文章编号: 1000-6788(2010)07-1207-11 中图分类号: N949 文献标志码: A

基于不等概率抽样的不完全信息条件下复杂网络抗毁性模型

吴 俊, 谭跃进, 邓宏钟, 李 勇, 刘 斌

(国防科技大学 信息系统与管理学院 管理系, 长沙 410073)

摘 要 为了填补随机失效与故意攻击之间的空白, 将复杂网络攻击信息的获取抽象成无放回的不等概率抽样问题, 建立了不完全信息条件下的复杂网络抗毁性模型. 其中网络攻击信息可以通过信息广度参数和信息精度参数调节控制, 随机失效或故意攻击是该模型的两个特例. 利用母函数方法解析推导出了任意度分布广义随机网络在随机不完全信息和优先不完全信息条件下的两个重要抗毁性度量参数 —— 临界移除比例和巨组元规模, 得到的解析结果可以分析和预测不完全信息条件下复杂网络的抗毁性. 以无标度网络为例对一般攻击信息参数组合进行了仿真分析, 发现随机隐藏少量节点信息将大幅度提高复杂网络的抗毁性, 获取少量重要节点的信息可以大幅度降低复杂网络的抗毁性.

关键词 复杂网络; 抗毁性; 不等概率抽样; 不完全信息

Model for invulnerability of complex networks with incomplete information based on unequal probability sampling

WU Jun, TAN Yue-jin, DENG Hong-zhong, LI Yong, LIU Bin

(Department of Management, College of Information Systems and Management, National University of Defense Technology, Changsha 410073, China)

Abstract To bridge the gap between random failure and intentional attack, a novel model for invulnerability of complex networks with incomplete attack information is proposed by considering the process of acquiring attack information as the unequal probability sampling. The attack information can be controlled by a tunable attack information accuracy parameter and a tunable attack information range parameter. The known random failure and the intentional attack are two extreme cases of our model. Using the generating function method, the analytical expressions of the critical removal fraction of vertices for the disintegration of networks and the size of the giant component under random incomplete information and preferential incomplete information are derived. The results allow us to make predictions on the invulnerability of complex networks under attack with incomplete information. Taking scale-free networks for example, the invulnerability of complex networks with general incomplete attack information is studied numerically. It is shown that hiding just a small fraction of vertices randomly can enhance the invulnerability and detecting just a small fraction of important vertices preferentially can reduce the invulnerability.

Keywords complex networks; invulnerability; unequal probability sampling; incomplete information

1 引言

21 世纪以来, 以信息技术的飞速发展为基础, 人类社会加快了网络化进程. 交通网络、通信网络、电力网络、物流网络 · · · · ·, 可以说我们被网络包围着, 我们赖以生存的网络越来越庞大, 越来越复杂. 但越来越频繁发生的事故也将一系列严峻的问题摆在我们面前: 这些网络到底有多可靠? 一些微不足道的事故隐患是

收稿日期: 2009-03-19
资助项目: 国家自然科学基金 (70501032, 70771111, 60904065)
作者简介: 吴俊 (1980–), 博士, 主要研究方向为复杂网络理论及其应用; 谭跃进 (1958–), 博士生导师, 主要研究方向为系统理论与综合集成; 邓宏钟 (1974–), 博士, 主要研究方向为复杂系统理论, 分布式人工智能和遗传算法; 李勇 (1979–), 博士研究生, 主要研究方向为复杂网络理论及其应用; 刘斌 (1982–), 博士研究生, 主要研究方向为复杂网络理论及其应用.

否会导致整个网络系统的崩溃? 在发生严重自然灾害或者敌对势力蓄意破坏的情况下, 这些网络是否还能正常发挥作用? 这些正是复杂网络抗毁性研究需要面对的问题. 随着复杂网络研究的兴起[1−8], 作为复杂网络最重要的研究问题之一, 复杂网络抗毁性研究的重大理论意义和应用价值日益凸显出来, 成为极其重要而且富有挑战性的前沿科研课题[9−11].

1999 年 Albert 等在《Nature》上发表经典论文研究了不同度分布复杂网络的抗毁性[12]. 他们考察了两种失效模式: 随机失效 (Random failure), 即随机地移除网络中的节点; 故意攻击 (Intentional attack), 即按照节点度从大到小的顺序移除节点, 通过观察节点移除过程中网络性能的变化以及网络状态的相变来刻画网络的抗毁性. Albert 等研究发现, 在随机失效下, 无标度网络相对随机网络有着更强的抗毁性, 但是无标度网络面对故意攻击显得异常脆弱. Albert 等的研究激起了大量研究人员对网络抗毁性的兴趣[13−19], 该论文在不到 10 年时间里已经被引用超过了 2000 次. 但目前大部分抗毁性研究工作都未能跳出 Albert 等原始工作的研究框架, 即研究不同网络在随机失效或者故意攻击下的抗毁性. 所谓随机失效和故意攻击, 从攻击信息角度来看, 就是零信息攻击和完全信息攻击. 显然, 在现实世界的复杂网络中, 随机失效和故意攻击只是两种极端情况, 我们面临更多的情况是不完全信息攻击, 即部分信息已知, 部分信息未知.

为了扩展现有基于随机失效和故意攻击的抗毁性模型, 本文将复杂网络攻击信息获取抽象成无放回的不等概率抽样问题, 建立不完全信息条件下的复杂网络抗毁性模型, 并对其进行解析和仿真分析.

2 不完全信息条件下复杂网络抗毁性模型

复杂网络在数学上可以描述成一个图 $G=(V,E)$, 其中 $V=\{v_1,v_2,v_3,\cdots,v_N\}$ 表示节点 (Vertex) 集合, $E=\{e_1,e_2,\cdots,e_W\}\subseteq V\times V$ 表示边 (Edge) 的集合, $N=|V|$ 表示节点数量, $W=|E|$ 表示边数量. 建立不完全信息条件下的复杂网络抗毁性模型需要解决两个问题: 1) 定量刻画攻击信息; 2) 明确不完全信息条件下的攻击模型.

