

Linear processes in high-dimension: phase space and critical properties

Iacopo Mastromatteo¹, Emmanuel Bacry¹, and Jean-François Muzy^{1,2}

¹Centre de Mathématiques Appliquées, CNRS, École Polytechnique,
UMR 7641, 91128 Palaiseau, France

²Laboratoire Sciences Pour l'Environnement, CNRS, Université de Corse,
UMR 6134, 20250 Corté, France

Abstract

In this work we investigate the generic properties of a stochastic linear model in the regime of high-dimensionality. We consider in particular the Vector AutoRegressive model (VAR) and the multivariate Hawkes process. We analyze both deterministic and random versions of these models, showing the existence of a stable and an unstable phase. We find that along the transition region separating the two regimes, the correlations of the process decay slowly, and we characterize the conditions under which these slow correlations are expected to become power-laws. We check our findings with numerical simulations showing remarkable agreement with our predictions. We finally argue that real systems with a strong degree of self-interaction are naturally characterized by this type of slow relaxation of the correlations.

1 Introduction

Stochastic linear models are instruments of paramount importance for describing physical, social or economic systems: despite being simple enough to be analytically tractable, they allow to accurately describe a large number of qualitatively different systems. In this work, our focus will be the study of linear models in the regime of high dimensionality, and the analysis of the effects induced by the collective regime of interactions on the overall stability and on the first and second-order properties of the system. Such analysis is related to multiple fields of activity previously considered in the literature:

- The properties of large ecosystems close to the equilibrium have been studied with similar methods, finding that universality controls the stability of large food webs [1]. Our analysis extends in particular the results obtained in [2], which consider the *dynamic* version of May's seminal model [1], to the case of a symmetric interaction kernel (see Sec. 4.1).
- Linear models for self-exciting marked point-processes are customarily used in order to model seismic activity [3] (the mark being used to account for the intensity of the activity). This work generalizes this approach by suggesting that

it is possible to see the components $i \in \{1, \dots, N\}$ as a set of *interacting* marks. From this perspective, our results suggest that the critical relaxation measured in these systems might be endogenously generated by the interaction among a large number of marks.

- Multivariate linear models are standard tools in order to account for the cross-correlation among economic variables [4], and in particular linear self-exciting point process can be used in order to assess the risk of contagion in a financial network [5].
- Linearly interacting point-process are used in order to model neural networks [6] made by a large number of components (although non-linear generalizations are customarily used in order to model inhibition).

In all these cases, one is interested in characterizing the statistical properties of a system composed of a large number $N \gg 1$ of linearly interacting entities $X(t) = \{X_i(t)\}_{i=1}^N$, t being interpreted as the time and i as an index labeling the different entities. A major concern is the limiting behavior of the system in the regime of high-dimensionality, as the limit $N \rightarrow \infty$ is expected to be fixed by universality. We shall effectively address this issue through three angles, namely *Stability*, *Endogeneity* and *Relaxation*. Let us describe more clearly these perspectives:

Stability. An exogenous noise can drive the system towards the instability point characterized by the divergence of one or more components of $X(t)$. Indeed, the interaction network generates feedback loops which may enhance the susceptibility of the system to external perturbations. Hence, we want to determine the critical interaction level beyond which such instability effects arise, dominating the behavior of the system.

Endogeneity. We will be interested in characterizing what proportion of the *average intensity*¹

$$\Lambda_i = \frac{\langle dX_i(t) \rangle}{dt} \quad (1)$$

is related to exogenous factors and what proportion derives from endogenous cross-contamination effects among the components.

Relaxation. As we will consider the components of the system to be individually characterized by short memory and fast relaxation, it will be interesting to study whether the lagged cross-correlations

$$c_{ij}(\tau) = \frac{\langle dX_i(t)dX_j(t+\tau) \rangle - \langle dX_i(t) \rangle \langle dX_j(t+\tau) \rangle}{dt^2} \quad (2)$$

display or not slow relaxation in the limit of large τ as a consequence of the collective regime of interactions.²

Strongly endogenous effects are commonly found in applications along with slow relaxation of the auto-correlation (between the same components of $X(t)$) or the cross-correlation (between different components of $X(t)$) functions. This is, for instance, the

¹ With a slight abuse of language, we will refer to Λ as to the vector of average intensities for both processes that we will consider (VAR and Hawkes), see also the footnote of page 6.

² While in principle one could quantify the amount of endogenous interaction by focusing on the cross-correlations, we choose to adopt the average intensity as a proxy. The results of Secs. 3, 4.1 and 4.2 show that both quantities appropriately signal the overall degree of self-interaction in the system.

case of high-frequency dynamics in financial markets. A growing number of empirical works, describing by means of linear models this high-frequency dynamics, are in fact revealing that financial correlations are characterized by long-range memory, signaled by power law relaxations with small exponents [7, 8, 9, 10]. This behavior can in principle be justified on the basis of the self-reflexive character of financial markets, by arguing that, as a consequence of the strongly endogenous regime in which markets operate, slow relaxation can be induced in the system. From this perspective, the results of Refs. [11, 12] can be used in order to investigate analytically the onset of such a critical regime under the implicit assumptions that each of the entities composing the system is *individually* poised at criticality. Our aim is to discuss the complementary scenario in which long-range correlations in the system are induced *as an effect of its interaction network*, by assuming that no component in the system singularly exhibits critical behavior. More precisely, we shall suppose that individual units are fastly relaxing and try to characterize the influence of the network structure on the long-time behavior ($\tau \rightarrow +\infty$) of the correlations $c_{ij}(\tau)$ in the limit $N \rightarrow +\infty$. The limit when the system can collectively encode *long-memory*, signaled by the divergence of the quantity $\hat{c}_{ij}(0) = \int_0^\infty d\tau c_{ij}(\tau)$ will be referred to as the issue of **Criticality** of the system.

Let us notice that, as a consequence of the assumption that individual units are fastly relaxing, the regime in which the $\tau \rightarrow \infty$ limit is taken before the $N \rightarrow \infty$ limit becomes trivial. On the other hand, taking the $N \rightarrow \infty$ limit before any other limit will allow us to access a very rich phenomenological behavior.

The organization of the paper is the following. We introduce our main framework for linear systems in continuous time in Section 2. VAR models are first introduced, followed by Hawkes processes. Then, we restrict our study to the particular case of factorizable linear systems. At the end of this Section, we present the methodology for analyzing these systems that will be used all along the paper on each model. Section 3 studies the case a deterministic regular network of interactions whereas Sections 4.1 and 4.2 study the case of random models within the framework of two tractable random matrix ensembles: the Gaussian Orthogonal Ensemble and the Regular Random Graph Ensemble. We draw our conclusions in Section 5, while the more technical parts of the discussion are relegated to the Appendices.

2 Linear systems in continuous time

In this preliminary section we define the classes of linear models on which we focus our analysis. Beside fixing the notations and the conventions that we will follow throughout the paper, we aim to highlight the main similarities among the two linear models that will be presented. In particular, we want to show that the relations defining their behavior are formally identical and are essentially due to their linear nature. More general linear models can be expected to share the same behavior, as long as relations such as (7), (10) and (11) hold.

2.1 Vector Autoregressive model

The first type of model that we are going to consider is an N -dimensional Vector AutoRegressive Model (VAR), a widespread model introduced in [4] and customarily used in econometrics in order to describe the dynamic relationship among the different components of an economic time-series (for a more exhaustive account of the vast literature concerning VAR processes, we address the reader to the surveys [13, 14]). The model is defined by a set of N processes $X(t) = \{X_i(t)\}_{i=1}^N$ evolving in discrete time, driven by Gaussian noises $\eta(t) = \{\eta_i(t)\}_{i=1}^N$, and interacting through a linear matrix kernel $\Phi(\tau)$. In vector notation, the process is defined by the relation

$$X(t) = \sum_{t'=-\infty}^{t-1} \Phi(t-t')X(t') + \eta(t) , \quad (3)$$

where we assume the interaction kernel $\Phi(\tau)$ to admit the discrete Fourier transform

$$\hat{\Phi}(\omega) = \sum_{\tau=-\infty}^{+\infty} e^{-i\omega\tau} \Phi(\tau) , \quad (4)$$

with components $\hat{\Phi}(\omega) = \{\hat{\Phi}_{ij}(\omega)\}_{i,j=1}^N$. We restrict ourselves to systems for which the interaction kernel $\Phi(\tau)$ is causal, i.e.,

$$\Phi(\tau) = 0, \quad \forall \tau < 0. \quad (5)$$

Stability of VAR models. Most of the time we shall consider that Φ satisfies the following so-called stability assumption :

- **(H1) Stability Assumption.** The spectral radius of $\hat{\Phi}_0$ is smaller than 1, which we can indicate by $||\hat{\Phi}_0|| < 1$. Equivalently, all the eigenvalues of $\hat{\Phi}_0$ have modulus smaller than one.

Indeed, under this last assumption, it is possible to prove (see e.g. [13]) that the infinite sum

$$\Psi(\tau) = 1 + \Phi(\tau) + \Phi(\tau) * \Phi(\tau) + \Phi(\tau) * \Phi(\tau) * \Phi(\tau) + \dots \quad (6)$$

is well-defined, and its Fourier transform can be written as the matrix

$$\hat{\Psi}(\omega) = (\mathbb{I} - \hat{\Phi}(\omega))^{-1} , \quad (7)$$

where \mathbb{I} denotes the N -dimensional identity matrix. The process $X(t)$ can then be written as the convolution

$$X(t) = \Psi(t) * \eta(t) , \quad (8)$$

where, by convention, the $*$ operator refers to the regular matrix product where all multiplications have been replaced by discrete convolutions. The process $X(t)$ is proved to admit a stationary state, whose associated probability measure expectation will be denoted by the symbol $\langle \dots \rangle$. We will be interested in computing first and second order properties of the process under such measure whenever stability assumption hold.

Continuous time VAR models. In order to emphasize the analogy with a Hawkes process, in the next parts of the discussion we will be considering the continuous-time version of this model, also employed in [15, 16] in order to model high-frequency financial data. In this case, Eq. (8) can be generalized to the continuous case by replacing the discrete convolution with a continuous one. In particular, we can define the continuous-time VAR by substituting the Gaussian noises $\eta(t)$ with a Wiener process with increments $d\eta(t)$. The continuous time process satisfies, by construction, the same properties of its discrete-time counterpart once we define the continuous-time Fourier transform as

$$\hat{\Phi}(\omega) = \int_{-\infty}^{+\infty} d\tau e^{-i\omega\tau} \Phi(\tau). \quad (9)$$

In particular, the analytical formulae for the average Λ and the cross-correlation matrix $c(\tau)$ are given by the equations below.

Endogeneity of VAR models. The expectation $\Lambda dt = \langle dX(t) \rangle$ can be expressed as

$$\Lambda = \hat{\Psi}(0)\mu, \quad (10)$$

where we write $\langle d\eta(t) \rangle = \mu dt$. The relation among endogenous and exogenous effects is determined by the matrix $\Psi(0)$, whose spectral norm determines the maximum output intensity Λ as a function of the norm of the driving vector μ .

Correlation of VAR models. The cross-correlation matrix

$$c(t-t')dtdt' = \langle dX(t)dX^T(t') \rangle - \langle dX(t) \rangle \langle dX^T(t') \rangle$$

can be obtained through the Fourier transform $\hat{c}(\omega)$, which is given by

$$\hat{c}(\omega) = \hat{\Psi}^*(\omega)\Sigma\hat{\Psi}^T(\omega), \quad (11)$$

where the symbol T denotes matrix transposition and $*$ indicates complex conjugation. We have assumed the covariance term Σ to be defined by the relation

$$dtdt'\Sigma\delta(t-t') = \langle d\eta(t)d\eta^T(t') \rangle - \langle d\eta(t) \rangle \langle d\eta^T(t') \rangle, \quad (12)$$

where $\delta(\tau)$ indicates the Dirac delta function. Moreover, we assumed the matrix Σ to be diagonal.

2.2 Hawkes process

The second class of models that we consider are Hawkes processes, a class of interacting point processes customarily used to describe self and cross-excitation phenomena [17, 18]. For a long time, Hawkes models have been extensively used to describe the occurrence of earthquakes in some given region [19, 20]. They are getting more and more popular in many other applications in which tracking how information diffuses through different “agents” is the main concern, e.g., neurobiology (neurons activity) [21], sociology (spread of terrorist activity) [22, 23] or processes on the internet (viral diffusion across social networks) [24, 25]. Their application to finance can be

traced back to Refs [26, 27], and has been followed by a still-ongoing spree of activity [28, 29, 30, 31, 32, 7, 8, 33, 34, 10, 9].

An N -dimensional Hawkes process is defined by a set of N counting processes evolving in continuous time $X(t) = \{X_i(t)\}_{i=1}^N$ ³. The probability for an event to be triggered is expressed by a stochastic intensity function $\Lambda(t) = \{\Lambda_i(t)\}_{i=1}^N$ which evolves according to the dynamics:

$$\Lambda(t) = \mu + \int_{-\infty}^t \Phi(t - t') dX(t'), \quad (13)$$

where the components of μ are commonly referred as *exogenous intensities* (or *baseline intensities*), and $\Phi(t)$ is a positive semidefinite, causal (in the sense of (5)), locally L^1 matrix kernel. Notice that, unlike in the VAR case, in order for the event probabilities $\Lambda(t)dt$ to be well-defined, we need to assume that

$$\Phi(\tau) \text{ satisfies component-wise positivity.}$$

Stability of Hawkes processes. If a Hawkes process satisfies the stability assumption (H1) specified above for the VAR model, then one can show that $X(t)$ is stationary and stable [17, 18]. Moreover, as in the VAR model, this condition implies that the infinite sum $\Psi(\tau)$ (Eq. (6)) is well-defined, and its Fourier transform $\tilde{\Psi}(\omega)$ is given by Eq. (7).

Endogeneity of Hawkes processes. As for the VAR model, the mean intensity $\langle \Lambda(t) \rangle dt = \langle dX(t) \rangle = \Lambda dt$ is expressed by Eq. (10). Again, the ratio of endogeneity versus exogeneity is determined by the matrix $\Psi(0)$, which sets the relation among endogenous intensity Λ and exogenous intensity μ .

Correlation of Hawkes processes. Again, as for the VAR model, the cross-correlation matrix

$$c(t - t') dt dt' = \langle dX(t) dX^T(t') \rangle - \langle dX(t) \rangle \langle dX^T(t') \rangle$$

can be expressed by Eq. (11) in which Σ is the diagonal matrix defined by (see [7])

$$\Sigma_{ij} = \Lambda_i \delta_{ij}, \quad (14)$$

where δ_{ij} stands for the Kronecker symbol, equal to one if $i = j$ and zero otherwise.

2.3 Hawkes processes versus VAR processes.

The previous results indicate that a Hawkes process is very reminiscent of a VAR process formulated in continuous time with non-negative $\Phi(\tau)$, although some important differences need to be emphasized.

