

基于随机矩阵的金融网络模型*

韩华 吴翎燕[†] 宋宁宁

(武汉理工大学理学院, 武汉 430070)

(2013年10月17日收到; 2014年3月27日收到修改稿)

随机矩阵理论运用于金融领域中研究金融相关系数矩阵的相关性, 相关系数矩阵是网络构建中的关键因素, 本文将随机矩阵理论与网络构建相结合, 研究基于随机矩阵的金融网络模型. 本文选取上海证券市场的股票数据, 将其中的股票数据分成四个阶段, 基于随机矩阵理论, 讨论金融相关系数矩阵和随机矩阵的特征值统计性质, 并在此基础上对现有的去噪方法进行改进, 建立更适合构建金融网络的相关系数矩阵, 并构建金融网络模型. 然后, 基于随机矩阵理论和网络的关键节点分析比较去噪前后的金融网络以及噪声网络, 发现对网络去噪后仍保留了原始网络的关键重要的信息, 而噪声信息对应的是原始网络中度比较小的节点所代表的信息. 最后, 基于去噪网络, 分析金融网络的拓扑结构, 如最小生成树、模体和社团结构, 发现改进后的金融网络的拓扑性质更加明显, 结构更加紧密.

关键词: 随机矩阵理论, 金融相关系数矩阵, 金融网络

PACS: 89.75.-k, 89.75.Fb

DOI: 10.7498/aps.63.138901

1 引言

金融市场条件随时间不断变化, 会造成两个股票的相关关系不是固定不变的, 而有限的时间序列估计相关关系又会受到一些噪声信息的干扰. 由于在实证的相关系数矩阵中会包含着很多不确定性和大量的噪声信息, 这对研究网络的拓扑结构造成一定的影响. 许多学者也在研究金融市场的动态变化规律. Li 等^[1] 引入一个多维的 Subdiffusion 模型, 描述了金融市场中每只股票的停滞期以及多只股票之间的依赖关系的行为, 说明了股票之间关系不是一成不变的, 并分析了金融市场的相关特点. Zhang 等^[2] 研究了金融市场的系统风险问题, 分析了金融风险系统的动态变化行为, 同时通过一系列的控制达到控制系统混乱的目的. 李蓉^[3] 利用随机矩阵理论的方法去除肺癌基因表达数据中的噪声并得出结论, 随机矩阵理论-等级聚类方法是一种有效的识别基因网络的新方法. 随机矩

阵理论也与相关系数矩阵的联系越来越紧密. 徐赞新等^[4] 借助随机矩阵理论的方法, 通过比较移动人群数据与随机数据在互相关矩阵谱分布上的差异, 指出了本征向量成员权重的时空模式与城市移动人群整体行为特征的波动过程非常符合. 与近年来, 随着经济物理学在金融时间序列分析中的应用, 随机矩阵理论越来越多的应用于金融领域中, 为分析相关噪声提供了有效的方法. 随机矩阵理论起源于核物理, 目前运用于量子场论、多元统计、无线通信以及金融领域中. 在金融领域中, 随机矩阵理论的应用也越来越多. Laloux 等^[5] 研究了如何去掉金融相关系数矩阵的噪声; Plerou 等^[6] 详细研究了相关系数矩阵的特征值、特征向量的分布、相关系数矩阵的改进措施及应用; Sharifi 等^[7] 和 Conlon 等^[8] 用随机矩阵理论作了资产组合的研究; Namaki 等^[9] 运用随机矩阵理论和相关系数分布行分析比较了在当地扰动和全球扰动下的德黑兰证券交易所和道琼斯工业平均指数. 近年来学者们也纷纷将随机矩阵理论运用于复杂网络中.

* 国家自然科学基金(批准号:71140015)和中央高校基本科研业务费(批准号:2013-Ia-007, 2013-Ia-040)资助的课题.

[†] 通讯作者. E-mail: 815007651@qq.com; hanhua@whut.edu.cn

Rojkova 等^[10]运用随机矩阵理论研究了路易斯维尔大学的互联骨干路由器和虚拟局域网的交通行为,分析真实和虚拟的交通率随着不同的交通时间序列而变化,将随机矩阵预测结果的特征值和逆参与比率应用于网络的交通拥挤机理中. Feher 等^[11]通过选取相关阈值构建基因相关网络,运用随机矩阵理论对网络社区进行检测,与其他层次聚类法比较发现此种方法划出的社团结构反映了其他方法的所有特性. Namaki 等^[12]运用随机矩阵理论分析了作为股票网络的市场模式的相关系数矩阵的最大特征值,为了更好地分析风险管理,消除了市场模式,并构建金融网络,分析了不同阈值下的金融网络的拓扑结构,发现网络具有无标度特性. Kumar 等^[13]将随机矩阵理论和网络的方法应用于调查 20 个财务指标的相关性和网络的性质,将 2008 年金融危机发生之前和发生时的结果进行比较分析,分析集群网络在不同阈值下的拓扑性质.

目前许多学者运用随机矩阵理论来分析相关系数矩阵,运用随机矩阵理论分析网络的相关性质,研究的主要是金融市场中的投资组合优化和风险管理等方面的问题. 而基于随机矩阵理论和相关系数矩阵,处理金融网络中的噪声问题并分析网络的拓扑性质,这方面的研究很少. 由于金融系统是高度复杂的动力学系统,金融网络中存在大量的噪声,这对研究网络的拓扑性质,分析网络的动力学行为不利,所以基于随机矩阵理论,研究去除网络中的噪声信息的方法仍需我们继续探讨.

