暗中低语：匿名社交网络分析

Gang Wang, Bolun Wang, Tianyi Wang, Ana Nika, Haitao Zheng, Ben Y. Zhao
Department of Computer Science, UC Santa Barbara

{gangw, bolunwang, tianyi, anika, htzheng, ravenben}@cs.ucsb.edu

摘要

近年来，社会互动和人际交往形式发生了重大变化。隐私意识的不断提高和诸如Snowden棱镜门事件的爆出，导致了新一代匿名社交网络和社交程序的快速发展。这些服务不使用传统认证方式，通过去除用户的强身份认证和模糊用户的社会关系的方式，鼓励用户和陌生人之间的交流，使得用户发表自己观点的同时不必担心遇到欺凌和报复。

尽管有数百万用户和数十亿的月度网页浏览量，但是几乎没有基于匿名社交APP，例如Whisper，改变社交互动的形式和内容的实证分析。在本文中，我们展示了对一个匿名社交网络进行的第一次大规模实证研究结果，对Whisper使用了整整3个月的网络跟踪，覆盖了超过100万独立用户编写的2400万个whispers。我们试图去了解匿名性和社交联系的缺乏如何影响用户的行为。我们从多个角度分析了Whisper，包括在缺乏持续不断的社交联系的情况下用户交互的结构、用户参与性和网络粘性随时间的变化。最后，我们策划并实施一个攻击，用以追踪Whisper用户的详细位置。我们已通知Whisper，他们已采取措施去寻找基于我们攻击的应对方法。

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences;

K.6 [**Management of Computing and Information Systems**]: Se- curity and Protection

主题

措施；设计；安全

关键词

匿名社交网络；图表；用户参与性；隐私

1. 引言

在过去十年中，Facebook，LinkedIn和Twitter等在线社交网络（OSN）彻底改变了我们的沟通方式。 通过将我们的线下社交方式转变为线上模式，这些在线社交网络极大地扩展了我们社交互动的数量和频率。

然而，在线社交网络的格局正在发生变化。发布在Facebook上的内容被用于审核求职者，离婚诉讼和解聘员工。此外，研究发现隐私挖掘产业发展迅速，社交网络的变化在鼓励更广泛的信息共享。 最后，在斯诺登棱镜门后这种发展趋势被加速了，许多文章都在提醒着广大互联网用户——他们的在线行为在NSA和其他实体的不断审查下。

所有这些因素都促成了新一波隐私保护通信和社交网络工具的迅速崛起。 这些快速增长的服务是伪匿名消息移动应用程序：SnapChat因为可以让照片在几秒钟内自毁上了热搜; Whisper允许用户匿名向公众发布他们的想法; Secret 允许用户与朋友分享内容，而不会泄露他们自己的身份。 这只是冰山一角，许多类似的服务越来越频繁出现，例如Tinder，Yik-yak和Wickr。

这些社交工具的匿名特性既吸引了狂热的支持者，也招来了众多批评者。 支持者认为，它们为举报者提供了有价值的渠道，可以避免遭到诉讼，还可以允许用户表达自己的观点而不必担心欺凌或报复[40,41]。 批评者认为，网络中问责机制的缺乏会导致并助长负面行为，例如人身攻击，威胁和谣言传播[2,4]。 然而，所有各方都认为这些工具对用户的沟通交流方式产生了巨大影响。

在本文中，我们通过对Whisper的详细测量和分析，描述了我们研究伪匿名社交网络的经验和发现。 Whisper是一款移动应用程序，允许用户在图像（例如Internet memes）上发布和回复公共消息，所有这些都使用匿名用户标识符。 Whisper不会将任何个人身份信息与用户ID相关联，不归档任何用户历史记录，也不支持用户之间的持久社交链接。 这些设计选择与Facebook等网络的设计选择极为相反。 然而，他们使Whisper成为最受欢迎的新社交网络之一，每月的页面浏览量超过30亿。 作为我们的工作数据集，我们从2014年2月开始，捕获了Whisper三个月里全部的数据流，其中包括超过100万个独立用户撰写的超过2400万条whispers和回复。

---

1 ）据我们所知，没有关于Whisper用户数的公开数据。

与具有身份验证性和社交联系的传统社交媒体相比，我们将研究重点放在Whisper匿名性的影响上。 鉴于Whisper与Facebook和LinkedIn等现有主流社交软件之间的巨大差异，我们的分析可能对未来社交网络的基础架构，消息网络中的用户隐私问题以及人们对社交行为的看法产生重大影响。 更具体地说，我们的研究还揭示了为什么匿名通信网络可以具有长期可持续性，即使这些网络消除了通常被认为是当今网络“粘性”关键的持久社会联系。我们的分析提供了几个重要发现。

- 第一，我们希望在没有社会联系的情况下理解用户交互。我们构建交互图并将其与Twitter和Facebook等传统社交网络进行比较。毫不奇怪，我们发现用户通信模式显示出高色散，低聚类，与现有系统有显着差异。对于每个用户，我们观察到他和他“朋友”的交互非常短暂，浓厚、长期的友谊是罕见的。
- 第二，我们对用户的研究表明，新用户的流量明显促进了平台内容的产生，用户明显分为短期（1-2天）和长期用户。我们通过将ML技术应用于用户仅1周的活动历史记录，可以将用户类型准确地分组。
- 第三，我们关于“删除的低语”的主要内容的研究表明，大多数被删除的低语都集中在成人内容上，而Whisper的审核小组通常会在攻击性的低语最初发布后的短时间内将它删除。
- 最后，we identified a significant attack that exposes cur- rent Whisper users to detailed location tracking. 我们描述了这次攻击和我们的实验。请注意，我们已向Whisper通知了此漏洞，他们正在采取积极措施来缓解此问题。

据我们所知，我们是第一个对Whisper和伪匿名消息系统进行详细研究的机构。这些移动平台上用户的快速增长的事实表明，伪匿名消息系统可能对当今已建立的OSNs提出了真正的挑战。 我们相信，我们的初步工作揭示了这些系统作为人际交往的新平台，并提供了对网络基础设施设计的深入了解，以支持Whisper和类似服务。

2. 背景和目标

在本节中，我们将简要介绍有关Whisper网络的背景信息，然后对我们研究的目标进行高度总结。

2.1 背景：Whisper网络

Whisper.sh是一款有两年历史的智能手机应用程序，已成为新一波伪匿名消息传递和社交通信服务的领导者，这些服务包括Snapchat，Secret，Tinder，Yik-yak，Ether和Wickr。 虽然详细功能可能有所不同，但这些软件都为用户提供了发表观点，分享秘密或八卦的服务，同时保证了用户的匿名和不可跟踪性。

作为仅限移动设备的服务，Whisper允许用户使用匿名昵称发送消息，接收回复。 自2012年推出以来，它大受欢迎，截至2014年初平均每月页面浏览量超过30亿[16]。它的功能非常简单：应用程序根据消息中的关键字找到一张背景图像，在这个背景图像上叠加每个用户的短文本消息（图1）。 由此产生的私语会以用户的随机或自选昵称发布给公众。 其他人可以匿名发送消息（Whisper的“喜欢”版本），或者在他们自己发布私语后进行公开回复。 此外，用户可以向私语作者发送私人消息以开始聊天，并且私人消息仅对消息双方参与者可见。

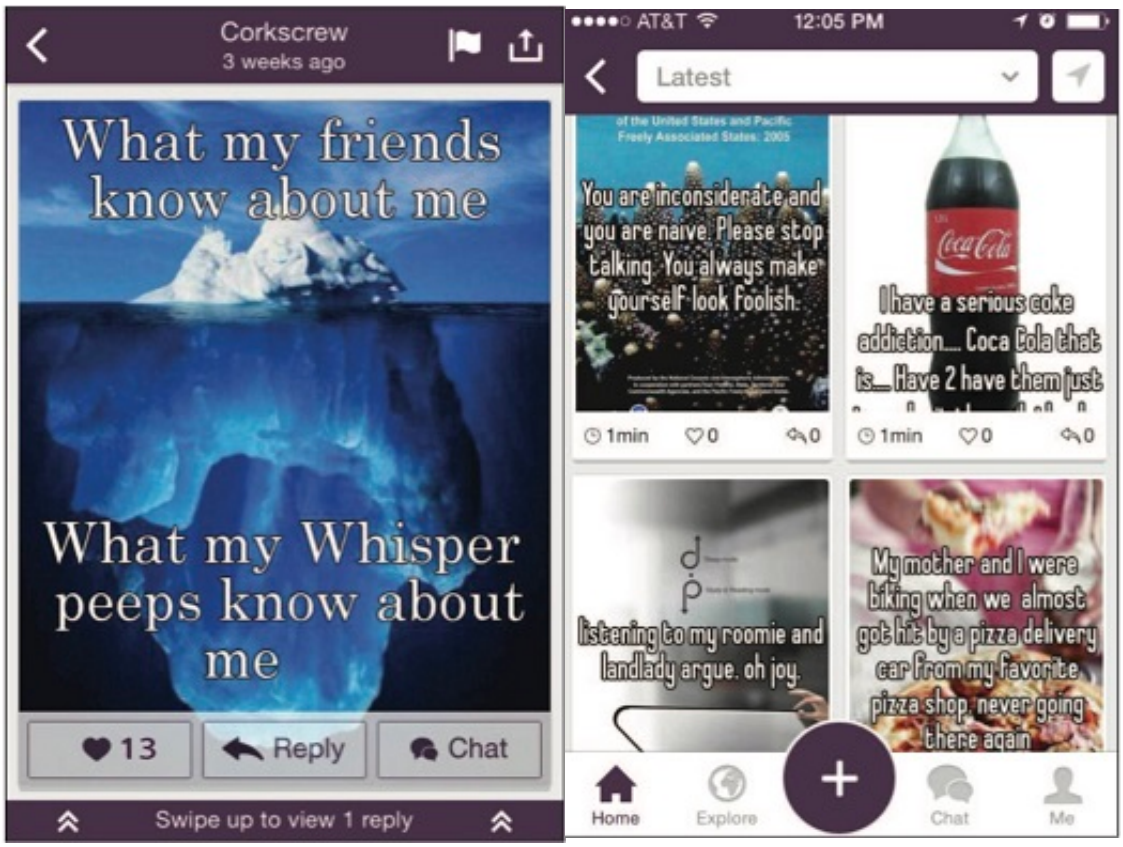


图1：一个样本私语消息的屏幕截图（左）和最新私语的公共流（右）。

用户匿名 Whisper的匿名性打破了Facebook或Google+等传统社交网络中的一些核心假设。 首先，Whisper用户仅通过随机分配（或用户选择）的昵称来识别，而不是与之相关联的任何个人信息，例如电话号码或电子邮件地址（2）。其次，Whisper服务器仅存储公共私语，用户的私人消息仅存储在他们的终端上。 没有搜索或浏览特定用户的历史低语或进行回复的功能。 第三，用户之间没有持久社交联系的概念。（例如，Facebook上的朋友，Twitter上的的追随者） 因此，它鼓励用户与各种陌生人交互，而不是与已知的“朋友”群体交互。

公共供稿 没有社交联系，用户可以浏览来自多个公共列表的内容，而不是他们的朋友（或追随者）的新闻源。 这些列表中包括一个最新的列表，其中包含最新的私语（系统方面）；附近的一个列表，显示附近区域用户发布的私语（半径范围约40英里）；一个流行的列表，只显示收到许多喜欢和回复的私语；一个精选列表，其中显示了由Whisper的内容管理员亲自挑选的部分流行语音。所有这些列表根据时间优先排序。

（2） 在服务器端，Whisper会将新用户与全局唯一标识符（GUID）相关联，并将其绑定到用户电话的DeviceID。 当通过iCloud切换到新手机时，用户可以转移他们的帐户（私人消息历史记录）。

2.2 目标

在目前的形式中，Whisper是一个研究伪匿名性对社交网络影响完美选择。 三个关键属性使其成为研究和分析的理想选择。 首先，Whisper是集中式的，即所有用户都可以访问单个数据流。 其次，Whisper适用于定期数据收集，即内容未加密且保持一段合适的时间。 第三，我们能够见到Whisper的管理团队，并获得了收集和分析Whisper公共数据流的许可。

在较高的层面上，我们的主要目标是了解用户如何在伪匿名社交网络上进行交互，匿名性又是如何影响用户行为，以及其对用户交互，用户粘性和网络稳定性的影响。 除了对Whisper网络结构的基本分析外，我们还可以将目标确定为几个具体问题。 首先，Whisper用户如何在匿名环境中互动，他们是否形成了类似于传统社交网络的社区？ 其次，Whisper缺乏身份认证的特性会不会消除用户之间的紧密联系？它是不是消除了在传统社交网络中影响用户长期参与的至关重要的粘性？ 考虑到缺乏用户特定的网络效应，它的短期历史记录是否可以用于模拟和预测用户参与度？最后，伪匿名对用户内容和用户隐私有何影响？

3. 数据和初步分析

在深入分析Whisper前，我们首先描述我们的数据收集方法和收集的数据集。 然后，我们对收集的数据集做一些高级分析。

3.1 数据收集

我们的目标是收集整个网络中发布的低语和回复。鉴于Whisper不存档历史数据，我们的方法是在很长一段时间内（2014年2月至5月）持续抓取新发布的低语。我们专注于“最新”列表，这是来自所有Whisper用户的最新低语的公共流。与其他公共列表（例如“附近”和“流行”）不同，“最新”列表提供对网络中整个窃听者流的访问。由于Whisper不提供第三方API，因此我们通过零散的Whisper网页来抓取“最新”列表。 每个下载的低语都包含一个低语WhisperID，时间戳，低语的纯文本，作者的GUID，作者的昵称，位置标签以及收到喜欢和回复的数量。作者的GUID并非旨在充当每个用户的持久ID，而是由于Whisper依赖于私人消息的第三方服务所以才以这种方式实现。根

