Journal of Physics: Complexity

PAPER • OPEN ACCESS

Inferring directional interactions in collective dynamics: a critique to intrinsic mutual information

Journal of Physics: Complexity

OPEN ACCESS

RECEIVED
22 October 2022

REVISED
19 November 2022

ACCEPTED FOR PUBLICATION
19 December 2022

PUBLISHED
29 December 2022

PAPER

Inferring directional interactions in collective dynamics: a critique to intrinsic mutual information

Pietro De Lellis[1], Manuel Ruiz Marín[2] and Maurizio Porfiri[3,*]

[1] Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples 80125, Italy
[2] Department of Quantitative Methods, Law and Modern Languages, Technical University of Cartagena, Cartagena, Murcia 30201, Spain
[3] Center for Urban Science and Progress, Department of Mechanical and Aerospace Engineering, and Department of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201,United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: mporfiri@nyu.edu

Keywords: information flow, statistical inference, intrinsic mutual information, transfer entropy, time-series analysis, collective behavior, leader–follower interaction

Supplementary material for this article is available online

Abstract

Pairwise interactions are critical to collective dynamics of natural and technological systems. Information theory is the gold standard to study these interactions, but recent work has identified pitfalls in the way information flow is appraised through classical metrics—time-delayed mutual information and transfer entropy. These pitfalls have prompted the introduction of intrinsic mutual information to precisely measure information flow. However, little is known regarding the potential use of intrinsic mutual information in the inference of directional influences to diagnose interactions from time-series of individual units. We explore this possibility within a minimalistic, mathematically tractable leader–follower model, for which we document an excess of false inferences of intrinsic mutual information compared to transfer entropy. This unexpected finding is linked to a fundamental limitation of intrinsic mutual information, which suffers from the same sins of time-delayed mutual information: a thin tail of the null distribution that favors the rejection of the null-hypothesis of independence.

1. Introduction

Information theory [1] has emerged as a powerful framework to study causal relationships underpinning the collective dynamics of complex systems. Without the need of a mathematical model to be identified or experimental manipulations to be conducted, information theory allows for deciphering the strength and direction of interactions between coupled units from mere experimental observations of their dynamics. For example, through the lens of information theory, researchers have clarified the differences between anatomical and functional networks in the brain [2, 3], quantified the role of media and policy on human decision-making [4, 5], identified physical pathways underlying climate change across the globe [6, 7], and detected leaders in groups of animals [8–10].

Most of these efforts rely on the notion of transfer entropy, formulated by Schreiber two decades ago to study pairwise, asymmetric interactions between coupled dynamical systems [11]. In its classical incarnation, transfer entropy measures the extent to which knowledge about the present state of a dynamical system (source) helps reduce the uncertainty in the prediction of the future of another dynamical system from its present (target). Transfer entropy can be readily calculated from raw time-series [12], as its computation only requires the determination of the joint probability mass function for the present and future of the target and the present of the source. Likewise, hypothesis-testing with transfer entropy is easy to perform [13, 14]; for example, permutation tests can be implemented to assess whether transfer entropy is different from zero with a given confidence level, so that the null-hypothesis of independence of the target from the source can be rejected.

Over the years, the seminal work of Schreiber has been extended along several threads that have made transfer entropy ubiquitous among theorists and practitioners. For example, Sun and Bollt have successfully addressed multivariate interactions in network systems, building on the notion of conditional transfer entropy [15]. Runge *et al* introduced the notion of momentary information transfer, which excludes misleading influence of autodependency to better detail the coupling strength between the units [16]. Likewise, Staniek has tapped into symbolic dynamics to improve the robustness of transfer entropy-based inference, especially when dealing with short time-series [17]. Despite this growing body of sound methodological efforts and successful applications to real datasets, there are still open questions about the theoretical interpretation and practical use of transfer entropy.

The work of James *et al* has brought an important critique to transfer entropy, by offering concrete examples of systems with exclusive OR interactions that defeat one's intuition [18]. Specifically, the authors point at potential 'interpretational errors, some quite subtle … including overestimating flow, underestimating influence, and more generally misidentifying structure when modeling complex systems as networks with edges given by transfer entropies.' At the core of the critique is the impossibility to mechanistically associate transfer entropy between two dynamical systems with the information flow or transfer between them.