假设我们要攻击网络中的 Nf 个节点, 其中 f 表示节点移除比例. 当我们不能获取网络的任何信息时, 我们只能在网络中随机的攻击 Nf 个节点, 这就是前面提到的随机失效. 但如果我们能获取网络的全部信息时, 那我们则会按照节点的重要程度的大小排序选择性地攻击 Nf 个节点, 这就是前面提到的故意攻击. 因此, 我们考虑将节点的重要度作为网络的攻击信息. 目前, 关于节点重要度的评估有很多方法, 例如介数法[20], 节点删除法[21], 节点收缩法[22] 等等. 但最简单也是使用最广泛的节点重要度指标就是节点的度.

令网络中节点 v_i 的重要度为 I_i, 将所有节点按照重要度 I_i 排序, 令节点 v_i 的序号为 r_i. 我们用 ∇_i 表示节点的信息获取状态, 即若 v_i 的重要度 I_i 已知, 则 $\nabla_i=1$, 否则 $\nabla_i=0$. 我们称所有已被获取信息节点的集合为 “已知区域 Ω”, 即 $\Omega=\{v_i|\nabla_i=1,v_i\in V\}$, 称所有未被获取信息节点的集合为 “未知区域 $\bar{\Omega}$”, 即 $\bar{\Omega}=\{v_i|\nabla_i=0,v_i\in V\}$. 这样, 攻击信息的度量问题就转化成已知区域 Ω 的确定问题. 如图 1 所示, 阴影部分表示攻击信息未知区域 $\bar{\Omega}$, 非阴影部分表示攻击信息已知区域 Ω, 其中 $\Omega=\emptyset$ 对应零信息攻击, $\Omega=V$ 对应完全信息攻击.

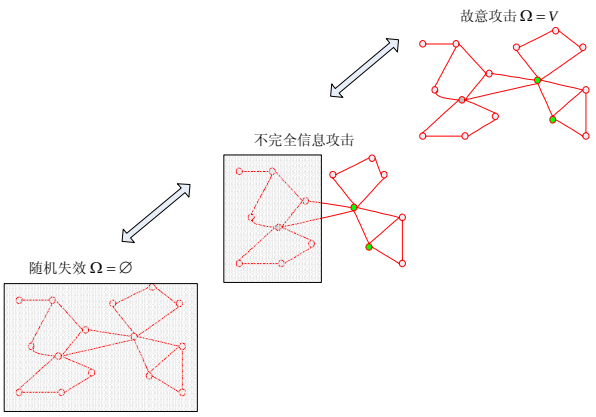

图 1 不完全信息攻击示意图

要确定 Ω 需要解决两个问题: 1) Ω 包含多少个节点, 即攻击信息的广度; 2) Ω 包含哪些节点, 即攻击信息的精度. 为了同时攻击控制信息的精度和广度, 我们把 Ω 的确定转化成 “不等概率抽样问题”. 不等概率抽样有时也被称为非概率抽样或非随机抽样, 它指总体中某一单元的入样概率与该单元的某一辅助变量大小成正比[23]. 在这里, 总体对应节点集合 V, 总体容量为 N; 样本对应已知区域 Ω, 样本容量为 $n=N\alpha$, 其中 $\alpha\in[0,1]$ 为攻击信息广度参数. α 越大, 获取的攻击信息越多. 考虑两种极端情况:

i) 当 $\alpha=0$ 时:

$$n=N\alpha=0 \tag{1}$$

即攻击信息量为零, 对应随机失效;

ii) 当 $\alpha = 1$ 时:

$$n = N\alpha = N \tag{2}$$

即攻击信息量为完全信息, 对应故意攻击.

构造节点 v_i 的辅助变量如下:

$$\pi_i = r_i^{-\delta} \tag{3}$$

其中 $\delta \in [0, \infty)$ 为攻击信息精度参数. 由式 (3) 可得单次抽样 $(n = 1)$ 节点 v_i 的入样概率为

$$\nabla_i = \frac{\pi_i}{\sum\limits_{t=1}^{N} \pi_t} = \frac{r_i^{-\delta}}{\sum\limits_{t=1}^{N} r_t^{-\delta}} \tag{4}$$

显然, δ 越大, 越可能获取到那些重要节点的信息, 即获取的攻击信息精度越高. 考虑两种极端情况:

i) 当 $\delta = 0$ 时:

$$\nabla_i = \frac{r_i^{-\delta}}{\sum_{t=1}^{N} r_t^{-\delta}} = \frac{1}{N} \tag{5}$$

即节点被获取信息的概率相等, 我们称这种攻击信息为随机不完全信息.

ii) 当 $\delta = \infty$ 时:

$$\sum_{t=1}^{N} r_t^{-\infty} = \sum_{t=1}^{N} t^{-\infty} = 1 + \sum_{t=2}^{N} t^{-\infty} = 1 \tag{6}$$

假设 $r_{i^*} = 1$, 则

$$\Delta_i = \begin{cases} 1, & i = i^* \\ 0, & i \neq i^* \end{cases} \tag{7}$$

即最重要的节点信息总是被优先获取, 我们称这种攻击信息为优先不完全信息.

为了避免重要度高的节点重复入样, 将攻击信息的获取过程抽象成如下无放回的不等概率抽样问题:

Step 1 按照 (4) 式中的概率抽取一个样本;

Step 2 将剩余节点按重要度排序并重新计算辅助变量 π_i 和入样概率 ∇_i;

Step 3 重复 Step 1 和 Step 2 直至抽出 n 个样本.

假设已经确定已知区域 Ω $(n = |\Omega| = N\alpha)$, 需要攻击网络中的 Nf 个节点, 节点被攻击后与其相连接的边随之移除. 考虑一种最简单的攻击模式: 先攻击已知信息节点, 再攻击未知信息节点, 即

i) 若 $f \leq \alpha$, 直接在已知区域 Ω 中按照节点的重要度从大到小依次攻击;

ii) 若 $f > \alpha$, 先把已知区域 Ω 中的节点全部攻击, 然后在未知区域 $\bar{\Omega}$ 随机攻击 $N(f - \alpha)$ 个节点.

3 不完全信息条件下复杂网络抗毁性解析分析

对于一般的复杂网络, 解析分析不完全信息条件下的抗毁性非常困难, 所以本文仅解析分析具有任意度分布的广义随机网络 [24] 的抗毁性, 即假设网络在满足度分布 $p(k)$ 条件下随机连接. 下面首先利用概率母函数方法 [13,24] 解析推导不完全信息条件下广义随机网络的两个重要抗毁性度量参数: 临界移除比例和巨组元规模, 进而解析分析两种特殊情况 (随机信息和优先信息) 下复杂网络的抗毁性. 为了便于推导, 这里选取节点的度作为节点的重要度指标.

