Information Transmission in Dynamical Networks:
The Normal Network Case

Giacomo Baggio, Virginia Rutten, Guillaume Hennequin, Sandro Zampieri

Abstract— Reliable information processing is a hallmark of many physical and biological networked systems. In this paper, we propose a novel framework for modelling information transmission within a linear dynamical network. Information propagation is modelled by means of a digital communication protocol that takes into account the realistic phenomenon of inter-symbol interference. Building on this framework, we adopt Shannon information rate to quantify the amount of information that can be reliably sent over the network within a fixed time window. We investigate how the latter information metric is affected by the connectivity structure of the network. Here, we focus in particular on networks characterized by a normal adjacency matrix. We show that for such networks the maximum achievable information rate depends only on the spectrum of the adjacency matrix. We then provide numerical results suggesting that non-normal network architectures could benefit information transmission in our framework.

Index Terms— Linear dynamical networks, information transmission over networks, digital communication, Shannon capacity, normal networks, matrix non-normality.

I. INTRODUCTION

A wide range of natural and engineered phenomena can be modelled as networks of interacting dynamical units. These units typically communicate with each other in order to accomplish some common goal. Seamless and profitable cooperation among units requires a robust and effective information transmission infrastructure at network level, especially in the presence of noise. Understanding how a network can enable efficient propagation of information across its units represents a central problem in many areas of engineering, biology and social sciences (see for instance [1]–[3], to cite just a few recent surveys). Perhaps, the most emblematic example is provided by the human brain wherein single neurons (or populations of neurons) are continuously rebroadcasting the information they receive to neighboring (populations of) neurons, in order to collectively execute complex tasks [4]–[6]. Unraveling the fundamental network principles at the core of efficient information transmission between neurons (or brain areas) is a current key challenge in neuroscience [3], [7]–[12].

G. Baggio is with the Department of Mechanical Engineering, University of California at Riverside, Riverside, CA, USA (email: gbaggio@engr.ucr.edu).

V. Rutten is with the Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge, UK and the Gatsby Computational Neuroscience Unit, University College London, London, UK.

G. Hennequin is with the Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge, UK.

S. Zampieri is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, Padova, Italy.

In this paper, we propose a novel framework for modelling digital communication and quantifying the amount of information that can reliably be propagated through networks of linear dynamical units. Our approach builds upon the assumption that information is transmitted periodically through the network and encoded in the impulse response of the system. Whenever the network units retain some memory of their previous states, past transmissions interfere with the current signal being transmitted, thereby limiting the amount of information that the network can reliably process. In contrast to other communication models proposed in the literature [13], [14], we explicitly incorporate this interference phenomenon in our model. We measure the efficiency of information transmission via Shannon's information rate and investigate how the structure of the network, described by its adjacency matrix, affects the latter information metric. Our primary aim is to dissect information transmission performance for the class of normal networks, i.e., networks characterized by a normal adjacency matrix. In particular, we show that, for such class of networks, the maximum achievable information rate is determined solely by the spectrum of the network adjacency matrix. Finally, numerical results seem to suggest that network non-normality could represent a valuable feature for enhancing the information transmission performance of a network. This corroborates recent works that have highlighted the beneficial role of matrix non-normality for the short-term memory storage in neuronal networks [15], [16] and in the controllability of large-scale networks [17], [18].

Notation: Given a matrix $A \in \mathbb{C}^{n \times n}$, $A^\top$ and A^* denote the transpose and Hermitian conjugate of A, respectively. $\mathrm{tr}(A)$ and $\det(A)$ stand for the trace and determinant of $A \in \mathbb{C}^{n \times n}$, respectively. $A \in \mathbb{C}^{n \times n}$ is said to be Hurwitz stable if all the eigenvalues of A have strictly negative real part. $A \in \mathbb{C}^{n \times n}$ is said to be normal if $AA^* = A^*A$, otherwise A is said to be non-normal. We write $A \geq 0$ ($A > 0$) to mean that the (Hermitian) matrix $A \in \mathbb{C}^{n \times n}$ is positive semidefinite (positive definite, resp.). We denote by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the (directed) graph with vertex (or node) set $\mathcal{V} = \{1, 2, \ldots, n\}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The (weighted) adjacency matrix $A \in \mathbb{R}^{n \times n}$ corresponding to the graph $\mathcal{G}$ satisfies $A_{ij} \neq 0$ iff $(j, i) \in \mathcal{E}$, where A_{ij} denotes the (i, j)-th entry of A. $\mathcal{N}(\mu, \Sigma)$ denotes the n-dimensional Gaussian distribution with mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in \mathbb{R}^{n \times n}$, $\Sigma \geq 0$. $\mathbb{E}[\cdot]$ stands for the expectation of a random variable. $\mathcal{L}_2^p[t_1, t_2]$ denotes the Hilbert space of p-dimensional square integrable functions in $[t_1, t_2]$, $t_2 > t_1$, equipped with the
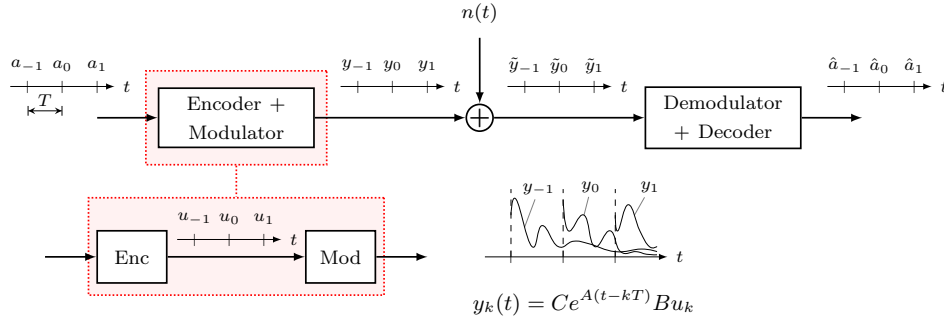
Fig. 1. Block diagram representation of the digital communication protocol considered in this paper.

inner product $\langle f, g \rangle_{\mathcal{L}_2} := \int_{t_1}^{t_2} f^\top(t)g(t)\mathrm{d}t$. Symbol $\delta(t)$ stands for the Dirac delta function and $\mathbf{1}(t)$ is the step function, i.e., $\mathbf{1}(t) = 0$ for $t < 0$ and $\mathbf{1}(t) = 1$ for $t \geq 0$.