据作者的GUID可以随时跟踪用户的帖子。在我们向Whisper的管理团队报告此问题后，他们在2014年6月删除了GUID字段。位置标记显示了城市和州级别的用户位置（例如，洛杉矶，加利福尼亚州），并且仅在低语作者启用这个功能的时候才能显示。对低语的回复我们采用相似的方法，唯一的区别是回复中带有被回复的内容。

爬取数据 我们采用了一个分布式网络抓取工具，它包含两个组件，一个主要抓取工具，可以提取最新的私语列表，还有一个回复抓取工具，用于检查过去的低语并收集与现有低语相关的所有回复序列。我们观察到Whisper服务器保持最新10K低语的一个队列。每30分钟运行一次主爬虫程序可确保我们捕获所有新的低语。相比之下，抓取回复的计算量更大。我们每隔7天抓取一次回复，并检查上个月写的所有私语的新回复。在实践中，我们发现很多低语在发布后一周内都没有收到任何后续回复。 我们在2014年2月6日至5月1日期间运行了我们的爬虫。在大约3个月的时间内，我们收集了9,343,590个Whispers，其中有15,268,964个回复和1,038,364个唯一GUID。感谢服务器端队列，我们可以收集到连续数据流，尽管有其间我们有少量中断用以更新爬虫代码。唯一值得注意的是，在4月20日的一次Whisper请求中，我们使用一组新的API调用我们的抓取工具，将它转移到另一个Whisper服务器。虽然这种转移减少了Whisper的负载，但是却没有收集到低语的位置标签。由于这仅影响了10天的数据，我们认为这对我们对基于位置的功能的分析几乎没有影响。

验证一致性 我们使用一个小实验进一步验证“最新”流的完整性。我们使用HTTP请求来同时抓取不同城市附近的6个位置的“附近”流：西雅图，休斯顿，洛杉矶，纽约，旧金山和芝加哥。我们用了6个小时来捕获这些流，确认了来自6个位置的2000多个耳低语都在同一时间段内出现在“最新”流中。

限制 我们的实验没有捕获两种类型的数据。第一，我们不会捕获只读取/利用低语但从不发布任何内容的用户。由于这些被动用户不会产生明显的用户交互，因此不太可能影响我们的大部分结论。第二，我们的数据仅限于可见的公共数据，我们无法访问用户之间的私人消息。因此，我们的结果代表了系统中用户交互的下限。正如我们稍后讨论的那样，我们认为公共互动和私人信息之间应该有很强的相关性。

3.2 初步分析

接下来，我们基于我们收集的回复和用户信息的数据集，提供一些高层面上分析的结果。我们为我们本节的结果设置了上下文，以便在后续部分中更详细地分析用户行为和匿名性。

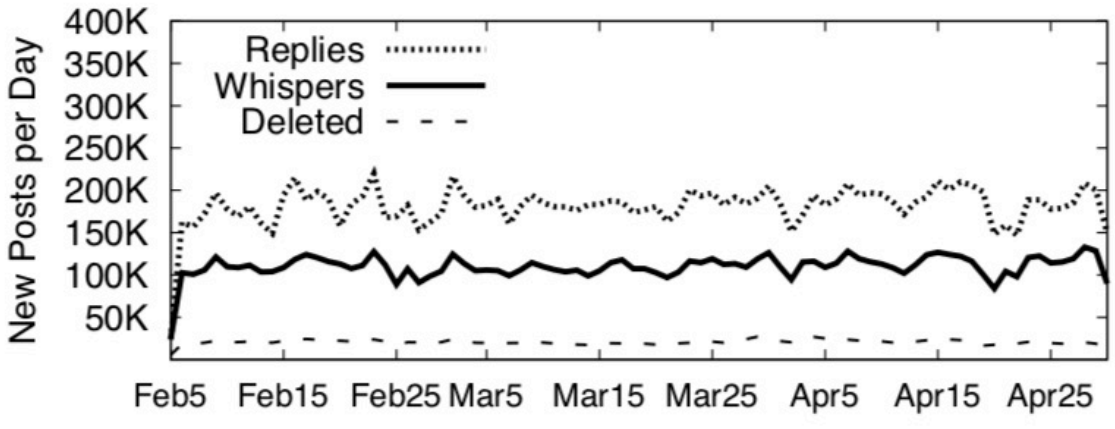
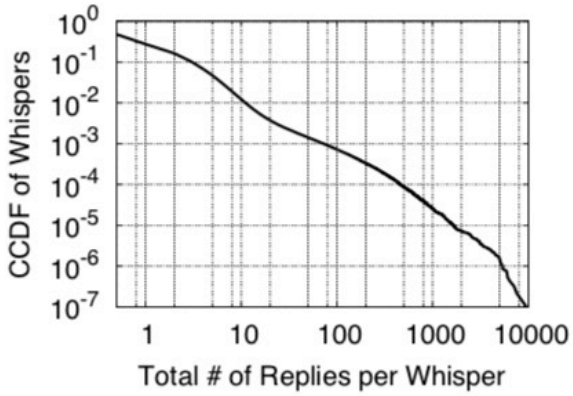


图2：Whisper上每天新产生低语，新的回复以及被删除的低语的数量

Whispers Over Time 首先我们来看一段时间内发布的低语。图2显示了我们研究期间每天发布的新低语和回复的数量。如图所示，Whisper中的新内容的发布数相对稳定，平均每天有100K新低语和200K回复。一个有趣的观测结果是，在任何时间范围内，对低语回复都比低语要多得多。

在我们的数据收集过程中，我们发现作者或Whisper管理员删除了很大一部分的低语。据我们所知，旧的Whispers不会"过期"地并停留在Whisper服务器上，但是可以在回复中被引用。但是，对于 被删除的低语，当我们尝试重新抓取他们的回复时，我们会收到"低语不存在"的错误。在每天发布的 100K大小的新低语中，大约18%最终被删除。我们将在§6中详细分析被删除的低语。



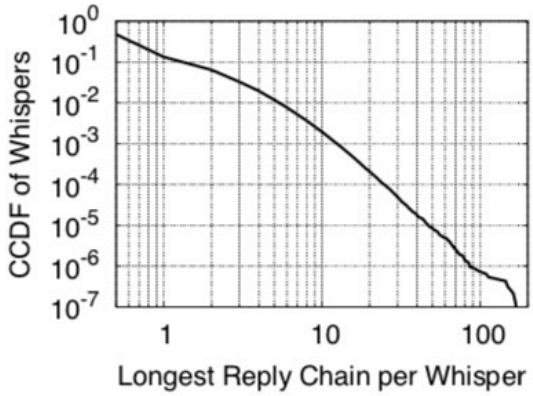
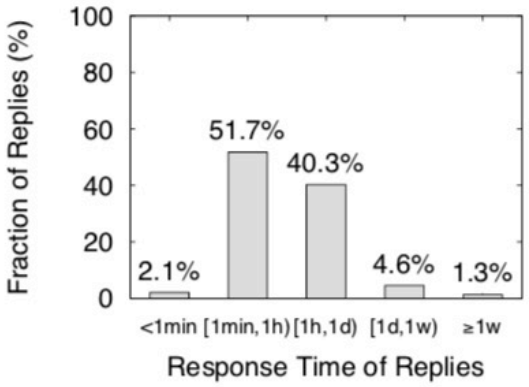


图3：每一个低语收到回复的总数 图4：每个低语的最长链长



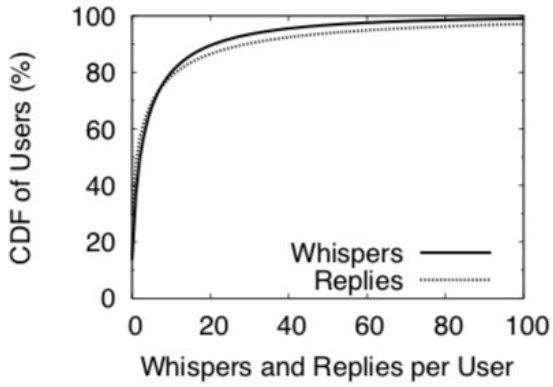


图5：每个回复与原低语之间的时间间隔 图6：每个用户发布的低语和收到的回复数量

回复 用户可以回复新的低语或其他之前发布的低语。用户的多个回复可以生成自己的回复，从而形成以原始低语为根的树结构。图3 和图4 显示了每个低语的总回复数和每个低语的最长链长（最大树深度）。不出所料，55%的低语没有得到回复。由于所有的低语都张贴在同一个公共列表中，因此每个低语只有很短的时间窗口来吸引用户的注意力。在被回复的低语中，大约25%的低语至少有2个回复链。以上这些在本质上成为用户之间对话的联系。

图5 描绘了回复发布时间的分布，这是每个回复与原始低语之间的时间间隔。 54%的回复是在原始低语发布的一小时内发布的，超过94%的回复在一天内发布。只有1.3%的回复在低语后一周或更长时间发布。这证实了我们的直觉——如果在发布后没有引起注意，那么以后就不太可能引起注意了。

用户 我们根据唯一的GUID查看每个用户生成的内容。图6 绘制了每个用户发布的低语和收到回复的数量。大多数用户（80%）总的发布低语或收到回复的数量少于10个。大约15%的用户只回复，但自己的低语，30%的用户只发了低语但没有收到回复。

内容分析 对低语内容的高层面分析表明，用户发布的内容高度个人化。搜索单数第一人称代词（例如，我，（人称代词I的宾格）我，我的，我自己）约占所有低语的62%。我们还发现情感关键词的大量使用。具体来说，40%的低语包含WordNet提供的1113个与人类情绪相关的关键词之一[33]。最后，人们经常提出寻求建议或同情的问题。根据低语中问号和疑问句的使用（例如，什么，为什么，哪些），大约20%的低语属于该问题。这三个类别有效覆盖了85%的低语。很明显，Whisper的匿名性鼓励用户在没有隐私暴露问题担心的情况下发布个人的私密内容。我们将仔细研究 §6 中的低语语“主题”。

4. 用户交互

我们的研究始于用户在Whisper上的互动。 由于Whisper用户无法建立持久的社交联系，这从根本上改变了用户交互和建立友谊的方式。 在本节中，我们将研究基于低语及其回复构建的双向交互过程，并寻求从三个不同层面理解用户交互。 首先，我们通过比较Whisper交互图的结构属性与传统OSN（例如，Face-book和Twitter）的结构属性来研究全球网络级别的交互。 其次，我们在Whis- per图中研究用户社区，并探索推动其形成的关键因素。 最后，我们研究用户级别的交互，以了解用户是否仍然在Whisper中建立了强关系（经常是交互的朋友）。

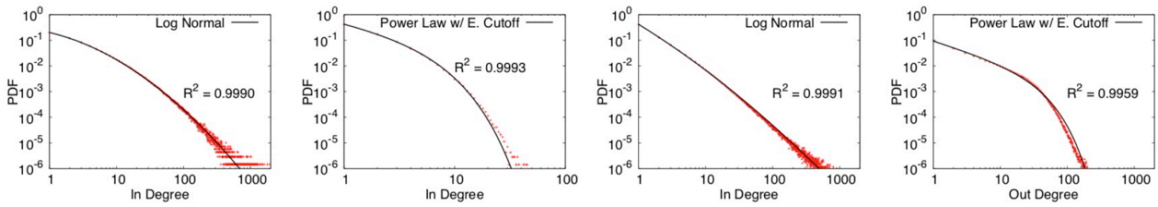
4.1 Whisper 交互界面

我们首先将Whisper的交互图与传统的在线社交网络（Facebook和Twitter）进行比较。 我们的目标是了解Whisper中缺少社会认证是否从根本上改变了用户在聚合网络级别的交互模式。 我们基于低语和后续回复构建了一个Whisper交互图，并将其结构与Facebook墙上帖和Twitter转推构建的图表进行比较。

构建交互图 我们基于低语和后续回复构建了Whisper交互图，这是Whisper中主要的公开可见交互。结果是一个相互作用的交互图，其中节点是用户，边表示回复操作。例如，如果用户A将回复B用户发布的低语，我们将构建从A到B的有向边。只有直接的回复用于构建边。我们从图中删除断开连接的单例节点。我们从3个月的数据集（Whisper-all）中生成主要交互图.