3.1 临界移除比例和巨组元规模

给定度分布 $p(k)$, 可得到其概率母函数

$$g_0(x) = \sum_{k=m}^{M} p(k) x^k \tag{8}$$

其中 m 为最小度, M 为最大度. 令沿随机选择的一条边到达的节点度为 k 的概率分布为 $p_E(k)$, 其不仅与度为 k 的节点数量成正比, 还与 k 本身成正比. 因此,

$$p_E(k) = \frac{kp(k)}{\sum_{k=m}^{M} kp(k)} \tag{9}$$

从而, 可得 $p_E(k)$ 的概率母函数为

$$g_E(x)=\sum_{k=m}^{M}p_E(k)x^k=\frac{\sum_{k=m}^{M}kp(k)x^k}{\sum_{k=m}^{M}kp(k)}=\frac{g_0'(x)}{g_0'(1)}x \tag{10}$$

当沿随机选择的一条边到达度为 k 的节点后, 还剩 $k-1$ 条边可以连接其它节点, 这就是剩余度的概念[24]. 其实所谓剩余度就是指节点的度减去 1, 引入剩余度可以方便推导. 令剩余度的概率分布为 $p_R(k)$, 由剩余度的定义易知

$$p_R(k)=p(k+1) \tag{11}$$

因此, 沿随机选择的一条边到达剩余度为 k 的概率分布为

$$p_1(k)=\frac{(k+1)p(k+1)}{\sum_{k=m-1}^{M-1}(k+1)p(k+1)}=\frac{(k+1)p(k+1)}{\sum_{k=m}^{M}kp(k)} \tag{12}$$

其概率母函数为

$$\begin{aligned}g_1(x)&=\sum_{k=m-1}^{M-1}p_1(k)x^k=\frac{\sum_{k=m-1}^{M-1}(k+1)p(k+1)x^k}{\sum_{k=m}^{M}kp(k)}\\&=\frac{\sum_{k=m}^{M}kp(k)x^{k-1}}{\sum_{k=m}^{M}kp(k)}=\frac{g_0'(x)}{g_0'(1)}=\frac{1}{<k>}g_0'(x)\end{aligned} \tag{13}$$

令度为 k 的节点未被攻击的概率为 $q(k)$, 则随机选择一个节点, 其节点度为 k 且没有被攻击的概率为 $w_0(k)=p(k)q(k)$, 其概率母函数为

$$F_0(x)=\sum_{k=m}^{M}p(k)q(k)x^k \tag{14}$$

同理, 沿随机选择的一条边到达一个剩余度为 k 且没有被攻击的节点的概率为 $w_1(k)=p_1(k)q(k+1)$, 其概率母函数为

$$\begin{aligned}F_1(x)&=\sum_{k=m-1}^{M-1}p_1(k)q(k+1)x^k=\sum_{k=m}^{M}\frac{kp(k)}{\sum_{k=m}^{M}kp(k)}q(k)x^{k-1}\\&=\frac{\sum_{k=m}^{M}kp(k)q(k)x^{k-1}}{\sum_{k=m}^{M}kp(k)}=\frac{F_0'(x)}{<k>}\end{aligned} \tag{15}$$

令 $h_1(k)$ 表示沿随机选择的一条边到达的连通片 s 的规模为 k 的概率, 其概率母函数为

$$H_1(x)=\sum_{k=0}^{\infty}h_1(k)x^k \tag{16}$$

当网络中含有巨组元时, 我们假设 $h_1(k)$ 不包括巨组元[24]. 所谓 "巨组元" 指的是包含网络中大多数节点的连通片, 即几乎一定 $|S|=\Theta(N)$[25−26]. 当沿随机选择的一条边到达的节点已被攻击时, 到达的连通片规模为零, 其概率为

$$h_1(0)=1-\sum_{k=m-1}^{M-1}p_1(k)q(k+1)=1-F_1(1) \tag{17}$$

当沿随机选择的一条边到达的节点未被攻击时, 到达的连通片规模大于零. 当到达节点的剩余度为 0 时, 到达的连通片规模为 1; 当到达节点的剩余度为 1 时, 到达的连通片规模为到达节点连接的分支规模再加 1; 当到达节点的剩余度为 2 时, 到达的连通片规模为到达节点连接的两个分支规模之和再加 1, $\cdots$ 依此类推. 因此, $H_1(x)$ 可表示为如下递归形式:

$$\begin{aligned}H_1(x)&=1-F_1(1)+xp_1(1)q(1)H_1(x)+xp_1(2)q(2)[H_1(x)]^2+\cdots\\&=1-F_1(1)+x\sum_{k=m}^{M}p_1(k)q(k)[H_1(x)]^k\\&=1-F_1(1)+xF_1[H_1(x)]\end{aligned} \tag{18}$$

同理, 令 $h_0(k)$ 表示随机选择的一个节点所属连通片的规模为 k 的概率, 其概率母函数为

$$H_0(x)=\sum_{k=0}^{\infty}h_0(k)x^k \tag{19}$$

当网络中含有巨组元时, 同样假设 $h_0(k)$ 不包括巨组元. 当沿随机选择的节点已被攻击时, 所属连通片规模为零, 其概率为

$$h_0(0)=1-\sum_{k=m}^{M}p(k)q(k)=1-F_0(1) \tag{20}$$

$H_0(x)$ 满足如下递归形式

$$H_0(x)=1-F_0(1)+xF_0[H_1(x)] \tag{21}$$

这样, 给定度分布 $p(k)$ 以及未被攻击概率 $q(k)$, 我们就可得到 $F_0(x)$ 和 $F_1(x)$, 将 $F_1(x)$ 代入 (18) 式即可解出 $H_1(x)$, 将 $F_0(x)$ 和 $H_1(x)$ 代入 (21) 式即可解出 $H_0(x)$, 从而可得到所有连通片规模的概率分布. 但实际上, 解析求解方程 (18) 是非常困难的, 大多数时候它都是超越方程, 没有显式解. 虽然很难解析得到连通片规模的概率分布, 但我们可以通过 (18) 式和 (21) 式得到临界移除比例 f_c 和巨组元规模 S.