³With abuse of notation, we are denoting quantities describing the VAR model with the same symbols adopted for the Hawkes process. We choose to do so in order to show more transparently the close relation among the two models, which satisfy extremely similar relations. We will specify explicitly which of the two frameworks we are considering whenever this notation results ambiguous.

First, while the μ in the autoregressive case identify the average increments of a multivariate Wiener process, in the Hawkes case they emerge as the exogenous component of the average intensity.

Secondly, in the autoregressive case the covariance of the noise Σ controlling the cross-correlations is independent of $\Phi(\tau)$ and μ , while in the Hawkes case it is *endogenously generated* and is thus fixed to the values $\Sigma_{ii} = \Lambda_i$.

Finally, the cross-correlation $c(\tau)$ of a Hawkes process is always singular: as we are considering counting processes with jumps of size one, the relation $dX_i(t) = (dX_i(t))^2$ holds. Then, the cross correlations can be decomposed as $c_{ij}(\tau) = \Lambda_i \delta_{ij} \delta(\tau) + c_{ij}^{(reg)}(\tau)$ where $c_{ij}^{(reg)}(\tau)$ is regular around zero. For VAR processes, such a singular contribution to the correlations emerges just if $\langle d\eta(t) d\eta^T(t') \rangle - \langle d\eta(t) \rangle \langle d\eta^T(t') \rangle$ contains a singular component centered at zero.

Let us point out that, with a slight abuse of language, we will refer to Λ and μ respectively as mean intensities and exogenous intensities even in the case of a VAR model.

2.4 Factorizable linear systems

2.4.1 An isotropical factorizable interaction kernel

As we are interested in studying how heterogeneity in the dynamic behavior of the system may emerge as an effect of the interaction network, we take models for which the interaction kernel $\Phi(\tau)$ can be factorized as

$$\Phi(\tau) = \alpha \phi(\tau) , \quad (15)$$

where α is a matrix and $\phi(\tau)$ is a scalar function of τ satisfying (without loss of generality)

$$\hat{\phi}(0) = \int_0^\infty d\tau \phi(\tau) = 1.$$

We additionally suppose that

$$\int_0^\infty d\tau \tau \phi(\tau) < \infty,$$

as we want to focus on the problem of whether long-range memory in correlations can be *endogenously* induced by a system whose interactions have short-range memory. Hence, we are supposing the different components of the system to react homogeneously in time (i.e., with the same speed) to innovations, while we confine the heterogeneity of the system to the interaction strengths, which are specified by the matrix α . Moreover, the α matrix is naturally interpreted as a weighted adjacency matrix specifying the overall strength of the interaction between the different components of the system⁴. We will consider exclusively cases for which the matrix α can be diagonalized, so that we will always be able to write

$$\alpha = U \lambda U^{-1} , \quad (16)$$

⁴Let us point a factorization assumption like (15) is very common in space-time Hawkes models for earthquakes [19].

where λ is a diagonal matrix of eigenvalues with elements equal to $\{\lambda_a\}_{a=1}^N$ and U is a suitable change of basis matrix, with entries denoted by $\{U_{ia}\}_{ia=1}^N$. Note that we are adopting the convention of using the indices i, j, \dots to denote the components of $X(t)$ and a, b, \dots for the eigenvalues of the α matrix. For simplicity, We will further require some supplementary conditions to hold for the system, namely

- **(H2) Unitarity Assumption.** The matrix U is assumed to be unitary (i.e. $U^\dagger = U^{-1}$, where † denotes Hermitian conjugation).
- **(H3) Homogeneity Assumption.** All the components of Σ (defined by (12) in the case of the VAR model and by (14) in the case of the Hawkes model) are assumed to be equal to $\bar{\Sigma}$, and the components of the the mean intensity vector are equal to $\bar{\Lambda}$.⁵

The diagonalization assumption (16) allows to write the components of the matrix $\hat{\Psi}(\omega)$ as

$$\hat{\Psi}(\omega) = \hat{\phi}(\omega)^{-1} U \left(\hat{\phi}(\omega)^{-1} - \lambda \right)^{-1} U^{-1}. \quad (17)$$

We will often employ this decomposition of $\hat{\Psi}(\omega)$ in order to link the distribution of eigenvalues of the matrix α with the value of the observables Λ and $c(\omega)$.

The unitarity **(H2)** and homogeneity **(H3)** assumptions above allow us to disregard the effect on the system of the heterogeneity in the angular components of the interaction matrix, permitting us to focus on the collective effect induced by the isotropical part of the interaction. Even though the global effect of the inhomogeneity of the system is an interesting problem on his own, the simpler case that we consider is a necessary first step in order to understand the collective effects of the large N limit on this type of systems.

In the homogeneous case, it will be useful to introduce the notation

$$\bar{\Lambda} = \frac{1}{N} \sum_i \Lambda_i \quad (18)$$

$$\overline{c(\tau)} = \frac{1}{N} \sum_i c_{ii}(\tau) \quad (19)$$

$$\overline{\overline{c(\tau)}} = \frac{1}{N^2} \sum_{ij} c_{ij}(\tau), \quad (20)$$

as all the information about the vector of average intensity and the lagged cross-correlation matrix is encoded in these averages over components.

Under the unitarity **(H2)** and homogeneity **(H3)** assumptions, it is easy to show that Eq. (11) leads to

$$\overline{\hat{c}(\omega)} = \frac{1}{N} \sum_i \hat{c}_{ii}(\omega) = \frac{\bar{\Sigma}}{N} \sum_a \frac{1}{\|\hat{\phi}(\omega)^{-1} - \lambda_a\|^2}, \quad (21)$$

⁵While in the translationally invariant case discussed in Sec. 3 the system will fulfill by construction the homogeneity assumption defined above, in the random cases presented in Secs. 4.1 and 4.2 this assumption will hold on average, allowing the system to enjoy the same properties.

so that the information about the eigenvectors encoded in U is not required in order to understand the behavior of the average over components of the autocorrelation function. **The behavior of autocorrelations is completely specified through the function $\phi(t)$ and the spectrum of the matrix α .**

2.4.2 Case of an exponential factorizable kernel

Finally, we want to introduce an useful benchmark case for the kernel function $\phi(\tau)$, which will be employed in order to characterize the long-time behavior of the system. Specifically, we will often particularize our results to the special case in which the interaction kernel has an exponential shape of the form

$$\phi(\tau) = \beta e^{-\beta\tau} \mathbb{1}_{\mathbb{R}^+}(\tau),$$

or equivalently

$$\hat{\phi}(\omega) = (1 + i\omega/\beta)^{-1},$$

where $\mathbb{1}_{\mathbb{R}^+}(\tau)$ denotes the indicator function on \mathbb{R}^+ . This particularly simple case will allow us to explore the main qualitative features of the results without losing analytical control on the solution. In this framework, the locations of the poles appearing in the Fourier transforms of the correlation function (11) are trivially dictated by the eigenvalues of the matrix α . The poles can be written as

$$\omega_a = i\beta(1 - \lambda_a) \quad \omega_a^* = -i\beta(1 - \lambda_a^*), \quad (22)$$

where the $\{\lambda_a\}_{a=1}^N$ denote as usual the eigenvalues of α .

Relaxation. In the exponential case, inverting the Fourier transforms appearing in Eq. (11) becomes straightforward, if one supposes the eigenvalues to be non-degenerate. One obtains in fact

$$\begin{aligned} c_{ij}(\tau) &= \beta\theta(-\tau) \sum_{abk} \frac{\lambda_a^*(2 - \lambda_a^*)}{2 - \lambda_b - \lambda_a^*} U_{ia}^* U_{ak}^{*-1} \Sigma_k U_{bk}^{-1} U_{jb} e^{\beta[1 - \lambda_a^*]\tau} \\ &+ \beta\theta(\tau) \sum_{abk} \frac{\lambda_b(2 - \lambda_b)}{2 - \lambda_b - \lambda_a^*} U_{ia}^* U_{ak}^{*-1} \Sigma_k U_{bk}^{-1} U_{jb} e^{-\beta[1 - \lambda_b]\tau}. \end{aligned} \quad (23)$$

Under **(H2)** and **(H3)**, above equation leads to a simplified form of the average autocorrelation, which – just as Eq. (21) above – is independent of the eigenvectors of α :

$$\overline{c(\tau)} = \frac{1}{N} \sum_i c_{ii}(\tau) = \frac{\beta\bar{\Sigma}}{N} \sum_a \frac{\lambda_a(2 - \lambda_a)}{(2 - \lambda_a - \lambda_a^*)} e^{-\beta[1 - \lambda_a]|\tau|}, \quad (24)$$

Each eigenvalue of α generates a decay mode indexed by a whose associated speed depends upon the distance between the corresponding λ_a and 1. In particular, the slowest mode $\lambda_{max} < 1$ controls the behavior of the correlations at large times, fixing the scaling of the correlation at large times to be proportional to $c_{ii}(\tau) \sim \exp(-\beta(1 - \lambda_{max})|\tau|)$. The *critical* regime in which the support of the distribution of λ_a touches

the instability point $\lambda = 1$ is of particular interest, and will be considered in specific sections (Secs. 3.4, 3.4.2, 4.1.5 and 4.2.4).

Genericity of exponential kernel. The main interest in considering the exponential kernel lies in the fact that it allows one to explore the long-time behavior of *any* short-range kernel. In fact, consider a generic kernel satisfying the *short-range* assumption

$$\int_0^\infty d\tau \tau \phi(\tau) < \infty . \quad (25)$$

This condition corresponds in Fourier space to the differentiability in zero of the function $\hat{\phi}(\omega)$, implying that it is possible to expand it as

$$\hat{\phi}(\omega) = 1 + i\omega/\beta + o(\omega) , \quad (26)$$

with $\lim_{\omega \rightarrow 0} o(\omega)/\omega = 0$. By back-transforming Eq. (11) in real space while keeping into account this expansion, one finds that the leading order term of the large time expansion of the cross-correlations is given by Eq. (23). Intuitively, as the limit of large times $\tau \rightarrow \infty$ corresponds to the small ω regime in Fourier space, this indicates that at large times one can neglect the $o(\omega)$ term in the expansion of the kernel, sticking with the term $1 + i\omega/\beta$ which corresponds in real space to the exponential kernel

$$\phi(\tau) = \beta e^{-\beta\tau} \mathbb{1}_{\mathbb{R}^+}(\tau) . \quad (27)$$

2.5 Methodology for analyzing factorizable linear systems

The next sections will analyze the behavior of factorizable linear systems (as defined in Section 2.4.1), addressing systematically, as explained in the introduction, the issues of stability, endogeneity and relaxation by focusing on a small set of relevant scalars encoding the collective information about the state of the system. In order to do so, we shall study the behavior of different observables. More precisely :

- **Stability: The spectral norm of α .** Due to Eq. (10), the value of the average intensities Λ_i for a fixed value of the input μ_i is determined by the largest eigenvalue of the matrix $\hat{\Psi}(0)$, which is precisely equal to $1/(1 - \|\alpha\|)$. In particular, for $\|\alpha\| = 1$ (i.e., the largest eigenvalue of α is equal to 1), endogenous effects spoil the validity of **(H1)**, compromising the stationarity of the system and inducing a divergence of the averages.
- **Endogeneity: $\bar{\Lambda}$.** We will be able to express $\bar{\Lambda}$ as a linear function of $\bar{\mu}$ thanks to the homogeneity hypothesis **(H3)**, encoding the information about $\Psi(0)$ in the ratio $\bar{\Lambda}/\bar{\mu}$. Hence, we will be able to interpret $\bar{\Lambda}/\bar{\mu}$ as a susceptibility of the system with respect to its driving input, while $\bar{\Lambda}/\bar{\mu} - 1$ can be thought of as expressing the ratio between the endogenous intensity against the exogenous one.
- **Relaxation: $\overline{c(\tau)}$ and $\overline{\overline{c(\tau)}}$.** Slow relaxation will be characterized by the behavior at large times of the auto-correlations $\overline{c(\tau)}$ and of the cross-correlations $\overline{\overline{c(\tau)}}$, distinguishing the cases in which these functions decay exponentially from the one in which they develop broad tails. These quantities can be thought as quantifying the response of the system to noise, as it is proportional to the noise covariance matrix Σ .

All these quantities will be analyzed in the $N \rightarrow \infty$ regime of large dimensionality, in which non-trivial dynamic effects are expected to emerge. We shall focus in particular on the issue of **Criticality**, which may induce long-range correlations (divergence of the L^1 norm of the self-correlation function).

Accordingly, the scaling in N that we have introduced for the above observables is such that they all attain a finite limit for $N \rightarrow \infty$. Nevertheless, specifying how such limit is approached requires a prescription encoding the dependence upon N of the interaction kernels $\Phi(\tau)$, of the components μ and, if needed, of the parameter Σ . In the next section we will use a deterministic prescription in order to scale the system with N , and we will explore the large N regime of a regular lattice of finite dimension (i.e., α is invariant under translation and the coefficients μ and Σ are constants). Only in Secs. 4.1 and 4.2 we will study the case in which the coefficients defining the model are free to fluctuate within two given statistical ensembles.

3 The deterministic case

3.1 A translationally-invariant network

In this section we will discuss the case of a translationally-invariant network of interactions for the process $X(t)$, whose components are arranged on the vertices of a regular lattice. This prescription is used in order to explore the effect of the interactions in a completely homogeneous network, disregarding the presence of irregularities in the system. Moreover, this framework allows to discuss the effect of the variation of the connectivity of the system, interpolating from the complete network (corresponding to the infinite dimensional limit, in which the interactions are broadly diluted through the system) to the low-dimensional case (in which strong fluctuations are induced as an effect of the topology), with the main advantage of controlling analytically the behavior of the system throughout the crossover.

In order to define the notion of spatial dimension, we preliminarily need to assume that a notion of geometry emerges naturally in the system, as its structure needs to be invariant under translations along a set of D directions. In particular, we consider the case in which each of the components is located on the vertices of a hypercube of dimension $D > 0$ (see Fig. 1). We will further assume the coordinates to have length L , so that we will be able to label them by using a vector index $\mathbf{i} \in \{0, \dots, L-1\}^D$, and to identify the size of the system N with the volume of the hypercube $N = L^D$.⁶ We will finally assume $\mu_{\mathbf{i}}$ and $\Sigma_{\mathbf{ii}}$ to be constants, i.e.,

$$\mu_{\mathbf{i}} = \bar{\mu} \text{ and } \Sigma_{\mathbf{ii}} = \bar{\Sigma}, \quad \forall \mathbf{i} \in [0, L-1]^D.$$

The periodicity condition will be enforced through the assumption

$$\alpha_{\mathbf{ij}} = \alpha_{\mathbf{i-j}} \quad \forall \mathbf{i}, \mathbf{j} \in [0, L-1]^D.$$

⁶These models can be thought as discretized versions of continuous-space ones such as the one employed in [35] in order to model the occurrence of earthquakes. In our language the continuous-space limit is recovered in the limit $N \rightarrow \infty$. Moreover, the results shown in Sec. 3.4 allow us to explore the regime in which long-range behavior is endogenously induced from the structure of the interactions, rather than enforced by construction in the parametric specification of the interaction model as it is assumed in [35].