为了进行比较，我们还根据我们之前工作中s收集的匿名数据集为Facebook和Twitter构建交互图[39,42]。这两个数据集都抓取了历史数据，这些数据涵盖了至少3个月内的用户交互。我们使用Facebook墙贴数据构建了一个定向交互图：如果用户A在用户B的墙上发布，我们创建了从A到B的有向边。对于Twitter，我们基于转发交互构建了图：如果用户A转发了来自用户B的消息，我们创建从A到B的有向边。为了匹配Whisper图3个月的时间覆盖，我们使用覆盖3个月时段的数据构建类似的Facebook和Twitter图。表1显示了所有三个交互图的关键统计数据。

等级分布和拟合函数 Whisper中的用户显示出比Facebook和Twitter中的用户高得多的平均度，这意味着用户与更大的其他用户样本进行交互。我们使用3个常用的社会图拟合函数，幂律 $(P(k) \propto k^{-\alpha})$，指数截止幂律来确定每个图的度分布的最佳拟合函数 $(P(k) \propto k^{-\alpha} e^{-\lambda k})$ 和对数正态 $(P(k) \propto e^{\frac{(\ln x - \mu)^2}{2\sigma^2}})$ [14,39]。我们遵循[10]中的拟合方法，并使用Matlab计算拟合参数和精度（R平方值），并在图7中显示结果。对于Whisper和Facebook图形，out-degree分布看起来类似于程度分布。为简洁起见，我们仅显示每个图的度内分布。直观地说，Facebook旨在模仿离线社交关系，而普遍的双向交互导致对称的出入度分布。对于Whisper，用户之间的用户交互很大程度上是随机的。相比之下，Twitter的入度和出度分布明显不同。众所周知，Twitter更像是一种信息传播媒介，而不是社交网络，而且互动是高度不对称的[25]。



(a) Whisper In (b) Facebook In (c) Twitter In (d) Twitter Out

图7：度发布和拟合结果

聚类系数 聚类系数是节点的直达节点数与它可能到达的所有节点数的比率。 它测量节点之间的本地连接级别。 Whisper图中的聚类系数（0.033）远小于Facebook（0.059）和Twitter（0.048）。 原因很明显：Whisper用户极有可能与陌生人互动，而这些陌生人不太可能互相交流。

平均路径长度 平均路径长度是图中所有最短路径对的平均值。图中所有成对的最短路径中的60个。考虑到我们图的数量，计算所有节点对的最短路径是不切实际的。 相反，我们在每个图中随机选择1000个节点，并计算从它们到图中所有其他节点的平均最短路径。 结果表明，Whisper图具有3个网络的最短平均路径长度。 这又是一个直观的结果，因为陌生人之间随机地相互作用，图中产生了大量的短路径，从而缩小了平均路径长度。 考虑到Whisper的高平均度，低聚类水平和较短的平均路径长度，Whisper展示了比Facebook和Twitter等“小世界”网络更多的随机图[38]的特性。

同配性 同配性系数测量图中节点连接到其他度值相近节点的概率。 同配性 > 0表示节点倾向于与其他度值相近节点连接，而同配性 <0表示节点连接到具有不同度值节点。 我们的结果表明，Whisper图的同配性系数非常接近于零（ - 0.011），这非常类似于随机图[29]。 相反，相似的用户倾向于在具有双向链接（例如，Facebook）的社交网络中聚集在一起，产生正值的同配性（0.116）。 在Twitter中，大量普通用户关注着名流和著名人物，从而产生更负值的分类（-0.025）。

4.2 交互图中的社区

接下来，我们分析了Whisper交互图中存在的社区结构。社区被定义为内部紧密联系但极少与外部关系网进行交互的节点集，具有高模块性。我们试图回答两个核心问题。首先，没有持久的社交联系，Whisper用户是否仍然可以在交互图中形成社区？其次，如果Whisper用户在交互图中形成了社区，推动Whisper用户形成社区的关键因素是什么？

社区检测。我们首先将社区检测算法应用于Whisper图，以检查是否存在社区结构。我们选择两种广泛使用的社区检测算法Louvain [7]和Wakita [37]。我们计算结果社区的模块性。模块化[28]是一个广为接受的社区检测度量标准，它衡量社区内连接的比例与随机连接时的预期分数之间的差异。模块化范围从-1到1，更高的值更强的社区。

为了使用图形捕获用户交互，我们基于两个节点之间的交互次数来权衡图形边缘。此外，我们将此分析重点放在最大的弱连接组件上，该组件包含所有节点的99%。应用Louvain为Whisper产生的社区平均模块化值为0.4902。在实践中，模块性> 0.3表示图形中的重要社区结构[24]。我们使用Wakita社区检测算法确认我们的结果，并找到0.409（也高于0.3）的结果模块性。作为参考，现有社交图的模块化分数包括Facebook（0.63），Youtube（0.66）和Orkut（0.67）[24]。毫不奇怪，Whisper中相对较弱的社区与其他观察结果相匹配，包括低聚类活动和弱关系。

用户社区 vs 地理位置 后续问题是，为什么Whisper中有社区？如果用户交互是随机的，那么不应该所有交互都是统一的吗？我们的假设是，这是由于Whisper中的“附近”功能，允许用户浏览（并可能回复）附近区域的人发布的低语。我们的直觉是，附近的流驱动用户更频繁地与同一地理区域中的其他人进行交互，从而有助于基于地理驱动的社区在交互图中形成。

为了测试这个想法，我们检查了每个社区中用户最多的地理区域。如果地理因素是形成社区的关键驱动力，那么社区应该由来自同一地点的用户主导。表2显示了Louvain及其相应顶级区域生成的前5个Whisper社区。这里我们使用“州”或“省”级别的位置标签。我们发现前5个社区都证实了这一点：大多数用户偏向于一个地区或几个地理位置相邻的地区（例如纽约，新泽西州和CT为C1）。

Community (size)	Top 4 Region (% of users)
C_1 (61,686)	NY (11), NJ (10), CT (4.8), CA (4.2)
C_2 (39,824)	England (61), Wales (3.5), CA (1.1), TX (0.9)
C_3 (28,342)	CA (62), TX(1.5), England (1.2), AZ (0.9)
C_4 (22,010)	IL (37), WI (21), IN (4.5), CA (1.5)
C_5 (16,017)	CA (64), England (1.4), TX (1.3), NY (0.8)

表2：前5大社区及其顶级地区。

为了在所有社区中量化这种现象，我们在图8中绘制了每个社区的前四个地理区域中的用户比例。Louvain生产了912个不同规模的社区，我们只考虑最大的150个社区，这些社区覆盖了> 90%的用户。结果再次证实了我们的假设，社区成员由顶级地区或前2个地区主导。这种强烈的交互地理位置证明了“附近”流在Whisper社区的形成中起着重要作用。虽然其他因素可能也有助于用户社区的形成（例如，用户的共享主题和兴趣，时区），但我们将对于它们的分析留给未来的工作。

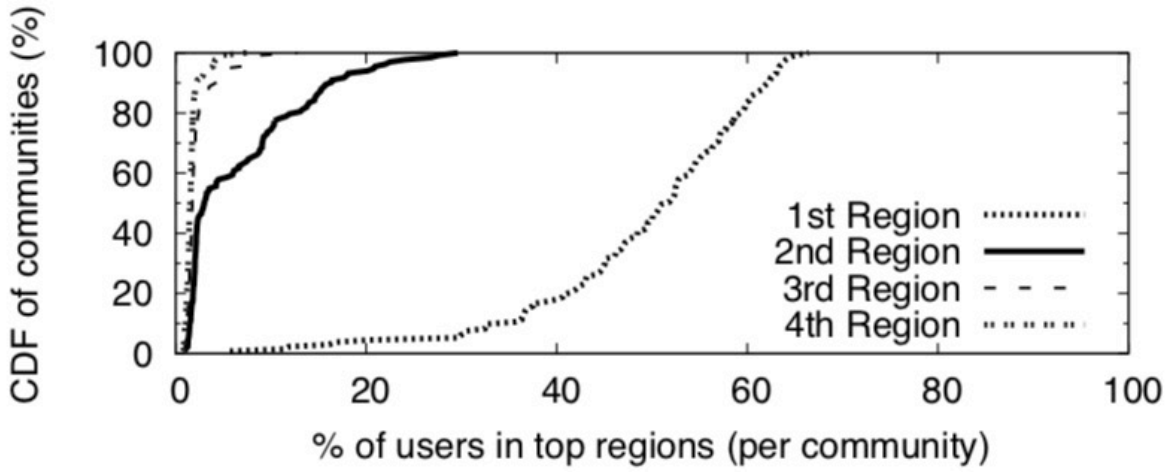


图8：每个社区的顶级区域中的用户百分比。 社区内的用户高度倾向于一个区域。

4.3 用户互动和持续交互

最后，我们在用户级别分析用户交互和隐式社交链接。 回想一下，Whisper缺乏持久的身份认证和社交链接会鼓励用户与陌生人互动。 在下文中，我们寻求两个关键问题的答案。 首先，用户是否拥有一组他们经常与之互动的固定“朋友”？ 尽管Whisper昵称具有匿名性，但这种友谊可能已经形成。 第二，这种离线友谊有多强烈呢？

单用户交互 我们通过寻找与他人互动频繁的用户群来寻找潜在的友谊（即强关系）。为方便起见，我们将用户与之交互的人（不论来源）称为她的熟人。对于每个用户，我们计算她熟人中的交互分布，并查找她与所有熟人的交互中的偏差。我们从每个用户的分布中选择几个点（50%,70%和90%），并在CDF中对它们进行聚合，以显示所涉及的熟人的百分比（图9）。为避免统计异常值，我们仅包含至少10次交互的用户。 我们发现用户交互在熟人之间的分布相当均匀。以90%的线为例，对于几乎所有用户（约90%），超过70%的熟人负责90%的互动。Whisper中这种相对较低的偏差与Facebook等传统OSN完全相反，在Facebook上，只有一小部分朋友（强关系）负责绝大多数用户的互动[39]。

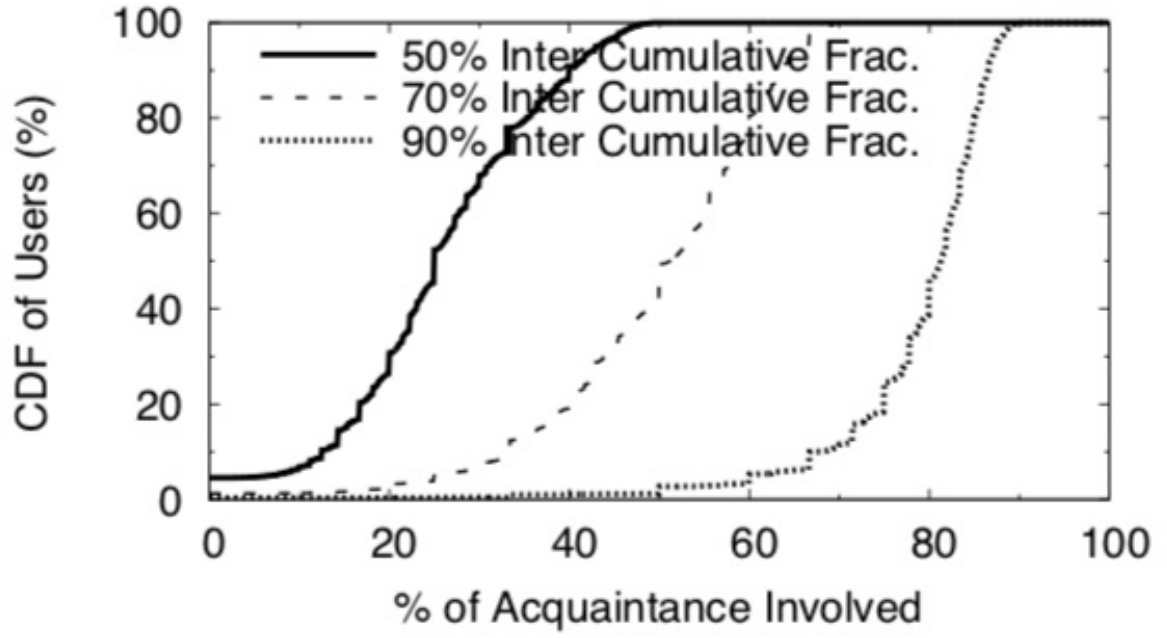


图9：对于不同的交互百分比，用户与熟人之间的交互分布。

通过低语进行的互动 在用户的熟人中，我们寻找潜在的强关系，即用户经常与之交互的熟人。图10显示了用户的熟人总数，用户不止一次交互的熟人，以及用户使用私发功能多次与之交互的熟人。在Whisper中，人们在同一个低语下进行不止一次互动是很常见的。但是，很难通过不同的低语与同一个人交谈，因为通过匿名昵称跟踪特定用户很困难。如图10所示，只有13%的用户拥有熟人，他们通过低语进行交互。 然后，我们选择那些通过低语进行交互的用户对以进行进一步分析。总共有503K这样的用户对。图11显示了这些用户对的生命周期（他们第一次和最后一次交互之间的时间间隔）以及他们在低语中的交互次数的热图。请注意，调色板是对数刻度 - 绝大多数用户对堆叠在左下角，表示短暂的，低交互关系。只有很小一部分异常值（右上角）实现了长期和频繁的互动。

