILITARY Explosives Chemistry Must Have Ebook
10	生存主义工程学：爆炸物与爆破
11	海王.and.the.Lost.Kingdom.2023年.1080p.网络版转录.1400MB.杜比数字5.1声道.x264编码

这些标题的度中心性得分较高，这意味着下载这些标题的用户往往也会下载其他多份敏感文档，而非单独访问这些文档。这些文档经常出现在共同下载模式中，属于更广泛兴趣簇的一部分，而不是孤立的下载内容。

此外，大多数文件的中介中心性得分均非零，这表明它们充当了不同网络区域之间的桥梁。某些文档可能成为用户探索更广泛高风险内容的入口或交汇点。相比之下，“*CIA 炸药破坏手册.pdf*”的度较低，且中介中心性得分为零，这表明访问该文件的用户群体较为有限，也不会连接不同的集群。

为了更好地了解用户行为，我们通过添加同一批用户下载的热门种子来扩展内容-内容网络。这

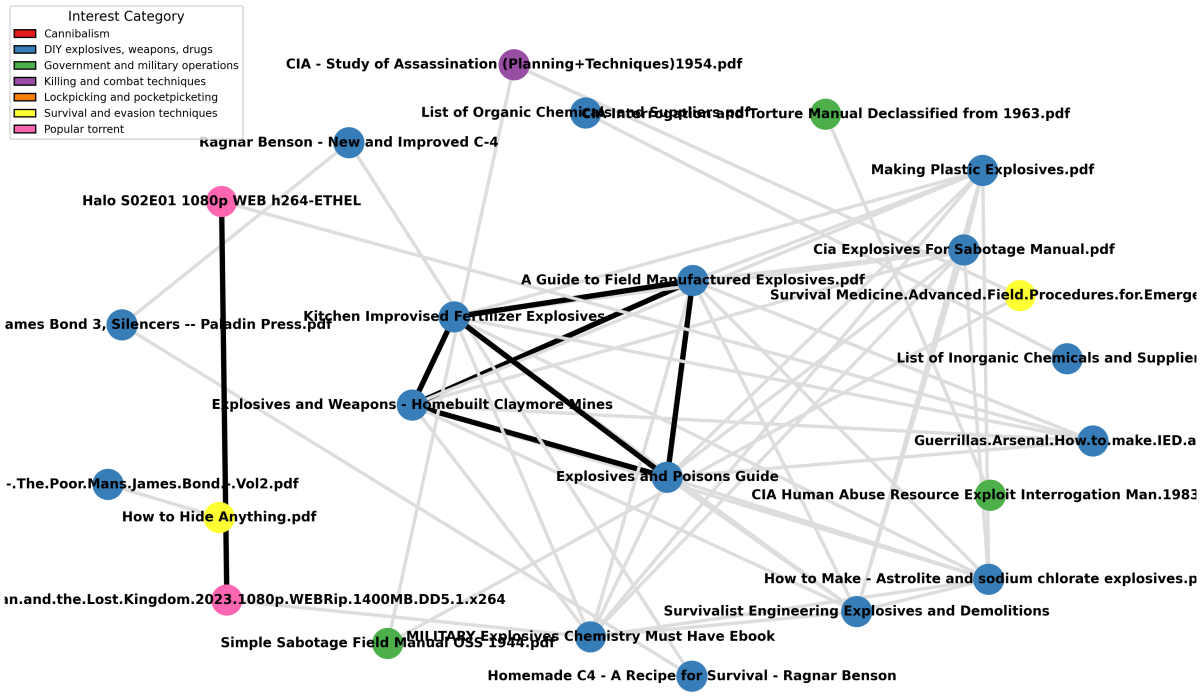


图12：扩展的内容到内容网络

个更广泛的网络（图12）将高风险文档与主流媒体下载联系起来，从而更深入地洞察用户兴趣。

图12揭示了多个层面的可操作洞见。首先，它有助于识别那些下载模式同时关联高风险文档和主流娱乐内容的用户，从而形成行为指纹。敏感文件与热门种子之间的关联，使人们能够基于主题共消费进行画像分析。