Building on this key limitation, James *et al* [19] have recently detailed information flow within a pair of dynamical systems, distinguishing multiple, co-existing modalities of information flow that have been erroneously compounded in the literature. Among them, 'intrinsic information flow' pertains to the predictive power that the present of the source alone has on the target's future, independent of the target's present: this quantity is what is routinely referred to as information flow (but seldom precisely measured). To quantify intrinsic information flow, the authors propose a cryptographic flow ansatz, which hypothesizes intrinsic flow to be equivalent to the secret key agreement between the two systems [20]. Such an ansatz is practically determined using an easy-to-compute upper bound, called intrinsic mutual information. The use of intrinsic mutual information in the study of leader–follower interactions has been explored by Sattari *et al* [21]. Through computer simulations of pairs and groups of self-propelled Vicsek-like particles [22], the authors have offered an important insight into information flow in collective dynamics, without the confounding effects that are brought about by classical information-theoretic metrics, such as transfer entropy.

While intrinsic mutual information constitutes a breakthrough in the quantification of information flow, its use as a tool for the inference of directional interactions has never been explored. Will the accurate quantification of information flow offered by intrinsic mutual information translate into an improved ability to detect directional influence? In this paper, we seek to provide an answer to this question through an integrated numerical and theoretical effort on the relationship between intrinsic mutual information and classical information-theoretic metrics (time-delayed mutual information and transfer entropy) to support hypothesis-testing in the inference of directional interactions.

We use a Boolean model of leader–follower interactions to compute exact, asymptotic expressions for information-theoretic metrics, which mitigate numerical artifacts related to the estimation of probability density functions from time-series. Similar to prior work on minimalistic models of collective dynamics [23–25], the model comprises a pair of directionally coupled Boolean units (a leader and a follower), subject to different intrinsic noises. The leader changes its state due to added noise, irrespective of the follower, while the follower responds to both the added noise and the leader. Upon gaining insight into the Boolean model, we examine simulation results from the modified Vicsek model by Sattari *et al* [21] to probe the generality of our claims and understand how the performance of the information-theoretic metrics vary with the size of the probability space where the inference is performed.

2. Results

2.1. Background on information-theoretic metrics for causal analysis

The most basic information-theoretic tool to study causal relationships between two dynamical systems is based on mutual information [1] (see section 4 for further details). Specifically, given two stationary, discrete stochastic processes $\{Y_t\}_{t\in\mathbb{Z}_{\geqslant 0}}$ and $\{Z_t\}_{t\in\mathbb{Z}_{\geqslant 0}}$, their (one-step) time-delayed mutual information is

$$\mathrm{MI}^{Z\to Y} = I(Z_t;Y_{t+1}) = \sum_{\substack{y_{t+1}\in\mathcal{Y}\\ z_t\in\mathcal{Z}}} \Pr(Y_{t+1}=y_{t+1}, Z_t=z_t)\log_2\frac{\Pr(Y_{t+1}=y_{t+1}|Z_t=z_t)}{\Pr(Y_{t+1}=y_{t+1})}. \tag{1}$$

Here, 'Pr' indicates the probability of an event; capital, lower case, and calligraphic letters are used for random variables, realizations, and sample spaces, respectively; commas are used for conjugation (logical AND); vertical bars are for conditioning of random variables; and semicolons are utilized to separate random variables when computing their mutual information I. To simplify the notation and avoid the excessive use of parentheses, we adopt the following operator precedence (high to low): 'comma,' 'semicolon,' and 'vertical bar.'

Mutual information of the pair (Z_t, Y_{t+1}) amounts to the reduction of uncertainty in the future state of Y (that is, Y_{t+1}) given the knowledge of the present state of Z (that is, Z_t). Being symmetric by construction, $\mathrm{MI}^{Z\to Y}$ will also correspond to the reduction of uncertainty in Z_t given the knowledge of Y_{t+1}. Importantly, a nonzero value of time-delayed mutual information can be registered even if the future state of Y is not directly influenced by the present state of Z, but their dynamics contain memory of their past states [21]. This drawback is resolved by transfer entropy, defined as the mutual information between Z_t and Y_{t+1}, conditional on Y_t, namely,

$$\mathrm{TE}^{Z\to Y} = I(Z_t; Y_{t+1}|Y_t) = \sum_{\substack{y_t, y_{t+1}\in\mathcal{Y}\\ z_t\in\mathcal{Z}}} \Pr(Y_{t+1}=y_{t+1}, Z_t=z_t, Y_t=y_t)\log_2\frac{\Pr(Y_{t+1}=y_{t+1}|Z_t=z_t, Y_t=y_t)}{\Pr(Y_{t+1}=y_{t+1}|Y_t=y_t)}. \quad (2)$$