We inform the reader that the present paper only reports some preliminary results. In particular, due to space limitation, some of the original proofs (specifically, proofs of Theorem 1, Theorem 2, and Corollary 1) have been omitted and will appear in a forthcoming and more complete publication.

II. MODELLING INFORMATION TRANSMISSION IN DYNAMICAL NETWORKS

We consider a continuous-time linear time-invariant dynamical system of the form

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{1}$$
$$y(t) = Cx(t), \tag{2}$$

where $x(t) \in \mathbb{R}^n$ denotes the state vector, $u(t) \in \mathbb{R}^m$ the input vector, and $y(t) \in \mathbb{R}^p$ the output vector at time $t \geq 0$. $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, and $C \in \mathbb{R}^{p \times n}$ denote the state, input, and output matrix, respectively. Throughout the paper, we will consider stable systems, namely systems where A is (Hurwitz) stable.

We interpret the dynamics expressed by (1)-(2) as those of a network, whose weighted adjacency matrix A corresponds to a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We structure B and C so as to single out two subsets of nodes $\mathcal{K} \subseteq \mathcal{V}$ and $\mathcal{T} \subseteq \mathcal{V}$, respectively:

$$B = [e_{k_1}, \ldots, e_{k_m}], \quad C = [e_{t_1}, \ldots, e_{t_p}]^\top,$$

where $\{e_i\}_{i=1}^n$ denotes the canonical vectors in $\mathbb{R}^n$ and $\{k_i\}_{i=1}^m \subseteq \mathcal{K}$, $\{t_i\}_{i=1}^p \subseteq \mathcal{T}$.

We model network information transmission by means of the digital communication protocol illustrated in the block diagram of Fig. 1. More precisely, we suppose that transmission occurs periodically every $t = kT$, $T > 0$, $k \in \mathbb{Z}$. T represents the transmission or sampling time of the communication channel. The to-be-transmitted piece of information is represented by a symbol a_k belonging to some alphabet $\mathcal{A}$ of finite cardinality. This symbol is mapped, through a suitable *encoding* procedure, to a vector $u_k \in \mathbb{R}^m$ (the input codeword). This vector then acts as an impulsive input $u(t) = u_k\delta(t)$ that transiently excites the linear dynamical system in (1)-(2). The subsequent

evolution of the system in the time interval $[kT, (k+1)T]$ constitutes the *modulation* stage of the protocol. In digital communication theory, a modulator is a function mapping a discrete symbol to a continuous-time trajectory. In our case, the modulator maps the input impulse into the p-dimensional square-integrable impulse response

$$y_k(t) = Ce^{At}Bu_k\mathbf{1}(t - kT) \in \mathcal{L}_2^p[kT, (k+1)T]$$

generated by the dynamical system (1)-(2). After this stage, the modulated signal is sent, via an additive white Gaussian noise (AWGN) channel, to the receiver. At the receiver side, the corrupted trajectory $\tilde{y}_k(t) = y_k(t) + n(t)$, with $n(t) \sim \mathcal{N}(0, \sigma^2 I)$, $\mathbb{E}[n(t)n^\top(s)] = \sigma^2 I\delta(t - s)$, for all $t, s \in \mathbb{R}$, is first demodulated and then decoded in order to obtain an estimate $\hat{a}_k$ of the transmitted symbol.

A crucial feature of our model relies on the fact that the memory of the system cannot be instantaneously erased at the end of every transmission.[1] This implies that previous modulated waveforms interfere with the waveform of the currently transmitted symbol, as depicted in the bottom of Fig. 1. In digital communication jargon, this phenomenon is known as inter-symbol interference (ISI). More precisely, for a transmission occurring at time $t = 0$, the "interference term" in our model is the sum of contributions in the time window $[0, T]$ made by modulated waveforms left over from previous transmissions at times $t = -kT$, $k \in \mathbb{Z}$, $k > 1$:

$$i(t) := \sum_{k=1}^{\infty} Ce^{A(t+kT)}Bu_{-k}, \quad t \in [0, T]. \tag{3}$$

In the next section, we develop a metric for quantifying information transmission performance within the digital communication framework introduced above.

III. AN INFORMATION TRANSMISSION METRIC

To measure the "efficiency" of information transmission over the previously described noisy digital communication channel, we turn to the notion of Shannon information rate, defined as

$$\mathcal{R} := \frac{\mathcal{C}}{T}, \tag{4}$$

where $\mathcal{C}$ denotes the Shannon capacity of the channel and $T > 0$ is the finite transmission time. We recall that Shannon

---

[1]This phenomenon arises naturally in any physical network whose units retain some memory of their previous state, e.g. in a neuronal network [6].

capacity $\mathcal{C}$ provides a tight upper bound on the number of bits per input symbol that can be sent reliably, i.e., with arbitrarily small decoding error probability, over the channel [19, Ch. 7]. Consequently, $\mathcal{R}$ measures the amount of information that can be sent reliably over the channel within a fixed transmission time interval. With reference to the communication channel described in the previous section, we will exploit the following additional assumptions:

A1) the available power at the sender just before the modulation stage is limited by $P > 0$, namely $\|u_k\|^2 \leq P$, $k \in \mathbb{Z}$;

A2) sufficiently many messages have been sent at the time of current transmission, such that the distribution of interference terms has reached an equilibrium;

A3) the input codebook is the same at every transmission and input codewords corresponding to different transmissions are independent of each other;

A4) the noise $n(t)$ is independent of current and past input codewords.

