

# 基于复杂网络的中小企板股市网络特性分析

马兴福,王 红,李园园

(山东师范大学 信息科学与工程学院,山东 济南 250014)

**摘 要:**基于复杂网络的理论及研究方法,以中小企板股票为节点、股票价格波动相关系数作为连边构建一个无向加权的股市网络。对该股市网络进行研究分析发现,中小板股市网络具有典型的小世界效应和无标度特性,这说明了在该股市网络中,单只股票的价格波动会很容易地影响到其它股票,新上市的股票更容易与具有很多连接的股票关联。并通过分析节点的影响系数和节点的介数,发现该股市网络存在对网络有重要影响的节点,这些节点的移除将会影响到证券市场的稳定性,因此有必要加强对这些股票的管理。

**关键词:**复杂网络;股市网络;小世界;无标度;中心性

**中图分类号:**TP393

**文献标识码:**A

**文章编号:**1673-629X(2012)04-0172-03

## Characteristic Analysis of Small and Medium Enterprises Board Stock Market Network Based on Complex Network

MA Xing-fu, WANG Hong, LI Yuan-yuan

(School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China)

**Abstract:** Based on the complex network theory and research methods, an undirected and unweighted stock market network is built taking the small and medium enterprises board stock as the node, the correlation coefficient of the stock price fluctuation as the boundaries. After studying and analyzing this network, it is found that the small stock market has the typical characteristics of small-world effect and the scale-free property. It indicates that in this network the fluctuation of a single stock price can easily affect other stocks and the new stocks are more easily correlated with those ones who have many connections. There existed some nodes which have a great impact on the network by analyzing the influence coefficient and the betweenness of nodes. The removal of these nodes will affect the stability of the stock market, so it is essential to enhance the management of these stocks.

**Key words:** complex network; stock market network; small-world; scale-free; centrality

## 0 引 言

复杂网络是复杂系统的抽象,现实中许多复杂系统都可以用复杂网络的相关特性进行描述和分析。其中网络中的节点表示复杂系统中的个体,而边则表示个体之间的关系<sup>[1]</sup>,如果网络中的两个节点之间存在着某种特定的关系,则它们之间存在着一条边将这两个节点相连接,否则两节点之间无边相连。比如,社会关系网络、科研合作网络、演员合作网络、计算机共享网络、新陈代谢网络、万维网、城市交通网络、电力网等等都可以作为复杂网络来研究。近几年来,对复杂网络的研究一直是许多领域的研究热点,也取得了许多

重要进展,其中最重要的特征是复杂网络的小世界效应和无标度特性<sup>[2]</sup>。

近年来,用复杂网络的理论来研究和解释证券市场中的各种现象受到了众多学者的广泛重视。文中利用复杂网络的理论及研究方法对中国股市网络进行了分析,发现该网络具有许多复杂网络的典型特性,比如小世界效应和无标度特性等。

## 1 股市网络的构建

文中以中小企板股票为节点、以股价波动相关系数作为连边权重构造一个全连接的加权网络。

记股票 $i$ 在第 $t$ 期的收盘价格为 $Y(t)$ 。股票 $i$ 从第 $(t-1)$ 期到第 $t$ 期的对数收益率定义为: $R_i(t) = \ln Y_i(t) - \ln Y_i(t-1)$ 。

用相关系数作为股票之间的相关性度量指标<sup>[3]</sup>,文中采用 Pearson 相关系数,因此股票 $i$ 和 $j$ 的相关系数 $C_{ij}$ 可表述如下:

收稿日期:2011-08-29;修回日期:2011-12-02

基金项目:国家自然科学基金(60970004);山东省研究生教育创新计划项目(SDYY10059);山东师范大学研究生重点课程项目

作者简介:马兴福(1988-),男,山东临沂人,硕士研究生,研究方向为复杂网络;王 红,教授,硕士生导师,研究方向为复杂网络、移动计算。

$$C(ij) = \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{(\langle R_i^2 \rangle - \langle R_i \rangle^2)(\langle R_j^2 \rangle - \langle R_j \rangle^2)}}$$

