

我国股票市场拓扑性及加权网络中行业主导性分析

金 秀, 姜 超, 孟婷婷, 庄霄威
(东北大学 工商管理学院, 辽宁 沈阳 110819)

摘 要: 采用聚类算法得到最小生成树, 研究不同行业股票的拓扑性结构. 在加权复杂网络模型的基础上, 建立了考虑网络时变特点的动态加权网络模型. 利用上证 180 指数作为研究样本, 研究中国股票市场的行业主导性. 实证结果表明: 我国证券市场存在明显的聚集性, 相同行业股票具有相似的特点; 金融保险和制造业行业相对更加活跃, 其他行业股票价格更容易受到这两种行业股票价格的影响; 证券市场中行业的主导性会随着时间的变化而改变, 新兴行业逐渐占据行业的主导地位.

关 键 词: 行业主导性; 聚类系数; 最小生成树; 拓扑性; 加权网络

中图分类号: F 830.91

文献标志码: A

文章编号: 1005-3026(2015)10-1516-05

Topology Property and Industry Dominance in Weighted Network for China's Stock Market

JIN Xiu, JIANG Chao, MENG Ting-ting, ZHUANG Xiao-wei

(School of Business Administration, Northeastern University, Shenyang 110819, China. Corresponding author: JIANG Chao, E-mail: 45786272@qq.com)

Abstract: The minimum spanning tree was constructed by using the clustering algorithm, which was used to study the topological structure of various industries. Based on the weighted network model, a dynamic weighted network model considering the characteristic of time variation was developed. Taking the SSE 180 index as sample, the market dominance of the Chinese stock market was studied. The empirical results showed that the Chinese stock market has notable aggregation, and the same industry stocks have similar characteristics. The finance/insurance and the manufacturing industry are relatively more active, and the prices of others are more susceptible to these two industries. The leading industry in the stock market will change with the time, and the emerging industry gradually dominates the overall industry.

Key words: industry dominance; clustering coefficient; minimum spanning tree; topological property; weighted network

近年来,人们运用复杂网络理论来研究金融市场,其中研究股票关联网络的基本拓扑性质和发掘股票市场的内在等级结构最为普遍. Gower 等^[1]利用复杂网络理论,提出了通过度量距离来构造最小生成树聚类分析方法. Mantegna 等^[2]通过相应的金融数据来刻画金融市场的整体结构和层次结构,首次将复杂网络理论引入到金融领域研究中. 黄飞雪^[3]将最小生成树引入房地产价格研究中,对金融危机发生前后城市的房屋销售价格指数进行比较分析. 这些研究没有考

虑市场的时序变化,属于静态网络的研究范畴.

股票市场是一个不断变化的复杂网络,静态网络研究已经不能满足投资者的需求. Reichl^[4]证明在金融系统中,对称性和平衡性的假设是难以成立的. 从风险管理角度来说,动态变化的复杂网络能更好地反应复杂系统的本质. Boginski 等^[5]研究了美国股票市场图及图的属性,证明了金融网络会随着时间的变化而发生改变. Tabak 等^[6]利用基于时变方法的复杂网络,研究巴西股票市场中不同部门的相互联系.

收稿日期: 2014-11-28

基金项目: 国家自然科学基金资助项目(71271047).

作者简介: 金 秀(1963-),女,辽宁沈阳人,东北大学教授,博士生导师.

通过复杂网络特性研究股票和行业间的关联性已取得一定成果,然而已有的研究大多仅限于单纯的统计性分析,只反映了股票市场的整体趋向及变化,投资者不能从中获得足够的有效信息.

本文以上证 180 指数为研究样本,构建加权股票网络结构.通过最小生成树分析股票市场的拓扑性,对股票进行分类;采用 β 系数、亲密度、聚类系数三种指标来观察随时间的变化不同股票和行业在市场中主导性的变化趋势.研究不同行业、不同股票主导性强弱,能够为投资组合管理提供有效的决策依据.

