# 抵制对社交图数据的重新识别挖掘

**摘要** 各种敏感的个人信息成为社交网络的隐私问题。然而，攻击者可利用社交图的特征来重新识别社交网络的目标实体。在本文中，首先分析一种新的基于桶攻击的攻击模型，根据它们的图结构特征重新识别社交网络中的社交个体。本文对于基于桶的攻击，提出了一种新的k-匿名方案。通过这种方案，社交个体完全受到k-匿名保护。实验说明了该方案的有效性。匿名网络的效用利用顶点度和中介中心性的结果来证明。

**关键词 社交网络 . k-匿名 . 大图 . 隐私 . 图的挖掘**

## 介绍

随着云计算的快速发展，社交网络数据经常发布给第三方用户。当外包社交网络数据时，它对保护隐私构成了巨大挑战。在传统的隐私保护中，数据匿名化已经被广泛采用，其涉及隐藏身份或相关的敏感数据。

图1显示了向第三方用户发布社交网络的示例。在该图中，通过在发布社交网络数据之前移除名称等标签来匿名化社交网络中的节点的名称。但是，潜在的攻击者可能仍然通过研究目标节点的拓扑结构来对发布的数据进行去匿名化处理。例如，攻击者可以根据邻居图定位与Bob对应的节点，从而知道他的工资是7000美元。在实际应用中，大量的社交节点可以被去匿名化。例如，Google+中86.9％-95.5％的用户是可去匿名化的，这意味着基于结构的去匿名化在实践中是强大的[7]。

图在很多应用中被广泛使用[3,4,22]。社交网络通常可以被定义为有向或无向图[13]。它的拓扑结构由实体（点）及其关系（边）组成，其中点是社交个体，边是社交个体之间的关系[2]。匿名社交网络的一种简单方法是隐藏网络中每个社交角色的标签。但是不能接受的是，直接外包给云时，这些关系常常暴露出来。诸如k-匿名方法的多种匿名方法已经被提出，周等人[21]定义了一个k-匿名社交网络，攻击者在知道任何目标的1-邻居图的情况下不能以高于1 / k的置信度重新识别目标。其基本思想是通过添加噪声边，使任何顶点的1-邻居图同构至少k-1个其他顶点1-邻居图。给定k个同构1-邻居图，每个人都有1 / k的概率被选为目标。

图2显示了三种基于图结构的攻击。例如，攻击者可以根据Bob的单跳邻居图（图2a）从匿名社交网络（图2b）重新识别Bob（目标）。这种攻击被称为1-邻居攻击，由Zhou等人提出[21]。图2c展示了一种2-匿名机制来保护隐私免遭1-邻居攻击。最近，在文献[9,17]中提出了1 \*-邻居和加权1-邻居攻击。如图2d所示，他们假设攻击者不仅知道目标的1-邻居图，而且还知道所有邻居的准确度。在这种情况下，所有顶点的1 \*-邻居的k-匿名可能会导致原始图形发生剧烈变化。例如，当Bob通过在原始图中添加三条边来保护时，由于它们具有1 \*-邻居图的唯一标志，所以还有两个顶点面临1 \*-邻居的攻击。由于累积效应，当针对图的所有顶点实现2-匿名抵制1\*-邻居时，最终图将变为完全连通图。如图1将社交网络数据外包给第三方用户。潜在的攻击者可能通过挖掘其子图结构特征结果来重新识别目标，由于它们在原始网络上的过度修改而降低了太多源图的实用性。

