许多网络攻击是从传播网络钓鱼网址开始的。当点击这些钓鱼网址时，受害者的私人信息会被泄露给攻击者。目前已经提出了几种机器学习方法来检测钓鱼网站。然而，对于检测具有规避性的钓鱼网址，即通过操纵模式假装为良性的钓鱼网址，仍然没有得到充分的探索。在许多情况下，攻击者i）重复使用准备好的钓鱼网页，因为制作一套全新的网页需要花费非同小可的费用，ii）喜欢不需要私人信息且比其他公司便宜的托管公司，iii）喜欢共享托管以获得成本效益，以及iv）有时使用良性域名、IP地址和URL字符串模式来规避现有的检测方法。受这些行为特征的启发，我们提出了一种基于网络的推理方法，以准确地检测出伪装成合法模式的网络钓鱼网址，即对逃避行为具有鲁棒性。在网络方法中，除非网络中的大多数邻居同时被回避，否则即使在回避之后，一个钓鱼网站仍然会被识别为钓鱼网站。在各种实验测试中，我们的方法始终显示出比最先进的方法更好的检测性能，例如，我们的方法的F-1为0.89，而基于特征的最佳方法为0.84。

通过www.DeepL.com/Translator（免费版）翻译

网络攻击对我们的社会造成了巨大的损害。许多网络攻击是从网络钓鱼开始的。网络钓鱼是为了欺骗人们将他们的敏感信息透露给攻击者。特别是，网络钓鱼的URL被伪装成人们看起来很熟悉的URL。粗心的人就会点击它们，导致他们的私人信息被泄露。因此，许多检测方法已经被开发出来，作为回应，攻击者开始考虑用合法模式进行伪装的规避技术（详情见第3节）[18, 29, 38, 39, 45]。因此，防止使用规避的网络钓鱼攻击是最重要的。已有人提出机器学习方法来检测网络钓鱼。它们可以被分为两种类型：基于内容和基于URL字符串。基于内容的方法下载并分析网页内容[33, 36, 44]。然而，它们需要非微不足道的计算来处理许多网页，并且对基于网络浏览器的攻击很弱（因为我们需要访问他们的网页）。最重要的是，要收集这样的训练数据并不容易。由于所有这些原因，基于内容的方法并不总是首选。基于字符串的方法主要依赖于URL字符串模式分析，因为众所周知，钓鱼网站的字符串模式非常容易区分[1-3, 6, 12, 19, 30, 36, 37, 44, 50]。因此，人们提出了许多检测钓鱼网站的词汇特征（见第2节）。众所周知，这些特征在检测网络钓鱼网址方面是有效的。由于基于字符串的方法在计算上是轻量级的，并能提供较高的准确性，许多研究者倾向于使用这种方法，因为其效率高[44]。一些研究人员依赖于IP地址和域名的黑名单。然而，众所周知，它的准确度很一般。

预测

图1：拟议方法的整体工作流程。在第一步，我们将收集到的URLs分割成单词，并删除无意义的单词，这些单词对应的是具有高频率但不携带有用信息的停止词。第二步，我们构建一个由URLs、域名、IP地址等组成的异质网络。在最后一步，我们运行定制化的信仰。几乎所有现有的基于字符串的方法都没有考虑规避问题[45]。逃避是指攻击者通过操纵钓鱼网站的模式来欺骗防御者的检测方法，从而创造出看似合法的钓鱼网站的技术。在这项工作中，我们考虑了钓鱼网站攻击的几个更关键的模式，以设计一种先进的基于字符串的检测方法，该方法优于现有的方法，并能有力地防止规避。首先，攻击者对成本效率很敏感[15]。在很多情况下，他们（部分）重复使用网络钓鱼的攻击材料，并且喜欢特定的托管公司，因为它们的政策比较宽松（比如不要求提供身份信息），价格也比其他机构相对便宜。当使用私人服务器而不是托管公司时，攻击者更喜欢共享托管，即一台服务器用于多个网络钓鱼攻击活动，也用于多个域名--在我们的数据中，15.8%的IP地址连接到多个域

