Superfamily classification of nonstationary time series based on DFA scaling exponents

Chuang Liu\(^1,2\) and Wei-Xing Zhou\(^1,2,3,4\)

\(^1\) School of Business, East China University of Science and Technology, Shanghai 200237, People’s Republic of China
\(^2\) Engineering Research Center of Process Systems Engineering, Ministry of Education, East China University of Science and Technology, Shanghai 200237, People’s Republic of China
\(^3\) School of Science, East China University of Science and Technology, Shanghai 200237, People’s Republic of China
\(^4\) Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, People’s Republic of China

E-mail: wxzhou@ecust.edu.cn

Received 11 March 2010, in final form 10 August 2010
Published 18 November 2010
Online at stacks.iop.org/JPhysA/43/495005

Abstract

The superfamily phenomenon of time series with different dynamics can be characterized by the motif-rank patterns observed in the nearest-neighbor networks of the time series in phase space. However, the determinants of superfamily classification are unclear. We attack this problem by studying the influence of linear temporal correlations and multifractality using fractional Brownian motions and multifractal random walks. Numerical investigations unveil that the classification of a superfamily phenomenon is uniquely determined by the detrended fluctuation analysis (DFA) scaling exponent \(\alpha\) of the time series. Only four motif patterns are observed in the simulated data, which are delimited by three DFA scaling exponents \(\alpha \approx 0.25\), \(\alpha \approx 0.35\) and \(\alpha \approx 0.45\). The validity of the result is confirmed by stock market indices and turbulence velocity signals.

PACS numbers: 89.75.Hc, 05.45.Tp, 89.65.Gh, 94.05.Lk

1. Introduction

In the last few decades, the understanding of the dynamics and the topological structures of complex systems has experienced extremely rapid progress. The behavior of a complex system is usually recorded and exhibited in the form of time series. New light has been shed on these studies from the perspective of complex networks \([1–3]\), which considers the interactions of different constituents of a complex system through a portfolio of time series. More recently, a wealth of methods have been proposed to construct complex networks from single time series.
[4–15], which enables us to study the dynamics of a complex system through the topological structure of the constructed network. These methods have been applied to investigate diverse natural and social systems [16–24], among which the visibility graph algorithm has attracted most attention [10, 11, 19–24].

Many complex biological, technological and social networks share similar universal statistical properties, such as the small-world characteristics [25, 26] and the scale-free distribution of node degrees [27]. The global properties can be used to classify networks into different universality classes from a macroscopic point of view [26]. On the other hand, complex networks with the same global statistics may exhibit different local structure properties. At the microscopic level, the building blocks of complex networks are subgraphs or motifs [28], and the network motif patterns of occurrence can be utilized to define superfamilies of networks [29]. The superfamily phenomenon is also observed in time series, which is able to distinguish different types of dynamics in periodic, chaotic and noisy processes based on the occurrence frequency patterns of network motifs [9].

In this work, we attempt to seek possible determinants that cause different motif-rank patterns emerging in complex nearest-neighbor networks constructed from a specific time series whose increments are stationary. Numerical investigations conducted on fractional Brownian motions (FBMs) and multifractal random walks (MRWs) uncover that the superfamily phenomenon is uniquely determined by the detrended fluctuation analysis (DFA) scaling exponent and the multifractality in MRWs has negligible impact. We also study the superfamily phenomenon using several stock indices of developed and emerging markets and the velocity signals of three-dimensional fully developed turbulent flows, and find that the above numerical conclusion holds as well.

2. Methods

2.1. Nearest-neighbor networks and motifs

The procedure to construct a nearest-neighbor network from a time series is described as follows [9]. The time series is embedded in a proper phase space, in which the time delay $\tau$ is calculated according to the mutual information method and the embedding dimension $d$ should be larger than $2D + 1$ to make sure that the attractors are unfolded, where $D$ is the correlation dimension of the system. To determine the time delay, we calculate the mutual information for different delays ranging from 1 to 100 and the time delay with minimal mutual information is selected [30]. The selected time decay differs from one time series to another. For a time series with length $\ell$, the nearest-neighbor network has $n = \ell - d\tau$ nodes. In this work, we choose $\ell = 10^4$ and $d = 10$ for the embedding dimension [9].

