灰色阴影：仔细观察灰色区域中的电子邮件

耶莲娜●伊萨琴科娃
Eurecom Sophia
Antipolis
06410 法国
jelena.isacenkova@gmail.com

戴维德●巴尔扎罗蒂
Eurecom Sophia
Antipolis
06410 法国
balzarotti @
eurecom . cn

摘要

每天，数百万用户花费大量时间浏览垃圾邮件文件夹中的邮件。由于时事通讯和自动通知占用户收件箱中 42% 的邮件，一些重要的电子邮件不可避免地会被错误分类为垃圾邮件。不幸的是，用户通常无法做出与安全相关的决定，而且工具也无法提供任何帮助来轻松区分无害的商业信息和肯定是恶意的信息。

大多数以前的研究都集中在垃圾邮件的检测上。相反，在本文中，我们研究了灰色电子邮件中经常被忽视的区域，即那些无法通过自动垃圾邮件过滤器以一种或另一种方式明确分类的邮件。特别是，我们通过将真实世界的电子邮件分组为批量电子邮件活动的集群来分析它们。我们的方法能够自动分类并将灰色电子邮件区域减少一半，误报率仅为 0.2%。

此外，我们确定了许多可用于预测活动类别的活动功能，并讨论了它们的有效性和局限性。我们的实验表明，灰色区域中的大部分电子邮件由合法的群发电子邮件组成：时事通讯、通知和营销优惠。后者似乎是一个大型电子营销行业，已发展成为一个用于发送合法群发电子邮件的复杂基础设施。据我们所知，这是第一次对此类电子邮件进行真实世界的实证研究。

关键词

灰色电子邮件、分类、僵尸网络生成的活动、商业活动、尼日利亚骗局、时事通讯、网络钓鱼、挑战响应系统

1. 介绍

如今，许多反垃圾邮件过滤器提供了针对大规模未经请求的电子邮件活动的良好保护。然而，随着垃圾邮件发送者改进了他们的技术以增加达到目标的机会，反垃圾邮件也允许免费制作此作品的全部或部分的数字或硬拷贝以供个人或课堂使用，前提是不制作或分发副本为了盈利或商业利益，副本带有本通知和第一页的完整引用。必须尊重非 ACM 拥有的本作品组件的版权。允许使用信用抽象。要以其他方式复制或重新发布，以在服务器上发布或重新分发到列表，需要事先获得特定许可和/或付费。请求权限权限@acm.org。

亚洲 CCS'14，2014 年 6 月 4-6 日，日本京都。

http://dx.doi.org/10.1145/2590296.2590344 .

解决方案在标记可疑电子邮件方面变得更加积极。

一方面，这种军备竞赛导致检测率稳步上升。另一方面，它也导致了误报的增加，只要重要消息被错误地标记为垃圾邮件，就会给用户带来严重后果。此外，在合法用户电子邮件和垃圾邮件之间存在一个难以自动分类的灰色区域。该区域通常包含最初征集的新闻通讯和商业报价，但用户不再感兴趣 [9]。更一般地说，它包括未被传统反垃圾邮件过滤器标记但用户不一定需要的邮件。 2012 年，Hotmail 估计灰色电子邮件是所有垃圾邮件投诉的 75% 的来源。 ReturnPath [28] 发布的另一份电子邮件情报报告指出，16% 的包含广告或营销信息的电子邮件通常被标记为垃圾邮件，因此， 永远不会到达用户邮箱。乍一看，很多人会认为这种“副作用”是一种优势。然而，据估计，只有三分之一的用户将此类邮件视为垃圾邮件，而三分之二的用户更愿意接收来自已知发件人的未经请求的商业电子邮件 [7]。最近的一份报告显示，尽管邮箱超载，但消费者仍然阅读了 18% 的已订阅营销电子邮件，并继续注册电子邮件优惠和邮件列表 [28]，因此时事通讯和自动通知总计达 42收件箱消息的百分比。由于这些原因，众所周知，大多数用户会定期检查他们的垃圾邮件文件夹，以确认没有重要邮件被反垃圾邮件过滤器错误分类。

不幸的是，这个过程非常耗时。反垃圾邮件解决方案在这方面不是很有帮助，并且通常不提供任何附加信息来帮助用户快速识别营销电子邮件、新闻通讯或用户可能感兴趣的“边界”案例。

更糟糕的是，当用户浏览他们的垃圾邮件以寻找看起来合法的东西时，他们需要决定哪些电子邮件是可以信任的，哪些是令人讨厌的，哪些可能构成真正的安全威胁。不幸的是，多项研究表明，大多数用户在做出此类与安全相关的决定时非常糟糕 [19]，这也是我们首先需要自动垃圾邮件过滤器的原因之一。例如，消息反滥用工作组最近在 2010 年进行的一项调查 [17] 报告说，访问过垃圾邮件的人中有 57% 承认是故意这样做的，因为

因为他们不确定可疑邮件是否是垃圾邮件。

虽然大多数现有研究都在处理如何有效、准确地区分垃圾邮件和非垃圾邮件的问题，但在本文中，我们重点关注区分这两个类别的细线。特别是，我们将研究限制在经常被忽视的灰色电子邮件领域 [31]，即那些无法通过自动垃圾邮件过滤器以一种或另一种方式明确分类的模棱两可的消息。我们假设垃圾邮件过滤器能够很好地检测大多数垃圾邮件，并且如果过滤器有“充分理由”相信邮件是未经请求的或包含恶意内容（例如，通过使用防病毒软件、黑名单，或通过匹配已知诈骗消息的签名），大多数用户没有理由仔细检查该决定。

我们通过分析挑战-响应反垃圾邮件解决方案的实际部署来开始我们的研究，以衡量这个灰色区域的范围。我们使用系统隔离的电子邮件作为灰色电子邮件类别的近似值，这些电子邮件已经排除了大部分垃圾邮件和垃圾邮件。根据我们的数据，在剔除明显的垃圾邮件和非垃圾邮件后，用户平均每天仍然手动查看五到六封邮件。平均而言，这些邮件中有 1.5% 带有附件，其中 9% 是恶意的。然而，其中一些消息也包含有趣的内容，用户平均每天阅读和白名单 1.5 条消息就是证明。我们还确信普通用户不太擅长区分垃圾邮件和非垃圾邮件。

在这些前提下，我们分析灰色区域中的消息，以提高我们对它们的理解以及使它们难以归类的原因。特别是，我们采用基于消息聚类、分类和基于图的细化的三阶段方法。提取的电子邮件功能应用于电子邮件活动的上下文而不是单个电子邮件。我们的技术能够自动对一半的灰色电子邮件进行分类，相当于所有电子邮件流量的 15%，误报率仅为 0.2%。此外，我们的结果显示了商业营销活动的许多有趣特征，这些特征构成了很大一部分灰色区域。据我们所知，这是对合法群发电子邮件的首次实证研究。

2. 背景

基于电子邮件的营销是广告和销售的常见做法，它既可用于与现有客户保持沟通，也可用于获取新客户。不幸的是，当用户的收件箱开始因各种类型的群发邮件而超载时，邮箱维护变得非常耗时。结果，引入了电子邮件过滤器以保护用户免受未经请求的消息的侵害，启动了价值数百万的反垃圾邮件保护行业和一场远未结束的战斗。事实上，尽管直接邮件营销的响应率（3.4%）高于电子邮件营销（0.12%）[8]，但电子邮件的低成本仍然使电子信息成为一种非常有吸引力的解决方案。

营销人员经常使用专业的营销工具来最大化他们的营销活动交付率。这些工具有助于清除客户列表中不存在的电子邮件、处理收件人投诉、避免陷入垃圾邮件陷阱，甚至提供详细的营销活动发送统计数据。今天开展电子邮件营销活动是一项复杂的操作，

