计算机科学
COMPUTER SCIENCE

独立级联模型下基于双区分集的观察节点选择方法

陈张缘, 陈崚, 刘维, 李斌

引用本文

陈张缘, 陈崚, 刘维, 李斌. 独立级联模型下基于双区分集的观察节点选择方法[J]. 计算机科学, 2025, 52(4): 280-290.

CHEN Zhangyuan, CHEN Ling, LIU Wei, LI Bin. Method for Selecting Observers Based on Doubly Resolving Set in Independent Cascade Model [J]. Computer Science, 2025, 52(4): 280-290.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240300127

独立级联模型下基于双区分集的观察节点选择方法

陈张缘　陈　峻　刘　维　李　斌
扬州大学信息工程学院　江苏 扬州 225000
(chenzhangyuan_001@163.com)

摘　要　随着互联网的发展,谣言信息可以在社交网络上快速传播,找到谣言源头有助于阻止负影响的传播,因此谣言源定位问题有着重要的研究价值。目前,最有效的源定位方法是基于观察节点的方法,但是现有选择观察节点的方法都没有考虑图的顶点分布的均匀性,并且都是预先设置观察节点的数量而没有根据图的拓扑特性来合理确定观察节点的个数。文中从节点预算阈值和节点的覆盖率阈值两个角度研究观察节点的放置策略,考虑了观察节点激活状态以及到源集合的区分距离,并提出了一种新的K-双区分算法。该算法首先根据双区分集概念选择初始观察节点,然后选择其中一个锚点根据提出的覆盖率和预算约束问题贪心地选择观察节点来达到预算和覆盖率阈值。在真实数据集上对所提算法进行了实验,在同一种源定位算法中对比多种选择观察节点的算法。实验结果表明,所提算法的源定位结果精确度和平均距离误差均优于对比算法,在大型数据集中只使用5%～10%的观察节点就可以达到很好的定位效果。

关键词:观察节点;社交网络;独立级联模型;双区分集;源定位
中图分类号　TP399

Method for Selecting Observers Based on Doubly Resolving Set in Independent Cascade Model

CHEN Zhangyuan,CHEN Ling,LIU Wei and LI Bin
College of Information Engineering,Yangzhou University,Yangzhou,Jiangsu 225000,China

Abstract　With the development of the Internet,rumor information may quickly spread on social networks. Finding the source of rumor can effectively help stop the spread of negative influences,so the source localization problem has great research value. At present,the most effective source localization method is based on observation nodes. But the existing selection of observation nodes has not considered the uniformity of node distribution in the network,and the number of observation nodes is pre-set without reasonably considering the topological characteristics of the graph. This paper studies the strategy of observer nodes placement from two perspectives:node budget threshold and node coverage threshold. Taking into account the activation status of observer nodes and the discriminative distance to the source set,a novel K-differentiation algorithm is proposed,which initially selects the initial observer nodes based on the concept of the doubly-resolving set. Subsequently,an anchor node is chosen to greedily select observer nodes based on a combination of coverage and budget differences to reach the specified thresholds for coverage and budget. The algorithm then addresses two problems related to budget and coverage proposed in this paper when choosing observer nodes for experimentation within the same source localization algorithm. Experiments are conducted in real world social datasets. The proposed algorithm for selecting observer nodes is compared with various alternatives within the same source localization algorithm. The results indicate that the accuracy of source localization and average error distance by our algorithm outperforms other algorithms and achieve excellent results in the large dataset with 5%～10% observer nodes.

Keywords　Observer nodes,Social networks,Independent cascade model,Doubly resolving set,Source localization

到稿日期:2024-03-18　返修日期:2024-08-14
基金项目:国家自然科学基金(61379066,61702441,61379064,61472344,61402395,61602202);江苏省自然科学基金(BK20140492,BK20160428);江苏省教育厅自然科学基金(12KJB520019,13KJB520026,09KJB20013);江苏省六大人才高峰(2011-DZXX-032)
This work was supported by the National Natural Science Foundation of China(61379066,61702441,61379064,61472344,61402395,61602202),Natural Science Foundation of Jiangsu Province,China(BK20140492,BK20160428),Natural Science Foundation of Jiangsu Provincial Department of Education,China(12KJB520019,13KJB520026,09KJB20013)and "Six Talents Peaks" of Jiangsu Province(2011-DZXX-032).
通信作者:陈峻(lchen@yzu.edu.cn)

1 引言

近年来，随着移动互联网的快速发展，社交网络已经成为人们生活和工作必不可少的组成部分。然而，这些平台在为信息交流提供便利的同时，也降低了谣言的传播成本，这无疑加速了谣言的传播。社交网络已逐渐成为被广泛攻击的目标以及被用作欺诈活动的媒介[1-2]。譬如，在商业和政治网络中，谣言和虚假新闻等负面信息会像流行病毒[3]那样迅速传播。这些有害的信息会误导人们的思维，使其对一些话题产生错误的认识，从而对社会产生不良影响。为了防止这些负面影响在社交网络中传播，我们需要从源头上抑制谣言信息的传播。源定位问题指在社交网络中找到最初传播谣言的几个节点[4]。

目前源定位方法主要有两种：一种是基于传播子图快照的方法[5]，另一种是基于观察节点的方法[6-9]。前者需要观察网络中所有节点的状态。然而，获取真实网络中每个节点的状态是非常困难和耗时的。因此，研究人员将注意力转向了基于部署观察节点的方法，并提出了高斯估计和有效距离[10]等多种方法。但这些方法大多是针对单源定位的研究方法，并不完全适用于多源定位。也有一些关于多源定位问题的研究[11-13]。Bao 等[14]提出了一种基于最大后验估计的社交网络谣言源定位的方法，该方法首先考虑全局和局部感染点、非感染点的影响，使用最大后验估计值来估计传播网络中节点为传播源的可能性。Yuan 等[15]提出了一种基于拓扑扩展的在线社交网络恶意信息源定位算法，该算法首先根据当前网络节点的状态挖掘隐藏信息，以选取候选源节点，然后基于 Jordan 中心性对恶意信息溯源进行定位。然而，关于如何选择最优观测节点部署策略的研究却很少。

现有的一些关于选取观察节点的研究首先是一些经典和常用的中心性方法。例如，Li 等[16]认为不同的选取节点策略可能会对源定位结果产生影响。他们对比了随机选取方法和度中心性方法，发现度中心性方法的平均跳数误差在一些数据集中的表现优于随机方法。Zhang 等[17]将度中心性、介度中心性以及特征向量中心性进行了对比，认为观测节点的覆盖范围可能是影响定位精度的关键因素。他们发现，定位精度与观测器的覆盖率密切相关，换句话说，高覆盖属性的观察节点可能有助于提高对源定位的精确度。Paluch 等[18-19]提出了一种被称作集合介度（Collective Betweenness）的基于最短路径的中心性度量，并通过实验证明了这种中心性度量在一些网络中优于其他中心性度量。但在对一些合成网络的实验中，如果网络节点感染率较高，则该方法的定位效果并不是很理想。此外，这些中心性度量算法往往并没有考虑到节点的预算或者将每个节点预算都设置成一样，没有考虑到每个节点在网络中的重要程度可能并不一样。

Spinelli 等[20]基于路径覆盖策略提出一种选择观察节点的方法，该方法在观察节点密度较小时表现较好，在他们的实验中可以覆盖网络中的所有节点。他们将集合 S 中任意两个观察者之间恰好有一条长度不超过 L 的最短路径的节点定义为单路径覆盖节点，双路径、三路径覆盖节点等的定义与之类似。Spinelli 等采用贪婪方法依次选择使单路径、双路径、三路径覆盖节点等数量最大化的观察节点，直到达到理想的观察节点密度。但文中考虑的路径覆盖范围和实验数据集都较小，为符合实际，应该在更大的数据集中进行实验。Zhang 等[17]提出基于邻居节点覆盖的方法，选择一组具有较多度为 1 的邻居节点作为观察节点。Zejnilovic 等[21-23]针对单源定位问题提出了一种根据源识别时间动态选择观察节点的方法，该方法在零方差设置中考虑了最小化检测精确源所需的观察者数量的问题，以及给定观察者预算最大化定位精度的问题。给定谣言开始的时间，他们解决了在树状图中最小化检测源所需观察节点数量的问题。但由于树是一种特殊的网络结构，该方法并不适用于实际应用中的网络。