假设网络中不存在巨组元, 则

$$H_1(1)=\sum_{k=0}^{\infty}h_1(k)=1 \tag{22}$$

$$H_0(1)=\sum_{k=0}^{\infty}h_0(k)=1 \tag{23}$$

从而, 可得平均连通片规模

$$<s>=H_0'(1)=F_0[H_1(1)]+F_0'(1)H_1'(1)=F_0(1)+F_0'(1)H_1'(1) \tag{24}$$

又由 (18) 式可知

$$H_1'(1)=F_1[H_1(1)]+F_1'(1)H_1'(1)=F_1(1)+F_1'(1)H_1'(1) \tag{25}$$

解得

$$H_1'(1)=\frac{F_1(1)}{1-F_1'(1)} \tag{26}$$

从而

$$<s>=F_0(1)+F_0'(1)H_1'(1)=F_0(1)+\frac{F_0'(1)F_1(1)}{1-F_1'(1)} \tag{27}$$

由 (27) 式可以看出, 当 $F_1'(1)=1$ 时平均连通片规模 $<s>$ 发散, 这意味着巨组元存在的临界点或者网络崩溃的临界点位于 $F_1'(1)=1$, 即

$$F_1'(1)=\frac{F_0''(1)}{<k>}=\frac{\sum_{k=m}^{M}k(k-1)p(k)q(k)}{\sum_{k=m}^{M}kp(k)}=1 \tag{28}$$

由于 $q(k)$ 仅和攻击信息以及节点移除比例 f 有关, 所以给定度分布 $p(k)$, 攻击信息参数 α、δ 后, 我们可由 (28) 式解出临界移除比例 f_c.

当网络中存在巨组元 S 时, 由 $H_0^{(x)}$ 定义可知

$$|S|/N=1-H_0(1) \tag{29}$$

又由 (21) 式可知

$$H_1(1)=1-F_1(1)+F_1(H_1(1)) \tag{30}$$

由上式解得 $H_1(1)=u$, 代入 (21) 式, 得

$$H_0(1)=1-F_0(1)+F_0(u) \tag{31}$$

因此, 可得巨组元规模

$$|S|/N=1-H_0(1)=F_0(1)-F_0(u) \tag{32}$$

其中 u 为方程

$$u=1-F_1(1)+F_1(u) \tag{33}$$

的最小非负实数解.

3.2 随机不完全信息条件下的复杂网络抗毁性

若 $f \le \alpha$, 在已知区域 Ω 中按照节点的度从大到小依次移除 Nf 节点. 令 $\tilde{K}$ 表示 Ω 中未被攻击节点的最大度, 则

$$q(k) = \begin{cases} 1, & k \le \tilde{K} \\ 1-\alpha, & k > \tilde{K} \end{cases} \tag{34}$$

其中, $\tilde{K}$ 由度分布 $p(k)$ 和 f 确定. 特别地, 若 $p(k) = (\gamma-1)m^{\gamma-1}k^{-\gamma}$ $(\gamma > 2)$, 则

$$\tilde{K} \approx m\,(f/\alpha)^{-1/(\gamma-1)} \tag{35}$$

若 $f > \alpha$, 先把已知区域 Ω 中的节点全部移除, 然后在未知区域 $\bar{\Omega}$ 随机移除 $N(f-\alpha)$ 节点. 因为 Ω 中节点也是随机选择的, 所以这种情况等价于在整个网络中随机移除 Nf 节点. 因此, 未被攻击概率为

$$q(k) = 1 - f \tag{36}$$

将 (36) 式代入 (28) 式, 可得随机失效条件下的临界条件

$$(1-f)\sum_{k=m}^{M} k(k-1)p(k) = \sum_{k=m}^{M} kp(k) \tag{37}$$

从而可得随机失效条件下的临界移除比例

$$f_c^{RF} = 1 - \frac{1}{\kappa - 1} \tag{38}$$

其中 $\kappa = < k^2 > / < k >$. 特别地, 若 $p(k) = (\gamma-1)m^{\gamma-1}k^{-\gamma}$ $(\gamma > 2$ 且 $\gamma \ne 3)$, 则

$$\kappa = \left(\frac{2-\gamma}{3-\gamma}\right)\frac{M^{3-\gamma} - m^{3-\gamma}}{M^{2-\gamma} - m^{2-\gamma}} \tag{39}$$

这与文献 [19] 中的结果一致.

若 $\alpha \le f_c^{RF}$, 那么必须至少移除 $N\alpha$ 节点才能使得网络崩溃, 即 $f_c \ge \alpha$. 此时等价于随机失效, 因此 $f_c = f_c^{RF}$.

若 $\alpha > f_c^{RF}$, 可知 $f_c < \alpha$. 将 (34) 式代入 (28) 式, 可得临界条件

$$\sum_{k=m}^{\tilde{K}} k(k-1)p(k) + (1-\alpha)\sum_{k=\tilde{K}+1}^{M} k(k-1)p(k) = \sum_{k=m}^{M} kp(k) \tag{40}$$

求解上式, 可得临界值 $\tilde{K}_c$, 再由 $\tilde{K}$ 和 f 的函数关系即可得到临界移除比例 f_c. 特别地, 若 $p(k) = (\gamma-1)m^{\gamma-1}k^{-\gamma}$ $(\gamma > 2$ 且 $\gamma \ne 3)$, 式 (40) 可写为

$$\frac{\alpha\tilde{K}^{2-\gamma} - 2m^{2-\gamma} + (2-\alpha)M^{2-\gamma}}{2-\gamma} = \frac{\alpha\tilde{K}^{3-\gamma} - m^{3-\gamma} + (1-\alpha)M^{3-\gamma}}{3-\gamma} \tag{41}$$

对于故意攻击 $(\alpha = 1)$, 式 (41) 可写为

$$\frac{\tilde{K}^{2-\gamma} - 2m^{2-\gamma} + M^{2-\gamma}}{2-\gamma} = \frac{\tilde{K}^{3-\gamma} - m^{3-\gamma}}{3-\gamma} \tag{42}$$

这与文献 [16] 中的结果一致.