越来越多的请求群发电子邮件落入收件人的垃圾邮件文件夹。事实上，这些文件夹通常包含无法被自动垃圾邮件过滤器明确分类的邮件。这个灰色区域占垃圾邮件投诉的 75% [9]，它包含合法和无害的群发电子邮件，以及可能导致计算机感染或个人数据被盗的恶意邮件。然而，用户似乎也无法有效地将一个类别与另一个类别区分开来 [17、19]，并且通常（70%）仅根据发件人字段和主题行做出决定。

为清楚起见，在本研究中，我们将群发电子邮件分为两类：合法邮件和垃圾邮件。第一个包括根据法律规定（例如 CAN-SPAM 法案 [1] 和电子隐私指令 [3]）发送的征求（订阅的新闻通讯和通知）或潜在征求（广告）消息。第二类包含未经请求的、恶意的或非法的促销电子邮件。

合法营销活动的分发成为一项大型业务，专业公司提供专业电子邮件营销工具作为服务，并为寻找新客户的营销人员销售分类电子邮件列表。此类列表的收集是合法的：当用户订阅某些服务并填写表格时，他们可能会选择或默认同意与第三方共享他们的信息。因此，在某些时候用户确实同意接收广告。

相关工作

许多过滤解决方案通常相互结合使用，用于检测和减少垃圾邮件。

在我们的方法中，我们专注于分析发件人的行为。帕塔克等人。 [20] 建议分析垃圾邮件发送者的发送行为，而 Ramachandran 等人。 [27, 26] 使用行为黑名单根据发件人的行为对发件人 IP 地址进行分类。Ramachandran 观察到垃圾邮件发送者表现出可识别的发送模式，可以根据这些模式构建行为指纹。钱等。 [25] 建议依赖 IP 集群的声誉，例如BGP集群，结合DNS信息，提高公网IP黑名单精度50%。郝等。 [10] 构建了一个名为 SNARE 的自动信誉引擎，旨在根据一些非内容电子邮件特征将合法发件人与垃圾邮件发送者区分开来。韦斯特等。 [30] 建立了一个信誉模型来预测垃圾邮件发送者的行为。该模型依赖于在部分知识情况下特别有用的空间和时间特征。 该模型设法对黑名单未识别的多达 50% 的垃圾邮件进行分类。这种网络级检测技术的优势在于，与典型的黑名单服务相比，它们对垃圾邮件活动的反应往往更快。

最接近我们研究的研究是由钱等人进行的。 [24]，其中作者提出了一种基于内容的无监督电子邮件活动聚类算法，并且还认识到在真实世界数据集中对合法活动进行分类的问题。特别是，他们试图通过使用某些关键字和每个活动的 IP 阈值来过滤掉合法的群发电子邮件。虽然后者在处理僵尸网络发送的活动时非常有效，但它对从网络邮件帐户发送的其他恶意活动无效。我们在研究中证明，此类活动往往会模仿合法活动的特征，并且难以仅根据发件人特征进行识别。

据我们所知，目前还没有对合法的群发电子邮件进行过已知的研究，而且只有少数人研究过灰色电子邮件现象［32，31，6］。 Yih等人。［31］认为即使使用最佳垃圾邮件过滤器过滤灰色电子邮件也是一项非常困难的任务。因此，作者提出单独对待灰色邮件，依靠用户反馈来标记邮件。他们在与我们在研究中使用的数据集类似的数据集上进行的实验表明，与按电子邮件处理相比，按活动对电子邮件进行分类可产生更高的精度和数据覆盖率。但是，电子邮件或活动类别可能取决于用户［6、7］。因此，Chang 等人。［6］研究了如何将用户反馈与用户偏好相结合以改进分类结果。尤恩等人。［32］提出了一种基于本体的技术，以根据用户行为提供个性化的灰色电子邮件过滤。尽管我们同意灰色电子邮件的个性化至关重要， 我们的结果还表明，用户反馈对于类别预测可能不可靠。由于垃圾邮件主要通过群发电子邮件发送，许多研究试图通过分析群发电子邮件活动来识别垃圾邮件（[16、21、29、13]）。卡尼奇等人。［13］通过渗透僵尸网络研究垃圾邮件活动，并从营销角度评估其转化率。 Li 等人首先提出了通过 URL 及其重定向对垃圾邮件进行聚类。[16]。帕塔克等人。［21］尝试使用 URL 对垃圾邮件活动进行聚类，但由于 URL 混淆，这被证明是一项具有挑战性的任务。托马斯等。［29］证实了这个问题并提出了一种实时过滤 URL 的新技术。最后，钱等人。［24］根据内容相似性和 Pitsillidis 等人确定了电子邮件活动。［23］提议从消息中提取的正则表达式中自动提取垃圾邮件活动模板。

正如我们在开头提到的，该领域以前的大部分工作都集中在识别垃圾邮件及其活动上。在本文中，我们排除了大部分垃圾邮件和合法邮件，而是关注它们之间的边界区域。

3. 方法

本节介绍我们在实验中使用的数据集以及我们用来处理和分析电子邮件的技术。由于不可能单独对每封电子邮件进行分类，我们采用了一种多层方法将它们分组到类似的活动中（该解决方案已被之前的几项研究证明是有效的［31、24、21］）。特别是，我们首先根据电子邮件标头对它们进行聚类。然后，我们根据一些活动属性提取一组特征，并使用它们来训练分类器以预测活动类别。最后，我们采用基于图的细化技术来进一步提高分类的覆盖率和精度。

3.1 数据采集

可用数据的数量和多样性对于成功识别电子邮件活动至关重要。消息应该从多个提要中收集，涵盖大量收件人、多个组织，并且持续很长时间［21，22］。我们的电子邮件数据集满足这些要求，因为它是从部署在数十个不同组织中的商业挑战-响应（CR）垃圾邮件系统收集而来的。 CR 过滤器是一种软件，可以自动回复任何以前未知的传入电子邮件发件人的挑战（在我们的例子中是 CAPTCHA）。如果发件人解决了 challenge，消息被传递给收件人，发件人被添加到白名单；如果没有，它将保留在隔离文件夹中，其收件人可以在其中手动查看并将其列入白名单/黑名单。由于在我们的研究中我们希望关注包含无法轻易归类为合法或垃圾邮件的电子邮件的边界区域，因此我们在 CR 系统中安装了一个传感器来拦截任何隔离邮件。这些电子邮件已成功通过许多传统的反垃圾邮件过滤器，包括病毒扫描程序、反向 DNS 和 DNS 黑名单验证。此外，用户之前从未与发件人进行过任何对话。因此，我们可以认为这个数据集是从明显的合法邮件和垃圾邮件中预先过滤出来的。

有时这个集合被称为灰色区域［6］，它存储不确定类别的电子邮件。此组中常见的电子邮件类别包括传统垃圾邮件和诈骗邮件、自动通知、时事通讯和商业优惠。由于这种多样性，用户需要不时地手动检查这些消息，以寻找任何有趣或丢失的电子邮件。

我们还对 CR 系统进行了检测以收集其他信息（见表 1）：用户打开的电子邮件和白名单消息（因此表明用户手动将它们分类为合法）。这提供了有关用户区分无害消息和有害消息的能力的见解。最后，我们的传感器收集了交付状态信息，例如对于 CR 系统发回的每封挑战电子邮件，发送、退回和交付。

在我们的实验中，我们依赖于从不同规模的公司收集的统计电子邮件数据。监测期为 6 个月，从 2011 年 8 月到 2012 年 1 月。在此期间，大约有 1100 万封邮件被传送到受监测的邮件服务器（表 1）。其中29.4%属于灰色信息类。为了保护参与研究的用户和公司的隐私，我们在实验中使用的数据不包括电子邮件正文，并且对标题进行了清理并以汇总形式进行了分析。