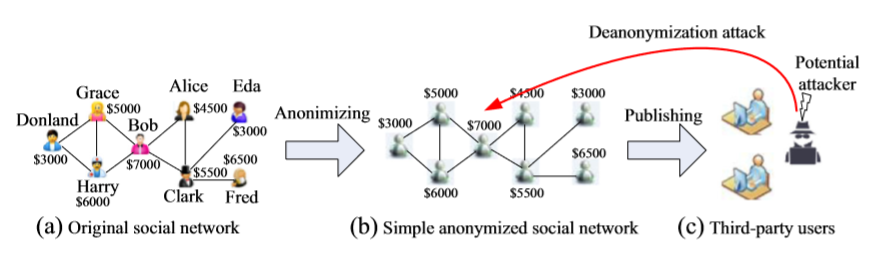


图 1：将社交网络数据外包给第三方用户。潜在攻击者可能通过挖掘其子图结构特征来重新识别目标

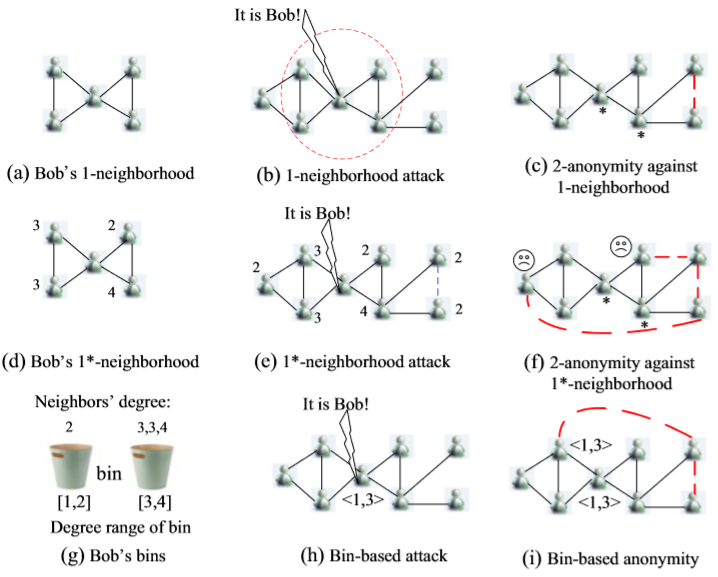


图 2：重新识别攻击。a-c说明1-邻居攻击和保护[21]，d-f显示了1 \*-邻居攻击和保护[9,17]。g-i是本文讨论的基于桶的攻击。桶作为一个根据度来容纳其邻居的容器。请注意，原始社交网络如图1a所示

我们确定了一种新的基于图结构的攻击，即基于桶的攻击，它可以应用于更实际的环境。先前的攻击，如1-邻居攻击和1 \*-邻居攻击需要准确地知道节点的邻居及其连接的度。但是知道邻居的大概度的数量会更容易。以图1a为例，鲍勃有四个邻居，他们的度分别是2,3,3和4。假设Bob有两个桶，度数分别为[1,2]和[3,4]。对于基于桶的攻击，攻击者只需知道第一个桶包含一个邻居（度为2），另一个桶包含三个邻居（度为3，3和4）正如图二所示的一样。根据桶的大小，攻击者可以从简单的匿名图中对Bob进行去匿名处理，因为图大小（即<1,3>）的特征在图中是可区分的（图2h）。

在本文中，我们确定了一个基于图结构的攻击和相应的k-匿名方案。与以前的工作不同，我们考虑了两个更实际的事实。首先，我们处理的是攻击者只拥有关于社交网络目标节点的粗略信息的去匿名攻击，而不是要求攻击者知道精确的邻居信息，如1 \*-邻居图，甚至边上的精确权重[9]。其次，更现实的是，除了社交网络中的关键实体之外，并非所有实体都需要完整的k-匿名保护。关键实体可以由数据提供者等相关方指定。本文主要针对给定关键实体列表的k-匿名方案，而不是确定关键实体的方法。K-匿名保护相当严格，成本特别高。完全k-匿名可能导致完全连接图并失去它的许多效用。因此，本文提出了一个关键实体的可行k-匿名保护机制，社交网络的其他角色可以通过其他方案来保护，如L-diversity[11]。