Gajewski 等[24]将上述的路径覆盖算法、邻居节点覆盖算法以及 CB 中心性算法与随机算法进行了实验对比。在不同的激活概率下，这 4 种算法中没有一种在源定位性能上能稳定地优于其他方法。在较大的激活概率下，随机算法甚至优于其他 3 种算法。Gajewski 等在实验中没有考虑到节点激活概率的随机性，这与实际应用的网络环境不符。此外，路径覆盖算法与 CB 中心性算法都涉及最短路径，而邻居节点覆盖算法则需要分组配对节点，并且需要重复遍历配对节点的邻居节点。这 3 种算法的时间复杂度都较高，因此不适用于大型网络的应用。

近来，还有一些关于运用双区分集选择观察节点的研究。Chen 等[25]从确定性的角度对源定位问题进行了研究，将其建模为最小预算双区分集（DRS）问题，并在循环图以及树中求解最小预算问题。Spinelli 等[20]利用观察节点到其他任意两个节点的距离的差异来区分源，如果一个观察节点到其他两个节点的距离不一样，则可以区分这两个节点。Wang 等[26]提出一个选择观察节点的指标，计算每个观察节点到其他任意两个节点的距离和以及距离差，并选择两者中较小的方差作为筛选值。他们发现选择方差筛选值较大的节点作为观察节点进行源定位的效果更好。Zhao 等[27]在选择观察节点时也采用了类似的方法。这些策略虽然在实验数据集中利用较少的观察节点就可以达到较高的精确度，但由于需要对所有的节点进行配对，使得时间复杂度随网络节点数呈指数上升，因此无法适用于大型数据集。此外，这些方法都无法根据网络拓扑性质以及源的个数来选择合适的观察节点个数，也就无法体现出观察节点的覆盖属性。

本文提出一种在基于观察节点的源定位问题中选择观察节点的新方法。该方法首先在网络中抽取若干个源的集合，根据双区分集的性质选取初始观察节点；然后针对源集合的覆盖率阈值和节点的预算阈值等不同问题，提出使用贪心策略选择观察节点的算法。

本文的主要贡献如下：

1）针对覆盖率和预算阈值的限制这两个条件，提出了两种观察节点选择问题，并提出使用贪心策略选择观察节点来解决这两个问题的算法。

2）在独立级联模型的源定位算法下多源定位问题中，所提方法可以根据源个数的不同，利用双区分集性质选择指定

源个数的区分观察节点，并且可以通过调整源集合参数来调整区分精度，从而降低算法的时间复杂度，在大型网络中也能保持较高的效率。

3）在真实网络的大型数据集中对所提算法的性能进行测试，并与 6 种对比算法进行比较。实验结果表明，与对比算法相比，该算法所选择的观察节点能够在多个覆盖率阈值、成本阈值下有效提升源定位效果。

本文第 2 章介绍传播网络的基于观察节点的源定位问题、独立级联（IC）传播模型以及观察节点选择问题；第 3 章介绍双区分集概念及其性质；第 4 章提出通过双区分集选择初始观察节点、针对两个问题选择观察节点的算法；第 5 章是对本文方法的实验评估及结果分析；最后总结全文。

2 问题的定义及传播模型

本章首先介绍基于观察节点的源定位问题定义与独立级联传播模型，然后介绍源定位中的观察节点选择问题。

2.1 基于观察节点的源定位问题定义

给出网络 $G=(V,E,\boldsymbol{P})$，V 表示节点集，E 表示边集，$\boldsymbol{P}=[p_{uv}]$ 表示边的权重矩阵，每条有向边 $(u,v)\in E$ 都有一个权重 $p_{uv}\in(0,1)$，表示节点 u 到邻居节点 v 的传播概率。设网络 G 被观察到的已被激活的顶点集合为 $O=(o_1,o_2,\cdots,o_m)$，$O\subset V$，$m<n=|V|$。如果影响力能够非常有效地传播开，最可能的传播源一定是一开始就极具影响力的那几个顶点。因此，给定一个正整数 k，传播源的定位问题就是求得具有 k 个顶点的传播源集合 S，其中 k 的大小是由选取观察节点的算法输入的源节点个数参数确定的，可以根据网络中节点数量选择合适数量的源节点，使得 S 所影响的范围 $I(S)$ 最大限度地包含 O 中的顶点。

2.2 IC 模型

本文研究独立级联（IC）模型下的基于观察节点溯源定位问题。IC 模型是一种常见的概率传播模型。在 IC 传播模型中，每个节点在传播过程中的每一个时刻处于两种状态之一：激活状态和未激活状态。节点在激活状态表示该节点受到影响，反之则没有。当传播开始时，所有传播源节点处于激活状态，然后由它们根据各边上的传播概率去激活其他节点。IC 模型的传播过程如下：当网络中的某节点被源节点 $s\in S$ 激活后，它会试图激活其未被激活的邻接节点 v，v 被 u 成功激活的概率为 p_{uv}。设节点 v 有多个邻居节点，这些邻居节点能否被成功激活是互相独立的，每个节点只有一次机会尝试激活其邻居节点。如果一个节点被激活，就不会再被其他节点激活，在下一时刻，它会试图去激活自己的邻居节点。重复上述过程，直至网络中不再有新的节点被激活，传播过程结束。

图 1 给出了 IC 模型中的影响力传播过程。其中，传播源为 a，它指向 3 个邻居节点，即 b，c 和 f。a 节点只有一次机会试图激活它的每一个邻居节点，而且对于 b，c，f 这 3 个节点，能否被成功激活是互相独立的。由于 a 对于 b 这个顶点的激活概率很低，因此 b 未被激活，而 f 和 c 被 a 成功激活，且始终保持激活状态不再改变。在下一时刻 c 和 f 又试图去激活自己的邻居节点 f，j 等。重复这样的激活过程，直至网络中没有新的可激活的节点。

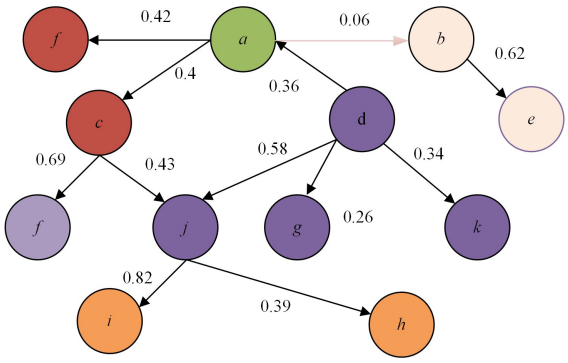

图 1 IC 模型传播示意图

Fig. 1 Schematic diagram of IC propagation model

2.3 观察节点选择问题

在基于观察节点的影响力源定位问题中，首先需要选择一定数量的观察节点。观察节点的分布和数量决定了源定位结果的质量。一个合理的观察节点集合应该能从传播路径的长度的角度正确地区分不同的源，从而使得源定位方法能够找到真正的源集合。这就需要观察节点对源节点集合具有很好的区分性。我们称一个观察节点集合能够区分的源节点集合的占比为它的覆盖率。显然，当观察节点较多时，覆盖率就越高。但是，选取每个节点作为观察节点有一定的成本。为此，我们定义两种观察节点的选择问题：1）给定一个覆盖率阈值 δ，目的是要找到覆盖率大于 δ 的开销最小的观察节点集合 O^*，即在给定的覆盖率阈值下的观察节点成本最小化；2）给定一个开销阈值 B，目的是要找到满足成本不超过 B 的观察节点集合 O^*，使得其覆盖面最大，即在给定成本预算下的观察节点的覆盖面最大化。

3 双区分集及其性质

为了找到源定位问题的合理的观察节点，首先给出双区分集的概念。

定义 1 设 x 为一个源节点，y 为一个观察节点，x 节点发出信息的时间为 $t(x)$，x 节点到 y 节点的距离为 $d(x,y)$，记 $F(x,y)=t(x)+d(x,y)$ 为 y 节点收到 x 节点所发出的信息的时间。

为了基于双区分集进行观察节点选择，我们以 $F(x,y)$ 为基础，给出如下定义。

定义 2 设 x,y 为两个源节点，如果有两个观察节点 a,b 使得 $F(x,a)-F(y,a)\neq F(x,b)-F(y,b)$ 成立，则称观察节点 a,b 可以双区分源节点 x,y，记为 $\{a,b\}\bowtie\{x,y\}$。

