

# Statistical characterization of an ensemble of functional neural networks

B.B.M. Silva<sup>1</sup>, J.G.V. Miranda<sup>1</sup>, G. Corso<sup>2</sup>, M. Copelli<sup>3</sup>, N. Vasconcelos<sup>4,5</sup>, S. Ribeiro<sup>5</sup>, and R.F.S. Andrade<sup>1,a</sup>

<sup>1</sup> Instituto de Física, Universidade Federal da Bahia, 40210-340 Salvador, Brazil

<sup>2</sup> Departamento de Biofísica e Farmacologia, Centro de Biociências, Universidade Federal do Rio Grande do Norte, 59072-970 Natal, Brazil

<sup>3</sup> Departamento de Física, Universidade Federal de Pernambuco, 50670-901 Recife, Brazil

<sup>4</sup> Departamento de Sistemas e Computação, Universidade Federal de Campina Grande, 58.429-900 Campina Grande, Brazil

<sup>5</sup> Instituto do Cérebro, Universidade Federal do Rio Grande do Norte, 59015-430 Natal, Brazil

Received 17 June 2012 / Received in final form 6 September 2012

Published online 24 October 2012 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2012

**Abstract.** This work uses a complex network approach to analyze temporal sequences of electrophysiological signals of brain activity from freely behaving rats. A network node represents a neuron and a network link is included between a pair of nodes whenever their firing rates are correlated. The framework of time varying graph (TVG) is used to deal with a very large number ( $>30\,000$ ) of time dependent networks, which are set up by taking into account correlations between neuron firing rates in a moving time lag window of suitable width. Statistical distributions for the following network measures are obtained: size of the largest connected cluster, number of edges, average node degree, and average minimal path. We find that the number of networks with highly correlated activity in distinct brain areas has a fat-tailed distribution, irrespective of the behavioral state of the animal. This contrasts with short-tailed distributions for surrogates obtained by shuffling the original data, and reflects the fact that neurons in the neocortex and hippocampus often act in precise temporal coordination. Our results also suggest that functional neuronal networks at the millimeter scale undergo statistically nontrivial rearrangements over time, thus delimitating an empirical constraint for models of brain activity.

## 1 Introduction

In the last decade, the complex network framework has been applied to investigate a large number of biological and social systems [1–3]. In this kind of investigation it is necessary to identify basic constituents of the original systems and the meaningful relationships among them. These elements will be mapped, respectively, onto network nodes (or vertices) and network edges (or links). It has been shown that many distinct measures used to characterize such networks also provide valuable quantitative insights about the way the constituents of the original system are related among themselves [4–6].

The anatomical structure of neurons and dendrites turns it quite easy to conceive the brain as a network [7]. However, the use of network theory to measure empirical properties of the neural system is quite recent [8]. Studies in this direction have indeed revealed that the brain as a network of synaptically connected neurons has small-world properties [9], which can be expressed at different spatial scales [10].

Other studies have changed the focus from form to function, using an alternative and complementary

approach to construct, characterize, and model functional neuronal networks (FNN). A FNN is usually obtained following a few steps: (1) define the network nodes as the units from which electrophysiological signals can be recorded (e.g. voxels in functional magnetic resonance imaging (fMRI), or electrodes in electroencephalography (EEG) or magneto-encephalography (MEG) recordings); (2) estimate a measure of functional association between the nodes (e.g. linear correlation, or mutual information); (3) generate an adjacency matrix based on some threshold criterion for the chosen measure [8]. This basic sequence was used to construct a FNN of the human brain [11], where each node corresponds to a small voxel from which the blood oxygen level-dependent (BOLD) signal was measured in a typical fMRI experiment. A network edge was introduced between the nodes  $i$  and  $j$  whenever the linear correlation among the BOLD signals from the corresponding voxels exceeded a suitable threshold value, yielding a power law distribution of node degree [12]. Similar distributions were found among neuron activity in in-vitro investigations, where the links are established by synchronization in calcium metabolism of cells [13]. Theoretical modeling has aimed at reproducing the typical experimental features in firing activity [14–17], but the

<sup>a</sup> e-mail: randrade@ufba.br

overall long term goal is far from being concluded: to understand the relationship between form and function, i.e. how does structural connectivity influences functional connectivity, and vice-versa.

The main task of this work is to construct and characterize FNN's related to neurophysiologic dynamics. We use extracellular records obtained from multi-electrode arrays (MEAs) of a few square millimeters chronically implanted in the brains of rats. The subjects were recorded for a few hours, during which they were freely behaving, exposed to novel objects, and spontaneously traversed their sleep-wake cycle. The use of MEA records for the FNN construction requires a suitable framework to bridge the difference between the typical time scales of firing spikes and neurophysiologic information processing.

The network nodes are identified as the measured neurons, while links among pairs of neurons are established requiring a lower bound in the correlation between in their firing rate. To deal with large record sets, we evaluate pairwise correlation between firing rates of neurons inside a moving window of duration  $W$  (on the order of one or two seconds), what leads to a very large number of networks (>30 000), each one of them reflecting the brain operating conditions within a restricted time interval. A statistical characterization of the constructed FNN's is performed on selected measures of the ensemble of networks. The data is analyzed separately for each behavioral state (waking, slow-wave sleep and rapid-eye-movement sleep).

The data used in this study is constrained to the brain areas probed by the electrodes, which differs from the whole-brain fMRI approach [11]. To maximize the number of nodes, we employ single-unit spiking activity. Since our focus is on robust collective properties of networks of neurons, however, we base the functional association among neurons on the correlation between their mean firing rates. The restriction in the number of distinct neurons used to set up the network data is compensated by the large number of analyzed networks using the long time records. Ever since the pioneering work of Sherrington and Adrian [18] early in the 20th century, spike rates have been a staple in neurophysiologic analysis, being extensively investigated with the aim of understanding, among other issues, the dynamics of sensory and motor processes. In summary, our approach differs from previously studied functional brain networks in that we depart from a small set of large networks to a much larger ensemble of small-size networks.

The rest of this work is organized as follows: in Section 2 we describe the methodology used to obtain the data and discuss data interpretation and proper time scales. Section 3 describes the FNN setup, whereas Section 4 presents our results on the statistics of the ensemble of FNN's. We discuss our concluding remarks in Section 5.