2 基本理论

2.1 金融相关系数矩阵

记股票 i ($i = 1, 2, \dots, N$) 的有效交易日价格序列: $\{P_i(1), P_i(2), \dots, P_i(N)\}$, 其中, N_i 表示股票 i 的有效交易天数, $P_i(t)$ 则表示股票 i 在第 t 个有效交易日的收盘价格, 定义股票 i 的对数收益率 (Logarithmic Return) 如下:

$$r_i(t) = \ln P_i(t) - \ln P_i(t-1). \quad (1)$$

对于任意两支股票 i 和 j , 由于它们的有效交易价格序列不同, 因此, 在计算相关性时, 取出它们的共同有效交易日时间序列: $\{u_1, u_2, \dots, u_m\}$, 其中, m 表示股票 i 和股票 j 的共同有效交易日总天数.

这样, 任意两支股票 i 和 j 之间的皮尔逊相关系数 (Pearson correlation coefficient) 为

$$\rho_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sqrt{(\langle (r_i)^2 \rangle - \langle r_i \rangle^2)(\langle (r_j)^2 \rangle - \langle r_j \rangle^2)}}, \quad (2)$$

其中, $\langle \dots \rangle$ 为统计平均, ρ_{ij} 的取值范围为 $[-1, 1]$. 若 $\rho_{ij} = -1$, 则表示两支股票完全负相关; 若 $\rho_{ij} = 1$, 则表示两支股票完全正相关, 若 $\rho_{ij} = 0$, 则表示两只股票不相关. 由此, 利用公式 (2) 和 (3) 可以得到证券市场的股票相关系数矩阵 C , 它的大小为 $N \times N$, 元素为 ρ_{ij} .

2.2 随机矩阵理论及去噪方法总结

考虑这样一个随机矩阵: $R = \frac{1}{L} A \times A^T$, 其中 A 是一个 $N \times L$ 随机矩阵, 由 N 个长度为 L 的独立的序列构成, 每个序列均服从 $N(0, 1)$ 分布. 矩阵 R 的统计性质是已知的. 特别是, 当 $N \rightarrow \infty$, $L \rightarrow \infty$, $Q = L/N (> 1)$ 固定的情况下, R 的特征值 λ 的分布函数 $P(\lambda)$ 有解析形式:

$$P(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda}, \quad (3)$$

其中 $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$, 且有

$$\lambda_{\min}^{\max} = \sigma^2 \left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \right),$$

这里 $\sigma^2 = 1$.

Wigner, Dyson 等人在研究量子系统中得出结论, 随机矩阵理论的预测代表了所有可能相互作用的一种平均. 偏离随机矩阵理论的一般预测就等同于考虑系统中的特殊非随机的属性, 未了解潜在的相互作用提供了线索.

随机矩阵理论指出, 通过比较相关系数矩阵 C 与随机矩阵的性质, 把 C 分为两个部分: 一部分是符合随机矩阵的性质 (噪声) 和一部分是差异部分 (真实信息), 从而对相关系数矩阵进行改进, 去除网络中的噪声信息.

运用随机矩阵理论改进相关系数矩阵, 从而达到金融系统的去噪, 目前已运用于投资组合优化和风险控制中. RMT 去噪方法种类很少, 主要包括 LCPB 法^[3], PG+ 法^[4] 和 KR 法^[5]. 这三种去噪方法都认为小于随机矩阵的最大特征值的市场特征值是噪声. 其中, LCPB 法用噪声特征值均值取代这些噪声特征值; PG+ 法用零取代噪声特征值; KR 法考虑了最小替代特征值的几个不同取值情况, 以过滤掉它们的影响.

3 相关系数矩阵及随机矩阵的特征值统计性质

通过对去噪等方法的总结, 比较发现, 去噪目的就是区分出相关系数矩阵中的噪声和潜在真实信息, 从而进一步提取出真实的信息. 在金融网络中, 我们则需要比较相关系数矩阵 C 和随机相关矩阵的性质差异, 找出它们之间的差异, 再提取噪声信息和真实信息.

Plerou, Loloux 等人最早在 1999 年开始利用随机矩阵理论研究了美国股票间波动的相关性, 运用相关系数矩阵和随机矩阵的特征值的概率密度分布图来分析相关系数矩阵和随机矩阵的性质差异, 比较分析相关系数矩阵中的噪声信息和潜在真实信息, 从而达到去噪的目的 [5,6].

我们比较相关系数矩阵 C 的特征值的分布 $P(\lambda)$ 和随机矩阵 R 的分布 $P_{nn}(\lambda)$. 本文搜集了 2000 年 1 月 1 日至 2011 年 4 月 1 日在上海证券交易所上市的 884 家 A 股公司的收盘价格数据, 在这里将选取其中的 2000 年 1 月 1 日至 2009 年 4 月 1 日的股票数据, 将其分成四个时间段来比较分析. 分别

为 2000 年 123 只股票, 共有 327 个有效交易天数记为阶段 I; 2001 年 174 只股票, 共有 261 个有效交易天数, 记为阶段 II; 2002 年至 2005 年 795 只股票, 共有 1045 个有效交易天数, 记为阶段 III; 2006 年至 2009 年 857 只股票, 共有 1051 个有效交易天数, 记为阶段 IV. 根据随机矩阵理论以及 (3) 式分别计算得到 λ_{\max} 和 λ_{\min} 的值. 如表 1 所示.

