

文章编号: 1000-5641(2017)02-0075-06

可视图复杂网络度分布拟合比较研究

张 蓉, 邹 勇

(华东师范大学 物理与材料科学学院, 上海 200241)

摘要: 可视图 (Visibility Graph, VG) 算法为研究时间序列的动力学特性提供了复杂网络的思想. 网络的度分布反映了时间序列的动力学特征. 通过自回归随机过程和分数布朗运动两种不同数据, 分别构建可视图. 对比结果表明, 在自回归随机过程中, 度分布可以用指数函数刻画; 而在分数布朗运动中, 度分布用幂律函数刻画更为合适. 这一结论不但适用于 VG 算法, 同时也适用于水平可视图 (Horizontal Visibility Graph, HVG) 算法.

关键词: 可视图; 自回归随机过程; 分数布朗运动; 拟合优度

中图分类号: O415 **文献标志码:** A **DOI:** 10.3969/j.issn.1000-5641.2017.02.010

Comparative regression analysis to degree distributions of visibility graph

ZHANG Rong, ZOU Yong

(School of Physics and Materials Science, East China Normal University,
Shanghai 200241, China)

Abstract: Visibility graph has provided much insight to study the dynamics of time series from the perspective complex network. We construct visibility graphs for time series from both auto-regressive stochastic and fractional Brownian motions. Our results suggest that degree distributions of the resulted complex networks of auto-regressive processes are characterized by exponential forms, while that of fractional Brownian motions obey power-law forms. Our conclusions hold for both the traditional visibility graph and its variant horizontal visibility graph.

Key words: visibility graph; autoregressive stochastic process; fractional Brownian motion; goodness of fit

0 引 言

复杂网络的理论及其应用自从 Watts 和 Strogatz 提出小世界网络以及 Albert 和 Barabási 提出无标度网络的概念以来, 便吸引了众多研究者的广泛兴趣并得到了快速发展^[1-4]. 近

收稿日期: 2016-03-18

基金项目: 国家自然科学基金 (11305062)

第一作者: 张 蓉, 女, 硕士研究生, 研究方向为非线性与复杂网络学.

E-mail: sandy58756071@sina.com.

通信作者: 邹 勇, 男, 副教授, 硕士生导师, 研究方向为复杂网络结构与动力学.

E-mail: yzou@phy.ecnu.edu.cn.

年来,有文献提出运用复杂网络的方法研究时间序列的动力学特性,并且取得了重要进展^[5-8],成为最近非线性时间序列研究的新热点.现在运用比较多的复杂网络方法有:Zhang和Small^[5]在2006年提出的周期图的网络方法;Lacasa等^[6]在2008年提出的可视图方法;Donner等^[7]在2010年提出的回归网络方法;Xu等^[8]在2008年提出的类似回归图的方法.运用这些不同的网络映射算法(即将数据映射为网络)时,主要是通过分析网络的平均最短距离、平均聚类系数以及度分布等刻画时间序列的网络特性,来展示传统时间序列分析方法得不到的结果.这些方法建立了一个跨越时间序列和复杂网络的新桥梁,将时间序列转化为形象的网络拓扑结构.

本文延用L. Lacasa提出的可视图(VG)算法. VG算法将时间序列中的数据直接定义为复杂网络中的节点,而节点间的连接关系由数据之间的线性可视关系决定.基于可视图方法,随着时间序列数据的不断增加,复杂网络的生成过程类似于Barabasi-Albert无尺度网络的动态生成过程,而且复杂网络中的中心(Hub)节点对应于时间序列数值特别大的数据. VG算法能够将任意的时间序列转化为网络,且为全连通网络,不依赖于阈值的选取,如周期时间序列能够转化成正则图、随机时间序列能够转化成随机图、分形时间序列能够转化成无标度网络.该方法已经在很多不同的领域内得到了很好的应用,比如太阳黑子方面^[9-10]、金融时间序列和飓风气候数据^[11-15].最近,又提出了水平可视图(HVG)算法^[16],同样得到了非常广泛的应用^[17-18].

VG算法可将某些分形时间序列转化为无标度网络(例如布朗运动序列),即网络的度分布符合幂律分布^[6].从数值角度看,计算网络的幂率指数不是一个简单的事情,在文献[6]中也并没有证明幂律分布是最优的度分布;另外,针对更加一般的随机时间序列,刻画幂率分布是否可行,以及在HVG算法中是否同样适用,目前尚不明了.进一步刻画网络度分布的拟合问题,无疑有助于理解网络化方法分析时间序列的基本问题.