## 2 Methodology – data acquisition and electrophysiology

The primary data we work with consists of experimental records of extracellular action potentials (spikes) detected by multi-electrode arrays in surgically implanted

rats [19–21]. We used single wire electrodes for extracellular recordings of action potentials. This technique is very different from intracellular recordings in the sense that the electrodes are not in contact with the cells, but rather placed in the extracellular space. Multichannel digital signal processors are used to distinguish and store the digital waveforms. Despite the distance between electrode tip and cell body, the technique allows for the recording of single units because the different waveforms of the putative neurons can be discriminated based on voltage and time differences [19,22]. The tools used for the reliable separation of units include the analysis of inter-spike intervals, as well as a principal component analysis of the waveforms [19,22]. Extracellular recordings performed in this way have been shown to be very stable, lasting many weeks and even months [19–22]. Discriminable neuronal units are typically obtained from 70–80% of micro-wires implanted. Typically, 2–4 units can be isolated from each of these micro-wires; on average, 2.3 units neuronal can be isolated by each micro-wire [19,22].

For each experimental data set, we denote by  $N$  the number of recorded neurons and by  $T$  the total recording time. We have used data produced by three implanted rats identified as *GE4*, *GE5*, and *GE6*, for which the respective values of  $(N, T(\text{s}))$  are  $(45, 8001)$ ,  $(51, 7952)$ , and  $(57, 8149)$ . The original electrophysiological data set indicates the instants of time in which each neuron fired. Based on this information, it is possible to produce the neuron firing rate in any chosen interval, which becomes the actual input of our network analysis.

During the recording time, the animals remained in complete darkness but were concomitantly videotaped using infrared cameras, so that it is possible to associate the spike firing rates to their behavioral state. We consider the subjects could assume three distinct behavioral states: awake (WK), slow wave sleep (SWS), and rapid eye movement (REM). A detailed description of the three states and the method used to discriminate SWS from REM states based on state-specific features of local field potentials is found in reference [23].

The electrodes were implanted in three distinct areas (hippocampus, somatosensory cortex, primary visual cortex) leading to the following number of registered neurons:  $(HP, S1, V1) = (4, 13, 28) - GE4$ ,  $(13, 16, 22) - GE5$ , and  $(22, 28, 7) - GE6$  [21]. In the current study, however, we disregard the anatomic localization of the neurons and treat the neuron population as a uniform ensemble. This approach is justified as we assume a spreading information hypothesis [24].

The spike firing process has a typical time scale ( $\delta < 10$  ms), while the inter spike interval  $\Delta$  is very rarely less than 5 ms [25]. Thus,  $\delta$  is too short to account for the slower processes of psychophysical and neurophysiologic nature that are likely to be related to any typical behavior observed in mammals. For instance, time intervals of the order of 100 ms roughly match the visual-motor delay that defines reaction time and represents a lower bound for meaningful behavior [26,27]. This information has been corroborated by the successful use of the 250 ms bin to

analyze neuronal ensemble activity of behavioral significance using linear correlations [20] as well as binary classifier including neural networks for pattern analysis [28]. Indeed, the classification quality achieved by binary classifiers is maximum and stable between 45–275 ms [29]. Therefore, we consider these values as guidelines in a procedure to coarse-grain the original information and construct FNN's at a larger scale. More specifically, our framework inquires whether correlations among the simultaneous firing rate of individual neurons are able to produce collective patterns in the brain that might be associated to the neurophysiologic dynamics.

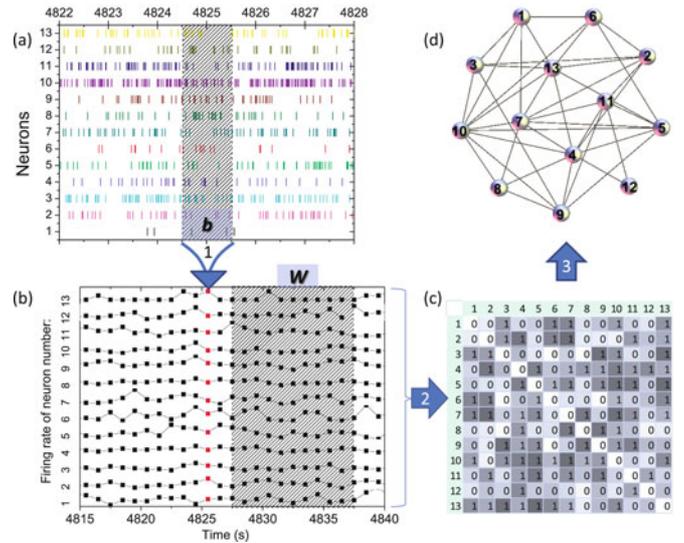
It is worthy to comment that some recent studies proposed a model based approach to approximate the functional connectivity at neuronal level [30,31]. However, there is still a lack of studies using this approach in neuronal data acquired from MEA's recording in a free behaving protocol, as used in this current paper. Finally, we would like to call the attention to the recent findings that strong firing correlations among neuronal pairs are typically concurrent with weak and yet significant correlations within the same neuronal population [32–34]. Despite such coexistence, these works have shown that the collective behavior of spiking neurons can be quantitatively described by maximum entropy models based on pair-wise correlations, which do not take into account higher-order correlations. These models predict that neuronal networks are dominated by strong pair-wise interactions, and that accurate reduced models can be built by dropping the weak interactions. This agrees well with the approach taken in our manuscript, which privileges strong correlations.

### 3 Network construction and time varying graph

For a given data set, each neuron  $i \in [1, N]$ , is associated with a spike time series  $x_i \in [0, T]$ . The elements of  $x_i$  are the firing times, i.e., the instants of time in which a spike in the activity of neuron  $i$  was recorded. The coarse-graining procedure starts by assigning the spikes in  $x_i$  into time intervals of size  $b \in [100 \text{ ms}, 1000 \text{ ms}]$ . Next we count, for each neuron  $i$ , the number of spikes inside each box, obtaining the bin series  $h_i = \{h_i(1), h_i(2), \dots\}$ , where  $h_i(q) \geq 0$  corresponds to the number of spikes fired by the  $i$ th neuron per  $b$  seconds in the  $q$ th time interval. This procedure maps the  $x_i$  series into a firing rate series  $h_i$ , used to construct a correlation matrix among neuron firing rates  $\omega_i = h_i/b$ . To this purpose, we consider a moving window of width  $W$ , encompassing  $N_W = W/b$  bins, and evaluate the Pearson correlation coefficient  $\rho_{i,j}$  between neurons  $i$  and  $j$  over a time interval  $W$  as

$$\rho_{i,j}(\bar{\tau}) = \frac{\sum_{q=\tau}^{\tau+N_W} (h_i(q) - \bar{h}_i)(h_j(q) - \bar{h}_j)}{\sqrt{\sum_{q=\tau}^{\tau+N_W} (h_i(q) - \bar{h}_i)^2 \sum_{q'=\tau}^{\tau+N_W} (h_j(q') - \bar{h}_j)^2}}. \quad (1)$$

$\bar{\tau}$  denotes the starting time of the window of duration  $W$  in which  $\rho$  is computed, and  $\tau = \bar{\tau}/b$ .  $h(q)$  contains the bins inside the  $q$ th window, while  $\bar{h}$  is the respective average.