Under these assumptions, we have the following result.

*Theorem 1 (**Shannon information rate**):* The Shannon information rate of the communication channel described in Sec. II under the input power constraint $P > 0$ is given by

$$\mathcal{R} = \frac{1}{2T} \max_{\Sigma \geq 0, \mathrm{tr}\, \Sigma \leq P} \log_2 \frac{\det\left(\sigma^2 I + \mathcal{O}\mathcal{W}\right)}{\det\left(\sigma^2 I + \mathcal{O}(\mathcal{W} - B\Sigma B^\top)\right)}, \tag{5}$$

where $\mathcal{O} := \int_0^T e^{A^\top t} C^\top C e^{At}\, \mathrm{d}t$ denotes the $[0, T]$-observability Gramian of the pair (A, C) and $\mathcal{W} := \sum_{k=0}^{\infty} e^{AkT} B\Sigma B^\top e^{A^\top kT}$ denotes the discrete-time controllability Gramian of the pair $(e^{AT}, B\Sigma^{1/2})$.

From the expression of $\mathcal{R}$ in (5), we note that the computation of the information rate boils down to solving a (generally non-convex) optimization problem over the space of positive semi-definite $m \times m$ matrices with trace less or equal than one. Further, it is worth observing that the numerator $\det(\sigma^2 I + \mathcal{O}\mathcal{W})$ can be thought of as a volumetric measure of the "energy" of the corrupted modulated signal $\tilde{y}(t)$ in $[0, T]$, whereas the denominator $\det\left(\sigma^2 I + \mathcal{O}(\mathcal{W} - B\Sigma B^\top)\right)$ can be thought of as a volumetric measure of the "energy" of the overall channel noise in $[0, T]$, consisting of the sum of the interference term $i(t)$ and the background noise $n(t)$.

IV. PRELIMINARY RESULTS

In this section, we present a number of auxiliary results concerning the simplification and monotonicity of the information rate formula derived in the previous section.

The first result asserts that in the maximization problem in Eq. (5), the maximizer always saturates the input power constraint.

*Lemma 1 (**Optimal Σ has trace equal to P**):* Consider the information rate in (5). It holds

$$\mathcal{R} = \frac{1}{2T} \max_{\Sigma \geq 0, \mathrm{tr}\, \Sigma = P} \log_2 \frac{\det(\sigma^2 I + \mathcal{O}\mathcal{W})}{\det(\sigma^2 I + \mathcal{O}(\mathcal{W} - B\Sigma B^\top))}.$$

Proof: Let us define

$$f(\Sigma) := \frac{1}{2T} \log_2 \frac{\det(\sigma^2 I + \mathcal{O}\mathcal{W})}{\det(\sigma^2 I + \mathcal{O}(\mathcal{W} - B\Sigma B^\top))}.$$

We will show that $\frac{\mathrm{d}f(\alpha\Sigma)}{\mathrm{d}\alpha} \geq 0$, $\alpha \in \mathbb{R}$, $\alpha > 0$. This implies that $f(\alpha\Sigma)$ is a monotonically increasing function of $\alpha > 0$, so that the optimal Σ maximizing $f(\Sigma)$ under the constraint $\mathrm{tr}\, \Sigma \leq P$ must satisfy the latter constraint with equality. To this end, we have

$$\begin{aligned}
f(\alpha\Sigma) &= \frac{1}{2T} \log_2 \frac{\det(\sigma^2 I + \alpha\mathcal{O}\mathcal{W})}{\det(\sigma^2 I \alpha\mathcal{O}(\mathcal{W} - B\Sigma B^\top))} \\
&= \frac{1}{2T} \log_2 \frac{\det(\alpha\mathcal{O}^{1/2}\mathcal{W}\mathcal{O}^{1/2} + \sigma^2 I)}{\det(\alpha\mathcal{O}^{1/2}(\mathcal{W} - B\Sigma B^\top)\mathcal{O}^{1/2} + \sigma^2 I)} \\
&= \frac{1}{2T} \sum_{i=1}^{n} \log_2(\alpha\lambda_i + \sigma^2) - \frac{1}{2T} \sum_{i=1}^{n} \log_2(\alpha\mu_i + \sigma_r^2),
\end{aligned}$$

where $\{\lambda_i\}_{i=1}^n$ and $\{\mu_i\}_{i=1}^n$ denotes the ordered eigenvalues of $\mathcal{O}^{1/2}\mathcal{W}\mathcal{O}^{1/2}$ and $\mathcal{O}^{1/2}(\mathcal{W} - B\Sigma B^\top)\mathcal{O}^{1/2}$, respectively. Taking the derivative of the previous expression w.r.t. α, we get

$$\frac{\mathrm{d}f(\alpha\Sigma)}{\mathrm{d}\alpha} = \frac{1}{T\ln 2} \sum_{i=1}^{n} \frac{\lambda_i}{\alpha\lambda_i + \sigma^2} - \frac{1}{T\ln 2} \sum_{i=1}^{n} \frac{\mu_i}{\alpha\mu_i + \sigma^2}.$$

Since $\mathcal{O}^{1/2}\mathcal{W}\mathcal{O}^{1/2} \geq \mathcal{O}^{1/2}(\mathcal{W} - B\Sigma B^\top)\mathcal{O}^{1/2}$, then $\lambda_i \geq \mu_i$, $i = 1, 2, \ldots, n$. This implies that

$$\frac{\mathrm{d}f(\alpha\Sigma)}{\mathrm{d}\alpha} = \frac{1}{T\ln 2} \sum_{i=1}^{N} \frac{\lambda_i - \mu_i}{(\lambda_i\alpha + \sigma^2)(\mu_i\alpha + \sigma^2)} \geq 0,$$

and this ends the proof. ∎

Another interesting property of $\mathcal{R}$ is described in the following lemma.

