

基于有向有限穿越可视图的时间序列伴生网络

汤胤, 易娜, 毛景慧

(暨南大学管理学院, 广东 广州 510632)

摘要: 提出两种新的可视图建网方法(有限穿越水平可视图和有向有限穿越可视图), 有限穿越水平可视图综合了水平可视图的算法简便性以及有限穿越可视图的抗噪声能力, 而有向有限穿越可视图则突破了原有算法的无向性, 考虑了时间的不可逆性. 与传统可视图的三种模型进行对比, 验证了两种算法的有效性, 最后将模型应用到股市进行了实证分析.

关键词: 时间序列; 复杂网络; 可视图

中图分类号: F713.36 文献标识码: A 文章编号: 1000-5781(2017)02-0156-07

doi: 10.13383/j.cnki.jse.2017.02.002

Derived network based on directed limited penetrable visibility graph for time series

Tang Yin, Yi Na, Mao Jinghui

(School of Management, Jinan University, Guangzhou 510632, China)

Abstract: This paper proposes two models: limited penetrable horizontal visibility graph(LPHVG) and directed limited penetrable visibility graph(DLPVG). LPHVG has both the simplicity of horizontal visibility graph(HVG) and the anti-noise performance of LPHVG. Directed limited penetrable visibility graph(DLPVG) takes into account the irreversibility of time and makes a breakthrough in the non-directional algorithms. The comparative experiments verifies the effectiveness of the two models, and the models are applied empirically to the real stock markets.

Key words: time series; complex networks; visibility graph

1 引 言

由于金融时间序列的混沌特性, 经典的数据挖掘算法对此类数据往往效果不佳. 近年来, 复杂网络应用于时间序列分析开始受到关注, 提出了一系列将时间序列映射到复杂网络的方法. 站在复杂网络这个新的研究角度, 可以借助大量复杂网络研究工具, 通过其拓扑结构特征的动态演变反推, 发现时间序列的全新特性, 从而对时间序列进行更加深刻的探索. 这些对于金融, 互联网, 生物医学等时间序列数据的分析都具有重要指导意义.

从时间序列到复杂网络的映射算法对时间序列的噪声, 随机性, 计算复杂度, 简便性和直观性等因素均有较高要求. Zhang 等^[1]提出了一种时间序列伴生网络, 网络节点对应为时间序列的周期, 而将周期内的时

收稿日期: 2015-12-13; 修订日期: 2016-06-02.

基金项目: 广东省科技发展专项资金资助项目(2016B040401003); 广东省重大科技专项资助项目(2016B010124008); 广东省科技发展专项资金资助项目(2016A010101016).

态关系作为弧, 可以观察到带噪声的周期信号与随机网络, 混沌时间序列与小世界网络和无标度网络的对应关系, 但实现方法比较复杂. 由此催生了可视图建网方法如可视图, 水平可视图, 有限穿越可视图等, 由于可视图简便直观, 成为研究时间序列一个重要分支. 其优点是伴生网络不仅继承了原时间序列的固有特征, 用于分析时间序列的动力学特性时候也比较直观.

Lacasa 等^[2,3]提出的可视图建网方法(visibility graph, VG), 经仿射变换之后仍然与初始的可视性保持一致. 其原理是网络中的每个节点对应离散时间序列数据中的每个时间点. 这些节点中, 将满足指定可视性规则的两个点连接起来, 此即为网络的边, 如图 1 所示, 在图 1(a)用 20 个直方条代表一个周期时间序列中的前 20 个数据点, 图 1(a)中直方条的高度代表每个时间点的数据值, 如果两个直方条的顶端相互可视, 这对应的两点在图 1(b)中的网络相连. 图 1(b)中的每个时点与时间序列中的数据点一一对应. 相互可视则定义为两点连接无需穿越其他数据点. 可视性准则: 若时间序列中任意两个数据点 (t_a, y_a) 和 (t_b, y_b) 相互可视, 则对任意的点 (t_c, y_c) , 满足

$$y_c > y_a + (y_b - y_a)(t_a - t_c)/(t_b - t_a), \quad \forall t_c \in (t_a, t_b). \quad (1)$$

