

基于遗传算法的抗毁性网络拓扑结构优化*

廖汉龙

系统工程学院, 管理科学与工程 19060023

hans.liaoh@gmail.com

摘要

复杂网络的抗毁性网络拓扑结构优化是重要的研究内容。本文主要以自然连通度指标来评价网络的抗毁性, 并基于此建立了优化模型。由于模型是非线性整数规划问题 (NIP), 属于NP难问题, 因此, 根据问题的具体特点, 本文设计了参数自适应的遗传算法对问题进行了求解。通过仿真实验, 该方法收敛速度较快, 且能够对网络结构进行有效优化。特别地, 针对不同的攻击场景, 我们对优化后的网络的抗毁性进行了具体分析, 得到了网络拓扑结构优化的一般规律。

1. 背景

广义上, 网络的抗毁性是指网络中的节点或边遭受故意攻击并失效的条件下, 网络维持其正常功能的能力[3]。而在实际应用中, 这种功能的能力主要体现在如何保持拓扑结构连通。因此, 复杂网络的抗毁性从一定意义上来说, 主要来源于节点之间替代途径的冗余性[2]。

复杂网络一般描述为图, 表示为 $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, 其中, 点集合为 $\mathbf{V} = \{v_1, v_2, \dots, v_N\}$, 边集合为 $\mathbf{E} = \{e_1, e_2, \dots, e_W\}$, 图的点和边的数目分别是 N , M , 即 $N = |\mathbf{V}|, W = |\mathbf{E}|$ 。图在实际的处理场景中, 一般表示为邻接矩阵 $A(\mathbf{G}) = \{a_{ij}\}$ 。图的自然连通度表示为 $\lambda = \ln(\frac{1}{N} \sum_{i=1}^N e^{\lambda_i})$, 其中, λ_i 指的是邻接矩阵的第 i 个特征根。自然连通度可以表示网络的冗余性, 且一般认为, 网络的自然连通度越高, 则网络的冗余性越高, 相应地, 网络具有更强的抗毁性。

2. 优化模型定义

我们从复杂网络内部结构属性出发, 选取自然连通度作为网络抗毁性测度, 优化复杂网络拓扑结构。为此, 对复杂网络图 \mathbf{G} 作如下基本假设:

- 复杂网络图是无权图, 即 $a_{ij} = \begin{cases} 0, & (v_i, v_j) \notin \mathbf{E} \\ 1, & (v_i, v_j) \in \mathbf{E} \end{cases}$;
- 复杂网络图 \mathbf{G} 是无向简单图, 即 $a_{ij} = \begin{cases} a_{ji}, & i \neq j \\ 0, & i = j \end{cases}$;

*本文为国防科大系统工程学院2019-2020学期智能优化算法课程论文, 任课教师: 陈宇宁老师。

- 复杂网络图是连通图，即其拉普拉斯矩阵的次小特征根 μ 大于0。当网络规模增大时，次小特征根求解时间过长，为了简便处理，我们定义函数 $C(\mathbf{G}) \begin{cases} > 0, & \mu > 0 \\ \leq 0, & \mu \leq 0 \end{cases}$ ， $C(\mathbf{G})$ 通过深度优先搜索为每一个节点进行遍历，若所有节点遍历结果都能联通任意节点，则图是连通图，否则是非连通图；

网络的抗毁性受很多因素的影响，其中最主要的因素是网络中边的数目。自然连通度关于边添加是严格单调递增的，这意味着如果没有边的数量限制，完全图将是抗毁性最优的网络。但是，在实际情况下，构造一个网络具有一定的成本约束，边的数量越多网络的成本越大。因此，有必要研究给定边数量 W 情况下网络的抗毁性问题，我们假设网络中边数量满足约束条件：

$$W = |\mathbf{E}| = \frac{1}{2} \sum_{i=0}^N \sum_{j=0}^N a_{ij}, \quad (1)$$

根据上述分析可建立复杂网络拓扑结构优化模型：

$$\begin{aligned} \max_{\mathbf{G}} \quad & \lambda = \ln\left(\frac{1}{N} \sum_{i=1}^N e^{\lambda_i}\right) \\ s.t. \quad & \sum_{i=0}^N \sum_{j=0}^N a_{ij} = 2W \\ & a_{ij} = 0 \text{ or } a_{ij} = 1 \\ & a_{ii} = 0 \\ & a_{ij} = a_{ji} \\ & C(\mathbf{G}) > 0 \end{aligned} \quad (2)$$