*Lemma 2 (**Scaling invariance of $\mathcal{R}$ w.r.t. P and σ^2**):* Consider the information rate in (5). For all $\alpha \in \mathbb{R}$, $\alpha > 0$, it holds

$$\mathcal{R}(P, \sigma^2) = \mathcal{R}(\alpha P, \alpha\sigma^2),$$

where we made explicit the dependence of $\mathcal{R}$ on P and σ^2.

Proof: First observe that if we replace Σ and σ^2 by $\alpha\Sigma$ and $\alpha\sigma^2$, respectively, then the Gramian $\mathcal{W}$ is replaced by $\alpha\mathcal{W}$. This in turn implies that the value of

$$f(\Sigma, \sigma^2) := \frac{1}{2T} \log_2 \frac{\det(\sigma^2 I + \mathcal{O}\mathcal{W})}{\det(\sigma^2 I + \mathcal{O}(\mathcal{W} - B\Sigma B^\top))}$$

is not affected by this change of variables. If $\Sigma^\star$ is the input covariance maximizing $f(\Sigma, \sigma^2)$ under the constraint $\mathrm{tr}\, \Sigma \leq P$, it holds

$$\begin{aligned}
\mathcal{R}(\alpha P, \alpha\sigma^2) &\geq f(\alpha\Sigma^\star, \alpha\sigma^2) \\
&= f(\Sigma^\star, \sigma^2) = \mathcal{R}(P, \sigma^2). \tag{6}
\end{aligned}$$

On the other hand, let $\Sigma^{\star\star}$ now denote the input covariance maximizing $f(\Sigma, \alpha\sigma^2)$ under the constraint $\mathrm{tr}\, \Sigma \leq \alpha P$. It holds

$$\mathcal{R}(P, \sigma^2) \geq f\left(\frac{1}{\alpha}\Sigma^{\star\star}, \sigma^2\right)$$

$$= f(\Sigma^{\star\star}, \alpha\sigma^2) = \mathcal{R}(\alpha P, \alpha\sigma^2). \qquad (7)$$

Eventually, a combination of (6) and (7) yields $\mathcal{R}(\alpha P, \alpha\sigma^2) = \mathcal{R}(P, \sigma^2)$. ∎

As a consequence of Lemmas 1 and 2, we have that $\mathcal{R}$ can be equivalently written in a more simplified form as

$$\mathcal{R} = \frac{1}{2T} \max_{\Sigma \geq 0, \operatorname{tr}\Sigma=1} \log_2 \frac{\det(I + \mathrm{SNR}\,\mathcal{O}\mathcal{W})}{\det(I + \mathrm{SNR}\,\mathcal{O}(\mathcal{W} - B\Sigma B^\top))}. \qquad (8)$$

where the input power constraint has been normalized to one and the symbol $\mathrm{SNR} := P/\sigma^2$ denotes the signal-to-noise ratio of the communication channel.

We conclude this section with a monotonicity result on the structure of the input and output node subsets.

Proposition 1 (Monotonicity of $\mathcal{R}$ w.r.t. $\mathcal{K}$ and $\mathcal{T}$): Let $\mathcal{K}_1, \mathcal{K}_2 \subseteq \mathcal{V}$ be two input nodes subsets and $\mathcal{T}_1, \mathcal{T}_2 \subseteq \mathcal{V}$ be two output nodes subsets. Suppose that the pair $(A, C_{\mathcal{T}_i})$, $i = 1, 2$, is observable, where $C_{\mathcal{T}_i}$ denotes the output matrix corresponding to subset $\mathcal{T}_i$. If $\mathcal{K}_1 \subseteq \mathcal{K}_2$ and $\mathcal{T}_1 \subseteq \mathcal{T}_2$, then

$$\mathcal{R}(\mathcal{K}_1, \mathcal{T}_1) \leq \mathcal{R}(\mathcal{K}_2, \mathcal{T}_2),$$

where we made explicit the dependence of $\mathcal{R}$ on the input subset $\mathcal{K}$ and output subset $\mathcal{T}$.

Proof: Let $\mathcal{K}$ and $\mathcal{T}$ be two input and output node subsets and let $B_\mathcal{K}$ and $C_\mathcal{T}$ denote the corresponding input and output matrices. Let us define

$$f(\Sigma, \mathcal{K}, \mathcal{T}) := \frac{1}{2T} \log_2 \frac{\det(I + \mathrm{SNR}\,\mathcal{O}_\mathcal{T}\mathcal{W}_\mathcal{K})}{\det(I + \mathrm{SNR}\,\mathcal{O}_\mathcal{T}(\mathcal{W}_\mathcal{K} - \Sigma_\mathcal{K}))},$$

