

文章编号: 1673-5196(2019)03-0101-07

聚类系数指标对复杂网络鲁棒性的影响分析

卢鹏丽¹, 董 璩¹, 曹 乐²

(1. 兰州理工大学 计算机与通信学院, 甘肃 兰州 730050; 2. 天水师范学院 电子信息与电气工程学院, 甘肃 兰州 741000)

摘要: 分析了采用度分布相同且聚类系数不同的三种类型网络(中性网络、同配网络和异配网络)在遇到随机故障或者蓄意攻击时,网络的初始聚类系数变化对网络鲁棒性的影响.实验分析表明,网络的初始聚类系数越大,网络在受到随机故障或蓄意攻击时网络中最大连通子图的直径和网络中最大连通子图的平均路径长度的起伏也就越大.初始聚类系数的变化在异配网中对网络鲁棒性的作用最明显,中性网次之,对同配网的鲁棒性不明显.

关键词: 复杂网络; 鲁棒性; 聚类系数

中图分类号: TP202 **文献标志码:** A

Analysis of influence of clustering coefficient as its index on robustness of complex network

LU Peng-li¹, DONG Men¹, CAO Le²

(1. College of Computer and Communication, Lanzhou Univ. of Tech., Lanzhou 730050, China; 2. School of Electronic Information and Electrical Engineering, Tianshui Normal University, Tianshui 741000, China)

Abstract: The complex networks can be divided into three types according to chaining rules of their nodes, namely neutral network, assortative network, and hetero-assortative network. In this paper, three types of network with identical degree distribution and different clustering coefficient are used to analyze the influence of their initial clustering coefficient on their robustness when they are subjected to random failure and deliberate attack. Experimental analysis shows that the larger the initial clustering coefficient of the network is, the larger the fluctuation of the diameter and average path length of the maximum connected subgraph in network will be when the network is subjected to random failures or deliberate attacks. And the effect of initial clustering coefficient in hetero-assortative network on the network robustness will be most obvious, the effect will be less for neutral network, and there will be no obvious effect for assortative network.

Key words: complex network; robustness; clustering coefficient

大数据在人们生活中扮演的角色越来越重要,复杂网络和复杂系统也得到人们进一步的重视.生活中的各种复杂系统都可以抽象作为复杂网络,复杂网络中节点数目众多,节点与节点之间的关系也千差万别.复杂网络的一项研究领域是网络部分结构失效对网络整体结构和功能的影响^[1],称为鲁棒性分析.

Albert 等^[2]分析了小世界网(WS 模型)和无标度网(BA 模型)在遭到蓄意攻击或随机故障时的网

络鲁棒性,并对万维网的鲁棒性进行了分析.结果显示小世界网在蓄意攻击和随机故障两种情况下的鲁棒性差异不是很大,无标度网和万维网对于随机故障的鲁棒性明显优于对蓄意攻击的鲁棒性,主要原因是两种网络的结构分布差异较大.Paolo 等^[3]建立了一种基于动态方法的模型,在动态模型下对 WS 小世界网和 BA 无标度网进行鲁棒性分析,提出了无标度网的不均匀性.Liu 等^[4]对中国九江炼油系统进行了鲁棒性分析,得出真实系统中具有均匀分布的网络鲁棒性更高.Bansal 等^[5]提出了同配网络、异配网络和中性网络的概念,并分析了三种网络的鲁棒性.Schultz 等^[6]提出变量梯度法对复杂网络进

收稿日期: 2017-09-27

基金项目: 国家自然科学基金(11361033)

作者简介: 卢鹏丽(1973-),女,甘肃酒泉人,博士,教授.

行稳定性的判断.Iyer 等^[7]除了采用介数中心性,还加入了紧密度和特征向量等全局指标,分析合成网络和真实网络遭到随机故障和蓄意攻击时的鲁棒性.

已有的文献多是对于复杂网络最大连通度的分析,本文主要采取最大连通子图的直径和最大连通子图的平均路径长度作为衡量标准^[8],全面分析了具有相同度分布且聚类系数不同的中性网络、异配网络和同配网络的鲁棒性.复杂网络的随机故障和蓄意攻击在文献[9-11]中已经有详细的描述,本文重点分析聚类系数在不同网络攻击中的表现.