从 (41) 式、(42) 式可看出, 若 $\gamma > 3$, 则当 $N \to \infty$, $M^{3-\gamma} \to 0$, 从而总是存在一个有限的临界移除比例 $f_c < 1$; 但当 $2 < \gamma < 3$ 时, 若 $\alpha < 1$, 则当 $N \to \infty$, $M^{3-\gamma} \to \infty$, 从而 $f_c \to 1$. 这意味着, 在标度指数 $2 < \gamma < 3$ 的无标度网络中, 当 $N \to \infty$ 时, 如果我们能隐藏部分节点的信息 $(\alpha < 1)$, 那么几乎需要移除所有节点才能使得网络崩溃 $(f_c \to 1)$. 将 (34) 式、(36) 式代入 (14) 式、(15) 式可得 $F_0(x)$ 和 $F_1(x)$, 再代入 (32) 式即可求得巨组元规模.

3.3 优先不完全信息条件下的复杂网络抗毁性

当 $f \le \alpha$ 时, 在已知区域 Ω 中按照节点的度从大到小依次移除 Nf 节点. 因为 Ω 中节点是按照度从大到小优先选择的, 所以这种情况等价于在整个网络中按照节点的度从大到小依次移除 Nf 节点, 即故意攻击 $(\alpha = 1)$. 令 $\tilde{K}$ 表示未被攻击节点的最大度, 则未被攻击概率可写为

$$q(k) = \begin{cases} 1, & k \le \tilde{K} \\ 0, & k > \tilde{K} \end{cases} \tag{43}$$

特别地, 若 $p(k)=(\gamma-1)m^{\gamma-1}k^{-\gamma}$ $(\gamma>2)$, 则

$$\tilde{K}\approx mf^{-1/(\gamma-1)} \tag{44}$$

将 (43) 式代入 (28) 式, 可得故意攻击条件下的临界条件

$$\sum_{k=m}^{\tilde{K}}k(k-1)p(k)=\sum_{k=m}^{M}kp(k) \tag{45}$$

求解上式, 可得临界值 $\tilde{K}_c$, 再由 $\tilde{K}$ 和 f 的函数关系即可得到故意攻击条件下的临界移除比例 f_c^{IA}. 特别地, 若 $p(k)=(\gamma-1)m^{\gamma-1}k^{-\gamma}$ ($\gamma>2$ 且 $\gamma\neq3$), 则临界条件可写为

$$\frac{\tilde{K}^{2-\gamma}-2m^{2-\gamma}+M^{2-\gamma}}{2-\gamma}=\frac{\tilde{K}^{3-\gamma}-m^{3-\gamma}}{3-\gamma} \tag{46}$$

当 $f>\alpha$ 时, 先把已知区域 Ω 中的节点全部移除, 然后在未知区域 $\bar{\Omega}$ 随机移除 $N(f-\alpha)$ 节点. 令 $\tilde{m}$ 表示 Ω 中节点的最小度, 则未被攻击概率可写为

$$q(k)=\begin{cases}\dfrac{1-f}{1-\alpha}, & k<\tilde{m}\\ 0, & k\geq\tilde{m}\end{cases} \tag{47}$$

其中 $\tilde{m}$ 由度分布 $p(k)$ 和 f 确定. 特别地, 若 $p(k)=(\gamma-1)m^{\gamma-1}k^{-\gamma}$ ($\gamma>2$ 且 $\gamma\neq3$), 则

$$\tilde{m}\approx m\alpha^{-1/(\gamma-1)} \tag{48}$$

若 $\alpha\geq f_c^{IA}$, 那么仅需移除不超过 $N\alpha$ 节点才能使得网络崩溃, 即 $f_c\leq\alpha$. 此时等价于故意攻击, 因此 $f_c=f_c^{IA}$.

若 $\alpha<f_c^{IA}$, 可知 $f_c>\alpha$. 将 (47) 式代入 (28) 式, 可得临界条件

$$(1-f)\sum_{k=m}^{\tilde{m}-1}k(k-1)p(k)=(1-\alpha)\sum_{k=m}^{M}kp(k) \tag{49}$$

求解上式, 可得到临界移除比例 f_c. 特别地, 若 $p(k)=Ck^{-\gamma}(\gamma>2$ 且 $\gamma\neq3)$, 式 (49) 可写为

$$\frac{(M^{2-\gamma}-m^{2-\gamma})(1-\alpha)+(\tilde{m}^{2-\gamma}-m^{2-\gamma})(1-f)}{2-\gamma}=\frac{(\tilde{m}^{3-\gamma}-m^{3-\gamma})(1-f)}{3-\gamma} \tag{50}$$

对于随机失效 ($\alpha=0$), 式 (50) 可写为

$$\frac{(2-f)(m^{2-\gamma}-M^{2-\gamma})}{2-\gamma}=\frac{(1-f)(m^{3-\gamma}-M^{3-\gamma})}{3-\gamma} \tag{51}$$

解得

$$f_c^{RF}=1-\frac{1}{\kappa-1} \tag{52}$$

其中

$$\kappa=\left(\frac{2-\gamma}{3-\gamma}\right)\frac{M^{3-\gamma}-m^{3-\gamma}}{M^{2-\gamma}-m^{2-\gamma}} \tag{53}$$

这与文献 [19] 中的结果一致.

从 (50) 式、(51) 式可看出, 当 $\gamma>3$ 时, 总是存在一个有限的临界移除比例 $f_c<1$; 当 $2<\gamma<3$ 时, 若 $\alpha=0$, 则当 $N\to\infty$, $M^{3-\gamma}\to\infty$, 从而 $f_c\to1$, 但是若 $\alpha>0$, 总是存在一个有限的临界移除比例 $f_c<1$. 这意味着, 在标度指数 $2<\gamma<3$ 的无标度网络中, 当 $N\to\infty$ 时, 只要我们能优先获取很少部分重要节点的信息 ($\alpha>0$), 那么也能通过移除部分节点使得网络崩溃 ($f_c<1$).

将 (43) 式、(47) 式代入 (14) 式、(15) 式可得 $F_0(x)$ 和 $F_1(x)$, 再代入 (32) 式即可求得巨组元规模.

4　不完全信息条件下复杂网络抗毁性仿真分析

在上一节中, 我们解析研究了两种特殊情况 (随机不完全信息和优先不完全信息) 下复杂网络的抗毁性. 本节中, 我们以无标度网络为例, 对一般攻击信息参数组合 (α,δ) 进行详细仿真分析.