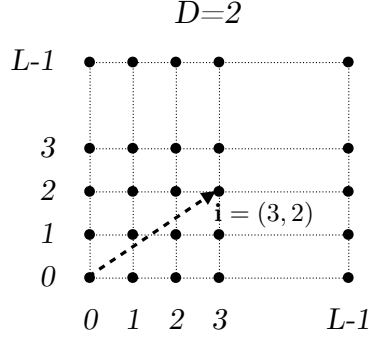


Figure 1: Sketch of a translationally-invariant network for $D = 2$. In this case the components of $X(t)$ are arranged on the vertices of a square lattice of length L .

Let us point out that the diagonalization assumption (16) as well as the Unitarity assumption **(H2)** and the Homogeneity assumption **(H3)** are all satisfied. Indeed, straightforward computations lead to

$$U_{\mathbf{ia}} = \frac{1}{\sqrt{N}} e^{\frac{2\pi i}{L} \mathbf{i} \cdot \mathbf{a}}, \quad (28)$$

$$\lambda_{\mathbf{a}} = \sum_{\mathbf{i}} e^{\frac{2\pi i}{L} \mathbf{i} \cdot \mathbf{a}} \alpha_{\mathbf{i}}, \quad (29)$$

where $\mathbf{a} \in \{0, \dots, L-1\}^D$.

3.2 Stability of translationally invariant systems

The condition of stability of the system (i.e. assumption **(H1)**) is equivalent to the condition

$$||\alpha|| = \max_{\mathbf{a}} |\lambda_{\mathbf{a}}| < 1$$

which we shall assume to be fulfilled both in the VAR and in the Hawkes case. In the latter case, or in the VAR model when all the entries of α are positive, such condition simplifies to $\lambda_0 < 1$. Let us point out that, in this case, the stability of the system is controlled by the parameter $||\alpha|| = \lambda_0$, which increases the degree of self-interaction of the system. In particular, a strongly susceptible behavior is detected in the regime of λ_0 close to 1, where the ratio $\bar{L}/\bar{\mu} - 1$ of endogenously-generated versus exogenously-generated intensity becomes extremely large (see (30)).

3.3 Endogeneity and Relaxation of translationally invariant systems

One can easily prove that

$$\bar{\Lambda} = \frac{\bar{\mu}}{1 - \lambda_0} \quad (30)$$

$$\overline{\hat{c}(\omega)} = \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \sum_{\mathbf{a}} \frac{1}{\|\hat{\phi}(\omega)^{-1} - \lambda_{\mathbf{a}}\|^2} \quad (31)$$

$$\overline{\overline{\hat{c}(\omega)}} = \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \frac{1}{\|\hat{\phi}(\omega)^{-1} - \lambda_0\|^2} . \quad (32)$$

Also notice that the expressions of $\bar{\Lambda}$ and $\overline{\hat{c}(\omega)}$ depend just upon the single eigenvalue λ_0 . This is because the corresponding eigenvalue is $U_{\mathbf{i}0} = N^{-1/2}(1, \dots, 1)$ so that by orthogonality one has $\sum_{\mathbf{i}} U_{\mathbf{i}\mathbf{a}} = \sqrt{N} \delta_{\mathbf{a}0}$. An identical phenomenon will occur in Sec. 4.2, due to the absence of fluctuations in the sum over row of the α matrix.

The autocorrelation function is more conveniently analyzed by rewriting it as

$$\overline{\hat{c}(\omega)} = \bar{\Sigma} \|\hat{\phi}(\omega)\|^{-2} \int_{-\infty}^{+\infty} d\lambda \frac{\rho(\lambda)}{\|\hat{\phi}(\omega)^{-1} - \lambda\|^2} , \quad (33)$$

after introducing the *spectral density*

$$\rho(\lambda) = \frac{1}{N} \sum_{\mathbf{a}} \delta(\lambda - \lambda_{\mathbf{a}}) . \quad (34)$$

The advantage of introducing the spectral density is that $\rho(\lambda)$ is expected to have a well-behaved limit for $L \rightarrow \infty$, which we can exploit in order to find the shape of correlations in the large size limit. In particular, if the maximum of the support of the spectrum is $\lambda_{max} < 1$, the auto-correlations decay exponentially at large times, while for $\lambda_{max} = 1$, the behavior of the spectrum close to $\lambda = 1$ may lead to a non-exponential behavior of the correlation in the $L \rightarrow \infty$ limit. These scenarios are illustrated by mean of some examples in the next Section.

3.4 Criticality for a non-negative α

We first want to investigate the behavior of the model in the vicinity of $\lambda_{max} = 1$ in the case in which $\alpha_{\mathbf{i}}$ is non-negative for all \mathbf{i} . Eq. (30) indicates that, for $\lambda_{max} \rightarrow 1$, the average intensity $\bar{\Lambda}$ diverges, and the ratio of endogenous-to-exogenous events explodes. This type of divergence will be common to all the critical cases which we will be investigating, and can be reabsorbed into a suitable definition of $\bar{\mu}$, similar to what is done in [11]⁷. Physically, this corresponds to the fact that a critical system operates

⁷Unlike in [11], where it is proven that the existence of a well-behaved $\lambda_{max} \rightarrow 1$ limit for the correlation in one dimension requires the kernel $\phi(\tau)$ to be long-ranged, we find that in the multidimensional setting even a short-ranged $\phi(\tau)$ may lead to a well-behaved limit for the correlations. Intuitively, for a fixed value of λ_{max} , a more densely wired system can redistribute potentially dangerous fluctuations among the components of the system, eventually avoiding the divergence of $c_{ii}(\tau)$.

in a regime where small input signals are translated into large outputs.

Interestingly, the average autocorrelation $\overline{c(\tau)}$ can have a finite limit even though the average intensity does not. Calculating it requires calculating the behavior of the spectrum $\lambda_{\mathbf{a}}$ close to the maximum eigenvalue, which is equal to $\|\alpha\| = \lambda_{\mathbf{0}}$. Then we can set $\mathbf{k} = 2\pi\mathbf{a}/L$, and calculate the limit

$$\lambda(\mathbf{k}) = \lim_{L \rightarrow \infty} \lambda_{\frac{\mathbf{k}L}{2\pi}} , \quad (35)$$

which is maximal for $\mathbf{k} = \mathbf{0}$, so that the system is stable for $\lambda_{max} = \lambda(\mathbf{0}) < 1$.

3.4.1 Case of a parabolic spectrum

If the function $\lambda(\mathbf{k})$ is twice differentiable around zero, then its gradient vanishes and its Hessian is negative definite. In this case it is easy to estimate (see App. A) the $L \rightarrow \infty$ limit of the density $\rho(\lambda)$ close to the point $\lambda = \lambda(\mathbf{0})$, which reads

$$\rho(\lambda(\mathbf{0}) - \epsilon) \approx \det(-H[\lambda(\mathbf{0})])^{-1/2} \frac{\epsilon^{D/2-1}}{\Gamma(D/2)} \left(\frac{1}{2\pi} \right)^{D/2} , \quad (36)$$

where $H[\lambda(\mathbf{0})]$ is the Hessian of $\lambda(\mathbf{k})$ calculated in $\mathbf{k} = \mathbf{0}$. Eq. (24) allows us to relate the exponent of ϵ in above expansion to the limiting behavior of the autocorrelations when $\bar{\alpha}$ approaches 1. In particular, the autocorrelations diverge approaching the instability point for $D = 1, 2$, while for $D > 2$ they result

$$\overline{c(\tau)} \sim \tau^{1-D/2} . \quad (37)$$

This implies that

- For $D = 3, 4$ the process can develop long range memory when $\lambda(\mathbf{0})$ tends to the instability point $\lambda(\mathbf{0}) = 1$.
- For $D \rightarrow \infty$ one finds instead that the system loses its power-law behavior, as the tails of the correlations become increasingly steeper.

Notice that the above result does not depend from the specific form of the interaction that we have chosen, but simply emerges from the non-negativity of α_i and the differentiability of $\lambda(\mathbf{k})$ around zero.

An example: next-neighbor interaction. As an example, one can study the behavior of a D -dimensional system with next-neighbor interactions, defined through $\alpha_i = \bar{\alpha}(2D)^{-1} \sum_{d=1}^D (\delta_{\mathbf{i}-\mathbf{e}_d} + \delta_{\mathbf{i}+\mathbf{e}_d})$, with \mathbf{e}_d denoting the d -th element of the canonical basis of \mathbb{R}^D . The spectrum of such a system is given by

$$\lambda(\mathbf{k}) = \frac{\bar{\alpha}}{D} \sum_{d=1}^D \cos(k_d) . \quad (38)$$

For such functional form of $\lambda(\mathbf{k})$ one can explicitly calculate the limiting value of the spectral density $\rho(\lambda)$ for $N \rightarrow \infty$ (App. A), which results

$$\rho(\lambda) = \frac{D}{\bar{\alpha}} \int_{-\infty}^{+\infty} \frac{dz}{2\pi} e^{iz\lambda D/\bar{\alpha}} J_0^D(z) \quad (39)$$

where $J_n(z)$ is the Bessel function of the first kind of order n calculated in z . One can verify by expanding $\rho(\lambda)$ close to the point $\lambda = \bar{\alpha} = 1$ that its limiting behavior is described by Eq. (36). Indeed, the procedure that we have given simply requires the calculation of $\lambda(\mathbf{k})$ around $\mathbf{k} = \mathbf{0}$, and hence can be used in order to tackle a larger class of problems.

The spectrum $\rho(\lambda)$ for a system with next-neighbor interactions is represented in Fig. 2 together with the average over components of the autocorrelation function $\overline{c(\tau)}$. We remark that the above results are valid not only in the regime in which the large size limit $N \rightarrow \infty$ is taken before the $\bar{\alpha} \rightarrow 1$ one, but also when $\bar{\alpha}$ tends to the critical value $\bar{\alpha} = 1$ slower than $\bar{\alpha} = 1 - K/N$ (see App. A). This scaling induces a maximum decay time for correlations, as it bounds the characteristic times of the various decay modes to be smaller than $N/K\beta$. Then, we expect the range of time over which the power-law decay of correlations is observed in a linear system with short range interactions to be at most of the order of the system size N .

3.4.2 Case of a non-parabolic spectrum ($D = 1$)

We want to stress here that a non-trivial behavior may also emerge as a consequence of the non-analyticity of the spectrum around its maximum. In fact, in the preceding section we have been able to argue that if the spectrum of a one-dimensional system $\lambda(\mathbf{k})$ is twice-differentiable, the $L \rightarrow \infty$, $\bar{\alpha} \rightarrow 1$ limit of the autocorrelation function is infinite (where we remind that in order not to obtain trivial results we are taking the $L \rightarrow \infty$ limit before the other ones). Suppose instead that the one-dimensional interaction matrix has tails proportional to $\alpha_i \sim |\mathbf{i}|^{-1-\gamma}$, as for example in the case

$$\alpha_i = \bar{\alpha} \left(\frac{(1 + |\mathbf{i}| \bmod L)^{-\gamma-1} + (1 - (L - |\mathbf{i}|) \bmod L)^{-\gamma-1}}{2\zeta(1 + \gamma)} \right), \quad (40)$$

where $\zeta(1 + \gamma)$ is the Riemann Zeta function calculated in $1 + \gamma$ for $\gamma > 0$. Then the large L limit of the spectrum can be calculated explicitly, and results

$$\lambda(\mathbf{k}) = \bar{\alpha}(e^{ik}\mathcal{P}[1 + \gamma, e^{ik}] + e^{-ik}\mathcal{P}[1 + \gamma, e^{-ik}]), \quad (41)$$

where $\mathcal{P}[1 + \gamma, z]$ is the Polylogarithmic function of order $1 + \gamma$ calculated in z . Its maximum $\lambda(\mathbf{0})$ is equal to $\bar{\alpha}$, but its first derivative at $\mathbf{k} = \mathbf{0}$ is not defined for $0 < \gamma < 1$. The analysis of the behavior of the spectrum in the vicinity of $\mathbf{k} = \mathbf{0}$ in that case reveals that the leading term in the expansion of $\lambda(\mathbf{k})$ around $\mathbf{0}$ is proportional to $||\mathbf{k}||^\gamma$, and thus in the limit $\bar{\alpha} \rightarrow 1$ induces a long-time behavior of the autocorrelations of the type

$$\overline{c(\tau)} \sim \tau^{-(1-\gamma)/\gamma}. \quad (42)$$

Intuitively, by diluting the interactions among a larger number of components, it is possible to tame the strong fluctuations arising in one-dimensional systems, obtaining a finite limit for the correlations. This also implies that long-range correlations are present for $1 > \gamma > 1/2$, when the dilution is less severe.

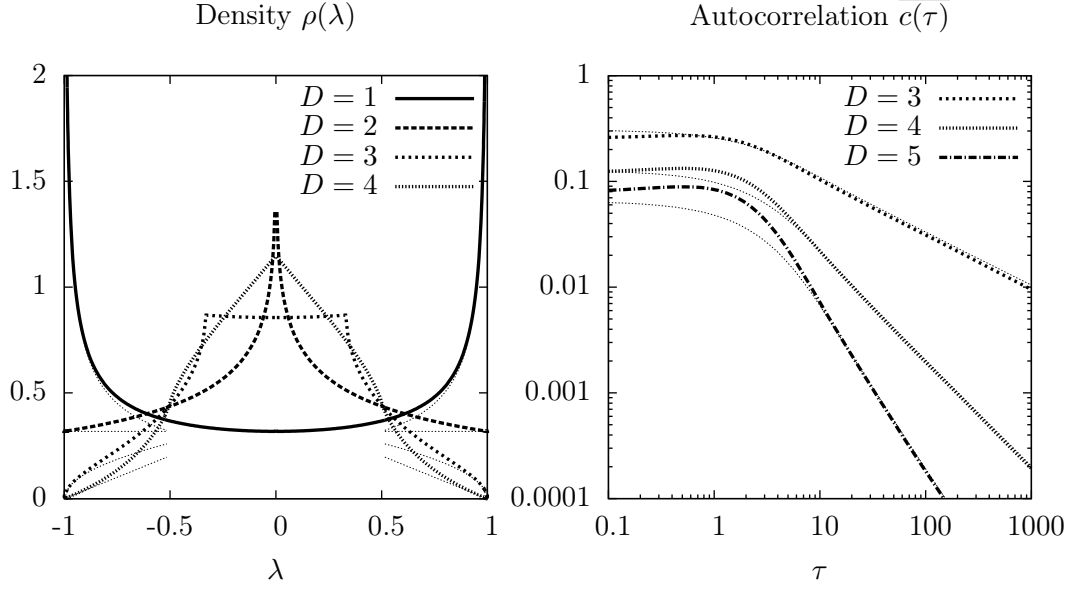


Figure 2: (*Left panel*) Spectrum $\rho(\lambda)$ of a translationally invariant system with next-neighbor interactions for an interaction strength $\bar{\alpha} = 1$ and different values of the dimension D (see Example of Section 3.4.1). While for $D = 1$ the spectrum diverges at the point $\lambda = \bar{\alpha}$, for $D = 2$ it tends to a constant, and for $D \geq 3$ it vanishes. The thick lines show the exact spectrum 39, while the soft dashed lines indicate the asymptotic predictions on the basis of Eq. (36). (*Right panel*) Average over components of the autocorrelation function for the same system at the critical value $\bar{\alpha} = 1$. While at $D \leq 2$ the correlation has no finite limit for $\bar{\alpha} \rightarrow 1$, we have represented its limiting value for $D = 3, 4, 5$. While the heavy curves refer to the predictions of Eq. (39), the soft dashed lines plotted for comparison are the asymptotic results of Eq. (36), predicting a decay of the type $\tau^{1-D/2}$.