Each point in the phase space is considered as a node of the network. For each node, four links are generated to its nearest neighbors that have not been connected to it. Specifically, a node is randomly chosen at the first step and the four nearest nodes are determined and connected to the former node. In the next step, each of the four nodes is connected to other four nearest nodes that have no connections to them. This procedure terminates until every node has been chosen to connect other four nodes. In this way, a nearest-neighbor network with $4n$ links is constructed and the average degree of node is exactly 8. The results presented below are averaged over many realizations. Certainly, different orders of chosen nodes in constructing connections result in different networks. However, our results are independent of the node order. Since all the constructed networks have the same size, we can directly compare the motif distributions without any reference to random networks [9, 28, 29].
A Figure 1. All six network motifs of size 4 in undirected nearest-neighbor networks. These motifs are labeled A, B, C, D, E, and F, respectively.

Following [9], we consider motifs with four nodes. Figure 1 shows all possible motifs that are connected to undirected graphs containing four nodes. In general, different networks have different motif distributions, and it is not necessary that all motifs appear. For instance, a chain-like network or a loop has only motif A, a star-like graph has only motif B, and a complete graph has only motif F. It is argued that distinct systems with the same motif occurrence rank may share some similar key elements and belong to a same superfamily [9, 29].

2.2. Detrended fluctuation analysis

The idea of DFA was invented originally to investigate the long-range dependence in coding and noncoding DNA nucleotide sequences [31]. Due to its efficiency and simplicity in implementation, the DFA is now becoming the most important method in the field [32]. The DFA algorithm can be implemented as follows.

For a given nonstationary time series \( \{z(i)|i = 1, 2, \ldots, \ell\} \) (say, FBMs, MRWs, turbulence velocity signals and stock market indices), define the profile \( y(i) \) of the increments \( x(i) = z(i) - z(i-1) \) as follows:

\[
y(i) = \sum_{j=1}^{i} [x(i) - \langle x \rangle], \quad i = 1, 2, \ldots, \ell, \tag{1}
\]

where \( \langle x \rangle \) is the sample mean of \( x(i) \). The series \( y(i) \) is covered by \( \ell_s \) disjoint boxes with the same size \( s \). When the whole series \( y(i) \) cannot be completely covered by \( \ell_s \) boxes, we can utilize \( 2\ell_s \) boxes to cover the series from both ends of the series. In each box, a polynomial trend function \( g(i) \) of the subseries is determined. The residuals are calculated by

\[
\epsilon(i) = y(i) - g(i), \tag{2}
\]

and the local detrended fluctuation function \( f_v(s) \) in the \( v \)th box is defined as the rms of the fitting residuals:

\[
[f_v(s)]^2 = \frac{1}{s} \sum_{i=(v-1)s+1}^{vs} [\epsilon(i)]^2. \tag{3}
\]
3. Numerical investigation of superfamily phenomenon in FBMs and MRWs

We synthesize FBMs using the wavelet-based algorithm to simulate FBMs [33] and MRWs [34]. For each input Hurst index $H$, which runs from 0.05 to 0.95 with a step of 0.05, 100 realizations of FBMs and MRWs are generated and the corresponding motif distribution for each time series is obtained. Figure 2 illustrates the dependence of the averaged motif ranks in descending order of the occurrence frequency with respect to the Hurst index $H$ for the two types of random walks. The insets of figure 2 compare the DFA scaling exponent $\alpha$ of the sample series and the input Hurst index $H$ for the FBMs and MRWs based on 100 realizations. Specifically, for each FBM (MRW) series with the input Hurst index $H$, we calculate the DFA scaling exponent $\alpha$ based on the DFA [31, 32]. It is found that $\alpha$ is very close to $H$ for the FBMs, while $\alpha$ is almost identical to $H$ when $H \geq 0.45$ but systematically deviates from $H$ when $H < 0.45$ for the MRWs [19].

