ISSN 1000-0054 清华大学学报（自然科学版） 2017 年 第 57 卷 第 12 期 3/16
CN 11-2223/N J Tsinghua Univ (Sci & Technol), 2017, Vol. 57, No. 12 1245-1253

在线社会网络中面向节点影响力的信息传播阻断模型

赵 宇[1,2]， 黄开枝[1,2]， 郭云飞[1]， 赵 星[1,2]

（1. 国家数字交换系统工程技术研究中心，郑州 450002；2. 移动互联网安全技术国家工程实验室，北京 100876）

摘 要：目前信息传播阻断模型是在网络中选择并删除 l 个最佳节点（边）使信息传播到的节点数量最小，该模型未考虑信息传播节点的影响力，导致选择的 l 个最佳节点（边）并不准确，阻断有效性较差。针对此问题，该文提出一种面向节点影响力的信息传播阻断模型，并设计了一种基于采样平均近似的求解方法。模型以网络中节点的影响力为有效性依据，通过选择并删除 l 个最佳节点来改变网络结构，使信息传播到的目标节点影响力之和最小；该模型为随机优化问题，首先利用采样平均近似将目标函数转化为确定性问题，其次进一步编码为混合整数规划问题，最后采用一种量子遗传算法解决该问题得到 l 个最佳节点并将其删除。仿真结果表明：相比于传统模型，通过本模型选择的 l 个最佳节点能够将信息传播的影响力控制在更小的范围，且处理时间更短。

关键词：在线社会网络；信息传播阻断；影响力最小；随机优化；混合整数编码

中图分类号：TN915.81 **文献标志码**：A

文章编号：1000-0054(2017)12-1245-09

DOI：10.16511/j.cnki.qhdxxb.2017.25.061

Information diffusion blocking model of node influence-oriented in online social network

ZHAO Yu[1,2], HUANG Kaizhi[1,2], GUO Yunfei[1], ZHAO Xing[1,2]

(1. National Digital Switching System Engineering and Technological R & D Center, Zhengzhou 450002, China;
2. National Engineering Laboratory for Mobile Network Security, Beijing 100876, China)

Abstract: Information diffusion blocking maximization is used to select and delete the best l nodes (edges) to minimize the number of nodes receiving information in the network. However, the model does not take into account the node's influence which blocks the information flow and lowers the efficiency. This paper presents an information diffusion blocking model that considers the node's influence with a method based on the sampling average approximation (SAA). The model is selects and deletes the best l nodes to change the network structure which minimizing the influence of the target nodes. The model is a stochastic optimization problem which is transferred into a deterministic problem using SAA. The problem is then encoded as a mixed integer programming (MIP) problem. Finally, a quantum genetic algorithm is used to select the best l nodes and remove them. Simulations show that the best l nodes selected by this model influence the information diffusion over a smaller range and the processing time is shorter than the traditional model.

Key words: social network; information diffusion blocking; minimum influence; stochastic optimization; mixed integer programming (MIP)

以微信和微博为代表的在线社会网络已经成为人们日常交流的重要工具，是民意集中表达与反映的平台。在给人们获取信息带来便利性的同时，该平台上也传播着大量有害信息，给人们正常生活造成了不良影响，甚至影响社会和谐，危害国家安全。在线社会网络具有规模大和结构复杂等特点，很难根除有害信息的产生，通过改变网络结构等方式来阻断信息传播是目前可行的解决途径，因此，对信息传播阻断方法的研究已经成为热点[1-2]。

信息传播阻断模型主要在信息传播源点数量和位置确定的条件下，研究选择并删除 l 个节点或边，使信息传播到的节点数量最少。目前信息传播阻断问题的研究主要分为 2 类：一类是减小邻接矩阵最大特征值使得信息传播的最少节点数量低于爆发门限，2012 年 Prakash 等[3]提出了消息大规模传播门限理论，证明了消息大规模传播条件主要由网络结构邻接矩阵的最大特征值和感染率决定，抑制信息传播只需使邻接矩阵的最大特征值（谱半径）减少到爆发门限以下，该结论为后续信息传播阻断研究提供了理论依据。如何通过删除节点或边最快地

收稿日期：2017-04-24

基金项目：国家“九七三”重点基础研究项目(2016YFB0801605)；
国家自然科学基金资助项目(61521003)

作者简介：赵宇(1984—)，男，博士研究生。

通信作者：黄开枝，教授，E-mail：huangkaizhi@tsinghua.edu.cn

减小谱半径是 NP-complete 和 NP-hard 问题[4]，文[5]以删除点或者边的代价最小为目标，设计了一种减小谱半径的贪心游走算法，得到近似度较高的解。文[6]基于谱半径提出描述节点阻断信息传播能力的概念 Shield Value，基于此概念设计了满足子模型特征的阻断函数，并提出了平衡优化质量和时间复杂度的 NetShield+算法。另外一类是以信息传播到的节点数量最小为直接目标，Khalil 等[7]以边活跃图模型为基础，得出通过删除边使信息传播范围最小化问题满足超模型特征的结论，基于该特征设计了有效的数据结构和最优近似算法，其阻断效果优于启发式算法。此外，Zhang 等[8]将删除对象调整为以群组为单元，通过删除或者免疫最佳的群组达到最佳阻断效果。

当前信息传播阻断问题研究的不足之处是其模型只研究传播到节点的数量，并未考虑节点间影响力的差异，导致阻断目标不精确，信息传播阻断的有效性较差。例如，对不健康信息进行阻断时，更应该考虑该信息对不同人群的危害，而不是只考虑传播的范围，如该信息在青少年人群中传播的影响可能远大于在老年人中传播的影响；对国际政治谣言进行阻断时，该谣言对于高级官员来说更加敏感，若传播到了偏远山区，即使接收人数较多，其产生的影响可能也相当有限。因此，阻断模型的目标并不应只局限于信息传播到的范围，更应考虑对网络节点产生的总体影响。