1 基本概念

$G(V, E)$ 表示一个无向无权的简单网络,其中 $V = \{v_1, v_2, \dots, v_n\}$ 表示 G 中节点的集合; $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$ 是 G 中边的集合,且 $|V| = n, |E| = m$; A 是其对应的邻接矩阵,如果节点 v_i 和 v_j 之间有边存在,则其元素 $A_{ij} = 1$, 否则 $A_{ij} = 0$.

定义 1(网络的直径 D) 一般定义两节点 (i, j) 间的最短距离 d_{ij} ^[12] 为连接两者的最短路径的边的数目;网络的直径为所有两点间的最大距离,记为 D ^[13], 即:

$$D = \max_{(i,j)} d_{ij} \tag{1}$$

定义 2(平均路径长度 L) 网络的平均路径长度 L 是所有节点对之间距离的平均值, 即:

$$L = \frac{1}{\frac{1}{2} N(N-1)} \sum_{i \geq j} d_{ij} \tag{2}$$

其中: N 为网络节点的总数目;平均路径长度 L ^[13] 描述网络中节点间的离散程度.

定义 3(聚类系数 C) 聚类系数 C 用来描述网络中节点的聚集情况,即网络有多紧密.一般地,假设网络中的一个节点 i 通过 k_i 条边与其他节点相连接, k_i 是节点 i 的邻居节点数目.如果 k_i 个节点之间互相连接,它们之间存在 $k_i(k_i - 1)/2$ 条边,而这 k_i 个节点之间实际存在的边数 E_i 与总的可能存在边数之比就是节点 i 的聚类系数 C_i , 即:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{3}$$

一个网络的聚类系数 C ^[13] 就是网络中所有节点的聚类系数的平均值, 即:

$$C = \frac{1}{N} \sum_{i=1}^n C_i \tag{4}$$

显然有 $0 \leq C \leq 1$, 只有在全连通网络中,聚类系数才能等于 1, 通常情况下一般均小于 1. 在完全随机网络中, $C \sim N^{-1}$, 其中 N 为网络节点的总数目.

定义 4(最大连通度 G_{max}) 最大连通度 G_{max} ^[14] 是指当网络受到攻击或者干扰时,在所剩仍具有连接能力网络中,其中所含节点数目最多的子网络中的节点数占所剩下节点数目的比例, 即:

$$G_{max} = \frac{N_{max}}{N'} \tag{5}$$

其中: N_{max} 是最大连通子图的节点个数; N' 是所有连通子图的节点数总和.

2 复杂网络的鲁棒性分析

2.1 复杂网络的结构

对于一个复杂网络,如果网络中连接度大的节点总是倾向于与连接度大的节点连接,那么这种网络称为同配网络;如果网络中连接度大的节点总是倾向于与连接度小的节点连接,那么这种网络称为异配网络;如果网络中两个节点之间是否有边相连与这两个节点的连接度无关,那么这种网络称为中性网络.图 1~3 形象地描述了中性网络、同配网络和异配网络在受到蓄意攻击和随机故障后网络的连通状况的仿真结果.其中蓄意攻击是指网络中的特定节点(即关键节点)发生故障以后网络的连通情况,而随机故障是指网络中任意节点发生故障以后网络的连通情况.

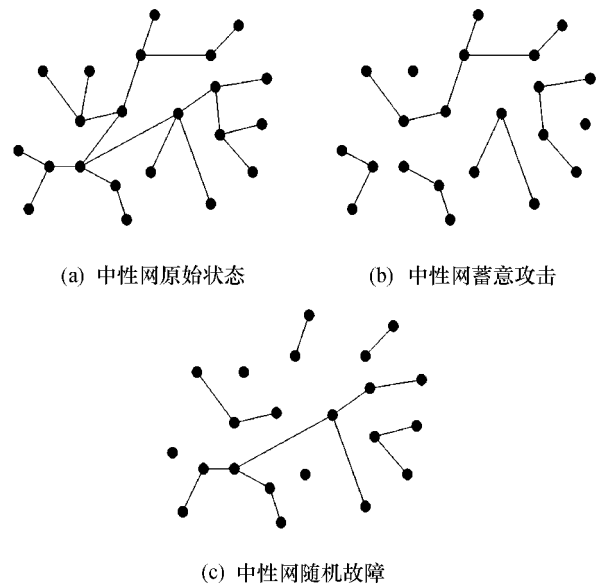


图 1 中性网在受到蓄意攻击或随机故障前后的连通状态
Fig.1 Connective state of neutral network before and after intentional attack or random fault

通过仿真结果的对照可知,蓄意攻击对网络连通度的影响明显大于随机故障对网络连通度的影响.中性网络节点之间的连接并无明确的规律,故对抗蓄意攻击和随机故障时表现出很大的不明确性.