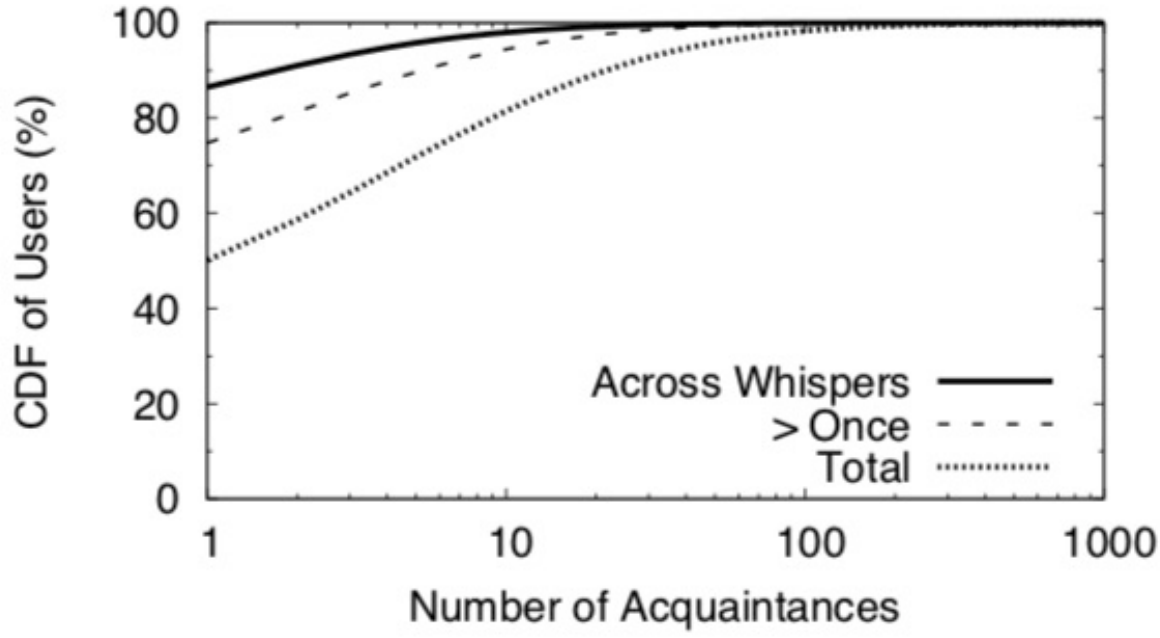


图10：用户熟人的数量，以及用户通过低语与之交互的人的数量。

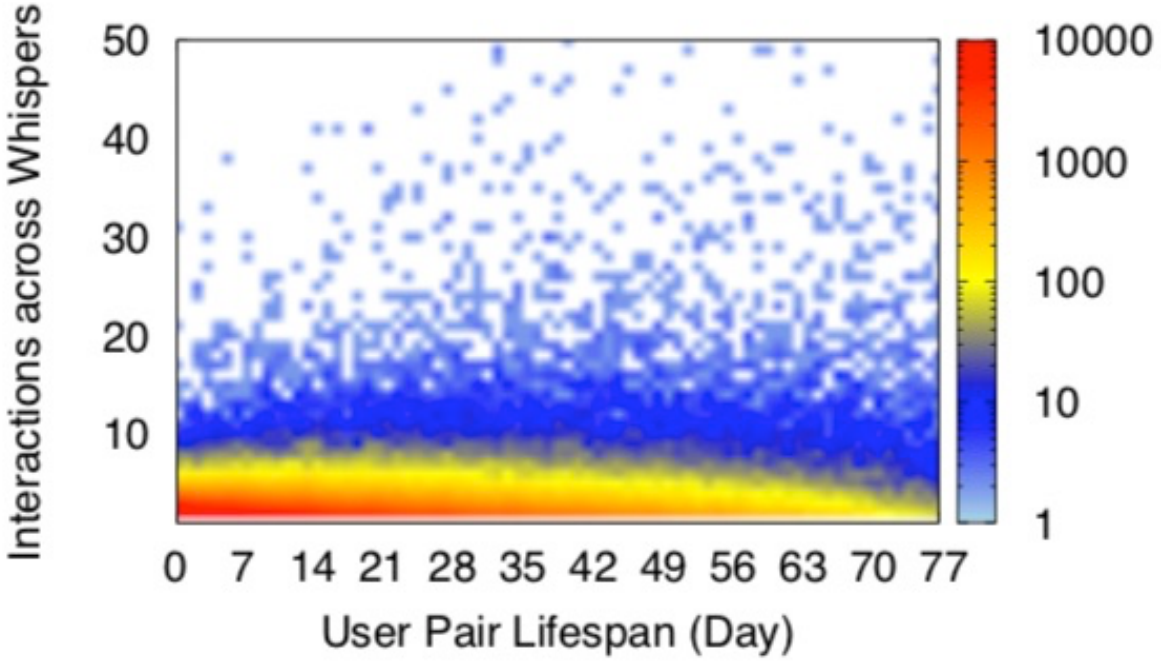


图11: 用户通过低语进行交互：生命周期与交互＃。

朋友还是随机遭遇的陌生人？ 尽管强关系是异常值，但有趣的是探索这些用户对如何通过低语不断地相互交互：这些离线朋友是否在公共信息中主动跟踪（使用昵称），或者这些只是偶然碰到？我们意识到这是一个非常难以确定的问题。但我们有一个关键的直觉：如果这些交互是真正随机的，那么这两个用户很可能共同位于同一地理区域，特别是那些Whisper用户稀疏的地区。然后，只要两个用户主动发帖，他们就有很好的机会在附近的列表中看到对方。 现在我们使用我们的数据来测试这种直觉。对于用户对交叉低语互动，我们首先检查他们的地理距离（3）我们发现在503K用户对中，90%有两个用户共同位于同一“州”，75%的距离 < 40英里，这是附近溪流的最大范围。图12显示了地理距离与用户对的交互频率之间的相关性。每个堆叠条加起来达到100%，每个类别代表具有不同交互级别的用户对（即通过低语的交互次数）。它表明频繁的交互更倾向于地理上彼此接近的用户。

图12 : 对于所有用户对，两个用户之间的距离和用户对的交互次数。

然后我们进一步检查这些位于附近区域的用户对（即距离<40英里）。更具体地说，我们分析了可能影响用户遭遇偶然事件的可能性的两个因素 -用户群体的地理区域和两个用户发布的低语总数（图13和图14）。直观地说，在相同的附近区域中较小的用户群体，一次又一次地在附近列表中遇到同一个人的机会更高。同样，两个用户发布的低语越多，他们就越有可能相遇并形成互动。此处，用户人口数是通过与配对用户具有相同城市级位置标签的用户总数来估算的。这两个结果证实了我们的猜测。随着用户人口密度的降低和用户帖子数量的增加，更频繁的用户对交互概率也会增加。

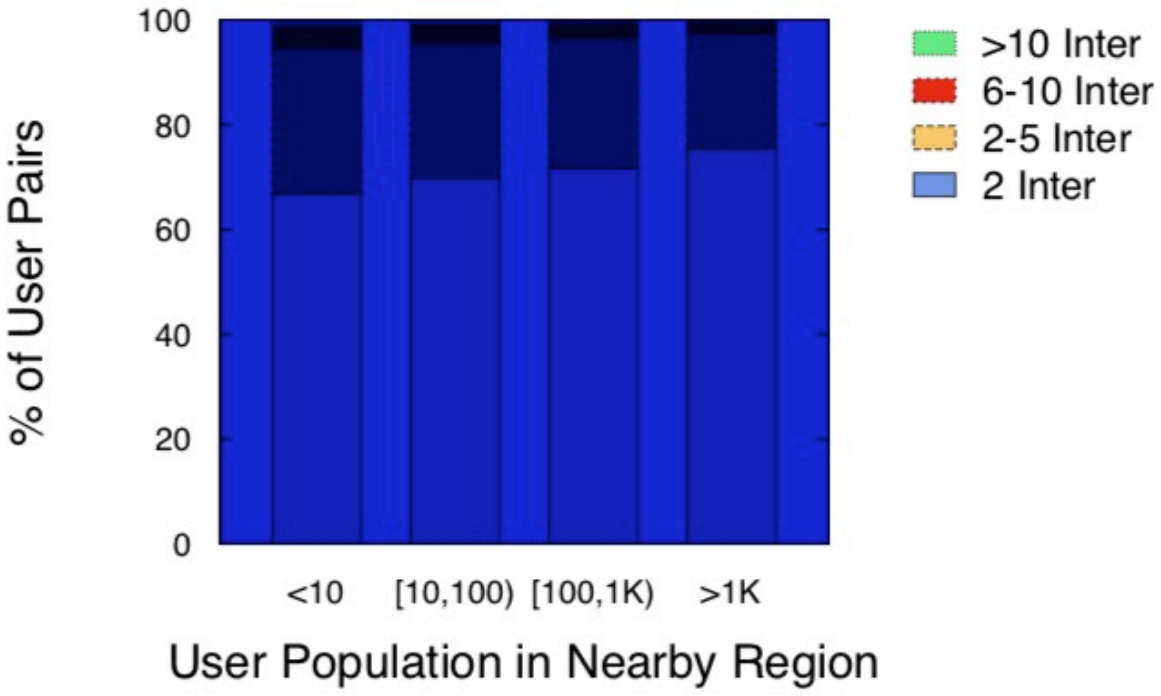


图13 : 对于附近的用户对，附近区域中的用户群与用户对的交互。

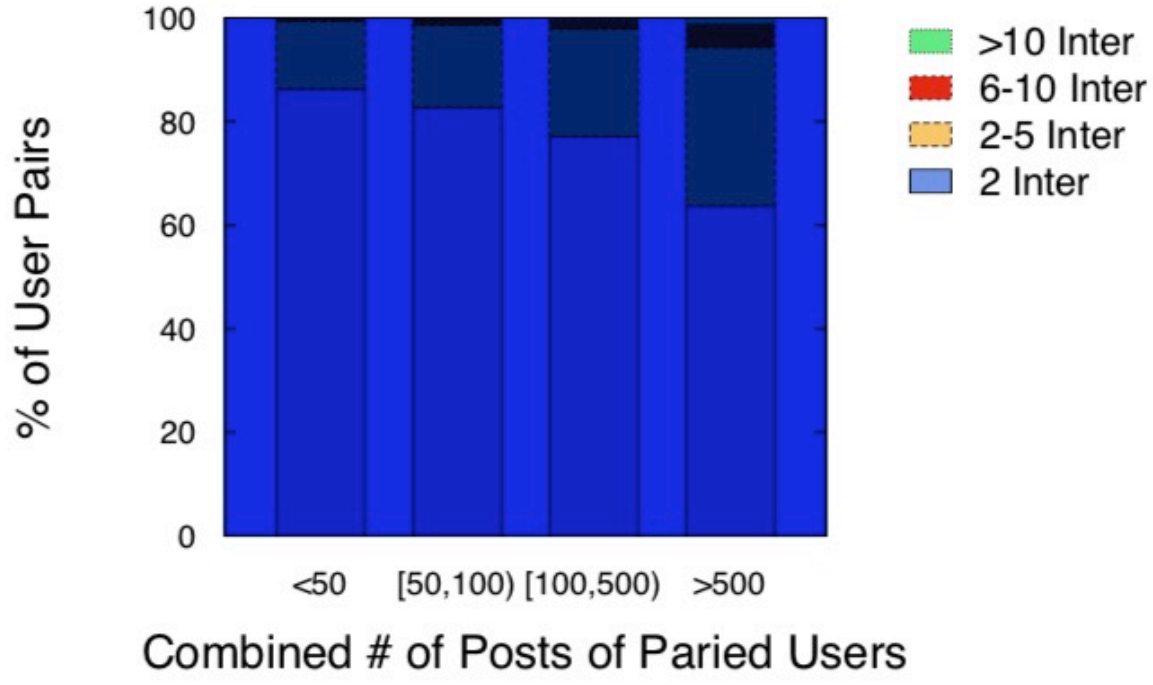


图14 : 对于附近的用户对，总低语数与用户对的互动次数。

总之，我们的分析表明，在Whisper中，强关系非常罕见。我们还发现强关系倾向于具有更高机会彼此相遇的用户对（即，共享位于具有稀疏用户群的区域中的活动用户）。因此，尽管可以通过Whisper交互发展强大的关系，但这种关系很可能受到地理密度和用户低语频率的严重影响。请注意，我们的分析仅依赖于公共交互，不包含私人消息。直觉上，我们认为用户的私密互动应该与他们的公共互动相关联，我们可以通过公共互动预测用户对与私人互动。之前的工作还证实，在建模关系强度时，公共互动比私人通信更具信息性[13,22]。

5. 用户参与度

到目前为止，我们的分析表明Whisper用户倾向于与陌生人而不是熟悉的朋友进行互动。负面后果是缺乏强关系通常会产生一个不那么“粘性”的网络，即用户离开的可能性更大[11]。这带来一个问题：如果没有强关系，Whisper用户可以长期保持参与性吗?