为此，本文提出了一种面向节点影响力的信息传播模型，并设计了一种基于采样平均近似的求解方法。在该模型中，阻断的有效性是以信息传播到的目标节点影响之和最小为目标，为此，模型引入了单个节点的影响力权值，通过选择并删除 l 个最佳节点，达到信息传播的有效阻断。

1 问题描述

在线社会网络用 $G=(V, E)$ 来表示，其中，V 和 E 分别代表网络中所有节点和所有边的集合。信息传播采用独立级联模型[9]，即接收到信息的节点通过连接边将信息传播给其邻居节点，节点接收该信息并传播的概率为 p_e，每条边的传播过程相互独立。在线社会网络中节点对同一条信息通常只转发一次，采用递进式的模型来描述此现象，即若节点已进入转发消息状态后将不会再回到等待接收消息状态，此条件保证节点的状态不会回退，而且即使网络结构中存在环路，传播路径也不会出现环路。

如图 1 为独立级联的传播过程，在 t_0 时刻，信息传播的源点节点 1 和 2 通过连接边将消息传播给邻居节点 3、4、5，其中节点 3 和 5 接收并转发该消息，用方块来表示，节点 4 没有继续转发该消息，用圆来表示；在 t_1 时刻，与节点 3 和 5 连接的节点为 6 和 9，其中节点 6 接收并转发该消息，用三角来表示，这一时刻节点 1 和 2 不会再接收并转发此消息；在 t_2 时刻，节点 6 转发该消息，但是在该信息传播过程中没有节点继续接收，这一传播过程最终接收消息的节点为 3、4、5、6。

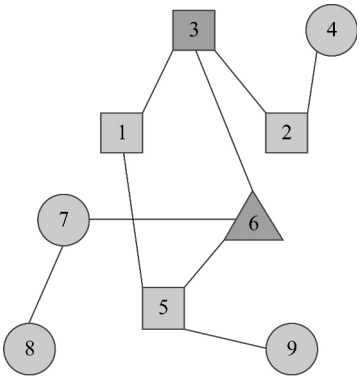

图 1 独立级联模型示意图

信息传播过程是在源点数量和位置固定的条件下，信息以独立级联的方式在网络中传播，最终传播到的节点数量用 $\sigma(G)$ 来表示：

$$\sigma(G)=\sum_{v\in T}x_v(G). \tag{1}$$

其中：$x_v(G)$ 表示图 G 中目标节点集 T 中的节点 v 感染概率，是 0～1 间的随机变量，其分布受独立传播概率和网络结构影响；目标节点集 $T\subseteq V$，V 为所有节点集。

信息传播阻断目标函数是研究删除哪些节点或者边使信息传播的范围最小：

$$\min_{l}(\sigma(G_l)). \tag{2}$$

其中：l 表示删除的节点集或者边集，G_l 表示删除节点集或者边集 k 后的网络结构。

目前信息传播阻断模型只以信息传播到的节点数量为目标，阻断的目标与实际需求存在较大偏差，阻断有效性较差。具体如图 2 所示：节点用圆点表示，圆点的大小代表节点影响力的大小，设 N_5 影响力为 $\alpha>1$，其他节点影响力为 1。假设信息源点传播的信息会被每一个节点接收并转发，若选择并删除 1 个最佳节点使得阻断效果最为有效，则发现不同的有效性目标会选择不同的删除节点。如果以阻断信息传播节点的数为目标，则选择删除 N_1

节点为最佳；如果以阻断信息传播节点影响力总和为目标，则选择删除 N_4 节点为最佳；进一步如果将节点的影响力大小都设置为 1，则信息传播阻断中以节点数量最小化为目标，就是以节点影响力总和最小化为目标的特例。

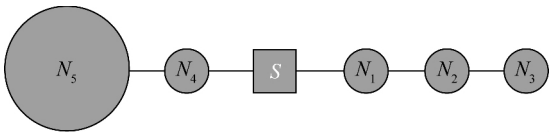

图 2　信息阻断有效性示意图

针对上述问题，本文引入单个节点的影响力权值，建立了面向影响力的信息传播阻断模型，并明确了信息传播的有效性以影响力总和为目标；然后，设计了一种基于采样平均近似的方法对该模型进行求解，选择并删除 l 个最佳节点，使信息传播阻断更为有效。

2　面向影响力的信息传播阻断模型

2.1　信息传播影响力定义

本文引入了单个节点的影响力权值 $\beta_v(G)$，该值与节点特征、网络结构和信息属性等内容相关，影响力的权值可根据实际情况灵活定义。如考虑不健康信息对青少年的影响，则该权值应该依据不健康信息对青少年人群节点产生的危害定义；如考虑谣言在异构网络中的传播范围，则该影响力权值应该以节点在网络中的连接性和中心性等条件为依据；影响力权值 $\beta_v(G)$ 可采用文[10-13]等的研究成果，由于其取值对建立信息传播阻断模型并无本质影响，故不针对影响力权值展开研究。基于单个节点的影响力权值和上述信息传播过程，定义网络中一组初始信息传播源点的最终影响力为

$$\sigma(G)=\sum_{v\in T}x_v(G)\beta_v(G). \tag{3}$$

其中：目标节点集合 T 通常情况下可设置为所有节点 V，针对特殊应用场景也可能存在关注特定节点集合的情况，此时设置 T 是所有节点 V 的子集。此外，本模型中初始信息传播集合 I 为固定值，并未在式(3)中体现。该信息传播影响力的定义具有一定的普适性，若将影响力权值都设置为 1，该定义就简化为信息传播的节点范围的定义。