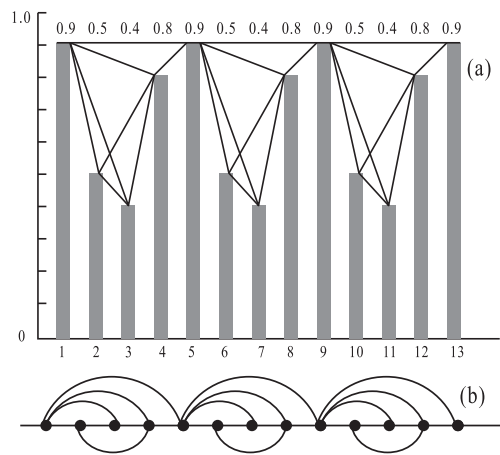


图 1 时间序列可视图建网

Fig. 1 Visibility graph for time series

基于 VG 模型, Luque 等^[4]提出了水平可视图建网模型(horizontal visibility graph, HVG), 几何上更简单统计量更少, 降低了复杂度, 如图 2 所示.

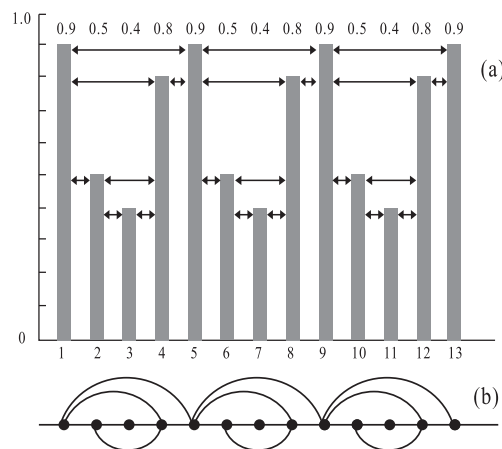


图 2 水平可视图建网

Fig. 2 Horizontal visibility graphs

HVG 模型中, 节点的定义与 VG 建网方法相同, 不同在于连边. 若两个直方条之间能画出一条不穿过其

它直方条的水平视线,则认为这两点在网络中是相连的. HVG 具有 VG 网络的相同性质,可用于区分混沌时间序列和随机时间序列,而且任意随机时间序列经 HVG 转化之后的伴生网络度分布都具有相同的幂律形式.

周婷婷等^[5]在 VG 模型基础上考虑可视范围的穿越概念,引入了有限穿越视距,提出了有限穿越可视图建网模型(LPVG, limited penetrable visibility graph). 算法仍基于 VG 可视性准则,其原理是定义有限穿越视距,网络中的两个节点如果可视,当且仅当网络中两节点间的可视线被位于它们之间的其他节点截断的次数 $n \leq N$. 如图 3(a) 所示,实线表示基于 VG 方法的水平连接线,虚线表示引入有限穿越视距新增的连线. 图3(b)中,实点上方为基于 VG 方法的水平连接线,实点下方表示新增的连线.