1 最小生成树及加权网络模型

1.1 最小生成树

1) 最小生成树(MST).规则网络是在一定规则下建立起来的,网络中任意两个节点的联系遵循既定的规则.MST 是一种规则网络图,边连接途中所有顶点且没有环,使边距离之和最短.最小生成树是一种无预先管理的聚类方式,能够通过移除最弱的连接来去除关联度较低的点,可以清楚刻画股票之间相关性的几何信息.

2) MST 的构建方法.金融市场上,许多股票被同时交易.为了评测两只股票同时演变的相似性和差异,引入了相关系数的概念,通过考察两只股票(记为股票 i 和股票 j) 日价格对数变化的相关系数来量化股票 i 和股票 j 之间价格变化的同步性程度.定义股票 i 在 t 时刻的收益率 $S_i(t)$ 为

$$S_i(t) = \ln(Y_i(t)) - \ln(Y_i(t-1)). \quad (1)$$

式中, $Y_i(t)$ 是股票 i 在时间 t 的日收盘价格.股票 i 和股票 j 在观察区间内价格的波动系数 ρ_{ij} 为

$$\rho_{ij} = \frac{\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle}{\sqrt{(\langle S_i^2 \rangle - \langle S_i \rangle^2) - (\langle S_j^2 \rangle - \langle S_j \rangle^2)}}. \quad (2)$$

式中, $\langle \cdot \rangle$ 表示对考察期内所有交易日的数据取平均值.由式(2)可知, $\rho_{ij} \in [-1, 1]$.

根据文献[7],股票的距离公式为

$$d_{ij} = \sqrt{2(1-\rho_{ij})}. \quad (3)$$

通过式(3), d_{ij} 可以构建一个股票距离矩阵 D , 且 $D_{ij} = d_{i,j} \in [0, 2]$. 根据 Kruskal 算法[8], 将距离矩阵 D 的非对角线上的元素按增序列排列, 把股票两两之间的距离用图形表示出来, 逐步将集合中所有的元素连接起来得到最小生成树.

1.2 加权网络的度量

加权网络是在无权网络的基础上,对每个节

点或边赋予其权重.加权网络中每条连边都带有一个度量连接强度的数值,为复杂网络中节点之间的关系和相互作用提供了更加细致的刻画手段.权重及其分布会对网络的性质和功能产生重要影响,有助于更深入地理解复杂系统的性质.

度量加权网络中节点的支配性和带动性的主要指标为节点亲密度、聚类系数和 β 系数.因此,本文选取这三种指标通过上证 180 的股票数据研究市场中各个行业的主导性质.

1) 亲密度.亲密度由 Sabidussi^[9] 提出,是衡量一个节点在网络中与其他节点亲密程度的重要指标.以 k 为起点 h 为终点的路径所包含的边的数量可表示为 $d_G(k, h)$, 则节点 k 的亲密度 $C(k)$ 为该节点到达所有其他节点的距离之和的倒数:

$$C(k) = \frac{1}{\sum_{h \in G} d_G(k, h)}. \quad (4)$$

由式(4)可知,亲密度越大表明节点的接近度越大,节点越居于网络的中心,它在网络中就越重要.位于节点中心的股票与其他股票的亲密程度越大,表明其影响力越大.

2) β 测度.复杂网络中的主导性可以定义为一个节点或者一个节点群在这个网络中占有一定比重,对其他节点或整个网络有较强的带动作用.在股票市场中,一支股票或者一个行业的股票在整个市场中所占有的权重越大,其所带来的影响就会越大.基于这个定义,考虑到方向和权重.这种基于主导关系的衡量方式称为 β 测度^[10], 表示为

$$\beta(i) = \sum_{j=1}^n \frac{\omega_{ij}}{\lambda_j}. \quad (5)$$

式中: ω_{ij} 表示两节点间边的权重; λ_j 为 j 在 ω 中的占优权重.由式(3)可知, d_{ij} 相当于两只股票向量之间的欧几里德距离,每支股票代表顶点, ω_{ij} 可以表示为距离的倒数,即 $\omega_{ij} = \frac{1}{d_{ij}} =$

$$\frac{1}{\sqrt{2(1-\rho_{ij})}}.$$

3) 聚类系数.在复杂网络中,聚类系数是对拓扑性质进行度量的重要物理量,是表示一个图形中节点聚集程度的系数.在金融股票市场中,股票的聚类系数越大,表明其他股票向这支股票的倾向性越高,即这支股票的主导性越大. Saramaki 等^[11] 定义的加权聚类系数模型为

$$c_i^\omega = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(\omega_{ij} + \omega_{ih})}{2} \alpha_{ij} \alpha_{ih} \alpha_{jh}. \quad (6)$$