此外，该网络还凸显了用户参与度的差异：有些用户会下载多个敏感文件和各种热门种子，而另一些用户则只是单点连接，可能属于随意或一次性接触。网络中的簇显示出主题重叠，例如游击战文献与军事主题娱乐内容之间的重叠（例如*Halo S02E01 1080pWEB h264-ETHEL*与*Guerrillas.Arsenal.How.to.make.IED.and.Bombs*之间）。这些模式能够依据主题共消费行为进行画像分析，相较于仅根据敏感内容

仅凭下载量。

为补充图12所呈现的直观洞察，表10按照与表8相同的结构，列出了扩展内容到内容网络中关键种子的网络指标。

表10：扩展网络：文档指标与对等节点发现。

标题	发现的对等节点	度	介数	关联标题 (#)
军事 爆炸物 化学必备电子书	4	0.391	0.035	1、2、3、4、6、7、8、10, 11
厨房自制肥料爆炸物-爆炸物	2	0.391	0.008	1、2、3、4、5、6、8、9、10
野外制造爆炸物指南-爆炸物.pdf	4	0.391	0.008	2、3、4、5、6、7、8、9、10
《爆炸物与毒物指南》	2	0.391	0.008	1、2、4、5、6、7、8、9、10
Explosives and Weapons - Homebuilt	4	0.391	0.008	1, 2, 3, 5, 6, 7, 8, 9,
自制阔剑地雷				10

这些指标印证了早前的发现：有关简易爆炸物的文档在度和中介中心性方面排名最高，并反复构成重点网络和扩展网络的核心。它们的核心地位凸显了其吸引力和主题重要性。值得注意的是，*MILITARY ExplosivesChemistry Must Have Ebook* 是唯一一份直接链接到热门种子的高排名文档，这表明有些用户既会接触高风险的指导性内容，也会下载DC漫画等主流媒体。这类交集有助于通过捕捉多种兴趣的下载模式来完善用户画像。将种子元数据与网络分析相结合，可以改进风险评估、升级检测和调查优先级排序，从而支持保护隐私的开源情报方法。

虽然更广泛的网络分析揭示了普遍的升级风险，但异常值也提供了有力的调查线索。如图7所示，孟买有一个IP地址因其异常活跃而格外突出——该用户下载了九本关于简易爆炸物的书籍，却没有访问任何热门种子，这表明其对高风险材料抱有刻意的兴趣，而非只是随意分享文件。结合其住宅互联网服务提供商归属（*Space Vision Digital Network Pvt. Ltd.*）及具体的地理位置，其活动表明升级风险较高，值得进一步审查。这些结果表明，公开可用的种子元数据如何支持基于行为的细分，并强化P2P画像。上下文元数据和精

细筛选对于使用开源数据进行风险评估、预警和确定调查优先级至关重要。

下一节将聚焦种子元数据对开源情报工作流程的技术价值。我们将评估其适用性和效率，涵盖可用性、结构、可靠性和处理工作量。

种子元数据和 IP 元数据的可用性都很高：海盗湾 保持在线，追踪器也能提供许多 IP 地址。IP 元数据来源支持可靠的地理定位和互联网服务提供商查询，有助于开展实时开源情报。

种子元数据属于半结构化数据——类别、上传者和磁力链接等字段较为一致，但文件标题会因上传者而异。这限制了自动化，因此需要进行预处理。哈希值本身提供的背景信息有限，而 Tracker 查询对于获取缺失数据往往不可靠。

IP 元数据的结构差异更大：通常会包含城市、国家、互联网服务提供商和自治系统字段，而地理坐标有助于缓解文本不一致的问题。尽管偶尔会出现错误，IP 元数据对于地理空间情报仍然可靠。不过，匿名化技术会降低地理定位的准确性；样本中约有 20% 的 IP 地址使用了隐私保护服务。虚拟专用网络的使用情况可能存在漏报，因此我们检查了互联网服务提供商/自治系统字段中的隐私信号。尽管如此，信誉良好的提供商通常能提供可靠的数据，但匿名化活动仍是一项主要局限。

使用 Selenium 和 BeautifulSoup 等工具抓取种子元数据效率很高。UDP Tracker 查询需要二进制套接字编程和稳健的解析，不过我们的系统可以重复使用。处理大量种子耗时较长，而且对等节点列表可能不完整，因此定期抓取可确保更高的准确性。

提取 IP 元数据需要额外的后处理，因为公共 API（例如$ip\text{-}api.com$ 和 $ip\text{-}info.io$）设有速率限制，且数据存在差异，因此若要扩大规模，就需要使用备用方案或付费选项。