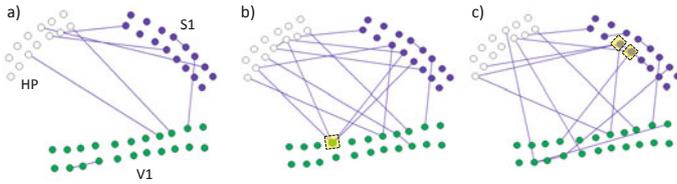


**Fig. 1.** (Color online) Stepwise illustration of the methodology: (a) recorded neuron firing spikes as function of time. The primary input is used to calculate the firing rate for each neuron in a bin of width  $b$ . (b) Sequence of firing rates. Red squares highlight the firing rates evaluated in (a), while the gray window  $W$  is used to compute Pearson's correlation (see Eq. (1)). (c) Resulting values of correlation matrix projected into four gray tones. Only pairs of strongly correlated neurons are set to 1. Such array is interpreted as the adjacency matrix of the network in panel (d).

We use the concept of time-varying graph (TVG) [35–37] to complete the construction of the time dependent FNN set. In the conventional graph theory, a graph is defined as  $G(\mathcal{V}, \mathcal{E})$  where  $\mathcal{V}$  and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  represent, respectively, the set of vertices (or nodes) and the set of edges  $e_{i,j}$  with  $i, j \in \mathcal{V}$ . A TVG is defined as  $G(\mathcal{V}, \mathcal{E}, \mathcal{T})$ , where  $\mathcal{T}$  defines which edges are available at each value of the discrete time variable  $\tau = \bar{\tau}/b$ . Actually, it is a generalization of a  $G(\mathcal{V}, \mathcal{E})$  graph with the inclusion of the edge presence function  $\mathcal{T} : \mathcal{E} \times \Gamma \rightarrow \{0, 1\}$ , where  $\Gamma \subseteq \mathbb{N}$  is a discrete time interval set called the system lifetime. According to this formalism, a FNN set is regarded as a single  $G(\mathcal{V}, \mathcal{E}, \mathcal{T})$  graph. To be consistent with our goal,  $\mathcal{T}$  must depend on a minimum amount  $\bar{\rho}$  of functional relationship among the neurons to be reached by the correlation coefficients  $\rho_{i,j}(\tau)$ . Therefore, we formally define

$$\mathcal{T}(e, \tau) = \begin{cases} 1, & \text{if } \rho_{i,j}(\bar{\tau}) \geq \bar{\rho}, \\ 0, & \text{if } \rho_{i,j}(\bar{\tau}) < \bar{\rho}, \end{cases} \quad \forall (e, \tau) \in (\mathcal{E}, \Gamma). \quad (2)$$

For the sake of definiteness we will call the obtained structure as a functional neuronal time varying graph (FNTVG). The main steps of our construction procedure are indicated in Figure 1. In Figure 2, we show the three subsequent elements of a FNTVG. In the next section, we use the following network measures to characterize all elements of a FNTVG during the graph lifetime  $\Gamma$ : the number of nodes connected to at least one node  $N_n(\tau)$ , the number of nodes in the largest component  $N_c(\tau)$ , the



**Fig. 2.** (Color online) Three elements (435, 436, and 437) of the *GE5* FNTVG showing how connections depend on instant firing rates. Nodes have been grouped according to their location in the brain: white  $\leftrightarrow$  hippocampus; blue (dark grey)  $\leftrightarrow$  somatosensory cortex; green (light grey)  $\leftrightarrow$  primary visual cortex. Besides this broad criterion, their relative positions have no correlation with their actual location inside the brain. Hubs defined according to equation (4) (yellow squares) also change with time.

total number of edges in the network  $E(\tau)$ , the diameter  $D(\tau)$ , the node degree  $k_i(\tau)$  and the node average mean path  $\ell_i(\tau)$ . The last two local measures can be worked out to provide both time-dependent network averages  $\langle z(\tau) \rangle$ , as well as time average of individual nodes  $\bar{z}_i$ , where  $z$  indicates  $k$  and  $\ell$ .

With such a *bona fide* FNTVG definition, we discuss important issues to implement the proposed framework. The first one is related to the proper time scales. With a typical firing time  $\delta (=10)$  ms, we get the maximal allowed firing rate  $\sim b/\delta$ . In our data, only in exceptional cases a neuron fires 4 or 5 times in a  $b = 100$  ms window. Therefore, the correlations among firing rates becomes very poor for such small values  $b \sim \delta$ . In fact, we still obtain an almost binary statistics. In the opposite limit, if we increase the value of  $b$  by a factor 10, the firing counts are 10 times as larger, but the typical scale of 250 ms for individual actions is surmounted by a factor 4. Therefore, the selected optimal value corresponds to  $(W(\text{ms}), b(\text{ms}), N_W) = (2500, 250, 10)$ . To illustrate the effects of other parameter choices we also report, in the next section, results for  $(W, b, N_W) = (1000, 100, 10)$ , and  $(2500, 100, 25)$ .

Moving windows depend both on the width as on the starting point, which we express as function of  $N_W$  and  $N_s$ , the distance between starting bins of two consecutive windows. For overlapping windows ( $N_s < N_W$ ), the same information is used repeatedly (as in a moving average), reducing statistical fluctuations. The cardinality of the set  $\Gamma$  is expressed as function of  $N_s$ ,  $b$ , and  $N_W$ , as  $|\Gamma| = (T/b - N_W + 1)/N_s$ . In this work, we restrict our analyzes to  $N_s = 1$ , so that  $|\Gamma| \sim 32\,000$  for all rats when  $(W, b, N_W) = (2500, 250, 10)$ .