本文首先简单介绍VG算法和HVG算法;其次,介绍本文的研究对象,即自回归随机过程(自回归随机时间序列)和分数布朗运动;最后,通过这两种可视化算法,把这两类数据转换为可视图,对比分析后发现,自回归随机过程的网络度分布符合指数分布,而分数布朗运动的网络度分布近似幂律分布.

1 可视图网络生成算法

VG算法的定义是:当给定的一组时间序列中,任意两点 (t_a, y_a) 和 (t_b, y_b) 对于两点间的任何点 (t_c, y_c) 都满足条件 $(t_a < t_c < t_b)$

$$y_c < y_a + (y_b - y_a) \frac{t_c - t_a}{t_b - t_a}$$

时,则认为这两点可视,即在网络中存在连边.现有结果表明可视图保留了原始数据的很多性质,比如周期时间序列得到规则图、随机时间序列转化为随机网络图、分形序列对应标度网络^[6]等.

HVG算法的定义是:如果任意两点 (t_a, y_a) 和 (t_b, y_b) 之间,对于任何点 (t_c, y_c) 都满足条件 $(t_a < t_c < t_b)$

$$y_a, y_b > y_c$$

时,则认为这两点可视,即在网络中相连.

2 两种时间序列介绍

2.1 自回归随机过程模型

自回归随机过程模型(简称为自回归模型)是利用前期若干时刻的随机变量的线性组合来描述以后某时刻随机变量的线性回归模型,它是时间序列中的一种常见形式.考虑一个时间序列 $x_1, x_2, x_3, \dots, x_n$, p 阶自回归模型(AR(p))表明序列中 x_t 是前 p 个时刻的线性组合及误差项的函数.本文仅研究AR(1)过程的序列,其一般形式的数学模型为 $x_t = \varphi_1 x_{t-1} + \epsilon_t$, 其中 φ_1 是模型参量, ϵ_t 是白噪声,其均值为0,方差为 σ ,且本文主要以 $\varphi_1 = (0, 0.3, 0.9, -0.5)$ 为例说明.

2.2 分数布朗运动

布朗运动过程是一种正态分布的独立增量连续随机过程,而分数布朗运动(Fractional Brownian Motion, FBM)的增量不再是独立的,这是对经典布朗运动的推广,同时也是理想的不规则扩散和分形随机行走的基础.分数布朗运动的赫斯特指数 H ($0 < H < 1$),当 $H = 1/2$ 时,运动是完全随机的;当 $H \neq 1/2$ 时,运动不是完全随机的,具有长程相关性.本文取不同的指数 H ($H = 0.3, 0.5, 0.7, 0.9$),利用Matlab7.0,应用快速傅里叶变换(Fast Fourier Transform, FFT)求解循环协方差矩阵,生成分数布朗运动的时间序列.本文对每一个 φ_1 和指数 H ,进行50次随机实验,每个序列长度为 $N = 5\,000$.

3 VG算法和HVG算法基于两种时间序列的应用

3.1 VG算法基于AR(1)的应用

度分布是复杂网络中一个重要的概念.如果遵循指数分布时,即网络的度分布为 $p(k) = e^{-\lambda k}$, k 表示度, $p(k)$ 是对应度 k 的概率大小, λ ($\lambda > 0$) 被称作率参数.通过拟合 $\ln(p(k))$ 对 k 的斜率得到率参数 λ .如果遵循幂律分布时,即有 $p(k) = k^{-\gamma}$, γ 被称作幂律指数,通过拟合 $\log_{10}(p(k))$ 对 $\log_{10}(k)$ 的斜率得到幂律 γ ,统计上,通常用 $p(k)$ 的累积分布 $F(k)$ 来拟合得到 γ .首先生成AR(1)过程在不同参数 φ_1 下的时间序列,然后构建VG网络.运用指数和幂率分别拟合度分布时的结果如图1和表1所示.