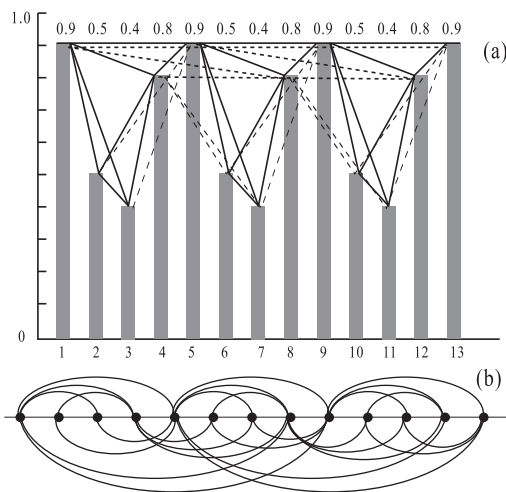


图3 有限穿越可视图建网

Fig. 3 Limited penetrable visibility graph

水平可视图建网算法简单高效,都能继承时间序列的内在形态特征,有限穿越可视图建网方法抗噪声能力强,但是这三种方法在实际应用中仍存在不足:可视图与水平可视图的抗噪声能力较弱,而有限穿越可视图的算法较复杂.更重要的是这三种建网方法都没有考虑时间序列的不可逆性,生成的无向网络显然不符合实践场景,这会进一步的深入研究带来问题.本文思路是在 LPVG 基础上改用水平视距构造模型,从而可以大幅降低计算复杂性.进一步地,考虑时间序列的不可逆性,使用有向可视图映射时间序列,可以更加准确地重现实践场景.本文首先讨论国内外可视图对时间序列建模的研究现状,然后针对当前研究的若干问题,提出一种改进的可视图建网方法有限穿越水平可视图(LPHVG),并进一步构造有向有限穿越可视图(DLPVG).算法的优点在于考虑了时间序列的不可逆性特征,既能够有效地提取出噪声时间序列,分形时间序列以及周期时间序列的复杂网络特征,又比已有的三种网络算法的抗噪能力更显著. LPHVG 和 DLPVG 两个模型的性能表现与前三个模型的比较也证实了这一点.为进一步论证,本文还给出了模型应用于股市实际数据的示例结果.

2 时间序列可视图复杂网络模型

2.1 有限穿越水平可视图复杂网络(LPHVG)

LPHVG 模型的基本思想是基于 VG 的可视性准则,定义有限穿越视距 N ,在网络中两个节点在如下情况下视为相连:两个直方条中间可以依据可视性规则画出一条水平连线,且两直方条间的水平连线被位于它们之间的其他直方条截断次数 $n \leq N$. 图 4 即是令 $N = 1$ 的一个 LPHVG 的建网例子.

可视性准则 $Y = \{y_i\}_{i=1,2,\dots,n}$ 是一个时间序列,这个序列包含 n 个序列点, N 是有限穿越可视距,如果在离散时间序列中随意取两个点 (t_a, y_a) 和 (t_b, y_b) ,这两个点相隔 m 个数据点而且相互可视,那么这两

个点之间存在 K ($0 \leq K \leq N$) 个数据点 (t_i, y_i) , 满足

$$y_i > y_a \text{ 或 } y_i > y_b, \quad \forall t_i \in (t_a, t_b), \quad (2)$$

其余 $m - K$ 个数据点 (t_j, y_j) , 满足

$$y_j < \min(y_a, y_b), \quad \forall t_j \in (t_a, t_b), \quad (3)$$

LPHVG 本质上与 LPVG 的算法相同, 但是具有较简单的几何可视准则, 相比于 HVG 又具有更广泛的可视性.

2.2 有向有限穿越可视图复杂网络 (DLPVG)

前面四种模型都没有考虑网络中边的方向性, 只能表达无向网络信息. 实际中许多复杂网络都是有向网络, 如电力系统网络, 食物链网络, 神经网络以及货物运输网络等. 基于 LPVG, DLPVG 算法则考虑了有向图的特性.

定义有限穿越视距 N , 即 $Y = \{y_i\}_{i=1,2,\dots,n}$ 为一个具有 n 个数据的时间序列, N 为有限穿越视距, 如果在离散时间序列中相隔 m 个数据点的两个点 (t_a, y_a) 和 (t_b, y_b) 相互可视, 那么这两个点之间存在 K ($0 \leq K \leq N$) 个数据点 (t_i, y_i) , 满足

$$y_i < y_a + (y_a - y_b)(t_a - t_i)/(t_b - t_a), \quad \forall t_i \in (t_a, t_b), \quad (4)$$

其余 $m - K$ 个点 (t_j, y_j) , 满足

$$y_j > y_a + (y_a - y_b)(t_a - t_j)/(t_b - t_a), \quad \forall t_j \in (t_a, t_b). \quad (5)$$

需要注意的是 DLPVG 和 LPVG 间的区别在于 DLPVG 的可视线具有方向性, 且度的计算方法以及度的分布有所不同. DLPVG 的度要分出度与入度分别计算, 且出度与入度的度分布在许多实际复杂有向网络中呈幂律分布^[6]. 图 5 即是在 LPVG 的基础上考虑有向图得到的 DLPVG 可视图建网. 采用 DLPVG 算法对时间序列进行建网, 能够抵抗一定的噪声, 另外结合入度分布和出度分布可以对时间序列的不可逆性进行进一步的研究和证明.