定义 3 设 A 为大小为 k 的源节点的集合，y 为一个观察节点，设 A 中的每一个源节点 x 发出信息的时间为 $t(x)$，x 到 y 的距离为 $d(x,y)$，记 $F(A,y)=\min\limits_{x\in A}F(x,y)$ 为 y 受到 A 所发出的信息影响的时间。

定义 4 设 $A\subset V$，$B\subset V$ 为两个不同的大小为 k 的源节点的集合，O 为观察节点集合，设 O 中至少有两个观察节点 o_1 和 o_2，使得 $F(A,o_1)-F(B,o_1)\neq F(A,o_2)-F(B,o_2)$ 成立，则称观察节点 O 可以 k-双区分源节点集合 A 和 B，记为 $O\bowtie$

$\{A,B\}$。

定义 5 设 $O\subset V$ 为观察节点集合，$T_k=\{S\mid S\subset V,|S|\leqslant k\}$ 为所有规模不大于 k 的源节点集的集合。若 T_k 中的任意两个不同的源节点集合 S_1 和 S_2 都可以被观察节点的集合 O 所双区分，即 $O\bowtie\{S_1,S_2\}$，则称 O 为 V 上可 k-双区分的观察节点集。

若节点的集合 O 要成为观察节点集，则它必须是 k-双可区分的。这是因为，如果 O 不是 k-双可区分的，则存在两个不同的源集合 S_1 和 S_2，使得 O 中任意两个顶点 o_1 和 o_2 皆有 $F_{S_1}(o_1)-F_{S_2}(o_1)=F_{S_1}(o_2)-F_{S_2}(o_2)$，则被观察节点的集合 O 不能区分集合 S_1 和 S_2 谁才是真正的源。因此，我们的目的是寻求一个规模最小的 k-双可区分顶点集合 O 作为观察节点集合。但是，如果用定义寻求这样的 k-双可区分顶点集合 O，需要检验 O 中所有顶点对 o_1 和 o_2，以及 T_k 中所有不同的源集合对 S_1 和 S_2，计算量很大。由此易知寻找规模最小的 k-双区分的问题是一个 NP-难问题。因此，我们可以利用双区分顶点集合的一些性质来求得该问题的近似解。

定理 1 对于网络 $G=(V,E)$，设 O 是 V 的一个 k-双区分集合，P 是 V 中的顶点集合且 $O\subset P$，则 P 也是一个双区分顶点集合。

证明：因为 O 是 V 的一个 k-双区分集合，则 V 中的任意两个不同的节点集合 S_1 和 S_2 都可以被观察节点的集合 O 所 k-双区分，即：存在顶点 $o_1\in O,o_2\in O$ 使得 $\{o_1,o_2\}\bowtie\{S_1,S_2\}$。由于 $P\supset O$，可知顶点 $o_1\in P,o_2\in P$，即 P 中存在顶点 o_1,o_2 使得 $\{o_1,o_2\}\bowtie\{S_1,S_2\}$。因此，$P$ 也是一个 k-双区分顶点集合。证毕。

定理 2 对于网络 $G=(V,E)$，设 $O=\{v_1,v_2,\cdots,v_k,\}$ 是 V 的一个顶点集合，设 S_1 和 S_2 为 V 中的两个不同的节点集合，记 $M(O,v_1)=\{\{v_1,v_i\}\mid i=2,\cdots,m\}$，当且仅当 $M(O,v_1)$ 中至少有一个顶点对 $\{v_1,v_i\}$ 满足 $\{v_1,v_i\}\bowtie\{S_1,S_2\}$ 时，O 可以 k-双区分 S_1 和 S_2，即 $O\bowtie\{S_1,S_2\}$。

证明：1）由于 $O\bowtie\{S_1,S_2\}$，证明存在 $v_i\in O$，满足 $\{v_1,v_i\}\bowtie\{S_1,S_2\}$。

因为 $O\bowtie\{S_1,S_2\}$，根据定义，至少存在顶点对 $\{v_j,v_i\}\subset O$ 满足 $\{v_j,v_i\}\bowtie\{S_1,S_2\}$。则有：

$$F(S_1,o_j)-F(S_2,o_j)\neq F(S_1,o_i)-F(S_2,o_i)\tag{1}$$

考察 $F(S_1,o_1)-F(S_2,o_1)$，必有：

$$F(S_1,o_1)-F(S_2,o_1)\neq F(S_1,o_i)-F(S_2,o_i)\tag{2}$$

或者

$$F(S_1,o_1)-F(S_2,o_1)\neq F(S_1,o_j)-F(S_2,o_j)\tag{3}$$

成立。否则式(1)不成立。若式(2)成立，则有 $\{v_1,v_i\}\bowtie\{S_1,S_2\}$；若式(3)成立，则有 $\{v_1,v_j\}\bowtie\{S_1,S_2\}$。

2）若存在 $v_i\in O$，满足 $\{v_1,v_i\}\bowtie\{S_1,S_2\}$，由定义 4 可以直接得出：$O\bowtie\{S_1,S_2\}$。证毕。

由定理 2 可知，我们要检查 O 是否可以 k-双区分 S_1 和 S_2，不必按照定义对 O 中所有的顶点对 $\{v_j,v_i\}$ 检查是否有 $\{v_j,v_i\}\bowtie\{S_1,S_2\}$，只需要在 $M(O,v_1)$ 中检查即可，这样可以大大减少计算量。我们称 v_1 为锚点。由于对 v_1 选择具有任意性，可以看到，O 中的任意一个顶点都可以用作锚点。

定理 3 对于网络 $G=(V,E)$，设 S_1 和 S_2 为 V 中的两个不同的节点集合，设顶点集合 $O=\{v_1,v_2,\cdots,v_k,\}$ 不能 k-双区分 S_1 和 S_2，v 是 V 的一个顶点，$v\notin O$，当且仅当 O 中至少有一个顶点 v_i 满足 $\{v,v_i\}\bowtie\{S_1,S_2\}$ 时，$O\cup\{v\}$ 可以 k-双区分 S_1 和 S_2，即：$O\cup\{v\}\bowtie\{S_1,S_2\}$。

证明：

1）由于 $O\cup\{v\}\bowtie\{S_1,S_2\}$，证明 O 中至少有一个顶点 v_i 满足 $\{v,v_i\}\bowtie\{S_1,S_2\}$。因为 $O\cup\{v\}\bowtie\{S_1,S_2\}$，$O\cup\{v\}$ 中至少有两个顶点 x,y 满足 $\{x,y\}\bowtie\{S_1,S_2\}$。但 $O=\{v_1,v_2,\cdots,v_k,\}$ 不能 k-双区分 S_1 和 S_2，O 的任何节点对 $\{v_j,v_i\}$ 都不能 k-双区分 S_1 和 S_2，则只有 $M(O\cup\{v\},v)=\{\{v,v_i\}\mid i=1,2,\cdots,m\}$ 中的某一个节点对 $\{v,v_i\}$ 能 k-双区分 S_1 和 S_2。

2）由于 O 中至少有一个顶点 v_i 满足 $\{v,v_i\}\bowtie\{S_1,S_2\}$，证明 $O\cup\{v\}\bowtie\{S_1,S_2\}$。

由定义 1 直接可以得出结论。证毕。

由定理 3 可知，如果 $O=\{v_1,v_2,\cdots,v_k,\}$ 不能 k-双区分 S_1 和 S_2，要选择 $V\backslash O$ 中的某一个顶点加入 O，使得 $O\cup\{v\}$ 能 k-双区分 S_1 和 S_2，可以不必按照定义对 $O\cup\{v\}$ 中所有的顶点对 $\{v_j,v_i\}$ 检查是否满足 $\{v_j,v_i\}\bowtie\{S_1,S_2\}$，只需要在 $M(O\cup\{v\},v)=\{\{v,v_i\}\mid i=1,2,\cdots,m\}$ 中检查即可，这样可以进一步大大减少计算量。