表 1 随机矩阵预测的最大和最小特征值

特征值	阶段 I	阶段 II	阶段 III	阶段 IV
λ_{\max}	2.6028	3.2997	3.5052	3.6214
λ_{\min}	0.1495	0.0337	0.0163	0.0094

计算出 C 的特征值 λ_i , λ_i 按大小排序 ($\lambda_{i+1} > \lambda_i$). 图 1 对比了四个阶段的相关系数矩阵的特征值分布 (记为 $P(\lambda)$) 和随机矩阵特征值分布 (记为 $P_{nn}(\lambda)$) 的情况.

我们注意到不管是哪一阶段, 相关系数矩阵的一部分特征值明显在 $[\lambda_{\min}, \lambda_{\max}]$ 范围内, 有一部分最大的特征值偏离预测值, 也有一些最小的特征值也偏离预测值. 特别地, 相关系数矩阵的最大特征值都比 λ_{\max} 大很多.

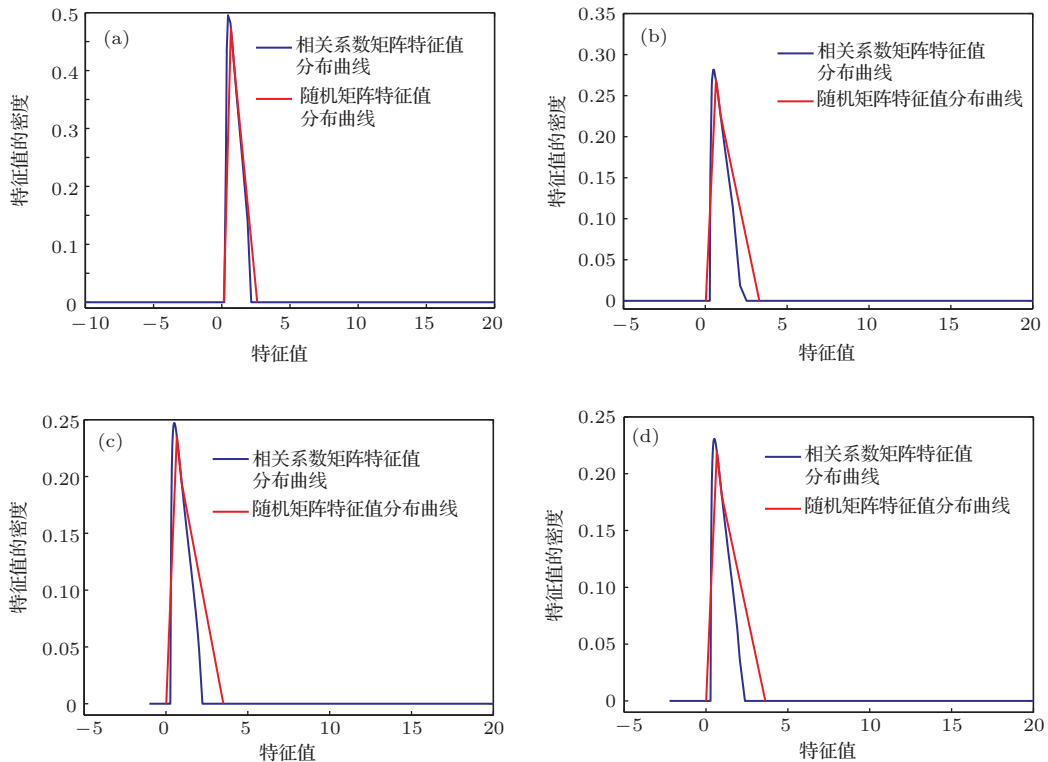


图 1 (网刊彩色) 四个阶段的相关系数矩阵和随机矩阵的特征值分布对比图 (a) 阶段 I; (b) 阶段 II; (c) 阶段 III; (d) 阶段 IV

通过比较金融相关系数矩阵 C 和随机矩阵的性质, 找出了它们之间的差异, 根据随机矩阵理论, 差异部分就是真实相关性的反应. 金融网络的去噪原理, 在于相关系数矩阵中符合随机矩阵的那部分性质, 说明了相关系数矩阵的元素的随机性, 即噪声信息, 需要去除; 而相关系数矩阵与随机矩阵的差异则反应了相关系数矩阵中的真实的相关性, 即真实信息, 去噪过程中保留此部分.

4 网络去噪方法的改进及网络模型构建

4.1 去噪方法的改进

运用随机矩阵理论, 计算得到随机矩阵的预测最大特征值和最小特征值为

$$\lambda_{\min}^{\max} = \sigma^2 \left(1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \right). \quad (4)$$

对于已有的三种去噪方法, 令金融相关系数矩阵全部特征值中属于 $[\lambda_1, \lambda_+]$ 的部分代表噪声. 通过以上比较分析金融相关系数矩阵和随机矩阵的特征值统计性质, 发现相关系数矩阵中有一部分性质符合随机矩阵的性质, 即属于 $[\lambda_-, \lambda_+]$ 的部分, 而其余部分为差异性. 据此, 我们令金融相关系数矩阵全部特征值中属于 $[\lambda_-, \lambda_+]$ 的部分代表噪声.