4 The stochastic case

While the definition of a periodic system such as the one analyzed above requires the existence of a notion of geometry in the space of the coordinates, we want now to focus our analysis on the disordered case in which no particular geometry is present in the system. This is typically a more realistic scenario for complex systems in which the interactions are not thought to be organized according to a peculiar spatial structure. In order to account for this lack of regularity, we model the parameters of the system as random variables extracted from a given *statistical ensemble*. In this type of framework, characterizing the behavior of a linear model requires understanding how the ensemble fluctuations in the parameters defining the model are inherited by the intensities Λ and the cross-correlations $c(\tau)$.

The first type of ensemble that we are going to consider is a suitable one in order to model disordered realizations of a VAR model, while it cannot be used to model Hawkes processes, as it assigns negative weights to the entries of the interaction matrix α with finite probability. Indeed, a great deal of results can be proved rigorously in this framework. Moreover, many of the following results can be used in order to have an intuitive grasp about more general scenarios, in which an analytical solution is not necessarily available.

4.1 The Gaussian Orthogonal Ensemble

4.1.1 Definition

We first consider a statistical ensemble in which the α matrices are drawn from the Gaussian Orthogonal Ensemble (GOE), in which each α is drawn according to a weight

$$P_N(\alpha) \propto \exp \left(\frac{N}{4\sigma_\alpha^2} \text{tr} \left[\left(\alpha - \mathbb{J} \frac{\bar{\alpha}}{N} \right) \left(\alpha^T - \mathbb{J} \frac{\bar{\alpha}}{N} \right) \right] \right), \quad (43)$$

where \mathbb{J} is the N -dimensional matrix with all components equal to one. This choice implies

$$\mathbb{E}[\alpha_{ij}] = \frac{\bar{\alpha}}{N} \quad (44)$$

$$\mathbb{V}[\alpha_{ij}] = \begin{cases} 2\sigma_\alpha^2/N & \text{if } i = j \\ \sigma_\alpha^2/N & \text{if } i \neq j \end{cases} \quad (45)$$

Where we are using the symbol $\mathbb{E}[\dots]$ in order to denote averages with respect to the measure defined by $P_N(\alpha)$ and $\mathbb{V}[\dots]$ in order to denote the variances with respect to that same measure. The factors of N appearing in the definition of the ensemble have been chosen in order for both the mean and the fluctuations of each of the N quantities $\{\sum_{j=1}^N \alpha_{ij}\}_{i=1}^N$ not to depend explicitly on N . Intuitively, this implies that if $\bar{\alpha}$ and σ_α^2 are finite in the large N limit, then the interaction strength on each of the N components is also finite. Moreover, it is possible to prove (see [36]) that the matrix α admits almost surely the decomposition

$$\alpha = \mathbb{J} \frac{\bar{\alpha}}{N} + O\lambda O^T, \quad (46)$$

with $OO^T = O^TO = \mathbb{I}$, which will be more useful than the decomposition (16) formerly introduced, due to the better symmetry properties of the O matrices. As usual, λ is a diagonal matrix of eigenvalues $\lambda = \{\delta_{ab}\lambda_a\}_{a,b=1}^N$, whose joint probability can be written as

$$p(\lambda_1, \dots, \lambda_N) \propto e^{-\frac{N}{4\sigma_\alpha^2} \sum_a \lambda_a^2} \prod_{a \neq b} |\lambda_a - \lambda_b|, \quad (47)$$

and O is an orthogonal matrix sampled uniformly and independently of λ on the N -dimensional Haar sphere [36].

Once the statistical ensemble for the interaction matrix α is fixed, it is necessary to prescribe a rule for the statistics of the vector quantities μ_i , Λ_i and Σ_{ii} appearing in the model. In particular, we will assume the μ_i to be independent and identically distributed variables, with mean $\mathbb{E}[\mu_i] = \bar{\mu}$ and variance $\mathbb{V}[\mu_i] = \sigma_\mu^2$. The parameters Λ_i will be indirectly fixed by the relation $\Lambda = \hat{\Psi}(0)\mu$. Finally, we additionally need to specify the statistics for the Σ_{ii} . We will take them to be independent and identically distributed with mean $\mathbb{E}[\Sigma_{ii}] = \bar{\Sigma}$ and variance $\mathbb{V}[\Sigma_{ii}] = \sigma_\Sigma^2$ in the VAR case, while in the Hawkes model the values of Σ will be indirectly fixed by the relation $\Sigma_{ii} = \Lambda_i$.

4.1.2 Stability in the GOE ensemble

The statistical ensembles defined above assigns strictly positive probability to α matrices associated with unstable systems (i.e., with finite probability it can be that $\|\alpha\| > 1$), spoiling the assumption **(H1)**. In order to focus our analysis to the stable realizations sampled from those ensembles, it is necessary to restrict the expectation $\mathbb{E}[\dots]$ to the set of matrices for which the largest eigenvalue is smaller than one. In particular, it will be useful to consider the probability measure $\overline{P}_N(\alpha)$, defined as:

$$\overline{P}_N(\alpha) = P_N(\alpha \mid \|\alpha\| < 1) = \frac{P_N(\alpha)}{P_N(\|\alpha\| < 1)} \mathbf{1}_{[0,1)}(\|\alpha\|). \quad (48)$$

We denote averages taken with respect to this measure with $\overline{\mathbb{E}}[\dots]$. As we are interested in studying the stability of these linear models in the regime of large N , we will characterize the large N behavior of $P_N(\|\alpha\| < 1)$, and in verifying whether $P_N(\|\alpha\| < 1)$ tends to a non-vanishing constant $P_\infty(\|\alpha\| < 1)$ in the limit of large N .

Indeed, the problem of stability in the GOE can be solved exhaustively by exploiting the results of Ref. [37, 38, 39] (also see Ref. [40] for a review), where it is shown that the behavior of the largest eigenvalue of the matrix α is dictated by the ratio between $\bar{\alpha}$ and σ_α .

- For $\bar{\alpha}/\sigma_\alpha > 1$, the largest eigenvalue λ_{max} is a Gaussian variable of mean $\bar{\alpha} + \sigma_\alpha^2/\bar{\alpha}$ and variance σ_α^2/N .
- If $\bar{\alpha}/\sigma_\alpha < 1$, the distribution of the rescaled random variable $(\lambda_{max}/\sigma_\alpha - 2)N^{2/3}$ converges to the Tracy-Widom distribution [41, 42].

This implies that, if $\bar{\alpha}/\sigma_\alpha > 1$ and $1 - \bar{\alpha} - \sigma_\alpha^2/\bar{\alpha} \gg N^{-1/2}$, then drawing an unstable sample matrix α becomes a large deviation event, and $P_\infty(\|\alpha\| < 1) = 1$. In the opposite case, if $\bar{\alpha}/\sigma_\alpha < 1$, it is enough that $1 - 2\sigma_\alpha \gg N^{-2/3}$ in order to have $P_\infty(\|\alpha\| < 1) = 1$ (for an exhaustive discussion accounting for this kind of large deviations see Ref. [43, 44]). In the special case in which the maximum eigenvalue

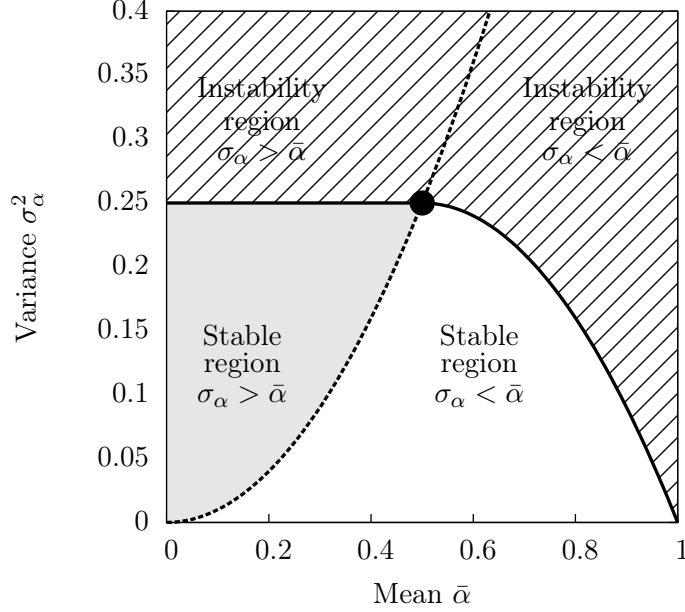


Figure 3: Phase space for the VAR model as a function of the parameters defining the statistical ensemble (in the GOE ensemble) for the matrix α (see Section 4.1). A region of stability (in which $P_{\infty}(\|\alpha\| < 1) = 1$) is separated from an unstable region (where $P_{\infty}(\|\alpha\| < 1) = 0$) by a critical line in which $P_{\infty}(\|\alpha\| < 1)$ is finite. While for $\bar{\alpha} > \sigma_{\alpha}$ the maximum eigenvalue is isolated, along the $\bar{\alpha} < \sigma_{\alpha}$ portion of the critical line (in bold) the maximum eigenvalue corresponds to the edge of the support for the density of eigenvalues.

is close to the instability point ($\lambda_{max} = 1 + O(N^{-1/2})$ for the case of $\bar{\alpha}/\sigma_{\alpha}$ large or $\lambda_{max} = 1 + O(N^{-2/3})$ for $\sigma_{\alpha}/\bar{\alpha}$ small), then $P_N(\|\alpha\| < 1)$ tends to a non-vanishing constant depending upon the precise values of $\bar{\alpha}$ and σ_{α} . This behavior is summarized in Fig. 3, where we show the shape of the stability region in the $(\bar{\alpha}, \sigma_{\alpha}^2)$ plane. Notice that this type of transition is precisely the one discovered in Ref. [1], which has been later related to the onset of a third order phase transition between the *pulled* and *pushed* phase of a Coulomb gas [45].

Additionally, an important qualitative difference emerges in the phases $\bar{\alpha} < \sigma_{\alpha}$ and $\bar{\alpha} > \sigma_{\alpha}$: while in the former case at large N the largest eigenvalue coincides with the support of the spectral density of eigenvalues, in the latter there is a gap among the support of the eigenvalue distribution and the largest eigenvalue (see Sec. 4.1.5 and Fig. 4). We will see that this difference will have a central role in determining the limiting behavior of correlations in the critical case.

4.1.3 Endogeneity and Relaxation in the GOE ensemble

Having established the stability condition for the matrices in the GOE, we enunciate our result for the averages of $\bar{\lambda}$, $\overline{\hat{c}(\omega)}$ and $\overline{\hat{c}(\omega)}$. The systematic procedure used to obtain these results is illustrated in App. B. We get:

$$\mathbb{E}[\bar{\lambda}] = \frac{\bar{\mu}}{\bar{\alpha}} \left(\mathbb{E} \left[\mathcal{F}^{(1)}(\bar{\alpha}, \{(1 - \lambda_a)^{-1}\}) \right] - 1 \right) \quad (49)$$

$$\begin{aligned} \mathbb{E}[\overline{\hat{c}(\omega)}] &= \bar{\alpha}^2 \frac{\bar{\Sigma}}{2N} \|\hat{\phi}(\omega)\|^{-2} \partial_{x^{(3)}}^2 \mathbb{E} \left[\mathcal{F}^{(3)}((\bar{\alpha}, \bar{\alpha}, 0), \{z_a^*(\omega)\}, \{z_a(\omega)\}, \{\|z_a(\omega)\|^2\}) \right] \\ &+ \bar{\alpha} \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \partial_{x^{(2)}} \mathbb{E} \left[\mathcal{F}^{(2)}((\bar{\alpha}, 0), \{z_a^*(\omega)\}, \{z_a^*(\omega)\|z_a(\omega)\|^2\}) \right] \\ &+ \bar{\alpha} \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \partial_{x^{(2)}} \mathbb{E} \left[\mathcal{F}^{(2)}((\bar{\alpha}, 0), \{z_a(\omega)\}, \{z_a(\omega)\|z_a(\omega)\|^2\}) \right] \\ &+ \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \mathbb{E} \left[\sum_a \frac{1}{\|\hat{\phi}^{-1}(\omega) - \lambda_a\|^2} \right] \end{aligned} \quad (50)$$

$$\mathbb{E}[\overline{\hat{c}(\omega)}] = \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \partial_{x^{(3)}} \mathbb{E} \left[\mathcal{F}^{(3)}((\bar{\alpha}, \bar{\alpha}, 0), \{z_a^*(\omega)\}, \{z_a(\omega)\}, \{\|z_a(\omega)\|^2\}) \right], \quad (51)$$

with $z_a(\omega) = (\hat{\phi}^{-1}(\omega) - \lambda_a)^{-1}$. The generating function $\mathcal{F}^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N)$ appearing in above formulae can be defined after considering a p -dimensional vector $\vec{x} = (x^{(1)}, \dots, x^{(p)})$ and a matrix $\{z_a^{(k)}\}_{(a,k) \in (1, \dots, N) \times (1, \dots, p)}$, in whose case it can be written as:

$$\begin{aligned} \mathcal{F}^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) &= \frac{\Gamma(N/2)}{\Gamma(N/2 - p)} \int_0^1 dt_1 \dots dt_p \prod_{k=1}^p t_k^{k-1} (1 - t_p)^{N/2 - p - 1} \\ &\times \prod_a \left(1 - \sum_{k=1}^p h^{(k)}(\vec{t}) x^{(k)} z_a^{(k)} \right)^{-1/2}, \end{aligned} \quad (52)$$

together with a set of auxiliary functions $h^{(k)}(\vec{t})$ given by

$$h^{(1)} = t_1 \dots t_p \quad (53)$$

$$h^{(2)} = (1 - t_1) t_2 \dots t_p \quad (54)$$

$$h^{(3)} = (1 - t_2) t_3 \dots t_p \quad (55)$$

$$\dots \quad (56)$$

$$h^{(p)} = (1 - t_{p-1}) t_p. \quad (57)$$

Eqs. (49), (50) and (51) above allow to express the observables as averages over the eigenvalue distribution of the generating functions $\mathcal{F}^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N)$. In App. B we illustrate how to express a generic observable within the model in terms of these generating functions, and show how to use this formalism in order to derive a systematic expansion in powers of $\bar{\alpha}$ for the momenta of any observable. For example, in the case

of the mean intensities, the first term of such an expansion, corresponding to the case $\bar{\alpha} = 0$, reads:

$$\mathbb{E}[\bar{\Lambda}] = \frac{\bar{\mu}}{N} \mathbb{E} \left[\sum_a \frac{1}{1 - \lambda_a} \right] \quad (58)$$

$$\mathbb{E}[\bar{\hat{c}}(\omega)] = \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \mathbb{E} \left[\sum_a \frac{1}{\|\hat{\phi}^{-1}(\omega) - \lambda_a\|^2} \right] \quad (59)$$

$$\mathbb{E}[\bar{\hat{c}}(\omega)] = \frac{\bar{\Sigma}}{N^2} \|\hat{\phi}(\omega)\|^{-2} \mathbb{E} \left[\sum_a \frac{1}{\|\hat{\phi}^{-1}(\omega) - \lambda_a\|^2} \right]. \quad (60)$$

Even more interestingly, the leading order term of the $1/N$ expansion of $\mathcal{F}^{(p)}$ can be computed analytically (see App. B), making possible the asymptotic estimation of the observables by means of the $\mathcal{F}^{(p)}$ generating functions.