在本节中，我们从观测每个用户的参与度出发来考虑这个问题。首先，我们调查用户参与度，以了解我们的数据集在3个月内的用户流失情况。其次，我们对一个机器学习分类器进行评估，并表明我们可以准确地预测用户发表第一篇文章后是否会继续使用该应用。我们使用实验寻找影响用户流失的关键因素。请注意，我们的分析仅限于发布至少1个低语或回复的“活跃”用户，不包括使用但不提供内容的被动用户。

5.1 用户参与度

首先，我们使用三个指标对用户活动进行基本分析：用户数量增长，新用户的内容贡献以及用户活跃周期分布。

用户数量增长 图15显示了我们数据集中随时间（11周）变化的用户总数。每个条形图显示该周刚刚加入（新）的新用户和我们在该周之前观察到的老用户（现有）的细分。我们观察到一个稳定的新用户进入网络的到达率，每周约80,000名新用户。回想一下，尽管用户数量在增长，但整个网络中的每日新帖（低语和回复）仍大致稳定（见图2）。这表明存在一定数量的用户“脱离”，即停止发布低语或回复。

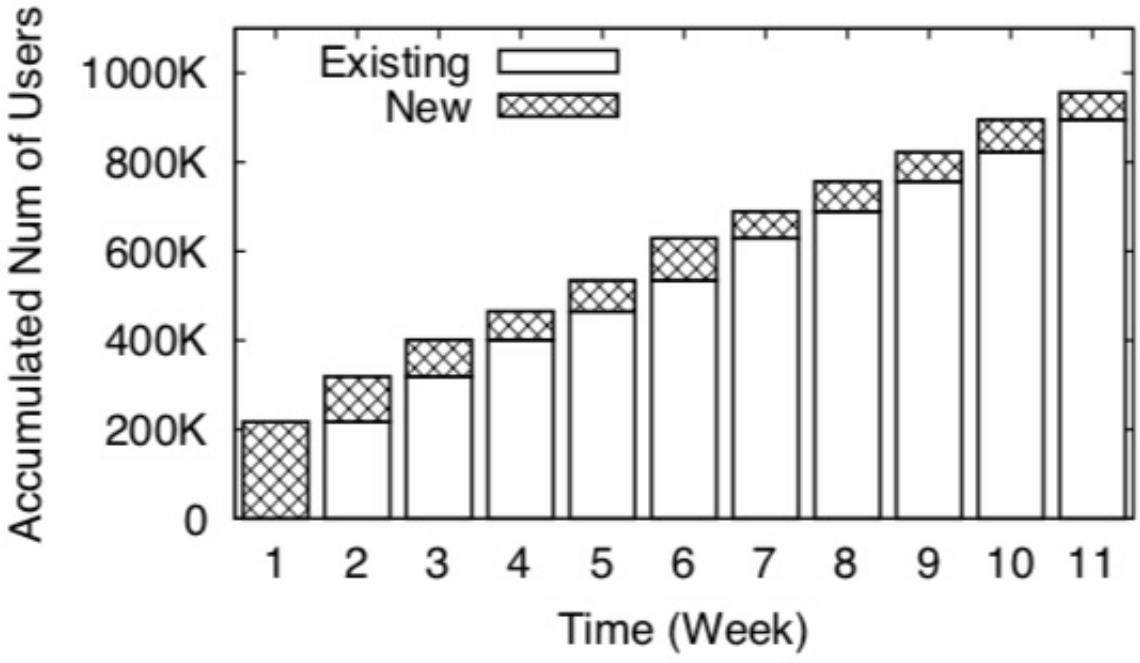


图15: 数据集中随时间变化的用户数量增长

新用户和老用户的内容贡献 这促使我们研究新用户和老用户对内容的相对贡献。图16显示了本周第一次出现的用户（新）和本周之前出现的用户（现有）的帖子细分（低语和回复）。我们发现新用户对整体低语流（> 20%）做出了重大贡献。但是，随着越来越多的用户从新用户过渡到“现有用户”，现有用户的内容生成不会随着时间的推移而显着增长。这证实了我们一开始的想法，即

一定比例的用户随着时间的推移而脱离。

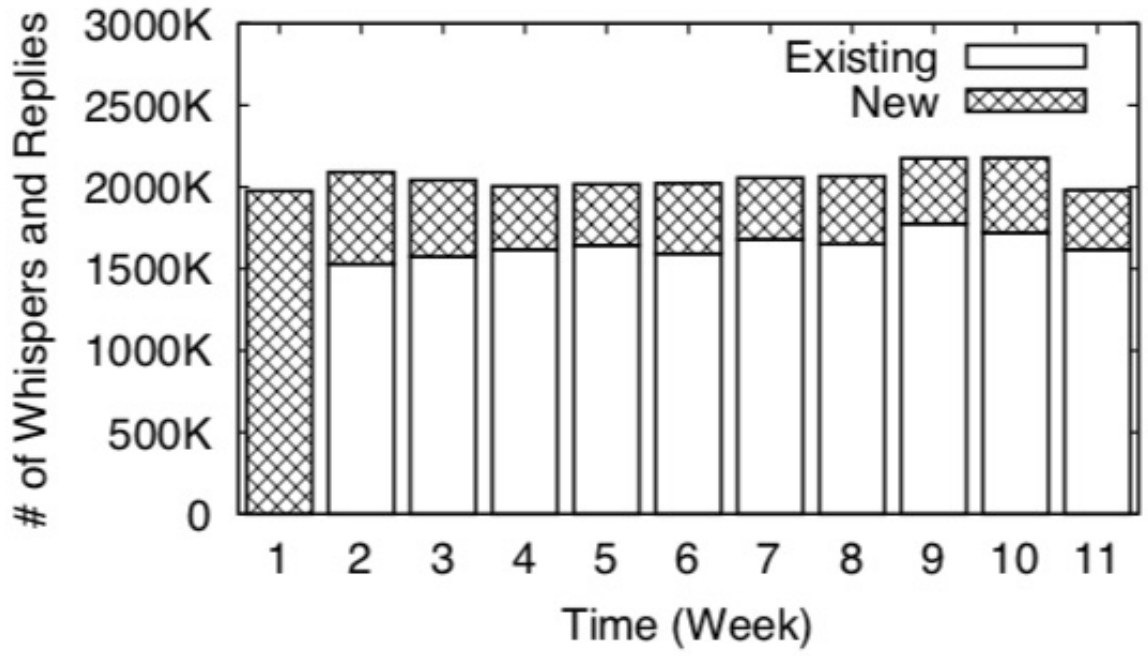


图16: 每周新老用户的低语和回复

单用户活跃期 接下来，我们关注个人用户，并检查用户在脱离前保持活跃状态的时间。更具体地说，我们计算他们停留在我们数据集中的活跃“生命周期”（他们发布的第一个和最后一个帖子之间的时间跨度）。鉴于我们专注于长期活动，我们会排除最近一个月加入我们数据集中的用户。因此，对于图17，我们仅考虑已在我们的数据集中至少一个月的用户（占所有用户的70.3%）。

图17显示了用户活动生命周期比率（PDF）的分布。用户显然分为两个极端：一个拥有极低比率的主要群集（0.03），代表那些在第一次发布后1或2天内迅速变为非活动状态的群体；另一个主要集群在1.00附近，表示在数据集中保持活跃状态的用户（至少1个月）。在其他用户生成的内容（UGC）网络中也观察到了类似的模式，例如博客和问答服务[17]。如果我们将活动比率的阈值设置为0.03，则这些“尝试并离开”用户占所有用户的30%。这解释了我们在图16中的观察现象——因为有很大一部分用户迅速脱离应用，尽管有大量新用户加入，但整体内容发布率仍保持稳定。

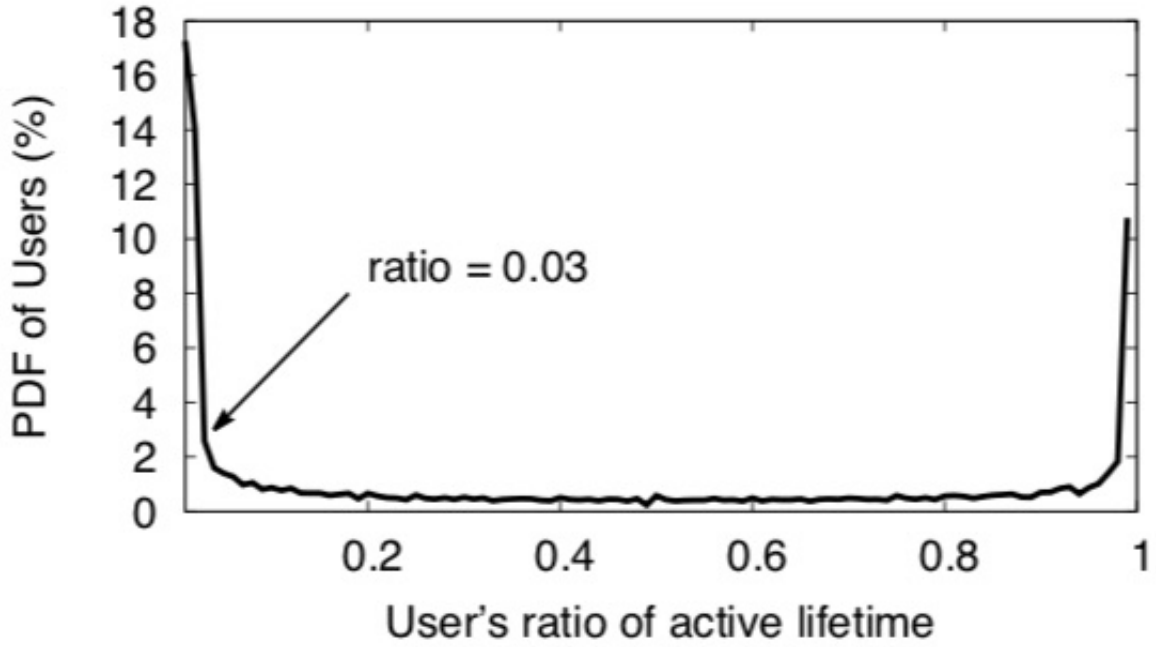


图17：用户在数据集中的有效生命周期

5.2 预测用户参与度

对上述分析得出的一个重要观察是，Whisper用户倾向于陷入两种行为极端中的一种 ——要么长时间保持活跃，要么快速转为非活跃状态（图17）。分布的双峰性质暗示了将用户分为两个集群的可行性。

在这里，我们尝试使用机器学习（ML）分类器来确定我们是否可以根据他们在第一次发布（在我们的数据集中）之后的早期行为来预测用户长期参与度。我们试图回答三个关键问题：首先，这种预测是否可能？其次，ML模型能够产生最准确的预测吗？第三，什么早期信号可以最强烈地表明用户有意离开？

我们采取三个步骤来回答上述问题。首先，我们根据用户在Whisper的前X天内的活动收集一组行为特征，理想情况下X值很小。其次，我们使用这些行为特征构建不同的机器学习分类器来预测用户长期参与度。最后，我们通过特征选择以确定什么早期特征可以最强烈地表明用户有离开倾向。

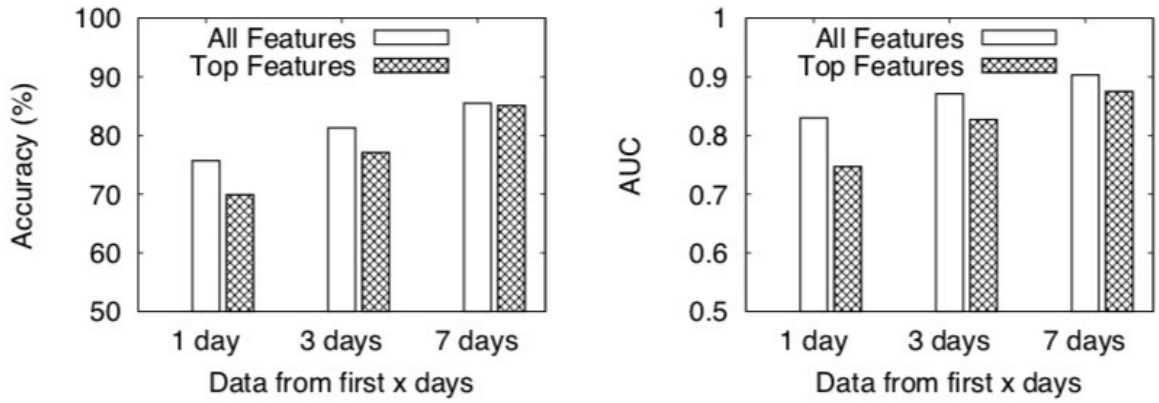
特征 我们探索了多个不同类别的特征（总共20个特征），以描述用户进入软件前X天的行为。在这些中特征，我们将选择最基本的功能。

- 内容发布特征（F1-F7）。 7个特征：用户发布的帖子总数，低语数，回复数，删除的低语数，以及用户至少发布一个帖子/低语/回复的天数。
- 交互特征（F8-F15）。 8个特征：总帖子中的回复率，熟人数量，双向熟人数量，所有回复的传出回复，与同一用户的最大互动次数，低语的获得回复的比率和低语得到的喜欢数量。
- 时间特征（F16-F17）：2个特征：用户低语收到回复的平均时间延迟;用户回复其他用户低语的平均时间延迟。
- 活动趋势（F18-F20）：3个特征：我们将每个用户的前X天平均分成三个桶并记录数量每个桶中的帖子（第一，中间和最后）。我们计算了2个特征，如Middle/First和Last/First。最后，我们计算三个桶中帖子数量是否单调减少。

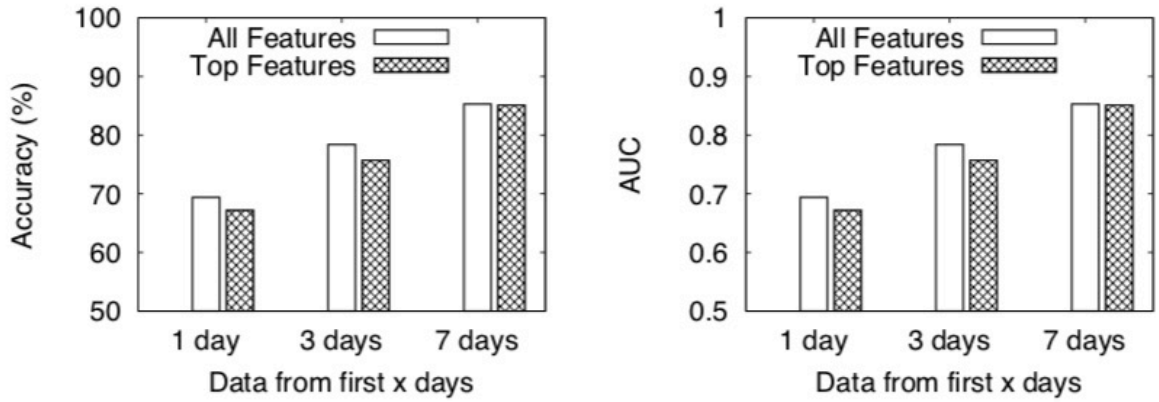
分类器实验 为了为我们的分类器构建训练集，我们使用于在我们的数据集中拥有至少一个月活动历史记录的用户（730K大小的用户）。我们从中选择了一组"短期"用户，他们试用该应用程序1-2天后快速脱离（不再发布帖子）。使用图17中的结果，我们从活动寿命比 < 0.03的那些用户中随机抽样50K大小的用户作为非活动集。然后，我们选择50K大小用户的随机样本，其活动生命周期比率 > 0.03，以形成活动集。