4 观察节点选择算法

4.1 选择观察节点的问题分类

对于一个候选的观察节点集合 O，要对 T_k 中所有的源节点集合的配对 S_i 和 S_j 检查是否满足 $O\bowtie\{S_1,S_2\}$，计算量很大。因此我们采用均匀抽样的方法，构建 m 个源节点集合 $S=\{S_1,S_2,\cdots,S_m\}$，满足 $\bigcup_{i=1}^{m}S_i=V$。记集合 $Q=\{\{S_i,S_j\}\mid 1\leqslant i<j\leqslant m\}$ 为 S 中集合的所有配对，记观察节点集合 O 所能 k-双区分的 Q 中源节点集的配对的集合为：

$$C(O)=\{\{S_i,S_j\}\mid O\bowtie\{S_i,S_j\}\}\tag{4}$$

定义观察节点集合 O 的覆盖率 $H(O)=\dfrac{|C(O)|}{|Q|}$；O 的总开销为 $w(O)=\sum_{v\in O}w(v)$，其中 $w(v)$ 为节点 v 的开销。

在实际选择观察节点时，要考虑到观察节点的开销和覆盖率，因此有两类不同的问题。

问题 1 给定一个覆盖率阈值 δ，目的是要找到覆盖率满足 $H(O)\geqslant\delta$ 的开销最小的集合 O^*：

$$O^*=\underset{H(O)\geqslant\delta}{\operatorname{argmin}}W(O)\tag{5}$$

即覆盖率不小于阈值 δ 且开销最小的观察节点集合 O^*。

问题 2 给定一个开销阈值 B，目的是要找到满足 $w(O)\leqslant B$ 且覆盖面最大的集合 O^*：

$$O^*=\underset{w(O)\leqslant B}{\operatorname{argmax}}H(O)\tag{6}$$

即开销不超过阈值 B 且覆盖面最大的观察节点集合 O^*。

4.2 覆盖面计算算法

本文采用贪心方法来选择观察节点，首先由两个节点构成初始观察节集合。在确定了初始观察节点集合 O 后，在此集合中任意选择一个锚点 v，利用锚点 v 根据定理 3 寻找能最大地覆盖集合 $Q=\{\{S_i,S_j\}\mid 1\leqslant i<j\leqslant t\}$ 且代价最小的顶

点加入可观察节点集合。为此，我们提出对锚点 v 和所有被考察的候选节点 u 计算其对于集合 Q 的覆盖面 $C(\{v,u\})$ 的算法，具体如算法 1 所示。

算法 1 *Cover*(v,V,Q)

Input：社交网络 G=(V,E)，传播概率矩阵 **P**=[$p_{v,u}$]，边(v,u)的传播概率$p_{v,u}$，锚点 v，候选源集的配对的集合 Q，抽样源集的集合 S

Output：{v,u}的覆盖顶点集合 C({v,u})，for all u∈V\{v}

Begin

1. For each node u in V do
2. C({v,u})=∅;
3. For each pair {S_1,S_2} in Q do
4. If F(S_1,v)−F(S_2,v)≠F(S_1,u)−F(S_2,u) then
5. C({v,u})=C({v,u})∪{{S_1,S_2}};
6. End if;
7. End for;
8. End for u;
9. Return(C({v,u}) for all u∈V\{v});

End

算法 1 复杂度分析：第 1 行的 for 循环的复杂度为 O(n)；第 3 行的 for 循环的复杂度为 O(m^2)，m 为抽样的源集合的个数。由于 m 可以被看作常数，因此算法 1 的复杂度为 O(n)。

4.3 初始观察节点选择算法

为了采用贪心方法来选择观察节点，我们提出如下算法来产生随机抽样源集合，以及由两个节点构成的初始观察节集合。首先随机产生若干个大小为 k 的源集合S_1，S_2，…，S_t，使得$\bigcup_{i=1}^{t} S_i = V$，同时对每一个顶点$v_i$随机分配一个发出谣言的时间$t_i$。易知，如果网络中有度为 1 的节点，则它一定被包含在能完全覆盖所有源集合的双区分集合之中。因此，算法首先选择度为 1 的节点 v 为初始观察节点。如果没有这样的顶点，就选择 V 中 $d(v)/w(v)$最大的顶点 v 为初始观察节点。在对 $V\backslash\{v\}$ 中所有的节点 u 计算其对于集合 Q 的覆盖面 $C(\{v,u\})$以后，选择$|C(\{v,u\})|/w(u)$最大的顶点 u 加入观察节点集合。产生随机抽样源集合、初始观察节点集合的算法如算法 2 所示。

算法 2 Initialized observed set

Input：社交网络 G=(V,E)，传播概率矩阵 P=[$p_{v,u}$]，边(v,u)的传播概率$p_{v,u}$，源集合的大小 k，节点 v 被选为观察节点的开销 w(v)，v∈V，节点 v 的度 d(v)，v∈V

Output：初始观察顶点集合 O，抽样源集的集合 S

Begin

1. m=0；S=∅;
2. While S≠V do
3. Generate source setS_m with size k randomly;
4. m=m+1; S=S∪{S_m};
5. End while;
6. 记集合 Q={{S_i,S_j}|1≤i<j≤m};q=|Q|;
7. Assign time that send rumors $t_i$for every node v_i in S randomly;
8. If there exists nodes with degree 1 in V then
9. Select node v with the smallest value of w(v);O=O∪{v};
10. else
11. Select v with the biggest value of $\frac{d(v)}{w(v)}$ in V;O=O∪{v}
12. Endif;
13. V′= V\{v};
14. Cover(v,V′,Q);
15. select u with the biggest value of |C({v,u})|/w(u);
16. O=O∪{u};
17. Return(O,S)

End

算法 2 复杂度分析：算法的 1－6 行选取源顶点集，复杂度为 O($k*n$)，n 为顶点的个数；7－16 行选取两个初始观察节点，复杂度为 O(n) 。由于 k 可以被看作常数，因此算法 2 的复杂度为 O(n)。

4.4 基于覆盖率阈值的成本最小观察节点选择算法

由于函数 $F(O)=|C(O)|$对于 O 具有单调性和次模性，因此我们提出贪心算法，对问题 1 求得成本最小化的观察节点集合。在用算法 1 对锚点 v 和所有被考察的候选节点 u 计算其对于集合 Q 的覆盖面$C(\{v,u\})$以后，我们用贪心法选择$|C(\{v,u\})|/w(u)$最大的顶点 u 加入观察节点集合。在将 u 加入观察节点集合后，相应地更新其他节点 w 的覆盖面 $C(\{v,w\})$，继续选择新的观察节点，直到集合 Q 被覆盖率达到阈值δ为止。

综上所述，问题 1 的算法框架如算法 3 所示。

算法 3 RC-ONS(Rate Constrained Observation Node Selection)

Input：社交网络 G=(V,E)，源集合的大小 k，节点 v 被选为观察节点的开销 w(v)，v∈V，覆盖率阈值 δ

Output：观察顶点集合 O

Begin

1. O=Initialized observed set;
2. Q={{S_i,S_j}|1≤i<j≤m};q=|Q|;
3. Select an archor node v in O randomly;
4. V′=V\{v};P=Q;H(O)=0;
5. While H(O)≤δ do
6. Cover(v,V′,P);
7. Select u with the biggest value of |C({v,u})|/w(u);
8. O=O∪{u};V'=V\{u};P=P\C({v,u});H(O)=H(O)+|C({v,u})|/q;
9. End while;
10. Return(O);

End

算法 3 的复杂度分析：第 1 行调用算法 2 产生初始观察顶点集合 O，复杂度为 O(n)，n 为顶点的个数；第 2－4 行初始化和选择锚点，复杂度为 O(1)；第 5－9 行构建观察节点集合，其中第 5 行的 While 循环的复杂度为 O(m^2)，m 为源集合的个数；第 6 行调用算法 1，复杂度为 O(n)；第 7 行选取使得$|C(\{v,u\})|/w(u)$最大的顶点，复杂度为 O(n)。因此，算法 3 总的复杂度为 O(m^2*n)。其中 m 可以被看作常数，因此算法 3 的复杂度为 O(n)。

4.5 基于开销阈值的覆盖面最大的观察节点集选择算法

类似于算法 3，我们使用贪心算法对问题 2 求得开销不

超过阈值 B 且覆盖面最大的观察节点集合 O^*。问题 2 的算法框架如算法 4 所示。

算法 4 CC-ONS(Cost Constrained Observation Node Selection)