名。其次，攻击者在良性的服务器、域名、IP地址和/或子串上创建钓鱼网址，以逃避现有的检测方法[15]。 传播方法，使其变得强大。
考虑到所有这些事实，我们设计了一个新颖的自然语言处理统一框架和基于网络的方法来检测钓鱼网站--其整体工作流程见图1。我们将每个URL视为一个句子，并将其分割成子串(字），考虑到URLs的语法和标点符号--URLs在英语中具有明确的语法。之后，我们建立一个由异质实体组成的大网络，如URL、域名、IP地址、权威名称服务器和子串，并执行我们定制的信念传播来检测钓鱼网站（见4.3.1节）。我们注意到，上面列出的相关工作不包括任何基于网络的推理方案。相反，类似的基于网络的推理方法已经被用于其他各个领域[7, 32]。然而，我们的方法与他们不同，它定义了边缘电位，当两个相邻的实体有不同的预测标签时，决定了一个惩罚。
我们的方法可以有效地推断出看似不相关的钓鱼网站实际上是相关的，并且对逃避行为具有鲁棒性。因为我们在这样一个异质实体的网络上进行推断，除非网络中的大多数邻居同时被规避，否则钓鱼网站的规避是不可能成功的（更详细的讨论与定理和证明见第5节），这是我们与现有工作相比的主要贡献。

通过www.DeepL.com/Translator（免费版）翻译

我们抓取了许多可疑的URL，还下载了其他研究人员发布的几个数据集[11, 46]。总的来说，我们有大约12万个虚假的和38万个良性的URLs。我们将我们的方法与最先进的基线方法进行比较，包括图卷积网络（GCN）和基于特征工程的方法。我们的方法在这些方法中显示出最好的检测性能。此外，在额外的规避测试中，我们的方法显示出比其他基线方法更好的F-1得分。由于规避会给攻击者带来非同小可的费用，使其无法访问良性域名、IP地址等，因此我们的稳健检测方法大大增加了攻击者进行规避的经济负担。我们的贡献可以总结为以下几点。

- 我们设计了一种新型的基于网络的推理方法，配备了我们提出的鲁棒的边缘势能分配机制。我们在边缘势能分配的基础上进行的网络推理优于许多基线方法，包括基于特征工程和基于网络的分类器。
- 我们提出的基于网络的方法有一个理论基础，即为什么它对规避具有鲁棒性（见第5节）。
- 我们用自己收集的和从其他工作中下载的大量URL进行了实验。我们的数据涵盖了各种各样的虚假/良性URL模式。
  在下文中，我们首先在第2节中回顾文献，在第3节中描述这项工作的动机。然后，在第4节和第5节中，我们设计了一种新型的基于网络的检测方法，对逃避行为具有鲁棒性，并分析了所提方法的理论鲁棒性。之后，我们在第6节中对有回避和无回避的网络钓鱼网址检测进行了广泛的实验。最后，在第7节和第8节，我们描述了抓取的数据并总结了我们的论文。作为参考，在附录A中，我们介绍了一组广泛用于检测钓鱼网站的词汇特征，并按照从表现最好的基线方法中提取的特征重要性降序排列。
  2 相关工作
  在本节中，我们将回顾钓鱼网站检测模型和攻击者的行为模式分析。
  2.1 检测钓鱼网站的方法
  为了应对网络钓鱼攻击，已经做了大量的工作[1, 3, 6, 12, 30, 33, 36, 37, 44, 50]。通常情况下，研究人员已经探索了机器学习技术来自动检测网络钓鱼的URL。为了保证

