

一种针对网络结构破坏下鲁棒影响力最大化问题的 Memetic 算法

王 帅¹⁾ 刘 静²⁾

¹⁾(中山大学智能工程学院 广东 深圳 518107)

²⁾(西安电子科技大学人工智能学院 西安 710071)

摘 要 复杂网络问题的建模与应用是目前研究的热点之一. 如何从网络中选择具有信息传播能力的成员, 即影响力最大化问题, 引发了广泛的关注. 目前的研究主要集中在如何将网络上的信息传播过程建模, 以及如何选择可靠的传播种子节点. 这些工作一定程度上帮助我们了解了网络上的信息传播过程, 但是已有研究并未考虑外界因素对于网络系统的影响, 比如结构损毁对于网络连通性以及网络上信息传播过程造成的干扰. 这使得目前的影响力最大化问题解决方案无法鲁棒地应对现实中的多种应用场景. 针对这一不足, 本文从结构受损的角度定义了鲁棒影响力最大化问题, 在已有工作的基础上设计了对应的性能衡量指标. 在该指标的指导下, 设计了尝试解决鲁棒影响力最大化(Robust Influence Maximization, RIM)问题的 Memetic 算法(Memetic Algorithm, MA), 记为 MA-RIM, 其目的是寻找具有鲁棒信息传播能力的种子节点. 多种网络上的实验结果验证了 MA-RIM 算法在解决鲁棒影响力最大化问题上的有效性.

关键词 复杂网络; 影响力最大化问题; 鲁棒性; 优化; Memetic 算法

中图法分类号 TP18

DOI 号 10.11897/SP.J.1016.2021.01153

A Memetic Algorithm for Solving the Robust Influence Maximization Problem Towards Network Structural Perturbances

WANG Shuai¹⁾ LIU Jing²⁾

¹⁾(School of Intelligent Systems Engineering, Sun Yat-Sen University, Shenzhen, Guangdong 518107)

²⁾(School of Artificial Intelligence, Xidian University, Xi'an 710071)

Abstract The complex network has attracted broad attention in recent researches. This field maintains a series of challenging problems, including the community detection, the robustness optimization, the social network analysis, and the information diffusion process. Much progress has been made in previous studies in summarizing the dynamics of different networked systems and enhancing the performance of networks. Meanwhile, the rapid development of social network services attracts millions of users to attend the network-based information spreading process. Consequently, a growing interest can be found on the influence maximization problem, which is generally defined as selecting system members with considerable ability to spread influence. Existing studies mainly concentrated on how to model the information spreading process on networked systems, and corresponding seed selection strategies to reach a better influence level. In these studies, the structure of network keeps almost unchanged, and the constructed model and obtained seeds may be only effective on current status. However, little attention has been paid to the complicated application environment of networks. Potential failures tend to cause damages on network structures. Such failures always hinder the related influence spreading process, and the

structure of networks tend to be disturbed as well. The existence of uncertainty and failures prevents a robust utilization of current solutions on realistic scenarios, and a thorough study is still required. Considering such deficiency, we intend to define the robust influence maximization problem (RIM), and a rational performance metric has been designed based on existing studies. In terms of the information diffusion process, the designed metric employs a numerical evaluation strategy to estimate the performance of seeds in a specific state of the inputted network. The reliability of such strategy was verified in several pilot studies. Then, in terms of the network destruction process, we focus on link-based attacks and the whole process of losing the connectivity is considered. A summation and a normalization operation are adopted to get the output of this metric, which guarantee the comparability of networks with different scales and seed sizes. Further, a changeable parameter is included in the metric, and a suggested configuration is presented via experimental validations. This configuration can ensure the generality of this metric towards different damage scenarios. Guided by the numerical evaluation result provided by this metric, a memetic algorithm, termed MA-RIM, has been devised to search for seeds that maintain the robust information spreading ability against structural failures. Considering the RIM is a discrete optimization problem, corresponding genetic operators and search mechanisms are elaborated to cater to the data feature. In MA-RIM, the population initialization operator is first conducted to get candidates that are with diverse gene information. The crossover operator is aimed at generating more potential candidates. The mutation operator manages to make minor changes on candidates to overcome the interference of local optima. And the local search operator intends to improve the performance of the whole population. With the help of the selection operator, those candidates with better performance are prone to be selected into the next generation. The experimental results on several network data validate the effectiveness and efficiency of the proposed method together with aforementioned operators. A possible method to select seeds with robust performance is given in this work.

Keywords complex network; influence maximization problem; robustness; optimization; memetic algorithm

1 引 言

复杂网络系统在现实生活中得到了广泛的应用,诸如电力网、路由网等系统是人类社会正常运转不可或缺的一部分^[1]. 一方面,在图论的基础上,网络模型对于网络系统的特性和功能分布进行客观描述;网络数据也常表现出一定的结构特征,诸如社团结构^[2]、随机连接特性^[3]、无标度特性^[4]等特征陆续被学者发现和总结,这些研究为进一步分析网络性能提供了可靠手段. 另一方面,网络系统常暴露在复杂的应用环境中,如何评估和优化网络系统针对攻击或破坏的鲁棒性也是时下研究热点之一^[5]. 这些工作进一步丰富和深入了对于复杂网络模型的研究,也为现实优化问题提供了理论参考.

近年来,随着互联网的普及与高速发展,各种社交网络平台进入了人们的视野,诸如微博、微信、

Facebook 等社交网络进一步丰富了人类的社交活动. 随着其用户规模的逐渐攀升,社交平台成为话题、观点、认知等信息传播的重要媒介;同时,网络社交平台具有信息传播高效、用户密集等传统媒介手段不可比拟的特点,因而线上营销逐渐成为一种重要的产品推广方式,受到了来自用户与厂商的关注. 由于资金等成本受限,厂商几乎不可能针对某个社交网络中的全部用户进行推广,而倾向于从网络中选出最具有影响力的部分用户,并在这些“种子”用户的帮助下对产品进行推广与营销,达到提升产品影响力的目的. 这一过程即可建模为影响力最大化问题^[6].

针对网络系统上的影响力传播过程,几种常见的信息传播模型包括独立级联模型(IC 模型)^[7]、权重级联模型(WC 模型)^[8]以及线性阈值模型(LT 模型)^[9]等. 与此同时,Kempe 等人在文献[7]中将网络上具有影响力的种子节点选择问题表示为一个离

散组合优化问题,并证明这是一个 NP 难问题.考虑到文献[7]中的影响力评估方案依靠于极其耗时的蒙特卡洛过程,难于在大规模网络上应用, Lee 等人在文献[10]中提出了一种针对于种子节点 2-hop 邻域的影响力快速评估方法,并验证了该方法的有效性.在这些研究的基础上,网络上的影响力最大化可被视为一个优化问题,即在一定性能评估因子的指导下,尝试从网络全部成员中寻找最优的种子节点组合.诸如爬山法^[7]、贪心算法^[11]等基于启发式搜索的方法可解决该问题,但存在搜索效率较低的缺点;同时,诸如度、中介数及社团等网络节点或团簇结构特性评估手段^[12-14]为种子节点的选择提供了额外信息,但可能存在种子影响区域重叠的不足.另外,一些基于种群的搜索方法也在影响力最大化问题上得到了应用,包括离散粒子群算法^[15]、Memetic^[16]算法均展示了较好的优化效果.这些优化方案根据网络的结构连接特点,结合网络中节点成员的特性信息,从不同的角度尝试解决网络上的影响力最大化问题.选择出来的具有影响力的种子节点集合为应对网络线上营销、区域流言控制、群众思想传播等现实问题提供了可靠的备选解.

但是,这些已有工作仅仅考虑了网络结构不发生改变的应用场景,所选种子节点的影响力有效性也仅限于当前的网络结构.对于一个网络系统来说,攻击和破坏难于避免,甚至时有发生.无论是针对网络节点^[5]还是网络链路的攻击^[17]都可能对网络性能造成严重的破坏.从网络结构的鲁棒性来讲,已有工作较为全面的研究,分析了结构攻击对于网络性能的影响机理^[5,18],设计了合理的数值化鲁棒性性能评估因子,并提出了一些行之有效的鲁棒性提升方案^[5,17,19].这些研究有力地推动了复杂网络理论在现实网络系统中的应用.针对影响力最大化问题,目前有一些前瞻性工作初步探讨了如何鲁棒地选择种子节点,但是其应用范围局限在诸如节点影响力传播概率^[20]、传播模型^[21]等系统参数不确定的情况.然而,这些工作还存在不足之处.其一,一般来讲,传播概率等因素是基于多次观察下得出的平均值,而且一些已有传播模型也考虑了不同重要性节点之间传播概率的差异^[7,15].对于网络中的整个影响力传播过程,网络中成员影响力传播能力的平均值,相对于每个节点单独的传播行为不确定性来说,可能更为重要.其二,以往的研究证明网络结构往往与网络的性能紧密相关^[3-4],对于一个社交网络来说,其网络结构的变化也会影响到该网络中成员的

互动行为^[22],而互动行为的改变很可能进一步对于影响力传播过程造成影响.因此,鲁棒的种子节点选择结果应该在一定程度上具有应对结构变化的能力,但这一重要特性在以往的研究中并未涉及.与此同时,如何在网络结构改变的条件下对于种子节点的影响力性能进行有效评估,以及相应的种子优化选择过程仍旧是目前研究没有解决的问题.

