Detection and prediction of the onset of human ventricular fibrillation: An approach based on complex network theory

Xiang Li* and Zhao Dong†

Adaptive Networks and Control Laboratory, Electronic Engineering Department, Fudan University, Shanghai 200433, China

(Received 12 July 2011; revised manuscript received 15 October 2011; published 12 December 2011)

Ventricular fibrillation is a life-threatening cardiac arrhythmia which deserves quick and reliable detection as well as prediction from human electrocardiogram time series. We constructed networks of human ventricular time series with the visibility graph approach to study the network subgraph phenomenon and motif ranks. Our results show that different dominant motifs exist as an effective indicator in distinguishing ventricular fibrillations from normal sinus rhythms of a subject. We verify the reliability of our findings in a large database with different time lengths and sampling frequencies, and design an onset predictor of ventricular fibrillations with reliable verifications.

DOI: 10.1103/PhysRevE.84.062901 PACS number(s): 87.19.Hh, 89.75.Hc, 89.75.Kd

Ventricular fibrillation as one of the main causes of sudden cardiac death deserves quick and reliable detection as well as prediction from human electrocardiogram time series. The past decades have witnessed extensive efforts and fruitful outcomes in developing new measures and approaches based on time-series analysis to overcome this challenge [1–6]. Recently, the exciting efforts have been devoted to mapping a time series into a network [7–9], and the topological analyses based on complex network theory [13–16] have been applied to help identify human electrocardiogram time series of patients with risk to sudden cardiac death, arrythmia, and congestive heart failure [7,10–12]. Note that these works have tried to classify different categories of time series with some indicators of global connectivity patterns, such as degree distributions [7], degree-degree correlation [10], and degree-mixing assortativity [10–12], of the associated networks, and the sensitivity to data lengths arouses more investigations on their reliability analysis [12].

On the other hand, the subgraph characteristics of associated human electrocardiogram networks deserve more attention, since it has been widely recognized that the rank distribution of subgraphs (or motifs) can help classify diverse networks [17,18]. One evidence in time series analysis comes from the latest work of Xu *et al.* that the time series of periodic, chaotic, and noise data have been classified with the different ranks of subgraph frequencies in the associated networks [9].

Therefore, this Brief Report does not follow the macroscopic network topological indicators in previous analysis of human electrocardiogram time series [7,10–12]. Here we apply the visibility graph method [8] to generate associated networks for the ventricular fibrillation (VF) time series in an MIT human heartbeat database, the Creighton University Ventricular Tachyarrhythmia Database (CUDB) [19]. We observe that the four-node subgraphs constructed from the time series of VF patients rank significantly different from those of normal sinus rhythm (NSR) data, and verify this conclusion with the dependence on different sampling frequencies and data lengths as well. Besides that, we designed an onset predictor of ventricular fibrillations with its effectiveness finally verified.

We briefly recall the visibility graph approach [8] in mapping a time series to a network as follows. Every data point of a time series is mapped to a node in its associated network. If two data points (t_a, y_a) and (t_c, y_c) in the time series at time t_a and t_c, respectively, are visible, that is, any other data (t_b, y_b) between them $(t_a < t_b < t_c)$ satisfies the following visibility criterion:

$$y_b < y_c + (y_a - y_c)\frac{t_c - t_b}{t_c - t_a}, \tag{1}$$

the corresponding two nodes are connected, yielding an undirected network. Figure 1 shows an example of the given time series which is mapped to an associated network with the visibility algorithm.

We apply the visibility graph method to the NSR and the ventricular fibrillation data in the CUDB, which has a total of 35 subjects (with the data files named from CU01 to CU35, one channel per file or subject with each channel 508 s long), and generate the associated networks. All signals in the database are sampled at 250 Hz with 12 bit resolution.

After mapping a time series of 10 s to an associated network, we use subgraph detection software [20] to find subgraphs in the network. In this Brief Report we only target the four-node subgraphs in the undirected networks, and all six types of four-node subgraphs are illustrated in Fig. 2.