我们的目标是仅根据用户在前X天的活动对两组用户进行分类，并使用1,3和7作为X的值。我们构建多个机器学习分类器，包括随机森林（RF），支持向量机器（SVM）和贝叶斯网络（BN），使用WEKA [19]中的这些算法的实现和默认参数。对于每个实验，我们运行10倍交叉验证并报告分类准确度和ROC曲线下面积（AUC）。准确度是指在所有情况下正确预测的比率。 AUC是另一种广泛使用的度量，具有更高的AUC表示更强的预测能力。例如，AUC > 0.5意味着预测优于随机猜测。

随机森林和SVM的实验结果如图18所示。贝叶斯结果与SVM的结果非常匹配，因此未来简洁起见我们省略了它。我们做了两个关键的观测。首先，行为特征可以有效地预测未来的参与度。即使仅使用用户的第一天数据（RF），准确度也很高（75%）。这证实了用户的早期行动可以作为其未来预测的指标。如果我们包含一周的数据，我们可以达到高达85%的准确率。其次，我们发现给予不同的分类器7天的数据，他们有相似的结果。然而，当他们被限制使用较少的数据（例如，1天）时，他们的结果会有所不同。随着数据量的减少，随机森林比SVM和贝叶斯网络产生更准确的预测。



(a) Predicting Inactive vs. Active (RF)



(b) Predicting Inactive vs. Active (SVM)

图18：使用随机森林和SVM的预测结果。 模型性能通过精度（左）和ROC曲线下面积（右）来评估

特征选取 最后，我们寻求确定最有力的信号来预测用户的长期参与度。为了找到答案，我们对20个特征进行特征选择。更具体地说，我们基于信息增益[18]对特征进行排序，这可以衡量特征对两类数据的区别能力。我们列出了表3中的前8个特征。正如预期的那样，预测能力相差很多，信息增益在前4个特征之后迅速下降（特别是1天）。为了验证他们的预测能力，我们使用他们的前4个特征重复每个实验。图18中的结果表明，前4个特征实现了整个分类器的大部分精度，而且复杂程度要低得多。

Rank	Observation Time Frame		
	1 day	3 days	7 days
1	Interact-F9 (0.15)	Post-F5 (0.27)	Post-F5 (0.46)
2	Interact-F11 (0.12)	Trend-F19 (0.18)	Post-F6 (0.31)
3	Interact-F10 (0.11)	Post-F6 (0.18)	Trend-F19 (0.28)
4	Interact-F12 (0.11)	Interact-F9 (0.16)	Post-F1 (0.27)
5	Trend-F18 (0.05)	Post-F1 (0.16)	Post-F7 (0.23)
6	Interact-F15 (0.04)	Post-F7 (0.13)	Trend-F20 (0.21)
7	Post-F1 (0.04)	Interact-F15 (0.12)	Interact-F15 (0.21)
8	Interact-F8 (0.04)	Interact-F11 (0.12)	Post-F2 (0.19)

表3：按信息增益排列的前8个特征及其类别（括号中显示的值）

然后我们仔细研究排名靠前的特征。首先，我们注意到，与3天和7天分类器相比，1天分类器依赖于不同的特征集。 1天模型严重依赖于交互功能。直观地说，该模型基于用户参与社交交互的积极程度来预测用户是否将继续参与。如果用户在第一天收到许多回复或主动回复其他用户，则该用户很有可能停留更长时间。对于3天和7天的模型，我们发现关键特征转移到用户的内容发布和活动趋势功能。这意味着一旦我们监控用户较长时间，用户留下或离开的意图更准确地反映在他的发布频率和数量，以及该活动是否随着时间的推移而下降。

使用通知的用户 刺激用户参与是任何新服务的关键目标。 Whisper已经部署的一个工具是推送通知，每天晚上7点到9点之间向用户的移动设备提供“当天的低语”。确切的通知时间每天都在Android和iOS设备之间变化。为了检查这些通知的影响，我们进行了一项小型实验。我们每天监控5部不同手机的通知时间，持续6天。我们在通知后的5分钟和10分钟间隔内查看Whisper流中的用户活动，并且发现与晚上7点到9点之间的其他5或10分钟窗口相比，新回复或低语没有统计上显着的增加。这意味着虽然这些通知可能会吸引用户阅读流行的低语，但新的低语或回复并没有显着增加。

6. Whisper的内容删除

匿名促进言论自由，但也不可避免地促进了负面的内容和行为[21,35]。与其他匿名社区一样，Whisper也面临着如何处理网络中负面内容（例如，裸露，色情或淫秽）的同样挑战。除了基于众包的用户报告机制外，Whisper还有专门的员工来检测低语[16]。我们的基本测量结果（§3.2）也表明这对系统有重大影响，因为我们观察到Whisper网站在实验的3个月内已经删除了大量的低语（> 170万）。 Whisper删除内容的比例（18%）远高于Twitter（<4%）等传统社交网络[1,30]。

在本节中，我们将详细介绍Whisper中的内容删除。首先，我们分析删除的低语的内容，以推断删除背后的原因。其次，我们分析删除的低语的生命周期，以了解低语被删除的速度有多快。第三，我们专注于删除的低语的作者，并将他们的行为与常态进行比较。

在我们开始之前，我们注意到虽然用户可以删除他们自己的低语，但我们认为服务器端内容审核是我们数据中大多数缺失的低语主要原因。直觉上，重新考虑并稍后删除自己低语的用户可能会在相对较短的时间内完成此操作。相比之下，我们的“已删除”数据集来自我们对每周运行一次的回复的后续抓取。实际上，由于最新流上的主爬虫每30分钟运行一次，我们预计大多数自我删除的低语甚至不会显示在我们的核心数据集中。

删除低语的内容分析 为了探究删除背后的原因，我们分析了删除的低语的内容。由于低语通常很短，自然语言处理（NLP）工具效果不佳（我们通过实验证实）。因此，我们采用基于关键字的方法：我们从所有低语中提取关键字，并检查哪些关键字与删除的低语相关联。首先，在处理之前，我们从关键字列表中排除常见的stopwords4。同样为了避免统计异常值，我们排除出现低于0.05%低声的低频词。然后，对于每个关键字，我们使用此关键字计算删除率，作为使用此关键字删除的低语的数量。我们按删除率对关键字进行排名，并检查顶部和底部关键字。 我们在我们的数据集中对所有900万个原始（不包括回复）的低语进行了此分析，其中1.7M后来被删除。这产生了按删除率排名的2324个关键字。我们在表4中列出了顶部和底部的50个关键字，并将它们手动分类为主题类别。毫不奇怪，许多删除的低语违反了Whisper关于色情信息和裸露的用户政策。相反，与个人表达，宗教和政治相关的主题最不可能被删除。

Topic	Top 50 Keywords Most Related to Deleted Whispers
Sexting (36)	sext, wood, naughty, kinky, sexting, bj, threesome, dirty, role, fwb, panties, vibrator, bi, inches, lesbians, hookup, hairy, nipples, freaky, boobs, fantasy, fantasies, dare, trade, oral, takers, sugar, strings, experiment, curious, daddy, eaten, tease, entertain, athletic
Selfie (7)	rate, selfie, selfies, send, inbox, sends, pic
Chat (7)	f, dm, pm, chat, ladys, message, m
Topic	**Top 50 Keywords Least Related to Deleted Whispers**
Emotion (17)	panic, emotions, argument, meds, hardest, fear, tears, sober, frozen, argue, failure, unfortunately, understands, anxiety, understood, aware, strength
Religion (10)	beliefs, path, faith, christians, atheist, bible, create, religion, praying, helped
Entertain. (8)	episode, series, season, anime, books, knowledge, restaurant, character
Life story (6)	memories, moments, escape, raised, thank, thanks
Work (5)	interview, ability, genius, research, process
Politics (1)	government
Others (3)	exactly, beginning, example

表4：与Whisper删除相关使用最多和使用最少的50个关键字的话题。

删除延迟 接下来我们分析一下低语的删除延迟，即在删除之前低语在系统中停留多长时间？回想一下，我们的回复爬虫每周工作一次，因此可以检测每周一次的删除低语。如图19所示，大多数（70%）删除的低语在发布后一周内被“删除”。一小部分（2%）的低语在删除前停留了一个多月。由于大多数低语在一周后失去了用户注意力（图5），我们认为这些删除不是众包标记的结果，而是由Whisper内容管理员删除。

为了获得更精确的低语删除视图，我们在一小组低语上进行一段频繁的爬行。 2014年4月14日，我们从最新的低语流中选择了200K新的低语，并在7天的时间内每3小时检查一次（重新抓取）这些低语。在我们的监测期间（一周），在200K低语中删除了32,153个低语。这些低语的寿命（每小时）更精确的分布如图20所示。我们发现低语删除的高峰在发布后3到9小时之间，绝大多数删除发生在发布的24小时内。这表明Whisper中的审核系统可以快速标记并删除令人反感的低语。但是，目前还不清楚这种反应速度是否足够，因为用户页面浏览的重点是最近的低语，3小时后的审核可能为时已晚，无法影响大多数用户看到的内容。

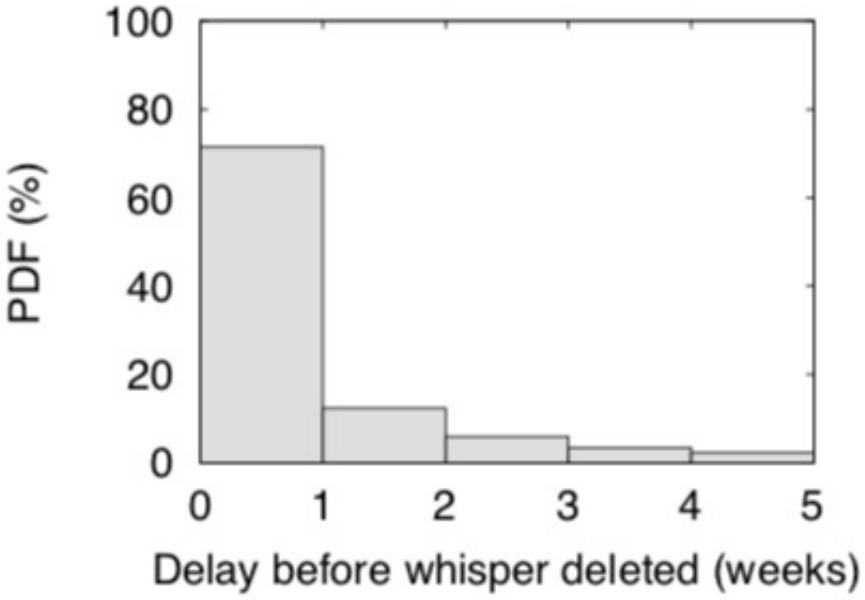


图19：删除速度（粗略统计）

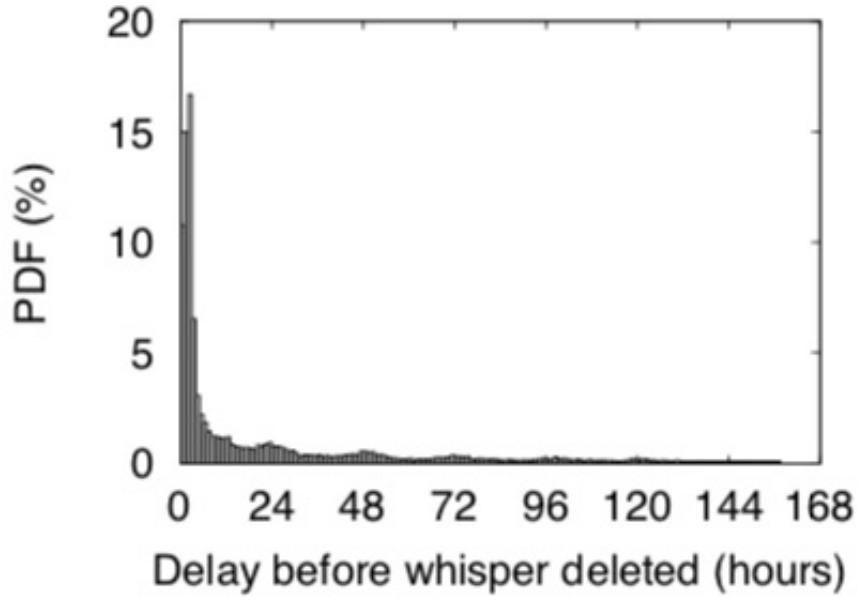


图20：删除速度（精确统计）

被删除的低语作者的特征 最后，我们仔细研究被删除的低语的作者，以检查可疑行为的迹象。在我们的数据集中有263K名用户（25.4%）至少删除了一个低语。删除的低语的分布在这些用户中高度倾斜：24%的用户负责所有删除的低语的80%。最糟糕的罪犯是一个在我们研究的时间段内删除了1230个低语的用户，而大约一半的用户只有一个删除（图21）。