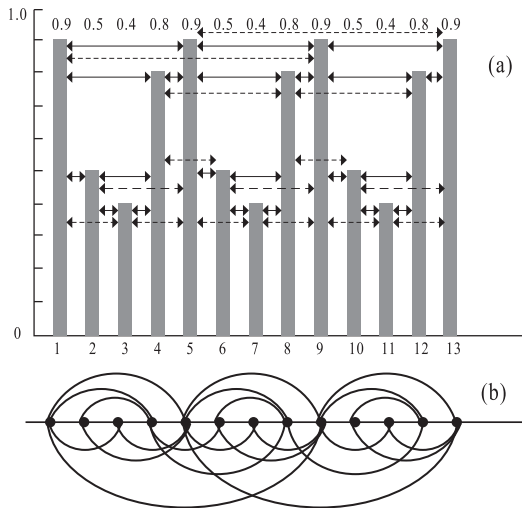


图 4 有限穿越水平可视图建网

Fig. 4 Limited penetrable horizontal visibility graph

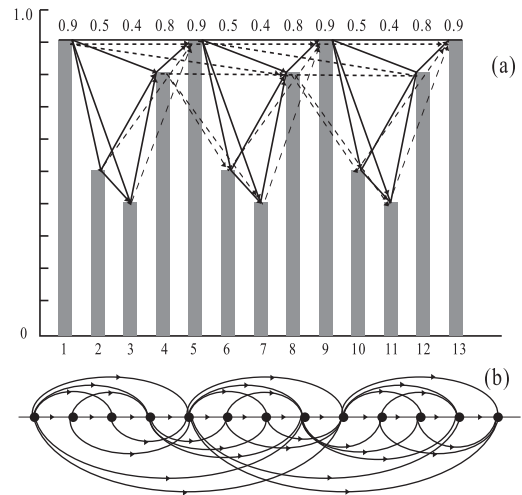


图 5 有向有限穿越可视图建网

Fig. 5 Directed limited penetrable visibility graph

3 可视化模型评价比较

本节通过计算网络度分布, 并用图形表现每种可视化方法的邻接矩阵, 对这五种方法进行初评价. 进而

运用这些算法分析周期时间序列, 和随机时间序列, 从识别不同信号的能力对比几种方法的优劣势.

3.1 邻接矩阵对比

从图6中的邻接矩阵图(其中 N 为节点编号)可发现, 由LPHVG算法构造的网络节点平均度值明显比VG和HVG都大, 只是比LPVG的平均度值略小, 这是基于VG和HVG的差别而得来的: 在网络连接性方面LPHVG比VG和HVG更紧密, 只是比LPVG会略小. 一般来说, 如果网络中一个节点的度越大就说明这个点在这个网络中越重要, 连接性越紧密也说明了该算法更好地继承了原有时间序列的一些典型特征. 因此可以推断出LPHVG在反应时间序列特征方面优于VG与HVG算法. 同时由于几何上只考虑水平视线, LPHVG算法复杂程度低于LPVG算法.