给定度序列 $w_1\geq w_2\geq\cdots\geq w_N$, 其中 $w_i=ci^{-1/(\gamma-1)}$, $m=w_N$ 为最小度, $M=c=w_1=mN^{1/(\gamma-1)}$, 为最大度, $\gamma>2$, 采用文献 [25–26] 中的配置模型构造随机无标度网络, 生成网络的度分布为 $p(k)=(\gamma-1)m^{\gamma-1}k^{-\gamma}$. 给定攻击信息参数组合 (α,δ), 在生成的随机无标度网络中按照不等概率抽样步骤确定已知区域 Ω, 然后按照攻击模型移除节点. 选择 $\kappa\equiv<k^2>/<k>\leq2$ 作为网络崩溃的临界值[19], 每移除一个节点后计算网络中的巨组元规模 $|S|$ 以及 κ, 并记录使得 $\kappa\leq2$ 需要移除的节点比例 T. 由于使用

配置模型构造随机无标度网络以及按照不等概率抽样确定已知区域 Ω 均有随机性, 所以我们对于特定网络参数独立执行 10 次配置模型, 对每一个网络独立确定 10 次已知区域 Ω, 最后计算平均值 $<|S|>$ 和 $<T>$ 作为巨组元规模和临界移除比例.

图 2 给出了无标度网络在不同攻击信息参数组合 (α,δ) 条件下巨组元规模 $|S|$ 随节点移除比例 f 变化图, 其中 $N=1000$, $m=2$, $\gamma=3.5$, 实线为解析结果, 与仿真结果非常吻合. 可以看出, 攻击信息对巨组元规模 $|S|$ 有显著影响. 如果攻击信息为零信息 ($\alpha=0$), 即使 50% 的节点被移除, 巨组元中仍然包含 30% 的节点; 如果攻击信息为完全信息 ($\alpha=1$), 20% 的节点被移除, 巨组元规模几乎接近零, 即网络崩溃. 此外, 我们还可以看出, 攻击信息精度比攻击信息广度对巨组元规模 $|S|$ 影响更大. 例如当 $\delta=2$ 时, 获取 30% 的重要节点信息 ($\alpha=0.3$), 基本等价于故意攻击, 即使仅获取 10% 的节点信息 ($\alpha=0.1$), 网络也变得非常脆弱; 但是, 如果 $\delta=0$, 获取 30% 的节点信息 ($\alpha=0.3$), 基本没有影响, 即使获取 80% 的节点信息 ($\alpha=0.8$), 网络抗毁性也非常强.

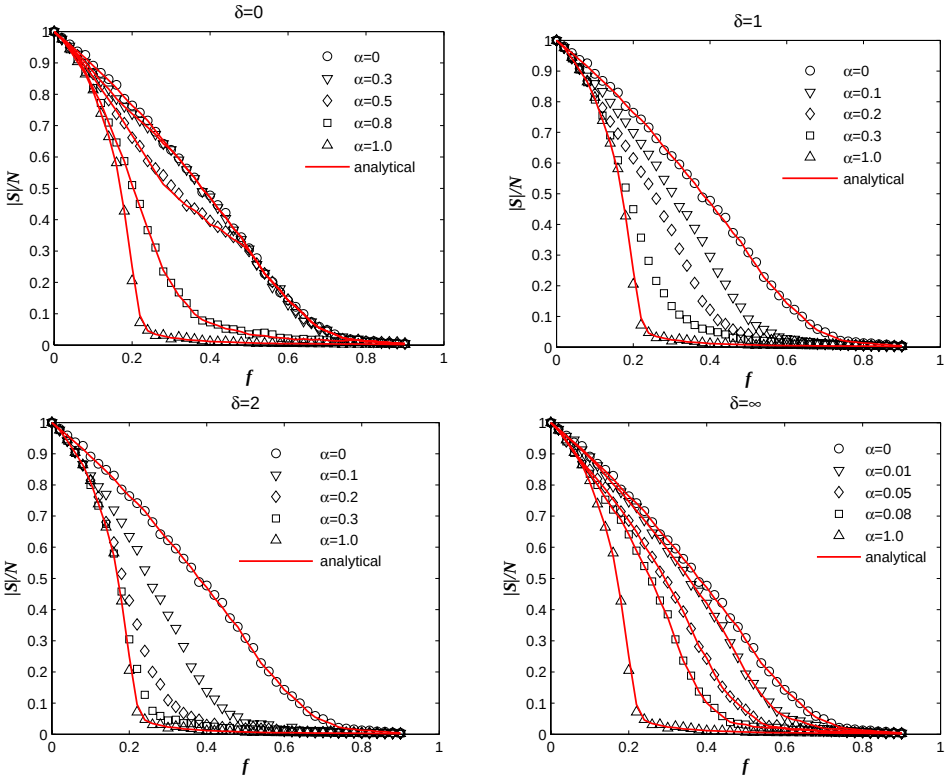

图 2 巨组元规模随节点移除比例变化图

为了直观展现攻击信息广度参数 α 和攻击信息精度参数 δ 对巨组元规模 $|S|$ 的影响, 在图 3 中我们给出了节点移除比例 $f=50\%$ 时, 巨组元规模 $|S|$ 关于 α、δ 的三维关系图以及等高线. 可以看出, 少量高精度信息就等价于大量低精度信息.