本文的其余部分安排如下：第2节介绍相关的工作。 所提出的方案在第3节中介绍。实验结果在第4节中提供。第5节总结了本文。

## 相关作品

在本节中，相关工作分别从两个方面进行评述：攻击机制和保护隐私的社交网络。

## 基于图结构的社交网络攻击

在外包社交网络数据之前，共同做法是匿名化，即除去名称，社会安全号码，邮寄邮件地址等标识符，但保留网络结构。图结构是社交网络的一个重要特征[6]。当攻击者拥有图结构的背景知识时，隐私保护可能仍然受到威胁并存在风险。

在基于邻域的攻击中，Zhou和 Pei首先提出了1-邻居攻击[21]。1-邻居图[21]被定义为，其中表示一组节点的集合，表示一组边集合 。攻击者通过比较1-邻居图的特征来重新识别目标节点。Wang等人提出了对1-邻居攻击的修正[17]。他们假设攻击者具有目标节点的1 \*-邻居图的内容。1 \*-邻居图[17]被定义为，其中是节点u的1-邻居图，而是u邻域的度数的序列。1 \*-邻居攻击意味着除了1-邻域图的结构之外，假设攻击者知道目标的单跳邻居的确切程度顺序。通过这个假设，攻击者可以以高于1 / k的概率从k-匿名社交网络重新识别目标。最近，在[9]中提出了加权1 \*-邻居图的攻击。他们进一步假设攻击者不仅知道目标节点的1 \*-邻居图，而且知道边上的权重。然而，其中一些假设太严格了，这会导致网络修改的成本非常高，甚至可能导致k-匿名失效。

最近提出了一些基于网络拓扑特征的去匿名的方法。算法Seed-and-Grow被提出用于通过从用户的种子集合中识别和扩展来从匿名社交图中识别用户。该算法首先识别种子子图，或者由攻击者做出来，或者由一小群用户串通而泄露，然后根据攻击者已有的用户社交关系知识扩展种子子图[12]。种子根据幂律度分布等社交网络的特点进行选择，然后进行扩展以识别其他用户[1]。从多个匿名网络的角度来看，网络对齐可以通过利用多个网络之间的相似性对社交网络进行去匿名化[8,19]。

## 针对基于图结构的攻击的隐私保护

现有的关于图外包中隐私保护的研究主要集中在基于特定结构特征的匿名化方面。通常通过修改网络结构来实现匿名化，或者通过添加或删除边缘来实现，这通常是指对原图的少量调整。在k-匿名机制中，匿名化的最终目标是确保k-匿名性，其中一个目标与k-1个参与者共享相同或相似的局部结构，以便找到目标的可能性为1 / k。

根据攻击者的先验知识，隐私保护可以采取不同的方案来确保安全性。对于1-邻居图攻击，添加一些边以使1-邻居图同构以抵御攻击并保护隐私信息[21]。此外，为了消除1 \*-邻居攻击，王等人提出了一种启发式的不可区分组匿名化（HIGA）方案[17]。针对1 \*-邻居攻击提出了概率匿名化方法。基于此，最近有一种加权的1 \*-邻居攻击和保护方案[9]。然而，这些方案由于高修改成本和效用损失而不能获得精确的k-匿名保护。针对基于共同朋友识别一对朋友的攻击，提出了一个新的k-匿名问题，它针对社交网络出版物中的共同朋友攻击[14]。他们设计了两个启发式算法来反映该图的效用，另外，他们设计了一个算法来确保k度匿名。

许多研究采用基于边编辑的模型，通过添加或删除边缘，根据隐私要求修改图形以满足某些特性。文献[18]提出了一个k度L-diversity模型，该模型保留了很高的实用性。该效用由平均路径长度来表示，该平均路径长度反映原始图到修改图后的失真。一种新的基于机器学习和优化技术的匿名图[10]被提出用于在非标记简单图中针对节点重新识别的隐私保护。