同配网中关键节点总是相互连接在一起,故同配网络在蓄意攻击时显得异常脆弱.而异配网在蓄意攻击时显示出很强的健壮性.

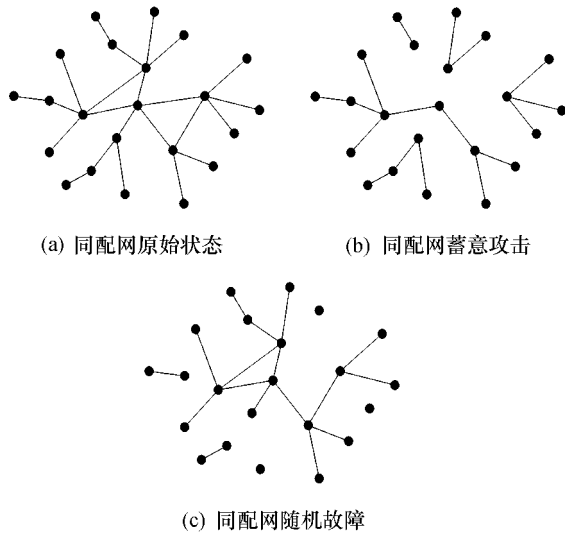


图 2 同配网在受到蓄意攻击或随机故障前后的连通状态

Fig.2 Connective state of assortative network before and after intentional attack or random fault

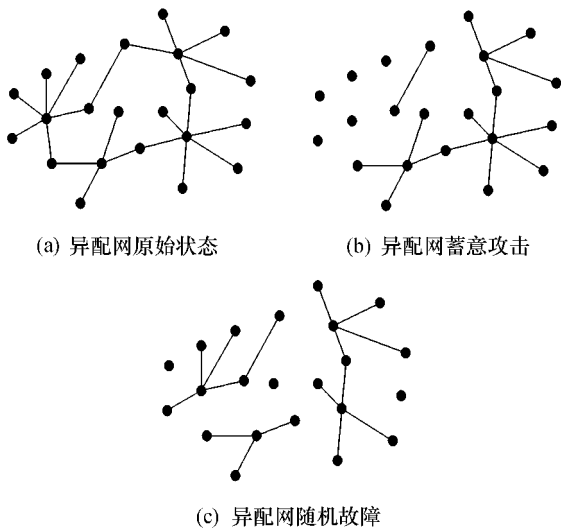


图 3 异配网在受到蓄意攻击或随机故障前后的连通状态

Fig.3 Connective state of hetero-assortative network before and after intentional attack or random fault

对多数实际网络进行研究显示,互联网以及蛋白质交换网络等生物网络是异配网络,而人际关系网以及电影演员合作网络等许多现实网络是同配网络,包括复杂网络中著名的无标度网络也属于同配网络.而不同的在线社会网络可能是同配、异配或者中性网络.例如包含 7 亿多节点的 Facebook 网络呈

现出同配性特征,大型在线社交网络 Cyworld 却是异配网络^[5].

2.2 复杂网络的鲁棒性

对于现实中的复杂系统,总是希望复杂系统拥有一定的鲁棒性,也就是复杂系统对外界的各种干扰具备一定的抗干扰能力.在实际生活中,系统面临各种各样的主观或者客观的干扰是不可避免的,鲁棒性和脆弱性分别是稳定指标与失效指标的角度来表征网络的特性,两者相辅相成.鲁棒性越大,其脆弱性就越小,即抗毁能力越强;鲁棒性越小,其脆弱性越大,即抗毁能力越弱.

先前的各种鲁棒性分析中都围绕着网络的最大连通度 G_{max} 进行.网络的鲁棒性通常与网络的最大连通子图有关,所以网络中最大连通子图的直径和平均路径长度是网络鲁棒性分析的指标.