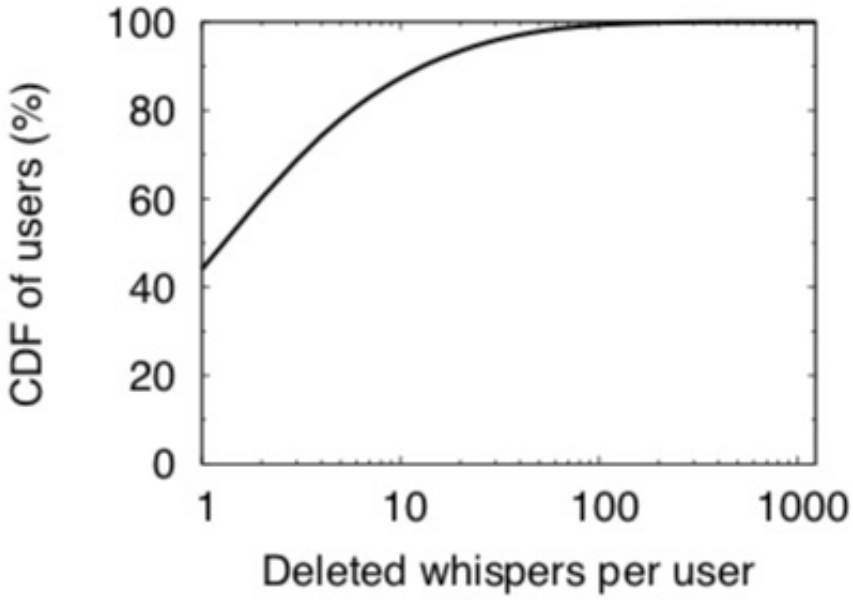


图21：每个用户删除的低语数

我们观察到一系列删除的低语中有重复低语的轶事。我们发现经常重新发布的重复低语很可能被删除。在我们的263K用户中，至少有1个被删除的低语，我们发现有25K用户发布了复杂的低语。在图22中，我们绘制了每个用户重复的低语数量与删除的低语数量。我们观察到围绕y = x直线的用户清晰聚类。这表明当用户发布许多重复的低语时，删除大多数或所有重复的低语的可能性更高。

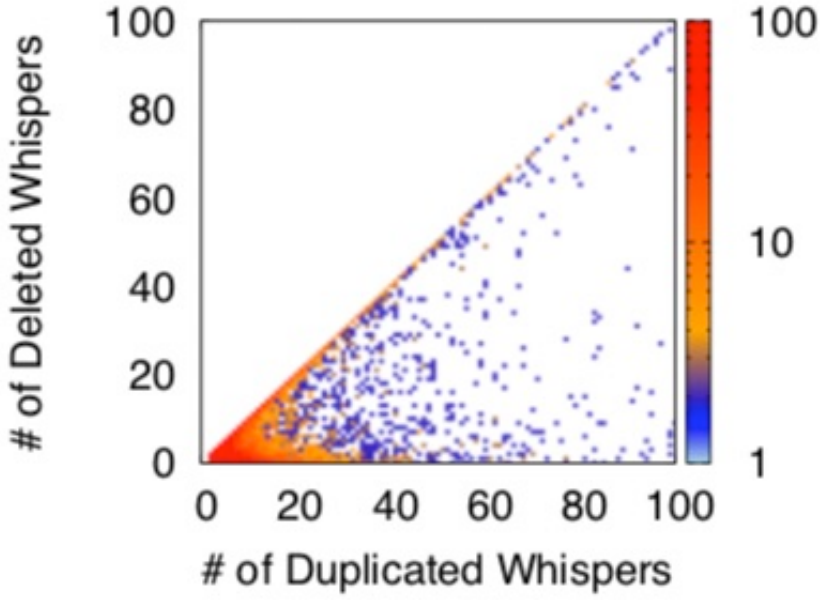


图22:复杂度 vs. 删除低语数

我们还观察到删除的低语的作者比普通用户更频繁地改变他们的昵称。图23显示了每个用户使用的昵称总数的分布。我们根据用户的删除次数对用户进行分类，并且还包括具有0次删除的用户基线。我们发现没有删除的用户很少会改变他们的昵称（如果有的话），但对于有许多删除的低语的用户来说，昵称更改的发生频率更高。我们推测用户可能会更改其昵称以避免被标记或列入黑名单。由于用户在使用应用程序时无法看到自己的GUID，因此他们可能会假设系统仅使用其昵称来识别它们。

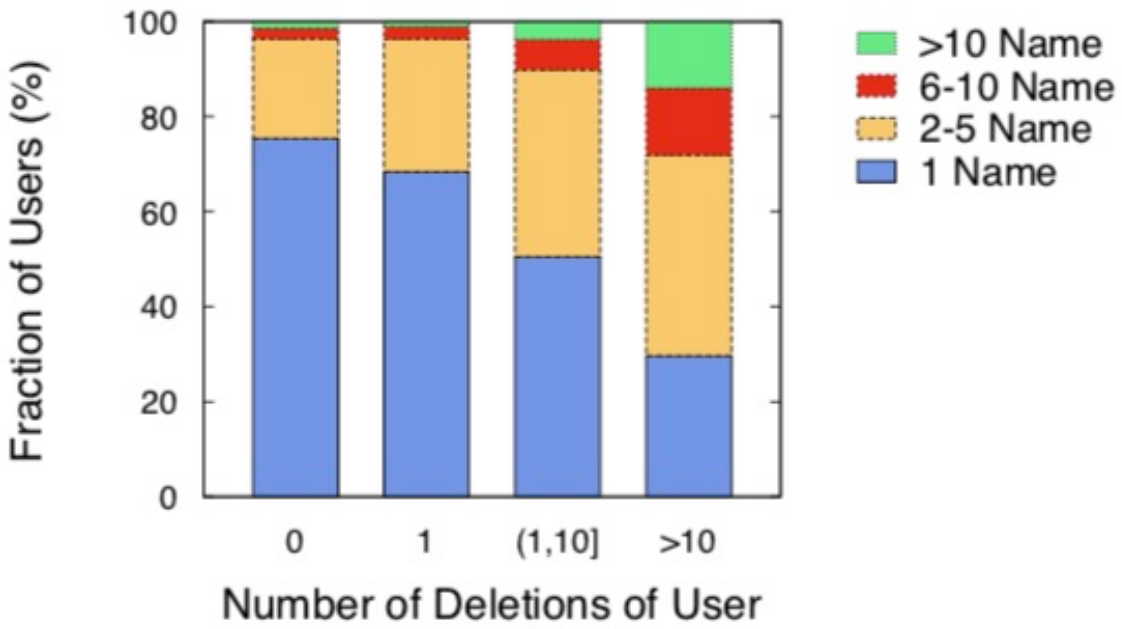


图23：用户删除的低语的数量vs. 昵称的数量。

7. Whisper用户实验

在我们的Whisper研究的最后一部分中，我们仔细研究了一个漏洞，该漏洞暴露了系统中Whisper作者的详细位置。实际上，这种攻击允许一个Whisper用户通过编写查询Whisper服务器的简单脚本，利用他们编写的低语准确跟踪（或潜在跟踪）另一个Whisper用户。此攻击表明了移动应用程序中用户隐私的固有风险，即使对于将用户匿名作为核心目标的应用程序也是如此。请注意，我们亲自会见了Whisper团队并告知他们这次攻击。他们支持这项工作，并已采取措施消除此漏洞。

在本节中，我们将介绍此位置跟踪攻击的详细信息。攻击利用了Whisper的“附近”功能，该功能返回附近张贴的低语列表，为每个低语附加“距离”字段。攻击从不同的有利位置产生了大量“附近”查询，并使用统计分析来反向设计低语作者的位置。我们通过实际实验验证了这种攻击的有效性。

7.1 精确定位

我们首先描述攻击的高级别：当用户（即受害者）发布新的耳语时，他将其位置暴露给Whisper服务器。附近区域的攻击者可以查询附近的列表以获得与耳语作者的“距离”。方法很简单：攻击者可以移动到不同的（附近）位置，并查询附近列表中与受害者的距离。使用多个距离测量，攻击者可以对每个作者的位置进行三角测量。 Whisper不会在查询中对位置进行身份验证使得这一过程变得更加容易，攻击者可以在舒适的起居室内发出来自不同位置的大量距离查询。

通过更多的努力，攻击者甚至可以通过在每次发出耳语时对他的位置进行三角测量来跟踪受害者随时间的移动。在实践中，这意味着攻击者可以在物理上去追踪受害者。虽然有效误差大约为0.2英里（详见下文），但推断受害者对特定兴趣点的移动绰绰有余。考虑到大多数Whisper用户是年轻人或青少年[4]，这种攻击可能导致严重后果。

距离粒度和误差。 实施这种攻击是非常重要的。 Whisper的设计团队一直意识到用户的位置跟踪风险，并在当前系统中构建了基本的防御机制。首先，它们对每个耳语应用距离偏移，因此存储在其服务器上的位置总是偏离实际作者位置一定距离。其次，附近函数返回的距离字段是粗粒度的整数值（以英里为单位）。这是Whisper最近在2014年2月所做的更改，之前附近的函数返回了十进制值的距离。第三，Whisper服务器在每个查询的答案中添加一个随机错误，即当我们从同一位置重复查询附近列表时，每个查询返回相同耳语的不同距离。具体的错误机制未知。

攻击细节 为了准确地确定用户位置，我们的方法是从不同的有利位置广泛测量“距离”，并使用大规模统计来推断用户的位置。具体来说，我们的攻击利用了Whisper的一个关键属性：服务器允许任何人以任意自我报告的GPS值作为输入查询附近的列表，并且不对这些查询施加速率限制。这有效地帮助我们克服返回距离的限制（即随机错误，粗粒度）。首先，我们可以通过从同一观察位置获取多个查询的平均距离来减少或消除每个查询噪声。其次，即使绝对距离仍然不准确，我们也可以根据不同位置的测量值来估计受害者的方向。然后，通过距离和方向，攻击者可以从更靠近受害者的位置重复测量，从而迭代地推断受害者的真实位置。

我们使用一个简单的例子来说明它是如何工作的。假设用户A（攻击者）在附近的列表中找到用户B（受害者）的私语，A想要查明B的位置：

1. 查询""附近""列表以获得其与受害者B的当前距离（d）（使用多个查询中的平均值）。
2. 为了估计方向，A需要额外的观察点。 我们选择8个点{A1，A2，... A8}均匀分布在以A为中心的圆上，半径为d（图24）。 从每个点，A查询“附近”列表以测量其与受害者的距离{d1，d2，...，d8}。 假设X是圆上的一个点，那么如果A到X是受害者的正确方向，则目标函数[29]达到最小值。
3. 然后攻击者使用AX和d移动到下一个位置，并重复步骤1和2.如果d <Thre1，则算法终止，或者两个连续轮次的不同距离d<Thre2。

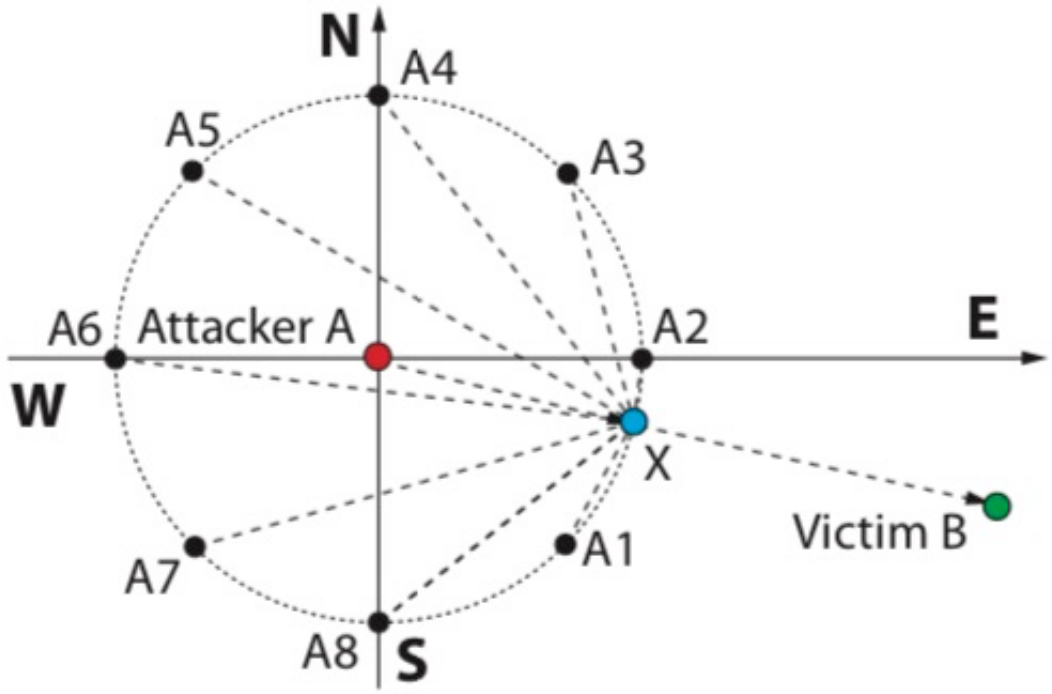


图24：估计受害者的距离和方向。

在实践中，攻击者可以使用伪造的GPS值编写所有查询脚本，从而不需要进行物理移动。

距离误差校正 最后，我们介绍了最后一步，它使用物理测量来校准并为位置数据添加一个额外的“校正”因子。

我们首先在预定义的物理位置L（在UCSB校园）发布目标耳语。然后我们使用来自一组观测点的附近列表测量到L的距离，每个观测点具有已知的地面真实距离L。地面真实距离范围覆盖1到25英里（以5英里为增量）并再次从0.1到0.9英里（以0.1英里为增量）。在每个增量处，我们使用8个观察点（如上所述）并使用每个观察点100次查询附近的列表。图25和图26绘制了地面实况距离与测量距离（每个位置25,50和100个请求）。对于超过1英里的距离，我们发现我们的估计低估了与受害者的真实物理距离。在1英里范围内，它显然高估了。真实距离和测量距离之间的映射用作生成我们的“校正因子”的指南，该校正因子应用于最终估计。