We observe that clear patterns emerge in figure 2. For the FBMs, there are four motif-rank patterns: ABCDEF when $H \leq 0.20$, ACBDFE when $0.25 \leq H \leq 0.30$, ACDBFE when $0.35 \leq H \leq 0.40$, and ACDFBE when $0.45 \leq H$. For the MRWs, there are three motif-rank patterns: ACBDFE when $H \leq 0.30$, ACDBFE when $0.35 \leq H \leq 0.40$, and ACDFBE when $0.45 \leq H$. Strikingly, when we compare the two cases using the DFA scaling exponent $\alpha$, the two plots tell the same story: the superfamily classification is determined uniquely by the DFA scaling exponent. To be more clear, for $H \leq 0.30$ in the case of MRWs in figure 2(b), $\alpha$ is less than 0.35 and greater than 0.25, which corresponds to the motif-rank pattern ACBDFE in figure 2(a). It is not surprising that the pattern ABCDEF is missing in the case of MRWs, since it is hard to obtain MRW series with the DFA scaling exponent $\alpha$ smaller than 0.20. We can thus focus on the FBM case in the following discussion.
In figure 2, we identify two key motifs, B and F, whose locations determine the associated motif-rank patterns. If we ignore B and F, the four patterns reduce to the same combination ACDE. This phenomenon is also observed in all the cases presented in [9]. The relative occurrence frequency (or rank) of motif B decreases with the increase of the DFA scaling exponent, and the rank of motif F increases with the DFA scaling exponent. We also find that, with the increase of the DFA scaling exponent, the ranks of C and D increase and the rank of E decreases. These trends can be quantitatively explained as follows, especially for B (sparse motif) and F (dense motif).

Motif B occurs when the neighbors of the point are all far from each other, while motif F occurs when the four points are all close enough to each other. Therefore, motif B represents the most irregular structure while motif F indicates the most transitive condition. For a large DFA scaling exponent, the time series is relatively smoother. The consequence is that data points with close time moments in the time series are mapped to close points in the phase space that are more probable to connect with each other. More F motifs are expected to appear. For small DFA scaling exponent, the time series is relatively rougher. It follows that close points in the time series are mapped to a sparse region in phase space. In the sparse region, nontransitive structures such as motif B will be common [9]. This approximate explanation applies to the observations in real systems given below.

4. Superfamily phenomenon in real systems

We have shown above that correlated time series have a unique superfamily with a specific motif-rank pattern ACDFBE while anticorrelated time series may exhibit four different motif-rank patterns based on the associated DFA scaling exponents. In this section, we investigate the superfamily phenomenon in two real systems using stock market indices and turbulence velocity. The similarity between stock markets and turbulence has been extensively studied [35, 36], which has stimulated the development of econophysics [37]. This comparative study aims to check if these two systems belong to the same superfamily.

4.1. Example 1: turbulence velocity signals

The velocity data set in three-dimensional fully developed turbulence has been obtained from a hot-wire probe in the S1 wind tunnel of ONERA by the Grenoble group from LEGI [38]. The mean velocity of the main flow is \( \langle v \rangle = 20 \, \text{m s}^{-1} \), and the rms velocity fluctuation is \( v_{\text{rms}} = 1.684 \, \text{m s}^{-1} \). Thus, the turbulence intensity is calculated to be \( I = v_{\text{rms}} / \langle v \rangle = 0.0842 \) and the velocity is large enough compared with turbulent fluctuations to use the Taylor’s frozen-flow hypothesis. The integral scale \( \approx 2 \, \text{m} \) is obtained from the autocorrelation of the velocity series (according to the Taylor’s frozen-flow hypothesis). The Kolmogorov microscale \( \eta \) is \( 0.195 \, \text{mm} \), the Taylor microscale is \( \lambda = 16.6 \, \text{mm} \), and the Taylor-scale Reynolds number is \( Re_\lambda = 1863 \) [39].

The measurement resolution of the experiment is \( 3.55 \times 10^{-5} \, \text{s} \). The total length of the velocity signal is about \( L = 1.7 \times 10^7 \). The signal is sampled at different time scales \( \delta t \) (in units of the measurement resolution). In our analysis, \( \delta t \) ranges from 1 to 200 with an increment of 1. For each \( \delta t \), the length of the sampled signal is about \( L_{\delta t} = L / \delta t \), and 100 random subseries with length \( \ell = 10^4 \) are extracted for analysis. The estimated DFA scaling exponents \( \alpha \) are shown in figure 3(a) as a function of the sampling time scale \( \delta t \), and the motif-rank patterns for each \( \delta t \) are illustrated in figure 3(b).