分类算法的有效性，拥有一套定义明确的特征是至关重要的。因此，我们在附录A中介绍了一套广泛使用的19个URL特征，这些特征是我们从相关论文中收集的。
在我们的实验中，所有这些特征都被一些基线方法所使用。所有提到的工作都不是基于基于网络的推理，而是基于特征工程。
Mao等人设计了一种基于网页内容特征的钓鱼网址检测方法，对逃避行为具有鲁棒性[33]。然而，在很多情况下收集这样的训练数据并不容易，因为网络钓鱼攻击不会持续很久，网页很快就会被删除，这是所有基于内容的检测方法的一个共同缺点[1]。
在[3，22，23]中，已经提出了几个序列（例如，在我们的环境中的URL）分类模型。其中一些具有先进的架构，以结合各种组件，如递归神经网络、卷积神经网络、词嵌入及其多个分层。我们使用他们的想法作为额外的基线。第一个使用长短期记忆（LSTM）单元，第二个模型使用一维卷积（1DConv），第三个基线同时使用（1DConv+LSTM）。
对于几个相关问题[7, 32]，已经使用了基于网络的方法。在[32]中，作者试图检测恶意域名（而不是URL），[7]中的作者提出了一种基于启发式的信念传播方法来检测恶意代码。这两项工作在如何创建网络方面有所不同，但使用相同的信念传播方法。这两种方法都对应于我们实验中标记为 "POL "的基线方法。Peng等人和Khalil等人也尝试用网络方法进行恶意域检测[21, 40]。然而，他们的方法并不直接适用于我们的钓鱼网址数据。
2.2 攻击者的行为模式
反钓鱼工作组的《钓鱼活动趋势报告》[15]是最有信誉的报告之一。我们分析了他们的季度报告。报告中最重要的两个观察结果是：i）有一些虚拟主机公司由于价格低廉和匿名性而受到攻击者的青睐，以及ii）许多钓鱼网址具有类似的字符串模式，因为它们是由类似的工具创建的，或者是从旧的钓鱼活动中重复使用的。还有许多其他有趣的观察结果如下。

- 使用免费主机供应商或网站建设者的钓鱼网站数量有所增加。据报道，81.7%的恶意网站是由免费主机提供商托管的[10]。这些免费的主机很容易使用，但也允许威胁者创建子域来欺骗用户。

通过www.DeepL.com/Translator（免费版）翻译

定义1. 规避是一种有效的技术，人们可以通过创建一个 "反明显 "的样本来干扰机器学习任务，例如，一个由良性域名或IP地址托管的钓鱼网站。这种规避可以通过各种方式进行。关于我们考虑的详细规避技术，请参考第6.6节。
Shirazi等人表明，现有的钓鱼网站检测方法受到了规避的不利影响，但没有提出对策[45]。具体来说，他们进行了规避测试，随机选择钓鱼网站的四个特征，并将所选特征改为其他良性值。在他们的非规避测试中，大多数分类器显示出较高的准确性。然而，在他们的规避测试中，表现最好的分类器的准确率（召回率）从82-97%下降到79-45%，有一个特征的变化，就下降到0%。
据我们所知，目前还没有人积极研究设计一种非基于内容的钓鱼网址检测方法，以稳健地应对规避行为。我们考虑了URL的许多方面，包括域名、IP地址、名称服务器和除内容外的字符串模式--因为收集钓鱼网站的内容需要付出非同小可的努力。最重要的是，我们的方法是基于它们的网络。直观地说，如果网络中的钓鱼网站的许多邻居仍然和以前一样，那么即使在规避之后，攻击者也不能干扰我们基于网络的推理任务（见第5节）。一些大

规模的规避行为仍然可以使我们的方法失效。然而，这需要非同小可的费用，从而降低了攻击者对这种规避的动机。

图2：寻找频率稳定点的肘部方法。在找到的稳定点之前的所有子串（词）都被认为是停止词。

虽然很难用金钱来衡量规避成本，但它包括各种无形的努力，例如利用良性的网络服务器来植入他们的钓鱼网页，在D日之前维持一个没有任何钓鱼活动的自定义域名，以防止它被列入黑名单，等等。特别是，它取决于安全环境和技能，直到攻击者成功利用良性服务器的管理员账户需要多长时间。