3.2 电子邮件集群

之前的几项研究已经涵盖了将电子邮件分组到活动中的任务（[15、23、16、24、21]）。以前的结果在识别电子邮件活动方面非常成功，但不幸的是，通常依赖于电子邮件正文的内容。我们的数据集仅限于电子邮件标题，因此迫使我们使用仅基于电子邮件主题的不同方法。这种技术的主要限制是电子邮件主题必须足够长，以尽量减少巧合匹配不同消息的机会。

将相似主题分组的明显解决方案是应用一些文本挖掘算法，但我们的输入文本很短，保持词序很重要。因此，我们决定使用一种基于“几乎精确”文本匹配的简单方法，扩展到包括具有可变部分的主题。后者可以是主题中的不同短语，包括随机词、标识符或用户名。我们使用长度递减（在 70 到 8 之间）的单词 n-gram，并使用允许跳过主题的不同部分的滑动窗口。我们的实现基于现有的 n-gram 提取库（Ngram Statistics Package [4]）、标准的停用词列表和一些自定义脚本来匹配提取的 n-gram 并将它们分配给集群。

表 1：一般统计数据

邮件服务器	13	白色电子邮件	2,806,415	解决的挑战	166,279
活跃用户	10,025	黑色电子邮件	5,066,141	用户将电子邮件列入白名单	42,384
消息总数	11,203,905	灰色邮件	3,331,349	用户查看了电子邮件	104,273

表 2：集群特性

A组	
发件人 IP	网络前缀的分配（/24）发件人姓名
电子邮件发件人姓名的分布	
Sender add.domain	邮件域名分配 Sender
add.prefix	电子邮件前缀的分布
B组	
拒绝	MTA 白色电子邮件中拒绝电子邮件的百分比
被列入白名单的电子邮件的百分比	
被退回的挑战	CAPTCHA 解决的退回挑战百分比
已解决挑战的百分比取消订阅标头	取消订阅标头的百分比
C组	
收件人数量-化数量	每封电子邮件的唯一收件人标准
每封电子邮件条目	
收件人的标题	地点 的 收件人的 电子邮件：
收件人/抄送/密件抄送/混合	
国家	基于原始IP的国家分布

该过程首先搜索最长的 n-gram（70），然后减少长度，直到找到足够多的相似匹配（每个集群的阈值为 30 封电子邮件）以创建集群。该算法对长主题有效，但对短主题有问题，因此将我们的分析限制在至少包含 10 个字符和 3 个单词的主题上。在这个阶段，我们成功地将 50% 的电子邮件聚集在 12,250 个集群中。簇大小在 30 到 8,468 条消息之间变化。

3.3 基于特征的分类

为了能够区分和分类已识别的集群，我们提取了一组分为三个类别的十一个特征（见表 2）。

A组：该组中的特征反映了集群内某个特征的相似性。这些值表示为 0 到 1 之间的范围，其中 0 表示聚类中的高分布（低数据相似性），1 表示低分布（高数据相似性）。特征相似度定义为：

$$a(C)=1-u/t$$

其中 u 是唯一或相似特征值的数量，t 是电子邮件总数。该组包含四个特征，用于测量发件人 IP 前缀和电子邮件地址的相似性，以及发件人姓名的相似性。特别是，我们将电子邮件域地址分为两部分：电子邮件前缀和电子邮件后缀。后缀通过删除数字差异进行分组（例如，abc10.com 和 abc22.com 之间）。当找到相似的后缀时，将它们合并，直到没有相似的值为止。电子邮件前缀使用 Levenshtein 距离算法的变体进行比较，其中根据电子邮件前缀本身的长度计算阈值。

这样，相似性得分被归一化以说明
事实上，例如，两个字符的区别
字符串在某种程度上相当于较长字符串的六个字符差异。

B 组：该组的特征反映了集群中具有特定特征值的消息的百分比。该组中有五个特征：CAPTCHA 已解决、拒绝、白色电子邮件、挑战被退回和取消订阅标头。第一个衡量发件人解决的挑战的百分比。被退回的挑战是由于收件人不存在或不接受来自发件人的电子邮件而未送达的电子邮件。每当一封电子邮件被发送给多个收件人时，我们还能够计算白色电子邮件的百分比（即，已经将发件人列入白名单的收件人的百分比）和传入电子邮件拒绝的百分比（即，被拒绝的收件人的百分比）被邮件传输代理拒绝 - 通常是因为服务器上不存在相应的地址）。最后，取消订阅标头功能会评估包含取消订阅标头的电子邮件的百分比。 后者通常用于商业消息和通知，为用户提供取消订阅列表的选项。

C 组：该组中的特征以不同的方式计算。每封电子邮件的收件人估计每封电子邮件的平均收件人数量。收件人的标头功能指示电子邮件收件人地址在电子邮件标头中的位置：收件人、抄送、密件抄送或在同一活动中使用多个位置时混合。最后，国家特征反映了集群中的国家数量（基于发件人 IP 地理位置）。

人工贴标

在执行我们的分类之前，我们需要建立一个训练集。显然，我们手动标记过程的结果取决于我们在实验中采用的垃圾邮件的实际定义。根据定义，垃圾邮件是通常批量发送的未经请求的电子邮件。然而，没有可靠的方法来验证某封电子邮件是否被征求，即收件人是否订阅了它。此外，垃圾邮件的概念在某种程度上是主观的，对所有用户而言可能并不相同。一些商业活动可能是未经请求的，因此可能被视为垃圾邮件。然而，当此类电子邮件由专业营销公司根据国家法规发送时，反垃圾邮件过滤器应如何处理它们就变得不清楚了。这也是为什么它们首先被视为灰色电子邮件的主要原因。

在本文中，我们采取保守的方法，将包含可能涉及恶意、欺诈或非法在线活动的潜在非法内容的活动标记为垃圾邮件。这包括不同的“商业模式”：非法产品销售商、恶意软件传播电子邮件、个人数据和凭据窃贼，或高级费用欺诈专家。最后，我们认为任何属于商业营销活动的电子邮件都是合法的（在一般反垃圾邮件

过滤器不应阻止它们，除非用户特别指示它们这样做）。

尽管即使有完整的电子邮件内容，电子邮件标签也可能很困难，但可以通过使用聚合的活动功能丰富电子邮件来促进这一点。所有活动功能都以聚合形式存储和查看，因此永远不会提供对任何不同电子邮件信息的访问。特定情况由电子邮件主题表示，这是一种文本信息，如果没有文本数据将很难聚合。当我们根据主题相似性对电子邮件进行分组时，我们还会保留活动主题的汇总副本。

在抽样期间，我们依赖于分析师的领域知识和附加信息（例如，每封电子邮件的平均收件人数量和发件国数量），这些信息对于一次只阅读一条消息的用户来说是无法获得的.通常一个主题足以做出标签决定，但如果不是，则分析人员会使用汇总的标题信息。例如，如果邮件主题类似于私人通信，但电子邮件已以 50 份相同的副本发送给不同的收件人，则这更有可能是骗局，而不是真正的私人邮件。同样，一条在线推广新产品或服务的消息从 30 多个不同的国家/地区发送了数千份，并且每封电子邮件有多个收件人，这可能是非法活动。

为了构建训练集，我们随机选择了 2,000 个活动并对其进行了手动标记。我们将 1,581 个（79%）标记为合法活动，将 419 个（21%）标记为垃圾邮件活动。此初步分类确认大部分垃圾邮件已从灰色数据集中过滤掉。

分类

使用上面介绍的十一个特征，我们训练了一个二元分类器。为了选择分类器，我们参考了 Kiran 等人提出的结果。 [14]，作者在其中证明，在垃圾邮件数据集上，集成分类器的性能优于单个分类器。基于这个结论，对于我们的分类任务，我们决定使用有监督的随机森林集成分类器。