2.2　模型建立

建立面向节点影响力的信息传播阻断模型：在网络中存在固定初始传播源点集 I，通过选择并删除 l 个最佳节点的方式改变网络结构，使得信息传播的节点影响力之和最小。在本模型中删除节点方法更为灵活，令删除节点动作作为操作管理动作集 $A=\{1,2,\cdots,L\}$，其中单个操作动作的对象根据需求可以设定为单个节点或一组节点 V_l，并要求各操作管理动作面向对象无交集且相互独立，则网络中所有节点集 V 可以表示为 $V=V_0\cup\left(\bigcup_{l=1}^{L}V_l\right)$，其中 V_0 是所有删除操作中没有包括的节点。

为了便于表示操作管理动作的执行情况，令向量 $\boldsymbol{Y}_L$ 为具体的操作策略，该向量表示管理动作集中删除动作 $\boldsymbol{y}_l$ 的执行情况，$\boldsymbol{y}_l$ 为 0～1 向量，$\boldsymbol{y}_l$ 取 1 时代表执行了对应的删除动作 l，$\boldsymbol{y}_l$ 取 0 时代表该删除动作 l 没有执行。另外，删除网络中的节点对网络结构造成了破坏，为每一个操作设定一定代价，令 c_l 表示删除节点动作 l 的代价，操作动作的总代价限制在一定阈值 C 以内。因此，基于节点影响力的信息阻断目标函数如下所示：

$$\begin{aligned}&\min_{\boldsymbol{Y}_L}\quad \sigma(G(\boldsymbol{Y}_L)),\\&\text{s.t.}\quad \sum_{l=1}^{L}c_l\boldsymbol{y}_l\leqslant C.\end{aligned} \tag{4}$$

其中：$G(\boldsymbol{Y}_L)$ 为执行了 $\boldsymbol{Y}_L$ 策略的删除动作后的网络结构图；$\sigma(G(\boldsymbol{Y}_L))$ 表示在 $G(\boldsymbol{Y}_L)$ 结构下，最终感染目标节点影响力之和，

$$\sigma(G(\boldsymbol{Y}_L))=\sum_{v\in T}X_v(G(\boldsymbol{Y}_L))\beta_v(G(\boldsymbol{Y}_L)). \tag{5}$$

在该模型中，求解目标函数得到的策略 $\boldsymbol{Y}_L$，通过执行策略 $\boldsymbol{Y}_L$ 删除 l 个最佳节点，便达到信息传播阻断效果最佳的目的。

2.3　模型性质分析

如何确定 l 个最佳节点并删除使信息传播影响力最小是 NP hard 问题[4]，通过对式(4)的分析发现其不满足子模和超模特征，具体见性质 1 和性质 2。

性质 1　传播影响力最小目标函数不满足子模特征。

证明：如图 3 所示，为便于量化，将删除节点的代价设置为 1，则删除节点的总代价直接对应着删除节点的数量。

一个函数具备子模型特征的定义如下：对于集合 $S\subseteq R\subset E$，$e\in E\backslash R$，若满足式(6)，则目标函数具备子模特征。

$$f(S\cup\{e\})-f(S)\geqslant f(R\cup\{e\})-f(R). \tag{6}$$

子模特征的直观解释是集合 R 增加一个元素 e 的边界收益要不大于其任何一个子集 S 增加一个

元素 e 的边界收益。

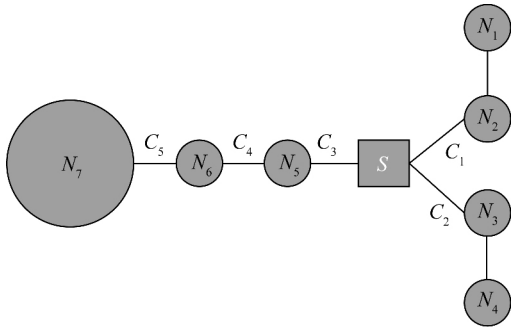

图 3 目标函数不满足子模特征示意图

采用反证法举例说明式(4)不满足子模型特征。首先构造删除节点的集合并计算删除节点后信息传播源点在整个网络中的影响力，假设删除节点的较小节点集 S 为单点 N_7，即 $S=\{N_7\}$，为了与子模函数一致，用 $f(x)$ 表示 $\sigma(G)$，那么 $f(S)=I-\alpha$，其中 I 是源点在整个网络中传播的最终影响力；接下来确定删除节点的较大集合 $R=\{N_6, N_7\}$，则删除节点集合 R 后的影响力为 $f\{R\}=I-\alpha-1$；选择删除的增量节点为 N_5，即 $e=N_5$。此时，

$$f(S\cup\{e\})-f(S)=-2,$$

而

$$f(R\cup\{e\})-f(R)=-1,$$

即

$$f(S\cup\{e\})-f(S)<f(R\cup\{e\})-f(R),$$

不满足式(6)，证毕。

性质 2 传播影响力最小目标函数不满足超模特征。

证明：若该目标函数具有超模特征，则对于集合 $S\subseteq R\subset E$，$e\in E\backslash R$，满足下式即可：

$$f(S\cup\{e\})-f(S)\leqslant f(R\cup\{e\})-f(R). \tag{7}$$

超模特征的直观解释是集合 R 增加元素 e 的边界收益要大于等于其任何一个子集 S 增加元素 e 的边界收益。

以图 4 为例，用反证法举例说明式(4)不满足边界收益减少特征。首先构造删除节点的集合并计算删除节点后源点传播信息后的影响力，假设删除的较小节点集 S 为单点 N_2，即 $S=\{N_2\}$，用 $f(x)$ 表示 $\sigma(G)$，那么 $f(S)=I-2$；然后选择删除节点的较大集合为 $R=\{N_2, N_5\}$，则删除节点后的影响力为 $f\{R\}=I-3$；选择删除的增量节点为 N_6，即 $e=N_6$。此时，