总之，种子和 IP 元数据是技术上可行但较为复杂的开源情报数据来源。它们的可靠性和结构各不相同，因此灵活的提取系统对于稳健的分析至关重要。

7. 结论

本研究探讨了如何利用种子元数据，系统性地支持旨在识别非法内容及传播高风险材料用户的开源情报工作流程。通过概念验证流水线，我们展示了

公开可获取的种子元数据——例如 Tracker 响应、IP地址地理定位和跨来源标记——如何揭示去中心化且部分匿名网络中的行为模式。通过整合这些元数据来源，该方法在符合隐私保护分析标准的同时，提升了用户画像、风险识别和调查优先级排序的能力。

本研究的核心研究问题如下：

研究问题： 我们的概念验证在多大程度上展现了利用种子元数据进行基于开源情报的用户画像分析的潜力？

概念验证表明，种子元数据可有效用于基于开源情报的用户画像分析。通过分析群体参与、匿名化使用情况和内容交互，它能够侦测行为模式、用户群组以及潜在的高风险信号。

该处理流程通过查询公共追踪器并收集节点数据，将超过 60,000 个唯一 IP 地址映射到特定种子，从而实现用户与内容关联、行为网络构建，以及对重复或异常活动的侦测。基于图的分析识别出共同下载模式和主题群组，并将内容类型与用户群组关联起来。尽管存在匿名化和种群可见性不完整等局限，研究结果证实，经过处理并结合上下文分析后，种子元数据作为开源情报来源具有巨大潜力。本研究还探讨了三个子问题：界定种子元数据及其可检索性（sRQ1）；通过网络映射考察用户与内容之间的结构关系（sRQ2）；评估这些数据在现实世界开源情报应用中的适用性（sRQ3）。下文将逐一系统探讨这些问题，以展示这一方法的可行性和实际应用价值。

sRQ1： 有哪些类型的Data可用于检测高风险行为，且能够获取和使用？

可公开访问的种子相关数据，包括内容元数据、群组 IP 地址和 IP 层级属性，可以被检索并整理，用于行为画像分析；尤其是在结合跨来源信号后，更是如此。

我们从公共索引网站、追踪器和 IP 情报服务中收集了种子元数据，包括种子标题、标签、文件结构、群集节点 IP、地理定位、组织数据和匿名化指标。对于 CEM 和武器手册等已标记内容，反复出现的参与模式表明，某些用户行为仍可通过群集活动追踪。这些发现证实了种子数据内容丰富且易于检索。查询公共 UDP 跟踪器，可以实时将种子映射到活跃的节点 IP，从而建立内容—用户关联。然而，要揭示更深层的行为关系，就需要超越孤立的活动，转向网络层面分析，由此引出 sRQ2。

sRQ2: 我们的概念验证能够多有效地利用公开数据绘制种子用户网络?

概念验证有效地绘制了用户网络，揭示了群集内基于共同下载模式和共同内容兴趣形成的结构关系、行为簇和核心参与者。

我们利用单部图投影和共同下载逻辑，构建了能够捕捉行为关联的用户网络和内容网络。在群集中频繁共同出现的 IP 地址往往会聚集在一起，这揭示了潜在的协作、共同兴趣或行为惯例。一些节点同时连接了高风险下载内容（如武器手册）和无害下载内容（如军事题材电视节目），表明两者在主题上存在重叠，值得进一步进行画像分析。中心性分析发现，反复出现且广泛参与群体的 IP地址可能是枢纽或惯犯。这种方法展示了如何通过网络投影分析种子元数据，从中发现行为模式。

这些发现表明，基于网络的方法可用于开源情报场景中的画像分析和升级检测。我们展示并讨论了种子数据如何揭示用户和内容之间的隐藏结构。这些数据不仅可用于个人画像分析，还可用于揭示群体层面动态。要全面评估这种方法在现实开源情报工作中的可行性，我们还必须评估数据的整体适用性，这将在 sRQ3 中讨论。

sRQ3： 就可用性、结构、可靠性和处理工作量而言，种子和 IP 元数据作为开源情报来源的适用性和效率如何?