这里连边的权重  $W_{ij} = C(ij)$  [4]。其中  $\langle R_i \rangle$  为股票  $i$  在  $n$  期内的平均收益率,即

$$\langle R_i \rangle = \frac{1}{n} \sum_{t=1}^n R_i(t)$$

$\langle R_i R_j \rangle$  为股票  $i$  和  $j$  在同一时间的收益率乘积的平均值。

其中相关系数  $C(ij)$  取值范围为  $[-1, 1]$ 。

文中数据选取从 2010 年 5 月到 2011 年 5 月持续交易的中小企

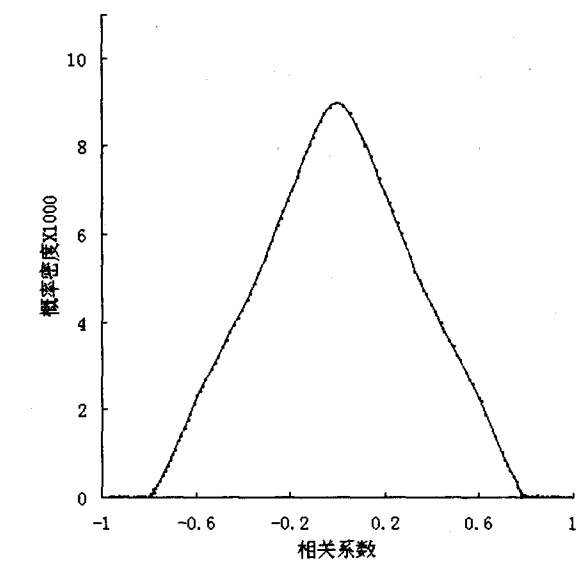


图 1 相关系数概率分布

2 小世界特性

复杂网络中两个主要的统计特性为:平均路径长度和聚类系数。如果一个网络具有较小的平均路径长度和较大的聚类系数,则该网络具有典型的小世界特性[5]。

在复杂网络中节点  $i$  的度  $K(i)$  表示与节点  $i$  有边相连的节点的数目。节点  $i$  和  $j$  直接的距离  $d(ij)$  定义为连接这两个节点的最短路径上的边数。网络的平均路径长度  $L$  定义为任意两个节点之间的距离的平均值[6],表示为:

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

其中  $N$  为网络节点数。

聚类系数用来表示网络中节点的聚集程度[7]。假设网络中的一个节点  $i$  有  $k_i$  条边将它和其他节点相连,这  $k_i$  个节点就称为节点  $i$  的邻居。显然,在这  $k_i$  个节点之间最多可能有  $k_i(k_i - 1) / 2$  条边。而这  $k_i$  个节

点之间实际存在的边数  $E_i$  和总的可能的边数之比就为节点  $i$  的聚类系数  $C_i$ ,即

$$C_i = 2E_i / (k_i(k_i - 1))$$

整个网络的聚类系数  $C$  就是所有节点  $i$  的聚类系数  $C_i$  的平均值[8]。这里  $0 \leq C \leq 1$ ,当  $C = 0$  时表明网络中的所有节点全部为孤立节点,即网络中没有任何的连边,当  $C = 1$  时表明网络中任意两个节点都直接相连。对于一个含有  $N$  个节点的完全随机网络,当  $N$  很大时,  $C = O(N^{-1})$ ,对于许多大规模的实际网络,它们的聚类系数虽远小于 1,但比  $O(N^{-1})$  要大得多,这些网络都具有明显的聚类效应。

这里引入阈值  $\theta$ ,当股票  $i$  和  $j$  的价格波动相关系数  $|C_{ij}|$  大于或等于  $\theta$ ,则节点  $i$  和  $j$  有边相连,否则两节点之间无连边。利用 Matlab 语言编程求出各个参数,其中网络平均路径长度用 Floyd 算法求出[9]。表 1 为在不同阈值下的平均路径长度和聚类系数。