Let us observe that a neuron  $i$  may not fire in the  $q$ th bin and remain in a silent state expressed by  $h_i(q) = 0$ . In the FNN construction, the presence of a large number of contiguous bins with  $h_i = 0$  may lead to a distorted correlation matrix in equation (1). Indeed, highly connected networks due to  $h_i = 0$  states would dominate the statistics, providing an undesired bias (for the ensemble of neurophysiologic states) that stems only from silence. There are several arguments pro and contra treating differently silence from active states in neuroscience. Although we

will not argue herein against any of them, we advise that the developed formalism seems to be tailored for neglecting the silent states. This becomes clear when we compare results where correlations were obtained under two different conditions: (a) imposing a very strict demand for correlations by demanding that more than 80% of the  $N_W$  values of  $h_i$  are non zero; (b) raising no objections to the presence of  $h_i = 0$  bins.

If  $h_i(q) = 0$  states are neglected, the effective number of elements in equation (1) is often smaller than  $N_W$ , i.e.  $N_W(\tau) < N_W$ . The measure of linear correlation provided by  $\rho_{i,j}(\tau)$  depends on the number of elements inside the considered window but, for large values of  $N_W$ , small fluctuations in the actual values of  $N_W(\tau)$  do not have much influence on the value of  $\rho_{i,j}$ . In such cases, the simplest strategy is to choose a constant value  $\bar{\rho} = R$  in equation (1), so that the additional step is to tune  $R$  to adequately extract relevant information from the data. Functional networks based on fMRI data [11] have used this strategy.

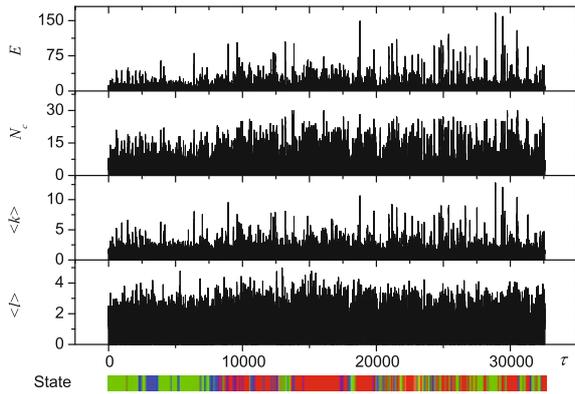
In the current study, not only  $N$  is much less than the number of fMRI voxels, but also  $N_W$  is restricted to the interval [10, 25]. Therefore it is wise adopt a significance level  $S$  of the correlation test that properly takes into account the effective sample size  $N_W(\tau)$ . For the sake of definitiveness, let us consider the classical statistical inference test for correlation [38] to set a hypothesis test for the correlation and a consequent significance level for the answer. The null hypothesis of the statistical test is: there is no correlation between the variables, or  $\rho_{i,j} = 0$ . Following the standard inferential statistics [38] the quantity  $\rho_{i,j}$  is associated with a  $t$  value of a Student distribution by the equation:

$$t = \rho_{i,j} \sqrt{(N_W - 2)/(1 - \rho_{i,j}^2)}. \quad (3)$$

Using this Student distribution, with  $(N_W - 2)$  degrees of freedom, we find the  $t$ -value of the test that can be interpreted as the probability that the results observed in a study could have occurred by chance if the null hypothesis was true. Let us denote the chosen significance level by  $S$ . Therefore, for  $t < S$  and a given  $N_W(\tau)$  we accept the null hypothesis. On the other hand, for  $t \geq S$ , we have no reason to conclude for an absence of correlation.

We compared our results with those obtained within the same framework by randomized surrogates originated by the following procedure: (1) each neuron  $i$  keeps its own set of firing rates; (2) for each value of  $i$ , the time order of firing rates  $h_i(q)$  was independently shuffled by sequentially choosing  $|\Gamma|$  random pairs  $(q_1, q_2)$  and switching  $(h_i(q_1), h_i(q_2)) \rightarrow (h_i(q_2), h_i(q_1))$ . This destroys any possible correlation among all neurons at a very low level.

Finally, we tested the legitimacy of using Pearson correlations in the current approach by evaluating  $Q(r)$ , the distribution of residues  $r$  obtained from the linear fits of our data. We have found that  $Q(r)$  can be well approximated by a normal Gaussian law.



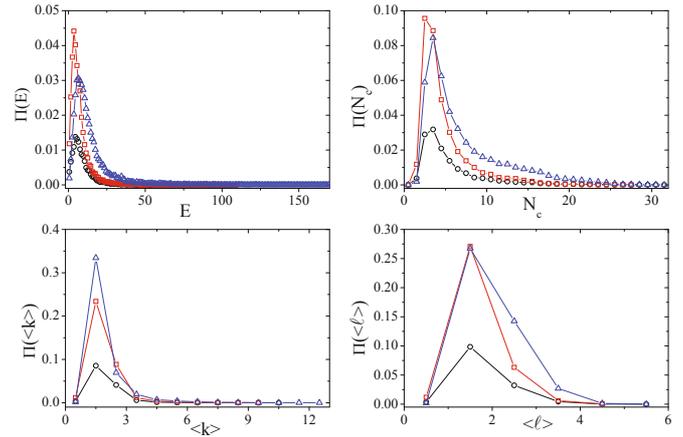
**Fig. 3.** (Color online) Values of  $E$ ,  $N_c$ ,  $\langle k \rangle$ , and  $\langle \ell \rangle$  as function of  $\tau$  for networks generated with data set  $GE6$ . Below the  $\tau$  axis, a gray (color) bar indicates the behavioral state (WK (dark gray, red), SWS(white, green), and REM(black, blue)) of the animal.