最近还提出了用于大规模数据集的分布式和可扩展解决方案。X.Zhang等人[20]提出了一种可扩展的两阶段自上而下的专业化方法，使用云上的MapReduce框架匿名化大规模数据集。它为大规模社交网络提供了一个可扩展的解决方案。在[15]中，作者们考虑了网络数据在多个数据持有者之间分配的分布式设置。目标是在不向任何数据持有者透露由其他数据持有者控制的节点之间的链接信息的情况下获得统一网络的匿名化视图。

## 提出k匿名方案

在本节中，我们提出了基于桶的k-匿名方案，它为社交网络的关键实体提供了严格的k-匿名保护。关键实体可以通过使用一些图的特征获得，如pagerank和H-index值[16]，它们也可以由数据提供者等相关方指定。提出的方案可以灵活地实现关键实体的k-匿名，但不关注如何选择它们。它可以应用于需要高强度保护的任何给定的顶点集合。所提出的方案的主要步骤包括选择任意顶点的k-1个候选顶点以进行匿名化;以及编辑该图以实现至少k个顶点以桶为单位保持相邻度相同的特征。因此，基于桶的攻击无法以高于1 / k的置信度重新识别目标。

## 根据桶特征选择候选点

为了实现图的给定顶点的k-匿名性，它需要找到至少k-1个节点来匿名化，这称为候选顶点。其基本思想是根据桶特征与目标节点的相似性排列所有其他顶点。为了不失一般性，我们将顶点u作为匿名目标。在下文中，详细说明用于u的选择候选顶点的过程。

给定原始图和需要k-匿名化的顶点u，u的邻居根据它们的程度分配到不同的箱。u的邻居属于一个箱子组成集合：

 （1）

其中是当前桶的度的范围。然后，目标顶点u的邻居度的特征可以根据这些桶表示为一个向量：

 （2）

其中是桶的总数，（0≤i≤ξ）是第i个桶的尺寸，即= 。此外，我们根据顶点u和v之间的桶矢量定义距离如下：

 （3）

其中γiu，γiv是第i个桶的u和v的大小。根据（3），可以通过对距离进行排序来获得k-1个候选顶点：

 （4）

当获得k-1个候选顶点时，最佳的桶大小将使编辑成本最小化：

 （5）

对于第i个桶，其目标大小根据（5）获得。那么，桶大小的目标矢量可以表示为：

 （6）

因此，我们选择给定顶点的k-1个候选顶点，以针对基于桶的攻击进行匿名化处理。该方法为实际应用提供了灵活的方案。数据所有者或提供者可以决定哪些顶点需要k-匿名保护。社交网络的其他角色可以使用诸如L-diversity [11]的方案来保护。在下文中，这些候选顶点的度数朝向桶大小的目标矢量被修改。

## 接近k匿名的方法对抗基于桶的攻击

为了对社交网络中的任何实体进行k匿名隐私保护，匿名图中至少有k个顶点，这些顶点对于攻击者的背景知识方面没有区别。3.1节给出了给定顶点选择k-1个候选顶点的方法以及相应的目标桶大小的矢量。下面介绍如何匿名化桶特征的详细信息。算法1演示了在社交网络中逼近目标顶点u的桶大小向量的过程。输入的参数是桶的数目，这些桶必须是与u的桶大小相同的桶。该算法的输出是匿名图。首先，对顶点u和目标向量之间的距离向量进行排序。然后，选择最小箱来接近目标尺寸。对于给定的桶 ，，如果当前桶大小大于目标桶大小（算法1的第5行），则顶点u减小第个桶大小，否则，增加其自身的桶大小（算法1的第7行）。添加边或删除边的操作可能会干扰匿名顶点。在算法1中，我们使用三个标志，即Lock桶，LockDegree和LockVertex，以避免修改边的连锁结果。当桶被用作候选顶点的匿名桶时，Lock桶被设置为真（算法1的第9行）。之后，桶大小不能再改变。对于匿名桶中的任何顶点，它们的度数在当前桶范围内受到限制，LockDegree设置为true。当一个候选顶点完成接近过程以达到桶大小矢量的目标时，其状态LockVertex被设置为true，以避免被其他匿名化过程（算法1的第14行）修改。

|  |
| --- |
| 算法1 匿名处理点的目标桶大小 |
| Input:,,,  Output:  1 ;  2 ;  3 foreach  do  4 if then  5 ;  6 else if then  7 ;  8 end  9 Lock桶true;  10 foreach  do  11 LockDegree true  12 end  13 end  14 LockVertextrue;  15 return ; |