Eqs. (49) and (58) state that, unlike in Eq. (30), all modes of α contribute to the average intensity due to the heterogeneity of the interaction network. As in that case, if any of the eigenvalues exceeds the value $\lambda = 1$, the system loses its stability, and the ratio of endogenous-to-exogenous intensity $\bar{\Lambda}/\bar{\mu} - 1$ explodes. The equations for the correlations for $\bar{\alpha} = 0$ are similar to the ones found in the deterministic case, except for the fact that the average cross-correlation for the non-diagonal terms $\hat{c}_{ij}(\omega)$ with $i \neq j$ is exactly equal to zero.

We remark moreover that by specializing the value of $\mathbb{E}[\hat{c}_{ii}(\omega)]$ to $\omega = 0$, we find an expression proportional to the one of $\mathbb{E}[\Lambda_i^2]$. This implies that the *memory* $\hat{c}_{ii}(0) = \int d\tau c_{ii}(\tau)$ is related to the fluctuations of the mean. In particular, whenever the process develops long-range memory (i.e., $\hat{c}_{ii}(0)$ diverges), then the fluctuations of the mean intensities are also bound to diverge.

Finally, we remind that in the cases in which $P_\infty(\|\alpha\| < 1)$ is equal to one, then $P_\infty(\alpha \mid \|\alpha\| < 1) = P_N(\alpha) \mathbb{1}_{[0,1)}(\|\alpha\|)$, so that it is possible to replace the unconditional measure with the actual one in the large N limit. This is the case in all the stability region $\bar{\alpha} + \sigma_\alpha^2/\bar{\alpha} < 1$, implying that, for example, the $\bar{\alpha} = 0$ value of the average auto-correlation function in the exponential kernel case is given by

$$\mathbb{E}[\bar{c}(\tau)] = \beta \bar{\Sigma} \int_{-2\sigma_\alpha}^{+2\sigma_\alpha} d\lambda \rho_W(\lambda) \frac{\lambda(2-\lambda)}{2(1-\lambda)} e^{-\beta[1-\lambda]|\tau|}, \quad (61)$$

where

$$\rho_W(\lambda) = \mathbb{E} \left[\frac{1}{N} \sum_a \delta(\lambda - \lambda_a) \right] = \frac{1}{2\pi\sigma_\alpha^2} \sqrt{4\sigma_\alpha^2 - \lambda^2} \quad (62)$$

is the well-known Wigner semicircle law with support in $[-2\sigma_\alpha, 2\sigma_\alpha]$ represented in Fig. 4 (see [46]). Fig. 5 shows a set of auto-correlation curves obtained by varying σ_α along the $\bar{\alpha} = 0$ line, comparing the results of Eq. (61) with the results of numerical simulations.

4.1.4 Fluctuations and law of large numbers for $\bar{\Lambda}$

A natural question arising in the study of systems with random interactions is the one related to the self-averagingness of the observables: Given a succession of realizations of the system of increasing N , does a law of large numbers hold for the quantities $\bar{\Lambda}$ and $\overline{c(\tau)}$? Indeed the mean intensity $\bar{\Lambda}$ depends upon the eigenvalues $\{\lambda_a\}_{a=1}^N$, which are strongly correlated variables. As a consequence, one may expect the natural scaling of fluctuations to be altered by the particular statistics of the eigenvalues. The fluctuations of the mean intensity $\bar{\Lambda}$ can be computed analytically by using the results of App. B, and result

$$\begin{aligned}\mathbb{E} \left[\overline{\Lambda \Lambda^T} \right] &= \bar{\mu}^2 (\bar{\alpha} \partial_{x(1)} + 1) \partial_{x(2)} \mathbb{E} \left[\mathcal{F}^{(2)}((\bar{\alpha}, 0), \{(1 - \lambda_a)^{-1}\}, \{(1 - \lambda_a)^{-2}\}) \right] \\ &+ \bar{\alpha}^2 \frac{\sigma_\mu^2}{2N} (\bar{\alpha} \partial_{x(1)} + 1) \partial_{x(2)}^2 \mathbb{E} \left[\mathcal{F}^{(2)}((\bar{\alpha}, 0), \{(1 - \lambda_a)^{-1}\}, \{(1 - \lambda_a)^{-2}\}) \right] \\ &+ \bar{\alpha} \frac{2\sigma_\mu^2}{N} \partial_{x(2)} \mathbb{E} \left[\mathcal{F}^{(2)}((\bar{\alpha}, 0), \{(1 - \lambda_a)^{-1}\}, \{(1 - \lambda_a)^{-3}\}) \right] \\ &+ \frac{\sigma_\mu^2}{N} \mathbb{E} \left[\sum_a \frac{1}{(1 - \lambda_a)^2} \right]\end{aligned}\tag{63}$$

$$\begin{aligned}\mathbb{E} \left[\overline{\Lambda \Lambda^T} \right] &= \frac{\bar{\mu}^2}{\bar{\alpha}^2} (\bar{\alpha} \partial_{x(1)} - 1) \mathbb{E} \left[\mathcal{F}^{(1)}(\bar{\alpha}, \{(1 - \lambda_a)^{-1}\}) - 1 - \frac{\bar{\alpha}}{N} \sum_a \frac{1}{(1 - \lambda_a)} \right] \\ &+ \frac{\sigma_\mu^2}{N} (\bar{\alpha} \partial_{x(1)} + 1) \partial_{x(2)} \mathbb{E} \left[\mathcal{F}^{(2)}((\bar{\alpha}, 0), \{(1 - \lambda_a)^{-1}\}, \{(1 - \lambda_a)^{-2}\}) \right],\end{aligned}\tag{64}$$

While Eq. (63) is useful to estimate the fluctuations of the individual Λ_i , Eq. (64) can be used in order to estimate the variance of $\bar{\Lambda} = N^{-1} \sum_i \Lambda_i$.

Considering the case $\bar{\alpha} = 0, \sigma_\alpha < 1/2$. In particular, we are able to estimate analytically the latter quantity along the $\bar{\alpha} = 0, \sigma_\alpha < 1/2$ line, as in such region the mean intensity reduces to a linear statistics of the eigenvalues. In that case, for N large enough the variance of $\bar{\Lambda}$ tends to

$$\begin{aligned}\overline{\text{Var}}[\bar{\Lambda}] &\rightarrow \frac{\sigma_\mu^2}{N} \int_{-2\sigma_\alpha}^{2\sigma_\alpha} d\lambda \frac{\rho_W(\lambda)}{(1 - \lambda)^2} \\ &+ \frac{\bar{\mu}^2}{N^2} \mathbb{P} \left\{ \int_{-2\sigma_\alpha}^{2\sigma_\alpha} d\lambda d\lambda' \frac{\rho_W(\lambda, \lambda')}{(1 - \lambda)(1 - \lambda')} \right\},\end{aligned}\tag{65}$$

where \mathbb{P} denotes the Cauchy principal value, and

$$\rho_W(\lambda, \lambda') = -\frac{1}{\pi^2} \left(\frac{1}{[(2\sigma_\alpha + \lambda)(2\sigma_\alpha - \lambda)]^{1/2}} \right)\tag{66}$$

$$\times \frac{\partial^2}{\partial \lambda \partial \lambda'} \left([(2\sigma_\alpha + \lambda')(2\sigma_\alpha - \lambda')]^{1/2} \log |\lambda - \lambda'| \right)\tag{67}$$

is an universal *spectral correlation*, which encodes the strongly interacting nature of the eigenvalue distribution [47, 48]. As both integrals appearing in Eq.(65) converge, the contribution of the σ_μ and $\bar{\mu}$ terms scale respectively as N^{-1} and N^{-2} .

Let us point out that we find by numerically evaluating Eq. (64) that this scaling extends to all the values of $(\bar{\alpha}, \sigma_\alpha)$ in the interior of the stability region. The scaling of fluctuations in the critical regime is indeed non-trivial: taking for example the case $\bar{\alpha} = 0$, one can check that both integrals appearing in Eq. (65) are formally divergent, indicating the emergence of a different scaling with N of the fluctuations. Finding the exact scaling at transition requires indeed a more sophisticated analysis, as one needs to take into account that $P_N(\alpha) \neq \overline{P_N}(\alpha)$ in order to compute the fluctuations of the largest eigenvalues dominating the divergence.

4.1.5 Criticality

Similar to what has been discussed above for a translationally invariant system, even in this random framework we can imagine the system to be able to develop slow correlation if the spectrum touches the instability point $\lambda = 1$. The situation is indeed more delicate in this case, as the maximum value may not coincide with the edge of the support of $\rho(\lambda)$ as it was in the former case. In particular when $\bar{\alpha} > \sigma_\alpha$ and $\alpha + \sigma_\alpha^2/\alpha = 1$ the large time behavior is dictated by an isolated exponential mode whose associated decay speed becomes extremely small. This scenario is similar to the one considered in Ref. [11]. The phase in which $\bar{\alpha} < \sigma_\alpha$ and σ_α is close to the instability point $\sigma_\alpha = 1/2$ leads instead to a richer dynamical behavior, and its phenomenology is, to the best of our knowledge yet unexplored. Fig. 4 summarizes this description by representing the spectrum in both of these cases.

Considering the case $\bar{\alpha} = 0, \sigma_\alpha = 1/2$. In order to be definite, let us consider the case $\bar{\alpha} = 0$ and $\sigma_\alpha = 1/2$. In that case and in the large N regime the realizations are stable with finite probability $P_\infty(|\alpha| < 1) = \mathcal{P}_1(0) \approx 0.8319$, where $\mathcal{P}_\beta(z)$ is the cumulative of the Tracy Widom distribution of index β calculated in z (see Refs. [43, 44]). The long time regime of the correlations is then dictated by the behavior of the density of eigenvalue close to 1. In particular, if one considers

$$\bar{\rho}_W(\lambda) = \overline{\mathbb{E}} \left[\frac{1}{N} \sum_a \delta_{\lambda - \lambda_a} \right], \quad (68)$$

the behavior of $\overline{\mathbb{E}} \left[\overline{c(\tau)} \right]$ depends upon the shape of $\bar{\rho}_W(\lambda)$ close to 1. Notice that, as $P_\infty(|\alpha| < 1) < 1$, one could expect that $\bar{\rho}_W(\lambda) \neq \rho_W(\lambda)$. However, at leading order in N , the two densities converge to the same one, and it is possible to use Eq. (61) even in the critical regime [43, 44]. Intuitively, this indicates that the number of eigenvalues exceeding the instability point $\lambda = 1$ is of order smaller than $O(N)$, implying that they give a negligible contribution to $\bar{\rho}_W(\lambda)$ in the large N limit. In the case of the exponential kernel, we find by plugging $\bar{\rho}_W(\lambda)$ into Eq. (24) that for N large enough

$$\overline{\mathbb{E}} \left[\overline{c(\tau)} \right] = \frac{\bar{\Sigma}\beta}{\pi} \int_{-1}^1 d\lambda \lambda(2-\lambda) \sqrt{\frac{1+\lambda}{1-\lambda}} e^{-\beta|\tau|(1-\lambda)} \sim \tau^{-1/2}. \quad (69)$$

Studying the tail behavior of the average autocorrelation is indeed delicate: one can rigorously find a scaling $c_{ii}(\tau) \sim \tau^{-1/2}$ just in the setting in which the limit of large N

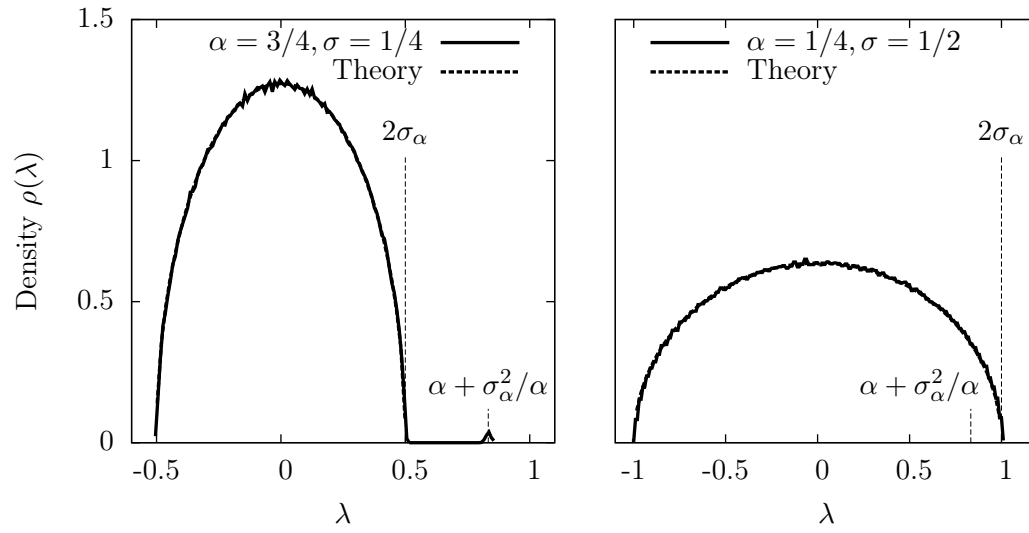


Figure 4: Spectrum $\rho(\lambda)$ of the Gaussian Orthogonal Ensemble in the cases $\bar{\alpha} > \sigma_\alpha$ (left panel) and $\bar{\alpha} < \sigma_\alpha$ (right panel) (See Section 4.1). We have averaged over 100 realizations of a system of size $N = 1000$ in order to obtain the curves depicted in the figure, which at this scale are almost exactly superposed to the Wigner semicircle.