对相关系数矩阵 C 进行谱分解有: $C = PAP^T$, 其中 P 为正交矩阵满足条件 ($PP^T = 1$), 而 Λ 为对角矩阵, P^T 表示矩阵的转置, 特征值矩阵 $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_K\}$, 对特征值 λ 可以按照从小到大的顺序进行重新排列, 即 $\lambda_1, \dots, \lambda_t, \lambda_{t+1}, \dots, \lambda_k, \lambda_{k+1}, \dots, \lambda_K$, 令 $\lambda_{k+1} > \lambda_{\max}$, 且 $\lambda_t < \lambda_-$.

对于已有的三种去噪方法, 将相关系数矩阵的特征值中比随机矩阵理论预测最大特征值小的特征值替换为常数值或者 0. 当替换为 0 时, 即 $\Lambda_N = \{0, \dots, 0, \lambda_{k+1}, \dots, \lambda_K\}$. 也就是说将小于 λ_{k+1} 的特征值都替换为 0.

在这里, 我们对已有的去噪方法进行改进, 我们令

$$\Lambda_N = \{\lambda_1 \dots \lambda_t, 0, \dots, 0, \lambda_{k+1}, \dots, \lambda_K\}, \quad (5)$$

即将相关系数矩阵在随机矩阵理论预测范围内的特征值替换为 0, 也就是说只将中间部分替换

为 0. 同样特征值矩阵也只留下了相应的特征向量, 则新的相关系数矩阵为

$$C_{\text{New}} = P_{\text{New}} \Lambda_{\text{New}} P_{\text{New}}^T, \quad (6)$$

其中 P_{New} 表示新的特征向量矩阵, 此外令对角元素 $\omega_{\text{New}, ii} = 1$, 确保 $\text{Tr}(C) = T(C_{\text{New}}) = N$.

为下面进一步研究去噪前后网络的比较, 我们还需构建噪声相关系数矩阵, 即只保留噪声信息的相关系数矩阵. 我们令 $\lambda_k < \lambda_{\max}$, 且 $\lambda_{t+1} > \lambda_-$, 令

$$\Lambda_Z = \left\{ \underbrace{0, \dots, 0}_{t \uparrow}, \lambda_{t+1}, \dots, \lambda_k, \underbrace{0, \dots, 0}_{K-k \uparrow} \right\}, \quad (7)$$

即将相关系数矩阵在随机矩阵理论预测范围外的特征值替换为 0. 同样特征值矩阵也只留下了相应的特征向量, 则噪声相关系数矩阵为

$$C_Z = P_Z \Lambda_Z P_Z^T, \quad (8)$$

其中 P_Z 表示噪声特征向量矩阵, 此外令对角元素 $\omega_{Z, ii} = 1$, 确保 $\text{Tr}(C) = T(C_Z) = N$.

4.2 构建金融网络模型

构建金融网络的具体步骤如下:

步骤 1 利用 (1), (2) 式计算原始相关系数矩阵;

步骤 2 基于改进的去噪方法对相关系数矩阵进行改进;

步骤 3 运用边相关系数阈值法^[14], 确定阈值;

步骤 4 建立金融网络模型.

5 金融网络的去噪效果分析

本文搜集的是 2000 年 1 月 1 日—2011 年 4 月 1 日在上海证券交易所上市的 884 家 A 股公司的收盘价格数据, 共有 4109 个交易日. 此数据是所本文搜集的上海的所有数据. 基于原始相关系数矩阵 C , 构建原始金融网络模型 (简称原始网络); 基于改进的相关系数矩阵 C_{New} , 构建去噪后的金融网络模型 (简称去噪网络); 基于噪声相关系数矩阵 C_Z , 构建噪声金融网络模型 (简称噪声网络).

5.1 基于关键节点对网络去噪前后的比较

无标度网络中存在一定数量的度值(连接数)极高的节点,即度最大的那些节点通常称为hub节点,也称为网络的关键节点^[15].少数hub节点拥有极其多的连接,而大多数节点只有少量的连接.少数hub点对网络的运行起着主导的作用.

为研究网络去噪前后的变化情况,首先选取上述上海的所有数据,即2000年1月1日至2011年4月1日的股票数据构建原始网络、去噪网络,以及噪声网络,分别找到网络的关键节点,并进行对比分析.如表2所示.

由表2可以看出,去噪网络与原始网络的度最大的节点主要是801, 867, 802, 866, 834,并且表中节点的度排名前十位的变化都不大,而这两个网络与噪声网络进行比较发现,噪声网络的关键节点的度相对于前两者,变化较大,度最大只有20,并且不再是原始网络和去噪网络的关键节点.

为进一步分析网络的关键重要信息和噪声信息,我们选取884只股票中的任意100只来构建金融网络,并找出原始网络、去噪网络以及噪声网络的关键节点,进行对比分析.并重复此过程五次,即重复选取五次100只股票进行分析.其中我们选取某一次的结果,如表3所示.

表2 网络的关键节点比较(a)

原始网络		去噪网络		噪声网络	
节点编号	节点的度	节点编号	节点的度	节点编号	节点的度
801	149	801	140	518	20
867	135	867	126	839	20
802	126	802	122	851	18
866	124	866	117	155	17
834	101	834	98	270	17
852	86	852	86	178	16
799	77	799	79	579	15
884	75	821	79	823	15
821	74	872	71	816	14
872	72	884	69	870	14

表3 网络的关键节点比较(b)

原始网络		去噪网络		噪声网络	
节点编号	节点的度	节点编号	节点的度	节点编号	节点的度
6	71	37	71	55	55
37	71	6	69	38	54
28	66	9	69	26	53
93	65	83	69	64	53
8	64	97	69	80	53
97	64	8	67	5	52
9	63	28	67	22	52
14	63	30	64	23	52
30	63	88	64	60	52
83	61	93	64	1	51

由表 3 可以看出, 去噪网络与原始网络的节点的度排名前十位的变化都不大, 度最大的节点主要是 6, 37. 而这两个网络与噪声网络进行比较发现, 噪声网络的关键节点相对于前两者, 变化较大, 原始网络和去噪网络的关键节点的度在噪声网络中都变得较小, 即不再是网络的关键节点. 重复选取五次, 发现得出的结果均类似于上述结果.

