

## 基于网络结构的股票相关性研究

侍冰雪<sup>a</sup>, 魏慧茹<sup>b</sup>, 朱韶东<sup>b</sup>, 朱家明<sup>b</sup>

(安徽财经大学 a. 金融学院; b. 统计与应用数学学院, 安徽 蚌埠 233030)

**摘 要:** 针对股票间相关性, 选取了中国股票市场的沪深主板、创业板、中小板中共 100 只股票在 2013 年 1 月 1 日至 2013 年 8 月 31 日的周收盘价进行数据分析。首先计算各股票的对数收益率, 建立样本股票之间的网络结构, 分析股票网络结构的统计特性; 然后重新对样本股票进行板块划分, 并与同一股票市场的网络结构研究结果进行对比分析。

**关键词:** 网络结构; 股票相关性; 相关系数; 股票板块划分

**中图分类号:** F224.0; F832.5

**文献标志码:** A

**文章编号:** 1671-8798(2015)01-0062-06

## Research on correlation between stocks based on network structure

SHI Bingxue<sup>a</sup>, WEI Huiru<sup>b</sup>, ZHU Shaodong<sup>b</sup>, ZHU Jiaming<sup>b</sup>

(a. School of Finance; b. School of Statistics and Applied Mathematics,  
Anhui Finance and Economics University, Bengbu 233030, China)

**Abstract:** According to the correlation between stocks, we select 100 kinds of stocks from Shanghai A and B shares, Shenzhen A and B shares, GEM, small and medium-sized plate stock in China stock market. We analyze their week closing price during January 1, 2013 to August 31, 2013. Firstly, we calculate the logarithmic rate of return of stock shares, then set up network structure and analyze the statistical characteristics of stock network structure. We not only divide the sample stock plate, but also explain the difference between our result and the actual. Finally, we analyze the research on the network structure of the same stock market comparatively.

**Key words:** network architecture; stock correlation; linear programming; plate division

股票间的相关性对于风险监管部门、金融投资者都具有重要意义。在研究股票间相关性时, 已不再局限于对影响资产现金流和资产折现率等经济基本面因素的研究, 而是越来越关注股票间相关程度, 将各种经济现象和复杂的社会活动抽象为网络分析。股票市场复杂系统对其他领域的影响日益加深, 经

---

**收稿日期:** 2014-11-12

**基金项目:** 国家自然科学基金项目(11301001); 国家级大学生创新创业训练计划项目(201410378194); 安徽省大学生创新创业计划项目(AH201410378311; AH201410378483)

**作者简介:** 侍冰雪(1994—), 女, 安徽省泗县人, 2012 级金融工程专业本科生。

**通信作者:** 朱家明, 副教授, 硕士, 主要从事应用数学与数学建模研究。

济、数学、社会等领域的学者也都开始运用网络结构来研究股票间的相关性,包括对成交量与收益率、市盈率与收益率之间关系的研究等。通过构建股票相关性网络,基于网络结构中的单个股票的平均路径和集聚系数对样本股票按相关性重新划分板块,对投资者来说,当持有的2只股票间相关性较大即具有同向波动性时,潜在风险很大,因此,这也对投资者的投资决策提供了良好的参考<sup>[1]</sup>。

## 1 构建股票间相关性网络

### 1.1 对数收益率

要研究不同市场股票间的相关性,首先采用随机抽样的方法,利用 Matlab7.0 编程,从中国股票市场选取了 20 只沪市 A 股、20 只沪市 B 股、10 只深市 A 股、20 只深市 B 股、20 只创业板股、10 只中小板股共 100 只股票作为研究对象。从投资者关心程度、反应股票当期价值等因素考虑,以收盘价作为原始指标,通过对股价对数化,得到每只股票的对数收益率作为相关性指标,即  $R_i(t) = \ln p_i(t) - \ln p_i(t-1)$ ,其中  $R_i(t)$  表示股票  $i$  第  $t$  周的对数收益率,  $p_i(t)$  为股票  $i$  的第  $t$  周收盘价。在计算股票的对数收益率之前,先对股票价格时间序列进行平稳性检验和消除趋势,然后建立度量模型,再分别计算每 2 只股票之间的相关性系数。本研究所用的股票周收盘价数据来自锐思数据库。