3. 模型求解

模型（2）是非线性0-1整数规划问题，属于典型的NP难问题，无法使用目前的一些求解器直接进行求解，因此我们考虑使用启发式的搜索方式。由于图拓扑结构的优化问题规模较大，若使用局部搜索方法存在优化速度慢和容易陷入局部最优的挑战，因此我们优先选择群智能算法。由于网络图的邻接矩阵为0-1矩阵，为了编码方式的便捷性，我们选择了遗传算法进行问题的搜索。

由于问题求解的规模较大，直接使用遗传算法容易出现收敛速度慢的问题，因此，有必要对遗传算法做改进，具体如下。

(1)局部搜索：在每一次的遗传搜索迭代结束之后，对种群中的每一个染色体进行局部搜索优化，即模因算法。局部搜索的邻域动作设置为对当前染色体所表示的邻接矩阵的任意一位取反操作，可表示为函数 $\mathcal{N}(\mathbf{G})$ ，局部搜索对染色体的更新具体表示为：

$$\mathbf{G}_k^{(next)} = \begin{cases} \mathcal{N}(\mathbf{G}_k), & \mathcal{F}(\mathbf{G}_k) < \mathcal{F}(\mathcal{N}(\mathbf{G}_k)) \\ \mathbf{G}_k, & \mathcal{F}(\mathbf{G}_k) \geq \mathcal{F}(\mathcal{N}(\mathbf{G}_k)) \end{cases} \quad (3)$$

其中， $\mathcal{F}(\mathbf{G})$ 表示染色体的适应度。我们固定每个染色体的局部优化次数，如50次。

(2)参数自适应:为了同时保证算法在训练时前期的多样性和后期的集中性，我们对交叉概率 P_c 和变异概率 V_c 做自适应性的调整：

$$P_c^{(i)} = P_c \rho_1^i \quad (4)$$

$$V_c^{(i)} = V_c \rho_2^i \quad (5)$$

公式（4）中的 ρ_1 和公式（5）中的 ρ_2 分别是小于1的常数， i 表示迭代次数。

4. 模型约束

由于优化模型含有大量的约束条件，且多数为较强的约束，比如对图的边的数目的限制为常数，此类约束将使得多数染色体不符合条件而被淘汰，在前期的迭代中导致子代染色体数目过少甚至为0。因此在本文中，我们结合使用修复不可行解和惩罚函数法对问题约束条件进行处理，具体如下：

（1）网络图的无向简单图约束。在种群初始化，交叉，变异和局部优化中，均可能造成图的邻接矩阵不满足无向简单图约束，如 $a_{ij} \neq a_{ji}$ ，因此，我们在每次涉及到改变图结构的操作之后，对邻接矩阵进行逐行扫描修复。

（2）网络图的边数目约束。由于几乎任何操作都将会改变图的边的数目，因此我们选择对图进行不可行解的修复。修复方式为：假设边的数目为 $W' = \frac{1}{2} \sum_{i=0}^N \sum_{j=0}^N a_{ij}$ ， $\Delta W = |W' - W|$ ，若 $W' - W > 0$ ，则随机选择邻接矩阵中的 ΔW 个元素 $\{a_1, a_2, \dots, a_{\Delta W}\}$ ，集合中的每一个元素值都为1，且在原邻接矩阵中的下标均满足 $j \geq i$ ，将集合中的元素的值变为0，而后做满足（1）中的无向简单图修复操作。反之，若 $W' - W < 0$ ，做相反操作。