Rephrasing James *et al* [19], transfer entropy is sensitive to both intrinsic dependencies between Z_t and Y_{t+1}, as well as the dependencies induced by Y_t. To filter the latter dependencies and precisely measure information flow, Sattari *et al* [21] proposed the use of intrinsic mutual information from Z to Y, defined as

$$\begin{aligned}\mathrm{IMI}^{Z\to Y} = \inf\Bigg\{ &\sum_{\substack{\bar{y}_t, y_{t+1}\in\mathcal{Y}\\ z_t\in\mathcal{Z}}} \Pr(Y_{t+1}=y_{t+1}, \overline{Y}_t=\bar{y}_t, Z_t=z_t)\\ &\times\log_2\frac{\Pr(Y_{t+1}=y_{t+1}|\overline{Y}_t=\bar{y}_t, Z_t=z_t)}{\Pr(Y_{t+1}=y_{t+1}|\overline{Y}_t=\bar{y}_t)} : \Pr(Y_{t+1}, Z_t, \overline{Y}_t)\\ &= \sum_{y_t\in\mathcal{Y}} \Pr(Y_{t+1}, Z_t, Y_t=y_t)\Pr(\overline{Y}_t|Y_t=y_t)\Bigg\}. \end{aligned} \quad (3)$$

Here, $\overline{Y}_t$ is an auxiliary variable taking values in $\mathcal{Y}$ and related to Y_t by means of the conditional probability $\Pr(\overline{Y}_t|Y_t)$—taking the form of an unknown (finite or infinite) $|\mathcal{Y}|\times|\mathcal{Y}|$ matrix. Computing the infimum over all possible conditional probabilities $\Pr(\overline{Y}_t|Y_t)$, intrinsic mutual information avoids including influence coming from the present state of both Z and Y when predicting the future state of Y.

Intrinsic mutual information has its theoretical roots in cryptography, whereby it can be viewed as an upper bound for the information shared by Z_t and Y_{t+1} that cannot be reconstructed or derived by Y_t. The definition of intrinsic mutual information begets the following, intuitive, inequalities:

$$\begin{aligned} 0 \leqslant \mathrm{IMI}^{Z\to Y} &\leqslant I(Z_t; Y_{t+1})\\ \mathrm{IMI}^{Z\to Y} &\leqslant I(Z_t; Y_{t+1}|Y_t), \end{aligned} \quad (4)$$

thereby implying that

$$\mathrm{IMI}^{Z\to Y} \leqslant \min\left\{\mathrm{TE}^{Z\to Y}, \mathrm{MI}^{Z\to Y}\right\}. \quad (5)$$

There is not an obvious relationship between time-delayed mutual information and transfer entropy: any of them can be larger than the other, since conditioning is not a subtractive operation. Intrinsic mutual information would reduce to time-delayed mutual information if the minimization process yielded a constant $\overline{Y}_t$, whereas it would be equivalent to transfer entropy in the case of $\overline{Y}_t = Y_t$ [21].

Albeit intrinsic mutual information has been shown to be more accurate in measuring information flow compared to transfer entropy [19, 21], this does not necessarily imply that it is a better instrument for inferring causal relationships. Indeed, a key step in the application of information-theoretic constructs to causal inference is hypothesis-testing, which requires contrasting observed values against data obtained under the null hypothesis of independence. In what follows, we clarify the relationship between intrinsic mutual information and the classical metrics of information flow on a minimalistic model of coupled Boolean units. For this model, all the information-theoretic quantities can be exactly computed, thereby enabling a comparison between time-delayed mutual information, transfer entropy, and intrinsic mutual information in terms of their ability to detect leader–follower interactions.

2.2. Boolean leader–follower model

Let us consider two Boolean random processes X_t^{L} and X_t^{F}, describing the state of the leader and follower, respectively. Their dynamics is given by

$$
\begin{aligned}
X_{t+1}^{\mathrm{L}} &= \begin{cases} X_t^{\mathrm{L}}, & \text{with probability } (1-\eta_{\mathrm{L}}), \\ 1-X_t^{\mathrm{L}}, & \text{with probability } \eta_{\mathrm{L}}, \end{cases} \\
X_{t+1}^{\mathrm{F}} &= \begin{cases} X_t^{\mathrm{F}}, & \text{with probability } (1-\eta_{\mathrm{F}})(1-w)+|1-X_t^{\mathrm{L}}-X_t^{\mathrm{F}}|w, \\ 1-X_t^{\mathrm{F}}, & \text{with probability } \eta_{\mathrm{F}}(1-w)+|X_t^{\mathrm{L}}-X_t^{\mathrm{F}}|w, \end{cases}
\end{aligned}
\tag{6}
$$

where $0<\eta_{\mathrm{L}}<1$, $0<\eta_{\mathrm{F}}<1$, and $0\leqslant w\leqslant 1$. Similar to the coupling of Vicsek-like models [21, 22], the gain w identifies the tendency of the follower to replicate the behavior of the leader at the previous time-step, with $w=1$ corresponding to the deterministic dynamics $X_{t+1}^{\mathrm{F}}=X_t^{\mathrm{L}}$ and $w=0$ to X_{t+1}^{F} being independent of X_t^{L}. Likewise, the parameters η_{L} and η_{F} capture the strength of the added noise in Vicsek-like models. Parameter η_{L} is the probability that the leader changes state in one time-step, whereas η_{F} is the probability that the follower changes state in the absence of coupling, that is, when $w=0$.