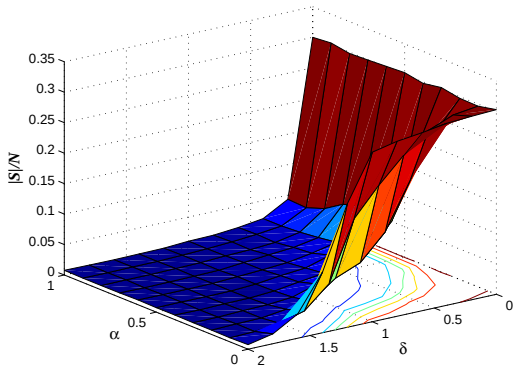

图 3 巨组元规模关于攻击信息参数的三维关系以及等高线图

图 4 给出了无标度网络在不同攻击信息参数组合 (α,δ) 条件下临界移除比例 f_c 随标度指数 γ 变化图, 其中 $N=1000$, $m=2$, 实线解析结果. 可以看出, 当 $\gamma\geq 3$ 时, 仿真结果与解析结果吻合良好, 但当 $\gamma<3$ 时, 仿真结果与解析结果稍有偏差. 这是因为当 $\gamma<3$ 时, 无标度网络的结构最大度 (Structure cut-off) $\sqrt{<k>N}$ 小于其自然最大度 (Natural cut-off) $mN^{1/(\gamma-1)}$, 这导致所生成的网络中自环和多重边的数量不能忽略, 而解析结果是在简单图假设下得到的. 文献 [15, 27] 对上述偏差进行过详细分析. 可以看出, 攻击信息对临界移除比例 f_c 有显著影响. 例如当 $\gamma=2.5$ 时, 如果攻击信息为零信息 $(\alpha=0)$, 则 $f_c=0.892$; 但如果 $(\alpha,\delta)=(0.2,2)$, 则 $f_c=0.430$, 这意味着如果能获取到 20% 比较重要节点的信息, 就可以大幅降低网络的抗毁性 (从 0.892 到 0.430); 如果攻击信息为完全信息 $(\alpha=1)$, 则 $f_c=0.215$; 但如果 $(\alpha,\delta)=(0.8,0)$, 则 $f_c=0.890$, 这意味着如果能随机隐藏 20% 的节点信息, 就可以大幅提高网络的抗毁性 (从 0.215 到 0.890).

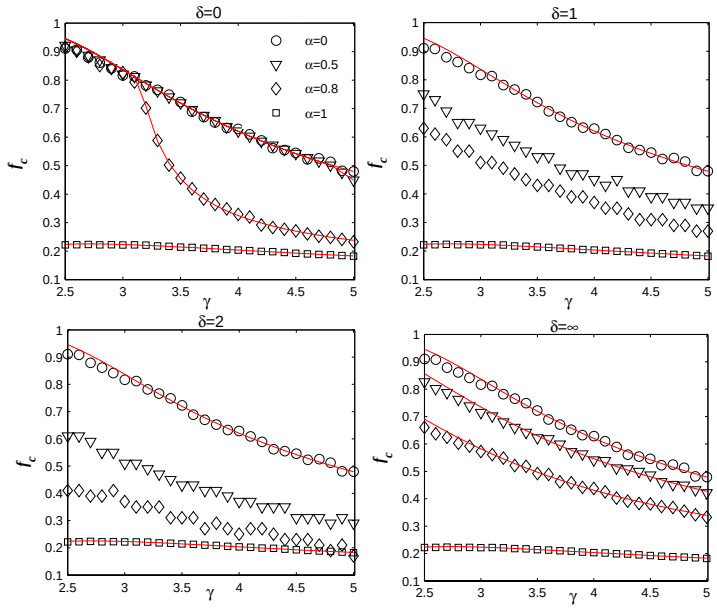

图 4 临界移除比例与标度指数关系图

为了直观展现攻击信息广度参数 α 和攻击信息精度参数 δ 对临界移除比例 f_c 的影响, 在图 5 中给出了标度指数 $\gamma=2.5$ 时, 临界移除比例 f_c 与 α、δ 的三维关系图以及等高线. 从中也可以看出, 少量高精度信息就等价于大量低精度信息.

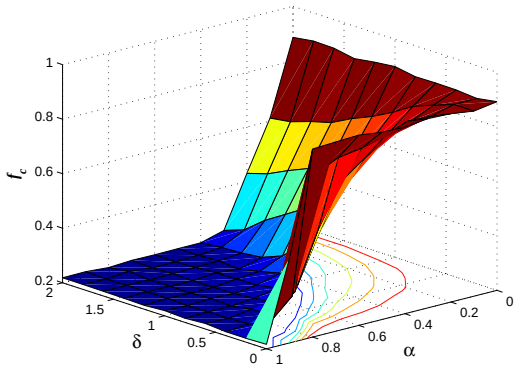

图 5 临界移除比例关于攻击信息参数的三维关系以及等高线图

5 结论与讨论

本文将复杂网络攻击信息获取过程抽象成无放回的不等概率抽样问题, 建立了不完全信息条件下的复杂网络抗毁性模型, 其中网络攻击信息可以用信息广度参数 α 和信息精度参数 δ 调节控制, 以前的随机失效及故意攻击是本文模型的两个特例. 利用概率母函数方法解析推导了随机信息和优先信息条件下具有任意度分布广义随机网络的两个重要抗毁性度量参数: 巨组元规模以及临界移除比例. 研究表明: 对于度分布为 $p(k) = (\gamma-1)m^{\gamma-1}k^{-\gamma}$ 的无标度网络, 当 $\gamma > 3$ 时, 总是存在一个有限的临界移除比例 $f_c < 1$. 但当 $2 < \gamma < 3$ 时, 若 $\delta = 0$, 则只要能隐藏部分节点的信息 ($\alpha < 1$), 那么几乎需要移除所有节点才能使得网络崩溃 ($f_c \to 1$); 若 $\delta = \infty$, 则只要能优先获取很少部分重要节点的信息 ($\alpha > 0$), 那么也能通过移除部分节点使得网络崩溃 ($f_c < 1$). 以无标度网络为例, 对一般攻击信息参数组合 (α, δ) 进行了仿真分析. 研究表明攻击信息对巨组元规模 $|S|$ 和临界移除比例 f_c 都有显著影响: 一方面随机隐藏少量节点信息就可以大幅提高网络的抗毁性, 另一方面获取少量重要节点的信息就可以大幅降低网络的抗毁性. 此外, 研究还发现攻击信息精度比攻击信息广度对 $|S|$ 和 f_c 影响更大, 当攻击信息精度很高时, 只需获取很少节点信息就能使得网络变得很脆弱; 反之, 当攻击信息精度很低时, 即使需获取大量节点信息, 网络抗毁性也很强.

需要指出的是, 本文仅仅考虑了基于节点的抗毁性并且假设节点被攻击后与之相连的边全部移除. 实际上, 很多时候节点很难被完全移除, 只是与其相连的部分边失效. 因此, 基于边的不完全信息条件下复杂网络的抗毁性还有待下一步继续研究. 此外, 目前大部分研究 (包括本文) 都以巨组元规模为网络性能指标, 以网络完全崩溃为临界条件, 以临界移除比例作为抗毁性指标. 但对于很多复杂网络来说, 要使其完全崩溃是非常困难的, 攻击少量的节点很难改变巨组元的规模. 这时, 可以考虑选择其他网络性能指标 (例如网络效率、连通节点对比例) 来代替巨组元规模, 以可调的阈值来代替网络崩溃作为临界条件.