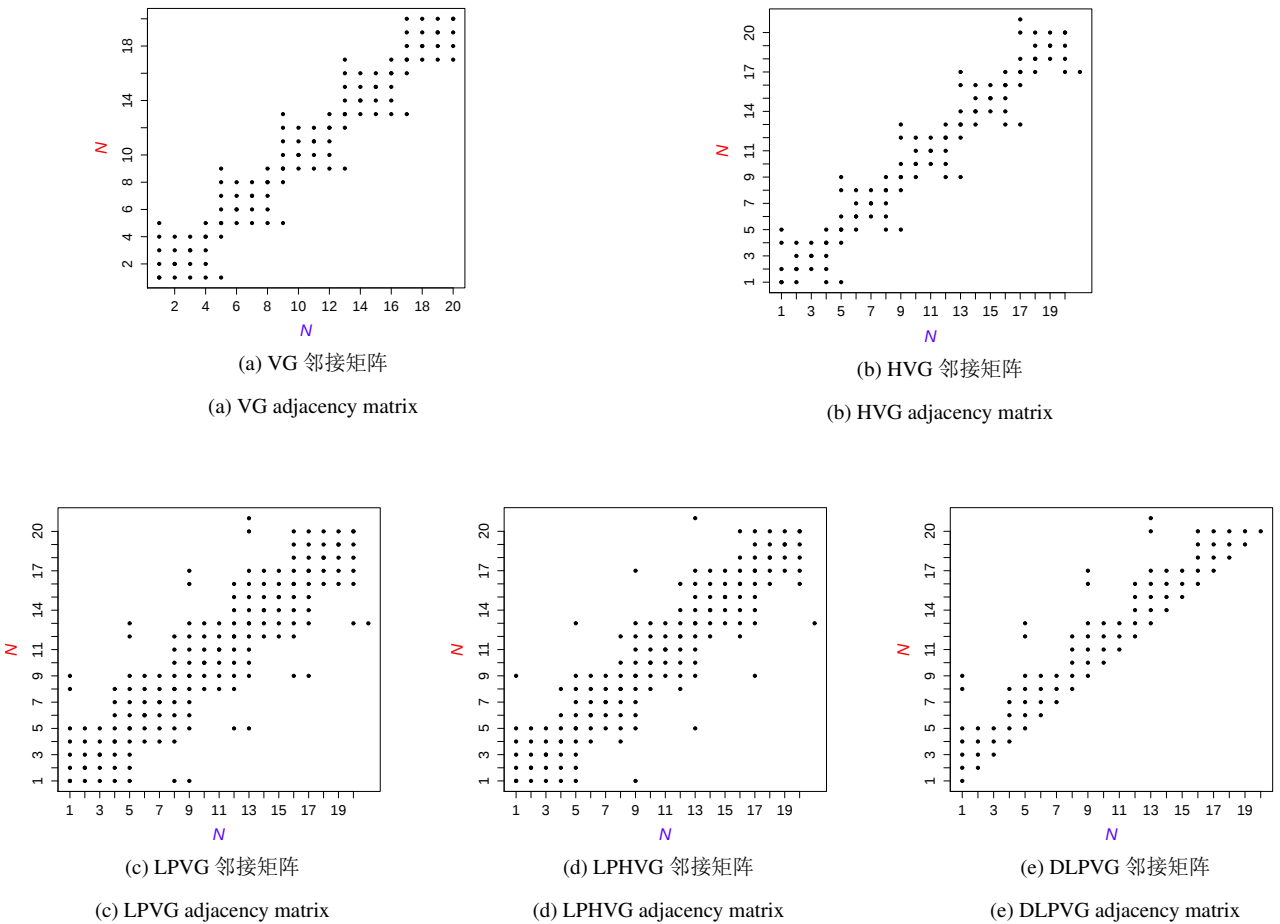


图6 邻接矩阵对比

Fig. 6 Comparison of adjacent matrix

复杂网络的邻接矩阵涵盖了网络的所有特性, 前面四种无向网络图都有一个对称的邻接矩阵, 而DLPVG邻接矩阵是一个非对称矩阵. 在网络连接紧密性和节点的度方面DLPVG与LPVG有着相似的特征值, 但出度和入度需分开计算.

3.2 周期时间序列评价

本文中采取的周期时间序列为

1) 正弦信号 $y = \sin(5\pi x)$, 采样间隔为 $\Delta x = 0.01$, 序列长度为 $n = 1\,000$;

2) 正弦信号 $y = \sin(2\pi x)$, 采样间隔为 $\Delta x = 0.01$, 序列长度为 $n = 1\,000$. Lacasa 等^[2]指出周期性时间序列由VG映射成复杂网络后, 变成规则网络; 随机时间序列则与随机网络相对应, 图7为这两种周期时间

序列由 LPHVG 与 DLPVG 算法转化成复杂网络之后的度分布图(其中 k 为节点度的取值, $p(k)$ 为节点度值为 k 的节点个数与网络总节点的比率). 经 LPHVG 转化之后, 时间序列网络度的值都集中在 4 这个点, 这属于一个规则网络. 而由 DLPVG 周期时间序列的度分布可看出, 不管是有向还是无向其度分布都有一个尖峰, 因此 DLPVG 也能识别出周期序列.