Input：社交网络 G=(V,E)，源集合的大小 k，节点 v 被选为观察节点的开销 w(v)，for all v∈V，开销阈值 B

Output：观察顶点集合 O

Begin

1. O=Initialized observed set;;
2. Q={{S_i,S_j}|1≤i<j≤m};
3. Select an archor node v in O randomly;
4. V′=V\{v};P=Q;cost(O)=0;
5. While cost(O)≤B do
6. Cover(v,V′,P);
7. Select u with the biggest value of |C({v,u})|/w(u);
8. O=O∪{u};V′=V\{u};P=P\C({v,u});cost(O)=cost(O)+w(u);
9. End while;
10. Return(O);

End

5 实验结果与分析

5.1 数据集

本文选择 3 种真实网络(Dolphins，Football，Facebook)进行模拟实验，以验证算法的有效性。

Dolphins 是一个描述海豚家族关系的网络，是由 Lusseau 等对栖息在新西兰 Doubtful Sound 峡湾的一个宽吻海豚群体进行了长达 7 年的观察而构造出的海豚关系网。网络中一个节点代表一个海豚，边表示两个海豚之间频繁接触。

Football 是一个美国橄榄球网络。在该网络中，每个节点代表参加美国 2000 年橄榄球赛季的一个高校代表队，连接两个节点之间的边则表示相应的两支球队之间至少进行过一场比赛。

Facebook 数据集由 Facebook 上的朋友列表组成。网络的节点为用户，链接用户的边表示他们的朋友关系。数据集通过将每个用户的 Facebook 内部 ID 替换为一个新值将其匿名化。数据集包括节点特征、社交圈子等。

表 1 列出了 3 种实际网络的主要的拓扑特性，如顶点的个数(N)、边的条数($|E|$)和顶点的平均度数($\langle d\rangle$)等。

表 1 真实网络参数描述

Table 1 Description of real network parameters

Name	N	$\|E\|$	$\langle d\rangle$
Dolphins	62	159	5.13
Football	115	613	10.66
Facebook	4039	88234	43.69

5.2 实验对比算法

实验中进行对比的算法如下：

1)随机算法(Random)：随机选择观察节点。

2)节点度算法(Degree)：选择传播网络中度最大的若干个节点作为观察节点。

3)介度算法(Betweenness)：选择传播网络中介度最大的若干个节点作为观察节点。

4)集合介度算法(CB)：一种基于最短路径的算法。通过计算每个节点经过其他任意两个节点的最短路径数量，再除以这两个节点之间最短路径数量，得到筛选值。选择传播网络中筛选值最大的若干个节点作为观察节点。

5)K 中位数算法(KMedian)[28]：选择传播网络中到其他所有节点距离最小的若干个节点作为观察节点。

6)距离方差区分算法(VD)：一种双区分算法，利用筛选每对节点和以及差的方差值来选择观察节点。

5.3 源定位算法

为了验证两种 ONS 算法(RC-ONS 和 CC-ONS)的有效性，我们在 3 个真实社交网络数据集上进行了测试，并和 6 种对比算法进行了性能比较。在使用每种算法找出观察节点集合后，我们使用同一种源定位算法找出 k 个源节点的集合，然后比较各种算法的观察节点所得到的源的精确度。在实验中，我们使用随机游走回溯算法进行源定位。考虑到观察节点成为源节点的可能性，我们定义了节点自身激活的概率。节点 v_i 自身激活的概率记为 p_{ii}，描述如式(7)所示：

$$p_{ii}=1-\max_{v_j\in AP_i} p_{ji} \tag{7}$$

其中，AP_i 代表节点 v_i 的被激活的父节点。然后，从观察集中的节点开始基于随机行走的回溯过程。用 C 表示候选源节点集，其最初就是观察节点集；然后从 C 中的每个激活节点开始随机行走。每次随机行走的过程描述如下。

1)在每个随机行走步骤中，为节点自身添加一条边。下一个可能的节点包括其活动父节点和自身。随机游走中的转移概率由式(8)和式(9)计算。

$$\begin{aligned}&p(A_{i\to i}\,|\,s_i=1)\\&=\frac{p(s_i=1|A_{j\to i})p(A_{i\to i})}{p(s_i=1|A_{j\to i})p(A_{i\to i})+\sum\limits_{v_k\in AP_i}p(s_i=1|A_{j\to i})p(A_{j\to i})}\\&=\frac{p_{ii}*\dfrac{1}{1+|AP_i|}}{p_{ii}*\dfrac{1}{1+|AP_i|}+\sum\limits_{v_k\in AP_i}p_{ki}*\dfrac{1}{1+|AP_i|}}\\&=\frac{p_{ii}}{p_{ii}+\sum\limits_{v_k\in AP_i}p_{ki}}\end{aligned} \tag{8}$$

$$p(A_{i\to i}\,|\,s_i=1)+\sum_{v_j\in AP_i}p(A_{j\to i}\,|\,s_i=1)=1 \tag{9}$$

其中，$A_{j\to i}$ 代表节点 v_j 尝试激活 v_i 的概率；$p(s_i=1|A_{j\to i})$ 表示父节点 v_j 试图激活 v_i 时，v_i 被激活的条件概率。在 IC 模型中，$p(s_i=1|A_{j\to i})$ 等于激活概率 p_{ji}，而 $p(s_i=1|A_{i\to i})$ 等于 v_i 激活自身的概率，表示为 p_{ii}。根据贝叶斯公式，不难得出后验概率的表达式(见式(8))。

2)用后验概率选择下一个节点，然后随机行走反向移动到所选节点。这个回溯过程就是一个反向扩散的过程。

3) 在候选集合中，每增加一个新的到达节点，就需要移除之前的节点。然后，从更新的候选集重复移动下一个随机游走步骤。当达到默认随机游走的步数或候选集不再变化时，上述过程停止。这样，候选集中剩余的节点被视为候选源节点。

独立地重复上述回溯实验，并记录一个节点被选为候选源节点的次数，称之为命中频率。最后按命中频率对所有节

点进行排序，选择前 k 个节点作为源节点。

5.4 源定位的结果比较的实验指标

实验中，使用每种算法找出观察节点集合，并使用上述源定位算法找出 k 个源节点的集合后，比较各种算法的观察节点所得到的源的精确度和距离误差，以此来衡量所找到的观察节点的质量。这里给出精确度和距离误差的两个标准。

1)距离误差。设 k 表示源的个数，d_k 表示所找到的源与真实源之间的最短距离。距离误差定义为：

$$Error\ distance=\frac{1}{k}\sum_{k}^{1}d_k$$

由上述定义可知，预测的源越接近真实的源，其距离误差就越小。最理想的情况是距离误差为零，这表示找到了真实的源。

2)精确度。设 n 表示实验次数，$N_c(i)$ 表示第 i 次实验时正确找到的真实源的个数，N_s 为真实源的个数。精确度的定义如下：

$$precision=\frac{n}{i=1}\sum_{n}\frac{N_c(i)}{N_s}$$

由上述定义可知，精确度越高表示找到的真实源的准确程度就越高。

5.5 阈值实验结果分析

5.5.1 不同成本阈值下的源定位平均距离误差和精确度的比较

本文在 3 个真实社交网络数据集上测试了 ONS 方法在不同成本阈值时的源定位结果的精确度。使用算法 CC-ONS 找出观察节点集合，设找到的观察节点的个数为 N，使用随机算法等对比算法，每个算法也产生个数为 N 的观察节点集合。然后使用同一种源定位算法找出 k 个源节点的集合，并比较各种算法的观察节点所得到的源的精确度。在不同的数据集中，由于其图形拓扑结构有较大差异，比如 Dolphins 只有数十个节点，而 Facebook 数据集有几千个节点，因此每个网络设置的预算也应该不一样。Dolphins，Football 以及 Facebook 预算上限分别为 16，24，100，每个节点的预算 $w(v)\in(0.3,0.5)$，每个节点分配发送谣言的时间 $t\in(1,5)$ 中的随机整数。k 和 m 的取值与节点数量有关，在小型数据集中，$k*m$ 的数值要接近网络节点数量以尽量覆盖网络。在大型数据集中，根据实际情况，我们不知道每个节点的特征，为了缩短算法运行时间，m 应取较小值。在真实数据集中的源定位精度如图 2 所示。