$$f(S\cup\{e\})-f(S)=-1,$$

而

$$f(R\cup\{e\})-f(R)=-1-\alpha,$$

即

$$f(S\cup\{e\})-f(S)>f(R\cup\{e\})-f(R),$$

不满足式(7)，证毕。

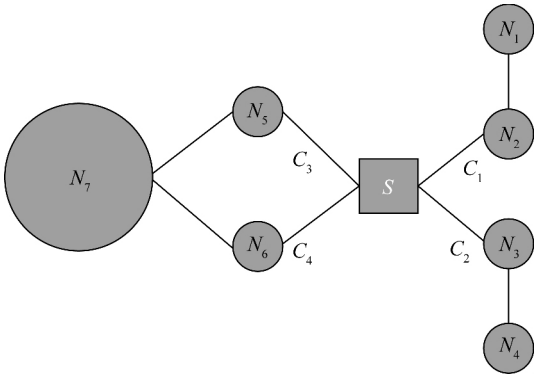

图 4 目标函数不满足子模和超模特征示意图

性质 3 贪心算法不能保证传播影响力最小目标函数的解近似最优。

证明：采用贪心算法的依据是目标函数满足子模或者是超模特征，若满足，则贪心算法就能够达到近似比为$(1-1/e-\varepsilon)$近似最优解。性质 1 和性质 2 证明了式(4)不满足子模和超模特征，因此采用贪心算法可能会出现非常差的结果。例如在图 2 中，假设通过删除 2 个节点使影响力最小，若采用贪心算法则采用每一轮删除一个影响力最佳的节点，最终结果为 N_2、N_4 2 个节点，删除节点后的影响力为 $I-4$；而全局最优解是删除节点 N_5、N_6，删除节点后的影响力为 $I-2-\alpha$，此时，α 的值越大，采用贪心算法的结果就越差，无法使用贪心算法得到近似最优解，为此本文设计了一种基于采样平均近似的方法对该模型进行求解。

3 基于采样平均近似的求解方法

信息传播阻断模型中的影响力总和为随机变量，模型为随机优化问题。因此，本文基于采样平均近似[14]的方法解决该问题。

3.1 采样平均近似

在复杂的网络结构中，即使消息发布源点已知，推测信息在整个网络的传播过程也是困难的，但表示节点感染概率的随机变量 $X_v(G)$ 在网络空间中的分布概率是确定的，该概率分布不依赖具体选择的删除节点策略。随机变量 $X_v(G)$ 的具体采样值可以通过一次信息传播过程来确定，信息传播路径所组成的网络即为传播视图 G'，为了能够快速地

得到网络视图，可采用文[9]提出的翻硬币方法，该方法假设每条边相互独立并依照一定概率 p_e 传播信息，传播信息的边连接而成传播路径，最终由传播路径连接而成传播网络。该信息传播网络结构即是快速生成的网络视图 G'，文[9]证明了翻硬币方法与级联传播模型得到的传播效果一致。

生成网络视图的过程中有两类比较特殊的节点：一类是在翻硬币过程中信息几乎没有传播到的节点，另一类是在信息传播中几乎每次都会传播到的节点。这两类节点不仅与每条边的独立传播有关，而且主要取决于网络的结构：几乎每次都会传播到的节点与源点间存在多条路径，信息传播到该节点的可能性非常高；而几乎没有被传播到的节点与源点的路径较远有关。那么可以对网络进行优化处理，剔除每次都传播不到的节点和压缩每次都会传播到的节点集。优化后的网络结构减少了处理对象，可降低处理时延。对网络 G 执行翻硬币方法 N 次，得到 G'_1，G'_2，…，G'_N 的网络视图集，该网络视图集作为训练视图集。在训练视图集中考虑采用 y 策略后，视图 $G'_k(y)$ 中节点 v 的影响力用确定值 $\upsilon_v^k(y)$ 表示。那么式(4)通过 SAA (sample average approximation)方法可得到：

$$\begin{aligned}&\min_{y} \quad \frac{1}{N}\sum_{k=1}^{N}\sum_{v\in T}\beta_v\upsilon_v^k(y),\\&\text{s.t.} \quad \sum_{l=1}^{L}c_l y_l \leqslant C.\end{aligned} \tag{8}$$

接下来分析采样结果与真实结果的差别。当训练视图数量 $N\to\infty$ 时，SAA 的结果会收敛于式(3)，当 N 的采样规模较小时，SAA 的结果并不是最优解，文[15]对随机路由问题进行了分析，该结论适用于本算法，对基于 SAA 的阻断算法的结果偏差进行分析如下。

对式(3)进行 M 次独立采样，每次采样的训练视图为 N 个。采样后会产生 $\hat{y}_1$，$\hat{y}_2$，…，$\hat{y}_m$ 个执行策略备选方案，对应着式(3)的目标结果为 Z_1，Z_2，…，Z_m，令

$$\bar{Z}=\frac{1}{M}\sum_{m=1}^{M}Z_m. \tag{9}$$

其中，$\bar{Z}$ 为 M 个样本 SAA 问题目标函数的最优平均值。$E[\bar{Z}]\leqslant\mathrm{OPT}$，OPT 为式(3)影响力最小化问题的最优解，那么 $\bar{Z}$ 成为式(3)最优解下界的统计估计量。