（3）图的连通性约束。对于图的连通性的修复较为困难，另外经过实验，我们发现随机生成的图，其是连通图的概率在95%以上，这意味着只有较少数的图不是连通图。因此在训练过程中，如果解不满足连通性约束，则放弃此解，即使用惩罚函数法，对不满足约束的解，其适应度值设为负无穷。

5. 仿真分析

5.1. 参数设置

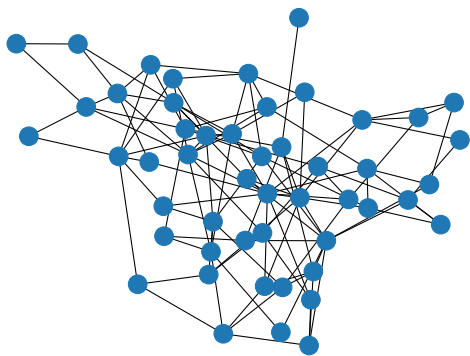
在仿真过程中，我们选择ER网络为研究对象[1]，基本参数设定为：节点数 $N = 50$ ，边数 $W = 150$ ，边连接概率 $p = 0.12$ ，生成的ER网络（连通图）拓扑结构与度分布如图1所示。采用遗传算法求解网络拓扑结构优化模型的相关参数说明及取值如表 1，仿真运行环境为：64位Windows10操作系统，python3.7，处理器为Intel i7 9570H，主频2.6GHz，内存16GB。

表 1. 算法基本参数

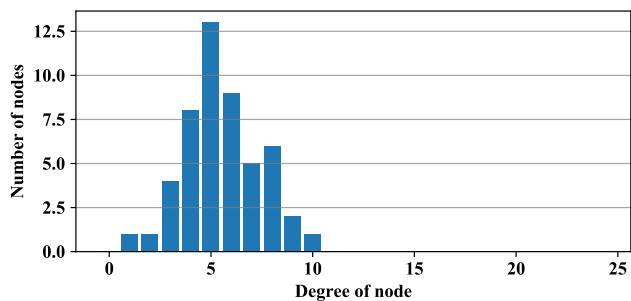
参数	含义	取值	参数	含义	取值
W	节点数目	50	ρ_1	交叉概率递减参数	0.995
POP_NUM	种群大小	100	V_c	变异概率	0.008 - 0.012
$ITER_TIME$	遗传迭代次数	60	ρ_2	变异概率递减参数	0.995
P_c	交叉概率	0.88-0.92	$ITER_LOCAL$	局部优化搜索迭代次数	50

5.2. 网络拓扑结构分析

以自然连通度作为复杂网络抗毁性测度优化网络拓扑结构，在保持边数量不变情况下，ER图1(a) 优化后的拓扑结构及其度分布见图2。随着进化代数的增加，复杂网络自然连通度值 λ 不断增加，最优值为6.57在第41代时获得，见图3。因此，遗传算法可用于求解复杂网络拓扑优化模型，而且运用了算法收敛速度快等优点。通过对复杂网络拓扑结构优化的仿真分析，可得到以下基本结论：