### 4 Results

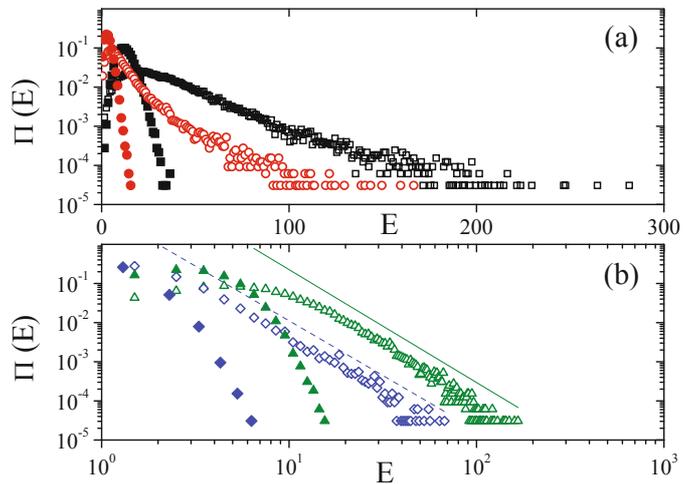
The elements of a FNTVG are such that  $\mathcal{V}$  depends on the original data set  $GE^*$ . Since  $N_s = 1$  always,  $\mathcal{E}$  and  $\mathcal{Y}$  depend on  $W, b$ , and  $\bar{\rho}$ . Very large values of  $\bar{\rho}$  set too restrictive conditions in equation (2), so that the FNTVG elements consist of a just a few connected nodes. By contrast, small values of  $\bar{\rho}$  do not distinguish the firing rate correlation, so that we just connect neurons that happen to be firing at the same time. Figure 3 shows selected network measures obtained from  $GE6$  data, with the optimal parameter values  $(W, b, N_W) = (2500, 250, 10)$ . We have discarded silent nodes and the value of  $\bar{\rho}$  is expressed by the significance level  $S = 0.99$ . The horizontal color bar indicates the animal activity as function of  $\tau$ .

The largest values attained by  $E$ ,  $N_c$ ,  $\langle k \rangle$ , and  $\langle \ell \rangle$  are 166, 30, 12, and 5. The behavior of  $\langle k \rangle$  and  $N_c$  makes it possible to identify periods where a large number of neurons are firing in a synchronized way (large  $\langle k \rangle$  and  $N_c$ ), as well as time intervals where most of them are at rest (low  $\langle k \rangle$  and  $N_c$ ). The association of synchronized firing rates with the values of  $\ell$  is not straightforward.  $\ell$  vanishes identically when no edge is present, and converges to 1 for a completely connected network. It goes through a maximum when the number of edges lies between the two limit cases, but this extreme does not depend only on  $E$ , but on the network topology and the way the edges are distributed among the nodes.

The corresponding probability distributions  $\Pi(x)$  of the FNTVG measures are presented in Figure 4. There we can observe, in a more convenient way, that the occurrence of these measures depends in a relative weak way on the behavioral state of the animal. The plots of  $\Pi(N_c)$ ,  $\Pi(\langle k \rangle)$ , and  $\Pi(\langle \ell \rangle)$  confirm that the corresponding measures are distributed over relatively short intervals with a characteristic scale. The same is not observed for the distribution of edges  $\Pi(E)$ , where the presence of long tails can be observed for all three distinct states. Thus, the FNTVG set  $\{E\}$  appears as the adequate record that might reveal fingerprints of a critical or a strong correlated



**Fig. 4.** (Color online) Probability distributions  $\Pi(E)$ ,  $\Pi(N_c)$ ,  $\Pi(\langle k \rangle)$ , and  $\Pi(\langle \ell \rangle)$  for the  $GE6$  data set. The results distinguish each of the assigned behavioral states in Figure 3: WK (blue triangles), SWS (red squares), and REM (black circles).



**Fig. 5.** (Color online) Distribution  $\Pi(E)$  for the  $GE6$  FNTVG with different values of  $S$ . In (a), squares and circles correspond to  $S = 0.95$  and  $0.99$ , respectively. Hollow and solid symbols indicate the results for the original data and shuffled surrogates. The mono-log scale reveals exponential tails for the shuffled data only. In (b), triangles and diamonds correspond to  $S = 0.990$  and  $0.999$ . Best fits to descending tails (solid and dashed lines) indicate power laws with exponents  $\gamma = 2.87$  and  $2.73$ , respectively.

behavior of the entire system. We investigated in detail the functional dependence of  $\Pi(E)$  during the  $\Gamma$  lifetime, taking into account the dependence of  $\Pi(E)$  on the significance level  $S$ . We remind that  $\Pi(E)$  is proportional to the number of occurrences where  $E$  pairs of neurons fire in a correlated way, i.e., how often in a freely behaving situation the animal requires its neurons to act together.

In Figure 5 we show, for different threshold values of  $S$ , the distributions  $\Pi(E)$  for the original and randomized data. Given the non random nature of the investigated data, it is fair to expect that  $\Pi(E)$  shows clear deviations from Poisson behavior are expected.

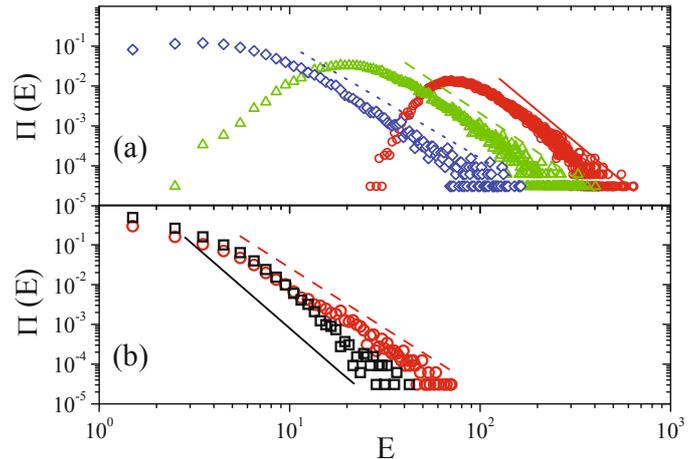
Panel (a) shows, in log-linear scale, the behavior  $\Pi(E)$  when  $S = 0.95$  and  $0.99$ . For the shuffled surrogates,  $\Pi(E)$  is characterized by exponential tails, while non-exponential fat tails characterize the distribution obtained by the actual data. This behavior becomes more evident when  $S$  increases, tails clearly deviate from an exponential decay even for the relatively small value  $S = 0.95$ . An overall effect produced by more restrictive control (larger  $S$ ) is to shift  $\Pi(E)$  to the left, for the average number of edges in the FNTVG becomes smaller. However, a proper tuning of the admitted level of Pearson correlation changes not only the average number of edges, but also reveals critical effects produced by neuron interaction.