In figure 3(b), we observe five motif-rank orders: ACDFBE for \( \delta t = 1 \), ACDBFE for \( 2 \leq \delta t \leq 4 \), ACBDFE for \( 5 \leq \delta t \leq 64 \), ABCDFE for \( 65 \leq \delta t \leq 82 \), and ABCDEF for
Figure 3. (a) Dependence of the DFA scaling exponent $\alpha$ of the resampled turbulence signals on the time scale $\delta t$. (b) Dependence of the motif-rank pattern on the resampling interval $\delta t$. The values with $\delta t$ larger than 120 are not shown.

Figure 4. Network motif ranks of turbulence signals with different sampling time scales: (a) $\delta t = 1$, (b) $2 \leq \delta t \leq 4$, (c) $5 \leq \delta t \leq 64$, (d) $65 \leq \delta t \leq 82$ and (e) $83 \leq \delta t$. (This figure is in colour only in the electronic version)

Taking into consideration the fact that the motif-rank pattern ABCDEF is very close to ABCDEF as we will show below, it is found that with the decrease of $\delta t$ (thus the increase of $\alpha$), the motif-rank pattern changes in an order of ABCDEF, ACBDFE, ACDBFE and ACDFBE. Furthermore, the switching values of $\alpha$ from one pattern to the other are also in excellent agreement with figure 2. Figure 3 also implies that the turbulence velocity signal possesses multiscale nature. However, the switching values of $\delta t$ cannot be related to the well-known turbulence scales (such as the dissipation subrange and the inertial subrange).

Plots (a)–(e) of figure 4 illustrate typical motif ranks in descending order for the networks induced from velocity series at different time scales. Figure 4(e) shows that the occurrence frequency of motif B (and motif A as well) decreases with $\alpha$ and the frequency of motif F (and D as well) increases with $\alpha$, as explained for the FBM case. Figure 4(e) confirms this
observation since the occurrence frequency of motif F decreases with \( \delta t \) and thus increases with \( \alpha \). There are also time series (say, \( \delta t = 55 \) in figure 4(c) and \( \delta t = 75 \) in figure 4(d)) that have different motif-rank patterns with their ‘neighbors’. However, this situation is rare, and averaging over 100 realizations with the same \( \delta t \) gives consistent and stable results. In addition, we find that the relative frequencies of motifs E and F are comparable in figure 4(d), which makes the motif-rank pattern ABCDFE marginal when compared with the pattern ABCDEF.

The above observations are further illustrated in figure 4(f). We find that with the increase of \( \delta t \), the motif frequencies of A, B, C and E increase, while the motif frequencies of D and F decrease. There are two decreasing quaternionic combinations, ACDE and ACFD. In addition, there are four intersections caused by B and C; B and E; B and F; and E and F, which correspond to the four switching values of \( \delta t \) separating the five motif-rank patterns shown in figures 4(a)–(e).

4.2. Example 2: stock market indices

Now we study the superfamily phenomenon of three stock market indices: the DJIA from 26 May 1896 to 27 April 2007 of size 30 147, the S&P 500 index from 29 November 1940 to 7 January 2008 of size 16 943, and the German DAX index from 1 October 1959 to 15 November 2002 of size 10 812. For these three data sets, it is hard to perform a similar analysis as for the turbulence signal due to the short length of the data. Therefore, we do not perform multiscale analysis on the stock index data. We convert the time series into three nearest-neighbor networks in the phase space and count the occurrence of the six motifs. The motif ranks for the three stock indices are shown in figure 5.

According to figure 5, the three stock indices have the same motif ranks ACDFBE. This is consistent with the regime of \( \alpha \geq 0.45 \) in the FBM case illustrated in figure 2(a). It is well known that stock index returns are basically uncorrelated with the DFA scaling exponent \( \alpha \approx 0.5 \) [37]. It is evident that the superfamily phenomenon of stock indices confirms the main conclusion of this work.
5. Conclusion

Time series can be mapped into nearest-neighbor networks, which allows us to classify time series into different superfamilies based on the motif occurrence patterns. We have investigated the motif patterns of FBM and MRW and found that the DFA scaling exponent of the time series plays a unique role in the superfamily classification. The multifractal nature of the MRWs has no influence on the motif ranks. Four different motif-rank patterns ABCDEF, ACDBFE, ACDFBE and ACDBFE are identified, which correspond to different ranges of the DFA scaling exponent. The results of similar analyses of velocity time series in fully developed turbulence and daily stock market indices are essentially consistent with this finding, except that an additional pattern ABCDEF appears between ABCDEF and ACDFBE with a very narrow range of exponent $\alpha$ for the turbulence case.