where $\Sigma_\mathcal{K} := B_\mathcal{K}\Sigma B_\mathcal{K}^\top$, $\mathcal{O}_\mathcal{T} := \int_0^T e^{A^\top t} C_\mathcal{T}^\top C_\mathcal{T} e^{At}\mathrm{d}t$ and $\mathcal{W}_\mathcal{K} := \sum_{k=0}^\infty e^{AkT}\Sigma_\mathcal{K} e^{A^\top kT}$. Under the assumption that the pair $(A, C_\mathcal{T})$ is observable (which in turn implies that $\mathcal{O}_\mathcal{T}^{-1}$ is non-singular), it is possible to rewrite $f(\Sigma, \mathcal{K}, \mathcal{T})$ as

$$f(\Sigma, \mathcal{K}, \mathcal{T}) = \frac{1}{2T} \log_2 \det\Big[(I + \mathrm{SNR}\,\mathcal{O}_\mathcal{T}\mathcal{W}_\mathcal{K})$$
$$\cdot (I + \mathrm{SNR}\,\mathcal{O}_\mathcal{T}(\mathcal{W}_\mathcal{K} - \Sigma_\mathcal{K}))^{-1}\Big]$$
$$= \frac{1}{2T} \log_2 \det\Big[I + \Sigma_\mathcal{K}^{1/2}(e^{AT}\mathcal{W}_\mathcal{K} e^{A^\top T}$$
$$+ \mathcal{O}_\mathcal{T}^{-1}/\mathrm{SNR})^{-1}\Sigma_\mathcal{K}^{1/2}\Big], \quad (9)$$

Since $\mathcal{T}_1 \subseteq \mathcal{T}_2$, we first note that $\mathcal{O}_{\mathcal{T}_1} \leq \mathcal{O}_{\mathcal{T}_2}$. We have the following chain of implications

$$\mathcal{O}_{\mathcal{T}_1} \leq \mathcal{O}_{\mathcal{T}_2} \;\Rightarrow\; \mathcal{O}_{\mathcal{T}_1}^{-1} \geq \mathcal{O}_{\mathcal{T}_2}^{-1}$$
$$\Rightarrow\; \widetilde{\mathcal{W}}_\mathcal{K} + \frac{\mathcal{O}_{\mathcal{T}_1}^{-1}}{\mathrm{SNR}} \geq \widetilde{\mathcal{W}}_\mathcal{K} + \frac{\mathcal{O}_{\mathcal{T}_2}^{-1}}{\mathrm{SNR}}, \; \forall \mathcal{K} \subseteq \mathcal{V},$$
$$\Rightarrow\; \Big(\widetilde{\mathcal{W}}_\mathcal{K} + \frac{\mathcal{O}_{\mathcal{T}_1}^{-1}}{\mathrm{SNR}}\Big)^{-1} \leq \Big(\widetilde{\mathcal{W}}_\mathcal{K} + \frac{\mathcal{O}_{\mathcal{T}_2}^{-1}}{\mathrm{SNR}}\Big)^{-1}, \forall \mathcal{K} \subseteq \mathcal{V},$$

where $\widetilde{\mathcal{W}}_\mathcal{K} := e^{AkT}\mathcal{W}_\mathcal{K} e^{A^\top kT}$. In view of (9), the latter inequality in turn implies that

$$f(\Sigma, \mathcal{K}, \mathcal{T}_1) \leq f(\Sigma, \mathcal{K}, \mathcal{T}_2), \quad \forall \mathcal{K} \subseteq \mathcal{V} \text{ and } \forall \Sigma \geq 0. \quad (10)$$

Next, let us define $\mathcal{S}_i := \{\Sigma_{\mathcal{K}_i} : \Sigma \geq 0, \operatorname{tr}\Sigma = 1\}$, $i = 1, 2$. Since $\mathcal{K}_1 \subseteq \mathcal{K}_2$, we have $\mathcal{S}_1 \subseteq \mathcal{S}_2$. By combining the latter fact with inequality (10),

$$\mathcal{R}(\mathcal{K}_1, \mathcal{T}_1) = \max_{\Sigma \geq 0, \operatorname{tr}\Sigma=1} f(\Sigma, \mathcal{K}_1, \mathcal{T}_1)$$
$$= \max_{\Sigma_1 \in \mathcal{S}_1} f(\Sigma_1, \mathcal{V}, \mathcal{T}_1)$$
$$\leq \max_{\Sigma_1 \in \mathcal{S}_1} f(\Sigma_1, \mathcal{V}, \mathcal{T}_2)$$
$$\leq \max_{\Sigma_2 \in \mathcal{S}_2} f(\Sigma_2, \mathcal{V}, \mathcal{T}_2)$$
$$= \max_{\Sigma \geq 0, \operatorname{tr}\Sigma=1} f(\Sigma, \mathcal{K}_2, \mathcal{T}_2) = \mathcal{R}(\mathcal{K}_2, \mathcal{T}_2),$$

and this concludes the proof. ∎

Remark 1: It is worth observing that the previous proposition in particular implies that $\mathcal{R}(\mathcal{K}_1, \mathcal{T}_1) \leq \mathcal{R}(\mathcal{V}, \mathcal{V})$. This means then the maximum information rate is always attained by choosing all network nodes as input and output nodes.

V. ANALYSIS OF NETWORK INFORMATION RATE

In this section, we investigate how the connectivity structure of the network, encoded by matrix A, affects the information rate $\mathcal{R}$ introduced in Sec. II. We start by analyzing the simplest possible case, i.e., the case of scalar A, and then we move to the case of normal A.