种子和 IP 元数据通常适合用于开源情报，且结构一致；不过，必须设法减轻匿名化、地理定位不准确以及数据完整性不足等问题。

种子和 IP 元数据通常易于获取，且在结构上适合用于开源情报分析，但仍面临数据噪声、完整性不一致以及 IP 地理定位受限等挑战。约 20% 的 IP 地址使用虚拟专用网络或代理服务器，使归因变得复杂；然而，匿名性检测本身也是一种有价值的开源情报信号。在与 CEM 分发相关联的 IP 地址中，匿名化比例超过 75%，呈现出明显的行为差异，可用于威胁优先级排序。匿名化的使用，尤其是将其与敏感内容关联起来时，能够揭示用户意图和技术成熟度，帮助调查人员聚焦于风险更高的用户画像。尽管存在依赖跟踪器和受应用程序编程接口限制等技术局限，种子和 IP 元数据仍可通过自动化方式获取，其一致性足以支持图分析，并能提供可靠的地理定位、组织和隐私信号。这些优势支持将此类数据整合到可扩展的开源情报流水线中，用于调查工作。

8. 局限性与未来工作

尽管取得了令人鼓舞的结果，当前研究仍受到若干局限的制约：

1. **手动且不可扩展的数据收集**：抓取与分析处理流程需要手动设置和干预，限制了其实时或长期监控的可扩展性。2. **种群可见性不完整**：仅依赖 UDP 跟踪器的响应，会遗漏仅通过分布式哈希表参与的对等节点，可能导致低估种子群组成。3.**匿名化造成的混淆**：观测到的 IP 地址中约有 20% 使用了隐私保护技术；由于元数据来源和检测逻辑的局限，一些匿名化的 IP 地址可能未被侦测到。4. **主题范围有限**：本案例研究有意将研究范围限制在单个上传者和一个关键词类别内。此方法上的选择旨在最大限度降低研究项目所有成员及相关组件意外接触非法内容的风险。

5. **关键词有限**：扩大关键词抽样范围可更可靠地实现行为泛化。6. **时间局限**：种子数据具有时效性。若不持续抓取，就无法准确追踪纵向行为或升级情况。

除了这些技术限制之外，隐私影响也是本研究的一项关键局限。尽管参与种子群会暴露 IP 地址，但这些地址并未与个人身份信息直接关联。大多数 IP 地址，尤其是动态 IP 地址，只能识别网络端点，而无法识别用户。若要将 IP 地址关联到某个个人，则需要访问互联网服务提供商用户记录，这超出了本项基于开源情报的研究范围。

不过，需要指出的是，本研究仅旨在提供概念验证，证明在犯罪相关种子共享数据的情境下，进行行为画像和部分去匿名化是可行的。通过整合额外的开源情报数据，甚至可能对使用虚拟专用网络的用户进行去匿名化，尽管这一过程仍然更加复杂。本研究并不旨在确切识别个人，而是侧重于分析行为模式，并检测可在网络层面观察到的高风险活动信号。

9. 未来工作

如前所述并已讨论，本研究旨在提出一种概念验证方法，用于分析并揭示种子数据共享情境下的非法行为。然而，为了增强这种方法的能力和相关性，未来研究应考虑

1.**自动化与扩展**：当前实现依赖手册式的基于批次的抓取和后处理，未来的工作可纳入从种子跟踪器进行实时数据收集、自动化 IP 信息丰富化以及实时网络图生成。这样一来，画像分析流程便能够扩展规模，并动态适应新兴威胁。

2. **DHT 集成**：可通过分布式哈希表（DHT）查询来补充对 UDP 跟踪器的抓取，以提升对等节点可见性，尤其是针对无跟踪器种子。基于 DHT 的发现机制能够提供更全面的群覆盖范围，并减少对可能受到限速的跟踪器的依赖。

3. **跨源开源情报融合**：未来的实现可以整合更丰富的 IP 级数据（例如历史活动、网络特征、设备指纹）或外部标记来源，以提高风险评估的准确性。下载时间、频率和跨种子导航等行为特征也有助于完善用户画像。4. **优先级排序与评分机制**：引入评分系统，根据群体参与、匿名化行为和内容敏感度对用户风险进行排序，有助于处理大型数据集并支持调查优先级排序。5. **扩展案例研究设计**：针对其他类型的高风险内容开展更多、更大规模的案例研究，有助于验证该处理流程在不同领域的可推广性，并深入了解不断演变的威胁生态系统。6. **伦理框架开发**：鉴于用户画像分析的敏感性，应开发隐私保护框架，以确保在调查场景中对开源数据进行合乎伦理的处理。