表 1 股市网络中相关统计特性

$\theta$	$L$	$C$
0.30	1.1435	0.9017
0.35	1.2463	0.8853
0.40	1.3674	0.8764
0.45	1.5319	0.7612
0.50	1.6158	0.6831

表 1 说明了我国中小企

3 无标度特性

在复杂网络中,另一个重要的统计特性为节点的度以及度分布。节点  $i$  的度  $k(i)$  定义为与该节点连接的其他节点的数目。节点的度越大,在某种程度上说明该节点越“重要”,反之,如果节点的度很小,则说明该节点在网络中所起到的作用较小,不会对网络产生很大的影响。在有向网络中,节点的度又分为入度和出度,节点的入度是指其他节点指向该节点的边的数目,而节点的出度则指从该节点指向其他节点的边的数目。网络中所有节点的度的平均值为网络的平均度,记为  $\langle k \rangle$ 。用分布函数  $P(k)$  来表示网络中的节点的度分布情况,其中  $P(k)$  表示的是一个随机选定的节点的度恰好为  $k$  的概率。研究表明,大量的真实网络的节点度分布是服从幂律分布的,即

$$P(k) \propto k^{-\gamma}$$

其中  $r$  为幂指数。节点度分布服从幂律分布,这说明绝大部分的节点的度相对较低,但存在少量的度很大的节点,这些度很大的节点称为网络的 Hub 节点,这些 Hub 节点在网络中起着重要作用。把节点度分布服从幂律分布的网络称为无标度网络,具有无标度特性<sup>[11]</sup>。

通过计算在不同阈值  $\theta$  下中小企业板股市网络节点的度分布情况,发现该股市网络在  $\theta = 0.6$  时节点度服从幂律分布,具有无标度特性,如图 2 所示。通过最小二乘法求出  $\theta = 0.6$  时的幂律指数  $r = 0.9217$ 。该网络的无标度特性说明了在中小板股市网络中具有极少度很大的超级节点,该节点对应的股票在该股市网络中具有重要的地位,因此有必要加强对这些股票的监管。同时该网络在动态增长的过程中也具有优先连接的特性,即遵从“富者更富”的法则或“马太效应”<sup>[11]</sup>,即新上市的股票更容易与具有很多连接的超级节点股票关联。

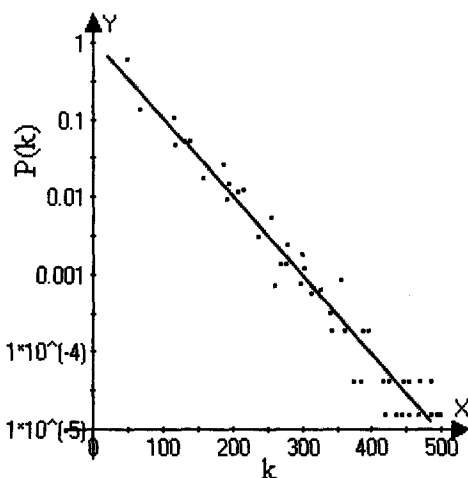


图 2 股票关联网络节点度分布

## 4 节点中心性

### 4.1 节点影响系数

在证券交易市场中,单只股票价格的波动可能会对其它股票价格的波动产生一定的影响,因此引入一个影响系数  $T_i$  来度量一个节点对其它节点的影响能力。 $T_i$  定义为节点  $i$  的所有连边的权重之和,可表示如下:

$$T_i = \sum_{j \in N} w_{ij}$$

通过计算各股票在股市网络中的影响系数,并对影响系数  $T_i$  的分布进行一元回归分析后得出  $T_i$  的分布是服从幂律分布的。通过最小二乘法计算得到该股市网络的影响系数分布的幂指数为 1.1300。这表明在该股市网络中存在少数影响力很大的节点,这些股票价格的波动很可能就会对它近邻股票的价格波动产

生很强的影响。

### 4.2 节点介数

节点的影响系数只是反映了该节点对其近邻节点的影响强度,是网络的局部性质。为了考察节点的全局性质以及节点在整个网络中的影响力,引入了节点介数<sup>[12]</sup>。