4 建议的方法

在介绍了我们的方法的整体工作流程后，我们描述了它的详细步骤和一些关键的可视化结果。

4.1 整体方法

图1显示了我们的整体工作流程。整个过程可以分为以下几个步骤。

(1) 我们从phishtank.com抓取许多URL，并下载其他作品的开放数据集。

(2) 如前所述，我们创建了一个由URLs、域名、IP地址、名称服务器和子串（词）组成的异质网络。我们使用标准的自然语言处理技术将URLs分割成子串（词），并在URL和子串之间画出边框。

(3) 我们运行我们定制的信念传播算法来推断未知URL的虚假/良性标签，这是我们的主要贡献。特别是，这种类型的推断被称为过渡性推断。在我们的案例中，训练样本和测试样本同时存在于一个网络中，测试样本的标签是根据网络结构从其他已知的训练样本的标签中推断出来的。

4.2 网络构建

我们做的是基于网络的分类，而不是基于特征工程的分类。如前所述，钓鱼网站有许多共同的字符串模式，而且各种实体相互交叉关联，因此我们创建了一个网络来表示多个实体（顶点）之间的复杂关系，如URL、其域名、IP地址、权威名称服务器和子字符串。- 我们在一个URL和它的域之间画一条边。

- 我们在一个域名和它所解决的IP地址之间画一条边。我们使用domains.google和virustotal.com来检索域名-IP地址解析历史。它们不仅返回当前的，而且还返回所有过去的带有时间戳的解析结果，从而实现正确的连接。有时，一个域名可以连接到多个IP地址。

图3：从你的数据中构建的网络。红色代表钓鱼网站，蓝色代表良性网站。其他颜色表示非URL实体--由于名称服务器的重要性低于其他实体，所以没有显示。请注意，存在着许多群集。顶点的大小代表一个顶点所属集群的强度（更确切地说，模块化[5]）。

- 我们在一个域和它的权威名称服务器之间画一条边。一般来说，一个域名存在多个权威的名称服务器，一个权威的名称服务器为多个域名提供解析服务。
- 如果一个URL（即句子）包含子串（即单词），我们在该URL和子串之间画一条边。对于这些边，如何将一个URL分割成子串是非常关键的。我们将在下一节中简短地描述这一点。

如何将一个URL分割成单词。一个URL是用来定位互联网上的资源的。它由几个部分组成：方案、用户名、密码、主机、端口号、路径和查询字符串--其中一些可能丢失。我们在每个部分使用我们定制的单词分割策略，如下所示。

- 方案指的是协议，例如，http和https。只可能有两个词。然而，由于这些词的频率非常高，我们在网络中不使用这两个词。我们将在短期内描述如何删除URL的这些停顿词。
- 用户名和密码可以在主机之前指定。我们使用标点符号来分割它们，即'//'、':'和'@'。一个例子是'http://username:password@example.com'。
- 主机名可以简单地用'.'分割成单词。
- 有时路径可以很长，用'/'隔开。我们使用所有可能的标点符号，如'/'、'.'、'!'、'&'、','、'#'、'$'、'%'和';'，将路径部分分割成单词。
- 查询字符串能够包含多个由'&'分隔的查询。'&'隔开，每个查询由一个查询名称和一个值组成，例如，'term=bluebird&source=browser-search'。我们使用两个标点符号'='和'&'来提取单词。

  由于URLs的语法定义得很好，提取单词可以非常有效地完成。然而，许多无意义的词也可以被提取出来。因此，在绘制URL和单词之间的边缘之前，应该把这些单词删除。在自然语言处理领域，众所周知，词的频率遵循Zipf定律--更确切地说，词的频率呈指数型衰减[47]。特别是，这种模式很好地描述了英语中的停止词。例如，在Brown Corpus of American English[13]中，最流行的停顿词 "the "的频率占据了所有单词出现次数的7%，第二流行的停顿词 "of "的频率为3.5%。我们发现，从URLs中提取的词也显示出类似的统计数据（参见图2）。之后，我们使用肘部方法[20]删除一些高频词。它决定将与连接两端的线段的垂直距离最大的点作为饱和点，在我们的数据中是800。我们删除所有频率值大于该点的词。