令 $\hat{y}$ 是式(3)的一个可行解，通常是一组规模为 N' 的采样视图的最优执行策略，则目标函数 $\sigma(G(\hat{y}))$ s.t. $\sum_{l=1}^{L}c_l y_l\leqslant B$ 是 OPT 的上界，上界的估计值在约束 $\sum_{l=1}^{L}c_l y_l\leqslant B$ 条件下为 $\sigma(G(\hat{y}))$，则最优解的上界与下界如下：

$$E[\bar{Z}]\leqslant\mathrm{OPT}\leqslant E[Z(\hat{y})]. \tag{10}$$

$E\left[Z(\hat{y})-\bar{Z}\right]$ 是最优解之差 $\mathrm{OPT}-\bar{Z}$ 的上界，$Z(\hat{y})-\bar{Z}$ 是最优解上界的无偏统计量。

3.2　混合整数规划编码

为求解确定性最优化问题，将式(8)编码为混合整数规划问题。当对网络视图进行删除节点操作时，$\upsilon_v^k(y)$ 会受到网络结构变化而产生变化，利用变量 x_v^k 替换 $\upsilon_v^k(y)$，从而将传播接收情况扩展到原有概率空间。编码后的混合整数规划目标函数如式(11a)所示，其中对于单个节点 v，从该节点到源点路径上的所有节点共计 M_A 个，其删除动作集合用 $A(v)$ 表示。

$$\min_{Y}\max_{X} \quad \frac{1}{N}\sum_{k=1}^{N}\sum_{v\in T}\beta_v x_v^k; \tag{11a}$$

$$\text{s.t.} \quad \sum_{l=1}^{L}c_l y_l\leqslant C; \tag{11b}$$

$$x_v^k\leqslant r\left(1-\frac{1}{M_A}\sum_{l\in A(v)}y_l\right),\ \forall v\notin V_0,\ \forall k; \tag{11c}$$

$$x_v^k\leqslant\sum_{(u,v)\in E_k}x_u^k,\quad \forall v\notin S,\ \forall k; \tag{11d}$$

$$0\leqslant x_v^k\leqslant 1,\ y_l\in\{0,1\}. \tag{11e}$$

其中，$r=\begin{cases}1, & \text{if } x_v^k\in y_l\odot V_l;\\0, & \text{if } x_v^k\notin y_l\odot V_l.\end{cases}$

目标函数明确后，建立删除操作策略与节点信息接收率的约束关系，最终编码为混合整数规划问题。在网络视图 G'_k 中没有直接删除节点 v 的情况下，其删除动作减少了从源点到目标节点 v 的传播路径数，降低了目标节点的影响概率，其影响关系如式(11c)所示，如果直接删除节点 v，则该节点的信息接收率值为 0，该动作由 r 控制；另外，针对所有节点，根据翻硬币的规则，传播视图中感染的目标节点必须与源点存在通路，即信息接收节点有已经接收并转发了信息的邻居节点，如式(11d)所示，由此建立起变量 x_v^k 和策略 y_l 的线性关系。

3.3　量子遗传算法

为快速和准确地解决节 3.2 中编码后的混合整数规划问题，可采用具有并行计算能力和全局最优解特征的智能算法，其中遗传算法是一种应用比较广泛的智能优化算法。Narayana[16] 为了提高遗传

算法的寻优能力，首次将量子计算理论与遗传算法进行结合，提出了量子遗传算法。江逸茗等[17]将量子遗传算法用于解决网络虚拟化环境下的监控问题，该量子遗传算法同样适用于求解节 3.2 中描述的混合整数规划问题，从而得出最佳的 l 个节点，具体步骤如下。

步骤 1：初始化。

量子比特状态为处于$|0\rangle$态、$|1\rangle$态以及$|0\rangle$和$|1\rangle$之间的任意叠加态，对应目标节点的状态为被删除，保留两者的叠加态可描述为

$$|\Psi\rangle=\alpha|0\rangle+\beta|1\rangle. \tag{12}$$

其中：α、β 是复数；$|\alpha|^2$ 和 $|\beta|^2$ 分别表示量子比特被观测为$|0\rangle$和$|1\rangle$态的概率，且两者各为 1。

量子遗传算法的运算对象的可行解可以看作是个体的染色体，每个染色体由多个量子比特组成，一个量子比特的概率幅可以定义为$[\alpha\ \ \beta]^{\mathrm{T}}$，而一个由 L 位量子比特组成的染色体的编码形式为

$$q=\left[\begin{array}{c|c|c|c}\alpha_1 & \alpha_2 & \cdots & \alpha_L\\ \beta_1 & \beta_2 & \cdots & \beta_L\end{array}\right]. \tag{13}$$

一个染色体可以同时描述 2^L 个状态，即覆盖了删除操作策略的所有空间，在观测时染色体将坍缩为一个确定的状态，即确定了删除的 l 个节点。

步骤 2：适应度计算。

在确定了染色体的编码以后，对染色体进行测量，方法是为每一个量子比特都生成一个随机数，若该随机数小于$|\alpha|^2$，则该量子比特位的测量值为 0，否则为 1；然后计算其适应度，个体执行策略 $\boldsymbol{Y}=[y_1,\ y_2,\ \cdots,\ y_L]$，可将式(11a)进行变换得出适应度函数：

$$\mathrm{Fit}(\boldsymbol{Y})=\left[\max\frac{1}{M}\sum_{k=1}^{M}\sum_{v\in T}x_v^k\right]^{-1}\varepsilon\left(\sum_{l=1}^{L}y_l-B\right). \tag{14}$$