式中, s_i 为顶点 i 的点权,根据文献[9],在加权网

络中,可定义 $s_i = \sum_{j \in \gamma(i)} \omega_{ij}$.

2 实证分析

2.1 数据选取

本文选取 2007 年 1 月至 2012 年 12 月上证 180 股票收盘指数构建网络矩阵,通过对亲密度、聚类系数、 β 测度的分析,比较上市 180 中 13 个行业的主导性. 13 个行业分别为金融保险、制造业、交通运输、采掘业、建筑业、信息技术、批发零售、电力煤气、房地产、传播文化、农林牧、综合类、

社会服务. 为了更好地研究股票的动态演化过程,将 2007 ~ 2012 年的数据通过递推的方式进行计算,观察 13 种行业在不同期间的中心性、亲密度、聚类系数的变化状况,从而分析其主导性的变化过程. 数据来源于 WIND 数据库.

2.2 实证结果及数据分析

1) 最小生成树分析. 剔除上证 180 中数据缺失半年以上的股票,根据式 (3) 将剩余的 124 只股票通过 matlab7.0 计算 d_{ij} ,使用 Gephi 软件,根据 Kruskal 算法对 d_{ij} 进行分析,得到 MST,如图 1 所示.

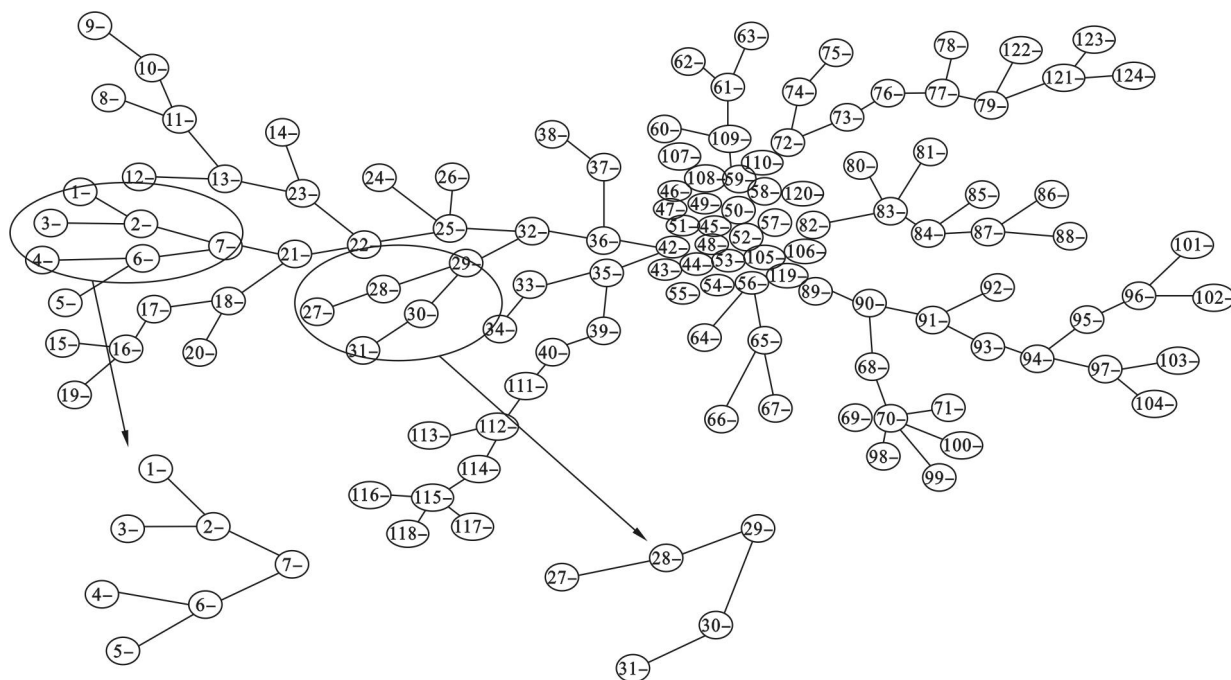


图 1 最小生成树

Fig. 1 Minimum spanning tree

由图 1 知,在 MST 中第 1 ~ 7 只股票形成一组,10 ~ 14 为一组,27 ~ 31 单独为一组,可得股票 1 ~ 7 都属于交通运输行业,股票 10 ~ 14 属于采掘业,股票 27 ~ 31 属于建筑业. 这表明隶属于同一种行业的股票会聚类在一起,与文献[8]提出的相同行业股票之间具有很强的聚类效应的结果相符.