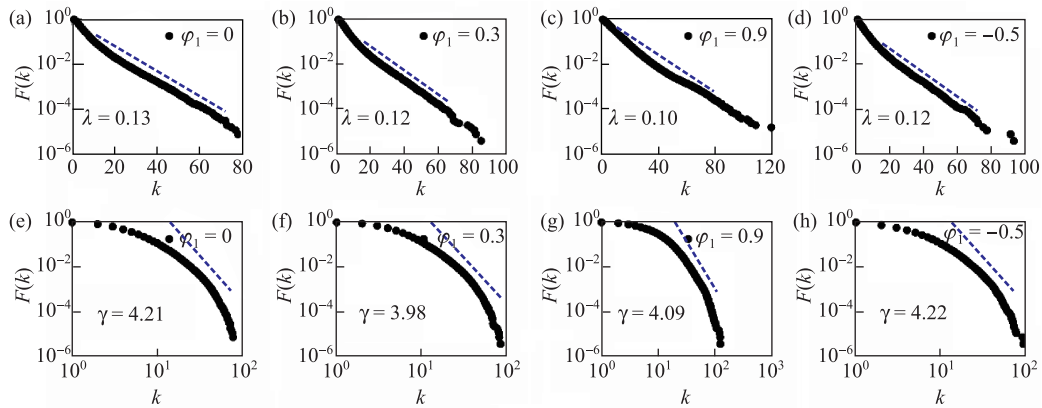


图1 图1(a)–图1(d) 半对数坐标下的度累积分布的指数拟合,虚线表示率参数 λ ;

图1(e)–图1(h) 双对数坐标下的度累积分布的幂率拟合,虚线表示幂律指数 γ

Fig. 1 (a)-(d) Exponential fitting to the cumulative degree distributions in semi-log plot, and dashed lines are slopes for λ ; (e)-(h) Similar to (a)-(d) but power law forms in double log plots

度分布斜线拟合的优劣运用相关系数平方 R^2 以及残差平方和SSE(Sum of Squares for Error)(误差项离差平方和)来刻画.当SSE越小, R^2 越接近1时,表示相关的拟合数值参考价

值越高, 拟合越优. 结合表 1 的结果, 得到结论: 自回归随机过程时间序列生成的 VG 度分布更接近指数形式. 类似, 在 HVG 下的结果如图 2 和表 2 所示, 得到相同的结论, 即自回归随机过程的时间序列生成的 HVG 可以用指数分布刻画.

表 1 AR(1) 在 VG 算法下指数与幂率拟合比较

Tab. 1 Comparison between exponential and power law scaling factors of VG graphs of AR(1)

	$\varphi_1 = 0$		$\varphi_1 = 0.3$		$\varphi_1 = 0.9$		$\varphi_1 = -0.5$	
	λ	γ	λ	γ	λ	γ	λ	γ
数值	0.13	4.21	0.12	3.98	0.10	4.09	0.12	4.22
R^2	0.998	0.957	0.999	0.919	0.986	0.953	0.998	0.959
SSE	0.655	2.575	0.123	9.074	3.829	5.290	0.440	2.458

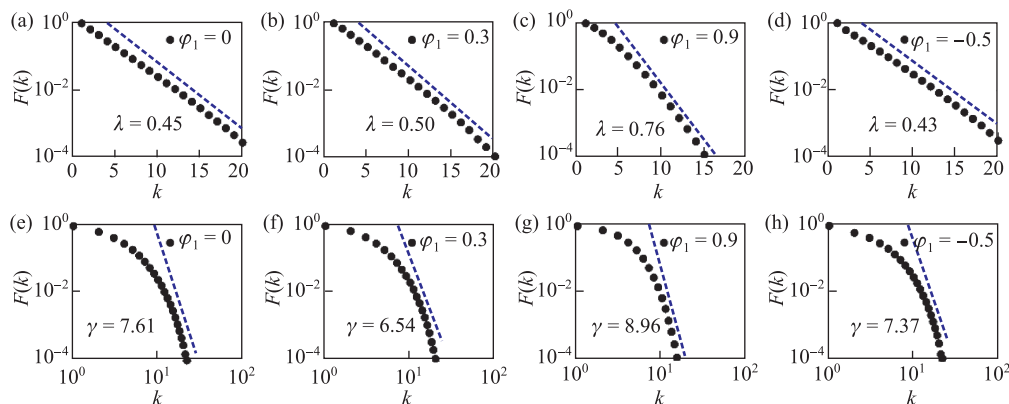


图 2 在 HVG 情形下结果, 其余信息如图 1 标注

Fig. 2 The same as Fig. 1, except for HVG

表 2 AR(1) 过程在 HVG 下的拟合比较, 其余信息如表 1 标注

Tab. 2 The same as Tab.1, except for HVG of AR(1)

	$\varphi_1 = 0$		$\varphi_1 = 0.3$		$\varphi_1 = 0.9$		$\varphi_1 = -0.5$	
	λ	γ	λ	γ	λ	γ	λ	γ
数值	0.45	7.61	0.50	6.54	0.76	8.96	0.43	7.37
R^2	0.995	0.947	0.993	0.911	0.994	0.955	0.996	0.952
SSE	1.515	1.599	1.839	3.273	0.951	1.227	0.6017	1.199

3.2 VG 算法基于 FBM 的应用

在 VG 算法下, 针对选取的不同参数 H 时, 得到的 VG 的度分布及拟合结果分别由图 3 和表 3 所示, 得到的结论是: 度分布更近似幂律分布.