其中：x_v^k 的取值由策略 $\boldsymbol{Y}$ 的具体取值和式(11b)等约束条件决定；$\varepsilon\left(\sum_{l=1}^{L}y_l-B\right)$为阶跃函数，适应度函数为 0。

步骤 3：量子旋转门。

为了对种群进行更新，采用量子旋转门机制。量子旋转门是一种具有酉性的矩阵，用于改变量子叠加态的概率幅，其定义为

$$\boldsymbol{U}(\theta)=\left[\begin{array}{cc}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{array}\right]. \tag{15}$$

在定义了量子旋转门以后，对某个染色体的第 l 个量子位$[\alpha_l,\ \beta_l]^{\mathrm{T}}$ 的更新过程为

$$\left[\begin{array}{c}\alpha_l'\\ \beta_l'\end{array}\right]=\left[\begin{array}{cc}\cos\theta_l & -\sin\theta_l\\ \sin\theta_l & \cos\theta_l\end{array}\right]\left[\begin{array}{c}\alpha_l\\ \beta_l\end{array}\right]. \tag{16}$$

其中：α_l'和β_l'分别表示经过变换后的第 l 个量子比特的概率幅；$\Delta\theta_l$ 表示该量子比特所对应的旋转门的旋转角，其定义如下：

$$\theta_l=s(\alpha_l,\ \beta_l)\Delta\theta_l. \tag{17}$$

其中：$s(\alpha_l,\ \beta_l)$决定量子旋转的方向，$\Delta\theta_l$ 决定量子旋转的角度。这 2 个变量的取值如表 1 所示。由于量子旋转的角度对算法的收敛速度影响较大，因此在算法运行的初期可以将 $\Delta\theta_l$ 的取值适当加大；在算法运行后期，为了精确求得最优解，可以适当减小 $\Delta\theta_l$ 的取值。

表 1 量子旋转门的调整策略

x_l	b_l	$\mathrm{Fit}(x)\geqslant\mathrm{Fit}(b)$	$\Delta\theta_l$	$s(\alpha_l,\ \beta_l)$			
				$\alpha_l\beta_l>0$	$\alpha_l\beta_l<0$	$\alpha_l=0$	$\beta_l=0$
0	0	否	0	0	0	0	0
0	0	是	0	0	0	0	0
0	1	否	0	0	0	0	0
0	1	是	δ	−1	±1	±1	0
1	0	否	δ	−1	±1	±1	0
1	0	是	δ	1	−1	0	±1
1	1	否	δ	1	−1	0	±1
1	1	是	δ	1	−1	0	±1

注：b 为当前最优解，b_l为最优解的第 l 位。

4 仿真分析

4.1 数据集和参数选取

实验数据集采用真实的网络数据集，利用这些网络数据集构建网络结构，并基于这些网络结构设置传播参数来模拟传播过程，进而验证本文设计的阻断方法。本文采用的网络数据集包括：1) Twitter，该社交网络是一种有向连接的网络；2) Slashdot，来自免费的开放网络社区，该数据集描述的是朋友之间的关系，而且个人的朋友关系可以对外公开；

3）Epinions，该数据集描述一种在线社会网络中人与人的信任关系。网络数据集如表 2 所示。

表 2　采用的数据集

数据集	节点数	边数
Twitter	81 306	1 768 149
Slashdot	77 360	905 468
Epinions	75 879	508 837

每类数据集具有不同的特点，Twitter 的节点关系更为紧密，网络的直径只有 7 跳，90%的有效直径为 4.5 跳；Slashdot 的节点的网络直径达到了 10 跳，90%的有效直径为 4.7 跳；Epinions 的节点关系在 3 个数据集中直径最大，达到了 14 跳，90%的有效直径为 5 跳。基于真实的网络数据结构，设定传播模型的相关参数，在所有网络中随机选择 5 个源点为信息发布的初始节点，并假设信息在网络中传播的感染率为 $p_e=0.2$。

节点的影响力在本仿真中以节点所处的网络结构特征为依据，标识节点影响力通常有节点的度或者介数等指标，而文[13]发现中尺度的网络结构指标 k-core 更能准确地反映节点对信息传播的作用，因此，使用的节点影响力权值为节点的 k-core 值。另外，对于网络结构来说，删除节点的直接代价是改变该点与其邻居的连接关系，因此约束代价选择以节点度为依据。

4.2　传播有效性分析

选择 3 种算法进行对比，其中第 1 种算法是启发式算法，依次删除度 degree 最大的节点；第 2 种是贪心算法 greedy-uc[9]，该算法是每一步都选择当前影响最大的节点；第 3 种是效率与代价最高比的贪心算法 greedy-cb[18]，该算法考虑了操作节点的代价，每一步选择节点时都选择影响力效果与代价的最大比值。

在 3 种网络结构中利用翻硬币的方法模拟信息传播，生成 $M=50$ 规模 $N=15$ 的采样对象，分别生成验证网络视图和测试网络视图，规模都为 750，即 $N\_valid=750$，$N\_test=750$。基于 SAA 的阻断算法在利用训练集中产生最佳执行策略，然后在各验证样本中进行验证。由于网络中节点的数量不同，设置删除代价的最大值为总节点度的 10%。