由此可见,通过对上证 180 股票价格的序列分析,采用聚类算法得到的最小生成树,可以很好地对股票进行组合分类,投资者通过观察图形,在投资组合的选择上可以有一个直观的参考数据. 在构建股市最小生成树的基础上,进一步通过加权网络模型的三个度量指标,将股票市场中的股票按行业分类,对其在市场中的主导性进行分析. 分析行业的主导性可以更加清晰地了解股票在当前市场中的表现,因此本文通过亲密度、聚类系数

和 β 测度三个指标对上证 180 中 13 种行业的主导性进行实证. 在此基础上,基于网络的动态演化过程又对 2007 ~ 2012 年这些行业主导性的变化趋势进行分析.

2) 行业主导性时变分析. 为了更清晰地了解各个行业的主导性变化过程,本文以 30 d 为一个观察窗口,对 2007 ~ 2012 年期间市场中 13 个行业的亲密度、聚类系数和 β 进行递推式计算,如图 2 所示. 由于图的大小和篇幅受限,在 124 支股票构成的图像中选取 2 个聚类组进行详细说明,其他聚类组的性质与这 2 组相同.

由图 2a 可知,2007 ~ 2008 年期间各个行业的聚类系数指标差别不大,相对平稳的波动表明股票市场在 2007 ~ 2008 年期间较为平稳. 2008 年中旬房地产行业的指标开始上升,同时制造业的指数在所有行业中达到最低,这表明我国曾经

的支柱性产业在受到国家政策和金融危机的双重打击时,投资者的投资方向迅速转移到了更为热

门的房地产行业中. 直到 2011 年,制造业的聚类系数指标才逐渐上升,恢复到了领先地位.