我们首先进行了交叉验证测试，其中我们将采样数据随机分成两组，分别包括 70% 和 30% 的数据。然后，我们在第一组上训练了随机森林分类器（配置有 500 棵树和每个拆分三个随机变量），并在第二组上测试了提取的模型。对于每个集群，该算法返回的分数介于 - 1（对于垃圾邮件）和 1（对于合法邮件）之间。接近零的分数表示分类器对样本不确定。由于我们的集合包含大小非常不同的类别，因此我们使用马修斯相关系数 (MMC) 来衡量分类质量。我们的模型实现了 0.75 的 MCC，其中值介于 [-1,..1] 之间，1 表示完美预测。该模型产生了 0.9% 的误报（即合法活动被错误分类为垃圾邮件）和 10% 的误报（即垃圾邮件被错误分类为合法活动）。 这些比率表明我们确定的属性集可有效区分两种类型的活动。我们还注意到，虽然我们的分类器可以很好地识别合法活动，但它更有可能错误分类垃圾邮件活动。第 5 节描述了对这种现象的进一步解释。

最后，我们应用从训练集中提取的模型来预测剩余未标记活动的分类。结果如表 3 所示。

表 3：活动分类结果

活动类型	人工取样	%	未标记	%
合法的	1,581	79%	8,398	81.9%
垃圾邮件	419	21%	1,852	18.1%
全部的	2,000		10,250	

表 4：每个活动类别的属性值

属性	合法的	垃圾邮件 最小值/平均值/最大值	灰色的
国家	1 - 1.2 - 6	7 - 29 - 123	1 - 5 - 80
IP	0.13 - 0.9 - 1	0 - 0.06 - 0.82	0 - 0.7 - 1
发件人邮箱领域	0.2 - 0.98 - 1	0 - 0.3 - 1	0 - 0.85 - 1
发件人邮箱字首	0.03 - 0.98 - 1	0 - 0.09 - 1	0 - 0.81 - 1
发件人	0 - 0.98 - 1	0 - 0.3 - 1	0 - 0.8 - 1
退订标头	0 - 0.5 - 1	0 - 0 - 0.3	0 - 0.3 - 1
反弹	0 - 0 - 1	0 - 0.1 - 1	0 - 0.1 - 0.9
验证码	0 - 0 - 1	0 - 0 - 1	0 - 0.1 - 1
白色电子邮件	0	0	0.001
拒绝	0 - 0 - 0.4	0 - 0.23 - 1	0 - 0.1 - 0.7
Rec.per 电子邮件	1 - 1 - 1.1	1 - 3 - 16	1 - 1.1 - 8
接受者标头	0.76 - 0.04 - 0.2	*收件人*、*密件抄送*、*混合股份* 0.3 - 0.1 - 0.6	0.4 - 0.33 - 0.3

3.4 基于图的细化

尽管我们使用分类器实现了相对较高的准确性，但我们仍然发现对于某些活动，我们的算法给出了不确定的结果。幸运的是，绝大多数活动都位于分类器分数的极端，接近 1（合法）或 -1（垃圾邮件）。在 [-0.8..0.8] 之间的范围内，活动变得更加稀缺。灰色区域内的灰色区域代表我们的技术无法自动分配明确类别的案例。

使用这两个阈值，我们可以细化我们的分类并将数据分为三类：合法（占活动总数的 77%）、垃圾邮件（16%）和灰色邮件 (6.4%)。三个类别中每个属性的最小值、平均值和最大值总结在表 4 中。由于大多数误报和漏报位于灰色区域，我们专注于通过使用图形来改进这些消息的分类基于技术。

特别是，我们构建了一个图表，其中节点代表活动，边对两个活动共享发件人 IP 地址和电子邮件域名的组合这一事实进行建模。这些链接创建了从同一邮件基础设施发送的活动网络。为避免在使用网络邮件提供商（欺骗与否）时活动之间可能出现的虚假连接，我们从图中删除了这些链接。

生成的图表包含 9,891 个关联的活动和 608 个孤立的子图。通过目视查看子图，我们注意到大多数由一个主要类别（垃圾邮件或合法节点）组成，有时与灰色节点混合（参见图 1 中的示例）。这似乎表明灰色活动也与同一组中的其他节点属于同一类，因为它们是使用相同的基础设施发送的。

此外，我们的图表包含一个巨大的组件——一个将所有活动的 52% 链接在一起的图表——无法确定它属于哪个类别。那里-

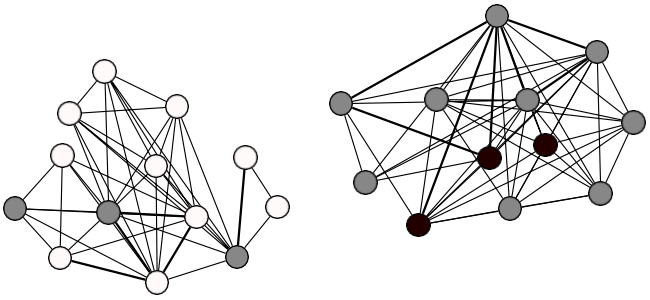


图 1：具有混合活动类别的子图：白色代表合法，灰色代表灰色，黑色代表垃圾邮件

表 5：使用图形分析细化活动分类。对 2,000 个抽样活动评估的分类错误

	随机森林	图分析
误报	0.9%	0.2%
假阴性	8.6%	7.6%
灰色地带	6.4%	2.9%

因此，我们应用社区发现算法 [5]，将所有节点分组为相互连接的社区，也称为组，将巨型组件分解为更小的部分。我们最终得到 660 个组，对于其中的大部分我们可以准确地将一个类关联起来。当灰色活动与任何其他类别在同一组中时，我们将灰色活动分配给其组的类别。

虽然这种技术适用于大多数组，但由于存在松散连接的节点，结果中仍会引入一些噪声。这些节点由于电子邮件重复使用合法活动的主题而错误地连接到一个组。为了删除这些连接，我们为每个节点计算一个称为聚类系数的图形度量。松散连接节点的系数等于 0，而紧密连接节点接近 1。因此，我们对所有聚类系数大于零且属于一组合法或垃圾邮件活动的灰色节点进行重新分类。为了确定组的类别，我们计算组中所有节点的分类器得分的平均值：高于 0.2 的组被认为是合法的，低于该阈值的组被认为是垃圾邮件。

使用这种方法，我们能够对超过一半的灰色活动（427）进行重新分类。这将误报率从 0.9% 降低到 0.2%（有关更多信息，请参见表 5）。整个数据集现在分为合法（80%）、垃圾邮件（17%）和灰色（2.9%）消息（合法活动增加 3%，垃圾邮件增加 1%）。同样，我们的方法对合法消息执行得更好。这是由于合法活动形成比恶意活动更强大的网络（随着时间的推移重复使用相同的邮件基础设施）。

4. 属性分析

在本节中，我们分析垃圾邮件和合法活动的特征，并将我们的发现与之前垃圾邮件研究 [21、24] 中提出的结果进行比较。

随机森林分类器提供了一些关于每个特征相关性的信息。有趣的是，最不重要的属性是 B 组中的属性，特别是集群中已列入白名单的电子邮件的百分比。最重要的是国家和 IP 地址的分布，其次是平均收件人数量，以及发件人电子邮件地址的相似性。后者被证明是有用的，因为垃圾邮件发送者经常更改发件人电子邮件，而合法的活动使用一个或多个重复模式。