2.3 算法介绍

复杂网络由于节点众多且结构复杂,网络在构造时很难出现构造的两个网络结构一样的情况,往往构造出来的网络结构之间较大的差异.为了更准确地分析聚类系数指标对复杂网络鲁棒性的影响,本文选取待分析网络时让待分析网络具有相同的节点度分布,使得构造出的网络之间结构差异较小.网络都由 1 000 个节点度已知情况下的节点,根据同配网、异配网和中性网的连接规律,将节点连接成所对应的同配网、异配网和中性网.为确保生成网络聚类系数的一般性,根据以上网络生成规则,生成 100 组同配网、异配网和中性网,并对它们的初始聚类系数进行了统计.统计发现节点数为 1 000 的中性网络,初始聚类系数主要分布在 0.001 到 0.003 之间,同配网的初始聚类系数主要分布在 0.008 到 0.012 之间,异配网的初始聚类系数主要分布在 0.001 5到 0.002 5 之间,故实验中采用了具有特殊初始聚类系数的网络作为待分析网络.

本文主要采取网络的最大连通子图的直径和最大连通子图的路径长度作为衡量标准,全面地分析了具有相同度分布且聚类系数不同的中性网络、异配网络和同配网络的鲁棒性,攻击方式分为随机故障和蓄意攻击,算法如下:

1) 随机生成具有 1 000 个节点的网络,计算网络的度分布.

2) 根据随机网络的度分布,对网络进行重连,生成多组相应的同配网络、中性网络和异配网络.

3) 计算多组同配网络、异配网络和中性网络的聚类系数,并在同一种网络中取出三组聚类系数不同的网络,作为待分析网络.

4) 分别按照随机故障和蓄意攻击两种方式确定需要在待分析网络中删除的节点,随机故障时随机选取节点进行删除,蓄意攻击时选取节点度较大的节点优先进行删除.同时将待删除节点以及待删除节点所连接的边删除.

5) 判断当前网络的最大连通子图,计算 f 、 D 和 L .其中, f 为受到攻击时节点数与原网络节点数的比值, D 为网络受到攻击后最大子图的直径, L 为网络受到攻击后最大子图的平均最短路径.

6) 计算待分析网络中的节点数.若节点数为 0,则进行下一步,否则返回步骤 4.

7) 算法结束.

3 实验与分析

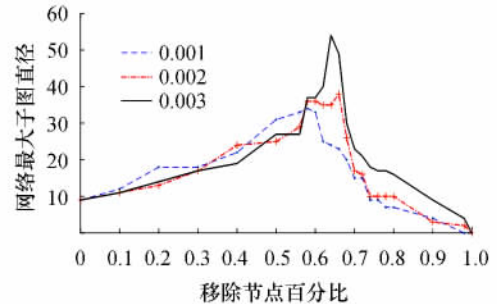
实验结果分别如图 4~6 所示,其中 L 为当前最大连通子图的平均路径长度, D 为当前最大连通子图的直径.图 4 分别为初始聚类系数为 0.001、0.002 和 0.003 的中性网络在随机故障和蓄意攻击时网络中 L 和 D 的变化情况.图 5 分别为初始聚类系数为 0.008、0.010 和 0.012 的同配网络在受到随机故障和蓄意攻击时网络中 L 和 D 的变化情况.图 6 分别为初始聚类系数为 0.001 5、0.002 0 和 0.002 5 的异配网络在受到随机故障和蓄意攻击时网络中 L 和 D 的变化情况.

3.1 中性网络

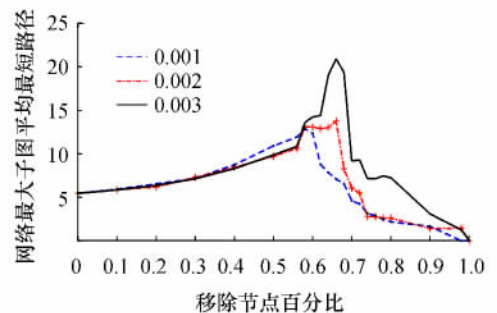
图 4a 比较了三种中性网络在受到随机故障时网络中 D 的变化情况.在丢失少量节点时,网络中 D 变化不明显.初始聚类系数越大的中性网络, D 的起伏越大.图 4b 比较了三种中性网络在受到随机故障时网络中 L 的变化情况.图 4c 比较了三种中性网络在受到蓄意攻击时网络中 D 的变化情况,且初始聚类系数越大的网络, D 越先出现起伏现象.图 4d 比较了三种中性网络在受到蓄意攻击时网络中 L 的变化情况,相比于网络随机故障,初始聚类系数越小的网络在被攻击时所产生的 L 的最大值要高于初始聚类系数大的网络.