The schematic in figure 1(A) shows how model (6) can be adapted to mimic the four interaction types considered in the paper by Sattari *et al* [21] that employed a modified Vicsek model, by means of a suitable selection of the noise parameters η_{L} and η_{F}. Indeed, the leader (follower) will have a natural tendency to change state or remain in the same state depending on η_{L} (η_{F}) being greater or smaller than $1/2$, respectively. This memory can be visualized as two self-loops of weight $|1/2-\eta_{\mathrm{L}}|$ and $|1/2-\eta_{\mathrm{F}}|$ for the leader and follower, respectively. Different from the leader, the follower dynamics is not only controlled by the intrinsic noise parameter. The follower changes its state also in response to the leader in the form of a tendency to copy the previous state of the leader that is modulated by w. Particular instances of the model, where the agents have no memory of their past state (η_{L} and/or η_{F} equal to 1/2), can be then represented with the absence of one or both self-loops.

Next, we formulate the system dynamics in terms of an ergodic, four-state Markov chain, for which we compute the stationary distribution in closed-form.

2.2.1. Transition matrix

The states of the leader and follower at time $t+1$ only depend on their state at time t, and therefore we can describe the time evolution of system (6) as a first-order homogeneous Markov chain with four states, defined as $1\equiv(X_t^{\mathrm{L}}=0,X_t^{\mathrm{F}}=0)$, $2\equiv(X_t^{\mathrm{L}}=0,X_t^{\mathrm{F}}=1)$, $3\equiv(X_t^{\mathrm{L}}=1,X_t^{\mathrm{F}}=0)$, and $4\equiv(X_t^{\mathrm{L}}=1,X_t^{\mathrm{F}}=1)$. We denote with $P\in\mathbb{R}^{4\times4}$ the transition probability matrix of the Markov chain, where its element ij is the probability that the chain takes the jth value at the next time-step given that the current value is the ith one.

For brevity, we detail how entry 11 of P is computed; other entries are analogously obtained. By definition,

$$
\begin{aligned}
P_{11} &= \Pr(X_{t+1}^{\mathrm{L}}=0,X_{t+1}^{\mathrm{F}}=0|X_t^{\mathrm{L}}=0,X_t^{\mathrm{F}}=0)=\Pr(X_{t+1}^{\mathrm{L}}=0|X_t^{\mathrm{L}}=0)\Pr(X_{t+1}^{\mathrm{F}}=0|X_t^{\mathrm{L}}=0,X_t^{\mathrm{F}}=0) \\
&= (1-\eta_{\mathrm{L}})\big((1-\eta_{\mathrm{F}})(1-w)+w\big),
\end{aligned}
\tag{7}
$$

where we have used the property that the next state of the leader is independent of the current state of the follower, and the property that the next states of the leader and follower are independent upon conditioning on their current states. Ultimately, we establish

$$
P=\begin{bmatrix}
(1-\eta_{\mathrm{L}})g_{1-\eta_{\mathrm{F}}} & (1-\eta_{\mathrm{L}})f_{\eta_{\mathrm{F}}} & \eta_{\mathrm{L}}g_{1-\eta_{\mathrm{F}}} & \eta_{\mathrm{L}}f_{\eta_{\mathrm{F}}} \\
(1-\eta_{\mathrm{L}})g_{\eta_{\mathrm{F}}} & (1-\eta_{\mathrm{L}})f_{1-\eta_{\mathrm{F}}} & \eta_{\mathrm{L}}g_{\eta_{\mathrm{F}}} & \eta_{\mathrm{L}}f_{1-\eta_{\mathrm{F}}} \\
\eta_{\mathrm{L}}f_{1-\eta_{\mathrm{F}}} & \eta_{\mathrm{L}}g_{\eta_{\mathrm{F}}} & (1-\eta_{\mathrm{L}})f_{1-\eta_{\mathrm{F}}} & (1-\eta_{\mathrm{L}})g_{\eta_{\mathrm{F}}} \\
\eta_{\mathrm{L}}f_{\eta_{\mathrm{F}}} & \eta_{\mathrm{L}}g_{1-\eta_{\mathrm{F}}} & (1-\eta_{\mathrm{L}})f_{\eta_{\mathrm{F}}} & (1-\eta_{\mathrm{L}})g_{1-\eta_{\mathrm{F}}}
\end{bmatrix}
\tag{8}
$$

where we have introduced the notations $f_\eta=\eta(1-w)$ and $g_\eta=f_\eta+w$. Obviously, all the rows of P sum to one.