参考文献

[1] 方锦清, 汪小帆, 郑志刚, 等. 一门崭新的交叉科学: 网络科学 (上) [J]. 物理学进展, 2007, 27(3): 239–343.
Fang J Q, Wang X F, Zheng Z G, et al. New interdisciplinary science: networks science (I) [J]. Progress in Physics, 2007, 27(3): 239–343.

[2] 方锦清, 汪小帆, 郑志刚, 等. 一门崭新的交叉科学: 网络科学 (下) [J]. 物理学进展, 2007, 27(4): 361–448.
Fang J Q, Wang X F, Zheng Z G, et al. New interdisciplinary science: networks science (II) [J]. Progress in Physics, 2007, 27(4): 361–448.

[3] 方锦清, 汪小帆, 刘曾荣. 略论复杂性问题和非线性复杂网络系统的研究 [J]. 科技导报, 2004, 22(2): 9–12.
Fang J Q, Wang X F, Liu Z R. On the study of complexity and nonlinear complex networks[J]. Science & Technology Review, 2004, (2): 9–12.

[4] 吴金闪, 狄增如. 从统计物理学看复杂网络研究 [J]. 物理学进展, 2004, 24(1): 18–46.
Wu J S, Di Z R. Complex networks in statistical physics[J]. Progress in Physics, 2004, 24(1): 18–46.

[5] 郑金连, 狄增如. 复杂网络研究与复杂现象 [J]. 系统辩证学报, 2005, 13(4): 8–13.
Zheng J L, Di Z R. A brief discussion on complex networks and complexity[J]. Journal of Systemic Dialectics, 2005, 13(4): 8–13.

[6] 陈禹. 人类对于网络的认识的新发展 [J]. 系统辩证学报, 2005, 13(4): 18–22.
Chen Y. New progress on the network for the human being[J]. Journal of Systemic Dialectics, 13(4): 18–22.

[7] 史定华. 网络 —— 探索复杂性的新途径 [J]. 系统工程学报, 2005, 20(2): 115–119.
Shi D H. Networks — A new approach for exploring complexity[J]. Journal of Systems Engineering, 2005, 20(2): 115–119.

[8] 汪秉宏, 周涛, 何大韧. 统计物理学与复杂系统研究最新发展趋势分析 [J]. 中国基础科学, 2005, 7(3): 37–43.
Wang B H, Zhou T, He D R. The trend of recent research on statistical physics and complex systems[J]. China Basic Science, 2005, 7(3): 37–43.

[9] 谭跃进, 吴俊, 邓宏钟. 复杂网络抗毁性研究综述 [J]. 系统工程, 2006, 24(11): 1–5.
Tan Y J, Wu J, Deng H Z. Invulnerability of complex networks: A survey[J]. Systems Engineering, 2006, 24(11): 1–5.

[10] 吴俊, 谭跃进. 复杂网络抗毁性测度研究 [J]. 系统工程学报, 2005, 20(2): 128–131.
Wu J, Tan Y J. Study on measure of complex network invulnerability[J]. Journal of Systems Engineering, 2005, 20(2): 128–131.

[11] 谭跃进, 吕欣, 吴俊, 等. 复杂网络抗毁性研究若干问题的思考 [J]. 系统工程理论与实践, 2008, 28(增刊): 116–120.
Tan Y J, Lü X, Wu J, et al. On the invulnerability research of complex networks [J]. Systems Engineering — Theory & Practice, 2008, 28(Suppl): 116–120.

[12] Albert R, Jeong H, Barabási A–L. Error and attack tolerance of complex networks[J]. Nature, 2000, 406: 378–382.
[13] Callaway D S, Newman M E J, Strogatz S H, et al. Network robustness and fragility: Percolation on random graphs[J]. Physical Review Letters, 2000, 85(25): 5468–5471.
[14] Gallos L K, Cohen R, Argyrakis P, et al. Stability and topology of scale-free networks under attack and defense srategies[J]. Physical Review Letters, 2005, 94(18): 188701.
[15] Holme P, Kim B J, Yoon C N, et al. Attack vulnerability of complex networks[J]. Physical Review E, 2002, 65(5): 056109.
[16] Cohen R, Erez K, Ben-Avraham D, et al. Breakdown of the internet under intentional attack[J]. Physical Review Letters, 2001, 86(16): 3682–3685.
[17] Paul G, Sreenivasan S, Stanley H E. Resilience of complex networks to random breakdown[J]. Physical Review E, 2005, 72(5): 056130.
[18] Vazquez A, Moreno Y. Resilience to damage of graphs with degree correlations[J]. Physical Review E, 2003, 67(1): 015101.
[19] Cohen R, Erez K, Ben-Avraham D, et al. Resilience of the Internet to random breakdowns[J]. Physical Review Letters, 2000, 85(21): 4626–4628.
[20] Freeman L C. A set of measures of centrality based upon betweenness[J]. Sociometry, 1997, 40(1): 35–41.
[21] Tsen F S P, Sung T Y, Lin M Y, et al. Finding the most vital edges with respect to the number of spanning trees[J]. IEEE Transcations on Reliability, 1994, 43(4): 600–602.
[22] 谭跃进, 吴俊, 邓宏钟. 复杂网络中节点重要度评估的节点收缩方法 [J]. 系统工程理论与实践, 2006, 26(11): 79–83.
Tan Y J, Wu J, Deng H Z. Evaluation method for node importance based on node contraction in complex networks [J]. Systems Engineering — Theory & Practice, 2006, 26(11): 79–83.
[23] 迟艳芹. 统计学原理与应用 [M]. 北京: 清华大学出版社, 2005.
[24] Newman M E J, Strogatz S H, Watts D J. Random graphs with arbitrary degree distributions and their applications[J]. Physical Review E, 2001, 64(2): 026118.
[25] Molloy M, Reed B. A critical point for random graphs with a given degree sequence[J]. Random Structures and Algorithms, 1995, 6(2/3): 161–179.
[26] Molloy M, Reed B. The size of the giant component of a random graph with a given degree sequence[J]. Combinatorics Probability Computation, 1998, 7: 295–305.
[27] Boguna M, Pastor-Satorras R, Vespignani A. Cut-offs and finite size effects in scale-free networks[J]. European Physical Journal B, 2004, 38(2): 205–209.