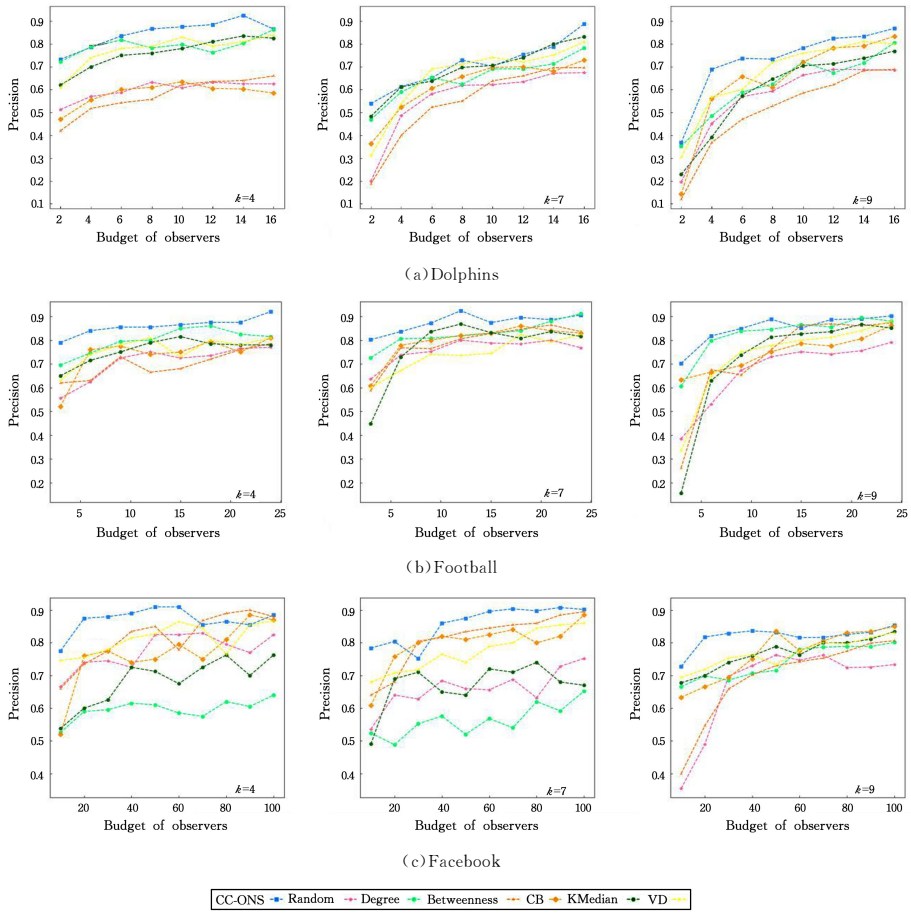

图 2 真实网络数据集上不同算法在不同预算阈值时的源定位精确度

Fig. 2 Source localization accuracies of different algorithms at different budget thresholds on real network datasets

图 2 中的实验结果表明，CC-ONS 在 3 种真实网络数据集上能显著提升源定位精度，尤其是在低预算阈值时。例如，在 Dolphins 数据集中，当观察节点预算阈值 $B<8$ 时，在 3 种 k 值下，CC-ONS 算法相较于其他算法源定位精度有较大提升；随着预算成本的上升，7 种算法的准确率都稳定上升。有趣的是，Random 算法在每个数据集中都不是表现最差的算法，

在一些度中心性明显的数据集中，Degree 算法的效果会略微超过 CC-ONS 方法。但在 Facebook 数据集中，当 $k=4$ 和 $k=7$ 时，Degree 算法的定位效果远远不如另外 6 种方法，甚至有较大波动；在 Dolphins 数据集中，当 $k=7$ 时，虽然 KMedian 算法和 VD 算法在几个成本阈值下精度略高，但这两种算法的时间复杂度远远高于 CC-ONS 算法；在 Dolphins 和 Football 数据集中，当$k=9$ 时，Degree 算法的精确度有时会略高于 CC-ONS 算法，这是由于这两个数据集的节点较少，度中心性比较明显。总体而言，CC-ONS 方法效果最好，CB 次之，VD 和 Degree 效果类似，Random 和 Betweenness 的效果最不理想。

我们还在 3 种真实数据集上测试了 CC-ONS 算法在不同预算阈值下的源定位结果的平均距离误差，实验结果如图 3 所示。从图中可以清楚地看出，在 Dolphins 数据集中，当预算阈值 $B\in(2,8)$ 的低预算区间时，CC-ONS 算法的距离误差达到了最好的效果，KMedian 和 CB 算法虽然在预算较高时才可以达到较好的效果，但在实际应用中，通常只能是低预算。在 Football 和 Facebook 数据集中也有类似的情形。虽然 Degree 算法在 Dolphins 和 Football 这类节点较少的数据集中的一些预算阈值下略微领先 CC-ONS 算法，但在 Facebook 这种大型数据集上的效果却是最差的。这是因为 Degree 算法十分依赖于网络节点的数量，并不适合大型数据集；CB 算法和 KMedian 算法虽然在高预算阈值下保持着和 CC-ONS 一样的距离误差，但它们都受限于时间复杂度，也不适合应用于大型数据集。

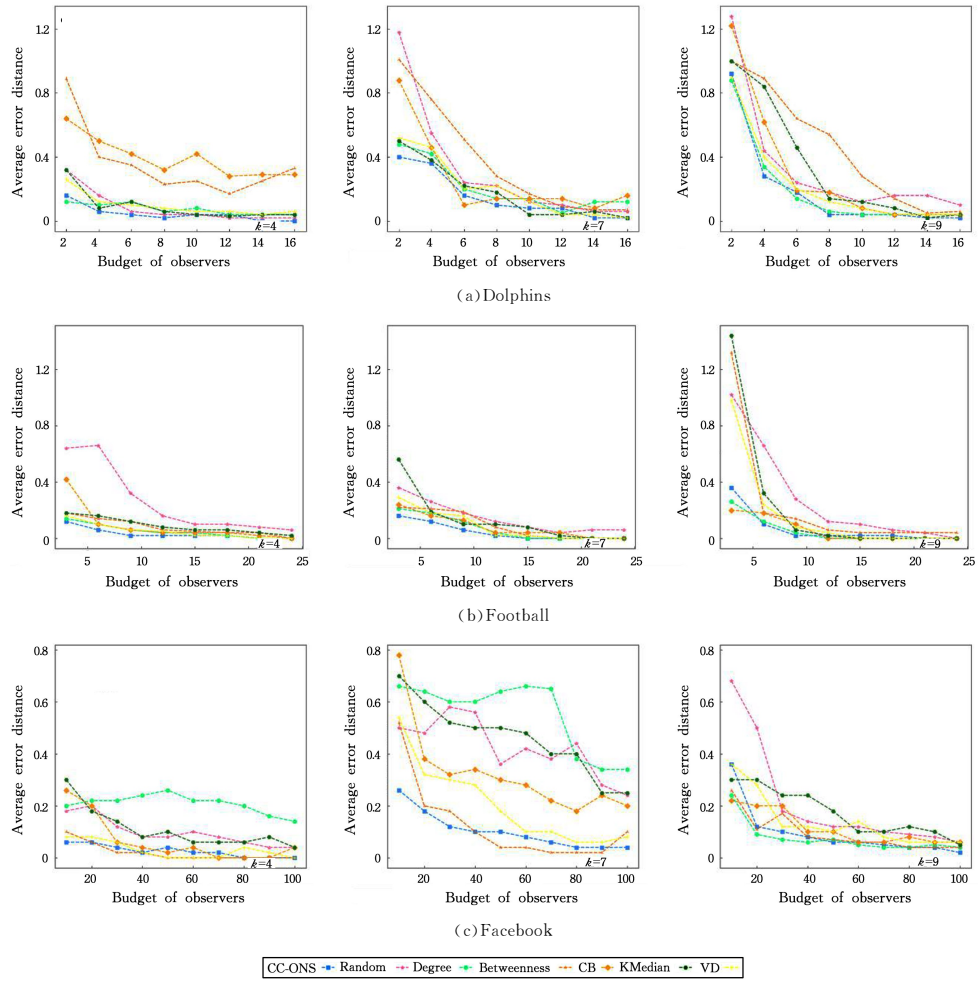

图 3 真实网络数据集上不同算法在不同预算阈值时的源定位平均距离误差

Fig. 3 Average error distance of source localization of different algorithms at different budget thresholds on real network datasets