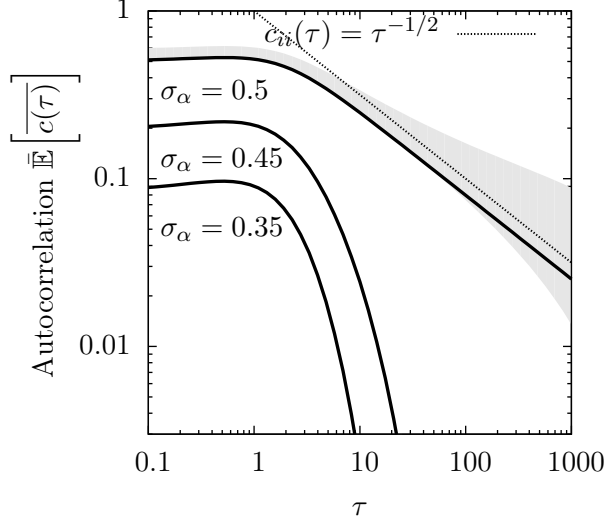


Figure 5: Autocorrelation function of the GOE ensemble for variable σ_α in the case $\bar{\alpha} = 0$ (See Section 4.1.3). The solid lines represent the theoretical predictions, while the shaded regions are two-sigmas error bars accounting for the results of the numerical estimations (300 realizations of systems of size $N = 1000$). Notice that the theoretical predictions for $\sigma_\alpha < 1/2$ are indistinguishable from the numerical results, while for $\sigma_\alpha = 1/2$ strong fluctuations arise in the system as a consequence of the critical regime that we are considering.

is taken before the one of large τ . In practice the finite N corrections establish a upper cutoff to the correlations, which decay exponentially beyond a time τ_{cut} whose location depends upon the corrections to the limiting law of $\bar{\rho}_W(\lambda)$. Moreover, fluctuations are large at any τ : while for $\sigma_\alpha < 1/2$ the variance of $\sum_i c_{ii}(\tau)$ has a finite limit for $N \rightarrow \infty$, in the case $\sigma_\alpha = 1/2$ such limit becomes infinite. The behavior of the autocorrelations in the critical case $\bar{\alpha} = 0, \sigma_\alpha = 1/2$ is also represented in Fig. 5, where we have compared the theoretical predictions with the results of numerical simulations. The divergence of $\hat{c}_{ii}(0) = \int d\tau c_{ii}(\tau)$ in this regime signals that long-range memory is induced in the system as an effect of the structure of the interaction matrix α . Notice that the behavior of $\hat{c}_{ii}(\omega)$ in $\omega = 0$ and the one of the fluctuations of the mean intensities Λ_i are related. In particular, if $\hat{c}_{ii}(0)$ diverges, Eq. (63) implies that the mean intensity Λ_i has infinite variance.

We remark that, as in the translationally-invariant case discussed above, the slow relaxation of the correlations is independent on the specific shape of the kernel $\phi(\tau)$, as long as the integral $\int_0^\infty d\tau \tau \phi(\tau)$ is finite.

4.2 Random graph ensembles

4.2.1 Definition

In this section we consider ensembles in which the interaction matrix α is generated by randomly assigning positive weights to the edges of a graph whose vertices are the N coordinates of the system. As in this setting the matrix α is by construction non-negative, we are free to use the following ensembles in order to build random realizations of a multivariate Hawkes proces.

There is indeed an important difference to underline among VAR processes and multivariate Hawkes models. In fact, while the Σ_{ii} appearing in a VAR model are independent of the Λ_i variables, in the Hawkes model their values need to coincide. This implies in particular that the expression of the correlations in the two cases can in general be different. Summarizing, while at large times a deterministic Hawkes process is always identical to an appropriately tuned continuous-time VAR, in the random case they can be different, as the noise in the Hawkes case is endogeneously generated.

As the ensemble for non-negative realizations of α , we take the one of the *c-regular random graphs*, in which each α is associated with an adjacency matrix \mathcal{G} uniformly chosen in the space of random undirected graphs of N vertices, each of which is connected exactly to c vertices. We then define α as $\alpha = (\bar{\alpha}/c)\mathcal{G}$. According to these conventions, the first two momenta of each entry result

$$\mathbb{E}[\alpha_{ij}] = \frac{\bar{\alpha}}{N} \quad (70)$$

$$\mathbb{V}[\alpha_{ij}] = \frac{\bar{\alpha}^2}{cN} \left(1 - \frac{c}{N}\right). \quad (71)$$

We will further suppose c to be an N -independent constant. In fact, if c grows with N , the bulk of eigenvalues of α shrinks with N , leading in the large N limit to an asymptotic spectrum consisting of $N - 1$ degenerate eigenvalues equal to zero and to a unique eigenvalue equal to $\bar{\alpha}$ [49, 50]. This type of behavior has been proved for several other ensembles of non-negative α matrices (e.g, [51, 52, 53]), and can be informally summarized by the statement that *ensembles of non-negative α matrices are expected to have a non-degenerate behavior in the large N limit just in the sparse case*. By *sparse*, we mean that given a random row $1 \leq i \leq N$, a number of entries independent of N should account for a finite fraction of the sum $\sum_j \alpha_{ij}$ with large probability.

4.2.2 Stability in the regular random graph ensemble

The issue of stability in this ensemble is trivial, as for $\bar{\alpha} < 1$ the system is always stable (the maximum eigenvalue does not fluctuate because the degree of each node of the graph is fixed), and the measure associated with the averages $\mathbb{E}[\dots]$ coincides with the unrestricted one $\mathbb{E}[\dots]$. Indeed, the maximum eigenvalue of the matrix α is associated with the eigenvector $U_{ia} = N^{-1/2}(1, \dots, 1)$. This scenario is analogous to the one analyzed in Sec. 3, in which the absence of fluctuations for the maximum eigenvalue led the Perron-Frobenius eigenvalue to be associated with the same eigenvector.

4.2.3 Endogeneity and Relaxation in the regular random graph ensemble

The expressions for the averages of the various observables in the random regular graph are similar to the ones calculated in Sec. (3) due to the absence of fluctuations in the largest eigenvalue of α . By using the decomposition (16) we can in fact show that

$$\mathbb{E}[\bar{\Lambda}] = \frac{\bar{\mu}}{1 - \bar{\alpha}} \quad (72)$$

$$\mathbb{E}[\overline{\hat{c}(\omega)}] = \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \mathbb{E} \left[\sum_a \frac{1}{\|\hat{\phi}(\omega)^{-1} - \lambda^a\|^2} \right] \quad (73)$$

$$\mathbb{E}[\overline{\hat{c}(\omega)}] = \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \frac{1}{\|\hat{\phi}(\omega)^{-1} - \bar{\alpha}\|^2} . \quad (74)$$

These expressions are valid for both the VAR and the Hawkes model, once one identifies Σ_{ii} with Λ_i . This is because even though the correlations should be calculated differently in each of the two cases, the absence of fluctuations for the maximum eigenvalue induces the same expression of the correlations.⁸ In the $N \rightarrow \infty$ limit, the form of the spectrum $\rho(\lambda)$ for this ensemble has been calculated in Refs. [49, 50]. It results:

$$\rho_c(\lambda) = \mathbb{E} \left[\frac{1}{N} \sum_a \delta_{\lambda - \lambda_a} \right] = \frac{1}{2\pi\bar{\alpha}^2} \frac{\sqrt{4\bar{\alpha}^2(c-1) - c^2\lambda^2}}{\bar{\alpha}^2 - \lambda^2} , \quad (75)$$

with support $\lambda \in [-2\bar{\alpha}\sqrt{c-1}/c, 2\bar{\alpha}\sqrt{c-1}/c]$. This explicit form of the spectral density can be used in order to evaluate numerically the diagonal terms $\bar{\Lambda}\bar{\Lambda}^T$ and $\overline{\hat{c}(\omega)}$.

4.2.4 Criticality

The behavior of the system in the vicinity of the point $\bar{\alpha} = 1$ can be analyzed easily due to the explicit expressions provided for the shape of the bulk (Eq. (75)) and the value of the maximum eigenvalue, which is always equal to $\bar{\alpha}$. As in the previous cases, we will refer explicitly to the case in which the kernel $\phi(\tau)$ is an exponential function ($\phi(\tau) = \beta e^{-\beta\tau} \mathbf{1}_{\mathbb{R}^+}(\tau)$), reminding that this choice doesn't induce any loss of generality at large times as long as the kernel is assumed to be short-ranged.

For $c > 2$ the edge of the bulk never touches the point $\bar{\alpha}$, indicating that the leading term of the large-time expansion of the auto-correlations is always proportional to $e^{-\beta(1-\bar{\alpha})|\tau|}$. Then the $\bar{\alpha} \rightarrow 1$ limit has the same type of behavior of the $\bar{\alpha} > \sigma_\alpha$ phase of the GOE ensemble already discussed. For $c = 2$, the edge of the spectrum touches the point $\lambda = \bar{\alpha}$, as the density takes the form

$$\rho_{c=2}(\lambda) = \frac{1}{\pi\bar{\alpha}^2} (\bar{\alpha}^2 - \lambda^2)^{-1/2} . \quad (76)$$

⁸Including degree fluctuations in the graph associated to α would lead to a different expression of the correlations. In particular, the value of auto- and cross-correlations in the Hawkes case would be sensitive to the *third* power of $\hat{\Psi}(\omega)$, as opposed to the *second* power of $\hat{\Psi}(\omega)$ which appears in the expressions which we have reported.

By inserting the above density into Eq. (73) one finds the type of divergence already encountered in the case of a one-dimensional system: in order to have a finite limit for $\bar{\alpha} = 1$, the leading behavior close to that point should be of the type $(1 - \lambda)^\gamma$ with $\gamma > 0$. This indicates that for $c = 2$ in the limit $\bar{\alpha} \rightarrow 1$, the correlations are bound to become infinite.

It would nevertheless be interesting to find an ensemble of non-negative matrices in which (i) the fluctuations of the maximum eigenvalue tend to zero in the large N limit, (ii) the maximum eigenvalue touches the bulk of the spectrum and (iii) the spectral density vanishes at the edge of the spectrum. This would allow the model to describe the onset of non-trivial correlations at the critical point $\bar{\alpha} = 1$ even for Hawkes processes. A natural candidate to this purpose would be the Erdős-Rényi ensemble of α matrices (informally, it can be thought of as an analytical continuation of the regular random graph to non-integer, fluctuating c). Moreover, the spectral density of the matrices sampled in that ensemble has been shown to be vanishing close to the edge of the bulk as in the Wigner case [54, 55, 56, 57, 58, 59], allowing in principle non-exponential correlations to arise in the critical case. The problem of the Erdős-Rényi ensemble lies indeed in points (i) and (ii), as the degrees of the nodes of the matrix α can have potentially large fluctuations, leading to a slow divergence of the maximum eigenvalue in the large N limit [60], spoiling the stability of the ensemble. Is it possible to generalize the Erdős-Rényi ensemble by reducing the amount of degree fluctuations, so to control the maximal eigenvalue of α ? We leave this issue as an interesting open question.

5 Conclusions

In this work we have investigated the first and second order properties of two linear models (VAR and Hawkes processes) in the regime of high dimensionality, providing the values of the observables characterizing their behavior under different setups. We have addressed in particular the issues of stability, endogeneity, and slow relaxation for both deterministic and stochastic realizations of these models, showing that under the assumptions of factorizability **(H2)** and homogeneity **(H3)**, all the information about these systems can be related to the spectrum $\{\lambda_a\}_{a=1}^N$ of the matrix encoding its interaction network. We have shown that, as opposed to the univariate setup considered in [11], collective effects can trigger slow relaxation of the correlation functions even in systems with short-memory interactions provided that the edge of such spectrum *smoothly* touches the instability point $\lambda = 1$. More generally, we find that in systems characterized by a strong degree of endogeneity, the dynamic properties are also expected to slow down, akin to what is observed for glassy systems [61], for which the proximity to a static transition point is signaled by a slow relaxation of the correlations. We additionally relate the memory of the system with the self-averaging behavior of the intensity Λ : in the stochastic case long-range correlations are measured just when the variance of the observed signal diverges on a realization-per-realization basis. On a more general footing, we find that random realizations of a large linear systems are always self-averaging unless the system is exactly poised at the critical point.

Our study also illustrates the tight relation among the VAR and the Hawkes pro-

cesses, whose first- and second-order properties are shown to coincide. A Hawkes process is in fact reminiscent of a VAR model in which the noise term is endogenously generated by the system. Such a fundamental difference also explain why random realizations of a Hawkes process differ from the ones sampled from a VAR, as in the former case the endogenous nature of the noise may enhance the level of fluctuations of the system, while in the latter the noise can be taken as completely exogenous. Another fundamental difference among the two models is the excitatory nature of the interactions for a Hawkes system as opposed to a VAR model, in which the sign of interactions is arbitrary: Our study supports the hypothesis that a given degree of frustration (i.e., negative interactions) is required in order to observe slow relaxations in short-memory linear systems.

Finally, from the perspective of empirical calibration, our study shows that the parametric fit of a system characterized by power law relaxations with a short range kernel naturally leads to a critical point, the only one which can accommodate such a slow behavior of correlations in data.

Acknowledgements

We warmly thank F. Lillo for long and fruitful discussions all along the preparation of this manuscript. We also thank J.-P. Bouchaud, J. Bun, S. Majumdar, P. Vivo for useful discussions. This research benefited from the support of the “Chair Markets in Transition”, under the aegis of “Louis Bachelier Finance and Sustainable Growth” laboratory, a joint initiative of École Polytechnique, Université d’Évry Val d’Essonne and Fédération Bancaire Française.

A Spectral properties of a finite-dimensional model

In Sec. 3 we have analyzed the behavior of a linear model whose components lie on the vertices of a hypercube of dimension D . In this appendix we show how to calculate the $N \rightarrow \infty$ limit of the spectral density

$$\rho(\lambda) = \frac{1}{N} \sum_{\mathbf{a}} \delta(\lambda - \lambda_{\mathbf{a}}) , \quad (77)$$

for a model with a next-neighbor interactions matrix $\alpha_{\mathbf{i}-\mathbf{j}}$. We will further match our result with the general prediction valid in the vicinity of $\lambda = \lambda_{\mathbf{0}}$ for any type of component-wise positive semi-definite $\alpha_{\mathbf{i}-\mathbf{j}}$.