由上述去噪网络和噪声网络分别与原始网络的关键节点的对比分析, 基于随机矩阵理论构建的去噪网络保留了原始网络的关键重要的信息, 而噪声网络过滤了原始网络的关键重要的信息. 并且, 噪声网络的关键节点只是原始网络中度比较小的节点. 由于网络的关键节点反映的是网络中的关键并且非常重要的信息, 则结合上述分析, 可以得出,

金融网络去噪保留原始网络的关键重要信息, 而噪声信息对应的是原始网络中除了关键节点以及连接度较大的其他节点反映的相关信息.

5.2 网络的拓扑结构分析

5.2.1 最小生成树

最小生成树是由 Mantegna 为了检测金融网络中的等级结构于 1999 年提出的. 一个由 n 个顶点的连通图的生成树是图的极小连通子图, 且包含原图中的所有 n 个顶点, 并且有保持图连通的最少的边. 最小生成树是连接图顶点的所有生成树中权值最小的生成树. 最小生成树是一类特殊的图, 是一种具有规则拓扑结构的规则网络 [16].

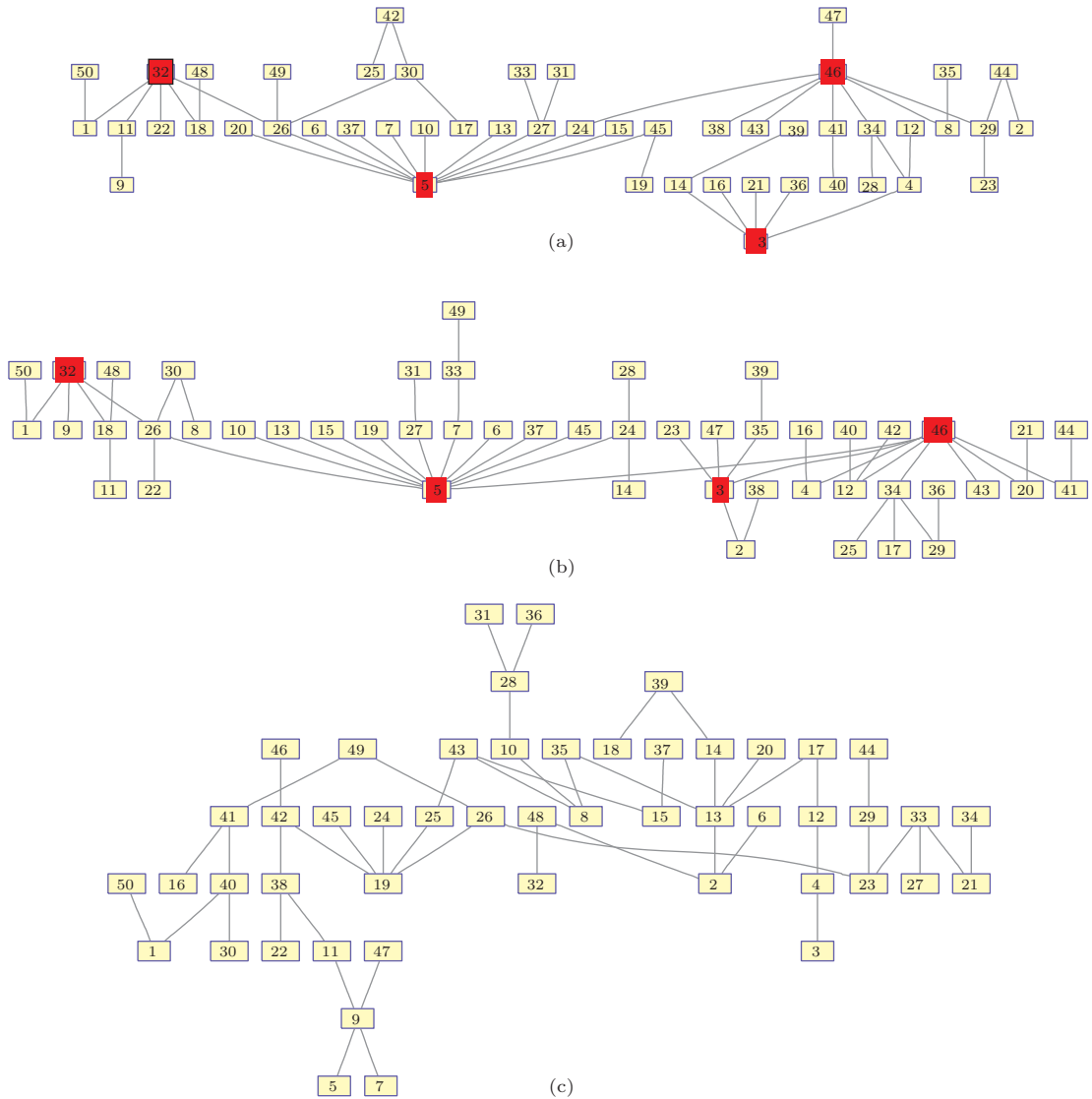


图 2 最小生成树 (a) 原始网络的最小生成树; (b) 去噪网络的最小生成树; (c) 噪声网络的最小生成树

在相关系数矩阵的基础上, 定义股票间的度量距离为

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}. \quad (9)$$

在这种定义下, d_{ij} 完全满足测量的三个公理: 1) $d_{ij} = 0$ 当且仅当 $i = j$; 2) $d_{ij} = d_{ji}$; 3) $d_{ij} \leq d_{ik} + d_{kj}$. 于是距离矩阵 D 可以用来确定连接 n 只股票的最小生成树.