#### 1.1.1 平稳性检验

本研究中的股票收盘价时间序列由如下随机过程生成<sup>[2]</sup>:  $X_t = X_{t-1} + \mu_t$ ,这里的  $\mu_t$  是一个白噪声,  $X_t$  表示股票  $i$  的第  $t$  周的平均对数收益率。

容易知道,股票收盘价时间序列有相同的均值  $E(X_t) = E(X_{t-1})$ 。为了检验该序列是否有相同的方差,可假设  $X_t$  的初值为  $X_0$ ,则易知:

$$\begin{aligned} X_1 &= X_0 + \mu_1 \\ X_2 &= X_1 + \mu_2 = X_0 + \mu_1 + \mu_2 \\ X_t &= X_0 + \mu_1 + \mu_2 + \cdots + \mu_t \end{aligned}$$

假定初始值  $X_0$  为常数,  $\mu_t$  是一个白噪声,因此  $\text{Var}(X_t) = t\sigma^2$ ,即  $X_t$  的方差与时间  $t$  有关而非常,故它是非平稳序列。

#### 1.1.2 采用一阶差分法消除趋势

然而,对  $X_t$  取一阶差分  $\Delta X_t = X_t - X_{t-1} = \mu_t$ ,由于  $\mu_t$  是一个白噪声,则序列  $\{\Delta X_t\}$  是平稳的。于是得到股票收盘价时间序列经过一阶差分法消除趋势后的时间序列,再用该序列中的股票收盘价来计算股票的对数收益率  $R_i(t)$ 。

### 1.2 绝对相关系数

计算 100 只样本股票对数收益率间的相关系数  $\beta$ ,并取绝对值。相关系数  $\beta \in [-1, 1]$ ,相关系数  $\beta$  为 1 时,表明对应的 2 只股票之间完全正相关,即一只股票价格变动会引起另一只股票同方向同幅度的变化; $\beta$  为 -1 时,2 只股票之间完全负相关; $\beta$  为 0 时,2 只股票之间不相关。由于市场上不同的 2 只股票之间不可能完全相关,因此,本研究将所有  $\beta$  为 1 的值看作股票与自身之间的关联系数。通过 Matlab7.0 作出不同股票对数收益率间的绝对相关系数概率分布图(图 1)。

图 1 表明,所选取的 100 只股票的周对数收益率的绝对相关系数大多数分布在 0.25~0.6 之间,在 0~0.25 和 0.6~1 的区间内分布相对较少。

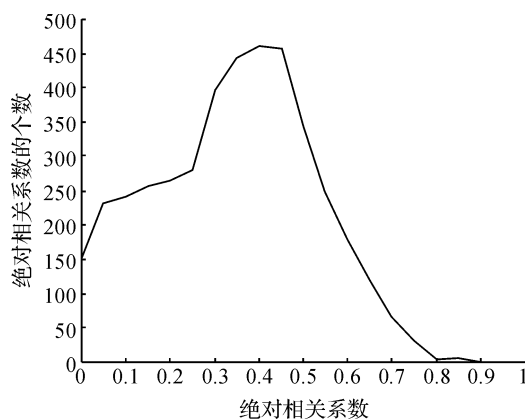


图 1 绝对相关系数分布图

Fig. 1 Probability distribution map of absolute correlation coefficient

### 1.3 股票相关性网络

#### 1.3.1 小世界网络结构

小世界网络模型<sup>[3]</sup>是一类既具有较短的平均路径长度又具有较高的集聚系数的网络的总称。定义  $d_{ij}$  为连接 2 个节点  $i$  和  $j$  之间的最短路径上的边数, 定义网络的平均路径长度  $S$  为任意 2 个节点之间的平均值, 即:

$$S = \frac{2}{N(N+1)} \sum_{i \leq N} \sum_{j \geq i} d_{ij} \quad (1)$$

式(1)中:  $N$ —网络节点数。公式(1)<sup>[4]</sup>包含了节点到自身的距离, 即 0。因股票间的相关系数不满足度量空间的条件, 为了进一步分析股票网络的结构性质, 故将绝对相关系数  $|\beta_{ij}|$  转化为相对应的距离  $d_{ij}$ , 转化公式<sup>[5]</sup>为:

$$d_{ij} = \sqrt{2(1-|\beta_{ij}|)} \quad (2)$$

股票价格之间的负相关仍属于相关关系, 当相关系数为正值时, 表明 2 只股票价格之间存在同向变动的关系; 当相关系数为负值时, 表明 2 只股票价格之间存在反向变动的关系。因此, 公式(2)中的  $\beta_{ij}$  取绝对值更合理, 另外, 该公式来源于欧式距离理论:

$$d_{ij} = \|r_i - r_j\|^2 = \sum_{k=1}^n (r_{ik} - r_{jk})^2 \quad (3)$$

其中  $r_i = \frac{R_i(t) - E(R_i)}{\sqrt{\text{Var}(R_i)}}$ ,  $\sum_{k=1}^n r_{ik}^2 = 1$ , 且  $d_{ij}$  满足测量公理:

$$\begin{cases} d_{ij} = 0, \text{当 } i = j \\ d_{ij} = d_{ji} \\ d_{ij} \leq d_{ik} + d_{kj} \end{cases}$$

#### 1.3.2 股票间相关性网络

由于负相关系数和正相关系数一样, 也是股票之间相关性的体现, 据图 1 绝对相关系数概率分布分析, 取阈值为 0.25, 即当股票对数收益率间的绝对相关系数大于或等于 0.25 时, 对应的股票节点之间有边连接, 小于 0.25 时没有边连接。利用 Netdraw 软件绘出 100 只股票的网络结构图<sup>[6]</sup>如图 2 所示。

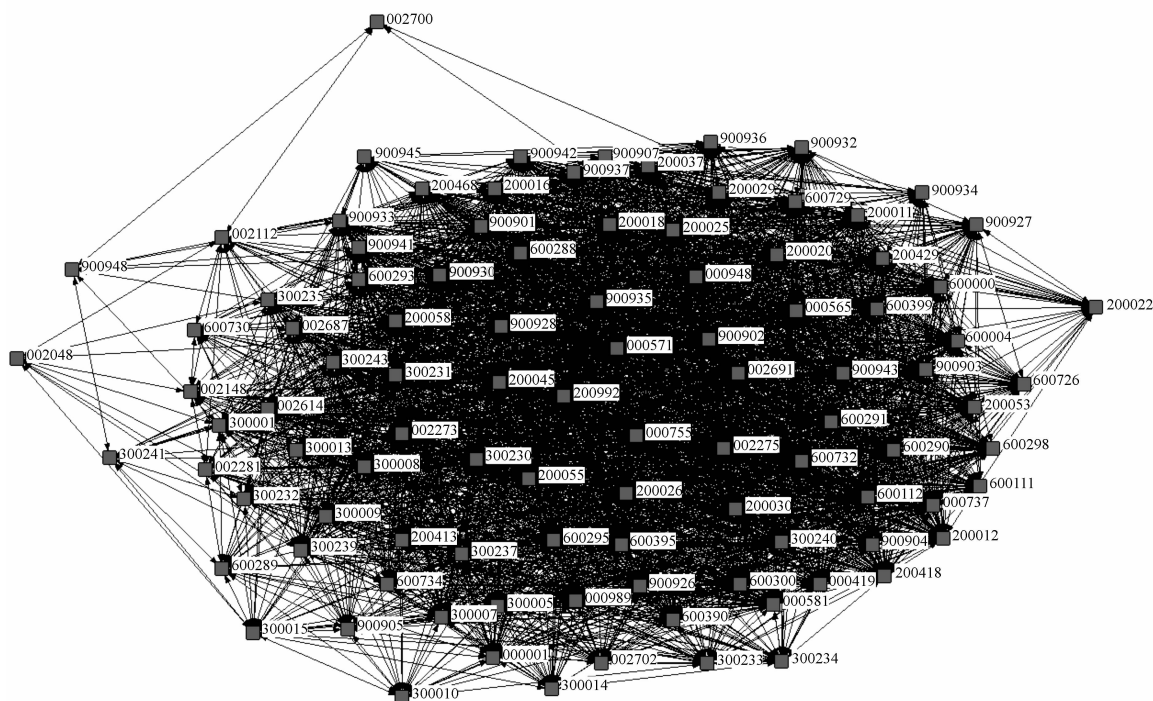


图 2 股票间网络结构图

Fig. 2 Network structure diagram of stocks

用集聚系数来描述网络<sup>[7]</sup>中节点之间结集成团的程度,假定网络中的一个节点  $i$  有  $k_i$  条边与其相连,这  $k_i$  个节点之间最多可能有  $k_i(k_i-1)/2$  条边,这  $k_i$  个节点之间实际存在的边数为  $E_i$ ,则