We start the work with one subject having the ventricular fibrillation time series in the CUDB. With the visibility criterion (1), we first map every 10-s episode (of the NSR or VF data) of CU01 to generate an associated network, and thus have 20 NSR networks and 20 VF networks from the total 508 s data of CU01, each of which has 2500 nodes. We calculate the percentage of each subgraph of Fig. 2 in the generated networks by the detection software in [20], and find the subgraph ranks in the descending order as shown in Fig. 3. Therefore, for the same one subject (CU01), we clearly observe that the electrocardiograms in the NSR (healthy) status and VF status exhibit significantly different subgraph ranks as CBADFE and ACBDFE, respectively. Especially, the dominant motif in the NSR networks is motif C while the one in the VF networks is motif A.

*lix@fudan.edu.cn.
†082021040@fudan.edu.cn

 ©2011 American Physical Society

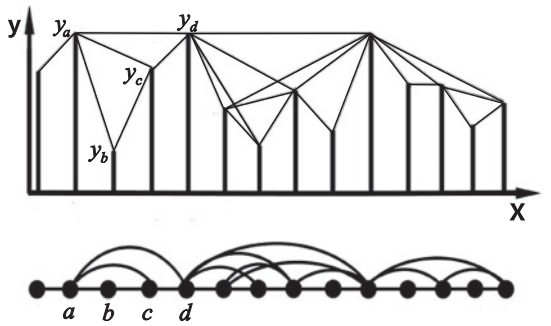

FIG. 1. An exampled time series and the associated graph derived from the visibility graph method. Each node in the graph corresponds to data in the time series. For example, the data point (t_b, y_b) can "see" other data points (t_a, y_a) and (t_c, y_c), so the corresponding node b is linked with a and c. The data point (t_b, y_b) cannot "see" data point (t_d, y_d), so the corresponding nodes b and d are not linked.

Note that motif A is a chain, while motif C is star structured. Such a significant difference of the dominant motifs between the NSR and VF networks comes from the different features of the NSR and ventricular fibrillation time series. In the NSR case, the R wave is the most prominent characteristic waveform with peak values in the electrocardiogram episode, and that the star-structured motif C prevails in the NSR networks. On the other hand, such R waves disappear in the VF data, which exhibits more aperiodic waves. Therefore, the chain-like motif A replaces motif C to dominate the VF networks.

We further extend our experiment to all 35 subjects of the whole CUDB database, and verify the discrimination between the NSR and VF data with the reference of the dominant motifs C and A in the associated networks. For simplicity we only focus on the most common three motifs in the generated networks. Randomly selecting 200 NSR 10-s episodes and 200 VF 10-s episodes from all 35 subjects in the database, we find that there are 131 NSR episodes having the most three common subgraphs ranked as CBA, and 48 NSR ones having the subgraph rank as CAB. On the other hand, there are 162 VF data presenting the subgraph rank of ACB, and 31 VF data with the rank of ABC, as listed in Table I. Therefore, motif C is the dominant subgraph in 179 of the total 200 NSR networks, while motif A is the dominant one in 193 of the total 200 VF networks. All of the NSR and VF networks are generated with the corresponding data episode of 10 s with 250 Hz sampling frequency.

Since ventricular fibrillation is a life-threatening cardiac arrhythmia, it requires both quick and reliable detection to save the patient's life. Many detection algorithms have been evaluated for their reliability by the two measures of sensitivity

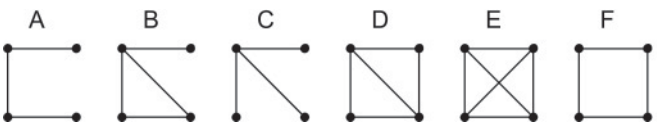

FIG. 2. All six types of four-node subgraphs in an undirected network, which are labeled from A to F.