如图 5 所示，整体阻断信息传播影响力的结果中基于 SAA 的阻断算法是最优的，尤其是在删除节点的中前期，主要原因是基于 SAA 的阻断算法考虑了全局视图，能够把潜在的影响力最大的节点考虑到后续的删除节点的范围内。3 种算法对比结果为，以节点的度为指标的启发式算法最差，基于目标增加量最多和目标增长率最高的贪心算法要优于基于度的启发式算法。

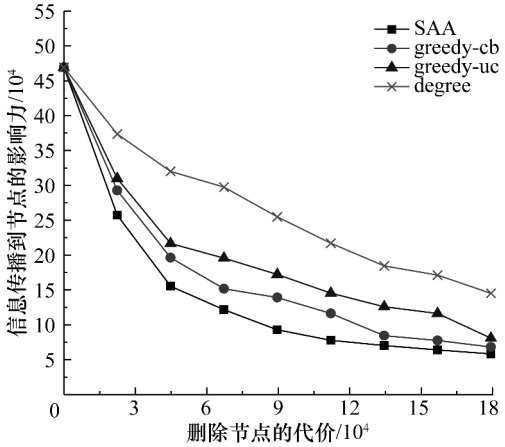

图 5　Twitter 网络中阻断影响力对比示意图

在 Slashdot 网络中的阻断效果如图 6 所示。在 Epinions 网络中的阻断效果如图 7 所示。

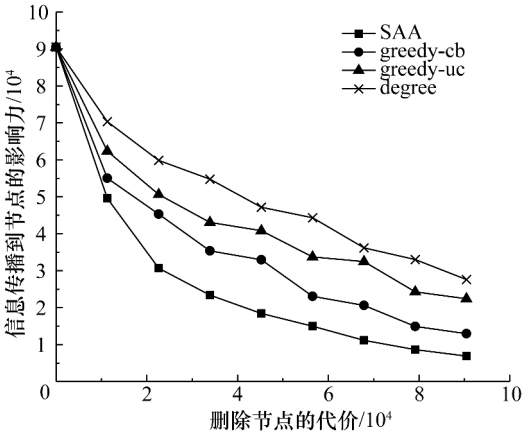

图 6　Slashdot 网络中阻断影响力对比示意图

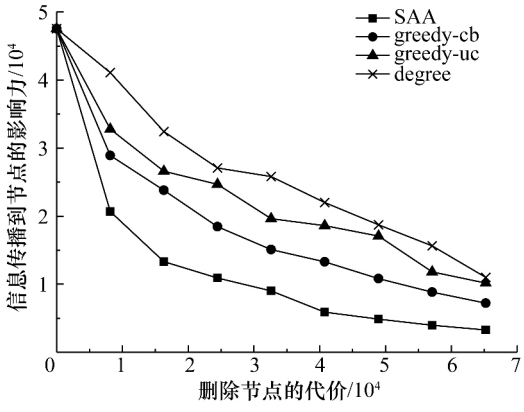

图 7　Epinions 网络中阻断影响力对比示意图

通过对比以上各图可以发现，贪心算法与基于SAA的阻断算法在Twitter数据中的最优解比较相近，而在其他网络中基于SAA的阻断算法要明显优于其他贪心算法。说明基于SAA的阻断算法与网络的紧密程度和结构相关，在Twitter网络中由于结构相对紧密，信息传播到相对远处的路径也较多，因此局部的最优解很可能就是全局的最优解。而在Slashdot和Epinions网络中，结构紧密度相对较差，网络中存在结构洞和较短的信息扩散路径，贪心算法容易造成局部最优解，因此，基于SAA的阻断算法要优于其他两类贪心算法。

对SAA的上界和下界分析如图8所示，设置删除代价的最大值为总节点度的10%，分别分析采样规模对3种网络结构的感染节点影响力的上界和下界的影响。

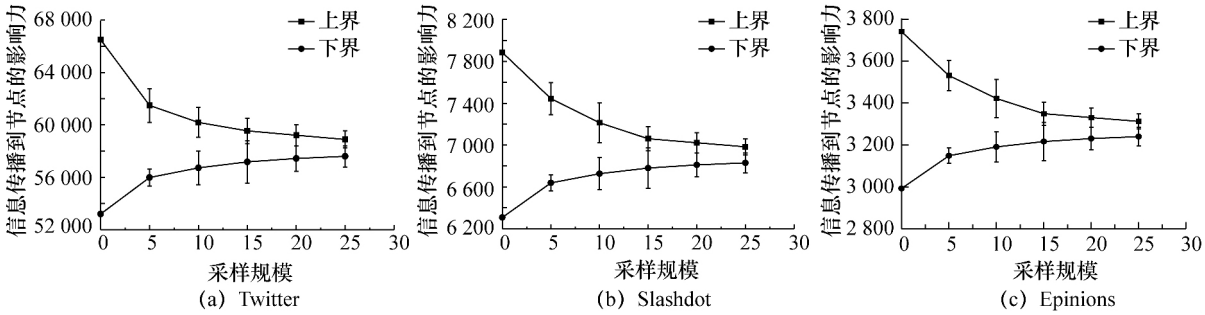

图8 3种网络中不同采样数量的影响

通过对3种结构的上界和下界的分析，可以发现3种网络的采样次数对结果的影响基本一致，当采样规模达到15的时候，上下界差与上界的比值最大为3.9%，因此，每次采样的网络视图规模为15时便可以满足需求。