$$C_i = \frac{2E_i}{k_i(k_i-1)} \quad (4)$$

式(4)中: $C_i$ —集聚系数。整个网络的集聚系数<sup>[8]</sup> $C$ 就是所有节点  $i$  的集聚系数  $C_i$  的平均值。当  $C$  为 0 时,所有节点均为孤立点,任意 2 个节点之间没有边连接,即所有股票之间都不存在联系;当  $C$  为 1 时,股票网络中任意 2 个节点之间都有边,即股票网络全局是耦合的。计算 100 只股票任意 2 个之间相关系数大于阈值 0.25 的股票节点之间的平均路径和集聚系数见表 1。

表 1 股票网络统计特征表

Table 1 Statistical characteristics of stock network

统计特征	值
集聚系数 $C$	0.608 2
平均路径 $S$	1.152 0

所建立的股票间相关性网络模型中,连接任意 2 个节点之间的最短关系链中的平均节点数为 1.152 0,在 1 和 2 之间,说明相关性股票网络中平均任意一个节点对之间通过 1 或 2 个节点就能连接。集聚系数为 0.608 2,据此认为股票网络中任意 2 个节点之间有边连接的概率为 60.82%。该集聚系数不是很大,说明按照股票市场板块从 8 个附表中依次抽样选取的 20 只沪市 A 股、20 只沪市 B 股、10 只深市 A 股、20 只深市 B 股、20 只创业板股、10 只中小板股之间的小世界性并不显著,样本股票收益率之间的波动关联性不紧密;较小的平均路径又说明各个股票收益率之间相互影响,虽然这些股票不属于同一板块,但是不同板块的股票之间的价格波动也会相互影响,股票之间有边连接的节点有较强的同步关联性。由较小的平均路径和不是很大的集聚系数可以认为 100 只样本股票相互之间有联系,但并不密切,这也与现实中该 100 只股票不属于同一板块的情况相一致。而杨治辉等<sup>[5]</sup>研究的沪深 300 成分股 13 个行业的 284 只股票构成的网络为小世界网络,黄玮强等<sup>[6]</sup>研究的上海 180 指数和深圳 100 指数成分股分别构成的股票市场的 PMFG 关联网络也均为小世界网络,对比可知,不同股票市场的股票间的关联度比相同市场的股票间的相关性要小。

## 2 对样本股票划分板块

### 2.1 集聚系数和平均路径

单个股票的集聚系数和最短路径能够说明该股票在网络中与其他股票之间的联系,集聚系数越大表明该股票与整个股票网络的联系越紧密,平均路径越短,表明该股票与网络中其他股票的连接越密切。以集聚系数  $C$  和最短路径为指标对样本股票进行分类,据此对样本股票重新划分板块。将各股票相关系数代入公式(1)和公式(4)求得每只股票的集聚系数和平均路径,如表 2 所示。

### 2.2 依据模糊 $C$ 均值聚类划分板块

模糊  $C$  均值聚类算法是一个简单的迭代过程,具体过程如下,定义目标函数为

$$J(U, V) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m (d_{ik})^2$$

显然  $J(U, V)$  表示了各类中股票的集聚系数  $C$  和平均路径  $S$  到聚类中心的加权距离平方和,权重是样本的  $x_k$  对第  $i$  类隶属度的  $m$  次方,聚类准则取为求的极小值:

$$(\min) \{J(UV)\}$$

其中聚类中心为: 
$$v_i = \sum_{k=1}^n (\mu_{ik})^m x_k / \sum_{k=1}^n (\mu_{ik})^m (i = 1, 2, \dots, c), (m > 1)$$

且有: 
$$\mu_{ik} = 1 / (d_{ik} / d_{jk})^{\frac{2}{m-1}}, d_{ik} = \|x_k - v_i\|$$