根据以上分析得出,当移除节点数目较少时,无论是随机故障还是蓄意攻击,网络的 D 和 L 都呈现出一个缓慢增值的趋势,但在移除节点数目到达一定数量时, D 和 L 的变化明显,蓄意攻击在移除节点 40%左右出现浮动,随机故障在移除节点 60%左右出现浮动,蓄意攻击对网络的破坏明显大于随机故障.中性网络在少量节点丢失时,网络最大子图的直径与平均最短路径都缓慢地增长.但在大量节点丢失时,网络的 D 和 L 都发生剧烈的变化.这是由

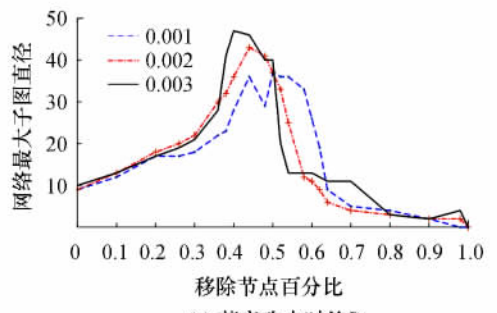
于中性网络的不确定性造成的,中性网络中节点之间的连边没有什么明确的关系,在受到大量节点丢失时,节点之间的离散程度明显变化很大,且聚类系数越大的中性网络,在丢失节点后网络中的 D 和 L 越大,节点间的离散程度越高,网络的鲁棒性越差.



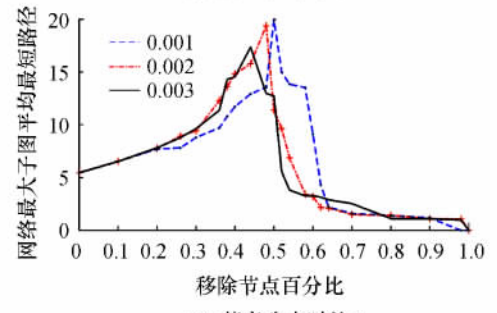
(a) 随机故障时的 D



(b) 随机故障时的 L



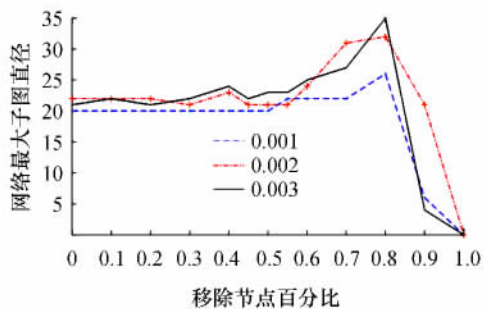
(c) 蓄意攻击时的 D



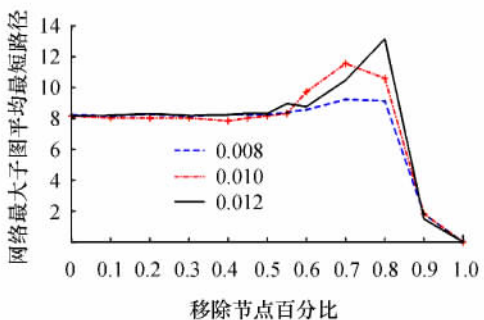
(d) 蓄意攻击时的 L

图 4 三种初始聚类系数不同的中性网络在受到攻击时网络中最大连通子图中的最大直径和平均最短路径变化情况

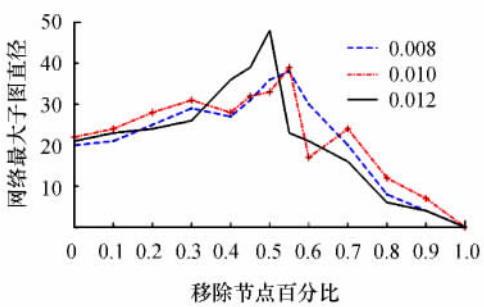
Fig. 4 The maximum diameter and average shortest path change in the most Dalian subgraph of the neutral network with three initial clustering coefficients in the network under attack