由于最小生成树结构规则简单, 节点数较少. 本文选取 884 只股票中的任意 50 只股票为研究对象, 分别得到原始网络、去噪网络以及噪声网络的最小生成树. 如图 2 所示.

由图 2(a) 和 (b) 中可以发现, 原始网络和去噪网络的最小生成树的 hub 节点变化不大, 都为节点 5, 46, 32, 3. 由 2(c) 可以看出, 噪声网络的最小生成树的节点结构变化很大, 去除了将原网络的关键重要信息, 而噪声网络的关键节点变为原始网络度很小的节点, 如节点 19. 说明基于全新的去噪方法构建的金融网络保留了网络的关键重要的信息, 与上述基于关键节点对网络去噪效果分析得到的结果一致.

5.2.2 模体结构

网络中的子图可以称之为模体, 需满足以下条件: 该子图在与实际网络对应的随机网络中出现的次数大于它在真实网络中出现的概率是很小的, 通常要求这个概率小于某个阈值 P (如 $P = 0.01$); 该子图在实际网络中出现的次数 N_{real} 不小于某个下限 U (如 $U = 4$); 该子图在实际网络中出现的次数 N_{real} 明显高于它在随机网络中出现的次数 N_{rand} , 一般要求 $N_{\text{real}} > 1.1N_{\text{rand}}$ [15]. 模体是网络的基本拓扑结构之一, 它的大小介于网络个体和社团之间, 一般由少数几个节点连接构成, 模体可以揭示网络的演化规律, 它是社团内部成员之间基本

的连接模式 [17], 其结构对应着个体之间的基本组成模式.

模体的 P -value 是指模体在随机网络中出现的次数大于在实际网络中出现的次数的频率. 模体的 P -value 越小, 该模体在网络中越重要.

任意一个模体 M_i , 设它在实际网络中出现的次数为 N_{real_i} , 在随机网络中出现的次数为 N_{rand_i} , N_{rand_i} 的平均值为 $\langle N_{\text{rand}_i} \rangle$, 标准差为 σ_{rand_i} , 则模体 M_i 在实际网络中的 Z -score 为

$$Z_i = \frac{N_{\text{real}_i} - \langle N_{\text{rand}_i} \rangle}{\sigma_{\text{rand}_i}}. \quad (10)$$

对 Z_i 进行归一化处理, 有

$$Z'_i = \frac{Z_i}{\sqrt{\sum_i Z_i^2}}. \quad (11)$$

Z -score 用来衡量模体的重要性, Z -score 越大, 模体在网络中越重要. Z -score 的归一化处理用来强调子图的相对重要性, 这在不同规模的网络做比较时是极其重要的, 一般地, 大型网络中模体的 Z -score 比小型网络中要大 [18].

在金融网络中, 模体的结构在一定程度上可以反映出股票间的某种联系, 对于研究金融市场中不同股票之间的相互作用以及投资者购买行为具有重要意义.

目前, 用来检测网络中模体的软件主要有 Mfinder 和 Fanmod, Fanmod 软件运用 rand-ESU 算法对模体进行检测, 该算法速度相对较快, 并且可以检测网络中三个节点到八个节点的模体, 可以检测出更多种类的模体, 本文中我们选用该算法检测金融网络中的模体.

我们以上海所有股票数据构建的沪市金融网络为对象, 检测了原始网络以及去噪网络的三个节点和四个节点的模体, 并进行比较分析. 如表 4 和表 5 所示.

表 4 原始网络的三节点模体检测分析

模体编号	形状	频率/%	平均频率/%	标准差	Z 值	P 值
238		8.5434	0.50106	0.0042387	18.974	0

表 5 去噪网络的三节点模体检测分析

模体编号	形状	频率/%	平均频率/%	标准差	Z 值	P 值
238		9.7303	2.2811	0.0020729	35.936	0

表6 原始网络的四节点模体检测分析

模体编号	形状	频率/%	平均频率/%	标准差	Z 值	P 值
31710		0.50952	0.0014356	4.4603×10^{-6}	113.91	0
13278		3.0123	0.19788	0.0016661	16.892	0
4958		15.042	1.3786	0.012734	10.73	0

表7 去噪金融网络的四节点模体检测分析

模体编号	形状	频率/%	平均频率/%	标准差	Z 值	P 值
31710		0.68736	0.024729	7.2508×10^{-5}	120.387	0
13278		3.6436	1.1894	0.00063665	38.548	0
4958		16.447	6.4233	0.0049277	20.342	0
4382		62.677	61.887	0.001868	4.2316	0

由表4和表5的比较可以看出,原始网络和去噪网络的三节点的模体都只有一种,Z值分别为18.974和35.936,去噪网络的三节点模体Z值增大.由表6和表7可以看到,原始网络的四节点模体为三种,去噪网络的四节点模体增加为四种,由两者前三种模体的Z值比较来看,去噪网络的模体Z值也都增大.综上可得到,去噪网络模体结构更加明显,在网络中的重要性增加.