特别是，我们发现发起国家的数量是最具指示性的参数，而之前的研究通常依赖于 IP 地址分布（例如 [24]）。

4.1 IP 和地理位置的作用

IP 地址变化通常被视为僵尸网络活动的有力指标，并经常用作检测垃圾邮件的可靠指标。然而，尚不清楚应该采用什么作为该指标的阈值，有多少不同的 IP 应该提醒我们注意分布式恶意活动，或者我们仅通过查看其 IP 地址分布就可以对电子邮件活动进行多准确的分类。

在之前的垃圾邮件活动研究中，Qian 等人。[24] 使用每个活动 10 个 IP 的阈值来区分垃圾邮件活动和合法活动。为了评估这个阈值，我们将它应用于我们的灰色数据集，如图 2（a）所示。该图绘制了垃圾邮件和合法活动的唯一 IP 前缀的分布。大约 90% 的合法活动确实低于 10 IP 阈值，而 90% 的垃圾邮件高于 - 导致全局错误率为 9.2%（准确地说，我们的测量是基于 /24 子网而不是单个IP 地址，因此实际错误率远高于 9.2%）。相比之下，这个错误比我们的分类器高 5 倍。

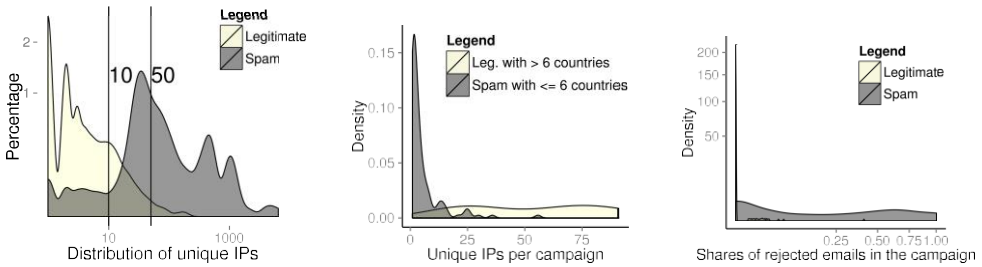
通过查看图 2（a），我们注意到超过 50 个 IP 前缀几乎没有剩余的合法活动，并且 99.8% 的合法活动低于此阈值。然而，一半的垃圾邮件活动位于阈值以上，另一半位于两个阈值（10-50）之间。这表明没有一个值可以以可接受的错误率将两个类分开。

当我们查看 IP 国家/地区分布时，结果会大大改善，因为一些合法的活动有很多 IP 前缀，但来自少数国家/地区。这可以用不同地点的几家营销公司分发的一个商业活动来解释。相比之下，绝大多数垃圾邮件活动都来自多个 IP 前缀和多个国家。事实上，通过使用六国阈值（由我们的分类器选择的阈值），我们仅错误分类了 0.4% 的合法活动和 12% 的垃圾邮件活动 - 导致总错误率为 2.8%。图 2（b）显示了分类错误。

最后，我们更仔细地调查了这组来源不多的垃圾邮件活动。有趣的是，它们中的大多数的分类器给出的分数很低，介于 0 和 -0.5 之间。图表细化对他们无效，因为这些活动根本没有出现在我们的图表中。在仔细的人工检查中，这些案例主要对应于网络钓鱼和尼日利亚诈骗。其中一些活动是使用网络邮件帐户以低容量和短时间发送的，因此隐藏在良性 IP 地址下。

4.2 面向接收者的属性

可以在三个不同的标题中指定电子邮件收件人：To、Cc 和 Bcc。有趣的是，我们没有发现使用抄送标头的活动，而且一些活动似乎会随时间随机更改收件人的位置



(a) 唯一 IP 前缀分布的对数图 (b) 活动和 6个国家的门槛 (c) 每个活动的电子邮件拒绝

图 2：活动中的属性分布

表 6：To/Bcc/Mixed 收件人报头分布

	到	*Bcc*	*混合的*
合法的	75%	5%	20%
垃圾邮件	30%	12%	58%
灰色的	20%	53%	27%

（我们将它们归类为 Mixed ）。我们还查看了每封传入电子邮件的收件人数量以及多个收件人电子邮件中不存在的电子邮件帐户数量（由于用户不存在而在 MTA-in 中被拒绝）。我们将这三个特征放在一起看，因为将它们结合起来通常比单独使用时更能提供信息。

大约 75% 的合法活动使用 To 标头（表 6），而垃圾邮件发送者通常在同一活动中混合使用不同的标头。 Bcc 标头被两种活动类型采用，尽管频率较低。然而，这在灰色活动中很常见：事实上，其中一半专门使用此标头来指定收件人。同样，这在前面提到的诈骗活动中很常见。

由于位于灰色区域的活动经常使用 Bcc 字段，因此它们的收件人列表较短，每封电子邮件平均只有 1.2 个收件人。相比之下，94% 的合法活动只有一个收件人，而垃圾邮件发送者往往每封电子邮件平均包含至少三个收件人。

但是，仅凭这些功能无法可靠地将垃圾邮件与合法邮件分开。例如，36% 的垃圾邮件活动每封电子邮件仅使用一个收件人，并且在 30% 的情况下在收件人标头中指定。有趣的是，通过将这两个标准与这些活动也具有高 IP 前缀分布的事实相结合，我们可以推断出它们源自受感染的机器或僵尸网络。

当活动中的某些消息被拒绝时，表明发件人的收件人列表未经验证或不是最新的。尽管有时用户在提供电子邮件地址时会出现拼写错误，但较高的拒绝率（如图 2（c）所示）以及多个收件人是垃圾邮件发送者活动的良好指标。事实上，在每封电子邮件有两个收件人的垃圾邮件活动中，只有 1% 的拒绝率低于 0.1。因此，这两个特征的组合对于活动分类表现良好。

4.3 时事通讯订阅标题

表 7：取消订阅活动中的标头

活动	标头存在	缺少标题
垃圾邮件	225 (10%)	2,013 (90%)
合法的	5,064 (51%)	4,948 (49%)
电子邮件		
垃圾邮件	2,710 (0.6%)	482,133 (99%)
合法的	506,352 (43%)	668,153 (57%)

我们的一项功能计算电子邮件中是否存在 List-Unsubscribe 标头。此标头专门用于指示批量电子邮件发件人以便单独处理此类电子邮件，并且通常指向可用于取消订阅邮件列表[1]的 URL 或电子邮件地址。建议常规批量发件人使用此标头。另一个针对群发电子邮件的建议是使用 Precedence: bulk header。但是，由于在我们的数据集中，此标头仅在少数消息中使用，因此我们将重点放在更常见的 List-Unsubscribe 标头上。

图 3 显示了使用取消订阅标头的每种活动类型的百分比。只有 10% 的垃圾邮件活动采用标头，仅占垃圾邮件总数的 0.6%。虽然合法的活动倾向于在他们的大部分电子邮件中使用标题，但大约一半的活动根本不使用它。这是由于几家不同的电子邮件营销公司在宣传同一活动，其中一些包含标题，而另一些则不包含。总的来说，大约一半的合法活动包含标头（表 7），所有合法活动中有 27% 的所有消息中都包含标头。

总之，我们发现垃圾邮件发送者很少使用取消订阅标头，但与此同时合法活动仅在一半的电子邮件中使用它。虽然此属性似乎是识别营销活动的一个很好的特征，但欺骗取消订阅标头非常容易，并且可以通过垃圾邮件发送者的最低额外成本来完成。