  图3显示了由拟议方法创建的网络。请注意，在集群的构建和地面真实的虚假/良性标签之间存在着很强的相关性，这证明了我们基于网络的推理方法是正确的，这一点将在不久之后描述。在这方面，我们工作的主要直觉是，除非一个集群中的大多数实体同时被规避，否则很难规避我们的自然语言处理和基于网络的方法。

  4.3 基于网络的推理

  我们采用循环信念传播（LBP）[4]进行基于网络的推理。我们在这一步的主要贡献是定义了一个比最先进的方法[7, 32]更先进的边缘潜力分配机制。因为这些方法通常不仅遵循邻居的多数投票，而且无论两个连接顶点的相似性如何，都会给出一个固定的边缘势，如果一个顶点有许多良性邻居，那么它主要被归类为良性。然而，我们希望正确地分类一个钓鱼网站顶点，即使它有许多良性邻居。因此，我们定义了一种更先进的边缘势能分配机制，以实现更复杂的分类，并达到规避的效果。我们将在第4.3.1节描述我们的边缘潜力定义。

  LBP是一种消息传递算法，用于解决基于网络的推理问题。让$x \in X$是一个隐藏变量，Nx是其邻近变量的集合，让$o \in O$是一个观察变量。在我们的语境中，观察变量意味着训练样本，隐藏变量意味着测试样本。我们用X和O分别表示所有隐藏变量和观察变量的集合。

  在我们的案例中，每个变量代表一个实体的虚假/良性标签。x在收集了来自Nx的所有信息后，向其他隐藏变量$y \in Nx$发送了一条信息。请注意，观察到的变量从未收到任何消息；它们只向其邻近的隐藏变量广播消息。在我们的案例中，训练集中的虚假和良性URL是观察变量。

  如前所述，我们需要计算一个消息$msg x \to y$（ℓ）从一个变量x到另一个变量y，关于一个虚假/良性标签$\ell \in L$，其中L={msg,$benign$}是一个所有可能标签选项的集合。存

在几种消息传递策略：sum-product, maxproduct, 和min-sum。我们使用最小和算法，该算法比其他两种算法具有更好的计算稳定性。对于一些高度顶点，在sum-product和max-product中，消息值往往会迅速衰减为零（即浮点下溢）。他们的乘积操作被简化为最小和算法中的和。最小和算法中的信息计算方法是：。

其中，ϕy（ℓ'）是变量y具有标签ℓ'的先验，而
$\psi xy(\ell, \ell')$是一个边缘势，表示x的标签是ℓ，y的标签是ℓ'的联合概率。请注意，在消息定义中存在一个对数函数，所以最小和相当于在对数空间中执行最大乘法，以获得更好的计算稳定性。
在多次交换消息后，我们首先计算出每个变量和标签对的成本，然后选择对每个变量产生最低成本的标签。当x的标签为ℓ时，其成本计算如下。
∑
$Cost\ (x, \ell) = \log\ (1 - \phi x\ (\ell)) + msgk \rightarrow x\ (\ell)$ 。 (2)
$k \in Nx$
那么，最小和算法所解决的问题的正式定义可以定义如下。
∑
argmin$g\ Cost\ (x, g(x))$, (3)
x
其中$g : X \rightarrow L$，其中X是一组隐藏变量，而
$L = \{phishy, benign\}$，是一个标签分配函数。值得一提的是，在我们的设定中，x可以是一个代表URL、域名、IP、名称服务器或单词的隐藏变量。我们的最终目标是推断出测试URL的标签。为此，我们也需要推断出其他非URL实体的标签，因为它们连接着URL。因此，最小和算法可以被描述为寻找这样的标签分配给隐藏变量的过程，使成本之和达到最小。
4.3.1 边缘势能分配。边缘势的定义ψxy（ℓ, ℓ'）是LBP方法的关键因素。[7]使用了同质性和异质性的启发式方法。例如，如果两个相邻的变量x和y具有不同（相同）的标签，如表1中的兼容性矩阵所示，他们会分配一个0.5 -ϵ（ resp. 0.5 +ϵ）的边缘潜力。ϵ通常被设定为非常小，例如0.001。我们使用两个标签，虚假的和良性的。表1：Polonium[7]中提出的基于同质性启发式的兼容性矩阵。