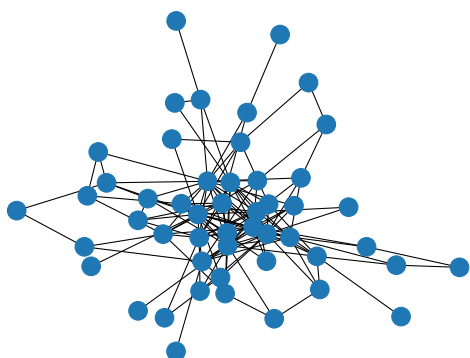


(a) 随机ER图网络拓扑结构

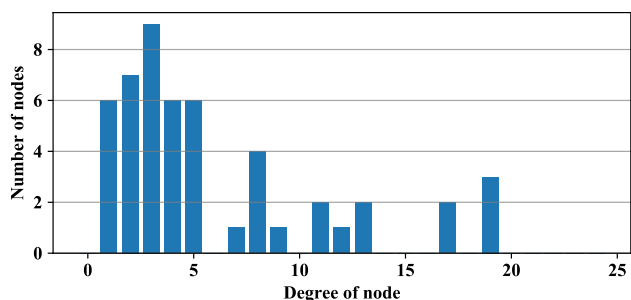


(b) 随机ER图度分布

图 1. 随机ER图的拓扑结构与度分布



(a) 优化后的ER图网络拓扑结构



(b) 优化后的ER图度分布

图 2. 优化后的ER图的拓扑结构与度分布

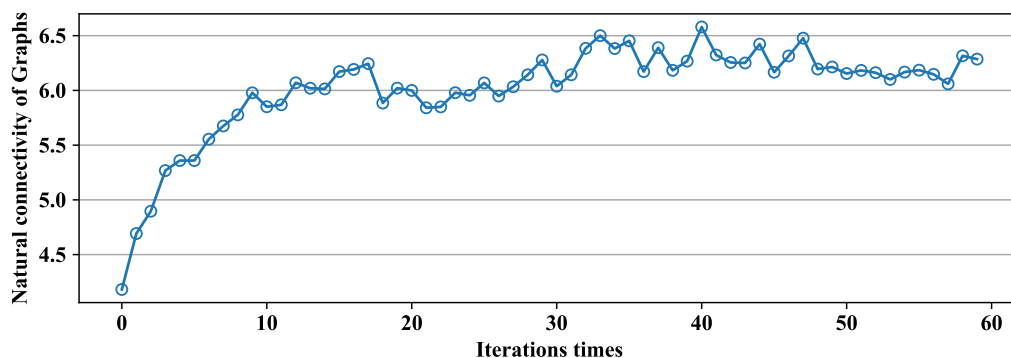


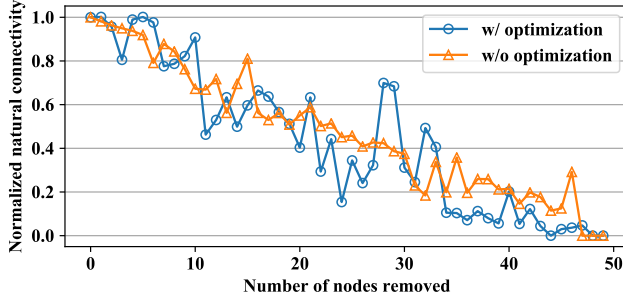
图 3. 自然连通度的迭代优化

(1) 仿真分析验证了复杂网络拓扑优化模型，以及采用遗传算法求解模型的有效性和可行性。随着拓扑结构的不断优化，自然连通度值 λ 不断增加，即表明网络抗毁性不断增强。

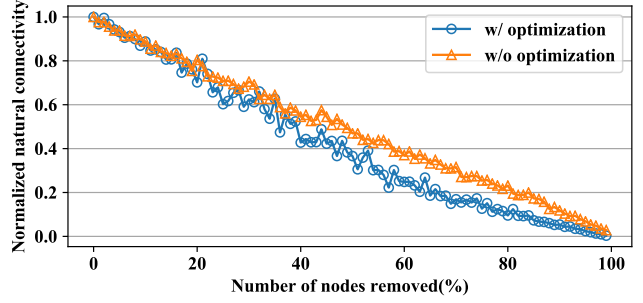
(2) 网络拓扑结构优化后，改变了网络度分度，其中度小节点比例增加，而度大节点比例减小。