增加桶大小的方法（算法1的第7行）是找到可行的顶点来创建新的边。但并不是所有的顶点都被允许编辑，包括添加和删除边。我们称一个被添加或删除边的顶点为目标顶点。如果一个目标顶点可以添加一个新的边，其度应该满足以下条件：

如果目标顶点可以连接到候选顶点，其度数必须在范围内，其中，是候选顶点所在的桶的度范围的开始值和结束值。

当在候选顶点和目的顶点之间添加边时，目标顶点的度数将增加1。如果其原始度数属于，则其度数将被覆盖。因此，它提供了1个单位来增加候选顶点的当前桶大小。还有另外两个条件是目标顶点不能被添加边。（1）目标顶点也是匿名节点时（LockVertex为true），当前候选顶点落入其锁定区域（Lock桶为true）。（2）目标顶点不是候选顶点，而是候选顶点的邻居，落入某个候选顶点的锁桶，其度数为。如果添加了一条边，它将修改某个邻居候选顶点的桶大小。

为了减小桶的大小（算法5的第5行），候选顶点可以用一些邻居来删除边。但这些邻居可能会被之前的匿名操作锁定。因此，一些边不允许被删除。如果候选顶点属于目标顶点的锁定桶，则边不能被删除，因为它会更改目标顶点的锁定桶大小。

如果满足以下条件，则不能从目标顶点v删除边缘：（i）的至少一个邻居顶点是匿名顶点；（ii）属于其邻居的锁桶；（iii）的等级等于，其中是锁定仓的开始程度值。（iii）的度为，其中是锁定桶的开始度的值。

如果当前候选顶点属于其邻居的锁定桶中，则无法删除该边。这些约束避免了修改图的链接结果，并且可以实现社交网络的给定实体的k-匿名。

算法1展示了对任何给定顶点u实现k-匿名的处理。算法的时间复杂度为，其中是桶的数量，N是顶点的平均度数。

图2i显示出了对图1a中的Bob相对应的节点实现针对基于桶的匿名的2-匿名的示例结果。选择与Grace对应的节点作为匿名候选节点。首先，Grace和Eda之间增加了一条边。那么，Grace的桶特征是<2,2>，这与Bob的不同。因此Eda和Fred之间又增加了一条边。因此，与Bob和Grace相对应的节点具有相同的桶特征<1,3>，这意味着他们已经实现了基于桶的攻击的2个匿名。

k匿名属性意味着发布的每个人的个人信息的内容不能与其他k-1个其它发布的个人信息区分开。理论上，攻击者将以1 / k概率重新识别目标。

## 评估结果

在本节中，我们首先演示原始图上修改边的数量的变化，以实现完全k-匿名。然后，我们展示效用度量的评估结果，例如不同k匿名的顶点度和中介中心性。这些评估是在两个真实世界的数据集和一个综合数据集上进行的。实验中使用了两个真实世界的数据集（http://snap.stanford.edu/data/index.html）。

|  |  |  |
| --- | --- | --- |
|  | #点 | #边 |
| Facebook | 4039 | 88234 |
| Wiki | 7115 | 103689 |
| Synthetic graph | 10000 | 179407 |

表 1：关于实验数据集的基本统计

## 修改边

在社交网络中，通常有一些关键实体需要更严格的隐私保护。在应用本文提出的方案时，关键实体可由数据所有者或云服务提供商在实际场景中提供，或使用H-index值或paperank值进行挖掘。本文主要关注给定关键实体的k-匿名方案，而不是决定谁是关键实体的方法。为不失一般性，我们在实验中随机选择四组顶点进行k-匿名化。