5. 电子邮件活动

在本节中，我们将介绍我们在灰色区域中识别出的四类电子邮件活动。我们已经将垃圾邮件与合法活动分开。我们进一步划分

[1] 通常，邮件正文中还包含取消订阅选项，但我们无法检查这种情况，因为我们无法访问电子邮件正文。

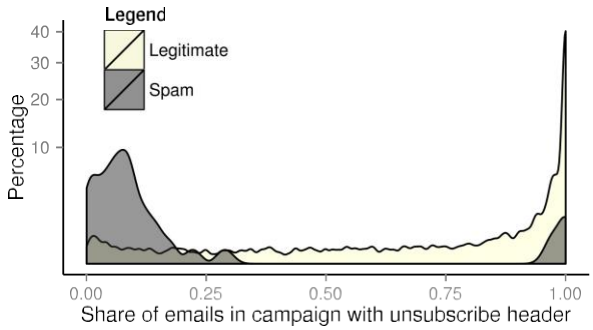


图 3：时事通讯订阅标题分布。仅绘制标题存在的情况

垃圾邮件分为两类：一类是由分布式和动态基础设施（可能由僵尸网络或受感染的机器发送）从少数 IP 发送的较小活动中生成的。

我们还将合法的活动分为两组。第一个由私人营销公司作为分发合法批量广告（即商业活动）的服务发送。第二个包括发送给订阅网络服务或邮件列表的用户的时事通讯，以及自动生成的通知（例如在线注册）。同样，第一个是由大型基础设施提供的，而第二个通常是从一组有限且恒定的 IP 地址发送的。

为了在我们的数据集中识别这四个类别，我们采用了一些简单的启发式方法。作为商业活动，我们将合法活动标记为属于第 3.4 节中描述的图中最大的互连组件。这些活动分布在许多不同的网络和域名上，因为此类活动有时由几个不同的电子营销服务提供商发送，从而形成一个相互关联的活动的大图。我们将其余分散的合法活动视为时事通讯和通知，因为它们依赖于更加静态和孤立的电子邮件传递基础设施。僵尸网络生成的活动取而代之的是来自六个以上不同国家的垃圾邮件集群和 20 多个独特的 /24 IP 前缀。最后，我们对剩余垃圾邮件活动中的 350 多个进行手动抽样，以识别诈骗和网络钓鱼活动。

所有类别都在图 4 中可视化，其特征的平均值总结在表 8 中。

5.1 商业活动

这是我们数据集中最大的类别，涵盖了 42% 的已识别活动，每个活动平均包含 148 封电子邮件。通过手动查看这些集群，我们确认这些消息主要是由专业电子邮件营销人员发送生成的。我们能够确定一些主要参与者（国内和国际），并且经常确认他们实际上经营合法业务。在他们的网站上，他们反复强调“他们不是垃圾邮件发送者”这一事实，他们只是向其他公司提供一种在当前立法范围内发送营销电子邮件的方式。事实上，他们还提供了一个在线程序，供用户选择退出并从未来的通信中删除。这些公司还使用广泛的 IP 地址范围来开展活动，可能是为了避免被列入黑名单。此外，我们发现其中一些非常有趣

表 8：每个活动类别的特征平均值。注意：用户操作仅针对具有操作的广告系列进行评估

属性	商业的	通讯	僵尸网络	骗局
国家	1.4	1.14	28.2	2.74
每封电子邮件的收件人	1.00	1.00	2.80	1.16
接受者 *到：*	0.75	0.77	0.31	0
标头（%）密件抄送：	0.07	0	0.12	0.83
混合：	0.18	0.22	0.57	0.17
发件人电子邮件前缀	0.97	0.98	0.12	0.94
发件人电子邮件域	0.96	0.99	0.31	0.97
IP分配	0.84	0.94	0.08	0.86
独特的IP	62		172	5
拒绝	00		0.24	0.02
发件人	0.97	0.98	0.34	0.95
反弹	0.01	0.02	0.09	0.14
退订标头	0.59	0.39	0.01	0
验证码	0.006	0.007	0	0.007
白色电子邮件	0.007	0.004	0.004	0.02
期间（天）	28	19	59	41
查看电子邮件	3.6	6	7.3	2.9
列入白名单的电子邮件	2.9	4	1.26	2.25
验证码已解决	19	26	1.7	7.6
活动	5,113	3,597	2,107	150

公司还提供可用于获取新客户的预编译电子邮件列表（已按用户兴趣分类）。

因此，电子邮件收件人既可以来自冷列表（即还不是客户的人），也可以来自当前客户列表。结果，不同的营销人员发送了许多不同的电子邮件活动，从而形成了一个大型的相互关联的活动网络，如图所示。由于发件人也依赖冷名单，因此确保收件人可以取消订阅未经请求的广告至关重要。事实上，商业活动的退订率最高。

平均而言，此类活动持续 26 天，但有些活动还会持续数月。不同的电子邮件营销公司通常参与发送单个活动，其中每个公司仅在特定时间范围内活跃。此外，每个营销服务提供商都有自己专用的 IP 地址范围，这就解释了为什么该组中的 IP 地址差异和活动的地理分布有时很高。作为比较，时事通讯（图 4，左上部分）使用的唯一 IP 地址平均比专业营销人员少三倍。

总而言之，商业活动可以高度分散，但与此同时，它们通常采用具有相似发件人姓名和电子邮件地址的一致电子邮件模式。

5.2 通讯活动

时事通讯发件人主要依赖静态和小型邮件基础设施。发件人通常是分发电子邮件的实际公司，通常具有较小且固定的 IP 地址范围。此类别包含前一个电子邮件的一半（可能是因为大多数合法邮件列表未进入隔离区，因为它们已被其客户列入白名单）并涵盖了周围

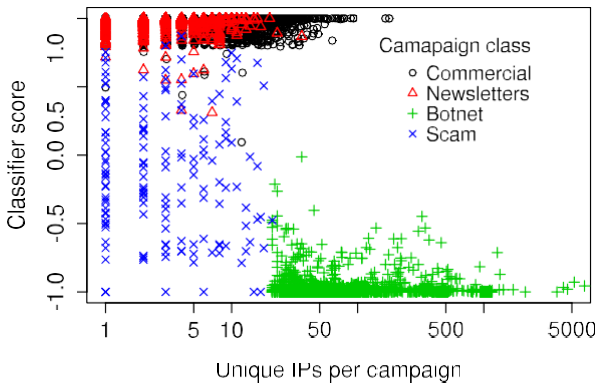


图 4：电子邮件活动类别分布

平均大小为 90 封电子邮件的活动总数的 30%。

人工检查似乎证实，这些活动主要由用户过去订阅的在线服务发送的通知和新闻通讯组成。发件人在地理上是本地化的（我们只遇到过分布式新闻通讯活动的一个例外）并且具有极其一致的发送模式。由于我们根据主题对活动进行分组，因此时事通讯往往会持续很短的时间。此外，他们通常使用有效的电子邮件收件人列表，并展示最低的 IP 地址、国家和发件人电子邮件地址变化。只有 Unsubscribe 标头的使用似乎不一致，因为只有 39% 的电子邮件使用它。然而，这可以解释为通知电子邮件通常不使用此标头 – 只有时事通讯需要订阅。 此类电子邮件标头中的一致模式表明发件人正在努力建立声誉并成功发送信件。毫不奇怪，这也是最常被用户列入白名单的类别。

5.3 僵尸网络生成的活动

不出所料，表 8 显示僵尸网络生成的活动具有高度动态的属性值，使它们成为最容易自动识别的类别。此类别包含的集群仅占所有活动的 17%（也是因为大多数垃圾邮件已被其他反垃圾邮件过滤器排除在灰色邮件之外）。僵尸网络活动具有最高的地理分布，因为它们是由来自世界各地的受感染计算机发送的：每个活动有 172 个独特的/24 个网络，平均分布在 28 个国家/地区。另一个普遍的特征是使用使用未经验证的电子邮件列表发送的多个收件人电子邮件。因此，这导致了最高的电子邮件拒绝率 (24%) 和最高的验证码请求退回率。 Unsubscribe 标头很少使用，发件人电子邮件地址相似度低。