我们提出的开源情报平台仅是更复杂平台的一个构建模块。未来，我们计划从暗网抓取信息，以搜索与非法内容相关的特定关键词。具体而言，我们知道海盗湾是一个可免费搜索的在线索引，收录不同类型的 Data（视频、音频、书籍等），且不包含露骨材料。然而，Btdigg 是另一个知名的种子分享平台。暗网用户利用这一未经审查的平台分享儿童性虐待材料。不过，要拦截这一特定的关键内容，我们必须先从论坛、Telegram 频道和聊天中抓取关键词，然后将这些发现用于我们提出的平台。

缩略语

应用程序编程接口 应用程序编程接口 自治系统 自治系统 行为证据分析 行为证据分析 CEM 儿童剥削材料 逗号分隔值 逗号分隔值 分布式哈希表 分布式哈希表 自己动手 自己动手 IP地址 互联网协议 互联网服务提供商 互联网服务提供商 开源情报 开源情报 P2P 点对点 概念验证 概念验证 研究问题 子研究问题 传输控制协议 用户数据报协议 虚拟专用网络

参考文献

[1] A. Lareida, B. Stiller，大型种子测量：以国家、网络和内容为中心的 BitTorrent 视频共享分析，载于：NOMS 2018——2018 IEEE/IFIP 网络运营与管理研讨会，IEEE，2018年，第1–9页。doi:10.1109/noms.2018.8406243。[2]M. Gartner, T. Krüger, D. Hausheer，利用 SCION 互联网架构加速 BitTorrent 文件传输（2023年）。doi:10.48550/ARXIV.2301.13499。[3]J. Pastor-Galindo, P. Nespoli, F. G. Mármol, G. M. Pérez，开源情报这座尚未开发的金矿：机遇、悬而未决的挑战与未来趋势，《IEEE Access》8（2020年）10282–10304。doi:10.1109/ACCESS.2020.2965257。[4] A. Yadav, A. Kumar, V. Singh，开源情报：网络安全领域的现状、应用及未来展望综述，《人工智能评论》56（11）（2023年）12407–12438。doi:10.1007/s10462-023-10454-y。

[5]N. Torres, P. Olivares，通过记录链接和准标识符攻击对评分数据集中的用户进行去匿名化，《Data》9（6）（2024年）75。doi:10.3390/data9060075。

[6]D. Van Puyvelde, F. Tabárez Rienzi，《开源情报的兴起》，《欧洲国际安全杂志》（2025）1–15，doi:10.1017/eis.2024.61。

[7]P.-D. Kovaci, N. M. Iacob，《将开源情报融入渗透测试》，《罗马尼亚网络安全期刊》6（1）（2024）59–65。doi:10.54851/v6i1y202406。

[8]M. Sidorova, A. Kuzmin，《开源情报及其在网络安全中的应用》，《俄罗斯新大学学报》系列《复杂系统：模型、分析、管理》(1) (2023年) 1–14。doi:10.18137/rnu.v9187.23.01.p.61。[9]M. Chaudhary, D. Bansal，《面向恐怖主义相关信息的开源情报提取：综述》，《WIREs数据挖掘与知识发现》12 (5) (2022年7月)。doi:10.1002/widm.1473。[10] M. Elguindy，《将数字取证方法应用于反恐开源调查》，《法律与新兴技术杂志》1 (1) (2021年) 11–64。doi:10.54873/jolets.v1i1.32。[11]M. M. Yamin, M. Ullah, H. Ullah, B. Katt, M. Hijji, K. Muhammad，《结合网络杀伤链的开源情报映射工具：面向对抗感知安全》，《数学》10 (12) (2022年) 2054。doi:10.3390/math10122054。[12] T. Katzner, E. Thomason, K. Huhmann, T. Conkling, C. Concepcion, V. Slabe, S. Poessel，《用于保护生物学的开源情报》，《保护生物学》36 (6) (2022年10月)。doi:10.1111/cobi.13988。[13]A. Kruspe, J. Niu, M. Stillman, P. Seeberger，《关于俄乌战争的开源情报（OSINT）推文数据集》，《国际ISCRAM会议论文集》(2024年11月)。doi:10.59297/377r3945。[14] Y. Pai U., K. P. Karani，《开源情报及其在下一代网络安全中的应用——文献综述》，《应用工程与管理通讯》(2021年) 1–