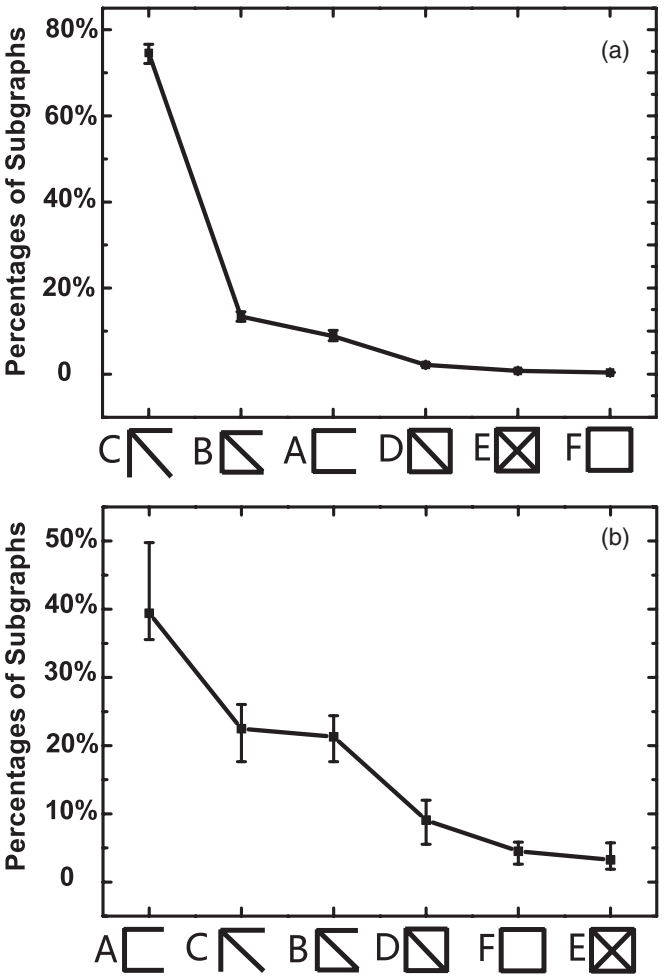

FIG. 3. The averaged percentages of six subgraphs in 20 networks generated from (a) the NSR episodes and (b) the VF episodes of CU01.

(SE) and specificity (SP) defined as below, the bigger, the more reliable:

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \quad \mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{2}$$

where TP is the number of true positive decisions (a VF case being correctly recognized as a VF one), FN is the number of false negative decisions (a VF case being wrongly recognized as an NSR one), TN is the number of true negative decisions (an NSR case being correctly recognized as an NSR one), and

TABLE I. The statistical number of 200 NSR and 200 VF episodes covering 35 subjects of the CUDB database, whose associated networks have the specified subgraph ranks of the most common three motifs in descending order, and each episode is 10 s with 250 Hz sampling frequency.

Specified Subgraph Ranks of the Most Common Three Motifs						
Data Type	CBA	CAB	ACB	ABC	Else	Total Numbers
NSR	131	48	0	0	21	200
VF	0	1	162	31	6	200

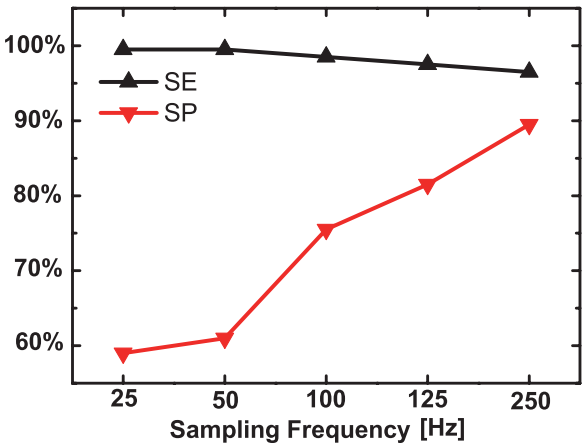

FIG. 4. (Color online) The reliability measures of sensitivity (SE) and specificity (SP) of 200 NSR and 200 ventricular fibrillation episodes with different sampling frequencies of 25, 50, 100, 125, and 250 Hz, respectively. All episodes are the same as those data in Table I with the same time length of 10 s but different sampling frequencies.

FP is the number of false positive decisions (an NSR case being wrongly recognized as a VF one).

We have found that motif C dominates in the associated networks of NSR time series, while motif A dominates in those of VF time series. Here we conjecture the dominance of motif A or motif C as an indicator in distinguishing the VF time series from the NSR electrocardiogram. With the statistical data in Table I covering total 35 subjects of the database, we calculate that the sensitivity and specificity are 96.5% and 89.5% (where TP = 193, TN = 179, FP = 21, FN = 7 for simplicity), respectively, indicating a very effective indicator in identifying a ventricular fibrillation patient from a healthy subject.