A. Scalar case

When $n = 1$, the simplified expression of $\mathcal{R}$ in (8) reduces to

$$\mathcal{R} = \frac{1}{2T} \log_2 \frac{2a + \mathrm{SNR}}{2a + \mathrm{SNR}\,e^{-2aT}}, \qquad (11)$$

where $B = C = 1$ and $A = -a$, $a \in \mathbb{R}$, $a > 0$, in view of stability. In this case, $\mathcal{R}$ is a monotonically decreasing function of the transmission time T. Hence, the optimal value of $\mathcal{R}$ is attained for $T \to 0$ and is equal to

$$\mathcal{R}_{\max} := \max_{T \geq 0} \mathcal{R} = \frac{1}{\ln 2} \frac{a\mathrm{SNR}}{2a + \mathrm{SNR}}. \qquad (12)$$

The latter quantity is a monotonically increasing function of a, so that its maximum value is attained for $a \to \infty$ and is equal to $\overline{\mathcal{R}}_a := \frac{\mathrm{SNR}}{2\ln 2}$ (see also the top plot of Fig. 2). Intuitively, for small values of a and T the ISI term becomes large enough to be detrimental for the rate. Lastly, as SNR increases, $\mathcal{R}$ in (11) increases as well until it saturates to the limit value $\overline{\mathcal{R}}_{\mathrm{SNR}} := \frac{a}{\ln 2}$ (see also the bottom plot of Fig. 2). Again, this finite upper bound is due to the ISI term which penalizes the transmission performance even in presence of vanishing background noise.

B. Normal network case

We derive here an explicit expression for the information rate of normal networks. We suppose that the dynamical network is fully controllable ($\mathcal{K} = \mathcal{V}$) and fully observable ($\mathcal{T} = \mathcal{V}$). The next result asserts that the information rate of a $n \times n$ normal network can be decoupled in the sum of the information rates of n independent scalar channels whose behavior depends on the eigenvalues of A.
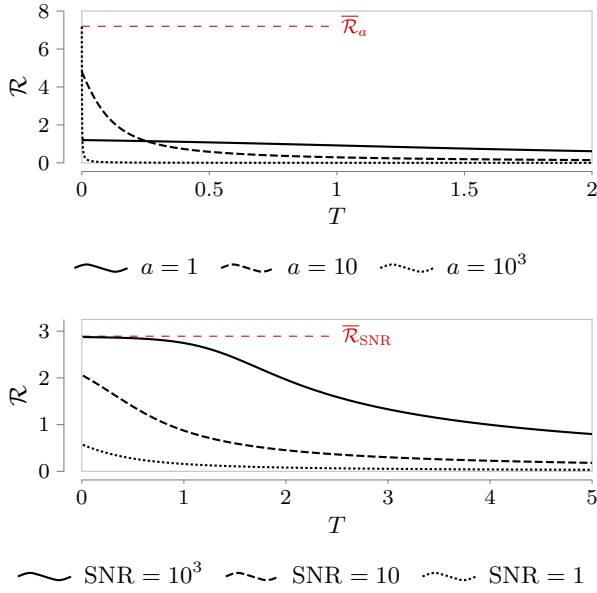
Fig. 2. Plots of $\mathcal{R}$ against transmission time T in the scalar case ($n = 1$). Top plot: SNR = 10. Bottom plot: $a = 2$. The dashed red line denotes the limit values $\overline{\mathcal{R}}_a = \lim_{a \to \infty} \mathcal{R}_T = 5/\ln 2$ and $\overline{\mathcal{R}}_{\mathrm{SNR}} = \lim_{\mathrm{SNR} \to \infty} \mathcal{R}_T = 2/\ln 2$.

Theorem 2 (Information rate for normal networks): If $\mathcal{K} = \mathcal{T} = \mathcal{V}$ and $A \in \mathbb{R}^{n \times n}$ is a normal and stable matrix with eigenvalues $\{\lambda_i\}_{i=1}^n$, then

$$\mathcal{R} = \max_{\substack{\{P_i\}_{i=1}^n, P_i \geq 0 \\ \text{s.t. } \sum_{i=1}^n P_i = P}} \sum_{i=1}^n \mathcal{R}(\mathrm{SNR}_i, \lambda_i), \qquad (13)$$

with $\mathrm{SNR}_i := P_i/\sigma^2$ and

$$\mathcal{R}(\mathrm{SNR}, \lambda) := \frac{1}{2T} \log_2 \frac{\mathrm{SNR} - 2\mathrm{Re}\,\lambda}{\mathrm{SNR}\, e^{2T\mathrm{Re}\,\lambda} - 2\mathrm{Re}\,\lambda}. \qquad (14)$$

From the above theorem we have the following interesting corollary.

Corollary 1 (Optimal rate for normal networks): If $\mathcal{K} = \mathcal{T} = \mathcal{V}$ and $A \in \mathbb{R}^{n \times n}$ is a normal and stable matrix, then the maximum of $\mathcal{R}$ is achieved for $T \to 0$ and the optimal signal-to-noise ratio allocation is

$$\mathrm{SNR}_i = \frac{\mathrm{Re}\,\lambda_i}{\mathrm{tr}\,A} \mathrm{SNR},$$

which yields

$$\mathcal{R}_{\max} = \max_{T \geq 0} \mathcal{R} = \frac{1}{\ln 2} \frac{\mathrm{tr}\,A\,\mathrm{SNR}}{2\,\mathrm{tr}\,A - \mathrm{SNR}}. \qquad (15)$$

It is worth pointing out that the expression in (15) coincides with the maximum information rate derived for the scalar case in (12), after replacing a with $-\mathrm{tr}\,A$. Therefore, the same considerations made for the scalar case hold also for the normal case.