25。doi:10.47992/ijaeml.2581.7000.0100。[15] Y.-W. Hwang, I.-Y. Lee, H. Kim, H. Lee, D. Kim, 《OSINT的现状与安全趋势》, 《无线通信与移动计算》2022 (2022年) 1–14。doi:10.1155/2022/1290129。[16] B. Akhgar, P. Bayerl, F. Sampson, 《开源情报调查：从战略到实施》, 施普林格国际出版社，2016年。doi:10.1007/978-3-319-47671-1。[17] B. Kumar, R. Verma, S. Tiwari, 《将开源情报用作可靠网络搜索工具》, SN Comput. Sci. 2 (5) (2021年) 402。doi:10.1007/S42979-021-00777-4。[18] T. O. Browne, M. Abedin, M. J. M. Chowdhury, 《利用人工智能开展开源情报（OSINT）应用研究的系统综述》, 《国际信息安全杂志》23 (4) (2024年) 2911–2938。doi:10.1007/s10207-024-00868-2。[19] E. Pelofske, L. M. Liebrock, V. Urias, 《利用开源情报Neo4j图数据库开展网络安全威胁搜寻与漏洞分析》(2023年)。doi:10.48550/ARXIV.2301.12013。[20] T. Day, H. Gibson, S. Ramwell, 《OSINT与非OSINT数据融合》，施普林格国际出版社，2016年，页码133–152。doi:10.1007/978- 3-319-47671-1 9。_[21] T. Dokman, T. Ivanjko, 《开源情报（OSINT）：问题与趋势》，载于《INFuture2019：数字时代的知识》，INFuture2019，人文与社会科学学院，萨格勒布大学信息与传播科学系，FF press，2020年。doi:10.17234/infuture.2019.23。[22] M. Schafer, M. Fuchs, M. Strohmeier, M. Engel, M. Liechti, V. Lenders, 《Blackwidow：监测暗网中的网络安全信息》，载于《2019年第11届国际网络冲突会议（CyCon）》，IEEE，2019年。doi:10.23919/cycon.2019.8756845。

[23] S. Szymoniak, K. Foks, 《开源情报的机遇与挑战：综述》，《科学与技术研究进展》18 (3) (2024) 123–139。doi:10.12913/22998624/186036。[24] S.-C. Chien, Y.-W. Kao, S.-M. Yuan, 《面向点对点环境的通用发布/订阅框架》，载于：2009年信息与服务科学新趋势国际会议，IEEE，2009年，第1257–1262页。doi:10.1109/niss.2009.211。[25]G. Jaideep, B. P. Battula, 《点对点网络现状、面临的安全问题及应对措施综述》，《国际应用工程研究期刊》11 (1) (2016) 616。doi:10.37622/ijaer/11.1.2016.616-620。[26]R. Mokadem, A. Hameurlain, A. M. Tjoa, 《在最小化维护开销的同时实现分层分布式哈希表系统中的资源发现服务》，《国际适应性与弹性自治系统期刊》3 (2) (2012) 1–17。doi:10.4018/JARAS.2012040101。[27] A. A. Jamal, W. J. Teahan, 《用于非结构化点对点网络资源发现的 Alpha 乘数广度优先搜索技术》，《国际先进科学、工程与信息技术期刊》7 (4) (2017) 1403。doi:10.18517/ijaseit.7.4.1451。[28] M. Shojafar, J. H. Abawajy, Z. Delkhah, A. Ahmadi, Z. Pooranian, A. Abraham, 《非结构化点对点网络中的高效分布式文件搜索》，《点对点网络与应用》8 (1) (2015) 120–136。doi:10.1007/S12083-013-0236-0。[29] BitTorrent 公司, 《2023年 BitTorrent 概览》，网页（2023年12月）。URL https://medium.com/@BitTorrent/2023-of-bittorrent-in-a-nutshell-e28af6cdfdc9[30] P. A. Watters, R. Layton, R. Dazeley, 《BitTorrent 上有多少材料属于侵权内容？一项案例研究》，《信息安全技术报告》16 (2) (2011) 79–87。doi:10.1016/j.istr.2011.10.001。[31] A. H. Schmidt, R. S. Antunes, M. P. Barcellos, L. P. Gaspary, 《通过监控 BitTorrent 网络来描述非法副本的传播》，IEEE，美国夏威夷州毛伊岛，2012年，第327–334页。doi:10.1109/NOMS.2012.6211915。