针对当前研究的局限与不足之处,本文从网络结构变化的角度出发,分析和研究如何鲁棒地解决网络上的影响力最大化问题.考虑到以链路为目标的结构毁坏行为在网络系统应用中时有发生^[17],本文主要针对链路攻击来探讨影响力最大化问题的鲁棒性.首先,在目前已有的网络结构鲁棒性衡量指标的基础上,本文将该指标与文献[10]中的影响力评估因子相结合,设计了一种可以评估种子节点在网络结构受损时的综合影响力性能指标.同时,本文通过实验分析确定了该评估因子的参数取值.该指标可以以数值形式评估某一个种子集合的影响力性能,进而判断种子节点选择的优劣;更进一步可以作为备选解的评估指标来引导相应的优化过程.另外,本文设计了一个 Memetic(Memetic Algorithm, MA)算法来解决鲁棒影响力最大化(Robust Influence Maximization, RIM)问题,记为 MA-RIM,尝试从网络中寻找鲁棒的种子节点集合来应对结构破坏对于影响力传播过程的挑战.在合成及实际网络数据上的实验证明,在所设计指标引导下得到的种子节点集合可以在网络受损过程中取得更佳的影响力传播效果.

本文第 2 节介绍种子节点影响力传播模型与评估方法,以及网络结构鲁棒性指标的背景工作;第 3 节给出种子节点在网络结构受损情形下的鲁棒性评估方法,并给出参数定标的过程;第 4 节给出 MA-RIM 算法的流程及具体算子,相应的实验结果在第 5 节中展示;最后,第 6 节对于全文工作进行总结.

2 种子节点影响力评估以及网络鲁棒性评估指标

一般来讲,社会网络结构可以用一个图 G 表示: $G=(V,E)$, 这里 $V=\{1,2,\dots,N\}$ 是网络中 N 个用户的集合; $E=\{e_{ij} \mid i,j \in V, i \neq j\}$ 代表了网络中不同用户间 M 条互动链接.本文对于网络的描述也遵循这一定义.

本节的内容包含两个部分:第一部分介绍影响力最大化问题中常用的传播模型与种子节点性能评

估方法;第二部分介绍网络鲁棒性问题中的结构评估因子。

2.1 影响力最大化问题中的传播模型与性能评估方法

所谓影响力最大化问题,即为从网络中选择 K 个节点作为种子集合 $S = \{s_1, s_2, \dots, s_K\}$,并由 S 在整个网络中进行影响力传播,其最终影响力范围达到最大值记为 $\sigma(S)$. 在种子节点尝试向外传播影响力的过程中,传播模型定义了信息传播的规则,IC 模型、WC 模型及 LT 模型等这几种常见的传播模型在传播规则上略有差异,现做一个简述. 为了方便运算,这些模型规定网络中的节点只有激活或未激活两种状态,若节点被激活则表示其接受了信息传播且可以继续将信息传给其他节点;若节点未激活状态则表明其还未被信息传播影响. 这里,节点的激活状态是不可跳变的,即节点只能从未激活态变为激活态,反之不可.

在 IC 模型中,在初始状态网络中只有种子集合 S 处于激活态,即 $S_0 = S$. 在传播过程中的第 t 步,在 S_t 中的每个节点 m 都有且仅有一次机会尝试激活它的每一个处于非激活态的邻居 n ,这里的激活概率为 p_{mn} . 当有多个节点尝试激活某一个非激活节点时,每个节点传播过程是独立的,且先后顺序随机确定. 而后,将在第 t 步被成功激活的节点保存在集合 Act_t 中,之后更新激活态节点集合 $S_{t+1} = S_t \cup Act_t$;若某一步骤 t' 得到的 $Act_{t'}$ 为空,则影响力传播过程停止, $\sigma(S)$ 即由最终激活的节点数量 $\|S_T\|$ 决定. WC 模型与 IC 模型的区别在于,在 IC 模型中,节点 m 向节点 n 传播影响的概率 p_{mn} 一般设置为常数(如 0.01、0.1 等,可根据不同的网络特点进行选择);而 WC 模型中的 p_{mn} 需要根据网络中的权重信息确定. LT 模型关注的是网络中节点的影响力积累过程,该模型规定每个节点 m 由自身的激活阈值来表示其被激活的难度,具体影响力传播过程同样从种子节点开始,每个非激活节点需接受到不小于自身激活阈值的影响力才可以跳变为激活状态. 类似地,若没有新的节点被激活,那么传播过程结束. 考虑到 IC 模型具有更为广泛的理论价值,且在以往研究中得到了更多的关注^[10-11,15-16],本文的研究也主要关注于遵循 IC 模型的影响力最大化问题.

一般来讲,可以采用蒙特卡洛过程对于种子集合 S 的影响力水平 $\sigma(S)$ 进行评估,但是该过程非常耗时,且具有一定的随机性,这一定程度上阻碍了该

方法在较大规模网络上的应用. 学者在文献[10]中提出了一种基于种子节点 2-hop 邻域(即邻居的邻居)的影响力水平快速评估方法,定义如下:

$$\hat{\sigma}(S) = \sum_{s \in S} \hat{\sigma}\{s\} - \left(\sum_{s \in S} \sum_{c \in C_s \cap S} p(s, c) (\sigma_c^1 - p(c, s)) \right) - \chi \quad (1)$$

这里 C_s 是种子 s 的 1-hop 邻居,即直接邻居, $p(s, c)$ 是激活节点(s)到非激活节点(c)间的激活概率, σ_c^1 是节点 c 的 1-hop 邻域范围, χ 表示新激活节点到原始种子节点的重复影响力,具体定义可参见文献[10]. $\hat{\sigma}(S)$ 的第一项式子评估了初始种子节点的 2-hop 影响力范围,而后在第二和第三项式子中扣除了从一个种子节点到另一个种子节点的潜在重复影响力. 文献[10]指出,种子节点的作用域主要集中在其 2-hop 邻域内,这种快速估计方法具有一定的合理性,可以可靠地评估种子的影响力性能. 另外,该估计方法具有计算代价较小的优势,因此可以用于引导具有一定规模网络上的种子节点搜索过程^[15-16].

2.2 网络鲁棒性性能的评估因子

网络系统常暴露在复杂的使用环境中,难免遭受各种各样的攻击和破坏. 在这一应用背景下,网络的鲁棒性性能显得尤为重要. 鲁棒的网络应尽可能保持自身功能不受攻击的影响. 以往的研究指出网络的性能往往依赖于其结构连通;因此,为了更加全面的反映网络在受到攻击后的性能稳定程度,衡量因子的设计过程需要考虑在结构受损过程中网络连通性的变化. 文献[5]中提出了一个鲁棒性衡量因子 R ,其评价过程如下:网络中的节点程序按照其重要性顺序(如节点度的大小)进行移除,记录每移除一个节点后的当前网络最大连通簇变化;最后通过归一化加和操作获得最终的数值化评价指标,具体定义如下:

$$R = \frac{1}{N} \sum_{Q=1}^N s(Q) \quad (2)$$

这里, $s(Q)$ 即为移除 Q 个节点后当前网络最大连通簇占比. N 为网络中节点总数, $1/N$ 是归一化算子,确保不同规模的网络可以被公平比较. 鲁棒的网络常常具有较大的 R 值,因子 R 也被成功应用于解决网络鲁棒性优化问题上^[5,18-19].

这里,因子 R 主要针对于基于网络节点成员的攻击过程;但在一些实际网络中,针对于网络节点的攻击可能难于实现. 比如电力网络,供电站(节点)往往处于严密的监控与保护下,完全破坏一个电站节点的难度非常大;与之相反,输电线路(链接或边)常暴露在环境中,因而切断一条输电边的难度会小很

多. 这种实际应用场景对于网络鲁棒性评估过程提出了新的挑战. 在文献[23]中, Zeng 等人针对于网络链路受到破坏的情况对于因子 R 进行扩展, 提出了因子 R_l , 定义如下:

$$R_l = \frac{1}{M} \sum_{P=1}^M s(P) \quad (3)$$

这里, $s(P)$ 即为移除 P 条边后当前网络最大连通簇占比. M 为网络中链路总数, $1/M$ 是归一化算子.

以 R, R_l 为代表的网络鲁棒性衡量因子以数值形式评估了网络的鲁棒性性能, 因此也可用于指导相应的优化过程. 但是, 这些指标主要针对于遭受攻击后的网络连通性变化, 并不能直接应用在网络影响力最大化问题上, 但仍可以为种子节点鲁棒影响力评估指标的设计提供有力的借鉴与参考.