7.2 Experimental Validation of the Attack

单目标实验 我们首先在UCSB校园的预定位置张贴耳语作为目标（受害者）。 然后我们从受害者的距离为1,5,10和20英里开始运行攻击算法。 我们的算法采用每个位置超过50个查询的平均距离，并且当连续轮次的估计距离<0.1英里或估计距离<0.5英里（基于图26）时终止。 我们重复每个实验10次，并在有和没有我们的距离误差校正因子的情况下测试性能。 结果显示在图27和图28中。

我们做了两个关键的观察。首先，算法非常准确。最终误差距离，即从估计的受害位置到地面实际位置的距离仅为0.1到0.2英里。在半径为0.2英里的情况下，攻击者已经可以有效地识别用户的重要兴趣点（例如，家庭，工作，购物中心），并使用移动性痕迹重建受害者的日常生活[3]。其次，结果表明，距离误差校正显着提高了算法的准确性，并减少了确定受害者位置所需的迭代次数。

地理位置不同的目标 为了确保我们的结果没有偏差并且特定于单个位置，我们应用从局部测量计算的校正因子（图25和图26）来在不同城市进行攻击。更具体地说，我们在圣巴巴拉和西雅图华盛顿，丹佛科罗拉多，纽约，纽约和爱丁堡苏格兰发布目标耳语。所有的耳语都是通过带有伪造GPS坐标的Android手机发布的。然后我们运行带有距离误差校正的算法。我们发现最终误差距离始终小于0.2英里，并且我们的校正因子可以推广以提高估计精度，而不考虑地理区域。

7.3 对策

仅通过向系统添加更多噪声不能减轻这种类型的统计攻击。 攻击者总是可以应用日益复杂的统计和数据挖掘工具来消除噪音并确定耳语的真实位置。 相反，关键是限制用户访问广泛的距离测量。 这意味着对查询的更多约束（例如，速率限制）到附近的列表。 例如，一种方法是强制执行每设备速率限制。 另一种方法是通过依赖客户端硬件（困难）或通过检测潜在攻击者的“不切实际”运动模式来检测虚假GPS值。 最后，最终的防御是完全删除“距离”字段。 虽然Whisper工程团队已经解决了这个问题，但我们并没有意识到他们采取的具体步骤。

8. 相关工作

在线社交网络 在过去几年中，研究人员对在线社交网络（OSN）进行了测量研究，包括Facebook [36,39]，Twitter [8,25]，Pinterest [12]和Tumblr [9]。今天的OSN存储了大量关于用户的敏感数据（例如，个人资料，朋友信息，活动痕迹），所有这些都会带来潜在的隐私风险。已经提出了各种技术来危害用户匿名并从社交网络数据推断用户的敏感信息[5,26,27,44]。我们的研究重点是匿名社交网络，它以消除持久身份和社交链接为代价来优先考虑用户隐私。

匿名在线社区 匿名在线服务允许用户发布内容和进行通信，而不会泄露他们的真实身份。研究人员研究了各种匿名平台，包括匿名论坛[32]，讨论板[6,23]和问答网站[21]。大多数早期的作品研究用户社区，重点关注内容和情感分析。最近，出现了匿名社交网络，特别是在移动平台上。最近的一项工作[31]对SnapChat进行了用户调查，以了解他们如何使用匿名社交应用。相比之下，我们的研究是第一个定量研究匿名Whisper网络中的用户交互，用户参与和安全隐患的研究。

设备本地化 我们用于本地化Whisper用户的攻击算法受到用于无线（移动）网络中的设备本地化的现有技术的启发[15,20,43]。 我们处理Whisper服务器注入的随机错误的方法与现有技术不同。此外，我们的贡献更多的是识别和验证安全漏洞，而不是本地化算法本身。

9. 总结和展望

匿名，仅限移动的消息传递应用程序（如Whisper）标志着从传统社交网络转向隐私意识的通信工具。据我们所知，我们的研究是第一个针对Whisper网络的社交互动，用户参与，内容审核和隐私风险的大型数据驱动研究。我们表明，如果没有强大的用户身份或持久的社交链接，用户就会与随机的陌生人交互，而不是定义一组朋友，从而导致长期用户参与的弱关系和挑战。我们表明，即使在匿名消息传递应用程序中，对用户隐私的重大攻击也是非常可行的。我们相信，这种通信工具中隐私权的转变仍然存在，我们对Whisper的研究中的见解为在此领域开发下一代系统的开发人员提供了价值。

Whisper不仅是一种社交沟通工具，也是一种共享匿名内容的网络。对Whisper中的主题和情感进行分析和建模将是未来工作的有趣主题。例如，用户是否以及如何围绕“主题”或“主题”建立社区？匿名帖子和对话如何影响用户情绪和情绪？ Whisper上的用户行为与现有内容网络（如Digg和Quora）的用户行为相比如何？

致谢

我们要感谢我们的牧羊人Alan Mislove和匿名审稿人的评论。 该项目部分得到了美国国家科学基金会资助的IIS-1321083，CNS-1224100，IIS-0916307，DARPA GRAPHS计划（BAA-12-01）以及国家部门的支持。 本材料中表达的任何观点，发现，结论或建议均为作者的观点，不一定反映任何资助机构的观点。

10. 参考文献

[1] ALMUHIMEDI, H., WILSON, S., LIU, B., SADEH, N., AND ACQUISTI, A. Tweets are forever: a large-scale quantitative analysis of deleted tweets. In Proc. of CSCW (2013).

[2] ANDREESEN, M. Public tweets. Twitter, March 2014. [3] ASHBROOK, D., AND STARNER, T. Using gps to learn

significant locations and predict movement across multiple

users. Personal Ubiquitous Comput. 7, 5 (2003), 275–286. [4] ASSOCIATED PRESS. Whispers, secrets and lies? anonymity

apps rise. USA Today, March 2014. [5] BACKSTROM, L., DWORK, C., AND KLEINBERG, J.

Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In Proc. of WWW (2007).

[6] BERNSTEIN, M. S., MONROY-HERNÁNDEZ, A., HARRY, D., ANDRÉ, P., PANOVICH, K., AND VARGAS, G. G. 4chan and/b: An analysis of anonymity and ephemerality in a large online community. In Proc. of ICWSM (2011).

[7] BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R., AND LEFEBVRE, E. Fast unfolding of communities in large networks. JSTAT 2008, 10 (2008).

[8] CHA, M., HADDADI, H., BENVENUTO, F., AND GUMMADI, K. Measuring User Influence in Twitter: The Million Follower Fallacy. In Proc. of ICWSM (2010).

[9] CHANG, Y., TANG, L., INAGAKI, Y., AND LIU, Y. What is tumblr: A statistical overview and comparison. CoRR abs/1403.5206 (2014).

[10] CLAUSET, A., SHALIZI, C. R., AND NEWMAN, M. E. Power-law distributions in empirical data. SIAM review 51, 4 (2009), 661–703.

[11] GARCIA, D., MAVRODIEV, P., AND SCHWEITZER, F. Social resilience in online communities: The autopsy of friendster. In Proc. of COSN (2013).

[12] GILBERT, E., BAKHSHI, S., CHANG, S., AND TERVEEN, L. “i need to try this!”: A statistical overview of pinterest. In Proc. of CHI (2013).

[13] GILBERT, E., AND KARAHALIOS, K. Predicting tie strength with social media. In Proc. of CHI (2009).

[14] GONG, N. Z., XU, W., HUANG, L., MITTAL, P., STEFANOV, E., SEKAR, V., AND SONG, D. Evolution of social-attribute networks: measurements, modeling, and implications using google+. In Proc. of IMC (2012).

[15] GONZALEZ, M. A., GOMEZ, J., LOPEZ-GUERRERO, M., RANGEL, V., AND OCA, M. M. GUIDE-gradient: A guiding algorithm for mobile nodes in wlan and ad-hoc networks. Wirel. Pers. Commun. 57, 4 (2011).

[16] GROVE, J. V. Secrets and lies: Whisper and the return of the anonymous app. CNet News, January 2014.

[17] GUO, L., TAN, E., CHEN, S., ZHANG, X., AND ZHAO, Y. E. Analyzing patterns of user content generation in online social networks. In Proc. of KDD (2009).

[18] GUYON, I., AND ELISSEEFF, A. An introduction to variable and feature selection. JMLR 3 (2003), 1157–1182.

[19] HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., AND WITTEN, I. H. The weka data mining software: an update. SIGKDD Explor. Newsl. 11, 1 (2009).

[20] HAN, D., ANDERSEN, D. G., KAMINSKY, M., PAPAGIANNAKI, K., AND SESHAN, S. Access point localization using local signal strength gradient. In Proc. of PAM (2009).

[21] HOSSEINMARDI, H., HAN, R., LV, Q., MISHRA, S., AND GHASEMIANLANGROODI, A. Analyzing negative user behavior in a semi-anonymous social network. CoRR abs/1404.3839 (2014).

[22] JONES, J. J., SETTLE, J. E., BOND, R. M., FARISS, C. J., MARLOW, C., AND FOWLER, J. H. Inferring tie strength from online directed behavior. PLoS ONE 8, 1 (2013), e52168.

[23] KNUTTILA, L. User unknown: 4chan, anonymity and contingency. First Monday 16, 10 (2011).

[24] KWAK, H., CHOI, Y., EOM, Y.-H., JEONG, H., AND MOON, S. Mining communities in networks: a solution for consistency and its evaluation. In Proc. of IMC (2009).

[25] KWAK, H., LEE, C., PARK, H., AND MOON, S. What is Twitter, a social network or a news media? In Proc. of WWW (2010).

[26] MISLOVE, A., VISWANATH, B., GUMMADI, K. P., AND DRUSCHEL, P. You are who you know: inferring user profiles in online social networks. In Proc. of WSDM (2010).

[27] NARAYANAN, A., AND SHMATIKOV, V. Robust de-anonymization of large sparse datasets. In Proc. of IEEE S&P (2008).

[28] NEWMAN, M. E. Modularity and community structure in networks. PNAS 103, 23 (2006), 8577–8582.

[29] NEWMAN, M. E. J. Assortative mixing in networks. Physical Review Letters 89, 20 (2002), 208701.

[30] PETROVIC, S., OSBORNE, M., AND LAVRENKO, V. I wish i didn't say that! analyzing and predicting deleted messages in twitter. CoRR abs/1305.3107 (2013).

[31] ROESNER, F., GILL, B. T., AND KOHNO, T. Sex, lies, or kittens? investigating the use of snapchat's self-destructing messages. In Proc. of FC (2014).

[32] SCHOENEBECK, S. Y. The secret life of online moms: Anonymity and disinhibition on youbemom.com. In Proc. of ICWSM (2013).

[33] STRAPPARAVA, C., AND VALITUTTI, A. Wordnet affect: an affective extension of wordnet. In Proc. of LREC (2004).

[34] STUTZMAN, F., GROSS, R., AND ACQUISTI, A. Silent listeners: The evolution of privacy and disclosure on facebook. Journal of Privacy and Confidentiality 4, 2 (2013).

[35] SULER, J., AND PHILLIPS, W. L. The bad boys of cyberspace: Deviant behavior in a multimedia chat community. Cyberpsy., Behavior, and Soc. Networking 1, 3 (1998), 275–294.

[36] UGANDER, J., KARRER, B., BACKSTROM, L., AND MARLOW, C. The anatomy of the facebook social graph. CoRR abs/1111.4503 (2011).

[37] WAKITA, K., AND TSURUMI, T. Finding community structure in mega-scale social networks: [extended abstract]. In Proc. of WWW (2007).

[38] WATTS, D. J., AND STROGATZ, S. Collective dynamics of 'small-world' networks. Nature, 393 (1998), 440–442. [39] WILSON, C., BOE, B., SALA, A., PUTTASWAMY, K., AND

ZHAO, B. User Interactions in Social Networks and Their

Implications. In Proc. of EuroSys (2009). [40] WORTHAM, J. New social app has juicy posts, all

anonymous. NY Times, March 2014. [41] WORTHAM, J. Whatsapp deal bets on a few fewer ‘friends’.

NY Times, February 2014. [42] XU, T., CHEN, Y., JIAO, L., ZHAO, B. Y., HUI, P., AND

FU, X. Scaling microblogging services with divergent traffic

demands. In Proc. of Middleware (2011). [43] ZHANG, Z., ZHOU, X., ZHANG, W., ZHANG, Y., WANG,

G., ZHAO, B. Y., AND ZHENG, H. I am the antenna: Accurate outdoor AP location using smartphones. In Proc. of MobiCom (2011).

[44] ZHELEVA, E., AND GETOOR, L. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In Proc. of WWW (2009).