将表 2 中股票的集聚系数和路径数据代入上述模型,在 Matlab7.0 中计算,可以求出股票聚类中心矩阵,进而得出最终的模糊分区矩阵。使用上述方法,可以根据中心坐标的特点,依据所选取的 100 只股票所属的聚类中心对股票分类,将 6 个类型分别作为新的股票板块。100 只股票的板块划分结果见表 3。

表 2 股票的集聚系数和平均路径

Table 2 Clustering coefficient and average path of stocks

代码	S	C	代码	S	C	代码	S	C	代码	S	C
002687	1.251 2	0.39	000581	1.107 2	0.75	200020	1.044 8	0.78	200418	1.144 9	0.71
002691	1.168 4	0.61	000737	1.101 2	0.77	200022	1.315 2	0.32	200429	1.093 9	0.70
002691	1.073 2	0.85	000755	1.068 7	0.82	200025	1.051 6	0.71	200468	1.130 3	0.64
002700	1.429 8	0.10	000948	1.160 2	0.64	200026	1.097 3	0.80	200992	1.054 0	0.79
002614	1.185 4	0.56	000989	1.150 7	0.75	200029	1.072 1	0.71	300001	1.202 8	0.56
002273	1.097 8	0.85	002048	1.367 2	0.10	200030	1.101 6	0.76	300005	1.181 5	0.57
002275	1.080 8	0.90	002112	1.264 2	0.27	200037	1.088 8	0.75	300007	1.142 8	0.70
002281	1.253 8	0.35	002148	1.255 5	0.30	200045	1.040 9	0.86	300008	1.192 0	0.53
000001	1.213 1	0.48	200011	1.181 1	0.62	200053	1.166 3	0.60	300009	1.233 7	0.49
000419	1.159 6	0.66	200012	1.154 1	0.61	200055	1.021 5	0.77	300010	1.314 4	0.24
000565	1.101 0	0.73	200016	1.128 4	0.65	200058	1.071 5	0.83	300013	1.214 3	0.47
000582	1.091 0	0.81	200018	1.127 2	0.70	200413	1.168 9	0.38	300014	1.266 9	0.30
600288	1.116 5	0.73	600300	1.125 0	0.75	600112	1.112 0	0.73	900902	1.058 6	0.71
600289	1.278 3	0.27	600390	1.189 5	0.60	600726	1.248 1	0.36	900903	1.118 5	0.67
600290	1.061 2	0.80	600395	1.123 6	0.73	600729	1.149 9	0.67	900904	1.140 7	0.65
600291	1.146 1	0.67	600399	1.183 6	0.62	600730	1.280 3	0.27	900905	1.248 9	0.37
600293	1.143 9	0.64	600000	1.123 4	0.66	600732	1.126 8	0.73	900907	1.097 5	0.77
600295	1.023 6	0.86	600004	1.113 7	0.64	600734	1.204 6	0.54	900926	1.128 0	0.70
600298	1.214 9	0.56	600111	1.201 4	0.50	900901	1.095 4	0.75	900927	1.261 9	0.43
900937	1.143 2	0.68	900941	1.163 3	0.62	900942	1.100 1	0.66	900943	1.132 1	0.69
300015	1.301 2	0.28	300230	1.053 9	0.90	300231	1.176 6	0.56	300232	1.230 6	0.40
300233	1.245 3	0.37	300234	1.257 2	0.33	300235	1.234 8	0.37	300237	1.194 3	0.55
300239	1.177 9	0.60	300240	1.099 8	0.81	300241	1.321 6	0.17	300243	1.077 1	0.87
900928	1.046 8	0.79	900930	1.065 9	0.83	900932	1.203 6	0.48	900933	1.136 6	0.67
900934	1.281 8	0.29	900935	1.095 0	0.76	900936	1.207 0	0.58	900945	1.287 9	0.23

表 3 股票板块划分结果

Table 3 Result of stock plate division

板块 1	板块 2	板块 3	板块 4	板块 5	板块 6
000565	002702	002614	002700	002691	002687
000571	002281	000001	002281	002273	000419
000581	002112	200011	300241	002275	000948
000737	002148	200413	900948	000755	000989
200026	200022	300001		200020	200012
200029	300010	300005		200025	200016
200030	300014	300008		200045	200018
200037	300015	300009		200055	200053
200429	300232	300013		200058	200418
300240	300233	300231		200992	200468
600288	300234	300237		300230	300007
600300	300235	300239		300243	600291
600395	600289	600298		600290	600293
600112	600726	600390		600295	600000
600732	600730	600399		900928	600004
900901	900905	600111		900930	600729
900902	900927	600734			900903
900907	900934	900932			900904
900934	900945	900936			900926
900941					900933
					900937
					900941
					900943