表 3 FBM 在 VG 算法下指数与幂率拟合比较

Tab. 3 Comparison between exponential and power law scaling factors VG graphs of FBM

	$H = 0.3$		$H = 0.5$		$H = 0.7$		$H = 0.9$	
	λ	γ	λ	γ	λ	γ	λ	γ
数值	0.07	2.47	0.03	1.97	0.01	1.56	0.004	1.12
R^2	0.989	0.990	0.965	0.995	0.939	0.990	0.962	0.986
SSE	0.527	0.121	6.826	0.105	15.520	0.429	11.770	0.452

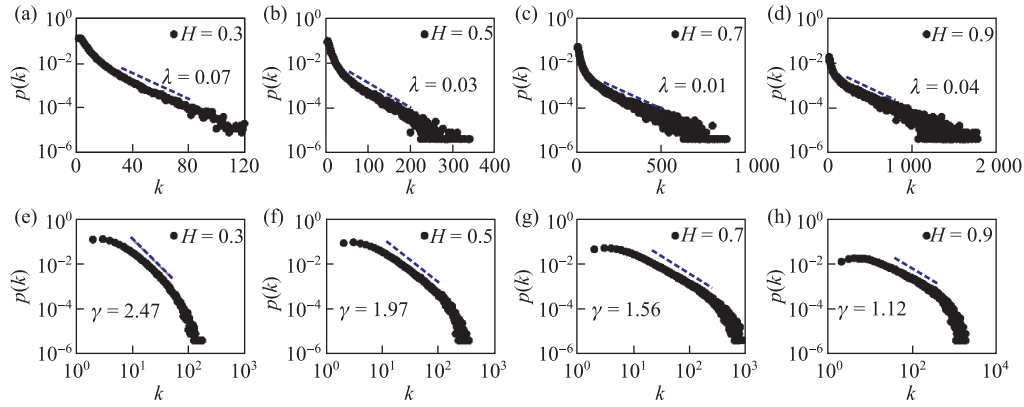


图 3 图 3(a)–图 3(b) 半对数坐标下的度分布的指数拟合, 虚线表示率参数 λ ; 图 3(e)–图 3(h) 双对数坐标下的度分布的幂率拟合, 虚线表示幂律指数 γ

Fig. 3 (a)-(d) Exponential fitting to the degree distributions in semi-log plot, and dashed lines are slopes for λ ; (e)-(h) Similar to (a)-(d) but power law forms in double-log plots

在 HVG 下, 可以得到相同的结论, 具体而言, 度分布和拟合结果如图 4 和表 4 所示.

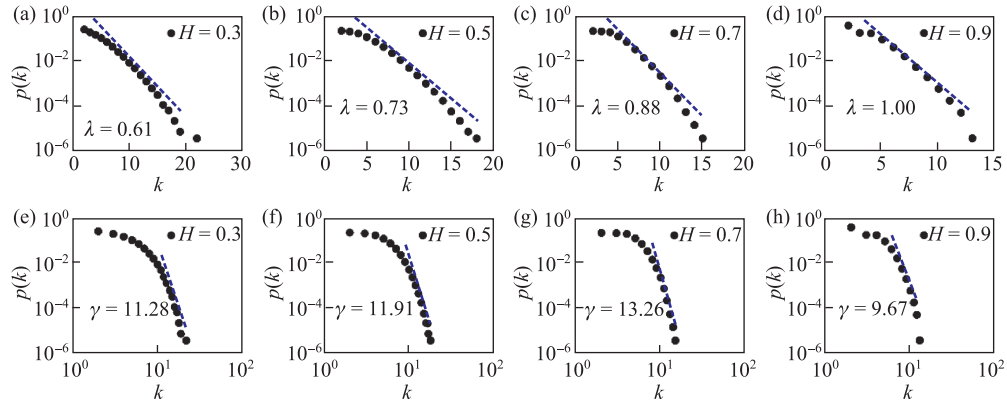


图 4 在 HVG 情形下结果, 其余信息如图 3 标注

Fig. 4 The same as Fig. 3, except for HVG

表 4 FBM 过程在 HVG 下的拟合比较, 其余信息如表 3 标注

Tab. 4 The same as Tab.3, except for HVG of FBM

	$H = 0.3$		$H = 0.5$		$H = 0.7$		$H = 0.9$	
	λ	γ	λ	γ	λ	γ	λ	γ
数值	0.61	11.28	0.73	11.91	0.88	13.26	1.00	9.67
R^2	0.975	0.980	0.970	0.980	0.948	0.971	0.953	0.979
SSE	4.689	0.210	6.815	0.274	9.658	0.328	6.999	0.143

4 结 论

本文将可视图方法应用在自回归随机过程和分数布朗运动中对比研究, 通过对指数拟合和幂律拟合的比较, 不论是在 VG 算法下, 还是在 HVG 算法下, 可得到以下结论.