图3演示了三个数据集上的修改边的数量。这些数据有两种趋势。（i）随着匿名化强度参数k的增加，修改边的数量增加。在k-匿名隐私保护中，攻击者无法以高于1 / k的置信度重新识别目标。更大的k意味着更强大的保护，但也表示更多边的修改。（ii）随着要匿名化的关键实体（分别为N = 100,200,300和400）的数量增加，修改的边的数量一直在增长。这些顶点完全是k-匿名保护的，并且至少有k-1个其他顶点具有相同的桶特征。对于N = 400，最多可以修改400 \* k \* d个顶点，其中d是顶点度。在我们的方案中，最大数量的修改边分别是Facebook，Wiki和合成数据集的2569，5006,1201。比较原始图的边数量，增加率仅为2.9％，4.8％和0.7％。下面显示了这些修改图的效用的进一步评估。

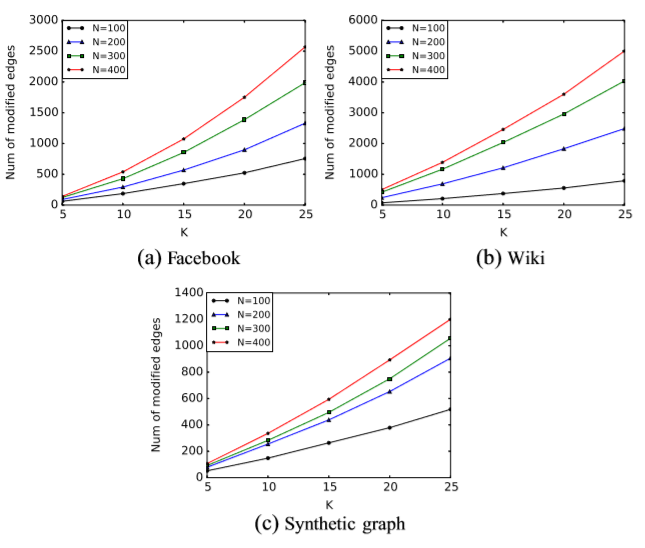


图 3：修改边的数量

## 效用

我们的数据的实用性可衡量匿名过程中的信息丢失和失真。丢失或失真的信息越多，公布的有用数据就越少。我们评估关于修改图的效用的三个指标，即顶点度和中介中心性。对于每个指标，我们不仅显示原始图和修改图的详细分布情况，还显示Kolmogorov-Smirnov（K-S）测试结果以定量检查变化。

采用Kolmogorov-Smirnov（K-S）统计是为了测试修正图与原始图指标分布的区别，两侧K-S检验的零假设是两个观测样本，即修改图和原始图中的指标分布，是从相同的分布得出的。K-S检验的结果是两个值：K-S统计值和p值。K-S统计值表示两个样本分布之间的差异，因此较低的K-S统计值表明更多的相同分布。P值表明拒绝零假设的明显水平。大于0.05的p值表明这两个样本是从95％置信度的相同分布中提取的。在我们的实验中，所有三个指标在不同k水平下的所有p值均大于0.05，这意味着通过K-S测试，修改后图与原始图在这些指标上没有区别。

1. 度的分布

在本节中，facebook，wiki和综合数据集的度分布分别如图4a，b和c所示。在本节中，facebook，wiki和综合数据集的度分布分别如图4a，b和c所示。与原始图相比，不同匿名化程度（k = 5，10，15，20和25）度的分布也会随之改变。图4d显示了K-S统计值。K-S检验证实可以接受相同分布的假设。

最近，一个相关的提议集中在社交图数据的k-匿名性上[9]。在[9]中，匿名社交网络的效用是通过最小节点度（MIN）与平均节点度（AVE）来度量的。为了与以前的工作进行比较，我们还对facebook数据集上的这些指标进行了统计。在原始的facebook图中，MIN值为1。我们提出的方案的MIN值保持为1。但是在[9]中，它从8变为22，k从5变为25。此外，我们还发现在原始图中k值增加的情况下，AVE值在我们的方案中增加地非常缓慢。这些结果说明我们的方案很好地保留了原始图的效用。