3 种子节点的鲁棒影响力性能评估

对于影响力最大化问题, 以往的工作忽略了潜在的网络结构破坏对于所选取种子节点信息传播能力的影响. 同样, 在实际网络的应用中, 决策者会倾向于性能鲁棒的种子节点集合, 以应对复杂多变的使用环境. 本小节具体介绍如何设计针对种子节点影响力鲁棒性的评估指标, 以及该指标中的参数定标过程.

3.1 评价指标设计

类似于其他网络系统, 社交网络也难免遭受攻击和破坏造成的结构损坏. 一般来讲, 若攻击者可以将一个关键用户从社会群体中完全移除, 这种针对于用户(节点)的攻击势必对于整个社交网络的功能以及网络上的影响力传播等社会行为造成极大的干扰. 这里, 社会关系网络常具有无标度特性^[4], 即仅有少部分用户或者节点具有较为显著的重要性. 这种针对于重要用户的攻击行为过于明显, 而且需要切断某个用户在该社会网络中的全部联系, 其实现难度也很大. 相对而言, 针对于用户间互连关系的攻击难度较小, 也更容易实现, 这种旨在破坏网络中链路的攻击在以往的研究中受到了很多学者的关注^[17, 23]. 针对于影响力最大化问题, 本文也从链路攻击的角度尝试对于种子节点的鲁棒性性能进行研究.

已有研究证明, 蓄意攻击对于网络性能有更加明显的破坏^[5, 19], 已有研究也提出了一系列评估网络中链路重要性的指标. 但考虑到一些网络系统的大规模特性, 诸如中介性系数^[24]等指标计算代价较大, 可能

并不适用. 由于链路的度^[23]也能较为准确的反映出其重要性, 本文主要采用这种因子来评估链路重要性. 对于一条链路 e_{ij} , 它的度定义为两个节点的度乘积, 即 $\text{degree}(e_{ij}) = \text{degree}(i) \times \text{degree}(j)$. 在获得网络中全部链路的重要性评估后, 攻击者按照其重要性对于链路进行排序, 优先攻击较为重要的链路以达到尽快破坏网络功能的目的. 若攻击者的资源非常充足, 那么网络中的全部链接都可能被移除; 但是较为常见的情况是攻击者只能完成对于部分排序靠前链路的攻击, 其分数记为 Per ($Per \leq 1$). 给定一个网络, 攻击过程按如下操作进行. 首先, 使用链路度评估网络中所有边的重要性, 根据获得的数值指标对于边进行排序, 然后按照重要性优先的顺序移除前 Per 分数的边.

在种子节点向外传播影响力的过程中, 若网络结构发生改变, 那么需要对于种子节点在当前网络的影响能力重新评估. 具体来讲, 若网络损失一条链接, 则需要重新计算所选种子节点集合的影响力水平. 由于移除不同链接网络中的传播行为之间相互独立, 因此可以遵循式(3)中的加和、归一化操作来求取种子节点在整个攻击过程中的综合影响力性能. 根据上述机制, 本文将种子节点在网络遭受链路攻击下的影响力性能指标定义如下:

$$R_s = \frac{1}{M \times Per} \sum_{P=1}^{M \times Per} \hat{\sigma}(S|P) \quad (4)$$

这里, 类似于式(3), M 为网络中链路总数, Per 是攻击者攻击的链路数量占 M 的分数. $\hat{\sigma}(S|P)$ 是选择的种子节点集合 S 在移除 P 条边的网络中的影响力估计值. R_s 是一个数值化的衡量因子, 其中的参数 Per 规定了链路移除的百分比; 针对某一个 Per 值, 若种子集合 S 可以取得较大的 R_s 值, 那么 S 可以在当前攻击下达到更好的影响力.

R_s 的设计一方面考虑了网络的结构攻击特征, 依照式(2)、(3)中网络性能评估方式, 对于种子节点在结构受损过程中的影响力性能进行记录并做归一化操作, 较大程度上保证了数值比较的公平性. 另一方面, 考虑到计算复杂度, 本文采用了式(1)中的影响力快速评估方法以数值形式估计种子集合 S 在受损网络中的性能. 这些评估策略保证了 R_s 可以较为全面地评估所选种子节点的鲁棒性性能, 进而为潜在的优化提供较为可靠的评估因子. 这里, R_s 中的参数 Per 还未确定, 而且 Per 的不同取值会对优化过程带来不同的偏好. 本文在下一个小节给出对这一问题的研究.

3.2 指标参数定标

R_s 中的参数 Per 代表了某个网络中链路移除的比值,确定某个 Per 后该因子就可以用于引导优化过程,完成种子节点的选择工作. 若 Per 的值设定较小,那么优化算法选择出来的备选解会偏向于网络未遭受攻击或者链路受损较少的情况,这样的解并不能保证在网络中大范围链路受损时有稳定的性能;反之亦然. 因此,如何确定一个合理的 Per 值,使种子选择过程综合考虑网络不同的受损情况,进而取得性能稳定、较为鲁棒的解,这对于鲁棒影响力最大化问题的解决具有重要的意义.

本文采用实验验证的方法尝试确定合理的 Per 取值. 本实验在两种最为常见的人工合成网络上进

行,包括基于 BA 模型生成的无标度(SF)网络^[4],以及基于 ER 模型生成的随机(ER)网络^[3],这些网络的规模 N 定为 500,平均度 $\langle k \rangle$ 定为 4. 在指标 R_s 的引导下尝试从对应网络中选择种子节点,这里种子节点集合的规模 K 定为 10,节点间的影响力传播概率 p 定为 0.01. 为了验证不同 Per 值对于种子选择结果的影响,这里 Per 值共取 11 组,从 0 开始,按照 0.1 为步长,直至最大值 1 为止,即 $Per = \{0, 0.1, 0.2, \dots, 1\}$. 按照某一组 Per 值选择出来的种子节点需要在其他组 Per 验证其影响力性能. 这个实验采用传统的遗传算法(Genetic Algorithm, GA)来完成对于种子节点的选择工作,对应的实验结果在表 1 和表 2 中展示.

表 1 ER 网络上不同 Per 值所选择种子的 R_s 性能测试结果(每一列代表的是选择种子时设置的 Per (边移除百分比),每一行代表的是不同 Per 值(也代表边移除百分比)下所选种子的影响力性能测试. 表中每行中的最优值用粗体表示,代表实验中某个 Per 值设置下种子节点可达到的最佳影响力性能. 最后一行的差值为当前优化目标的种子集合与各个 Per 值设置下可达的最优性能之间差距的总和. 表中的数值结果是 5 次独立实验均值)

测试	目标										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	10.91	10.79	10.77	10.72	10.69	10.63	10.57	10.57	10.56	10.56	10.58
0.1	10.64	10.74	10.72	10.68	10.67	10.62	10.57	10.57	10.56	10.55	10.57
0.2	10.49	10.61	10.64	10.62	10.61	10.60	10.56	10.56	10.55	10.54	10.56
0.3	10.39	10.55	10.59	10.61	10.59	10.57	10.55	10.55	10.54	10.53	10.54
0.4	10.33	10.48	10.51	10.54	10.55	10.55	10.53	10.52	10.52	10.51	10.52
0.5	10.28	10.42	10.46	10.50	10.50	10.51	10.51	10.50	10.50	10.49	10.49
0.6	10.24	10.36	10.41	10.45	10.45	10.47	10.48	10.47	10.47	10.46	10.46
0.7	10.21	10.32	10.36	10.40	10.41	10.42	10.44	10.44	10.44	10.43	10.43
0.8	10.19	10.29	10.33	10.36	10.37	10.38	10.40	10.40	10.41	10.41	10.40
0.9	10.17	10.26	10.29	10.33	10.33	10.34	10.36	10.36	10.37	10.38	10.37
1	10.15	10.23	10.27	10.29	10.30	10.31	10.32	10.33	10.34	10.34	10.35
差值	2.08	0.89	0.68	0.58	0.64	0.63	0.75	0.75	0.82	0.78	0.91

表 2 SF 网络上不同 Per 值所选择种子的 R_s 性能测试结果(该表中仅展示最终计算差值的结果,计算方式同表 1. 表中的数值结果是 5 次独立实验均值)

优化时的 Per	差值
0	0.56
0.1	0.47
0.2	0.50
0.3	0.42
0.4	0.67
0.5	1.26
0.6	1.25
0.7	1.39
0.8	1.57
0.9	1.72
1	1.74

在表 1 中,表中第一行数值代表 GA 在搜索种子过程中使用的 Per 取值,第一列数值代表所选取的种子节点进行影响力评估时的 Per 取值. 由这个表可知,在实验中不同的 Per 取值会对所选种子节点集合的性能产生影响. 在优化过程中,若 GA 在搜索过程和测试过程中设置的 Per 值相同,那么该种子集合与取其他 Per 值获得的结果相比可以达到

更好的影响力性能;但与此同时,某个 Per 取值获得的解在其他 Per 值下的性能表现可能会出现明显的下滑,尤其是当优化的 Per 取值与测试的 Per 取值相差较大时. 但是,给定一个网络数据,决策者的期望是只获得一组种子集合备选来完成影响力传播任务,而这些种子应该具备应对多种结构破坏的能力. 这也就是说,优化算法选择出来的种子备选集合应该在多种 Per 取值条件下都取得较好的表现,即对于结构攻击有较好的鲁棒性.