2.2.2. Stationary probability distribution

Since none of the elements of P is zero, all the states are aperiodic and positive recurrent, that is, the Markov chain is ergodic [26]. The unique stationary distribution,

$$
\pi_\infty^{\mathrm{LF}}(i,j)=\lim_{t\to+\infty}\Pr(X_t^{\mathrm{L}}=i,X_t^{\mathrm{F}}=j)
\tag{9}
$$

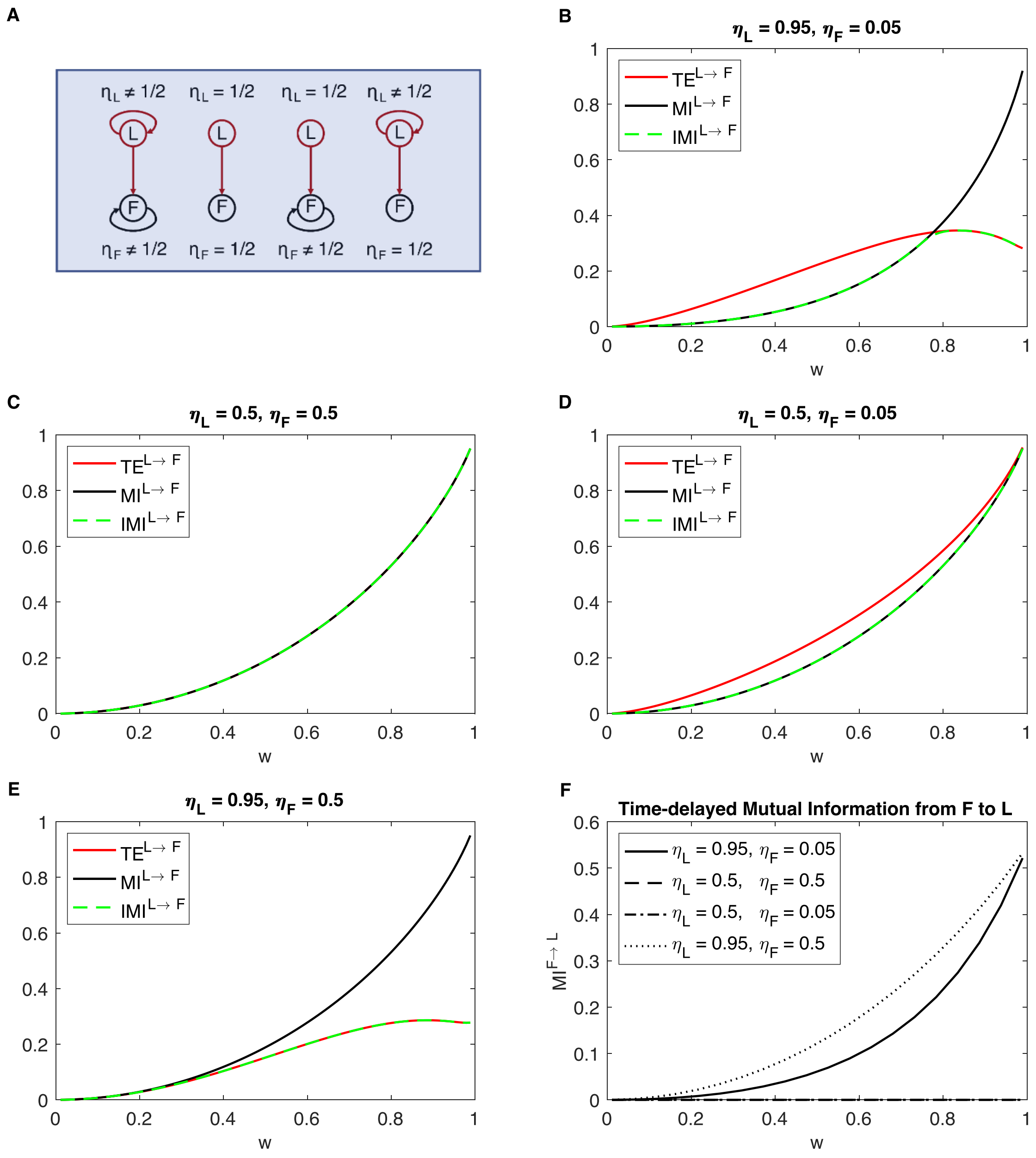

Figure 1. Information-theoretic analysis of model (6). Panel (A) shows a schematic of the model for different combination of the leader and follower parameters η_{L} and η_{F}. Panels (B)–(E) report time-delayed mutual information, transfer entropy, and intrinsic mutual information (black solid line, red solid line, and green dashed line, respectively) from leader to follower for model (6) as functions of the coupling gain w for four pairs of noise parameters η_{L} and η_{F}: B, $\eta_{\mathrm{L}} = 0.95$ and $\eta_{\mathrm{F}} = 0.05$; C, $\eta_{\mathrm{L}} = 0.5$ and $\eta_{\mathrm{F}} = 0.5$; D, $\eta_{\mathrm{L}} = 0.5$ and $\eta_{\mathrm{F}} = 0.05$; and E, $\eta_{\mathrm{L}} = 0.95$ and $\eta_{\mathrm{F}} = 0.5$. Panel (F) reports time-delayed mutual information from follower to leader for model (6) as a function of the coupling gain w for the same four pairs of noise parameters η_{L} and η_{F}: solid, $\eta_{\mathrm{L}} = 0.95$ and $\eta_{\mathrm{F}} = 0.05$; dashed, $\eta_{\mathrm{L}} = 0.5$ and $\eta_{\mathrm{F}} = 0.5$; dotted-dashed, $\eta_{\mathrm{L}} = 0.5$ and $\eta_{\mathrm{F}} = 0.05$; and dotted, $\eta_{\mathrm{L}} = 0.95$ and $\eta_{\mathrm{F}} = 0.5$. Note that in B the green curve is superimposed to the black one for low values of w and to the red one for large values of w; in C, all curves are indistinguishable; in D, the black and dashed green curves are indistinguishable; in E, the red and green curves are indistinguishable; and in F, dashed and dotted-dashed curves are identically zero. We remark that transfer entropy and intrinsic mutual information from follower to leader are identically zero.