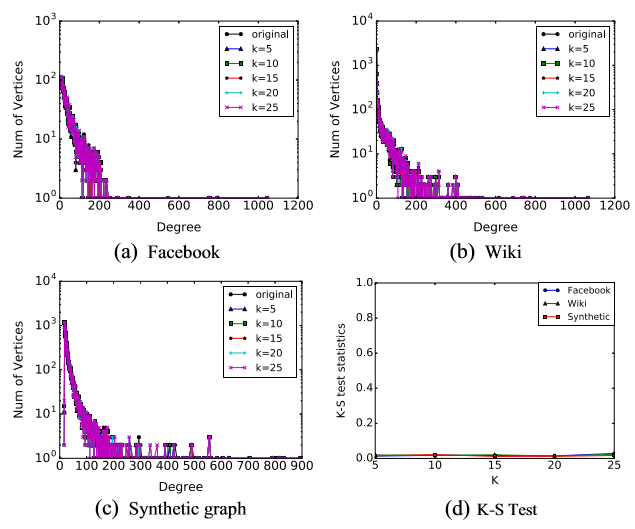


图 4：度的分布及其K-S测试

1. 点的中介中心性

中介中心性分布是图的一个重要特征。对于连通图中的每对点，存在连接两个点的最短路径。点的中介中心性被定义为穿过点的最短路径的数量：

 （7）

其中是图中最短路径的总数，是穿过顶点的最短路径的数量。

图5显示了中介中心性和他们的T-S测试结果的分布。在这些比较子图中，所有匿名参数k的分布非常相似。特别是，在图5中，合成图的分布几乎相同，这由只有0.7％的修改的图3c中的结果证实。

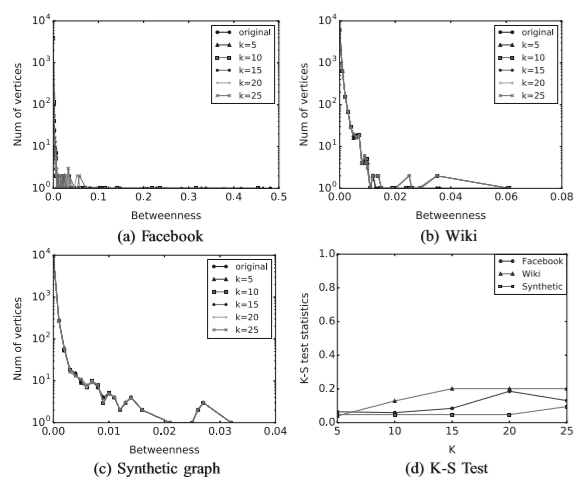


图 5：中介中心性分布及其K-S测试

## 结论

在本文中，我们首先通过放宽背景知识的要求来识别基于桶的攻击。在不了解1-邻近的度的情况下，基于桶的攻击可以重新识别具有1-邻近的度分布的大概特征的目标，本文将其表述为桶。为了抵制这种攻击，我们提出了一个有效的k匿名方案。与概率k匿名保护不同，该方案实现了对关键实体的完全k-匿名保护。实验结果表明，使用所提出的方案的匿名社交网络都具有非常低的边的修改率，并且就顶点度和中介中心性来说，该方法保留了相当大的原始效用。

## 参考文献

[1] Chiasserini, C.F., Garetto, M., Leonardi, E.: Social network de-anonymization under scale-free user relations. IEEE/ACM Trans. Networking 24(6), 3756–3769 (2016)

[2] Gao, J., Liu, P., Kang, X., Zhang, L., Wang, J.: PRS: Parallel relaxation simulation for massive graphs. Comput. J. 59(6), 848–860 (2015)