5.2.3 社团结构

社团结构定义为内部节点之间的连边数目比节点与外部连边数目大的网络子图,它是网络的中观结构,也是研究网络拓扑结构和网络组成的重要特征.一般而言,社团可以包含模块、类、群、组等各种含义^[19].目前,用来划分社团的算法主要有Kemighan-Lin算法、基于Laplace图特征值的谱平分法、以及GN算法等.

其中,GN算法弥补了一些传统算法的不足.并且,Newman等人引进了一个衡量网络划分质量的标准——模块性^[20],它是用来衡量社团结构划分好坏的重要测度,模块性的衡量标准Q定义如下:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tre} - \|e^2\|, \quad (12)$$

其中,元素 e_{ij} 表示网络中连接两个不同社团的节点的边在所有边中所占的比例;这两个节点分别位于第*i*个和第*j*个社团;Tre表示矩阵对角线上各元素之和,即 $\text{Tre} = \sum_i e_{ii}$,它给出了网络中连接某一个社团内部各节点的边在所有边的数目中

所占比例; a_i 表示每行(或者列)中各元素之和,即 $a_i = \sum_j e_{ij}$,它表示与第*i*个社团中的节点相连的边在所有边中所占的比例; $\|x\|$ 表示矩阵*x*中所有的元素之和.在实际网络中,划分不同个数的社团,会相对应一个Q值,找到它的局部峰值.Newman等人发现这些峰值的位置与所期望的划分位置密切相关,也可认为该峰值为最优Q值.

为进一步比较去噪前后网络的拓扑结构,这里我们根据上述上海证券市场的股票数据的四个阶段以及上海股票总数据,分别构建金融网络模型,运用GN算法对原始网络以及去噪网络进行社团划分,划分的最优Q值如表8所示.

表8 原始网络与去噪网络的最优Q值比较

最优Q值	阶段I	阶段II	阶段III	阶段IV	总数据
原始网络最优Q值	0.013	0.013	0.323	0.204	0.119
去噪网络最优Q值	0.035	0.014	0.331	0.348	0.128

从表8中得到,四个阶段的去噪网络的Q值都比原始网络的Q值大,说明去噪网络的社团结构更加稳固,更有利于分析网络的拓扑性质.下面根据上海证券市场的股票的总数据,划分社团,在原始网络和去噪网络的Q值分别为0.119以及0.128时,可以划分10个社团如图3(a)和(b)所示.

从图3(a)和(b)中看到,运用GN算法对去噪网络的社团划分,除了第一主要社团以外,其他社团分布比原始网络的社团分布更加均匀.以去噪网络的社团为对象,按照行业统计,去噪网络的最大社团,即第一主要社团涉及的行业主要有制造业,

金融业, 以及采掘业, 如中国石油、中信证券、中海集运等公司. 主要社团节点数占据整个网络的比例达到85.3%, 社团结构清晰. 可以看出, 上海金融市场呈现高度集中式发展, 少数巨大的社团构成了网络的主要部分.

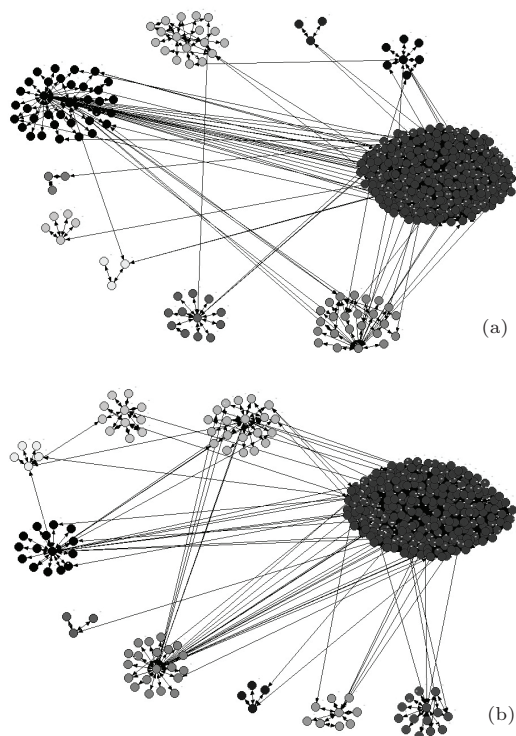


图3 社团结构 (a) 原始金融网络的社团结构; (b) 去噪金融网络的社团结构

6 结 论

本文基于随机矩阵理论, 通过对上海证券市场的金融相关系数矩阵以及随机矩阵的特征值统计性质的分析, 提出了全新的去噪方法. 运用网络的关键节点、最小生成树对原始网络、去噪网络以及噪声网络进行了对比分析, 发现去噪网络保留了原始网络的关键重要的信息, 即保留了上海金融市场的重要信息, 与此相反, 噪声网络过滤了原始网络重要的信息. 通过对原始网络和去噪网络的模体结构和社团结构的分析比较, 结果得出去噪网络的模体结构在网络中的重要性增加, 这更有利于研究金融市场中不同股票之间的相互作用以及投资者的购买行为. 并且去噪网络的社团分布更加均匀, 表现出上海金融市场呈现集中式发展的趋势, 符合实际市场的发展情况. 基于改进的去噪方法, 去噪网络均表现出更良好的拓扑性质. 构建的网络的拓扑结构更加紧密, 可以更好地反应金融网络的内在

性质.