for all $i, j \in \{0, 1\}$, can be computed as the left eigenvector with unitary eigenvalue of P [26], normalized such that its elements sum to 1. Therefore, we determine

$$\pi_{\infty}^{\mathrm{LF}}(0,0) = \pi_{\infty}^{\mathrm{LF}}(1,1) = \frac{a + w - 2\eta_{\mathrm{L}} w}{2(2a + w - 2\eta_{\mathrm{L}} w)}, \tag{10a}$$

$$\pi_{\infty}^{\mathrm{LF}}(0,1) = \pi_{\infty}^{\mathrm{LF}}(1,0) = \frac{a}{2(2a + w - 2\eta_{\mathrm{L}} w)}, \tag{10b}$$

where $a=\eta_\mathrm{F}+\eta_\mathrm{L}-2\eta_\mathrm{F}\eta_\mathrm{L}-\eta_\mathrm{F}w+2\eta_\mathrm{F}\eta_\mathrm{L}w$.

From equation set (10), it follows that the stationary probabilities for the leader and follower are all equal to $1/2$, whereby $\pi_\infty^{\mathrm{LF}}(i,0)+\pi_\infty^{\mathrm{LF}}(i,1)=\pi_\infty^{\mathrm{LF}}(0,i)+\pi_\infty^{\mathrm{LF}}(1,i)=1/2$ for all $i\in\{0,1\}$, similar to a Vicsek model for which none of the agents has a preferential heading direction.

2.3. Information-theoretic metrics

From the transition matrix (8) and its stationary distribution (10), we compute the stationary joint probability distributions $\lim_{t\to+\infty}\Pr(X_{t+1}^\mathrm{F},X_t^\mathrm{L},X_t^\mathrm{F})$ and $\lim_{t\to+\infty}\Pr(X_{t+1}^\mathrm{L},X_t^\mathrm{L},X_t^\mathrm{F})$. The calculation uses the definition of conditional probability so that, for example,

$$\Pr(X_{t+1}^\mathrm{F},X_t^\mathrm{L},X_t^\mathrm{F})=\Pr(X_{t+1}^\mathrm{F}|X_t^\mathrm{L},X_t^\mathrm{F})\Pr(X_t^\mathrm{L},X_t^\mathrm{F}). \tag{11}$$