在复杂网络中,节点介数反映了相应的节点在整个网络中的作用力和影响力,是一个重要的全局几何量,在中小企业板股票关联网络中,对节点介数进行重新定义,节点  $i$  的介数  $b(i)$  定义为网络中所有最短路径中经过该节点的路径的数目。可以通过计算删除该节点以及与该节点的连边后,网络中最短路径长度发生变化的最短路径数来得到该节点的介数。通过计算网络节点介数发现该网络节点介数也是服从幂律分布的,如图 3 所示。通过最小二乘法求出幂律指数  $r = 0.1233$ 。这说明在中小企业板股票网络中存在节点介数很大的节点,这些节点对整个股市的稳定性具有很大的影响,称这些节点为该网络的中心节点。移除或删除这些节点将会影响到整个网络,影响到证券市场的稳定性。

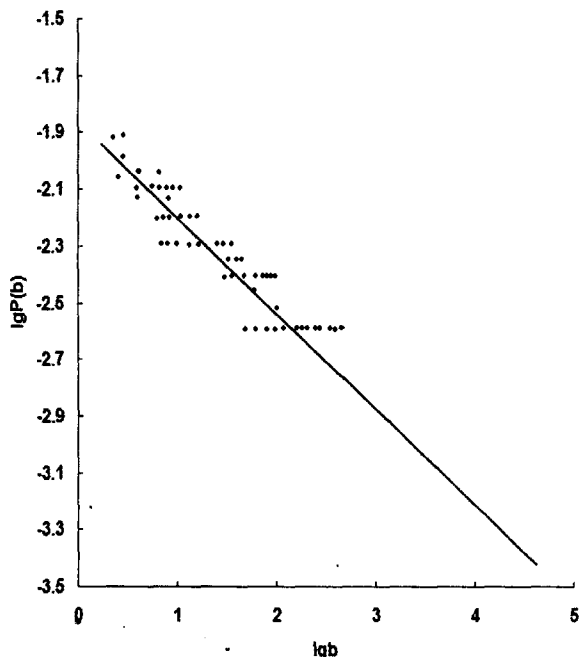


图 3 股票关联网络节点介数分布

## 5 结束语

利用复杂网络的理论及研究方法对中国中小企业板股票网络进行研究,发现在该股票关联网络中,单只股票的价格波动会对其他股票的价格波动产生影响,尤其是那些和它在相同集团的股票。通过分析该股票关联网络的度分布,发现中小企业板股票网络同样具

(下转第 178 页)

表 2 实验结果

Method performance	BPN		LDA		PSO-SVM		T(s)
	Training (%)	Testing (%)	Training (%)	Testing (%)	Training (%)	Testing (%)	
约简前	85	100	100	60	100	100	0.21
约简后	70	80	85	70	100	100	0.10

3) 在相同的参数下,使用没有经过关联分析处理的数据对 SVM 多类分类器进行训练和诊断时间为 0.21s,是经过处理后的两倍多,当数据样本和故障类型增加时,效果会更显著。经过 PSO 参数寻优,准确性达到 100%,明显优于 BP 神经网络和线性判别分析,可以很好地应用于设备故障诊断。

#### 4 结束语

在汽轮机组故障诊断中,采集到的数据中存在一定的冗余属性和重复样本,利用关联分析,计算属性间相关系数,可以对其进行约简处理以达到降低数据维数的目的,使得样本更具有代表性,实现简单,效率更高。当数据样本和故障类型增加时,效果会更显著。此外,应用粒子群方法选择参数能避免支持向量机的过学习,一定意义上寻得最优参数,提高了分类准确率。文中将进一步研究,并把关联分析直接应用于电厂实时数据库,以达到预期效果。

#### 参考文献:

[1] 梁娜. 基于数据挖掘的火电厂故障诊断研究[D]. 保定:

(上接第 174 页)