为了数值展示式(4)中 Per 值对于种子选择过程的影响,我们在表 1 结果的基础上进行了更进一步的数据分析,旨在评估获得的种子节点在多种测试条件下的性能. 对于每一行数据,几组由 GA 选择出的种子节点可取得的最优影响力指标加粗显示;一般当种子选择和性能测试过程采用相同或接近的 Per 值时达到最优影响力. 然后,针对于每一列数据,将某个 Per 值对应的 GA 搜索结果在不同测试实验中获得性能指标与该列最优指标相减,得到该

Per 值作为评估因子的性能差值；再将所有测试情况的差值加和，对应结果在最后一行中展示。表 2 中的结果也采用相同的计算方式。

表 1、表 2 中的差值计算结果可能对于 Per 的最终定标有所启发。我们可以发现 Per 取值为 0.3 时得到种子集合在 ER 和 SF 这两种常见的合成网络上都可以取得较为稳定的影响传播能力；在 Per 值改变的条件下，这些种子集合即使不能达到对应情况的最优解，但仍可取得相对而言可以接受的性能。另外，考虑到很多现实网络都具有随机连接^[3]或者无标度^[4]的特性，这里的实验结果可能对于实际网络系统中的影响力最大化问题解决有一定的现实意义。因此，在后续尝试通过优化算法寻找性能鲁棒的种子节点的实验中，本文将影响力性能评估因子 R_s 中的参数 Per 定为 0.3。同样，使用者可以根据所处理网络的特征或者人为偏好对于参数 Per 进行修改，使得获得的种子节点的影响力性能倾向于网络结构较为完整的应用场景或者网络结构受损严重的应用场景。

4 MA-RIM 算法

4.1 算法框架

解决鲁棒影响力最大化问题的核心在于从网络中选择合适的种子节点集合，使这些种子在网络遭受攻击的情况下保持较为稳定的影响力传播性能。该问题可被看作是一个离散优化问题，该问题的决策变量或基因就是欲选择的节点集合 $\{S_1, S_2, \dots, S_K\}$ ，且每个种子互不相同。本文欲采用 Memetic 算法，在种子鲁棒影响力传播因子 R_s 的引导下，完成这一优化问题，该算法记为 MA-RIM。与传统的 GA 相比，MA-RIM 具有一个基于影响力最大化问题设计的局部搜索算子，以达到搜索个体邻域、进一步提升种群适应度水平的目的。

算法 1. MA-RIM 算法的框架。

输入： G_0 ：输入网络； K ：种子节点集合规模

$MaxGen$ ：最大进化代数； $PopSize$ ：种群规模

p_c ：交叉算子执行概率； p_m ：变异算子执行概率

p_l ：局部搜索算子执行概率

输出： $S^* = \{S_1, S_2, \dots, S_K\}$ ：最优种子集合

1. 执行初始化操作，在 G_0 结构信息的基础上产生规模为 $PopSize$ 的初始化种群 P_0 ， $g=0$ ；
2. 随机选择 P_g 中没有被选中的两个个体 P_i 和 P_j ，以概率 p_c 尝试执行交叉算子，并将产生的新个体放入

临时种群 P_t 中，直到所有个体被选中；

3. 对于 P_g 和 P_t 中的每个个体 P 以概率 p_m 执行变异算子；
4. 对于 P_g 和 P_t 中的每个个体以概率 p_l 执行局部搜索算子；
5. 用当前种群中的最优解尝试更新 P_{g+1} 中的第一个个体， P_{g+1} 中其余 $PopSize - 1$ 个个体按照轮盘赌选择法从 P_g 和 P_t 中选择得到， $g=g+1$ ；
6. 若 g 大于 $MaxGen$ ，输出当前种群中的最优个体，否则，转 2；

MA-RIM 算法的执行流程简述如下。首先，该算法通过随机选择与度优先选择两种策略相结合的方式产生一系列种子备选来初始化种群。然后，算法执行交叉算子，交换当前种群中个体间的基因信息，进而产生更多个体，进一步扩大种群规模。在此基础上，算法执行变异算子，随机改变个体的某一位基因，帮助搜索过程跳出局部最优。此外，算法还需执行局部搜索算子，该算子利用 Per 设置为 0.3 时种子集合的结构特征搜索每个个体的邻域空间，达到进一步提升种群中每个个体适应度水平的目的。每次迭代结束于选择算子，算法选择那些表现较好的个体进入下一代种群，并保存或尝试更新搜索过程中出现的最优解。重复执行上述步骤，直到满足迭代终止条件。搜索过程中的最优解即为算法的最终输出。MA-RIM 的流程总结在算法 1 中。各个遗传算子具体的流程在下一小节给出。

4.2 遗传算子

MA-RIM 主要包括这样几个遗传算子：初始化算子、交叉算子、变异算子和局部搜索算子。这些算子在 MA-RIM 的执行过程中发挥着不同的作用，都对于算法的性能有一定的影响。

MA-RIM 的搜索过程开始于初始化算子，这个算子仅执行一次，其目的是根据输入网络 G_0 的结构信息得到一些种子节点备选解来构成初始化种群。具体来讲，对于前半规模 $PopSize/2$ 的种群，个体的初始化随机完成，即每个个体随机从 G_0 的 N 个节点中选择 K 个作为该个体的基因，这里需要保证 K 个种子互不相同。对于后半种群，个体的初始化操作需要考虑 G_0 中每个节点的度信息（即每个节点连接的邻居总数）。在文献[7, 15-16]中指出，社交网络中邻居数量较多的关键节点往往具有更好的影响力传播性能。这里，根据节点的度信息，使用轮盘赌方法选择节点作为个体的基因。在此规则下，度较大的节点有更大的可能被选择，这些节点可能具

有更佳的性能,但是为了丰富种群在解空间中的分布,一些度较小的节点仍有可能被选择成为备选种子. 初始化算子旨在产生广泛散布于解空间中的备选种子解,为进一步搜索提供了充足的样本,能一定程度上克服局部最优解的干扰,有助于接下来寻优操作的顺利进行.

交叉算子的目的是交换两个个体间的部分基因信息,从而产生更多的个体,并进一步丰富种群中个体的基因型. 这里,每个个体的基因都是长度为 K 的离散序列. 本文的交叉方式采用概率为 p_c 的单点交叉,对于种群中随机选择的两个个体 i 和 j ,先随机产生一个位于 $[1, K-1]$ 区间内的整数 $posi$,再将 i 与 j 中 $posi$ 前后的基因信息交换,进而产生两个新个体 i' 和 j' . 注意,如新个体 i' 或 j' 存在重复的种子节点,则执行修正操作,将重复的节点替换为另一个从网络中节点;该节点的选择随机进行,在选择时需保证种子集合的有效性. 交叉算子的一个例子如图 1 所示.

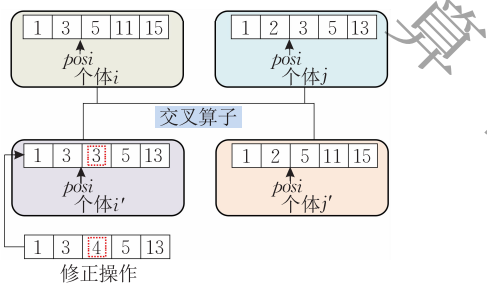


图 1 交叉算子的一个例子(图中的两个个体 K 设置为 5,交叉位置 $posi$ 随机取为 2. 注意在新产生的个体 i' 中有重复的种子节点,这时需要执行修正操作,将重复的节点替换为一个其他节点,以保证种子的有效性)

此外,MA-RIM 还需执行变异算子,对于种群中的每个个体,该算子以 p_m 的概率执行,将个体基因的某一位替换为网络中的其他节点,这个节点的选择同样随机进行,同时保证该个体的基因中不存在重复节点. 变异算子随机改变种群中个体的部分基因信息,这里有可能产生新的种子节点组合方式,一定程度上提升算法的搜索能力.

本文提出的 MA-RIM 与传统的 GA 算法最大的区别在于对网络结构信息的利用. 这里我们先对表 1、表 2 得到的实验结果进行分析. 在 3.2 小节 的实验中, Per 为 0.3 的衡量因子引导的优化过程可以获得性能较为鲁棒的种子集合,这里以 SF 网络为例,表 3 针对于该情况下选择出的节点在度、节点间距离与其他 Per 值得出结果之间的差异进行数值分析.