Herein, the conditional probability on the right-hand-side of the equation is obtained from matrix (8), upon marginalizing with respect to the state of the leader at $t+1$, that is,

$$\begin{aligned}\Pr(X_{t+1}^\mathrm{F}=x_{t+1}^\mathrm{F}|X_t^\mathrm{L}=x_t^\mathrm{L},X_t^\mathrm{F}=x_t^\mathrm{F})=&\Pr(X_{t+1}^\mathrm{L}=0,X_{t+1}^\mathrm{F}=x_{t+1}^\mathrm{F}|X_t^\mathrm{L}=x_t^\mathrm{L},X_t^\mathrm{F}=x_t^\mathrm{F})\\&+\Pr(X_{t+1}^\mathrm{L}=1,X_{t+1}^\mathrm{F}=x_{t+1}^\mathrm{F}|X_t^\mathrm{L}=x_t^\mathrm{L},X_t^\mathrm{F}=x_t^\mathrm{F}).\end{aligned} \tag{12}$$

Complete expressions are reported in table 1 and utilized to compute closed-form, asymptotic expressions of classical information-theoretic quantities (time-delayed mutual information and transfer entropy) and of intrinsic mutual information as functions of the coupling gains between the unites and the strengths of the added noises.

2.3.1. Classical metrics

The computation of time-delayed mutual information from leader to follower ($\mathrm{MI}^{\mathrm{L}\to\mathrm{F}}$) and vice versa ($\mathrm{MI}^{\mathrm{F}\to\mathrm{L}}$) can be undertaken from (1), using the expressions in table 1. Therein, the conditional probabilities should be written using the definition of conditional probability as $\Pr(X_{t+1}^\mathrm{F}=x_{t+1}^\mathrm{F}|X_t^\mathrm{L}=x_t^\mathrm{L})=\Pr(X_{t+1}^\mathrm{F}=x_{t+1}^\mathrm{F},X_t^\mathrm{L}=x_t^\mathrm{L})/\Pr(X_t^\mathrm{L}=x_t^\mathrm{L})$ and similarly for the follower-to-leader interaction. Then, any of the probabilities appearing in the expression of time-delayed mutual information can be retrieved from table 1 through marginalization; for example,

$$\Pr(X_{t+1}^\mathrm{F}=x_{t+1}^\mathrm{F},X_t^\mathrm{L}=x_t^\mathrm{L})=\Pr(X_{t+1}^\mathrm{F}=x_{t+1}^\mathrm{F},X_t^\mathrm{L}=x_t^\mathrm{L},X_t^\mathrm{F}=0)+\Pr(X_{t+1}^\mathrm{F}=x_{t+1}^\mathrm{F},X_t^\mathrm{L}=x_t^\mathrm{L},X_t^\mathrm{F}=1). \tag{13}$$

Likewise, transfer entropy from leader to follower ($\mathrm{TE}^{\mathrm{L}\to\mathrm{F}}$) and vice versa ($\mathrm{TE}^{\mathrm{F}\to\mathrm{L}}$) can be calculated via (2), by replacing for the joint distributions in table 1.

In figures 1(B)–(E), we display time-delayed mutual information and transfer entropy from leader to follower as functions of the coupling gain w for four pairs of noise parameters η_L and η_F that exemplify interaction types from figure 1(A). In agreement with numerical results on the modified Vicsek model by Sattari *et al* [21], we observe the following. First, in the presence of a self-loop for the follower (figures 1(B) and (D)), time-delayed mutual information can be less than transfer entropy. This surprising finding is related to the onset of a synergistic information flow, whereby simultaneous knowledge about the present of the leader and follower improves the predictive power about the future of the follower, compared to mere access to the present of the follower. Second, in the absence of self-loops in both the leader and the follower (figure 1(C)), time-delayed mutual information and transfer entropy are equivalent, which is due to the lack of memory in the dynamics. Third, in the presence of a self-loop only for the leader, transfer entropy is less than time-delayed mutual information (figure 1(E)), in agreement with one's intuition about the role of transfer entropy in mitigating redundant information from the follower's own dynamics.

From the follower to the leader, transfer entropy is always zero, since the follower does not provide any predictive power about the future state of the leader once the present state of the leader is known. On the other hand, time-delayed mutual information can be different from zero due to the shared history of the follower and the leader. In figure 1(F), we report time-delayed mutual information from the follower to the leader for the same cases considered in figures 1(B)–(E). Predictably, without a self-loop in the leader, time-delayed mutual information is zero: the leader does not have a memory and, as such, no information is shared in a common history with the follower.