A.1 Spectral density of a next-neighbor model

The next-neighbor interaction kernel is defined by

$$\alpha_{\mathbf{i}} = \frac{\bar{\alpha}}{2D} \sum_{d=1}^D (\delta_{\mathbf{i}-\mathbf{e}_d} + \delta_{\mathbf{i}+\mathbf{e}_d}) \quad (78)$$

with \mathbf{e}_d denoting the d -th element of the canonical basis of \mathbb{R}^D . By using Eq. (28), it is straightforward to prove that

$$\lambda(\mathbf{k}) = \frac{\bar{\alpha}}{D} \sum_{d=1}^D \cos(k_d) \quad (79)$$

with $\mathbf{k} = 2\pi\mathbf{a}/L$. The limit $N \rightarrow \infty$ can be calculated by writing the number of eigenvalues with density smaller than λ , which results

$$\begin{aligned} \mathcal{N}_{<}(\lambda) &= \frac{1}{N} \sum_{\mathbf{a} \in \{0, \dots, L-1\}^D} \mathbb{1}_{\lambda_{\mathbf{a}} \in (-\infty, \lambda)} \\ &\xrightarrow{N \rightarrow \infty} \frac{1}{(2\pi)^D} \int_{[-\pi, \pi]^D} d\mathbf{k} \mathbb{1}_{\lambda(\mathbf{k}) \in (-\infty, \lambda)} . \end{aligned} \quad (80)$$

By differentiation and successive exponentiation of a delta function, one finds

$$\rho(\lambda) = \frac{d\mathcal{N}_{<}(\lambda)}{d\lambda} = \frac{1}{(2\pi)^D} \int_{-\infty}^{+\infty} \frac{dz}{2\pi} \int_{[-\pi, \pi]^D} d\mathbf{k} e^{iz(\lambda - \frac{\bar{\alpha}}{D} \sum_{d=1}^D \cos(k_d))} \quad (81)$$

$$= \int_{-\infty}^{+\infty} \frac{dz}{2\pi} e^{iz\lambda} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} dk e^{-iz \frac{\bar{\alpha}}{D} \cos(k)} \right)^D \quad (82)$$

$$= \int_{-\infty}^{+\infty} \frac{dz}{2\pi} e^{iz\lambda} J_0^D(|z|\bar{\alpha}/D) , \quad (83)$$

which, after a change of variables, is the result displayed in Eq. (39).

A.2 Edge density for a positive definite interaction matrix

We now want to compute the spectral density for a generic interaction matrix of the form α_{i-j} which is positive semi-definite and whose limiting spectrum is twice-differentiable in the limit $N \rightarrow \infty$. In this case, the maximum of the spectrum is achieved for $\mathbf{k} = \mathbf{0}$, so that one can expand it as

$$\lambda(\mathbf{k}) = \lambda(\mathbf{0}) + \frac{1}{2} (\mathbf{k}^T H[\lambda(\mathbf{0})] \mathbf{k}) + o(\mathbf{k}^2) \quad (84)$$

Under this approximation, one can write that the number of eigenvalues larger than λ is given by

$$\mathcal{N}_{>}(\lambda) \xrightarrow{N \rightarrow \infty} \frac{1}{(2\pi)^D} \int_{\mathbf{k} \in [-\pi, \pi]^D} d\mathbf{k} \mathbb{1}_{\lambda_{\mathbf{a}} \in (\lambda, \lambda_0]} \quad (85)$$

$$\approx \frac{1}{(2\pi)^D} \int_{\mathbf{k} \in \mathbb{R}^D} d\mathbf{k} \mathbb{1}_{-1/2(\mathbf{k}^T H \mathbf{k}) \in [0, \lambda_0 - \lambda)} . \quad (86)$$

Above expression is the volume of a D -dimensional ellipsis, and it results

$$\mathcal{N}_{>}(\lambda(\mathbf{0}) - \epsilon) \approx \frac{\det^{-1/2}(-H[\lambda(\mathbf{0})])}{(2\pi)^D} \frac{(2\pi\epsilon)^{D/2}}{\Gamma(D/2 + 1)} . \quad (87)$$

By differentiation with respect to ϵ , one recovers Eq. (36).

A.3 Scaling of $\bar{\alpha}$ and N close to the critical point

During our discussion, we have always supposed the limit $N \rightarrow \infty$ to be attained before any other one, notably the one $\bar{\alpha} \rightarrow 1$ corresponding to the critical regime of a linear system. In this last part of the appendix, we want to show that it is possible to relax this assumption, and provided that $\bar{\alpha}$ tends to one slower than

$$\bar{\alpha} = 1 - \frac{K}{N}, \quad (88)$$

we obtain a finite limit for the auto-correlations $\overline{c(\tau)}$. In fact, for $\lambda_{\mathbf{a}}$ real, we can always write Eq. (21)

$$\overline{c(\tau)} \leq \overline{c(0)} \leq \frac{\kappa}{L^D} \sum_{\mathbf{a} \in \{0, \dots, L-1\}^D} \frac{1}{(1 - \lambda_{\mathbf{a}})} \quad (89)$$

$$= \frac{\kappa}{L^D} \sum_{\mathbf{a} \in \{0, \dots, L-1\}^D} \frac{1}{1 - \bar{\alpha} D^{-1} \sum_{d=1}^D \cos(2\pi a_d / L)} \quad (90)$$

where κ is a constant equal to $\beta \bar{\Sigma}$. The terms in which the cosine is negative can be bound by 1, and thus give a finite contribution. By symmetry of the cosine function, one can additionally restrict the remaining terms to the subset $\mathbf{a} \in \{0, \dots, L/4\}^D$, where $\cos(x) \leq 1 - Cx^2$, and $C = 1/3$. It results that

$$\overline{c(\tau)} \leq \frac{\kappa'}{L^D} \sum_{\mathbf{a} \in \{0, \dots, L/4\}^D} \frac{1}{1 - \bar{\alpha} D^{-1} \sum_{d=1}^D (1 - 4\pi^2 a_d^2 / L^2)}. \quad (91)$$

The potential divergence at small \mathbf{a} comes only from the $\mathbf{a} = \mathbf{0}$ term, and can be reabsorbed by a scaling of $\bar{\alpha}$ slower than $\bar{\alpha} = 1 - K/L^D$, in whose case one can safely approximate the above sum by an integral. Such integral can be bound by

$$\overline{c(\tau)} \leq \frac{\kappa''}{L^D} \int_1^{L/4} dr \frac{r^{D-1}}{(1 - \bar{\alpha}) + 4\pi^2 \bar{\alpha} \frac{r^2}{L^2}}, \quad (92)$$

where κ'' is a constant taking into account an integration over the angular components arising from a transformation in polar coordinates and the small r contribution to the sum. This integral converges for $D > 2$ even for $\bar{\alpha} = 1$. Summarizing, provided that the divergence at small distances is tamed by an appropriate scaling of $\bar{\alpha}$, the large length behavior is dictated only by dimensionality, regardless of the scaling of $\bar{\alpha}$.

B Observables of a random linear system

In Sec. 4.1 we have been interested in calculating the momenta of several observables (the correlations $\hat{c}_{ij}(\omega)$, the mean intensities Λ_i or the exogenous intensities μ_i) according to the expectation $\mathbb{E}[\dots]$ associated with the GOE statistical ensemble. In this appendix, we will describe the systematic method which has been developed in order to calculate all these quantities. As a preliminary note, we remark that

in all such ensembles, the momenta of Λ_i and $c_{ij}(\tau)$, are by construction invariant under permutation of the indices. That is why, without loss of generality, complete information about the system can be encoded in the scalar expressions appearing in Sec. 4.1.3. For brevity, we list in this appendix just the results that we get for the intensities $\overline{\mathbb{E}}[\Lambda_i]$ and the average correlations $\overline{\mathbb{E}}[\hat{c}_{ij}(\omega)]$ in the VAR case with random drive. In fact, the procedure that we illustrate can be generalized straightforwardly to the other scenarios, leading to the expressions reported in Sec. 4.1.3.

First, in the VAR case we remark that the decomposition (46) leads to an the expression for $\hat{\Psi}(\omega) = (\mathbb{I} - \hat{\Phi}(\omega))^{-1}$ of the type

$$\hat{\Psi}(\omega) = \hat{\phi}(\omega)^{-1} O \left[(\mathbb{I} \hat{\phi}(\omega)^{-1} - \lambda)^{-1} \sum_{n=0}^{\infty} \left(\frac{\bar{\alpha}}{N} \right)^n \left((O^T \mathbb{J} O) (\mathbb{I} \hat{\phi}(\omega)^{-1} - \lambda)^{-1} \right)^n \right] O^T. \quad (93)$$

This allows to write explicitly the expression for the vector Λ and the correlation matrix $\hat{c}(\omega)$, which become

$$\Lambda = \hat{\Psi}(0) \mu = \sum_n \left(\frac{\bar{\alpha}}{N} \right)^n O \hat{K}(0) [M \hat{K}(0)]^n O^T \quad (94)$$

$$\hat{c}(\omega) = \|\hat{\phi}(\omega)^{-1}\|^2 \sum_{n,m} \left(\frac{\bar{\alpha}}{N} \right)^{n+m} O \hat{K}^*(\omega) [M \hat{K}^*(\omega)]^n O^T \bar{\Sigma} O [\hat{K}(\omega) M]^m \hat{K}(\omega) O^T, \quad (95)$$

where M is the symmetric matrix $M = O^T \mathbb{J} O$ and where $\hat{K}(\omega) = (\mathbb{I} \hat{\phi}(\omega)^{-1} - \lambda)^{-1}$. The relevant scalars for the averages of these quantities become

$$\begin{aligned} \overline{\mathbb{E}}[\bar{\Lambda}] &= \frac{\bar{\mu}}{N} \sum_n \left(\frac{\bar{\alpha}}{N} \right)^n \overline{\mathbb{E}} \left[\text{tr} \left\{ (M \hat{K}_0)^{n+1} \right\} \right] \\ \overline{\mathbb{E}}[\hat{c}(\omega)] &= \frac{\bar{\Sigma}}{N} \|\hat{\phi}(\omega)\|^{-2} \sum_{n,m} \left(\frac{\bar{\alpha}}{N} \right)^{n+m} \overline{\mathbb{E}} \left[\text{tr} \left\{ \hat{K}^*(\omega) (M \hat{K}^*(\omega))^n (\hat{K}(\omega) M)^m \hat{K}(\omega) \right\} \right] \\ \overline{\mathbb{E}}[\overline{\hat{c}(\omega)}] &= \frac{\bar{\Sigma}}{N^2} \|\hat{\phi}(\omega)\|^{-2} \sum_{n,m} \left(\frac{\bar{\alpha}}{N} \right)^{n+m} \overline{\mathbb{E}} \left[\text{tr} \left\{ \hat{K}^*(\omega) (M \hat{K}^*(\omega))^n (\hat{K}(\omega) M)^m \hat{K}(\omega) M \right\} \right]. \end{aligned}$$

These expressions are more conveniently computed after introducing the generating function

$$F^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) = \sum_{n^{(1)}} \cdots \sum_{n^{(p)}} \prod_{k=1}^p \left\{ \frac{x^{(k)}}{N} \sum_a z_a^{(k)} \left(\sum_i O_{ia} \right)^2 \right\}^{n^{(k)}}, \quad (96)$$

which satisfies the identities:

$$F^{(p)}(0, \{z_a^{(k)}\}_{a=1}^N) = 1 \quad (97)$$

$$\begin{aligned} \frac{1}{n!} \frac{\partial^n}{\partial (x^{(k)})^n} F^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) \Big|_{x^{(k)}=0} &= \left\{ \frac{1}{N} \sum_a z_a^{(k)} \left(\sum_i O_{ia} \right)^2 \right\}^n \\ &\times F^{(p-1)}(x^{(1)}, \dots, x^{(k-1)}, x^{(k+1)}, \dots, x^{(p)}, \\ &\quad \{z_a^{(1)}\}, \dots, \{z_a^{(k-1)}\}, \{z_a^{(k)}\}, \dots, \{z_a^{(p)}\}) . \end{aligned} \quad (98)$$

By using these function, it is not hard to show that all the trace expressions can be computed as derivatives with respect to the $x^{(k)}$ variables. Indeed, in order to recover the expression listed in Sec 4.1.3 it's still necessary to average the generating function $F^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N)$ over the distribution of the symmetric matrices O over the Haar measure. In order to do that, we need to use a result proved in the next appendix, where it is shown that

$$\mathbb{E} \left[\prod_a \left(\sum_i O_{ia} \right)^{2n_a} \right] = N^n \frac{\Gamma(N/2)}{4^n \Gamma(N/2 + n)} \prod_a \frac{(2n_a)!}{n_a!} , \quad (99)$$

where $n = \sum_a n_a$. Then, we can expand $\mathbb{E} \left[F^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) \right]$ as

$$\begin{aligned} \mathbb{E} \left[F^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) \right] &= \mathbb{E} \left[\sum_{n^{(1)}} \dots \sum_{n^{(p)}} \prod_{k=1}^p \left\{ \frac{x^{(k)}}{N} \sum_a z_a^{(k)} \left(\sum_i O_{ia} \right)^2 \right\}^{n^{(k)}} \right] \\ &= \mathbb{E} \left[\sum_{\vec{n}^{(1)}} \dots \sum_{\vec{n}^{(p)}} \prod_{k=1}^p \prod_{a=1}^N \left[\left(\frac{x^{(k)} z_a^{(k)}}{4} \right)^{n_a^{(k)}} \frac{1}{n_a^{(k)}!} \right] \right] \\ &\times \frac{\Gamma(N/2)}{\Gamma(N/2 + \sum_k n^{(k)})} \prod_k \Gamma(n^{(k)} + 1) \frac{(2 \sum_k n^{(k)})!}{(\sum_k n^{(k)})!} , \end{aligned}$$

where the $\vec{n}^{(k)} = (n_1^{(k)}, \dots, n_N^{(k)})$ variables have been introduced by expanding the multinomials appearing in the first line. Then we can expand the products of Gamma functions as products of Beta functions by using

$$\begin{aligned} &\frac{\Gamma(N/2)}{\Gamma(N/2 + \sum_k n^{(k)})} \prod_k \Gamma(n^{(k)} + 1) \\ &= \frac{\Gamma(N/2)}{\Gamma(N/2 - p)} B(n^{(1)} + 1, n^{(2)} + 1) B(n^{(1)} + n^{(2)} + 2, n^{(3)} + 1) \dots B \left(\sum_k n^{(k)} + p, N/2 - p \right) \\ &= \int_0^1 dt_1 \dots dt_p, t_1^{n^{(1)}} (1 - t_1)^{n^{(2)}} t_2^{n^{(1)} + n^{(2)} + 1} (1 - t_2)^{n^{(3)}} \dots t_p^{\sum_k n^{(k)} + p - 1} (1 - t_p)^{N/2 - p - 1} \end{aligned}$$

in order to write

$$\begin{aligned} \overline{\mathbb{E}} \left[F^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) \right] &= \frac{\Gamma(N/2)}{\Gamma(N/2-p)} \int_0^1 dt_1 \dots dt_p \prod_{k=1}^p t_k^{k-1} (1-t_p)^{N/2-p-1} \\ &\times \sum_{\vec{n}^{(1)}} \dots \sum_{\vec{n}^{(p)}} \prod_{k=1}^p \prod_{a=1}^N \left[\left(\frac{x^{(k)} h^{(k)}(\vec{t}) z_a^{(k)}}{4} \right)^{n_a^{(k)}} \frac{1}{n_a^{(k)}!} \right] \frac{(2 \sum_k n^{(k)})!}{(\sum_k n^{(k)})!}, \end{aligned}$$

where the functions $h^{(k)}(\vec{t})$ are defined as in Sec. 4.1.3. Finally, one needs to employ recursively the identity

$$\sum_{n=0}^{\infty} t^n \frac{(2(n+m))!}{n!(n+m)!} \frac{(2m)!}{m!} (1-4t)^{-1/2-m} \quad (100)$$

in order to sum the series appearing in the above expression for $\overline{\mathbb{E}} \left[F^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) \right]$ and obtain

$$\begin{aligned} \overline{\mathbb{E}} \left[F^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) \right] &= \frac{\Gamma(N/2)}{\Gamma(N/2-p)} \int_0^1 dt_1 \dots dt_p \prod_{k=1}^p t_k^{k-1} (1-t_p)^{N/2-p-1} \\ &\times \prod_a \left(1 - \sum_{k=1}^p h^{(k)}(\vec{t}) x^{(k)} z_a^{(k)} \right)^{-1/2} \\ &= \mathcal{F}^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N). \end{aligned} \quad (101)$$

You can notice that all the observables listed in Sec. 4.1.3 are conveniently expressed in terms of the derivatives of $\mathcal{F}^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N)$. We finally remark that:

- Finding the derivatives $\mathcal{F}^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N)$ calculated in $\vec{x} = 0$ is straightforward, as the integrals in the \vec{t} variables factorize. This property can be used to find systematically the terms of a small $\bar{\alpha}$ expansion for the expression of the observables, and in particular has been used in order to write down the $\bar{\alpha} = 0$ expressions listed in Sec. 4.1.3.
- The leading term of the large N expansion of $\mathcal{F}^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N)$ is particularly simple, and results:

$$\mathcal{F}^{(p)}(\vec{x}, \{z_a^{(k)}\}_{a=1}^N) = \prod_k \left(1 - \frac{\bar{\alpha}}{N} \sum_a z_a^{(k)} \right)^{-1}, \quad (102)$$

indicating that, provided that the $z_a^{(k)}$ variables are of order 1, the generating function is also of order 1.