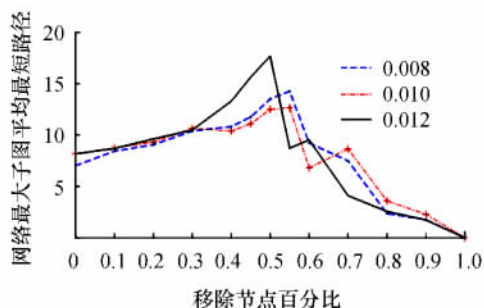
(a) 随机故障时的 D



(b) 随机故障时的 L



(c) 蓄意攻击时的 D



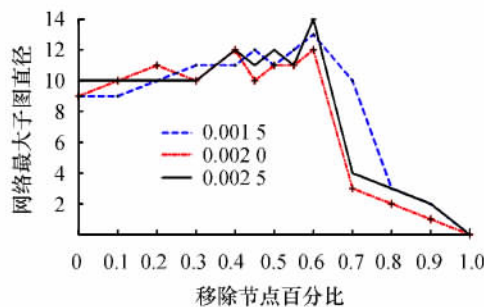
(d) 蓄意攻击时的 L

图 5 三种初始聚类系数不同的同配网络在受到攻击时网络中最大连通子图中的最大直径和平均最短路径变化情况

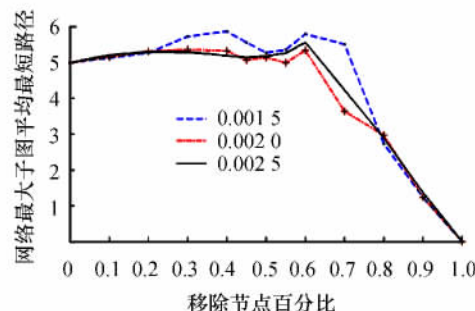
Fig.5 the maximum diameter and average shortest path change of the most Dalian pass subgraph in the assortative network under three different initial clustering coefficients when attacked

3.2 同配网络

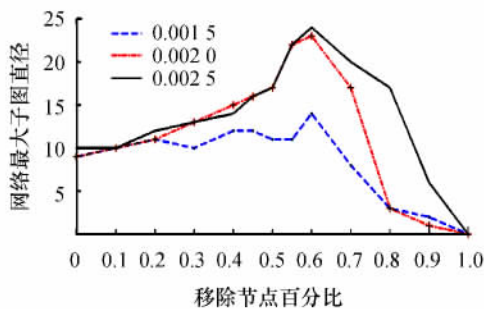
同配网络的 D 和 L 在随机故障中体现出很高的稳定性.在丢失大量节点之前, D 一直保持着



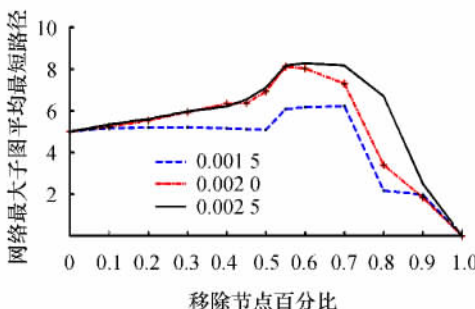
(a) 随机故障时的 D



(b) 随机故障时的 L



(c) 蓄意攻击时的 D



(d) 蓄意攻击时的 L

图 6 三种初始聚类系数不同的异配网络在受到攻击时网络中最大连通子图中的最大直径和平均最短路径变化情况

Fig.6 The maximum diameter and average shortest path change of the most Dalian pass subgraph in the disassortative network under three different initial clustering coefficients when attacked

稳定的状态.随着丢失节点数目的增加(在丢失节点数 50%时)网络中最大连通子图的直径才有微弱的变化,且初始聚类系数越大的同配网络,直径的起伏

越大.在丢失节点数 75%左右时,网络被分割为大量小碎片,直径剧烈下降.这是由于同配网络中连接度大的节点优先与连接度大的节点连接,这种网络结构类似于无标度网,所以同配性网络在随机故障时,网络的 D 和 L 都呈现出一种稳定的态势.且初始聚类系数越大,网络在随机故障时出现的直径最大值越高,网络的离散程度越高.同配网络在遭到蓄意攻击时,网络的 D 和 L 变化剧烈,且在丢失节点占比 35%左右时就出现明显的变化,所产生的 D 和 L 的最大值都远高于同配网络在受到随机故障时所产生的 D 和 L .在同配网中初始聚类系数越大,网络中 D 和 L 的浮动越大,节点间的离散程度越高,网络的鲁棒性越差.

3.3 异配网络

图 6a 比较了三种异配网络在受到随机故障时网络中 D 的变化情况.图 6b 比较了三种异配网络在受到随机故障时网络中 L 的变化情况.图 6c 比较了三种异配网络在受到蓄意攻击时 D 的变化情况.图 6d 比较了三种异配网络在受到蓄意攻击时,网络中 L 的变化情况.初始聚类系数越大,异配网络在遭到蓄意攻击时网络的直径起伏越大.