We also consider the dependence of the above reliability results on different sampling frequencies and time lengths. As shown in Figs. 4 and 5, we clearly observe that the longer

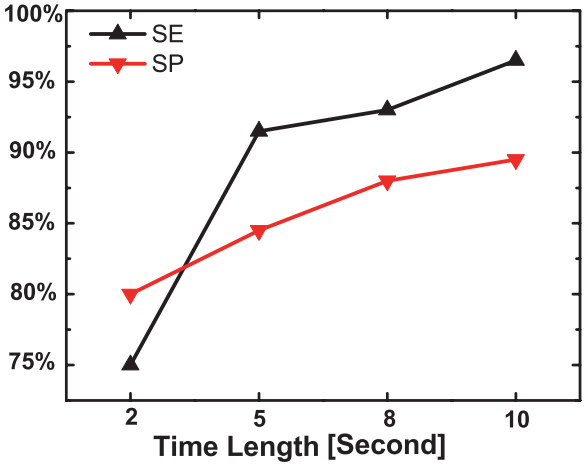

FIG. 5. (Color online) The reliability measures of sensitivity (SE) and specificity (SP) of 200 NSR and 200 ventricular fibrillation episodes with different time lengths of 2, 5, 8, and 10 s, respectively. All episodes are the same as those data in Table I with the same sampling frequency 250 Hz but different time lengths.

the time length (the higher the sampling frequency), the more reliable the identification.

Another more challenging problem is to predict the onset of VF based on the ventricular time series data. We have observed that the different dominant motifs of C and A work as a reliable indicator in distinguishing a ventricular fibrillation from the normal sinus rhythm of a subject. Therefore, when a subject experiences a transition from the NSR physiological state to the VF state, we conjecture that the percentage of motif C decreases while that of motif A increases significantly during the transition, with which we put a step forward in further predicting the onset of VF.

To design the onset predictor of VF, we set each data episode as five successive ECG beats in a time series. Therefore, after generating an associated network for each beat with the visibility graph method, we calculate $M(i)$ as the difference between the percentages of motif C and motif A in the associated network for each beat i,$i = 1, 2, 3, 4, 5$, yielding the mean difference in a 5-beat episode as $\overline{M} = \sum_{i=1}^{5} \frac{M(i)}{5}$. We first target the subject of CU15 in the CUDB database, having 405 s of NSR data before the onset of VF. We examine the CU15 data in every 60-s interval and calculate $\overline{M}$ of each 5-beat episode as shown in Fig. 6. We clearly observe that at the 5-beat NSR episode 405 s before the VF onset [Fig. 6, inset (a)], the corresponding $\overline{M} = 37.84\%$. While the one at the 5-beat episode 105 s before the VF onset [Fig. 6, inset (b)] is 3.99%, and that of after the VF onset is -13% [Fig. 6, inset (c)]. Therefore, we set the alarm threshold as 5% and design the onset predictor of VF as follows:

if $\overline{M} < 5\%$, the onset predictor predicts that a VF will happen.

In the CUDB database there are a total of 40 NSR records before the onset of VF with 32 subjects involved, and the data

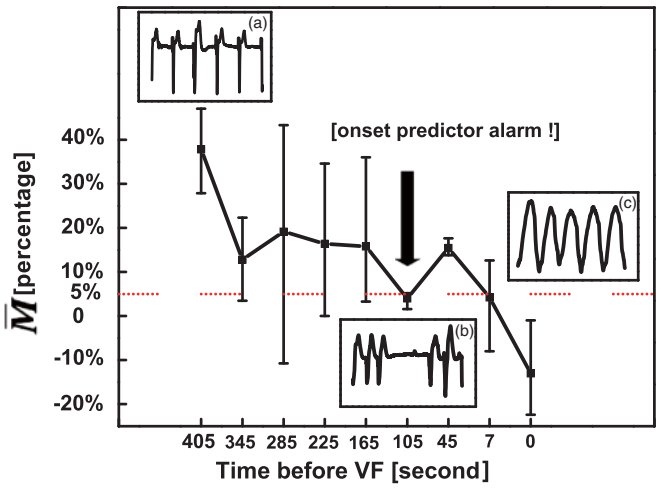

FIG. 6. (Color online) The mean difference $\overline{M}$ between the percentages of motif C and motif A in associated networks generated from the NSR 5-beat episode before the onset of VF in CU15. The whole data series is 405 s before the VF onset, and every 5-beat episode is examined in 60-s intervals. Inset (a): The 5-beat NSR episode is 405 s before the onset of VF; (b) the 5-beat NSR episode is 105 s before the onset of VF; and (c) the VF episode is 0.4 s after the onset of VF. The (red) dashed line is the alarm threshold of our designed onset predictor.