(1) 对于自回归随机过程, 相较于幂律拟合, 指数拟合的残差平方和 SSE 更小, 相关系数平方 R^2 更接近 1 时, 拟合更优, 因此自回归时间序列可视图的度分布更接近指数分布.

(2) 对于分数布朗运动时间序列, 相较于指数拟合, 幂律拟合的 SSE 更小, R^2 更接近 1 时, 拟合更优, 因此分数布朗运动时间序列可视图的度分布更接近幂律分布.

传统时间序列分析的优劣性值得探讨, 尤其需要重视与回归网络、周期图等方法比较研究; 在气候、金融和大脑等实际数据中的应用还有待于进一步研究.

[参 考 文 献]

- [1] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks [J]. *Nature*, 1998, 6684(393): 440-442.
- [2] ALBERT R, BARABÁSI A L. Statistical mechanics of complex networks [J]. *Review of Modern Physics*, 2002, 74(1): 47-97.
- [3] 汪小帆. 复杂网络理论及其应用 [M]. 北京: 清华大学出版社, 2006.
- [4] 何大彻. 复杂系统与复杂网络 [M]. 北京: 高等教育出版社, 2009.
- [5] ZHANG J, SMALL M. Complex network from pseudoperiodic time series: topology versus dynamics [J]. *Physical Review Letters*, 2006, 96(23): 238701-238704.
- [6] LACASA L, LUQUE B, BALLESTEROS F, et al. From time series to complex networks: The visibility graph [J]. *Proceedings of the National Academy of Sciences*, 2008, 105(13): 4972-4975.
- [7] DONNER R V, ZOU Y, DONGES J F, et al. Recurrence networks-a novel paradigm for nonlinear time series analysis [J]. *New Journal of Physics*, 2010, 12(2): 129-132.
- [8] XU X K, ZHANG J, SMALL M. Superfamily phenomena and motifs of networks induced from time series [J]. *Proceedings of the National Academy of Sciences*, 2008, 105(50): 19601-19605.
- [9] ZOU Y, SMALL M, LIU Z H, et al. Complex network approach to characterize the statistical features of the sunspot series [J]. *New Journal of Physics*, 2014, 14(1): 1-8.
- [10] ZOU Y, DONNER R V, MARWAN N, et al. Long-term changes in the north-south asymmetry of solar activity: A nonlinear dynamics characterization using visibility graphs [J]. *Nonlinear Processes in Geophysics*, 2014, 21(1): 1113-1126.
- [11] YANG Y, YANG H. Complex network-based time series analysis [J]. *Physica A*, 2008, 387(s 5-6): 1381-1386.
- [12] LACASA L, TORAL R. Description of stochastic and chaotic series using visibility graphs [J]. *Physical Review E*, 2010, 82(3): 4881-4888.
- [13] GAO Z K, JIN N D. Flow-pattern identification and nonlinear dynamics of gas-liquid two-phase flow in complex networks [J]. *Physical Review E*, 2009, 79(6): 1019-1027.
- [14] YANG Y, WANG J, YANG H, et al. Visibility graph approach to exchange rate series [J]. *Physica A*, 2009, 388(20): 4431-4437.
- [15] ELSNER J B, JAGGER T H, FOGARTY E A. Visibility network of United States hurricanes [J]. *Geophysical Research Letters*, 2009, 36(16): 554-570.
- [16] LUQUE B, LACASA L, BALLESTEROS F, et al. Horizontal visibility graphs: exact results for random time series [J]. *Physical Review E*, 2009, 80(2): 593-598.
- [17] GUTIN G, MANSOUR T, SEVERINI S. A characterization of horizontal visibility graphs and combinatorics on words [J]. *Physica A*, 2011, 390(12): 2421-2428.
- [18] NUÑEN A, LACASA L, VALERO E, et al. Detecting series periodicity with horizontal visibility graphs [J]. *International Journal of Bifurcation & Chaos*, 2012, 22(7): 1250160(10pages). DOI: 10.1142/S021812741250160X.

(责任编辑: 李 艺)