平均而言，僵尸网络活动持续时间最长，其中一项与毒品相关的活动在我们整个六个月的实验期间发送缓慢。帕塔克等人。 [21] 还研究了垃圾邮件活动的长度，重新
在数据集跨度上移植最大长度为 99 天
宁 150 天。我们的活动比这长得多，可能是由于不同的数据集（我们直接从用户邮件服务器收集，而不是从开放中继收集）、不同的电子邮件
分组方法（相似主题与 URL），或垃圾邮件发送者行为随时间的变化。

尽管这些活动具有易于识别的特征，但用户对这些电子邮件表现出惊人的高兴趣。此类别的每个活动的电子邮件查看次数最多，这表明用户通常对在黑市上推广和销售的产品感到好奇 [18]。

5.4 诈骗和网络钓鱼活动

这些活动包含网络钓鱼和尼日利亚诈骗电子邮件。欺诈者使用威胁性信息或试图用巨额金钱收益引诱他们来欺骗他们的受害者。此类别的特征在很大程度上类似于商业活动的特征，因此很难在不分析电子邮件正文的情况下自动分离这些活动。事实上，这些活动中的大多数都属于我们分类器的灰色区域。这就是我们需要手动验证此集合的原因。这类威胁更有可能通过基于内容的检测技术来识别，例如，通过查看电子邮件地址和电话号码 [12]，或正文中包含的 URL [21、29]。

我们仅发现 12,601 封此类电子邮件，平均活动规模为 84 封电子邮件。网络钓鱼活动通常使用众所周知的公司名称（例如银行、eBay、Paypal）来欺骗电子邮件地址，而尼日利亚诈骗者主要依赖网络邮件帐户 [12]。在这种情况下，许多发件人解决了 CAPTCHA 挑战——确认这些骗局背后通常有真人。验证码被破解的 IP 地址大多位于西非国家，如尼日利亚或象牙海岸。此类别中的所有消息均不包含取消订阅标头。

不幸的是，用户似乎经常成为此类攻击的受害者，因为他们在这些活动中打开甚至将消息列入白名单。

6. 用户行为

我们的数据集还提供了有关用户对隔离电子邮件执行了哪些操作的信息。特别是，我们收集了有关已阅读、添加到用户白名单或黑名单的消息以及后来由发件人解决的验证码的信息。这些数据可以让我们了解普通用户识别可疑电子邮件的能力。

表 9 显示了三个用户操作统计信息。正如预期的那样，用户活动主要涉及合法和灰色活动。事实上，用户浏览此文件夹中的电子邮件的主要原因是发现错过的通知或未送达的良性消息。然而，也有很大一部分用户打开了垃圾邮件，可能是被一些欺骗性的主题所吸引。如表 8 和图 5 (a) 所示，最高的活动收视率是由僵尸网络生成的活动产生的，甚至超过了时事通讯。在我们为期六个月的实验中，用户查看了超过 3,888 封垃圾邮件，导致五分之一的用户至少查看过一封垃圾邮件，并且平均打开
其中 5 个。

在对僵尸网络生成的电子邮件被阅读并列入白名单的活动进行人工检查后，我们确认这些活动正在宣传非法产品，例如

[2] 不幸的是，从我们的数据集中，我们无法得知有多少用户下载了附件或点击了邮件正文中包含的链接。

表 9：用户对活动执行的操作

	看过	列入白名单	验证码解决了
合法的	42%	12%	3.5%
垃圾邮件	25%	6%	0.2%
灰色的	40%	17%	10%

毒品和盗版软件。这可能表明两件事：要么用户在区分合法电子邮件和有害电子邮件方面存在问题，要么一些用户真正对垃圾邮件发送者宣传的产品感兴趣。很难得出结论，因为这两种假设可能对不同的用户都是正确的，但很明显，他们中的大多数人并没有意识到打开恶意电子邮件所涉及的安全威胁。

同时，我们应该将报告的已查看电子邮件统计数据与实际列入白名单的电子邮件数量进行比较——这一操作可以解释为相当于点击几个网络邮件服务提供的“不是垃圾邮件”按钮。每个僵尸网络生成的活动中列入白名单的电子邮件数量（1.26 封电子邮件，表 8）是所有类别中最低的，这表明大多数用户成功区分了它们。但是，我们注意到诈骗/网络钓鱼活动在每个活动中被列入白名单的电子邮件数量与商业活动几乎相同（2.25 对 2.9）。这表明用户可能难以区分这些类别。重要的是要记住，此类别是由领域专家手动抽样的，对于典型用户而言并非如此，因为他们中的大多数人都没有接受过培训，并且更有可能落入此类欺诈行为。

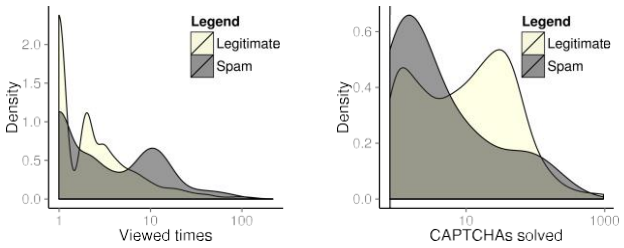
为了进一步衡量这种现象的严重程度，我们计算出某个用户将合法电子邮件列入白名单的概率为 0.36%，将垃圾邮件列入白名单的概率为 0.0005%。这些数字可能看起来很低，但当乘以用户数量和收到的消息数量时，它们会迅速增加。总的来说，每个合法活动平均有 3.9 封电子邮件被列入白名单，而每个垃圾邮件活动有 1.1 封电子邮件。

我们要回答的最后一个问题是，发件人在活动中解决了一些验证码这一事实是否可以很好地表明其合法性。不幸的是，事实并非如此，原因有二。首先，大多数合法的活动发件人都是自动化工具，因为大量灰色电子邮件由时事通讯、在线通知和营销电子邮件组成。其次，尽管总体趋势是用户在合法类别中解决了更多的验证码（图 5（b）），但如表 8 所示，与其他类别相比，诈骗和网络钓鱼活动的验证码解决率（7.6）也较高。最后，僵尸网络生成的活动解决少数验证码的罕见案例对应于垃圾邮件发送者向欺骗地址发送的挑战，如 Isacenkova 等人先前所述。 [11].

总而言之，用户对灰色电子邮件生成的操作是错误的，因此用于预测是不准确的。他们经常打开甚至具有潜在危险的电子邮件，而忽略安全风险。这些结果与 Onarlioglu 等人进行的用户研究中测试的结果一致。 [19].

7. 未分类的电子邮件

我们的活动分类涵盖了隔离区中一半的电子邮件，误报率为 0.2%。人们可能想知道在我们的聚类方法之外剩下的 50% 里面有什么。钱等。 [24] 得出结论，大多数合法电子邮件不应被分类



(a) 查看邮件 (b) CAPTCHAs solved 图 5：每个活动的用户操作数

由于人类产生的内容的独特性而进入集群。此外，由于大部分垃圾邮件和合法电子邮件已从我们的数据集中过滤掉，因此确切比例可能会有所不同。

我们可以通过假设合法活动发送者始终依赖稳定的托管基础设施（如第 3.4 节所述）来尝试近似未聚类部分的内容。在这种情况下，对于每个合法的活动，我们都可以尝试在未分类的电子邮件集中查找由相同子网和域名发送的邮件。使用这种技术，我们发现 26% 的电子邮件来自同时负责合法活动的发件人。近 40% 的邮件来自网络邮件提供商。垃圾邮件集在非集群集中的匹配数量非常少，这是意料之中的，因为大多数这些电子邮件都是从随时间变化的受感染机器发送的。