$\psi xy\ (\ell, \ell')$	Phishy	Benign
讨厌的	0.5 + ϵ	0.5 - ϵ
良性	0.5 - ϵ	0.5 + ϵ

表2: 我们的兼容性矩阵M用于最小和算法。x和y是指矢量表示。sim(x, y)是两个矢量之间的相似度。

$\psi xy(\ell, \ell')$	Phishy	Benign
Phishy	min(ths+, 1 - sim(x, y))	max(ths-, sim(x, y)
虚的	max(ths-,sim(x, y))	min(ths+, 1 - sim(x, y))

在推理过程中，对于每个实体，将分配一个标签作为预测结果。最终的标签分配在很大程度上受到边缘势能定义的影响。
与[7]相比，我们加入了更多的因素，如实体之间的相似性和改进的兼容性矩阵，以得出可靠的边缘势--我们在第5节中很快证明，可靠的相似性定义可以导致我们方法中的规避-稳健性。类似性可以通过各种嵌入方法来测量，比如Doc2Vec[24]和Node2Vec[16]。我们在第4.3.2节中讨论了如何计算URL、其域名、IP地址、权威名称服务器和单词的向量表示。为了计算基于这些矢量表示的相似性，我们采用了几种不同的相似性测量方法，包括余弦

相似性和各种内核。我们提出的边缘潜力的定义如表2所示。在表中，我们用黑体字表示实体的向量表示，sim(x, y)表示两个向量之间的相似性，可以用各种方式定义。以下是两个这样的例子。
cos(x, y)基于余弦相似度。
sim(x, y) =
基于RBF核。

之后, 我们使用受铰链损失[42]启发的概念来分配边缘潜力值. 例如, 表中的min(ths+, 1 - sim(x, y))是将两个实体具有相同标签时的最小边势限制在ths+。当sim(x, y)较低(resp. high)时, 建议的定义会施加一个封闭于1(resp. ths+)的大(resp. small)惩罚。因此，与现有方法相比，建议的机制能够分配更复杂的边缘电位。
在将我们的兼容性矩阵应用于其他应用时，应该非常小心。回顾一下，我们使用的是最小和算法，因此在我们的兼容性矩阵M中，当ℓ和ℓ'相同时，我们分配0（在和-积和最大-积算法中对应1）。对于和-积和最大-积算法，应该使用1 - M。