[32] W. A. Al-Khater, S. Al-Maadeed, A. A. Ahmed, A. S. Sadiq, M. K. Khan, 《网络犯罪检测技术综合综述》，IEEE Access 8 (2020) 137293–137311。doi:10.1109/access.2020.3011259。[33] N. A. Mutawa, J. Bryce, V. N. Franqueira, A. Marrington, 《应用于数字取证的行为证据分析：基于点对点网络儿童色情案件的实证分析》，载于：2015年第十届可用性、可靠性与安全国际会议，IEEE，2015年，第293–302页。doi:10.1109/ares.2015.49。[34] S. K. Gautam, H. Khajuria, R. R. Gupta, B. P. Nayak, 《印度网络空间儿童性虐待材料传播的最新趋势》，《国际网络战与恐怖主义期刊》12 (1) (2022) 1–15。doi:10.4018/ijcwt.297857。[35] H.-E. Lee, T. Ermakova, V. Ververis, B. Fabian, 《儿童性虐待材料侦测：综合调查》，《国际法医学：数字调查》34 (2020) 301022。doi:10.1016/j.fsidi.2020.301022。[36] T. Riebe, 《网络安全开源情报的隐私顾虑与接受因素：一项代表性调查》，Springer Fachmedien Wiesbaden，2023年，第221–248页。doi:10.1007/978-3-658-41667-6_14。[37] A. C. Kigerl, 《侵权国家：预测各国的软件盗版率、BitTorrent 跟踪器托管情况及点对点文件共享客户端下载量》，《国际网络犯罪学期刊》7 (1) (2013)。[38] G. Bissias, B. Levine, M. Liberatore, B. Lynn, J. Moore, H. Wallach, J. Wolak, 《五种点对点网络上的接触型犯罪者特征及儿童剥削材料贩运》，《儿童虐待与忽视》52 (2016) 185–199。doi:10.1016/j.chiabu.2015.10.022。[39] L. Yu, M. Chen, 《多粒度尺度下 BitTorrent 对等节点的地理分布》，载于：2010年国际计算机与通信技术会议（ICCCT），IEEE，2010年，第93–98页。doi:10.1109/iccct.2010.5640419。[40] D. Li, Y. Zhang, S. Jia, D. Liu, Y. Jin, Y. Wu, 《一种用于提升网络流量稳定性的 BitTorrent 流量优化方法》，《信息》10 (12) (2019) 361。doi:10.3390/info10120361。

[41] Z. Balkibayeva，《用于证据目的的元数据提取与分析方法》，乌兹别克斯坦法律与数字政策期刊 2 (5) (2024年) 31–44。doi:10.59022/ujldp.233。[42]J. A. D. Barreto、H. Dos Santos、V. P. Goncalves、D. L. Guidoni、G. P. Rocha Filho、R. I. Meneguette，《But2：一种通过被动搜索追踪和识别 BitTorrent 网络上共享的非法文件的新工具》，载于：2024 IEEE拉丁美洲通信会议（LATINCOM），IEEE，2024年，第1–6页。doi:10.1109/latincom62985.2024.10770622。[43]O. Lajam、S. Mohammed，《利用网络编码优化点对点内容分发效率：原理、挑战与未来方向》，《网络与计算机应用期刊》223 (2024年) 103825。doi:10.1016/j.jnca.2024.103825。[44] I.-S. Hwang、A. Rianto、R. Kharga、M. S. Ab-Rahman，《基于软件定义网络的本地感知型NG-PON2上的全球点对点BitTorrent实时流量》，IEEE Access 10 (2022年) 76884–76894。doi:10.1109/access.2022.3192439。[45]A. Nicolini、C. Lorenzetti、A. Maguitman、C. Chesñevar，《点对点网络中的智能查询处理：语义问题与路由算法》，《计算机科学与信息系统》16 (2) (2019年) 409–442。doi:10.2298/csis180411002n。[46] R.-G. Simion、M.-L. Pura，《一个BitTorrent分布式哈希表爬虫》，载于：2019年第13届IEEE应用计算智能与信息学国际研讨会（SACI），IEEE，2019年。doi:10.1109/saci46893.2019.9111636。[47] A. Lareida、R. Pernischova、B. B. Rodrigues、B. Stiller，《将 .torrent 内容消费抽象为二模图，并将其投影到内容网络（connet）》，载于：2017年IFIP/IEEE综合网络与服务管理研讨会（IM），IEEE，2017年，第151–159页。doi:10.23919/inm.2017.7987275。