lengths range from 13 to 496 s. As a comparison, we also examined all 18 NSR records from the NSRDB database in [21] to evaluate the reliability of our proposed onset predictor. With these 58 records, the onset predictor of VF achieves a sensitivity (SE) of 85% and a specificity (SP) of 88.8%, and the predictor can predict an imminent VF from 5 to 290 s prior to a true onset of VF.

To summarize this Brief Report, we have applied the visibility graph approach to generate associated networks of human ventricular fibrillation time series with extensive data of the CUDB database, whose subgraphs rank and dominant motif are significantly different from those of healthy ones. Based on this reliable finding, we have further designed an onset predictor of ventricular fibrillations, and verified the effectiveness with the CUDB and NSRDB databases. Therefore, we arrive at the conclusion that different dominant motifs work as an effective and reliable indicator to distinguish and predict (the onset of) ventricular fibrillations from normal sinus rhythms of a subject. We should state that our method fails to distinguish VF and ventricular tachycardia (VT) at present, which deserves our efforts in designing new mapping methods, and the ECG analysis based on complex network theory is worth more attention in answering the remaining questions [22].

The authors are grateful to the anonymous reviewers for their valuable comments in helping improve this paper. This work was partly supported by the Natural Science Foundation Program No. 60874089 and the NCET Program No. 09-0317 of China.

[1] L. Sherman, C. Callaway, and J. Menegazzi, Resuscitation **47**, 2 (2000).
[2] A. Amann, R. Tratnig, and K. Unterkofler, IEEE Trans. Biomed. Eng. **54**, 1 (2007).
[3] M. C. Wu, E. Watanabe, Z. R. Struzik, C. K. Hu, and Y. Yamamoto, Phys. Rev. E **80**, 051917 (2009).
[4] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, Phys. Rev. E **66**, 026702 (2002).
[5] N. Wessel, H. Malberg, R. Bauernschmitt, and J. Kurths, Int. J. Bifurcation Chaos **17**, 10 (2007).
[6] A. Voss, S. Schulz, R. Schroeder, M. Baumert, and P. Caminal, Philos. Trans. R. Soc. Sect. A **367**, 277 (2009).
[7] J. Zhang and M. Small, Phys. Rev. Lett. **96**, 238701 (2006).
[8] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, Proc. Natl. Acad. Sci. USA **105**, 4872 (2008).
[9] X. Xu, J. Zhang, and M. Small, Proc. Natl. Acad. Sci. USA **105**, 19601 (2008).
[10] J. Zhang, J. Sun, X. Luo, K. Zhang, T. Nakamura, and M. Small, Physica D **237**, 2856 (2008).
[11] Z. Shao, Appl. Phys. Lett. **96**, 073703 (2010).
[12] Z. Dong and X. Li, Appl. Phys. Lett. **96**, 266101 (2010).
[13] R. Albert and A. Barabasi, Rev. Mod. Phys. **74**, 1 (2002).
[14] G. Chen and X. Wang, IEEE Circ. Syst. Mag. **3**, 1 (2003).
[15] A. Barabasi, Nat. Phys. **1**, 68 (2005).
[16] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[17] R. Milo, R. Shen-Orr, S. Itzkovitz, N. D. Chklovskii, and U. Allon, Science **298**, 5594 (2002).
[18] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science **303**, 5663 (2004).
[19] http://www.physionet.org/physiobank/database/cudb/.
[20] We use FANMOD as the network motif detection tool, which is available at the URL http://theinf1.informatik.uni-jena.de/wernicke/motifs/index.html.
[21] http://www.physionet.org/physiobank/database/nsrdb/.
[22] D. Yang and X. Li (submitted).