[3] Gao, J., Song, B., Ke, W., Hu, X.: Balanceali: multiple PPI network alignment with balanced high coverage and consistency. IEEE Trans. Nanobioscience 16(5), 333–340 (2017)

[4] Gao, J., Wang, J., Zhong, P., Wang, H.: On threshold free error detection for industrial wireless sensor networks. IEEE Trans. Ind. Inf. (2017). https://doi.org/10.1109/TII.2017.2785395

[5] http://snap.stanford.edu/data/index.html

[6] Hu, R., Zhu, X., Cheng, D., He, W., Yan, Y., Song, J., Zhang, S.: Graph self-representation method for unsupervised feature selection. Neurocomputing 220, 130–137 (2017)

[7]Ji,S.,Li,W.,Srivatsa,M.,Beyah,R.:Structural data de-anonymization:theory and practice.

IEEE/ACM Trans. Networking 24(6), 3523–3536 (2016)

[8] Koutra, D., Tong, H., Lubensky, D.: Big-align: fast bipartite graph alignment. In: Proceeding of IEEE 13th international conference on data mining (ICDM), pp. 389–398 (2013)

[9] Liu, Q., Wang, G., Li, F., Yang, S., Wu, J.: Preserving privacy with probabilistic indistinguishability in weighted social networks. IIEEE Trans. Parallel Distrib. Syst. 28(5), 1417–1429 (2017)

[10] Maag, M.L., Denoyer, L., Gallinari, P.: Graph anonymization using machine learning. In: IEEE 28th international conference on proceeding of advanced information networking and applications (AINA), pp. 1111–1118 (2014)

[11] Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M.: l-Diversity: privacy beyond kanonymity. In: Proceedings of international conference on data engineering, pp. 1–12 (2006)

[12] Peng, W., Li, F., Zou, X., Wu, J.: A two-stage deanonymization attack against anonymized social networks. IEEE Trans. Comput. 63(2), 290–303 (2014)

[13] Su,Z.,Xu,Q.,Qi,Q.:Bigdata in mobile social networks:aQoE-oriented framework.IEEENetw.

30(1), 52–57 (2016)

[14] Sun, C., Philip, S.Y., Kong, X., Fu, Y.: Privacy preserving social network publication against mutual friendattacks.In:Proceeding of IEEE 13th international conference on data mining,pp.883–890(2013)

[15] Tassa, T., Cohen, D.J.: Anonymization of centralized and distributed social networks by sequential clustering. IEEE Trans. Knowl. Data Eng. 25(2), 311–324 (2013)

[16] Tang, J., Lou, T., Kleinberg, J.: Inferring social ties across heterogenous networks. In: Proceedings of ACM international conference on Web search and data mining, pp. 743–752 (2012)

[17] Wang, G., Liu, Q., Li, F., Yang, S., Wu, J.: Outsourcing privacy-preserving social networks to a cloud. In: Proceedings of Infocom, pp. 2886–2894 (2013)

[18] Yuan,M.,Chen,L.,Philip,S.Y.,Yu,T.:Protecting sensitive labels in social network data anonymization. IEEE Trans. Knowl. Data Eng. 25(3), 633–647 (2013)

[19] Zhang, J., Philip, S.Y.: Multiple anonymized social networks alignment. In: Proceeding of IEEE 13th international conference on data mining (ICDM), pp. 599–608 (2015)

[20] Zhang,X.,Yang,L.T.,Liu,C.,Chen,J.:Ascalable two-phase top-down specialization approach for data anonymization using mapreduce on cloud. IEEE Trans. Parallel Distrib. Syst. 25(2), 363–373 (2014)

[21] Zhou, B., Pei, J.: Preserving privacy in social networks against neighborhood attacks. In: Proceedings of international conference on data engineering, pp. 506–515 (2008)

[22] Zhu, X., Li, X., Zhang, S., Ju, C., Wu, X.: Robust joint graph sparse coding for unsupervised spectral feature selection. IEEE Transactions on Neural Networks and Learning Systems 28(6), 1263–1275 (2017)