表 3 SF 网络上根据不同 Per 值选择出来的种子节点在 网络中的属性信息

优化时的 Per	平均度	度的标准差	节点间平均距离
0	25.0	8.10	9.4
0.3	23.7	9.09	9.6
0.5	18.8	8.98	11.4
0.7	18.7	8.98	11.4
1.0	18.6	9.11	11.6

如表 3 所示,在 $Per=0$ 即网络没有遭受攻击的情况下,种子节点的平均度都较大,种子节点间的平均距离较小;随着 Per 值的增大,所选择的种子节点平均度有下降的趋势,而且这些节点间的度差异以及节点之间的距离有所上升. 当 Per 达到 1 时,所选择的种子节点相对于其他选择结果具有最小的平均度,最大的标准差,以及最大的节点间距离. 这种情况的产生有迹可循,在实验中,攻击由切断网络中的链接完成,那些较为重要的边会被优先攻击,而这些边通常将邻居较多即度较大的节点相连. 在网络不遭受攻击的情况下,度较大的节点具有更多的邻居和更大的潜在影响范围,因此这些关键节点会被优先考虑成为影响力传播的种子. 但是,当表 2、表 3 中的攻击发生时,网络会丢失很多较为重要的边链接,随之而来的是这些关键节点的影响能力下降. 此时,越大规模的攻击会造成更多的连通性损失,也会更加削弱度较大的关键节点在信息传播过程中的作用. 值得注意的是,一些在原始网络中并不是很重要的节点在这种情况下可以表现出更好的影响传播能力. 其原因在于这些节点的链接并不是非常重要,可以在较长的时间内免于遭受攻击,由此保证了信息传播的正常进行. 由于这种机制,在种子节点选择过程中应该考虑一些度较小的节点,以此保证在发生攻击时所选种子依旧保持一定的影响能力. 单独针对 $Per=0.3$ 的情况来说,其种子的平均度比较大,但是种子之间也存在较大的度差异,同时种子节点间也保持了一定的距离.

表 3 中的结果可以为局部搜索算子的设计提供一些参考. 其一,算子的搜索过程应该考虑节点更多的局部特征,如节点的 2-hop 邻域内的邻居节点;另外,算法在进化过程的初期可以全局地考虑一些度较大的节点进入种子集合,进而尽快提升个体的适应度水平(影响传播能力);但随着进化迭代的执行,搜索步骤应逐渐减小对于度较大节点的重视,以免陷入局部最优以及影响最终优化效果.

采用上述策略,本文将 MA-RIM 中的局部搜索算子设计为两部分. 第一部分是种子的节点局部邻

域搜索,考虑将每个种子替换为其邻居或邻居的邻居.这里为了避免评估时间过长,邻居的邻居按照概率 p_{micro} 进行选择.若替换操作可以提升种子集合的适应度水平,则保留替换后的节点.第二部分是种子节点的全局搜索,尝试按照概率 p_{macro} 用网络中度较大的几个节点(占网络规模的 2%)替换当前种子集合中的某一个点,该点的选择依照轮盘赌进行,度越小则该种子被选择的概率越大.若替换后的种子集合具有更好的适应度,那么接受替换操作.该算子中的 p_{micro} 和 p_{macro} 在实验中均设置为 0.5.

与依靠随机交换或选择的交叉算子和变异算子不同,局部搜索算子在式(4)中鲁棒影响力指标因子的指导下完成搜索过程,只保留对于影响传播能力有贡献的新个体.这样的搜索过程虽然比较耗时,但是可以直接提升每个个体乃至整个种群的适应度水平,促进算法寻优过程的收敛.局部搜索算子的具体流程在算法 2 中给出.

算法 2. 局部搜索算子.

输入: P : 当前种群; p_i : 局部搜索算子执行的概率
 p_{micro} : 个体邻域搜索概率; p_{macro} : 网络全局搜索概率
 gen : 当前进化代数; $MaxGen$: 最大进化代数

输出: P' : 经过局部搜索的种群

1. $P' = \emptyset$;
2. FOR P 中的每个个体 p DO;
3. IF $rand() < p_i$ DO: /* $rand()$ 产生 $[0,1]$ 区间内的随机数 */
4. FOR p 中的每个种子节点 s DO;
5. $Nei = \emptyset$;
6. FOR s 的每个邻居 s_n DO;
7. 将 s_n 放入 Nei ;
8. 对于 s_n 的每个邻居 s_{nn} , 若 $rand() < p_{\text{micro}}$, 则将 s_{nn} 也放入 Nei 中;
9. END FOR;
10. 尝试用 Nei 中的每个节点替换 s , 若 R_s 值提升则将替换后的种子集合保存在 p 中;
11. END FOR;
12. IF $rand() < p_{\text{macro}} \times ((MaxGen - gen) / MaxGen)$ DO;
13. 得到网络中度大小在前 2% 的节点集合 Top ;
14. 根据每个种子度的大小,利用轮盘赌从 p 的 K 个种子中选择出一个种子 s ;
15. 尝试用 Top 内的每个节点替换 s , 若 R_s 值提升,则将替换后的种子集合保存在 p 中;
16. END IF;
17. END IF;
18. 将 p 放入 P' ;
19. END FOR;

在每轮迭代的最后,MA-RIM 算法利用轮盘赌方法将较优个体选择进入子代种群.这里一些性能较差的个体同样有被选择的机会,其目的是一定程度上提升种群的多样性.另外,子代种群的第一个个体保存的是当前搜索过程找到的全局最优解,用来防止种群适应度退化.这样的选择机制对于算法避免局部最优也有一定的贡献.

MA-RIM 算法主要分为以下几个部分:(1)初始化算子.针对 $PopSize$ 个个体进行初始化操作,其复杂度为 $O(PopSize)$,同时算子仅执行一次;(2)交叉算子.每轮迭代最多执行 $PopSize/2$ 次,整个迭代过程中该算子的复杂度为 $O(PopSize \times MaxGen)$;(3)变异算子.每轮迭代最多执行 $PopSize$ 次,整个迭代过程中该算子的复杂度为 $O(PopSize \times MaxGen)$;(4)局部搜索算子.每轮迭代最多执行 $PopSize$ 次,每次执行需要的复杂度为 $O(K \times \langle k \rangle + 1)$,整个迭代过程中该算子的复杂度为 $O(PopSize \times MaxGen \times K \times \langle k \rangle)$.综上,MA-RIM 算法的复杂度为 $O(PopSize \times MaxGen \times K \times \langle k \rangle)$.

5 实验结果

为了综合检验所设计算法 MA-RIM 的性能,我们先在具有 500 个节点的 SF、ER 两种人工合成网络上进行实验.首先,实验验证了不同方法选择出来的不同规模的种子节点集合的影响能力.本文实现了 GA、基于 Betweenness (Bet)^[13]、基于 PageRank (Pag)^[12]、基于 Degree (Deg) (度)^[14] 这几个种子选择方法提供对比.实验中,采用了式(4)中的 R_s 与式(1)中的 $\hat{\sigma}(S)$ 对于种子集合的性能进行测试. MA-RIM 的参数设置如下: $MaxGen$ 设为 150, $PopSize$ 设置为 50, p_c 设置为 0.6, p_m 设置为 0.5, p_i 设置为 0.5; GA 采用类似的参数设置:迭代次数设置为 150,种群规模设置为 50,交叉、变异概率分别设置为 0.6 和 0.5.另外,也实现了贪心算法 (Greedy)^[6,11],在式(4)的引导下尝试搜索种子集合.

以种子节点集合规模 $K=5, 10, 20$ 为例的结果在图 2 中展示.由图可得,MA-RIM 得到的种子集合相比于其他对比的方法可以取得更好的 R_s 值,这意味着这些种子节点可以在网络遭受攻击时保持更好的信息传播能力.这里的结果展示,若仅依靠原始网络的结构信息进行种子节点选择(如中介数、度等),那么获得的种子节点往往不能很好地应对网络结构受损的情况.随着期望 K 值的增长,MA-RIM

相比于其他方法的优势会更加明显. 当 K 值较小时, 即使简单地选择网络中重要性较高的几个节点作为种子, 也可达到较好的影响力传播性能; 而当 K 值较大时, 网络中有限的关键节点无法满足种子集合规模, 而且种子之间会出现影响范围交叠的情况, 这增加了种子选取的难度, 使用基于进化策略的优化方案可以更好地应对这种情况. 贪心算法的搜索主要针对单个个体进行, 这样的方法寻优能力有限, 获得的结果性能并不突出. 另外, 使用度等网络结构信息进行种子节点选择时并没有考虑潜在的结构损毁过程, 这可能也影响了这些种子在网络遭受结构损毁时的影响力传播性能. 对比 GA 和 MA-RIM, 二者采用基于种群的搜索策略, 均包括了对于不同个体间的信息交换(交叉算子)及个体的变异(变异算子)操作; 但是, MA-RIM 还包括对于个体的局部搜索策略, 能更为有效地利用种群内的基因信息, 因而具有更好的寻优能力. 这里的实验结果验证了算法 2 中给出算子的有效性.