5.5.2 不同覆盖率阈值下的源定位平均距离误差和精确度的比较

我们还在 3 个真实社交网络数据集上测试了 RC-ONS 算法在不同覆盖率阈值时的源定位结果的精确度。由于 RC-ONS 算法可以通过选择不同的源集合抽样个数 m 来均匀采样，因此降低了算法时间复杂度。在 Dolphins 数据集的实验中，当 $k=2$ 时，取源集合抽样个数 $m=30$；$k=3$ 时，取 $m=20$；$k=4$ 时，取 $m=15$。在 Football 数据集中，$k=3$ 时，取 $m=30$；$k=4$ 时，取 $m=25$；$k=5$ 时，取 $m=20$。在 Facebook 数据集中，在 3 种 k 值下，m 的值均取 15。实验结果如图 4 所示，从图中可以看出，随着覆盖率的提高，所有算法的源定位精确度总体上都有所提升，RC-ONS 算法在覆盖率阈值 $\delta\in(95,100)$ 时，定位精确度相比其他算法有较大的提升，可以看出，将所有源集合配对提升了源定位算法的精确度。在 Football 数据集中，虽然 Degree 算法和 CB 算法在 $\delta=100$ 时效果和 RC-ONS 方法差距不大，但是 RC-ONS 在较低覆盖率时就保持了较高的定位精确度，这说明 RC-ONS 可以在低预算的情况下取得较高的源定位精确度。

本文还在不同数据集上测试了源检测结果的平均距离误差，实验结果如图 5 所示。

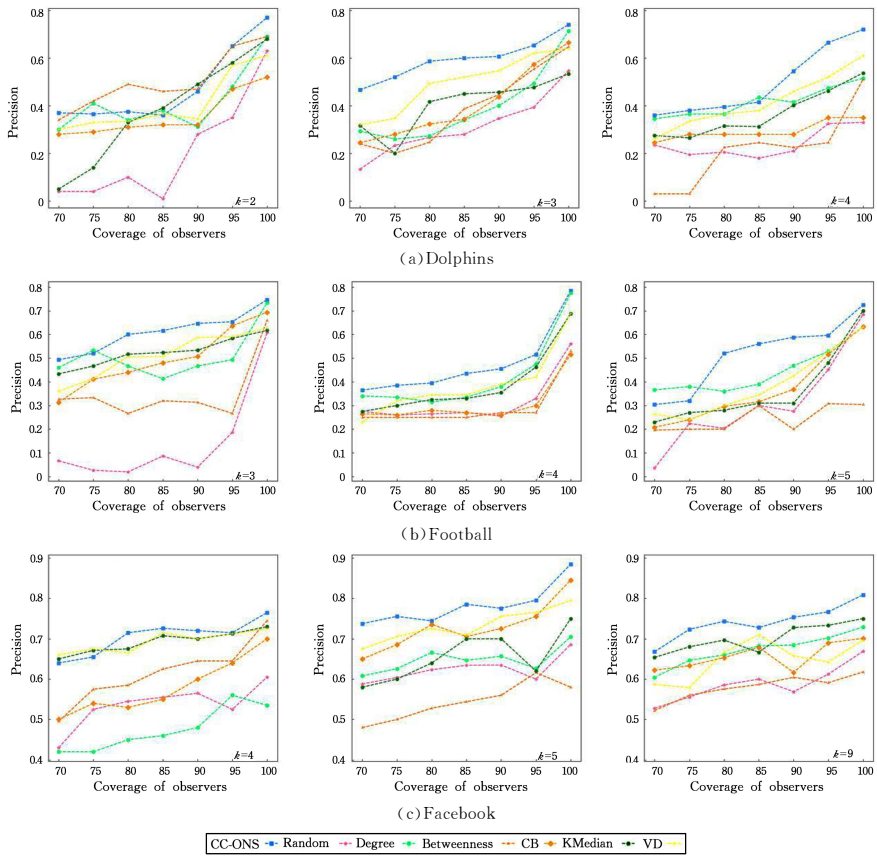

图 4 3 个网络数据集上不同算法在不同覆盖率阈值时的源定位精确度

Fig. 4 Source localization accuracies of different algorithms at different coverage thresholds on three network datasets

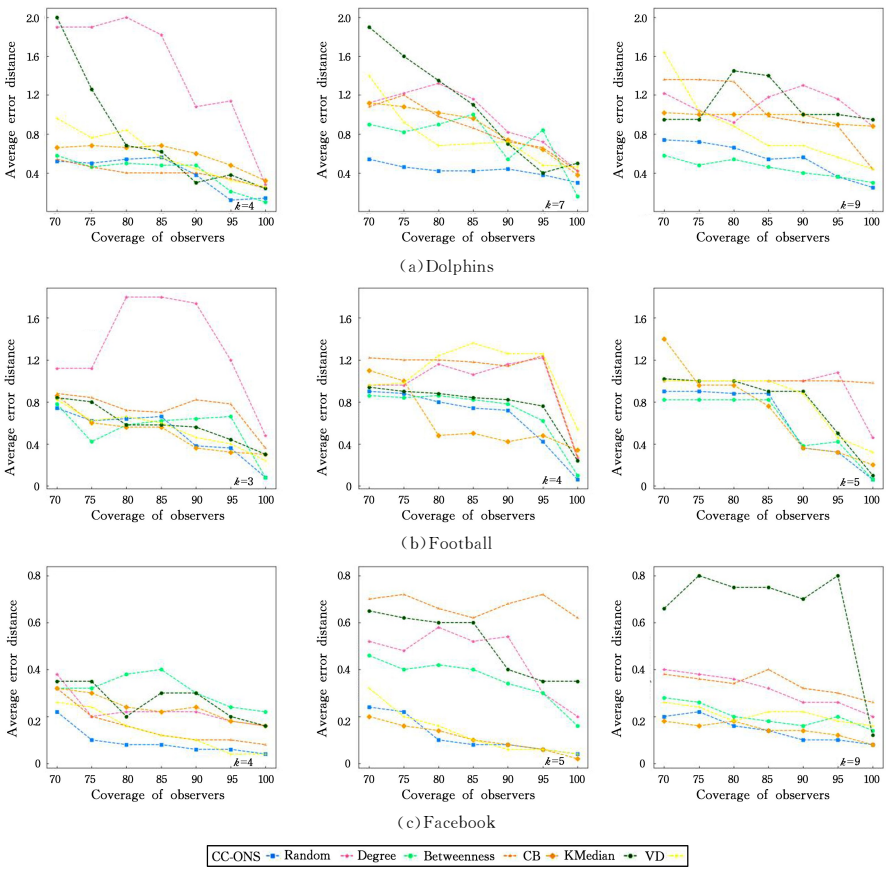

图 5 3 个网络数据集上不同算法在不同覆盖率阈值时的源定位平均距离误差

Fig. 5 Average error distance of source localization of different algorithms at different coverage thresholds on three network datasets

从图中可以明显地看出，随着覆盖率阈值 δ 的增加，7 种算法的平均距离误差总体上都在减少，RC-ONS 算法在不同的数据集上都展现出了很好的效果。在 Dolphins 数据集中，Degree 算法略优于 RC-ONS 算法，这是因为度中心性在节点较少的数据集中比较有优势。在实验中我们还发现，当覆盖率阈值 $\delta\in(90,100)$ 时，7 种算法的平均距离误差都大幅减小，这是全面覆盖的影响导致的，少数的边缘节点需要较多的观察节点覆盖，这也减小了平均距离误差。

结束语 在社交网络中，为了控制谣言的传播，我们往往要寻找负面信息的源头。目前，最有效的源定位方法是基于观察节点的方法，但是现有选择观察节点的方法都是预先设置观察节点数量，而没有根据图的拓扑特性来合理确定观察节点的个数。本文从节点预算阈值和节点的覆盖率阈值两个角度来研究观察节点的放置策略，将预算阈值和覆盖率阈值相结合提出了两类问题，并提出了两种基于 K-双区分集合的观察节点选择算法。所提算法通过贪心算法选择覆盖率和预算比值最大的节点来筛选出解决两类问题的观察节点。对所提方法进行了实验，结果表明其源定位精确度均高于其他算法。未来我们会在合成网络中进行实验，并根据覆盖率和成本阈值提出一种最小预算全覆盖算法，进一步证明选择的观察节点的有效性。

参 考 文 献

[1] KHULLER S, RAGHAVACHARI B, ROSFIELD A. Landmarks in grapghs[J]. Discrete Applied Mathematics, 1996, 70(3): 217-229.

[2] DEVARAPALLI R K, BISWAS A. Rumor Detection and Tracing its source to Prevent Cyber-Crimes on social Media[C]// Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques and Applications. 2021: 1-30.