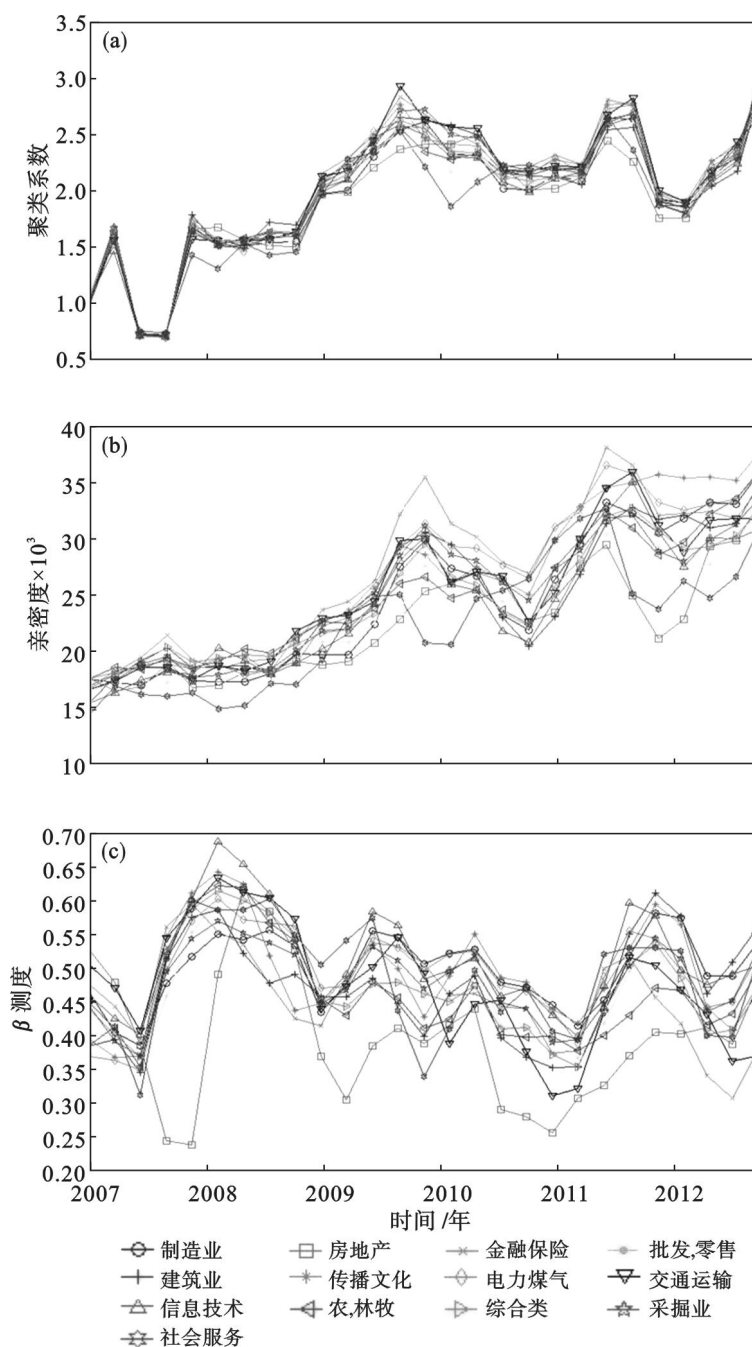


图 2 时间对聚类系数,亲密度及 β 测度的影响

Fig. 2 Effect of time on clustering coefficients, closeness centrality and β

由图 2b 可知,2007 ~ 2012 年间,金融保险行业几乎一直领先于其他各行业. 只有在 2008 年金融危机时期,各银行信誉度受到重挫,其亲密度指标落后于更加平稳的传统支柱产业的制造业. 新兴的信息技术业的亲密度一直在不断地上升,表明各行业中企业的快速发展与信息技术的联系将会越来越紧密.

由图 2c 可知,上下波动最大且频繁的就是房

地产行业. 2007 ~ 2008 年期间,房地产行业是受到国家政策影响最大的行业,其主导性指标在 2007 年中期达到最低值,尔后开始迅速反弹. 这表明投资者在房地产政策发布初期,都处于谨慎的观望状态,当房地产行业适应新政策后,投资者才将资金重新投放回房地产行业. 其他行业的主导性指标都趋于平稳状态,出现较小的波动.

3 结 论

1) 上证 180 指数的股票存在明显的聚集性, 具有相同特性的股票会聚类在一组. 投资者应该保持持有股票的多样性, 避免同时持有同一聚类中的多支股票, 防止一损俱损问题的出现.

2) 金融保险和制造业行业相对更加活跃, 说明我国的主导性行业依旧十分依赖传统的支柱产业. 主导性越高表明其影响力越大, 其他行业的股票价格更容易受到这两种行业股票价格的影响.

3) 证券市场中行业的主导性会随着时间的变化而改变, 传统支柱产业依旧主导着所有行业的动向. 但新兴行业的崛起, 尤其是信息科技类的迅速发展将逐渐改变这一现状. 不同时期股票市场会反映出不同的信息, 主导性强的股票所反馈的动态信息会对投资者的决策起着重要作用.

参考文献:

- [1] Gower J C, Ross G J S. Minimum spanning trees and single linkage cluster analysis[J]. *Applied Statistics*, 1969, 18(1): 54 - 64.
- [2] Mantegna R N. Hierarchical structure in financial markets [J]. *The European Physical Journal B*, 1999, 11(1): 193 -

197.