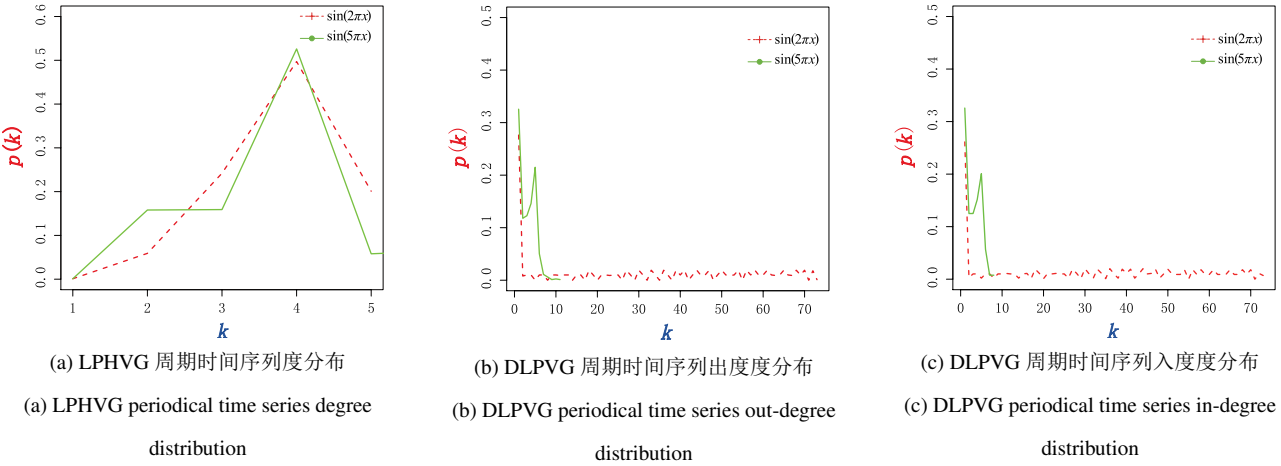


图 7 周期时间序列度分布

Fig. 7 Periodical time series degree distribution

表 1 DLPVG 周期时间序列入度的度值分布表($T = 5\pi$)

Table 1 Periodical time series in-degree distribution by DLPVG ($T = 5\pi$)

k	1	2	3	4	5	6	7	8
in-degree	86	33	33	40	53	15	2	2

表 2 DLPVG 周期时间序列出度的度值分布表

Table 2 Periodical time series out-degree distribution by DLPVG($T = 5\pi$)

k	1	2	3	4	5	6	7	8	9	10	11
out-degree	265	96	100	119	175	41	9	5	1	2	1

对比 DLPVG 算法的出度和入度分布, 得出的结论有: DLPVG 周期时间序列出入度分布的形状基本类似, 但经过 DLPVG 映射后各节点的出度的度值会比入度的度值稍微大一些. 表现在图中, 即周期为 5π 的时间序列, 出度分布在 $k = 5$ 左右第二个尖峰的时候概率比入度的大, 而且出度分布会延伸到 $k = 10$ 以后, 入度分布中最大的度取值则不超过 10. 针对有向图, 表1给出了更具体的实验数值. 出入度反映了时间序列的不可逆性, 具有先后性这一性质. 说明伴生网络纳入有向性后, 出入度及其分布的不同也证实了本文研究在理论方面的贡献.

4 模型应用于股市

股票价格是随机时间序列数据, 并且常含有白噪声. 本节通过 LPHVG 与 DLPVG 模型对股票时间序列的映射提供一个新的角度进行解读. 选取民生银行 2011-01-04—2013-05-10 的股价数据, 图 8 是这 5 种模型分析个股日交易数据得出的度分布函数图形.

图 8 进一步验证了之前对五种模型的评价结论. 经转化的随机网络模型应呈不规则的多峰分布特征, 由上图可看出, 对于有白噪声干扰的随机时间序列, LPVG, LPHVG 的识别能力较强, 抗噪声能力优于 VG 与 HVG. 此外, 通过 DLPVG 模型分析, 经转化的随机网络模型的出入度都呈现出幂律分布的特征, 另外节

点的出度值一般会比入度值稍微大些,而且网络度分布也较分散(出度分布会延伸到 $k = 40$ 以后,入度分布中最大的度取值是不超过 40 的),这跟时间序列的不可逆性也有很大关系,对于生活中很多时间序列的研究有着重要的启发意义.