异配网络的 D 和 L 在随机故障中体现的相对稳定.在丢失大量节点前,网络的 D 和 L 一直保持着一种相对稳定的状态,但有小的起伏.随着丢失节点数目的增加(在丢失节点的数 60%左右时)网络中的最大连通子图的直径开始剧烈下降,且下降的程度与初始的聚类系数无关.异配网络在少量的节点遭到蓄意攻击时,网络的最大连通子图的直径缓慢增长,在移除节点占比 50%左右时,直径明显地增长,且初始聚类系数越大的网络直径的变化越大.网络的最大连通子图的平均最短路径呈现出一种缓慢增加的趋势,在移除节点占比 60%左右时出现极值,平均路径长度开始缓慢下降.在异配网络的蓄意攻击中可以明显地看出聚类系数对网络中 D 和 L 的影响较大,聚类系数越大的异配网 D 和 L 的变化越明显,节点间的离散程度越高,网络的鲁棒性越差.

3.4 综合分析

中性网络和异配网络的初始网络直径都在 9 左右,初始平均路径长度都在 5 左右,且不会随着初始聚类系数的变化而发生变换.同配网的初始网络直径和初始平均路径长度相对较大,初始直径在 22 左右,初始平均路径长度在 8.10 左右,也不会随着初始聚类系数的变化而改变.

由表 1 分析可知,在度分布相同的情况下,同配

网的初始聚类系数明显大于中性网和异配网,中性网和异配网的初始聚类系数相似.为了进一步分析聚类系数指标对不同系统的鲁棒性的影响,取初始聚类系数相同的中性网和异配网进行比较.图 7 为

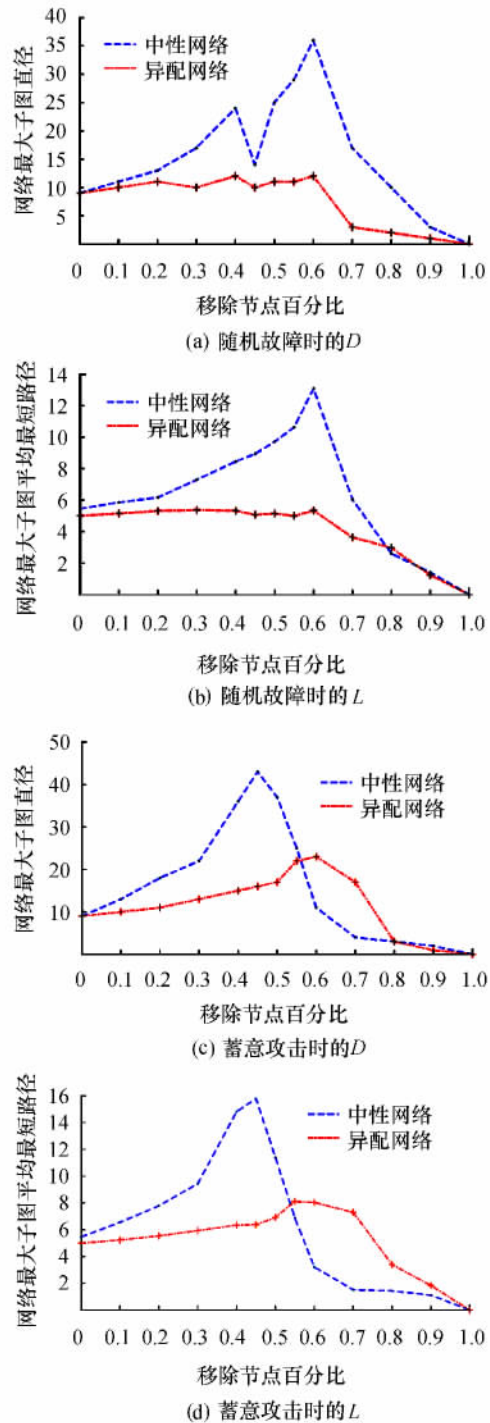


图 7 相同聚类系数的中性网和异配网在受到攻击时网络中最大连通子图中的最大直径和平均最短路径变化情况

Fig.7 The maximum diameter and the average shortest path change in the most Dalian pass graph in the network when the neutral network and the disassortative network with the same clustering coefficients are attacked

表1 三种不同网络在不同聚类系数下网络的初始直径和初始聚类系数

Tab.1 Initial diameter and initial clustering coefficient of three different networks under different clustering coefficients