尽管这种启发式方法只能粗略估计剩余 50% 的消息中包含的内容，但它仍可用于（作为更复杂系统的一部分）自动将营销活动与更危险的垃圾邮件形式区分开来。

8. 讨论和结论

在本文中，我们提出了一个系统来识别和分类灰色电子邮件的活动。作为该集合的近似值，我们选择使用挑战响应反垃圾邮件过滤器的隔离文件夹，因为它已经清除了明显的垃圾邮件和非垃圾邮件。

我们的分析揭示了预测性最强和最弱的电子邮件活动类属性。我们还证明了以前用于电子邮件活动分类的技术 [24] 在我们的设置中没有提供可接受的结果，确认灰色区域包含最难分类的消息。此外，我们确认并扩展了先前关于僵尸网络活动的研究的一些发现 [21]。

我们的系统可以以不同的方式使用。首先，它可以帮助了解大型商业活动的运作方式、它们的来源以及它们与其他未经请求的电子邮件有何不同。它还可以作为一种输入，自动将营销活动和时事通讯放在一个单独的文件夹中，以便用户可以清楚地区分这些消息与其他形式的垃圾邮件。事实上，我们研究中的用户经常打开僵尸网络生成的电子邮件，并且在处理诈骗和网络钓鱼消息时特别容易出错；我们认为，专门用于合法群发电子邮件的单独文件夹会在用户和恶意邮件之间创建一个额外的层，从而使用户在查找丢失和错误分类的电子邮件时可以专注于群发文件夹。有趣的是，在我们完成我们的

研究表明，GMail [2] 部署了一个类似的解决方案，用于将用户新闻通讯、通知和其他商业电子邮件分为不同的类别。

我们还发现，我们基于发件人行为的分类方法适用于除诈骗之外的任何活动。我们相信后者将在很大程度上受益于基于内容的电子邮件分析，例如URL 或电子邮件/电话集群。最后，我们证明了通过使用基于图形的细化方法，通常可以仅根据发件人信息识别合法的电子邮件活动，并将其归类为新闻通讯或商业广告。在合法群发电子邮件的实证研究方向上，这是一个特别有希望的结果。

9. 致谢

导致这些结果的研究已根据赠款协议 nr 获得欧盟第七框架计划 (FP7/2007-2013) 的资助。257007。我们还感谢 MailInBlack 提供我们研究中使用的数据。

10. 参考文献

[1] CAN-SPAM 法案：控制对垃圾邮件的攻击 2003 年非请求色情和营销法。

[2] 收件箱标签和类别标签。 http://gmailblog.blogspot.fr/2013/05/a-new-inbox-that-puts-you-back-in.html。

[3] 关于隐私和电子通信的指令 2002/58，涉及电子通信领域的个人数据处理和隐私保护（隐私和电子通信指令）。 ，2002.

[4] S. Banerjee 和 T. Pedersen。 ngram 统计包的设计、实现和使用。 ITPC，2003 年。

[5] V. D. Blondel，J.-L.纪尧姆、R.兰比奥特和 E. 列斐伏尔。大型网络中社区的快速展开。统计力学杂志：理论与实验，2008 (10)：P10008，2008。

[6] M.-W.张，W.-T。 Yih 和 R. Mccann。针对灰色邮件的个性化垃圾邮件过滤。中国经济研究会，2008 年。

[7] D. 休闲。垃圾邮件：它如何伤害电子邮件和降低互联网生活，2003 年。

[8] 直销协会。响应率报告，2012 年。

[9] D. M. A. DMA。电子邮件送达率审查白皮书，2012 年。

[10] S. Hao、N. Syed、N. Feamster、A. Gray 和 S. Krasser。使用 SNARE 检测垃圾邮件发送者：时空网络级自动信誉引擎。在过程中。第 18 届 USENIX 安全研讨会会议，第 101-118 页。USENIX 协会，2009 年。

[11] J. Isacenkova 和 D. Balzarotti。挑战响应垃圾邮件过滤器的真实世界部署的测量和评估。 ACM SIGCOMM，IMC，2011 年。

[12] J. Isacenkova、O. Thonnard、A. Costin、D. Balzarotti 和 A. Francillion。诈骗丛林内部：深入了解 419 诈骗电子邮件操作。国际捕鲸委员会，2013 年。

[13] C. Kanich、C. Kreibich、K. Levchenko、B. Enright、G. M. Voelker、V. Paxson 和 S. Savage。Spamalytics：垃圾邮件营销转换的实证分析。中国化学会，2008 年。

[14] *P. Kiran 和 I. Atmosukarto。垃圾邮件还是非垃圾邮件——这就是问题所在。技术。代表，华盛顿大学，2009 年。*

[15] A. Kolcz 和 A. Chowdhury。通过上下文强化指纹识别。中国经济研究会，2007 年。

[16] F. Li 和 M. han Hsieh。垃圾邮件发送者集群行为和基于组的反垃圾邮件策略的实证研究。中国经济研究会，2006 年。

[17] MAAWG。电子邮件安全意识和使用情况报告，2012 年。

[18] D. McCoy、A. Pitsillidis、G. Jordan、N. Weaver、C. Kreibich、B. Krebs、G. M. Voelker、S. Savage 和 K. Levchenko。 Pharmaleaks：了解在线制药联盟计划的业务。USENIX，2012 年。

[19] K. Onarlioglu、U. O. Yilmaz、D. Balzarotti 和 E.基尔达。洞察用户应对网络攻击的行为。国家安全局，2012 年。

[20] A. Pathak、Y. Hu 和 Z. Mao。从独特的有利位置窥视垃圾邮件发送者的行为。在过程中。第 1 届 Usenix 大规模漏洞利用和紧急威胁研讨会，第 1-9 页。 USENIX 协会，2008 年。

[21] A. Pathak、F. Qian、Y. C. Hu、Z. M. Mao 和 S.兰詹。僵尸网络垃圾邮件活动可能会持续很长时间：证据、影响和分析。计量学，2009 年。

[22] A.皮西利迪斯，C.卡尼奇，G.M.沃尔克，K. Levchenko 和 S. Savage。品尝者的选择：垃圾邮件提要的比较分析。国际媒体中心，2012 年。

[23] A.皮西利迪斯，K.列夫琴科，C.克雷比奇，C.卡尼奇，G. M. Voelker、V. Paxson、N. Weaver 和 S. Savage。僵尸网络柔道：与垃圾邮件作斗争。国家安全局，2010 年。

[24] F. Qian、A. Pathak、Y. C. Hu、Z. M. Mao 和 Y. Xie。基于无监督学习的垃圾邮件过滤案例。计量学，2010 年。

[25] Z. Qian、Z. M. Mao、Y. Xie 和 F. Yu。在用于垃圾邮件检测的网络级集群上。在 NDSS，2010 年。

[26] A. Ramachandran 和 N. Feamster。了解垃圾邮件发送者的网络级行为。 ACM SIGCOMM 计算机通信评论，第 36 卷，第 291-302 页。美国计算机学会，2006 年。

[27] A. Ramachandran、N. Feamster 和 S. Vempala。使用行为黑名单过滤垃圾邮件。在过程中。第 14 届 ACM 计算机和通信安全会议，第 342-351 页。美国计算机学会，2007 年。

[28] 返回路径。电子邮件智能报告，2012 年第 3 季度。

[29] K. Thomas、C. Grier、J. Ma、V. Paxson 和 D. Song。实时 url 垃圾邮件过滤服务的设计与评估。 IEEE 安全与隐私研讨会，2011 年。

[30] A. G. West、A. J. Aviv、J. Chang 和 I. Lee。使用时空信誉减少垃圾邮件。技术报告，DTIC 文件，2010 年。

[31] W.-t。 Yih、R. McCann 和 A. Kolcz。通过检测灰色邮件改进垃圾邮件过滤。中国经济研究会，2007 年。

[32] S. Youn 和 D. McLeod。使用个性化本体对灰色电子邮件进行垃圾邮件决策。国资委，2009 年。