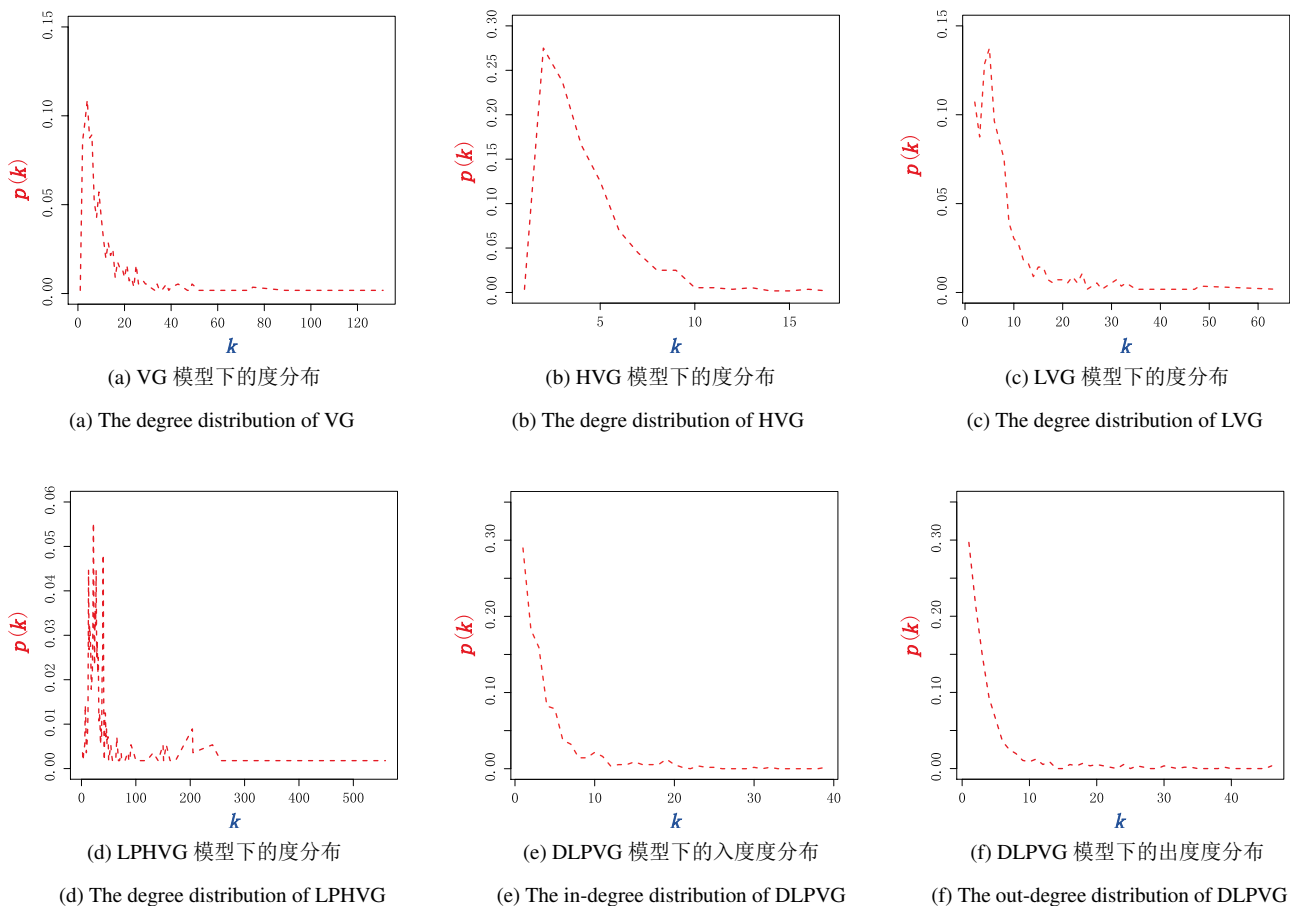


图 8 个股股价伴生网络度分布

Fig. 8 Stock price time series degree distribution of derived network

5 结束语

本文在经典可视图网络模型基础上,提出 LPHVG 和 DLPVG 两种新的模型. LPHVG 综合 LPVG 的抗噪声性以及 HVG 的简便性,实验结果表明 LPHVG 算法的伴生复杂网络比 HVG 能够更好地反应时间序列的特征,同时算法也比 LPHVG 算法简单一些. 基于此提出 DLPVG,解决了对时间序列的不可逆性问题,提供了更直观优化的伴生网络转换方法.

本文只将时间序列的不可逆性用到 LPVG,对于有向网络的每个点都有入度和出度值,后续可以对出入度以及点与点之间的出入度 in-in, out-out, in-out, out-in 的关联性进行研究. 另外,将 VG, HVG, LPHVG 模型引入有向性,将有助于各个学科的时间序列研究.

参考文献:

- [1] Zhang J, Small M. Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters*, 2006, 96(23): 238–701.