Remark 2: In light of Proposition 1 and Remark 1, the results in this subsection (Theorem 2 and Corollary 1) can be also interpreted as *fundamental limitations* for the information rate of normal networks regardless of the choice of $\mathcal{K}$ and $\mathcal{T}$. Specifically, (13) provides an upper bound

to the achievable information rate of any normal network w.r.t. *any* choice of input and output node subsets $\mathcal{K}$, $\mathcal{T}$. Furthermore, the expression of $\mathcal{R}_{\max}$ in (15) corresponds to the maximum information rate that can be attained by a network described by a normal matrix A for a given SNR level. Observe also that in the ideal scenario of infinite SNR, the optimal information rate reduces to $-\frac{1}{\ln 2} \mathrm{tr}\,A$.

VI. NUMERICAL EXAMPLES

To illustrate the results outlined in the previous section, we consider the following n-dimensional Toeplitz line network
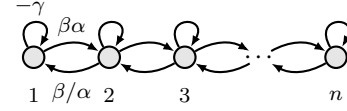
described by the adjacency matrix

$$A = \begin{bmatrix} -\gamma & \beta/\alpha & & \\ \beta\alpha & -\gamma & \ddots & \\ & \ddots & \ddots & \beta/\alpha \\ & & \beta\alpha & -\gamma \end{bmatrix} \in \mathbb{R}^{n \times n}, \qquad (16)$$

with $\alpha, \beta, \gamma > 0$ such that $\gamma > 2\beta$ to ensure stability. The parameter $\alpha > 0$ regulates the "anisotropy" of the network, so that it can be thought of as an index of network non-normality.
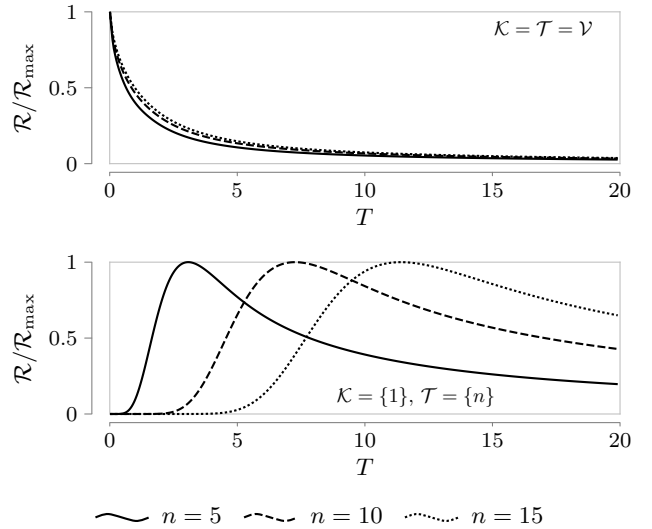

Fig. 3. Normalized plots of $\mathcal{R}$ against transmission time T for the line network in (16) with SNR = 1, $\alpha = 1$, $\beta = 1$, and $\gamma = 2.5$. Top plot: $\mathcal{K} = \mathcal{T} = \mathcal{V}$. Bottom plot: $\mathcal{K} = \{1\}$, $\mathcal{T} = \{n\}$. Notice that the curves obtained for $\mathcal{K} = \mathcal{T} = \mathcal{V}$ coincides (up to a normalization factor) with the ones of the scalar case in Fig. 2, after replacing a with $-\mathrm{tr}\,A$.

We first focus on the normal case, i.e., we set $\alpha = 1$. In Figure 3 the "normalized" time behavior of $\mathcal{R}$ for unit SNR and different values of n is illustrated. From this figure, it is worth noticing that in the fully observable and fully controllable case ($\mathcal{K} = \mathcal{T} = \mathcal{V}$) $\mathcal{R}$ is a monotonically decreasing function of T, in agreement with the expression derived in Theorem 2 (top plot); whilst the monotonic

behavior breaks down if the input and output nodes are selected to be the most distanced ones ($\mathcal{K} = \{1\}$, $\mathcal{T} = \{n\}$), in that the information rate exhibits a peak for an optimal T different from zero (bottom plot). Further, observe that the latter optimal transmission time scales proportionally to n. Figure 4 displays the behavior of $\mathcal{R}_{\max} = \max_{T \geq 0} \mathcal{R}$ against the noise variance σ^2 for a unit input power constraint, $n = 8$ and four different choices of $\mathcal{K}$ and $\mathcal{T}$. From the plot, it can be noticed that the upper bound in Corollary 1 (obtained by choosing $\mathcal{K} = \mathcal{T} = \mathcal{V}$) is quite tight in case $\mathcal{K} = \{1\}$, $\mathcal{T} = \mathcal{V}$ and $\mathcal{K} = \mathcal{V}$, $\mathcal{T} = \{1\}$, whereas it becomes rather uninformative in case of most distanced input and output nodes, i.e., $\mathcal{K} = \{1\}$, $\mathcal{T} = \{n\}$.
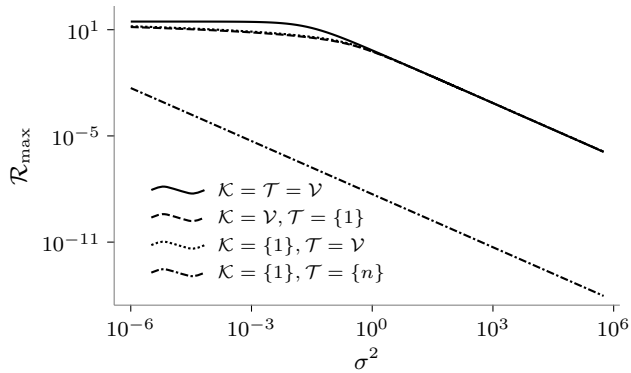

Fig. 4. Logarithmic plot of $\mathcal{R}_{\max}$ against σ^2 for $P = 1$ and different choices of $\mathcal{K}$ and $\mathcal{T}$ for the line network in (16) with $n = 8$, $\alpha = 1$, $\beta = 1$ and $\gamma = 2.5$. The curves obtained for $\mathcal{K} = \mathcal{V}$, $\mathcal{T} = \{1\}$ (dashed line) and $\mathcal{K} = \{1\}$, $\mathcal{T} = \mathcal{V}$ (dotted line) are almost overlapping.