[3] REN J, LIU M, LIU Y, et al. Optimal resource allocation with spatiotemporal transmission discovery for effective disease control[J]. Infectious diseases of poverty, 2022, 11(1): 34.

[4] LIU P, LI L, FANG S, et al. Identifying influential nodes in social networks: A voting approach[J]. Chaos, Solitons & Fractals, 2021, 152: 111309.

[5] SHAH D, ZAMAN T. Rumors in a Network: Who's the Culprit? [J]. IEEE Transactions on Information Theory, 2011, 57(8): 5163-5181.

[6] PINTO P C, THIRAN P, VETTERLI M. Locating the source of diffusion in large scale networks[J]. Physical Review Letters, 2012, 109(6): 68702.

[7] LI X, WANG X, HAO C, et al. Locating the Epidemic Source in complex Networks with sparse Observers [J]. Applied Sciences, 2019, 9(18): 3644.

[8] SPINELLI B, CELIS L E, THIRAN P. A General Framework for Sensor Placement in Source Localization[J]. IEEE Transactions on Network Science and Engineering, 2019, 6(2): 86-102.

[9] CELIS L E, PAVETIĆ F, SPINELLI B, et al. Budgeted sensor placement for source localization on trees[J]. Electronic Notes in Discrete Mathematics, 2015, 50: 65-70.

[10] LOKHOV A, MÉZARD M, OHTA H, et al. Inferring the origin of an epidemic with a dynamic message-passing algorithm[J]. Physical Review E, 2014, 90(1): 012801.

[11] WANG Z X, SUN C, RUI X, et al. Localization of multiple diffusion sources based on overlapping community detection[J]. Knowledge-Based Systems, 2021, 226: 106613.

[12] LIU Y, LI W, YANG C, et al. Multi-source detection based on neighborhood entropy in social networks[J]. Scientific Reports, 2022, 12(1): 1-12.

[13] ZHANG Z, YUE K, SUN Z, et al. Locating Sources in Online social Networks via Random walk[C]// 2017 IEEE International Congress on Big Data (Big Data Congress). IEEE, 2017: 337-343.

[14] BAO Z Q, CHEN W D. Rumor Source Detection in Social Networks via Maximum-a-Posteriori Estimation [J]. Computer Science, 2021, 48(4): 243-248.

[15] YUAN D Y, GAO J, YE M X, et al. Malicious Information Source Locating Algorithm Based on Topological Extension in Online Social NEtwork[J]. Computer Science, 2019, 46(5): 129-134.

[16] LI W, GUO C, LIU Y, et al. Rumor source localization in social networks based on infection potential energy[J]. Information Science, 2023, 634, 172-188.

[17] ZHANG X Z, ZHANG Y B, LV T Y, et al. Identification of efficient observers for locating spreading source in complex networks[J]. Physic A, 2015, 442: 100-109.

[18] PALUCH R, LU X, SUCHECKI K, et al. Fast andaccurate detection of spread source in large complex networks[J]. Scientific Reports, 2018, 8(1): 2508.

[19] PALUCH R, GAJEWSKI L, HOLYST G, et al. Optimizing sensors placement in complex networks for localization of hidden signal source: A review[J]. Future Generation Compute Systems, 2020, 112: 1070-1092.

[20] SPINELLI B, CELIS L E, THIRAN P. Observer placement for source localization: The effect of budgets and transmission variance[C]// 54th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2016. IEEE, 2016: 743-751.

[21] ZEJNILOVIC S, GOMES J, SINOPOLI B. Sequential observer selection for source localization[C]// 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE, 2015: 1220-1224.

[22] ZEJNILOVIC S, XAVIER J, GOMES J, et al. Selecting observers for source localization via error exponents[C]// 2015 IEEE International Symposium on Information Theory (ISIT), IEEE, 2015: 2914-2918.

[23] ZEJNILOVIC, GOMES J, SINOPOLI B. Network observability and localization of the source of diffusion based on a subset of nodes[C]// 51st Annual Allerton Conference on Communica-

tion, Control, and Computing (Allerton). IEEE, 2013: 847-852.
[24] GAJEWSKI L, PALUCH R, SUCHECKI K, et al. Comparison of observer based methods for source localization in complex networks[J]. Scientific Reports, 2022, 12: 5079-5079.
[25] CHEN X, HU X, WAN C. Approximation for the minimum cost doubly resolving set problem[J]. Theoretical Computer Science, 2016, 609: 526-543.
[26] WANG H J, ZHANG F F, SUN K J. An algorithm for locating propagation source in complex networks[J]. Physics Letters A, 2021, 393: 127184.
[27] ZHAO J, HAO K, CHEONG H. Early identification of diffusion source in complex networks with evidence theory[J]. Information Science, 2023, 642: 119061.
[28] BERRY J, HART, W, PHILLIPS C, et al. Sensor placement in municipal water networks with temporal integer programming models [J]. Journal of Water Resources Planning and Management, 2006, 132(4): 218.

CHEN Zhangyuan, born in 2000. His main research interests include data mining and complex network.

CHEN Ling, born in 1951, professor, Ph.D supervisor, is a member of CCF (No. 85182M). His main research interests include data mining complex network and artificial intelligence.

（责任编辑：杨雪敏）

2025CCF 分部工作计划会在厦门召开

2025 年 3 月 15 日至 16 日，2025CCF 分部工作计划会在厦门召开，CCF 厦门会员活动中心承办。来自全国 40 余个城市会员活动中心（分部）、会员与分部工委、秘书处等 90 余人齐聚一堂，共同回顾 2024 年会员与分部工作，对 2025 年主要工作和方向进行了介绍和讨论。会议由 CCF 会员与分部工委主任李贝主持。

会上李贝介绍了 2024 年会员与分部工作情况。2024 年获得优秀分部及单项奖的分部进行了分享。

CCF 秘书长唐卫清针对分部工作的目标和未来发展方向做了讲解。他强调，2025 年 CCF 将以提升服务质量、深化国际影响为核心目标，重点推进会员留存率优化、数字化服务升级及计算机博物馆建设等。

CCF 副秘书长、业务总部总经理束庆山就苏州业务总部开展的活动情况及 CCDE 内容做了详细介绍，通过认证体系、科创赛事、产学合作及高端会议，持续推动计算领域技术转化、教育普及与产业融合，2025 年将进一步深化生态布局，服务国家数字化战略。

CCF 副秘书长、东阳西西艾弗计算文化与发展研究院院长臧根林详细介绍了计算机博物馆进展情况。目前博物馆主体结构已封顶，计划 2026 年夏季开馆。目前征集藏品超 4600 件，其中包括“两弹一星”项目用过的磁鼓、天河一号超级计算机、硅晶柱等珍贵展品。

CCF 助理秘书长、东北办事处主任陈国威就“2024 年各办事处工作介绍及 2025 年主要工作计划”“CCF 教师能力提升计划”“第四届 CCF 东北论坛”分别做了介绍。

CCF 会士、常务理事、公益工委主任卜佳俊对 CCF 公益日及相关品牌活动做了介绍。CCF 将持续开展多种形式的公益活动，让更多的会员有用技术“做好事、行善举”的公益意识，提升学会的社会美誉度。

CCF 会员部主任富蕾介绍了 CCF 分部活动情况与新增权益。CCF 通过优化分部权益方案，加大奖励力度，同时强化活动审核力度，推动分部高质量完成会员发展与服务目标，进一步凝聚地方会员力量。

CCF 太原主席赵鹏做了第三届 CCF 黄河科技大会（CHHC 2025）筹备情况介绍，大会将以“智汇产业，数聚生态”为主题，通过高端论坛、产业对接、产教融合活动，打造“计算＋文化”的产业生态盛宴，助力区域数字经济与科技创新。

与会人员围绕着“科协新规下分部发展会员的挑战与应对”、“分部的组织能力建设”等两个议题展开了热烈的研讨与交流。各位发言人积极献言献策，针对会员发展工作提出了切实可行的思路和具体措施。

本次会议通过两天的研讨与交流，凝聚共识、激发创新。随着计算机博物馆开馆在即、公益服务纵深推进、教师培训全面铺开，CCF 各分部正以更开放的姿态服务会员、更丰富的会员活动吸引会员，为学会的高质量发展注入新动能。

据 CCF 微信公众号