Acknowledgments

This work was partially supported by the Program for New Century Excellent Talents in University (NCET-07-0288) and the Fundamental Research Funds for the Central Universities.

References

[1] Albert R and Barabási A L 2002 Rev. Mod. Phys. 74 47–97
[2] Newman M E J 2003 SIAM Rev. 45 167–256
[3] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D U 2006 Phys. Rep. 424 175–308
[4] Zhang J and Small M 2006 Phys. Rev. Lett. 96 238701
[5] Li P and Wang B H 2006 Chin. Sci. Bull. 51 624–9
[6] Li P and Wang B H 2007 Physica A 378 519–26
[7] Zhang J, Sun J F, Luo X D, Zhang K, Nakamura T and Small M 2008 Physica D 237 2856–65
[8] Li C B, Yang H and Komatsuzaki T 2008 Proc. Natl Acad. Sci. USA 105 536–41
[9] Xu X K, Zhang J and Small M 2008 Proc. Natl Acad. Sci. USA 105 19601–5
[10] Lacasa L, Luque B, Ballesteros F, Luque J and Nuño J C 2008 Proc. Natl Acad. Sci. USA 105 4972–5
[11] Luque B, Lacasa L, Ballesteros F and Luque J 2009 Phys. Rev. E 80 046103
[12] Kostakos V 2009 Physica A 388 1007–23
[13] Shirazi A H, Jafari G R, Davoudi J, Peinke J, Tabar M R R and Sahimi M 2009 J. Stat. Mech. P07046
[14] Marwan N, Donges J F, Zou Y, Donner R V and Kurths J 2009 Phys. Lett. A 373 4246–54
[15] Donner R V, Zou Y, Donges J F, Marwan N and Kurths J 2010 New J. Phys. 12 033025
[16] Yang Y and Yang H J 2008 Physica A 387 1381–6
[17] Gao Z K and Jin N D 2009 Phys. Rev. E 79 066303
[18] Gao Z K and Jin N D 2009 Chaos 19 033137
[19] Ni X H, Jiang Z Q and Zhou W X 2009 Phys. Lett. A 373 3822–6
[20] Lacasa L, Luque B, Luque J and Nuño J C 2009 Europhys. Lett. 86 30001
[21] Elsner J B, Jagger T H and Fogarty E A 2009 Geophys. Res. Lett. 36 L16702
[22] Yang Y, Wang J B, Yang H J and Mang J S 2009 Physica A 388 4431–7
[23] Liu C, Zhou W X and Yuan W K 2010 Physica A 389 2675–81 (arXiv:0905.1831)
[24] Quan M C, Jiang Z Q and Zhou W X 2010 J. Phys. A: Math. Theor. 43 335002 (arXiv:0910.2524)
[25] Watts D J and Strogatz S H 1998 Nature 393 440–2
[26] Amaral L A N, Scala A, Barthelemy M and Stanley H E 2000 Proc. Natl Acad. Sci. USA 97 11149–52
[27] Barabási A L and Albert R 1999 Science 286 509–12
[28] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U 2002 Science 298 824–7
[29] Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M and Alon U 2004 Science 303 1538–42
[30] Fraser A and Swinney H 1986 Phys. Rev. A 33 1134–140
[31] Peng C K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 Phys. Rev. E 49 1685–9
[32] Kantelhardt J W, Koscielny-Bunde E, Rego H H A, Havlin S and Bunde A 2001 Physica A 295 441–54
[33] Abry P and Sellan F 1996 Appl. Comput. Harmon. Anal. 3 377–83
[34] Bacry E, Delour J and Muzy J F 2001 Phys. Rev. E 64 026103
[35] Ghashghaie S, Breymann W, Peinke J, Talkner P and Dodge Y 1996 Nature 381 767–70
[36] Mantegna R N and Stanley H E 1996 Nature 383 587–8
[37] Mantegna R N and Stanley H E 2000 An Introduction to Econophysics: Correlations and Complexity in Finance
  (Cambridge: Cambridge University Press)
[38] Anselmet F, Gagne Y, Hopfinger E J and Antonia R A 1984 J. Fluid Mech. 140 63–89
[39] Zhou W X and Sornette D 2002 Physica D 165 94–125