In Fig. 5 we compared the behavior of $\mathcal{R}$ for a normal network ($\alpha = 1$) and a non-normal one ($\alpha = 3.5$) in the fully observable and fully controllable case. From the figure, two important observations can be made: (i) in the non-normal network case the optimal transmission time is different from zero, even though $\mathcal{K} = \mathcal{T} = \mathcal{V}$; (ii) the value of $\mathcal{R}_{\max}$ in the non-normal case is higher that the one obtained in the normal case, even though the two network matrices share the same spectrum. The latter fact in particular suggests that non-normality could represent a valuable feature for increasing the information transmission performance of a network.
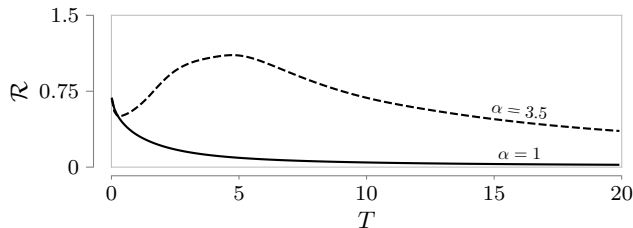

Fig. 5. Plot of $\mathcal{R}$ against transmission time T for the line network in (16) with $n = 10$, SNR $= 1$, $\beta = 1$, $\gamma = 2.5$ and two different values of α.

VII. Concluding remarks and perspectives

In this work, we have proposed a new communication framework to describe the propagation of information in networks of interconnected linear dynamical systems. We measured information transmission performance by means of the Shannon information rate, and analyzed how the connectivity structure of the network influences this information metric. We proved that for a normal network, the information rate is entirely determined by the eigenvalues of the adjacency matrix (Theorem 2 and Corollary 1). We validated our theoretical results with a simple yet paradigmatic numerical example, which further suggested that network non-normality could substantially enhance the quality of information transmission. Further research is needed to better explore this conjecture. This could be done by specifically investigating how network diameter, recently shown to be related to the non-normality of the adjacency matrix [20], affects the Shannon information rate.

References

[1] G. Tkačik and A. M. Walczak, "Information transmission in genetic regulatory networks: a review," *J. Phys. Condens. Matter*, vol. 23, no. 15, p. 153102, 2011.

[2] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM Sigmod Rec.*, vol. 42, no. 2, pp. 17–28, 2013.

[3] A. Avena-Koenigsberger, B. Misic, and O. Sporns, "Communication dynamics in complex brain networks," *Nat. Rev. Neurosci.*, vol. 19, no. 1, p. 17, 2018.

[4] G. Hennequin, T. P. Vogels, and W. Gerstner, "Optimal control of transient dynamics in balanced networks supports generation of complex movements," *Neuron*, vol. 82, no. 6, pp. 1394–1406, 2014.

[5] M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy, "Neural population dynamics during reaching," *Nature*, vol. 487, no. 7405, p. 51, 2012.

[6] T. P. Vogels, K. Rajan, and L. F. Abbott, "Neural network dynamics," *Annu. Rev. Neurosci.*, vol. 28, pp. 357–376, 2005.

[7] J. M. Beggs and D. Plenz, "Neuronal avalanches in neocortical circuits," *J. Neurosci.*, vol. 23, no. 35, pp. 11 167–11 177, 2003.

[8] T. Akam and D. M. Kullmann, "Oscillations and filtering networks support flexible routing of information," *Neuron*, vol. 67, no. 2, pp. 308–320, 2010.

[9] W. L. Shew, H. Yang, S. Yu, R. Roy, and D. Plenz, "Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches," *J. Neurosci.*, vol. 31, no. 1, pp. 55–63, 2011.

[10] T. P. Vogels and L. Abbott, "Gating multiple signals through detailed balance of excitation and inhibition in spiking networks," *Nat. Neurosci.*, vol. 12, no. 4, p. 483, 2009.

[11] T. P. Vogels and L. F. Abbott, "Signal propagation and logic gating in networks of integrate-and-fire neurons," *J. Neurosci.*, vol. 25, no. 46, pp. 10 786–10 795, 2005.

[12] V. Rutten and G. Hennequin, "Limits on fast, high-dimensional information processing in recurrent circuits," in *Computational and Systems Neuroscience (Cosyne)*, 2017, conference abstract.

[13] C. Kirst, M. Timme, and D. Battaglia, "Dynamic information routing in complex networks," *Nat. Commun.*, vol. 7, p. 11061, 2016.

[14] U. Harush and B. Barzel, "Dynamic patterns of information flow in complex networks," *Nat. Commun.*, vol. 8, p. 2181, 2017.

[15] S. Ganguli, D. Huh, and H. Sompolinsky, "Memory traces in dynamical systems," *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 48, pp. 18 970–18 975, 2008.

[16] S. Lim and M. S. Goldman, "Noise tolerance of attractor and feedforward memory models," *Neural Comput.*, vol. 24, no. 2, pp. 332–390, 2012.

[17] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 40–52, 2014.

[18] F. Pasqualetti and S. Zampieri, "On the controllability of isotropic and anisotropic networks," in *Proc. 53rd IEEE Conf. Decis. Control (CDC)*, 2014, pp. 607–612.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. John Wiley & Sons, 2012.

[20] G. Baggio and S. Zampieri, "On the relation between non-normality and diameter in linear dynamical networks," in *Proc. European Control Conf. (ECC)*, 2018, (to appear).