网络类型	初始聚类系数	初始直径	初始路径长度
中性网	0.001	10	5.20
	0.002	9	5.45
	0.003	10	5.46
同配网	0.008	20	8.10
	0.010	22	8.16
	0.012	24	8.14
异配网	0.001 5	9	4.93
	0.002 0	10	4.89
	0.002 5	9	4.87

初始聚类系数都为 0.002 的中性网和异配网在受到随机故障和蓄意攻击时网络中 L 和 D 的变化情况。

图 7a 比较了中性网络和异配网络在受到随机故障时网络中 D 的变化情况。图 7b 比较了中性网络和异配网络在受到随机故障时网络中 L 的变化情况。图 7c 比较了中性网络和异配网络在受到蓄意攻击时 D 的变化情况。图 7d 比较了中性网络和异配网络在受到蓄意攻击时,网络中 L 的变化情况。图 7 中网络的节点数目、节点度分布和初始聚类系数都相同,只因为节点之间连线的方式存在差异就使得网络的鲁棒性有着巨大的差异。

由以上分析可知,聚类系数越大,网络中 D 和 L 的起伏越大,网络的鲁棒性越差。对于异配网,初始聚类系数对网络受到随机故障和蓄意攻击时的影响最大,中性网络次之,同配网络虽然有很大初始聚类系数,却在随机故障和蓄意攻击时对其最大连通子图的直径和平均路径长度都没有很大的影响。

4 结语

通过分析聚类系数对复杂网络中最大连通子图的直径和平均路径长度的影响可知,在度分布相同的情况下,聚类系数越大,网络的鲁棒性越差。且聚类系数在不同的网络中所体现出作用的大小也不同,在异配网中聚类系数对网络的鲁棒性的作用明

显,中性网次之,同配网中受到聚类系数的影响最小。除聚类系数对网络中最大连通子图的直径和最大连通子图的平均路径长度的影响外,聚类系数对网络中其他相关方面的影响将来需要做进一步的研究和验证。

参考文献:

- [1] PATEL S J, PATTEWAR T M. Software birthmark based theft detection of JavaScript programs using agglomerative clustering and frequent subgraph minming [C]//Embedded System (ICES), 2014 International Conference on. [S. l.]: IEEE, 2014.
- [2] ALBERT R, JEONG H, SI A L. Error and attack tolerance of complex networks [J]. Nature, 2000, 406(4): 378-382.
- [3] CRUCITTI P, LATORA V, MARCHIORI M, et al. Error and attack tolerance of complex networks [J]. Physica A, 2004, 340: 388-397.
- [4] LIU Suyu, RONG Gang. Analysis on refinery system as a complex task-resource network [J]. Chinese Journal of Chemical Engineering, 2013, 21(3): 253-262.
- [5] 毛 凯. 复杂网络结构的稳定性与鲁棒性研究 [J]. 计算机科学, 2015, 42: 85-88.
- [6] MAO J, GAO J, LIU Y. Power allocation over finding cognitive MIMO channels: an ergodic capacity perspective [J]. Trans Veh Technol, 2016, 61: 1162-1173.
- [7] 陆靖桥, 傅秀芬, 蒙在桥. 复杂网络的鲁棒性与中心性指标的研究 [J]. 计算机应用与软件, 2016, 33: 302-310.
- [8] MYLES G, COLLBERG C. Software watermarking via opaque predicates: Implementation analysis and attacks [J]. Electronic Commerce Research, 2006, 6(2): 155-171.
- [9] MYLES G, COLLBERG C, HEIDEPPEM Z, et al. The evaluation of two software watermarking algorithms [J]. Software: Practice and Experience, 2015, 35(10): 923-938.
- [10] MELIÁN C J, BASCOMPTE J. Complex networks: two ways to be robust [J]. Ecology Letters, 2010, 5(6): 705-708.
- [11] 刘飞飞, 蔺靖娜, 刘潇潇. 基于贝叶斯复杂网络的复杂网络攻击方法研究 [J]. 计算机工程与应用, 2017(53): 18-25.
- [12] CHRISTIAN S. Shortest-path queries in static networks [J]. ACM Computing Surveys, 2014, 46(4): 1-31.
- [13] 刘宏鲲, 周 涛. 中国城市航空网络的实证研究与分析 [J]. 物理学报, 2007, 56(5): 106-112.
- [14] 刘一奎, 刘天琪, 李 茜, 等. 一种高聚类系数的无标度网络演化模型 [J]. 网络安全技术与应用, 2015(1): 55-56.