Panel (b) illustrates, in log-log scale, the typical changes in the tails of  $\Pi(E)$  for larger values of  $S$ . We use the same value  $S = 0.99$  (as in panel (a)) and  $S = 0.999$ . Visual inspection suggests that both tails might be consistent with power-law behavior within a range of  $E \gtrsim$  one order of magnitude. We used a least square fitting procedure to adjust the data points in the distribution tail to a power law function  $\Pi(E) = AE^{-\gamma}$ . The results indicate that  $\gamma$  decreases as the significance  $S$  increases. This leads to the conclusion that, even if the tails do not strictly follow power laws, the decay of  $\Pi(E)$  slows down when  $S$  increases. We observe that  $S = 0.999$  stands as an upper bound. Still more restrictive conditions wash out important effects of actual interaction mechanism from the FNTVG, leading to an incomplete picture of the neuron system.

A quantitative assessment to the question whether  $\Pi(E)$  follows a power law was conducted by a non-parametric Kolmogorov-Smirnov (KS) test. This approach measures the likelihood that a given data distribution follows a chosen functional dependence. It has great importance in those cases where the range of variation of the experimental data is relatively small [39,40]. This is precisely the situation we face in this work. The largest number of links in a network with 57 nodes is  $\sim 1600$ , but the restriction imposed by large values of  $S$  reduces this number to 166 when  $S = 0.99$ , and 67 when  $S = 0.999$ .

When  $S = 0.999$ , we obtained a positive result of the KS test (score 0.16) when we compared the actual values of  $E$  with those of an artificially produced power law distribution with exponent  $\gamma = 2.7$  in the interval [5, 67]. This range of  $E$ , slightly larger than a decade, corresponds to the region of the distribution tail. The KS results remain valid for  $\gamma$  in the interval [2.6, 2.8]. We also obtained positive results for  $S = 0.99$ . The test scored 0.055 for the comparison of  $\Pi(E)$  with a power law distribution with  $\gamma = 2.8$  in the interval [13, 150]. When  $S \leq 0.95$ , the same procedure leads to negative outcomes of the KS test. To check the alternative behavior, we also performed the KS test in the opposite situation, i.e., we tested the tails of  $\Pi(E)$  for exponential decay  $A \exp(-\beta E)$ . The results were negative for any value  $S \in [0.9, 0.99]$  and suitable range of  $\beta$  values.

We performed further statistical tests and explored other parameter values to check our procedure. The results confirm that the FNTVG formalism is a suitable tool



**Fig. 6.** (Color online) (a) Distribution  $\Pi(E)$  for the FNTVG with the same data set and parameter values used in Figure 5, but with  $h_i = 0$  bins. The symbols indicate the same values of  $S$  as in Figure 5. New  $h_i = 0$  distributions are shifted to the right, but the tails decay with almost the same  $S$  dependent exponent  $\gamma$ . (b) Distribution  $\Pi(E)$  for the same data set in Figure 5 and  $S = 0.99$ . Squares and circles correspond to different choices of  $(W, b, N_W) = (1000, 100, 10)$  and  $(2500, 100, 25)$ .

for the analysis of neurophysiologic processes. This is illustrated by a discussion of the behavior of  $\Pi(E)$ , as it provides a clear signature of the FNTVG properties. We report results obtained when  $S = 0.99$ , since the larger average number of edges reduces the fluctuation magnitude.

To test the robustness of the FNTVG framework with respect to bootstrapping procedure, we considered random assemblies of elements of the original set. We generated several samples with different number  $N_B$ , for instance  $N_B = 30\,000$ ,  $25\,000$ , and  $20\,000$ . We repeated the procedure 5000 times for each such value of  $N_B$ , and noticed that the values of  $\gamma$  change in a narrow interval of amplitude  $< 0.05$  around the value obtained for the complete set.

In Figure 6a we show the main features of  $\Pi(E)$  with the presence of  $h_i = 0$  bins for  $S = 0.950$ ,  $0.990$ , and  $0.999$ , which can be directly compared to those shown in Figure 5. It is possible to note that: (i) the average number of edges increases, shifting  $\Pi(E)$  the right; (ii) the relative number of FNTVG elements with small number of links decreases, so that the new  $\Pi(E)$  shows a much sharper decrease in the region of small number of edges; (iii) the exponent of power-law tail of the new  $\Pi(E)$  remains very close to the previous one. Therefore, for the purpose of characterizing the power-law behavior, the presence of  $h_i = 0$  bins is almost irrelevant. This suggests further that the power-law tail is actually intrinsic to the system, not an artifact induced by the framework. Note also that the effect described in (ii) becomes more relevant when  $S$  is small. Finally, for the largest value  $S = 0.999$ , the distributions with  $h_i = 0$  bins become very close to that without  $h_i = 0$  bins for all values of  $E$ , with exception of a shift to the right. This indicates that, though abundant, silent bins are not sufficient to change the features of the large pair-wise correlations under very restrictive conditions.

This is the main reason why we indicated that neglecting  $h_i = 0$  bins in the devised framework is a more adequate choice. From now on we always discard  $h_i = 0$  bins.

Figure 6b illustrates the behavior of  $\Pi(E)$  for the choices  $(W, b, N_W) = (1000, 100, 10)$  and  $(2500, 100, 25)$ . The first choice weakens the average correlation and increases the number of FNTVG elements with a very small number of nodes and edges (1–10), where the complex network measures provide no useful statistical treatment. We note a strong increase in the value of  $\gamma$ . In the second choice, the effect of reducing  $b$  is compensated by concomitantly increasing  $N_W = 25$ , keeping  $W = 2500$  ms as in the optimal parameter values. The number of small networks lies between the two previous quoted conditions.  $\Pi(E)$  is shifted to the left, but the exponent of the distribution tail suffers only minor changes. This indicates that, although the considered measurements are localized in time, the patterns they produce via FNTVG are revealed only by coarse grained procedures.

Let us discuss some properties related of other FNTVG measures. In particular, the node degree distribution  $p(k, \tau)$  could identify a possible scale-free network topology associated to a critical state [11]. As shown in Figure 3, the values of  $k_i$  are quite small, so that it becomes difficult to identify any power law. Therefore, we develop another strategy to identify the presence of hubs and the existence of a hierarchical structure in the FNTVG.

To this purpose we used the value of  $k_i$  to classify the position of each neuron according to the firing activity for each value of  $\tau$ . Then we count the number of occurrences where each node  $i$  has a particular high value  $k_i$ , which measures its tendency to play the role of a hub. For each value of  $\tau$ , we identified hubs as nodes  $i$  satisfying the condition

$$k_i(\tau) > \langle k(\tau) \rangle + 2\sigma(\tau), \quad (4)$$

where  $\sigma(\tau)$  indicates the corresponding standard deviation of node degree.