金融市场的演化规律还有很多, 将新的去噪方法运用到其他金融市场中, 有待我们继续探索研究. 考虑到金融相关系数矩阵在投资组合以及风险控制中均有十分重要的作用, 而其中用到的实证相关系数矩阵又存在大量的噪声信息, 因此基于新的去噪方法, 研究金融市场的投资组合以及风险控制是我们下一步的工作.

参考文献

- [1] Li G H, Zhang H, Luo M K 2012 *Chin. Phys. B* **21** 128901
- [2] Zhang X D, Liu X D, Zheng Y, Liu C 2013 *Chin. Phys. B* **22** 030509
- [3] Li R, Yan P L, Chen J, Li J, Li J, Zhang K W, Zhong J X 2009 *Acta Phys. Sin.* **58** 6703 (in Chinese) [李蓉, 颜平兰, 陈健, 李俊, 李金, 张凯旺, 钟建新 2009 物理学报 **58** 6703]
- [4] Xu Z X, Wang Y, Si H B, Feng Z M 2011 *Acta Phys. Sin.* **60** 040501 (in Chinese) [徐赞新, 王钺, 司洪波, 冯振明 2011 物理学报 **60** 040501]
- [5] Laloux L, Cizeau P, Bouchaud J P, Potters M 1999 *Phys. Rev. Lett.* **83** 1468
- [6] Plerou V, Gopikrishnan P, Rosenow B, Amaral L A N, Guhr T, Stanley H E 2002 *Phys. Rev. E* **65** 1
- [7] Sharifi S, Crane M, Shamaie A, Ruskin H 2004 *Phys. A: Stats. and Theo. Phys.* **335** 629
- [8] Conlon T, Ruskin H J, Crane M 2007 *Physics A* **382** 565
- [9] Namaki A, Jafari G R, Raei R 2011 *Physics A* **390** 3020
- [10] Rojkova V, Khali Y, Elmaghraby A, Kantardzic M 2007 *IEEE Int. Symp. on Sig. Proc. and Inform. Technol.* Giza, December 15–17, 2007 p647
- [11] Feher K, Whelan J, Mueller S 2011 *Stats. Appl. in Gene. and Mol. Biol.* **10** 44
- [12] Namaki A, Shirazi A H, Raei R, Jafari G R 2011 *Phys. A* **390** 3835
- [13] Kumar S, Deo N 2012 *Phys. Rev.* **E86** 026101
- [14] Boginski V, Butenko S, Pardalos P 2005 *Comput. Stats. and Data Anal.* **48** 431
- [15] Wang X F, Li X, Chen G R 2006 *The. and Appl. of Comp. Net.* (Beijing: Tsinghua University Press) p37 (in Chinese) [汪小帆李翔陈关荣 2006 复杂网络理论及其应用 (北京: 清华大学出版社) 第 37 页]
- [16] Mantegna R N 1999 *Eur. Phys. J. B* **11** 193
- [17] Shen-Orr S S, Milo R, Mangan S, Alon U 2002 *Nat. Genet.* **31** 64
- [18] Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S 2008 *Science* **303** 1538
- [19] Gulbahce N, Lehmann S 2008 *Bio. Ess.* **30** 934
- [20] Newman M E J, Girvan M 2004 *Phys. Rev. E* **69** 026113
- [21] Han H, Liu W L, Wu L Y 2013 *Acta Phys. Sin.* **62** 168904 (in Chinese) [韩华, 刘婉璐, 吴翎燕 2013 物理学报 **62** 168904]

Financial networks model based on random matrix*

Han Hua Wu Ling-Yan[†] Song Ning-Ning

(Department of Science, Wuhan University of Technology, Wuhan 430070, China)

(Received 17 October 2013; revised manuscript received 27 March 2014)

Abstract

Random matrix theory is applied to study the correlation between different financial correlation coefficient matrices in the financial field. Correlation coefficient matrix is a key factor for constructing a network. In this paper we relate the random matrix theory to the network construction to study the financial networks model in terms of the random matrix. We select the stock data of Shanghai stock market, and divide them into four stages. We discuss the statistical properties of eigenvalues in financial correlation coefficient matrix and random matrix based on the random matrix theory, and improve the existing denoising method to construct the correlation coefficient matrix and to make it more suitable for building financial networks. After that we can build the financial network model. Then we analyze and compare the original financial network, the denoising financial network and the noise financial network in terms of the random matrix theory and the key node of networks. It is found that the primary important information is still in the original network, and the noise information corresponds to the information which the nodes of small degree in the original network include. Finally we analyze the topological structure of the financial networks, such as the minimum spanning tree, the motif and community structure. We also find that the topological properties of the improved financial networks are more remarkable and the topological structure is more compact.

Keywords: random matrix theory, financial correlation coefficient matrix, financial networks

PACS: 89.75.-k, 89.75.Fb

DOI: 10.7498/aps.63.138901

* Project supported by the National Natural Science Foundation of China (Grant No.71140015), and the Fundamental Research Fund for the Central Universities of China (Grant No. 2013-Ia-007, 2013-Ia-040).

[†] Corresponding author. E-mail: 815007651@qq.com; hanhua@whut.edu.cn