- [3] 黄飞雪. 金融危机对中国 70 城市房价影响的关联聚集效应[J]. *管理评论*, 2011, 23(6): 3 - 8.
(Huang Fei-xue. Correlative cluster effect of financial crisis's impact on the housing price index of 70 Chinese cities[J]. *Management Review*, 2011, 23(6): 3 - 8.)
- [4] Reichl L E. A modern course in statistical physics[M]. New York: Wiley, 1998.
- [5] Boginski V, Butenko S, Pardalos P M. Statistical analysis of financial networks [J]. *Computational Statistics & Data Analysis*, 2005, 48(2): 431 - 443.
- [6] Tabak B M, Serra T R, Cajueiro D O. Topological properties of stock market networks; the case of Brazil[J]. *Physica A*, 2010, 389(16): 3240 - 3249.
- [7] Bonanno G, Caldarelli G, Lillo F, et al. Topology of correlation-based minimal spanning trees in real and model markets[J]. *Physical Review E*, 2003, 68(1): 1 - 4.
- [8] Huang W Q, Zhuang X T. A network analysis of the Chinese stock market[J]. *Physica A*, 2009, 388(14): 2956 - 2964.
- [9] Sabidussi G. The centrality index of a graph [J]. *Psychometrika*, 1996, 31(4): 581 - 603.
- [10] Van den Brink R, Gilles R P. Measuring domination in directed networks[J]. *Social Networks*, 2000, 22(2): 141 - 157.
- [11] Saramaki J, Kivela M, Onnela J, et al. Generalizations of the clustering coefficient to weighted complex networks [J]. *Physical Review E*, 2007, 75: 027105 - 4.

(上接第 1510 页)

参考文献:

- [1] Qu P, Mason S J. Using tabu search on the single machine multi-orders per job scheduling problem [C]//IIE Annual Conference and Exhibition. Houston: Institute of Industrial Engineers, 2004: 1831 - 1835.
- [2] Sobeyko O, Monch L. Genetic algorithms to solve a single machine multiple orders per job scheduling problem [C]//2010 Winter Simulation Conference. Piscataway: IEEE, 2010: 2493 - 2503.
- [3] Jampani J, Mason S J. Column generation heuristics for multiple machine, multiple orders per job scheduling problems[J]. *Annals of Operations Research*, 2008, 159(1): 261 - 273.
- [4] Mason S J, Chen J S. Scheduling multiple orders per job in a single machine to minimize total completion time [J]. *European Journal of Operational Research*, 2010, 207(1): 70 - 77.
- [5] Erramilli V, Mason S J. Multiple orders per job compatible batch scheduling [J]. *IEEE Transactions on Electronics Packaging Manufacturing*, 2006, 29(4): 285 - 296.

- [6] Erramilli V, Mason S J. Multiple orders per job batch scheduling with incompatible jobs[J]. *Annals of Operations Research*, 2008, 159(1): 245 - 260.
- [7] Chen J S. Optimization models for the flow-shop scheduling problem with multiple orders per job [C]//The 40th International Conference on Computers and Industrial Engineering. Piscataway: IEEE, 2010: 1 - 6.
- [8] Zhou B H, Zhai Z Q. Lifetime distribution model of port facilities with pitting corrosion of stochastic processes [J]. *Applied Mechanics and Materials*, 2011, 44: 46 - 50.
- [9] Shen L X, Wu Y B. Single machine past-sequence-dependent delivery times scheduling with general position-dependent and time-dependent learning effects [J]. *Applied Mathematical Modeling*, 2013, 37(7): 5444 - 5451.
- [10] Qu P, Mason S J. Metaheuristic scheduling of 300 mm jobs containing multiple orders [J]. *IEEE Transactions on Semiconductor Manufacturing*, 2005, 18(4): 633 - 643.
- [11] Pfund M, Fowler J W, Gadkari A, et al. Scheduling jobs on parallel machines with setup times and ready times [J]. *Computers & Industrial Engineering*, 2008, 54(4): 764 - 782.