由表3可知,利用基于模糊C均值聚类的股票板块划分方法对100只股票进行重新划分的结果与实际存在差别,如板块1中的000565、000571、000581、000737在现实中属于深市A股,200026、200029、200030、200037、200429在现实中属于深市B股,300240属于创业板股,而利用本研究方法则划分在同一板块中;再比如000565、000571、000581、000737、000001、000755及000419、000948、000989在现实中同属于深市A股,但在本研究的划分结果中却分别属于板块1、2、5、6。

股票板块指某些股票因具有某一共同特征而被人为地归类在一起组成的群体。现实中可按照上市公司所属行业、上市公司经营业绩、上市公司股本规模、上市公司所具有的独特概念等标准来划分,一家上市公司可以同时属于多个不同的板块,因而具备了多重身份,故本研究划分的板块和现实中主板、二板的划分有所差异。本研究划分的板块中的股票依据在样本股票网络结构中的集聚系数和路径来划分的,反映的是与样本股票网络关系的紧密程度或者说与样本股票中其他股票的关联度,板块1~6中的股票与样本股票中的其他股票的关联度依次降低。与其他股票关联度较低的股票比与其他股票关联度较高的股票在股票市场上具有更好的独立性,股票投资者可以此作为组合投资、分散风险的参考之一<sup>[9]</sup>。

### 3 结 语

通过研究中国股票市场100只样本股票在2013年1月1日至2013年8月31日的周收盘价数据,以对数收益率为基础建立股票间相关性网络,结合实际分析了100只样本股票网络的集聚系数和平均路径。与一些学者所研究的属于同一股票市场的股票网络结构对比,可得出以下结论:1)该100只股票在现实中来自不同板块,但却具有较小的平均路径,说明股票收益率之间存在相关性;不是很大的集聚系数说明不同市场的股票之间相关性不是很强。2)其他学者研究的同一股票市场的股票网络结构的小世界性大都很强,说明同一市场内股票价格波动相关性较大,而本研究中来自不同市场的股票网络结构的小世界性并不显著。

在本研究的样本股票市场的复杂关系的基础上重新划分板块,划分结果给股票投资者提供了参考。另外,除了股票投资之外,也可以通过对构成基金的股票组合建立股票网络结构,通过计算集聚系数和路径长度,判断是否具有小世界性或者小世界性是否显著,来检验基金中股票组合的合理性。当基金组合中的股票间小世界性较强且关联度过大时,可考虑重新构建股票组合以降低投资风险;而当不具有小世界性或者小世界性很弱时,表明该基金组合较合理,相对风险较小。

### 参考文献:

- [1] 韦艳华,张世英. 金融市场的相关性分析: Copula-GARCH模型及其应用[J]. 系统工程, 2004(4): 7-12.
- [2] 李子奈,潘文卿. 计量经济学[M]. 北京: 高等教育出版社, 2010.
- [3] Watts D J, Strogatz S H. Collective dynamics of 'small world' networks[J]. Nature, 1998(393): 440-442.
- [4] Lim G, Kim S, Kim J, et al. Structure of a financial cross-correlation matrix under attack[J]. Physical A-Statistical Mechanics and its Applications, 2009(388): 3851-3858.
- [5] 杨治辉,贾韩梅. 股票收益率相关性的网络结构分析[C]//第十三届中国控制会议论文集. 烟台: 中国自动化学会控制理论专业委员会, 2011: 5732-5736.
- [6] 黄玮强,庄新田,姚爽. 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(3): 94-103.
- [7] Albert R, Barabasi A L. Statistical mechanics of complex networks[J]. Reviews of Modern Physics, 2002, 74(1): 47-97.
- [8] 牟廉明. 数据挖掘中聚类方法比较研究[J]. 内江师范学院学报, 2003(2): 16-20.
- [9] Harmo Y, Masulis R W, Nag V. Correlations in price changes and volatility across international stock markets[J]. Review of Financial Studies, 1990, 3(2): 281-307.