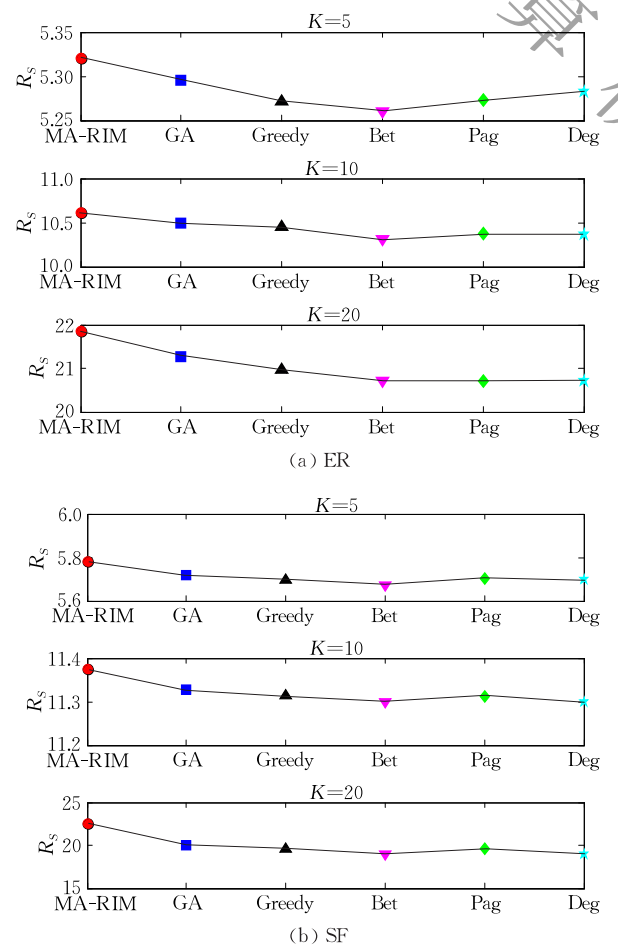


图 2 ER 与 SF 网络上选择的不同种子集合性能, 以种子集合 $K=5, 10, 20$ 为例(图中结果是 5 次独立实验得到的均值)

以 $K=20$ 为例, 表 4 数值展示了几种算法获得的种子集合在具有 500 个节点的 ER、SF 两种网络受到结构损毁时取得的影响力性能($\sigma(S)$)及算法所需运行时间(记为 T , 单位为 s); 这里的影响力性能通过蒙特卡洛方法估计, 结果取 5000 次实验的均值. 实验中对比了同样依靠蒙特卡洛过程完成种子节点选择的 CELF 算法^[11], 给出了 MA-RIM、GA、Pag、Deg 这几种方法获得的结果, 并以 MA-RIM 获得的结果为基准进行显著性水平为 0.05 的 t 检验, 验证不同方法获得的结果是否具有显著的性能差异. 这里, CELF 依靠节点的真实传播性能进行种子选择, 这样的操作可以保证选择过程的有效性; 但是其性能评估过程依靠蒙特卡洛过程, 需进行大量的实验来获得可靠估计, 该过程非常耗时, 难于应对网络遭受结构打击时的情形. 由表可知, 在两种网络上 MA-RIM 选择的种子节点均可以取得较好的鲁棒影响力传播性能, 获得的结果接近 CELF 达到的效果, 且明显优于其他方法. 同时, MA-RIM 算法较 CELF 相比也可以节省大量的运算时间, 验证了该算法的有效性. 同时, 表 4 中的结果也侧面验证了 R_s 的合理性, 证明该因子可以对于种子节点的信息传播能力进行合理估计.

表 4 两种合成网络上不同方法选择 20 个种子节点的影响力性能及所需运算时间对比(以 MA-RIM 为基准进行了显著性检验, 若 MA-RIM 明显优于/劣于某种方法则标记为+/-, 若性能接近则标记为 $\approx$)

	ER		SF	
	$\sigma(S)$	T	$\sigma(S)$	T
MA-RIM	26.43	7350	74.33	7270
CELF	26.51(-)	13460	74.36($\approx$)	18240
GA	24.34(+)	5921	68.51(+)	6020
Pag	22.32(+)	2	65.21(+)	3
Deg	21.95(+)	1	65.25(+)	1

另外, 表 4 中的结果也展示了 ER、SF 两种合成网络的差异. 具体来讲, 选择同样数目的种子节点, SF 网络上的种子可以取得更大的影响力, 这种现象与两种网络的结构特征有关. ER 网络度分布较为均匀, 算法选择的种子节点具有较好的信息传播能力, 但是较其他节点相比优势并不明显; 而 SF 网络具有幂律分布特性, 算法选择的种子节点常具有较高的度, 对比于其他节点有较为明显优势. 这种结构特性造成了表中结果的数值差异. 与此同时, 这里的实验结果验证 MA-RIM 可以应对这两种常见的合成网络, 具有较好的通用性.

更进一步, 在节点规模为 1000 和 1500 的合成

网络上验证了算法的性能,不同方法获得的种子集合 R_s 性能指标在表 5 和表 6 中展示,同时给出了以 MA-RIM 获得的结果为基准的显著性检验结果. 由表可知,在较大规模的网络上,本文提出的算法明显优于其他方案,展示出良好的可扩展性. 随着网络规模的进一步增大,网络的结构包含更多的信息,这增加了鲁棒种子节点的选择难度. 这时,单纯依靠网络结构信息往往无法选择出可靠的种子集合;而采用 R_s 引导的算法可以获得传播性能更加鲁棒的解.

表 5 较大规模的 ER 网络上不同方法选择种子节点的 R_s 性能对比

	K	MA-RIM	GA	Bet	Pag	Deg
1000	5	5.355	5.291(+)	5.172(+)	5.242(+)	5.228(+)
	10	12.031	11.601(+)	10.374(+)	10.452(+)	10.452(+)
	20	22.951	21.594(+)	20.757(+)	20.893(+)	20.822(+)
1500	5	6.015	5.191(+)	5.144(+)	5.172(+)	5.172(+)
	10	12.971	11.457(+)	10.338(+)	10.388(+)	10.342(+)
	20	23.626	21.821(+)	20.669(+)	20.886(+)	20.778(+)

表 6 较大规模的 SF 网络上不同方法选择种子节点的 R_s 性能对比(若算法与 MA-RIM 没有明显性能差异标记为 $\approx$)

	K	MA-RIM	GA	Bet	Pag	Deg
1000	5	6.125	6.016(+)	5.855(+)	5.859(+)	5.847(+)
	10	13.736	12.015(+)	11.703(+)	11.716(+)	11.706(+)
	20	26.927	24.584(+)	23.123(+)	23.196(+)	23.196(+)
1500	5	6.079	6.079($\approx$)	6.079($\approx$)	6.079($\approx$)	6.079($\approx$)
	10	14.447	13.025(+)	12.041(+)	12.102(+)	12.102(+)
	20	26.961	24.514(+)	23.897(+)	23.956(+)	23.921(+)

在表 7 中,本文以具有 500 个节点的 SF 网络为例,在 $K=5,10$ 和 20 的情况下,在数值上对比了利用 R_s 作为优化指标获得的种子在原始网络中的影响力性能 $\hat{\sigma}(S)$. 从表中可知,在 R_s 引导下取得的种子节点集合在原始网络未遭受破坏的情况下也可以保持较好的影响力传播性能,取得与实验中最佳方法性能相近的解. 另外,这里的实验结果也验证了 PageRank 这一网络结构指标在解决影响力最大化问题中的重要价值,利用这一指标选择出来的网络节点常可以取得实验中最佳的 $\hat{\sigma}(S)$ 值. 网络中节点 PageRank 指标的评估过程也依靠于节点间的信息流通,其机理与社交网络用户间影响力传播过程有相似的地方,这可能是 PageRank 指标在这几个实验中得到较好结果的原因之一. 在以往的工作中^[12,15],一些学者也指出该指标可以帮助决策者选择网络中信息传播能力较强的节点. 总结来讲,图 2、表 4 至表 7 的实验结果验证了 MA-RIM 可以成功解决鲁棒影响力最大化种子节点选择任务,且性能优于一些已有的种子节点选择方案. 这些结果

验证了 4.2 小节中给出的遗传算子的有效性. 另外,在参数 Per 设定为 0.3 的性能指标 R_s 引导下的优化结果在原始网络信息传播过程中也可获得较为满意的结果,这验证了 3.2 小节中衡量因子参数定标过程的合理性.