- In the degenerate case in which all the z_a are equal to z , then the generating function $\mathcal{F}^{(p)}(\vec{x}, \{z^{(k)}\}_{a=1}^N)$ is *exactly* (i.e., to all orders in N) equal to

$$\mathcal{F}^{(p)}(\vec{x}, \{z^{(k)}\}_{a=1}^N) = \prod_k \left(1 - \bar{\alpha} z^{(k)} \right)^{-1} \quad (103)$$

C Momenta over the Haar sphere

In this appendix we want to compute the momenta of random symmetric matrices O distributed according to the Haar measure associated to the group $O(N)$. In order to do that, we can exploit the Weingarten formalism [62, 63], which is used to compute generic expressions of the type

$$\mathbb{E}[O_{i_1 a_1} \dots O_{i_{2n} a_{2n}}] = \sum_{p, q \in \mathcal{P}_2(2n)} \delta_{i_1 i_{p(1)}} \dots \delta_{i_{2n} i_{p(2n)}} \delta_{a_1 a_{q(1)}} \dots \delta_{a_{2n} a_{q(2n)}} W^O(p, q) \quad (104)$$

where $\mathcal{P}_2(2n)$ is the set of all possible pairings of $\{1, \dots, 2n\}$ and $W^O(p, q)$ is the element (p, q) of the orthogonal Weingarten matrix [62, 63]. In particular, we want to exploit this formula in order to compute the expectation

$$\begin{aligned} & \mathbb{E} \left[\prod_a \left(\sum_i O_{ia} \right)^{2n_a} \right] \quad (105) \\ &= \sum_{i_1, j_1} \dots \sum_{i_n, j_n} \mathbb{E} \left[\underbrace{O_{i_1 1} O_{j_1 1} \dots O_{i_{n_1} 1} O_{j_{n_1} 1}}_{n_1 \text{ terms}} \dots \underbrace{O_{i_{n-n_N+1} N} O_{j_{n-n_N+1} N} \dots O_{i_n N} O_{j_n N}}_{n_N \text{ terms}} \right] \\ &= N^n \sum_{p, q \in \mathcal{P}_2(2n)} \delta_{a_1 a_{q(1)}} \dots \delta_{a_{2n} a_{q(2n)}} W^O(p, q). \quad (106) \end{aligned}$$

Its value can be obtained by solving the simpler problem of computing momenta of the components of a random vector $\vec{O}_i = (O_{i1}, \dots, O_{iN})$ distributed uniformly on the N -dimensional unit sphere. The expression for its momenta is in fact proportional to Eq. (106):

$$\mathbb{E} \left[\prod_a (O_{ia})^{2n_a} \right] = \sum_{p, q \in \mathcal{P}_2(2n)} \delta_{a_1 a_{q(1)}} \dots \delta_{a_{2n} a_{q(2n)}} W^O(p, q). \quad (107)$$

Hence, once the value of Eq. (107) has been computed, the value of Eq. (106) can be recovered after multiplying by N^n .

The value of Eq. (107) is easily obtained by constructing the generating function

$$Z[\vec{J}] = \int d\vec{u} e^{\vec{J} \cdot \vec{u}} \delta(|\vec{u}|^2 - 1), \quad (108)$$

which can be evaluated exponentiating the delta function:

$$Z[\vec{J}] \propto \int \frac{dz}{2\pi} d\vec{u} e^{\sum_{a=1}^N J_a u_a - iz(\sum_i u_a^2 - 1)} \quad (109)$$

$$= \int_{-\infty}^{+\infty} \frac{dz}{2\pi} e^{iz} \prod_a \left(\int_{-\infty}^{+\infty} du_a e^{J_a u_a - iz u_a^2} \right) \quad (110)$$

$$= \int_{-\infty}^{+\infty} \frac{dz}{2\pi} e^{iz} \prod_a e^{\frac{J_a^2}{4iz}} \left(\frac{\pi}{iz} \right)^{1/2} \quad (111)$$

$$= \int_{-\infty}^{+\infty} \frac{dz}{2\pi} e^{iz} \sum_{n=0}^{\infty} \frac{(\sum_a J_a^2/4)^n}{n!} \frac{\pi^{N/2}}{(iz)^{n+N/2}} \quad (112)$$

$$= \pi^{N/2} \sum_{n=0}^{\infty} \frac{(\sum_a J_a^2/4)^n}{n!} \int_{-\infty}^{+\infty} \frac{dz}{2\pi} e^{iz} \frac{1}{(iz)^{n+N/2}} \quad (113)$$

$$= \frac{\pi^{N/2}}{2} \sum_{n=0}^{\infty} \frac{(\sum_i J_a^2/4)^n}{\Gamma(n+1)\Gamma(n+N/2)}, \quad (114)$$

One gets finally

$$Z[\vec{J}] \propto 2\pi^{N/2} \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(n+1)\Gamma(n+N/2)}, \quad (115)$$

with $x = \sum_a J_a^2/4$.

By using this result, it is easy to calculate the average components of a vector uniformly distributed on the unit sphere:

$$\begin{aligned} \frac{1}{Z(0)} \frac{\partial^{2n} Z(\vec{J})}{\partial J_1^{2n_1} \dots \partial J_N^{2n_N}} \Big|_{\vec{J}=0} &= \mathbb{E} [(O_{i1})^{2n_1} \dots (O_{iN})^{2n_N}] \\ &= \frac{\Gamma(N/2)}{4^n \Gamma(N/2 + n)} \prod_{p=1}^m \frac{(2n_p)!}{n_p!}, \end{aligned} \quad (116)$$

where $n = \sum_{a=1}^N n_a$.

References

- [1] Robert M May. Will a large complex system be stable? *Nature*, 238:413–414, 1972.
- [2] Viktor K Jirsa and Mingzhou Ding. Will a large complex system with time delays be stable? *Physical review letters*, 93(7):070602, 2004.
- [3] Yoshihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.
- [4] Christopher A Sims. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48, 1980.

- [5] Yacine Aït-Sahalia, Julio Cacho-Diaz, and Roger JA Laeven. Modeling financial contagion using mutually exciting jump processes. Technical report, National Bureau of Economic Research, 2010.
- [6] Patricia Reynaud-Bouret, Vincent Rivoirard, Christine Tuleau-Malot, et al. Inference of functional connectivity in neurosciences via hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*, 2013.
- [7] Emmanuel Bacry, Khalil Dayri, and Jean-Francois Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5):1–12, 2012.
- [8] Vladimir Filimonov and Didier Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.
- [9] Stephen J Hardiman, Nicolas Bercot, and Jean-Philippe Bouchaud. Critical reflexivity in financial markets: a hawkes process analysis. *The European Physical Journal B*, 86(10):1–9, 2013.
- [10] Emmanuel Bacry and Jean-François Muzy. Hawkes model for price and trades high-frequency dynamics. *arXiv preprint arXiv:1301.1135*, 2013.
- [11] Pierre Brémaud, Laurent Massoulié, et al. Hawkes branching point processes without ancestors. *Journal of applied probability*, 38(1):122–135, 2001.
- [12] Thibault Jaisson and Mathieu Rosenbaum. Limit theorems for nearly unstable hawkes processes. *arXiv preprint arXiv:1310.2033*, 2013.
- [13] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.
- [14] Mark W Watson. Vector autoregressions and cointegration. *Handbook of econometrics*, 4:2843–2915, 1994.
- [15] M Potters, J-P Bouchaud, and L Laloux. Financial applications of random matrix theory: Old laces and new pieces. *Acta Physica Polonica B*, 36:2767, 2005.
- [16] Iacopo Mastromatteo, Matteo Marsili, and Patrick Zoi. Financial correlations at ultra-high frequency: theoretical models and empirical estimation. *The European Physical Journal B*, 80(2):243–253, 2011.
- [17] A.G. Hawkes. Point spectra of some mutually exciting point processes. *J. R. Statist. Soc. B*, 33:438–443, 1971.
- [18] A.G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
- [19] Y. Ogata. Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155(2-4):471–507, August 1999.
- [20] Agnès Helmstetter and Didier Sornette. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of geophysical research*, 107(B10):2237, 2002.
- [21] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Hal e-print, 00789127, To appear in Journal of Mathematical Neuroscience*.
- [22] E.. Lewis and G. Mohler. A nonparametric em algorithm for multiscale hawkes processes. *Preprint*, 2010.
- [23] G.O. Mohler, M.B. Short, P.J. Brantingham, F.P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106:100–108, 2011.

- [24] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.
- [25] S.H Yang and H Zha. Mixture of mutually exciting processes for viral diffusion. *Proceedings of the 30th International Conf. on Machine Learning*, 28, 2013.
- [26] Valérie Chavez-Demoulin, Anthony C Davison, and Alexander J McNeil. Estimating value-at-risk: a point process approach. *Quantitative Finance*, 5(2):227–234, 2005.
- [27] Patrick Hewlett. Clustering of order arrivals, price impact and trade path optimisation. In *Workshop on Financial Modeling with Jump processes, Ecole Polytechnique*, pages 6–8, 2006.
- [28] Clive G Bowsher. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912, 2007.
- [29] Luc Bauwens and Nikolaus Hautsch. *Modelling financial high frequency data using point processes*. Springer, 2009.
- [30] Kay Giesecke, Lisa R Goldberg, and Xiaowei Ding. A top-down approach to multivariate credit. *Operations Research*, 59(2):283–300, 2011.
- [31] Ioane Muni Toke. “market making” in an order book model and its impact on the spread. In *Econophysics of Order-driven Markets*, pages 49–64. Springer, 2011.
- [32] Paul Embrechts, Thomas Liniger, Lu Lin, et al. Multivariate hawkes processes: an application to financial data. *Journal of Applied Probability*, 48:367–378, 2011.
- [33] E Bacry, S Delattre, M Hoffmann, and JF Muzy. Some limit theorems for hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.
- [34] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.
- [35] Yoshihiko Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.
- [36] Madan Lal Mehta. *Random matrices*, volume 142. Academic press, 2004.
- [37] Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.
- [38] Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in mathematical physics*, 272(1):185–228, 2007.
- [39] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.
- [40] G Ben Arous and A Guionnet. Wigner matrices, 2011.
- [41] Craig A Tracy and Harold Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.
- [42] Craig A Tracy and Harold Widom. The distribution of the largest eigenvalue in the gaussian ensembles: $\beta = 1, 2, 4$. In *Calogero—Moser—Sutherland Models*, pages 461–472. Springer, 2000.

- [43] David S Dean and Satya N Majumdar. Extreme value statistics of eigenvalues of gaussian random matrices. *Physical Review E*, 77(4):041108, 2008.
- [44] David S Dean and Satya N Majumdar. Large deviations of extreme eigenvalues of random matrices. *Physical review letters*, 97(16):160201, 2006.
- [45] Satya N Majumdar and Grégory Schehr. Top eigenvalue of a random matrix: large deviations and third order phase transition. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(1):P01012, 2014.
- [46] Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564, 1955.
- [47] E Brézin and A Zee. Universality of the correlations between eigenvalues of large random matrices. *Nuclear Physics B*, 402(3):613–627, 1993.
- [48] CWJ Beenakker. Universality of brézin and zee’s spectral correlator. *Nuclear Physics B*, 422(3):515–520, 1994.
- [49] H Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.
- [50] Brendan D McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203–216, 1981.
- [51] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of erdős-rényi graphs ii: Eigenvalue spacing and the extreme eigenvalues. *Communications in Mathematical Physics*, 314(3):587–640, 2012.
- [52] László Erdős, Antti Knowles, Horng-Tzer Yau, Jun Yin, et al. Spectral statistics of erdős-rényi graphs i: Local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013.
- [53] Charles Bordenave, Pietro Caputo, Djalil Chafaï, et al. Spectrum of large random reversible markov chains: heavy-tailed weights on the complete graph. *The Annals of Probability*, 39(4):1544–1590, 2011.
- [54] AJ Bray and GJ Rodgers. Diffusion in a sparsely connected space: A model for glassy relaxation. *Physical Review B*, 38(16):11461, 1988.
- [55] GJ Rodgers and AJ Bray. Density of states of a sparse random matrix. *Physical Review B*, 37(7):3557, 1988.
- [56] G Biroli and R Monasson. A single defect approximation for localized states on random lattices. *Journal of Physics A: Mathematical and General*, 32(24):L255, 1999.
- [57] Guilhem Semerjian and Leticia F Cugliandolo. Sparse random matrices: the eigenvalue spectrum revisited. *Journal of Physics A: Mathematical and General*, 35(23):4837, 2002.
- [58] Sergey N Dorogovtsev, Alexander V Goltsev, José FF Mendes, and Alexander N Samukhin. Spectra of complex networks. *Physical Review E*, 68(4):046109, 2003.
- [59] Reimer Kühn. Spectra of sparse random matrices. *Journal of Physics A: Mathematical and Theoretical*, 41(29):295002, 2008.
- [60] Michael Krivelevich and Benny Sudakov. The largest eigenvalue of sparse random graphs. *Combinatorics, Probability & Computing*, 12(1):61–72, 2003.
- [61] Jorge Kurchan and Laurent Laloux. Phase space geometry and slow dynamics. *Journal of Physics A: Mathematical and General*, 29(9):1929, 1996.

- [62] Benoît Collins and Piotr Śniady. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.
- [63] Benoît Collins and Sho Matsumoto. On some properties of orthogonal weingarten functions. *Journal of Mathematical Physics*, 50(11):113516, 2009.