有无标度特性,即具有“马太效应”现象,新上市的股票更容易与那些度很大的节点相联。同时,在该股票关联网络中存在着对整个网络具有重要作用的节点,这些节点所对应的股票价格的大幅度波动会对整个股市网络产生较大的影响,这些节点的移除也会影响到股市网络的稳定性。因此在证券交易市场,应该加强对这些股票的监管,确保证券交易市场的稳定性。

#### 参考文献:

- [1] Newman M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(3): 167-256.
- [2] 孙博文,孙百瑜,刘天立,等. 中国股市的拓扑结构及其复杂性质研究[J]. 黑龙江大学自然科学学报, 2006, 23(3): 366-369.
- [3] 熊胜君,杨朝军. 沪深股票市场行业效应与投资风格效应的实证研究[J]. 系统工程理论与实践, 2006(4): 44-49.
- [4] Kim H J, Lee Y, Kahng B, et al. Weighted scale-free network in financial correlations[J]. J Phys Soc Jpn, 2002, 71(9):

华北电力大学, 2007.

- [2] Chen Kaiying, Chen Longsheng. Using SVM based method for equipment fault detection in a thermal power plant[J]. Computers in Industry, 2011, 62(1): 42-50.
- [3] 郭小芸,马小平. 基于粗糙集的故障诊断特征提取[J]. 计算机工程与应用, 2007, 43(1): 221-224.
- [4] Han J, Kamber M. 数据挖掘概念与技术[M]. 范明, 孟小峰, 译. 北京: 机械工业出版社, 2001.
- [5] 任江海, 黄焕宇, 孙婧昊, 等. 基于相关性分析及遗传算法的高维数据特征选择[J]. 计算机应用, 2006, 26(6): 1403-1405.
- [6] 连可, 陈世杰, 周建明, 等. 基于遗传算法的 SVM 多分类决策树优化算法研究[J]. 控制与决策, 2009, 24(1): 7-12.
- [7] Wu Chih-Hung, Ken Yun, Huang Tao. Patent classification system using a new hybrid genetic algorithm support vector machine[J]. Applied Soft Computing, 2010, 10(4): 1164-1177.
- [8] 徐庆伶, 江西莉. 一种基于支持向量机的半监督分类方法[J]. 计算机技术与发展, 2010, 20(10): 115-117.
- [9] 向昌盛, 周子英. 支持向量分类机的参数选择方法研究[J]. 计算机技术与发展, 2010, 20(9): 95-97.
- [10] 姚全珠, 蔡婕. 基于 PSO 的 LS-SVM 特征选择与参数优化算法[J]. 计算机工程与应用, 2010, 46(1): 134-136.
- [11] 汪江. 汽轮机组振动故障诊断 SVM 方法与远程监测技术研究[D]. 南京: 东南大学, 2005.
- [12] 江伟, 罗毅, 涂光瑜. 基于多类支持向量机的变压器故障诊断模型[J]. 水电能源科学, 2007, 25(1): 52-55.

2133-2136.

- [5] 汪小帆, 李翔, 陈关荣. 复杂网络理论及应用[M]. 北京: 清华大学出版社, 2006.
- [6] 李耀华, 姚洪兴. 金融危机下股市网络的结构特性研究[J]. 成都信息工程学院学报, 2010, 25(1): 107-111.
- [7] Watts D J, Strogatz S H. Collective dynamics of small world networks[J]. Nature, 1998, 393(6): 440-442.
- [8] 黄玮强, 庄新田. 中国股票关联网络拓扑性质与聚类结构分析[J]. 管理科学, 2008, 21(3): 94-102.
- [9] Kim H J, Kim I M. Scale-free network in stock market[J]. J Kor Phys Soc, 2002, 40(6): 1105-1108.
- [10] 兰旺森, 张所地. 基于复杂网络的中国股市房地产版块股票强相关性研究[J]. 数学的实践与认识, 2009, 39(4): 62-66.
- [11] 鲁巍巍, 林正春. 基于复杂网络理论的沪深 A 股分析[J]. 科学技术与工程, 2009, 9(11): 2859-2862.
- [12] Albert R, Barabasi A L. Statistical mechanics of complex network[J]. Rev Mod Phys, 2002, 74(1): 47-97.