表 7 500 个节点 SF 网络上不同方案选择 K 个种子节点的影响力性能对比(表中数据是在 $\hat{\sigma}(S)$ 评价下的数值结果)

K	MA-RIM	GA	Bet	Pag	Deg
5	6.580	6.580	6.500	6.580	6.559
10	12.417	12.421	12.357	12.434	12.357
20	23.625	23.606	23.624	23.626	21.438

另外,本文也对两种基于种群搜索方法(MA-RIM 和 GA)的收敛过程进行了分析. 以 SF 网络上的实验为例,两种算法的收敛曲线在图 3 中给出. 图 3(a)展示了优化过程中 R_s 的变化趋势,由图中变化曲线可知,两个基于种群的搜索算法 MA-RIM 与 GA 都可以逐渐收敛,且 MA-RIM 的最终优化结果可以略好于 GA. 另外,值得注意的是 MA-RIM 可以更快地收敛到最优解,这里的实验结果体现了局部搜索算子在搜索过程中的作用. 另一方面来讲,图 3(b)中 $\hat{\sigma}(S)$ 在进化过程中的变化趋势也反映了优化过程对于种子节点信息传播能力的影响. 在进化初期, R_s 和 $\hat{\sigma}(S)$ 两个指标可以得到同时提升,但是在后续以 R_s 为目标的寻优过程中,MA-RIM 和

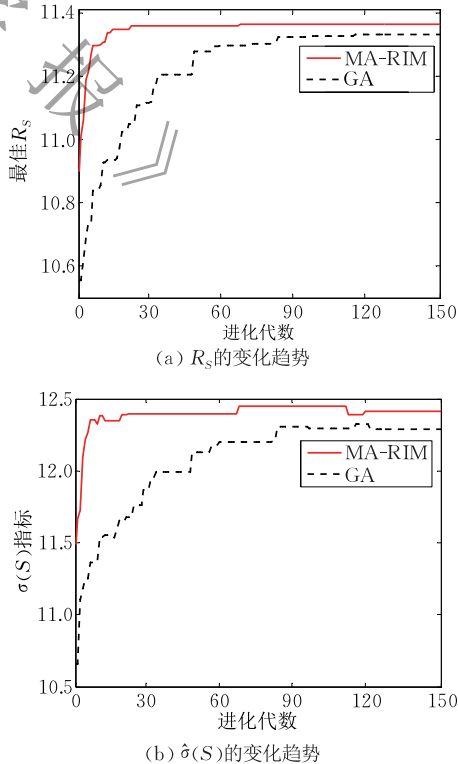


图 3 SF 网络上的实验在进化过程中种子影响力的性能变化趋势

GA 所选种子节点的 $\hat{\sigma}(S)$ 性能指标都有所下降,这说明种子节点在网络受攻击时的影响力提升需要以原始网络中的部分影响能力为代价.但是在表 7 中的数值结果证明了这部分损失相对而言并不明显,而这些代价可以有效提升所选种子节点在影响力传播过程中应对网络结构破坏的鲁棒性.对于决策者来说,性能鲁棒的解决方案往往更加具有应用价值.针对于影响力最大化问题来讲,鲁棒的种子节点不但能够较好的完成原始网络中的信息传播任务,而且在网络遭受攻击、连通性受损的条件下也能一定程度上确保信息传播任务的完成.因此,本文提出的种子节点选择方案可以更为全面地应对诸如文献 [11-15] 中提出的社会网络影响力传播问题.

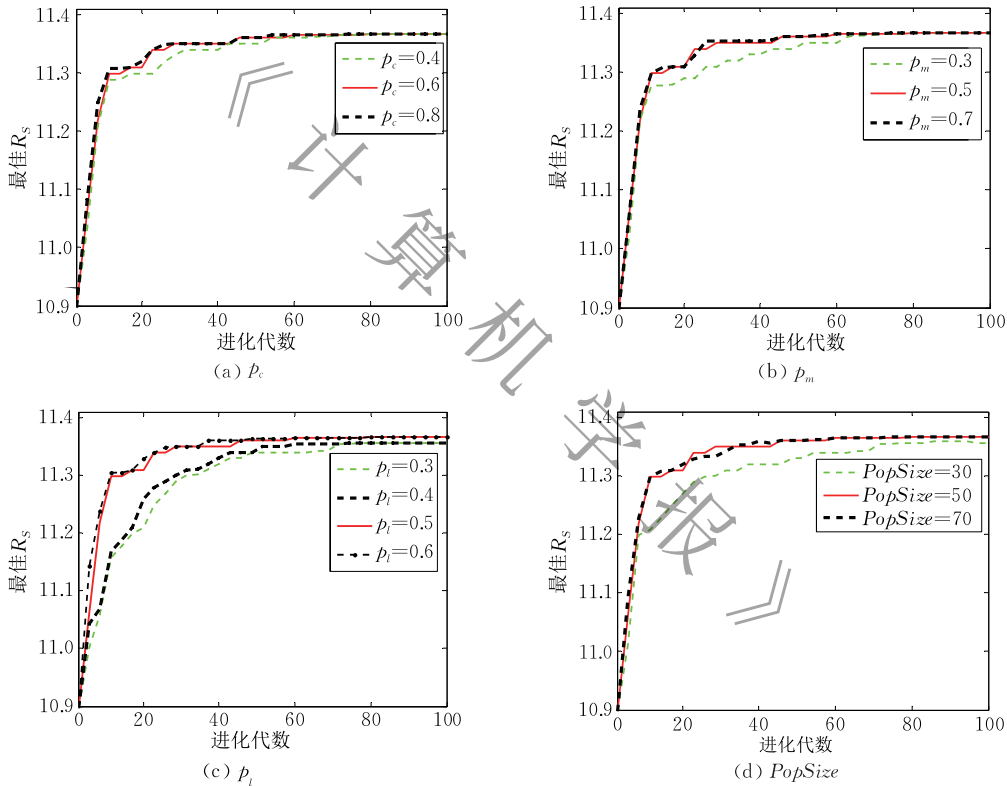


图 4 SF 网络上不同参数设置的算法收敛曲线

除了上述的 SF、ER 人工合成网络外,本文还在两种实际网络上验证了算法的性能.第一个是文献 [25] 给出的柏林某区域物流运输网络(Frid),该网络包含 224 个代表货运站的节点,376 条代表运输线路的边;第二个是文献 [26] 给出的某高校计算机专业学生社会关系网络(CS),该网络包含 1025 个代表不同学生的节点,1043 条代表学生间互动关系的边.以 $K=5, 10, 20$ 的实验为例,不同种子选择方案在这两个网络上获得的结果分别在表 8 和表 9 中展示.如该表所示,在 Frid 和 CS 两种实际网络上,

为了验证实验中设计参数的合理性,图 4 中的实验在具有 500 个节点的 SF 网络上对比了不同 $p_c, p_m, p_l, PopSize$ 参数设置下算法的收敛曲线.由图可知, p_c, p_m 对于算法性能影响较小,不同参数设置的搜索性能比较接近. p_l 对最终结果有一定的影响,较小 p_l 能降低局部搜索算子的执行次数,这有助于减少算法所需的计算代价,但是会降低其性能;同样,过高的 p_l 会带来额外的运算开销,且对于搜索性能的贡献有限.过小的种群规模 $PopSize$ 会抑制算法的搜索性能,而过大的 $PopSize$ 会造成运算代价的提升.合理的参数设置可以保证算法在搜索能力和运算代价之间取得均衡,这里的实验验证了本文采用参数的合理性.

本文提出的算法 MA-RIM 仍然可以取得较好的结果,算法选择的种子可以在网络受攻击时达到最佳的信息传播能力,同时在原始网络中也可以保持较好的影响力.

表 8 Frid 网络上不同 K 设置时各个方案获得的实验结果

K		MA-RIM	GA	Bet	Pag	Deg
5	R_s	5.270	5.255(+)	5.148(+)	5.181(+)	5.134(+)
	$\hat{\sigma}(S)$	5.285	5.278	5.259	5.362	5.343
	R_s	10.525	10.473(+)	10.301(+)	10.372(+)	10.274(+)
10	$\hat{\sigma}(S)$	10.540	10.537	10.477	10.673	10.549
	R_s	21.121	20.874(+)	20.562(+)	20.701(+)	20.692(+)
20	$\hat{\sigma}(S)$	21.001	20.980	20.796	21.117	21.117

表 9 CS 网络上不同 K 设置时各个方案获得的实验结果

K	MA-RIM	GA	Bet	Pag	Deg
5	R_s	5.639	5.632(≈)	5.326(+)	5.575(+)
	$\hat{\sigma}(S)$	6.140	6.137	5.702	6.314
10	R_s	11.202	11.191(+)	10.562(+)	11.121(+)
	$\hat{\sigma}(S)$	12.073	11.963	10.969	12.152
20	R_s	23.137	22.195(+)	20.935(+)	22.138(+)
	$\hat{\sigma}(S)$	23.356	23.329	21.774	23.458

这里获得的结果可以为决策者在解决实际问题时提供一定的参考. 对于 Frid 网络来说, 信息传播能力较强的种子节点可作为物流中转中心的备选, 这样的节点可以在短时间内对于相连的运输节点产生影响, 在一些情况下提升整个系统的运输效率^[21,25]. 对于 CS 网络来说, 从该网络中选择的种子节点一般为该社会网络中的关键个体, 比如担任领导职务或者科研业务骨干成员, 通过这些节点可以更加快速地在网络中传播诸如任务指派、学术前沿介绍等信息, 借此加强管理者对于整个网络的管控. 综上, 无论对于常见的、具有明显度分布特征的 ER、SF 等合成人工网络, 还是一些实际网络, 本文设计的算法都可为鲁棒影响力最大化问题的解决提供一定的参考与借鉴.