实体的矢量表示。我们描述了如何计算各种实体的可靠向量表征。众所周知，这些嵌入方法可以有效地发现潜在的

图4：用DeepWalk进行的成对顶点相似性的例子。
DeepWalk和我们的网络推理的例子，ths+ = ths- = 0.7。
(a) 我们用广度优先搜索法选择了PageRank最高的URL和它附近的其他199个URL，分别为虚假类和良性类。总共有400个URL在相似度图中。(b) 从相似性中，我们基于网络的推理能够几乎正确地推断出。
实体之间的关系[16, 24-27, 35, 41, 52]，在存在回避的情况下，这与我们基于网络的检测很适合。
基于词嵌入的方法。在自然语言处理领域，已经提出了各种语义嵌入方法，如Word2Vec[35]和Doc2Vec[24]。正如我们前面提到的，我们将URLs分割成单词，所以我们可以直接应用这些方法来计算URLs和单词的向量表示。然而，在这种方法中我们不能直接计算域名、IP地址和名称服务器的向量表示，因为它只考虑字符串。然而，受局部线性嵌入（LLE）[43]的启发，我们提出了一种启发式方法，将域名、IP地址或名称服务器表示为其邻居矢量的平均矢量。LLE说，一个实体的向量表示是其邻居向量的加权组合，例如，在我们的案例中是等权的。为此，我们首先计算域名的平均矢量表示，然后是IP地址等，给定URL的矢量表示由Word2Vec或Doc2Vec计算。
基于网络嵌入的方法。另一种可靠的寻找矢量表征的方法是使用网络嵌入方法。许多这样的方法都是由社会网络研究人员提出的。这种方法的一个优点是，我们可以同时找到所有实体的向量表示，因为它们可以直接在我们的网络上运行。我们使用Node2Vec[16]和DeepWalk[41]。在图4（a）中，我们展示了一个成对的相似性图，直观地证明了我们的嵌入和基于相似性的边缘潜力分配。然而，我们看到绿圈中的一小部分虚假的和良性的配对有很高的相似性。这可以通过我们提出的边缘潜力分配机制来纠正，如图4（b）所示。
5 我们基于网络的方法的规避稳健性
在这一节中，我们正式证明一个隐藏变量的虚假/良性标签遵循其相似邻居的多数标签，这提高了对规避的稳健性。
例子1. 假设ths+ = ths- = 0，一个由隐藏变量ÃÃ组成的小网络。
由一个隐藏变量u和它的m邻居Nu={v1, - - ,vm}组成的小网络。

图5：为了便于讨论，假设u是一个隐藏变量，其他变量的标签是固定的。每条边都用sim(u, vi)来注释。我们的方法认为u是虚假的，尽管u有更多良性的邻居。

图6：有两个聚类。一般来说，虚假集群和良性集群之间的联系并不强（参考图3）。'Domain1'在回避后与'IP2'有联系。然而，它们之间的联系很弱，而且在嵌入之后。sim(x, y)很低，其中x=$Domain1$，y=$IP2$。因此，我们的兼容性矩阵对它们的不同标签给予了较低的惩罚，信念传播仍然可以将'Domain1'识别为虚假的。

让ℓu成为u的虚假/良性标签。当ℓu=argminℓ Íj sim(u, vj)- I (vj, ℓ)，其中I (vj, ℓ)∈{0, 1}是一个指标函数，表示如果vi有一个标签ℓ，则公式（3）的最小和算法被优化。

证明。ℓu是由公式（2）推断出来的。特别是方程中的第二项，Ív∈$Numsgv\rightarrow u$（ℓ），对决定其标签是有意义的，而且$msgv\rightarrow u$（ℓ）在假定网络中只被ψv（ℓv, ℓu）所支配（参考公式（1））。Ív∈Nu ψvu (ℓv, ℓu)当ℓu遵循考虑向量相似性的多数标签时是最小的，因为ψvu（ℓv, ℓu）是由表2中的sim（v, u）决定。□

例1(Lemma 1的例子)。在图5中，有我们在Lemma 1中使用的小网络。为了便于讨论，假设只有u是隐藏变量，其他都是观察变量。最佳的最小和解是$g(u)$=$Phishy$，因为sim（u, v1）>
j>1 j(u, vj)，并且m (u, os) = Íj>1 j(u, vj) 小于 m (u, $Benign$) = j(u, v1)。

对于更大的一般网络，这个定理可以推广到下面的定理。

表3：每个数据集的虚假和良性URL的数量。请注意，Sorio和Ahmad的数据集已经被贴上了地面真实的标签，所以我们没有对它们使用virustotal. 存在重叠的URLs，所以URLs的总数小于它们的总和。

数据集 VirusTotal
阈值 # 钓鱼网站
URL # 良性
网址
美国银行 4/7 4,610 9,408
eBay 4/5