4.3 预处理对算法时间的影响

预处理包括2种处理方式：一种是剔除不相关的节点；另一种是将感染关系最为紧密的节点进行压缩处理。以Slashdot数据为例，该数据集有77 360个节点和905 468条边，经过预处理后，该数据集减少至18 456个节点和452 332条边。比较进行预处理和没有进行预处理的基于SAA算法的运行时间，以采样50次规模15为例。如图9所示，没有预处理的计算时间是进行预处理的计算时间的3～10倍。运算时间与解空间的大小相关，但是并不是线性关系。在解空间相对较小时，运算时间随着删除节点的代价增加而增加，而删除节点代价大于一定值时，其运算时间随着删除节点代价的增加几乎保持不变，并且能够在较大的解空间内保持较小的运行时间。

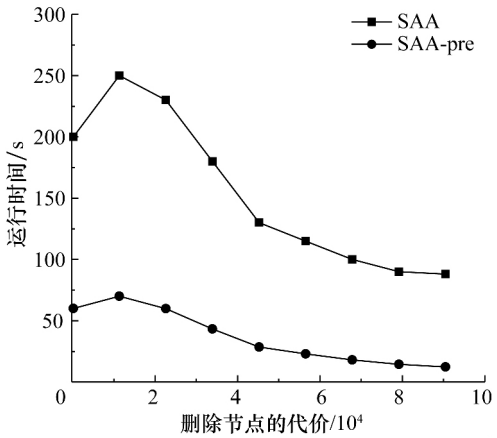

图9 预处理与正常计算时运算时间对比图

5 结 论

本文针对社会网络中阻断信息传播的问题，提出了一种面向节点影响力的信息传播阻断模型，该模型的目标是使信息传播的影响力之和最小。本文证明了该模型的目标函数不满足子模量和超模特征，导致贪心算法并不能够保证近似最优解。然后，使用采样平均近似方法以全局的角度考虑阻断模型，将该问题转化为确定性问题，并分析了该最优解的界限。最后，进一步将采样结果编码为混合整数规划问题，采用一种量子遗传算法进行求解。仿真结果表明，该方法阻断信息传播影响力的效果优于贪心算法，并且网络结构上的优化方法可以有效降低算法运行时间。

参考文献 (References)

[1] 陈卫. 社交网络影响力传播研究[J]. 大数据，2017，1(3)：201503.
CHEN Wei. Research on influence diffusion in social netwok [J]. *Big Data*，2017，1(3)：201503.(in Chinese)

[2] Nowzari C, Preciado V M, Pappas G J. Analysis and control of epidemics: A survey of spreading processes on complex networks [J]. *IEEE Control Systems*, 2016, **36**(1): 26-46.

[3] Prakash B A, Chakrabarti D, Valler N C, et al. Threshold conditions for arbitrary cascade models on arbitrary networks [J]. *Knowledge and Information Systems*, 2012, **33**(3): 549-575.

[4] Tong H, Prakash B A, Eliassi-Rad T, et al. Gelling, and melting, large graphs by edge manipulation [C]// Proceedings of the 21st ACM International Conference on Information and Knowledge Management. Hawaii, USA: ACM, 2012: 245-254.

[5] Saha S, Adiga A, Prakash B A, et al. Approximation algorithms for reducing the spectral radius to control epidemic spread [C]// Proceedings of the 2015 SIAM International Conference on Data Mining. Vancouve, Canada: Society for Industrial and Applied Mathematics. 2015: 568-576.

[6] Chen C, Tong H, Prakash B A, et al. Node immunization on large graphs: Theory and algorithms [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, **28**(1): 113-126.

[7] Khalil E B, Dilkina B, Song L. Scalable diffusion-aware optimization of network topology [C]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2014: 1226-1235.

[8] Zhang Y, Adiga A, Saha S, et al. Near-optimal algorithms for controlling propagation at group scale on networks [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, **28**(12): 3339-3352.

[9] Kempe D, Kleinberg J M, Tardos É. Maximizing the spread of influence through a social network [J]. *Theory of Computing*, 2015, **11**(4): 105-147.

[10] Liu Y, Tang M, Zhou T, et al. Identify influential spreaders in complex networks, the role of neighborhood [J]. *Physia A: Statistical Mechanics and Its Applications*, 2016, **452**: 289-298.

[11] Xia Y, Ren X, Peng Z, et al. Effectively identifying the influential spreaders in large-scale social networks [J]. *Multimedia Tools and Applications*, 2016, **75**(15): 8829-8841.

[12] Zhang J X, Chen D B, Dong Q, et al. Identifying a set of influential spreaders in complex networks [J]. *Scientific Reports*, 2016, **6**: 27823.

[13] Kitsak M, Gallos L H, Havlin S, et al. Identification of influential spreaders in complex networks [J]. *Nature Physics*, 2010, **6**(11): 888-893.

[14] Rubinstein R Y, Kroese D P. Simulation and the Monte Carlo Method [M]. New York: John Wiley & Sons, 2016.

[15] Verweij B, Ahmed S, Kleywegt A J, et al. The sample average approximation method applied to stochastic routing problems: A computational study [J]. *Computational Optimization and Applications*, 2003, **24**(2-3): 289-333.

[16] Narayanan A. An introductory tutorial to quantum computing [C]// Proceedings of the IEEE Colloquium on Quantum Computing Theory, Applications and Implications. London, England: IEEE, 1997: 1-3.

[17] **江逸茗，兰巨龙，周慧琴. 网络虚拟化环境下的资源监控策略** [J]. **电子与信息学报**, 2014, **36**(3): 708-714.
JIANG YiMing, LAN JuLong, ZHOU Huiqin. Resource monitoring policy for network virtualization environment [J]. *Journal of Electronics & Informaion Technology*, 2014, **36**(3): 708-714. (in Chinese)

[18] Leskovec J, Krause A, Guestrin C, et al. Cost-effective outbreak detection in networks [C]// Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Jose, CA, USA: ACM, 2007: 420-429.