[49]J. Han、D. Choi、T. Chung、C.-N. Chuah、H.-c. Kim、T. T. Kwon，《根据内容间关系预测内容消费》，《网络与计算机应用期刊》132 (2019年) 1–9。doi:10.1016/j.jnca.2019.01.030。[50]H. Sainani、J. M. Namayanja、G. Sharma、V. Misal、V. P. Janeja，《通过地理上下文特征增强进行IP信誉评分》，ACM管理信息系统汇刊 11 (4) (2020年) 1–29。doi:10.1145/3419373。[51] S. Bagui、X. Fang、E. Kalaimannan、S. C. Bagui、J. Sheehan，《使用时间相关特征对虚拟专用网络流量流进行分类的机器学习算法比较》，《网络安全技术期刊》1 (2) (2017年) 108–126。doi:10.1080/23742917.2017.1321891。[52]M. Lotfollahi、M. Jafari Siavoshani、R. Shirali Hossein Zade、M. Saberian，《Deep Packet：一种使用深度学习进行加密流量分类的新方法》，《软计算》24 (3) (2019年) 1999–2012。doi:10.1007/s00500-019-04030-2。[53]R. Mateless、H. Zlatokrilov、L. Orevi、M. Segal、R. Moskovitch，《Ipvest：利用机器学习对隐藏在单个IP地址后的网络实体IP流量进行聚类》，IEEE网络与服务管理汇刊 18 (3) (2021年) 3647–3661。doi:10.1109/tnsm.2021.3062488。[54] Y. Ren、H. Li、R. Li、H. Zhu、L. Sun，《理解互联网中的用户级IP区块》，安全与通信网络 2022年 (2022年) 1–11。doi:10.1155/2022/1941697。[55] L. Wu、J. Harrington、C. Kuwanoe、C. C. Zou，《利用BitTorrent的力量发动分布式拒绝服务攻击》，安全与通信网络 4 (8) (2011年) 860–870。doi:10.1002/sec.215。[56] P. Schutz、S. Gal、D. Chatzopoulos、P. Hui，《为在线应用实现去中心化索引与引导》，IET区块链 1 (1) (2021年) 3–15。doi:10.1049/blc2.12001。[57] K. Jarosz、L. Opiola、L. Dutka、R. G. Slota、J. Kitowski，《提高去中心化协作数据共享系统的数据可用性与容错能力》，载于：第17届计算机会议论文集

《科学与智能系统》，FedCSIS 2022第30卷，IEEE，2022年，第563–566页。doi:10.15439/2022f183.[58] J. D. Martin，《盗版、公共访问与保存：探索公共种子索引的可持续可访问性》，《信息科学与技术协会论文集》53 (1) (2016年) 1–6。doi:10.1002/pra2.2016.14505301123.[59] C. Zhang, P. Dhungel, D. Wu, K. W. Ross，《揭开BitTorrent生态系统的面纱》，《IEEE并行与分布式系统汇刊》22 (2011) 1164–1177。doi:10.1109/TPDS.2010.123.[60] S. Kumar，《爆炸物法证检验》，《新兴技术与创新研究杂志》（2020年）。网址：https://api.semanticscholar.org/CorpusID:225871127[61] T. G. Ferreri, A. Weir，《紧急医疗服务、简易爆炸装置与恐怖主义活动》，StatPearls，2019年。网址：https://api.semanticscholar.org/CorpusID:213048233[62] E. Drummond，m2t，GitHub代码仓库（2014年）。网址：https://github.com/erindru/m2t[63] O. van der Spek，《BitTorrent的UDP Tracker协议》，网页，BitTorrent改进提案15（2017年1月）。网址：https://www.bittorrent.org/beps/bep_0015.html