(3) 根据图1(b)和图2(b)可知，ER网络度分布呈二项分布特点，而优化后度分布由低到高基本呈现递减

趋势，即度由低到高，而节点比例呈下降趋势。因此，ER网络拓扑结构优化之后，度小的节点倾向于和度大节点相连。

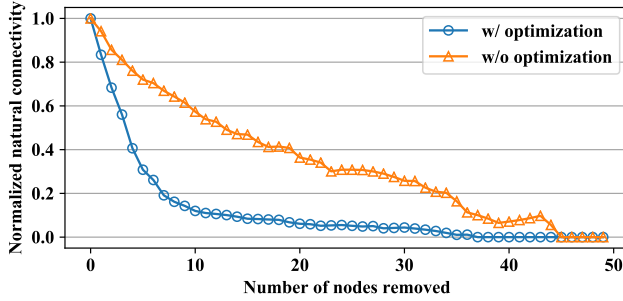


(a) 随机去除节点

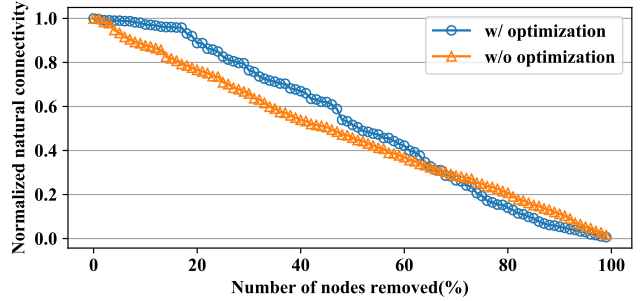


(b) 随机去除边

图 4. 随机去除网络的节点与边



(a) 恶意去除节点



(b) 恶意去除边

图 5. 恶意去除网络的节点与边

6. 抗毁性分析

复杂网络拓扑优化的目的是提高网络的抗毁性，上述分析以自然连通度为网络抗毁性测度建立拓扑优化模型，并采用遗传进化算法求解并提高了网络的抗毁性。在此，分别采取随机（去点、去边）攻击和恶意（去点、去边）两种种策略，针对复杂网络拓扑优化前后实施攻击，分析复杂网络抗毁性变化。

通过仿真分析可知，复杂网络拓扑优化后面临不同攻击策略具有以下特点：1）在随机去点攻击策略中（图4(a)），优化后网络抗相对于优化前在收到抗攻击是更加不稳定，若增大网络规模，将呈现更高的抗毁性。

2）在随机去边攻击策略中（图4(b)），优化前后网络抗攻击能力相似，而当去边比例大于约为0.5，优化后网络抗攻击能力下降速度较优化前稍快。

3）在按度去点攻击策略中（图5(a)），该攻击方法将按节点度从大到小依次去除，优化后网络在面临该攻击方法表现出较差的抗毁能力，去除少数（约10%）度较大的节点网络抗毁性则大幅下降（约90%），网络几乎已经瘫痪；而优化前网络则表现出相对较强的抗毁攻击能力。

4) 在按介数去边攻击策略中(图5(b)), 按边的介数大小依次去边, 优化后网络面临该攻击方法表现出相对较强的抗毁能量, 去除少数(约20%)介数较大边, 对网络抗毁性影响有限。分析其原因在于, 优化后网络中度小的节点倾向于和度大节点相连, 从而导致连接度大节点之间边的介数也相对较大, 因此边的冗余性较强; 而针对优化前的网络结构, 反之亦然。

7. 总结

复杂网络拓扑优化是复杂网络研究中重要内容, 论文建立了复杂网络拓扑优化模型, 提出了基于遗传进化算法的求解方法, 并通过仿真分析验证了模型与方法的可行性。但通过优化前后复杂网络抗毁性对比分析, 发现拓扑优化过程中存在矛盾: 尽管网络拓扑优化提高了网络抗毁性, 以及复杂网络面临随机去点攻击和按介数去边攻击抗毁能力, 但在面临按度去点攻击时, 网络表现出较差的抗毁攻击能力。因此, 在实际应用当中, 需要根据复杂网络面临的现实威胁选择不同的网络拓扑结构, 以及优化其拓扑结构

参考文献

- [1] L. M. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021–1024, 1994.
- [2] 吴俊. 复杂网络拓扑结构抗毁性研究. PhD thesis, 长沙: 国防科学技术大学, 2008.
- [3] 谭跃进, 吴俊, 邓宏钟, 朱大智. 复杂网络抗毁性研究综述. *系统工程*, 24(10):1–5, 2006.