6 总结和展望

本文在目前网络上影响力最大化问题研究的基础上, 进一步将鲁棒性的概念引入这一问题. 针对网络结构遭到破坏的应用场景, 旨在从网络中选择具有稳定信息传播能力的种子节点. 首先, 结合已有工作, 本文提出了一个能够数值化衡量种子在网络结构受损时信息传播能力的指标因子, 并通过实验将因子中的参数定标. 更进一步, 设计了一个基于种群搜索机制的鲁棒影响力最大化问题优化算法——MA-RIM, 该算法将网络节点的局部邻域信息与全局重要节点信息相结合, 以达到获得可靠种子节点集的目的. 在实验中, 该算法展示了较好的性能, 为一些实际问题的解决提供了可选择的解决方案.

在 R_s 的设计中, 本文通过在合成网络上的实验确定了该因子的参数. 这个参数一定程度上决定了优化过程的偏好. 对于不同的优化问题、网络输入、应用场景, 可能需要采用不同的参数设置, 而非实验中采用的固定参数. 我们将在未来的工作中继续探究这一问题. 另外, 如何在大规模社交网络^[27-28]、多层网络^[29]、有向网络^[22,30]等具有更为复杂结构特性

的网络系统中鲁棒地解决影响力最大化问题同样值得关注.

参 考 文 献

[1] Newman M E J. Networks: An Introduction. Oxford, UK: Oxford University Press, 2010

[2] Girvan M, Newman M E J. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(12): 7821-7826

[3] Erdős P, Rényi A. On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 1960, 5(1): 17-61

[4] Barabási A L, Albert R. Emergence of scaling in random networks. Science, 1999, 286(5439): 509-512

[5] Schneider C M, Moreira A A, Andrade J S, et al. Mitigation of malicious attacks on networks. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(10): 3838-3841

[6] Bond R, Fariss C, Jones J, et al. A 61-million-person experiment in social influence and political mobilization. Nature, 2012, 489: 295-298

[7] Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network//Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, USA, 2003: 137-146

[8] Chen W, Wang Y, Yang S. Efficient influence maximization in social networks//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, 2009: 199-208

[9] Rahimkhani K, Aleahmad A, Rahgozar M, Moeini A. A fast algorithm for finding most influential people based on the linear threshold model. Expert System Application, 2015, 42(3): 1353-1361

[10] Lee J, Chung C. A fast approximation for influence maximization in large social networks//Proceedings of the 23rd ACM SIGKDD International Conference on Companion on World Wide Web. Seoul, Korea, 2014: 1157-1162

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

[12] Zhang Xian-Li, Tang Jian-Xin, Cao Lai-Cheng. Influence maximization algorithm based on reverse PageRank. Journal of Computer Applications, 2020, 1: 96-102(in Chinese)
(张宪立, 唐建新, 曹来成. 基于反向 PageRank 的影响力最大化算法. 计算机应用, 2020, 1: 96-102)

[13] Yang J, Liu J. Influence maximization-cost minimization in social networks based on a multiobjective discrete particle

- swarm optimization algorithm. *IEEE Access*, 2018, 6(1): 2320-2329
- [14] Qiu Li-Qing, Jia Wei, Fan Xin. Influence maximization algorithm based on overlapping community. *Data Analysis and Knowledge Discovery*, 2019, 31(7): 94-102(in Chinese) (仇丽青, 贾玮, 范鑫. 基于重叠社区的影响力最大化算法. *数据分析与知识发现*, 2019, 31(7): 94-102)
- [15] Gong M, Yan J, Shen B, et al. Influence maximization in social networks based on discrete particle swarm optimization. *Information Sciences*, 2016, 367: 600-614
- [16] Wang S, Liu J, Jin Y. Finding influential nodes in multiplex networks using a memetic algorithm. *IEEE Transactions on Cybernetics* 2021, 51(2): 900-912.
- [17] Wang S, Liu J. Constructing robust community structure against edge-based attacks. *IEEE Systems Journal*, 2019, 13(1): 582-592
- [18] Buldyrev S, Parshani R, Paul G, et al. Catastrophic cascade of failures in interdependent networks. *Nature*, 2010, 464: 1025-1028
- [19] Zhou M, Liu J. A memetic algorithm for enhancing the robustness of scale-free networks against malicious attacks. *Physica A*, 2014, 410: 131-143
- [20] He X, Kempe D. Stability and robustness in influence maximization. *ACM Transactions on Knowledge and Discovery Data*, 2018, 12(6): 1-34
- [21] Chen W, et al. Robust influence maximization//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, USA, 2016: 795-804
- [22] Wang S, Liu J. A multi-objective evolutionary algorithm for promoting the emergence of cooperation and controllable robustness on directed networks. *IEEE Transactions on Network Science and Engineering*, 2018, 5(2): 92-100
- [23] Zeng A, Liu W. Enhancing network robustness against malicious attacks. *Physical Review E*, 2012, 85: 066130
- [24] Brandes U. A fast algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 2001, 25(2): 163-177
- [25] Farid A. Symmetrica: Test case for transportation electrification research. *Infrastructure Complexity*, 2015, 2(1): 1-10
- [26] Nooy W, Mrvar A, Batagelj V. *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press, 2004
- [27] Mao Jia-Xin, Liu Yi-Qun, Zhang Min, Ma Shao-Ping. Social influence analysis for Micro-Blog user based on user behavior. *Chinese Journal of Computers*, 2014, 37(4): 791-800 (in Chinese) (毛佳昕, 刘奕群, 张敏, 马少平. 基于用户行为的微博用户社会影响力分析. *计算机学报*, 2014, 37(4): 791-800)
- [28] Goyal A, Lu W, Lakshmanan L. CELF++: Optimizing the greedy algorithm for influence maximization in social networks //*Proceedings of the 20th ACM SIGKDD International Conference on Companion on World Wide Web*. Hyderabad, India, 2011: 47-48
- [29] Wang S, Liu J. Community robustness and its enhancement in interdependent networks. *Applied Soft Computing*, 2019, 77: 665-677
- [30] Liu X, Stanley H E, Gao J. Breakdown of interdependent directed networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(5): 1138-1143



WANG Shuai, Ph. D., lecturer. His research interests include complex networks and evolutionary algorithms.

LIU Jing, Ph.D., professor. Her research interests include evolutionary computation, complex networks, fuzzy cognitive maps, multiagent systems and data mining.

Background

The complex network has attracted much attention in recent studies, and social network services are a kind of popular-used network system. Given different kinds of social networks, the determination of effective seeds that maintain a good ability to spread information, i.e., the influence maximization problem, is a challenging problem in online social marketing. Powerful seeds are desired in plenty of fields including the product promotion, rumor control, and

theory propagation.

Existing studies successfully simplify the information spreading processes in networks with the help of several propagation models. Further, the influence ability of seeds has also been defined in a numerical form. Based on these studies, the influence maximization on social networks can be tackled as an optimization problem, and a series of algorithms have been devised to select influential seeds from networks.

However, all of these existing algorithms only consider that the network system keep steady in the seed selection process, and the stability or robustness of selected seeds against structural perturbances is not touched.

For networked systems, attacks and errors may happen from time to time, and a robust network is expected to still perform well under structural failures. For the influence spreading process, some point studies have investigated the robustness of seeds towards the uncertainty of system parameters such as different spreading models or activation probability. In terms of the robustness, structural perturbances cannot be omitted for its high threats. But the method that evaluates the robustness of seeds under structural perturbances and how to solve this robust influence maximization problem remain to be addressed.

For tackling these problems, in this paper, a measure which evaluates the robustness of seeds in an information propagation process (the independent cascading model) is defined. Then, guided by this measure, the robust seed selection task is taken as an optimization problem, and a problem-directed memetic algorithm has been designed to find possible solutions to the problem. The experimental validations on several synthetic and real-world networks demonstrate the effectiveness of the proposed algorithm, and show that robust seeds can be found at the cost of a certain of the influence ability.

This paper concentrates an important performance of solutions in the influence maximization problem, which may provide better solutions for decision makers to deal with realistic dilemmas with complicated scenarios.