We compared the observed results with those of a corresponding random network. Diversely from the usual approach, our FNTVG is not like a large network to which we can fit a Poisson curve to check for the randomness of the distribution, but contains thousands of elements with a number of connected vertices that varies from 2 to 56. Therefore we change our approach: we count the number of hubs in each FNTVG element using equation (4) and relate this result with the expected number of hubs of a Poisson distribution. The fraction of vertices satisfying equation (4) in a Gaussian distribution  $\sim 0.0228$ . The probability of the occurrence of two hubs will be the square of this value since these two events are independent. We will use these estimations for the Poisson distribution where the actual reference value depends on the  $\lambda$  free parameter which, in our case, corresponds to  $\langle k \rangle$ . Since  $\langle k \rangle \geq 1$ , we just have to verify whether the desired Poisson values are below the Gaussian estimate. Indeed, the expected Gaussian value that corresponds to a high  $\langle k \rangle$  is a Poisson upper bound, which makes our statistics conservative. We computed the number of hubs for the optimal parameter values and  $S = 0.99$ . To avoid a small network

bias we have performed our calculation only for networks with four or more nodes. If we include all elements of the FNTVG, the probability of occurrence of one and two hubs are, respectively,  $0.41 \pm 0.143$  and  $0.109 \pm 0.083$ . These results are indeed very high compared to the random estimations 0.0228 and 0.00053, which make us conclude that the presence of hubs among elements of our FNTVG is not random-like.

## 5 Discussion and final remarks

This work follows the general approach of previous investigations that used concepts and tools from statistical physics and complex systems to characterize records of brain activity. The approach based on fMRI measurements [11] and the statistics of spike avalanches in MEA records [41] provided sound evidences of collective behavior of distant neurons in the brain. The framework developed herein, based on similar methods of the former, advanced in characterizing several aspects of brain activity captured by the FNTVG elements: topology, hub structure, edge and degree distribution, minimum path features. All of these measures can be interpreted in terms of correlation among neuron activity in different brain areas.

It is important to recall that, although we concentrated the detailed discussion on the results for the rat *GE6*, we worked with recorded data of two further subjects (*GE4* and *GE5*), as stated in the first paragraph of Section 2. Despite the fact that the proportion of implanted electrodes in the distinct areas of the brain varies for each subject, we consistently obtained similar qualitative results for all data-sets. For all three subjects, we also considered sub-networks consisting only of nodes in each of the three brain regions. The network features are also very similar to those reported for the entire node set. However, as the number of nodes is at most the half of the entire set ( $< 30$ ), the results are subject to much larger fluctuations. Nevertheless, our results indicate that the proposed framework leads to similar properties, independently of the specific neuron type.

Deciphering the meaning of correlation in neuron activity is a major task in neuron sciences [30,31]. Correlations among firing neurons can be easily evaluated, but it is still difficult to identify a basic synchronization mechanism. One of the current approaches considers that different actions (movement, learning, thinking, etc.) may be explained by simple models involving two or more groups of neurons, but its validation is far from achieving consensus [31]. Based on firing rates rather than on single spikes, our analyzes lead to results of different nature compared to those from individual neuron activity. They can be used to detect neuronal ensemble activity of behavioral significance [20] resulting from an assembly of neuron states that last much longer than the typical firing time scale.

The most important results refer to fairly clear fat tails in  $\Pi(E)$  and the large number of hubs. All statistical tests and several parameter variations confirmed the existence of fat tails that converge to power law behavior when  $S \gtrsim 0.99$ . Unless we perform random shuffling

that completely destroys correlations, all effects related to changing parameter values in the FNTVG construction, sub-sampling with non overlapping bins, bootstrap, or inclusion of silent states are not able to remove this signature of long range correlation in the recorded signal. This suggests that tails are not an artifact, and that they express themselves only over a short interval as a consequence of the experimental set-up that restricts the number of nodes in the networks. As this power-law behavior can be seen as a new complex time dependent pattern that emerges from a critical behavior of a strong correlated system, we call it an *edge scale-free FNTVG*. In addition, as the number of possible edges grows with the square of the number of nodes, we can search for power-law distributions even if the number of recorded neurons in the FNTVG is only  $\sim 50$ .

The second quoted feature allowed for the identification of a large number of hubs and for noticing that the specific neurons playing this role dynamically changes with time. This hint to a further specific feature of brain operation: the place of hub is shared in a heterogeneous way among the pool of recorded neurons. What remains almost constant in the brain is not the actual function of individual neurons, but aspects of the network structure itself.

This may hint at a critical operating condition for the brain and touches an interesting discussion in neurophysiology about the criticality of the senses. It has been argued that sensory systems that work at criticality should optimize the amplification of the input signal as well as enlarge the input signal range [42,43]. Recent works have demonstrated the criticality in the statistics of neuronal spikes [44,45], but current experimental limitation restrict the number of recording neurons to  $\sim 200$  in these papers. Although the main objective of this work is not to discuss criticality in neurophysiology, our results hint positively to the question of the existence of criticality in electrophysiological neural data.

Finally, we stress that the FNTVG has a natural interpretation in neurophysiology, which is interested in tracking the activity of an ensemble of neurons during learning, behavioral tasks, or the sleep-wake cycle [44,46]. In this way, it is well suited to be a mathematical tool of choice for the investigation of brain activity extracted from a neuronal population.

This work was supported by the following Brazilian agencies: CNPq (individual grants and also through Instituto Nacional de Ciência e Tecnologia – Sistemas Complexos (INCT-SC), and Instituto Nacional de Ciência e Tecnologia – Interfaces Cérebro-Máquina (INCeMaq)), FINEP, FAPESB/PRONEX, FACEPE/PRONEX, FACEPE/PRONEM, and FAPERN. Financial support of the Pew Latin-American Program in Biomedical Science is also acknowledged.

## References

- R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002)
- M.E.J. Newman, *SIAM Rev.* **45**, 167 (2003)
- S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, 2003)
- M.E.J. Newman, A.-L. Barabási, D.J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, 2006)
- S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, *Phys. Rep.* **424**, 175 (2006)
- L.F. Costa, F.A. Rodrigues, G. Travieso, P.R. Villas Boas, *Adv. Phys.* **56**, 167 (2007)
- Principles of Neural Science*, edited by E.R. Kandel, J.H. Schwartz, T.M. Jessel, 4th edn. (McGraw Hill, New York, 2000)
- E. Bullmore, O. Sporns, *Nat. Rev. Neurosci.* **10**, 186 (2009)
- O. Sporns, D.R. Chialvo, M. Kaiser, C.C. Hilgetag, *Trends Cogn. Sci.* **8**, 418 (2004)
- O. Sporns, G. Tononi, R. Kötter, *PLoS Comput. Biol.* **1**, e42 (2005)
- V.M. Eguíluz, D.R. Chialvo, G.A. Cecchi, Marwan Baliki, A. Vania Apkarian, *Phys. Rev. Lett.* **94**, 018102 (2005)
- D. Fraiman, P. Balenzuela, J. Foss, D.R. Chialvo, *Phys. Rev. E* **79**, 061922 (2009)
- P. Bonifazi, M. Goldin, M.A. Picardo, I. Jorquera, A. Cattani, G. Bianconi, A. Represa, Y. Ben-Ari, R. Cossart, *Science* **326**, 1419 (2009)
- L. de Arcangelis, C. Perrone-Capano, H.J. Herrmann, *Phys. Rev. Lett.* **96**, 0281071-4 (2006)
- A. Levina, J.M. Herrmann, T. Geisel, in *Advances in Neural Information Processing Systems*, edited by Y. Weiss, B. Schölkopf, J. Platt (MIT Press, Cambridge, MA, 2006), Vol. 18, pp. 771–778
- A. Levina, J.M. Herrmann, T. Geisel, *Natl. Phys.* **3**, 857 (2007)
- L. de Arcangelis, H.J. Herrmann, *Proc. Nat. Acad. Sci.* **107**, 3977 (2010)
- C. Sherrington, E.D. Adrian, *Nobel Lectures, Physiology or Medicine* (Elsevier Publishing Company, Amsterdam, 1965), pp. 1922–1941
- J. Kralik, D. Dimitrov, D. Krupa, D. Katz, D. Cohen, *Methods* **25**, 121 (2001)
- S. Ribeiro, D. Gervasoni, E.S. Soares, Y. Zhou, S.C. Lin, P. Pantoja, M. Lavine, M.A.L. Nicolelis, *PLoS Biol.* **2**, 126 (2004)
- S. Ribeiro, X. Shi, M. Engelhard, Y. Zhou, H. Zhang, D. Gervasoni, S.C. Lin, K. Wada, N.A. Lemos, M.A.L. Nicolelis, *Front. Neurosci.* **1**, 43 (2007)
- M.A.L. Nicolelis, D. Dimitrov, J.M. Carmena, R. Crist, G. Lehew, J.D. Kralik, Steven P. Wise, *Proc. Natl. Acad. Sci. USA* **100**, 11041 (2003)
- D. Gervasoni, S.C. Lin, S. Ribeiro, E.S. Soares, J. Pantoja, M.A.L. Nicolelis, *J. Neurosci.* **24**, 11137 (2004)
- M.A.L. Nicolelis, M.A. Lebedev, *Nat. Rev. Neurosci.* **10**, 530 (2009)
- P. Dayan, L.F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, 2001)
- M.I. Posner, *Chronometric explorations of Mind* (Oxford University Press, USA), 1986
- E. Brenner, J.B.J. Smeets, *J. Mot. Behav.* **29**, 297 (1997)
- N.A.P. Vasconcelos, J. Pantoja, H. Belchior, F.V. Caixeta, J. Faber, M.A.M. Freire, V.R. Cota, E.A. de Macedo, D.A. Laplagne, H.H. Gomes, S. Ribeiro, *Proc. Natl. Acad. Sci. USA* **108**, 15408 (2011)

29. N.A.P. Vasconcelos, W. Blanco, J. Faber, H.M. Gomes, T.M. Barros, S. Ribeiro, in *The Relevance of the Time Domain to Neural Network Models*, edited by A.R. Rao, G.A. Cecchi (Springer Series in Cognitive and Neural Systems, Springer, 2011)
30. I.H. Stevenson, J.M. Rebesco, L.E. Miller, K.P. Körding, *Curr. Opin. Neurobiol.* **18**, 582 (2008)
31. J.W. Pillow, J. Shlens, L. Paninski, A. Sher, A.M. Litke, E.J. Chichilnisky, E.P. Simoncelli, *Nature* **454**, 995 (2008)
32. E. Schneidman, M.J. Berry, R. Segev, W. Bialek, *Nature* **440**, 1007 (2006)
33. G. Tkačik, E. Schneidman, M.J. Berry, W. Bialek, [arXiv:q-bio.NC/0611072](https://arxiv.org/abs/q-bio.NC/0611072), (2006)
34. E. Ganmor, R. Segev, E. Schneidman, *J. Neurosci.* **31**, 3044 (2011)
35. N. Santoro, W. Quattrociochi, P. Flocchini, A. Casteigts, F. Amblard, in *3rd AISB Social Networks and Multiagent Systems Symposium (SNAMAS)*, 2011, edited by AISB (York, UK), pp. 32–38
36. W. Quattrociochi, F. Amblard, E. Galeota, *Social Network Analysis and Mining* **2**, 229 (2012)
37. V. Kostakos, *Physica A* **388**, 1007 (2009)
38. R.R. Sokal, F.J. Rohlf, *The Principles and Practices of Statistics in Biological Research* (W.H. Freeman and Co., New York, 1995)
39. A. Clauset, C.R. Shalizi, M.E.J. Newman, *SIAM Rev.* **51**, 661 (2009)
40. A. Klaus, S. Yu, D. Plenz, *PLoS ONE* **6**, e19779 (2011)
41. T.L. Ribeiro, M. Copelli, F. Caixeta, H. Belchior, D.R. Chialvo, M.A.L. Nicolelis, S. Ribeiro, *PLoS ONE* **5**, 14129 (2010)
42. O. Kinouchi, M. Copelli, *Nat. Phys.* **2**, 348 (2006)
43. W. Shew, H. Yang, T. Petermann, R. Roy, D. Plenz, *J. Neurosci.* **29**, 15595 (2009)
44. S. Ribeiro, M.A.L. Nicolelis, *Learn. Memory* **11**, 686 (2004)
45. T. Petermann, T.C. Thiagarajana, M.A. Lebedev, M.A.L. Nicolelis, D.R. Chialvo, D. Plenza, *Proc. Natl. Acad. Sci.* **106**, 15921 (2009)
46. M.A. Wilson, B.L. McNaughton, *Science* **265**, 676 (1994)