Finally, we comment that the role of the coupling gain is non-trivial. While time-delayed mutual information seems to increase with the coupling gain for different choices of the noise parameters, transfer entropy could decrease for sufficiently large values of w, as in figure 1(B). In such a case, the follower will tend to systematically replicate the behavior of the leader, whose dynamics is, however, evolving in response to its own history. As a result, the information flow from the leader to the follower could be hindered by larger values of w. Compact expressions for time-delayed mutual information and transfer entropy are in

Table 1. Stationary joint probability distribution of X_{t+1}^{F}, X_t^{L}, and X_t^{F} and of X_{t+1}^{L}, X_t^{L}, and X_t^{F} for the computation of closed-form, asymptotic expressions of information-theoretic metrics for model (6).

$(x_{t+1}^{\mathrm{F}}, x_t^{\mathrm{L}}, x_t^{\mathrm{F}})$	$\mathrm{Pr}(X_{t+1}^{\mathrm{F}} = x_{t+1}^{\mathrm{F}}, X_t^{\mathrm{L}} = x_t^{\mathrm{L}}, X_t^{\mathrm{F}} = x_t^{\mathrm{F}})$
$(0,0,0),(1,1,1)$	$g_{1-\eta_{\mathrm{F}}}\pi_\infty^{\mathrm{LF}}(0,0)$
$(0,0,1),(1,1,0)$	$g_{\eta_{\mathrm{F}}}\pi_\infty^{\mathrm{LF}}(0,1)$
$(0,1,0),(1,0,1)$	$f_{1-\eta_{\mathrm{F}}}\pi_\infty^{\mathrm{LF}}(0,1)$
$(0,1,1),(1,0,0)$	$f_{\eta_{\mathrm{F}}}\pi_\infty^{\mathrm{LF}}(0,0)$
	$\mathrm{Pr}(X_{t+1}^{\mathrm{L}} = x_{t+1}^{\mathrm{L}}, X_t^{\mathrm{L}} = x_t^{\mathrm{L}}, X_t^{\mathrm{F}} = x_t^{\mathrm{F}})$
$(0,0,0),(1,1,1)$	$(1-\eta_{\mathrm{L}})\pi_\infty^{\mathrm{LF}}(0,0)$
$(0,0,1),(1,1,0)$	$(1-\eta_{\mathrm{L}})\pi_\infty^{\mathrm{LF}}(0,1)$
$(0,1,0),(1,0,1)$	$\eta_{\mathrm{L}}\pi_\infty^{\mathrm{LF}}(0,1)$
$(0,1,1),(1,0,0)$	$\eta_{\mathrm{L}}\pi_\infty^{\mathrm{LF}}(0,0)$

general not feasible; second-order Taylor expansions in terms the noise parameters are presented in the supplementary note 1.

2.3.2. Intrinsic mutual information

Obviously, intrinsic mutual information from follower to leader is zero ($\mathrm{IMI}^{\mathrm{F}\to\mathrm{L}}$), since transfer entropy is zero and intrinsic mutual information cannot be larger than transfer entropy. The computation of intrinsic mutual information from leader to follower ($\mathrm{IMI}^{\mathrm{L}\to\mathrm{F}}$) requires an auxiliary stochastic process $\{\overline{X}_t^{\mathrm{F}}\}_{t\in\mathbb{Z}_{\geqslant 0}}$, which is univocally related to $\{X_t^{\mathrm{F}}\}_{t\in\mathbb{Z}_{\geqslant 0}}$ via

$$\begin{aligned}\mathrm{Pr}(\overline{X}_t^{\mathrm{F}} = 0 | X_t^{\mathrm{F}} = 0) \equiv \alpha,\\ \mathrm{Pr}(\overline{X}_t^{\mathrm{F}} = 0 | X_t^{\mathrm{F}} = 1) \equiv \beta,\end{aligned} \tag{14}$$

where $\alpha, \beta \in [0,1]$ are the parameters upon which conditional mutual information is optimized.

Next, we can easily compute the joint distribution $\mathrm{Pr}(X_{t+1}^{\mathrm{F}}, X_t^{\mathrm{L}}, \overline{X}_t^{\mathrm{F}})$ in terms of α, β, and values listed in table 1. For completeness, we report the resulting joint distribution in table 2. Following analogous steps to transfer entropy computation, but using table 2, we obtain $I(X_{t+1}^{\mathrm{F}}; X_t^{\mathrm{L}} | \overline{X}_t^{\mathrm{F}})$ as a function of α and β. By taking the minimum over α and β in the compact unit square, we calculate intrinsic mutual information.

Results in figures 1(B)–(E) indicate that intrinsic mutual information is very well approximated by the minimum between time-delayed mutual information and transfer entropy for any selection of the noise parameters. Only in figure 1(B), where time-delayed mutual information and transfer entropy cross for a coupling of $w = 0.776$, we observe a narrow window of the coupling gain in which intrinsic mutual information is lower than both the classical metrics (for $w \in [0.770, 0.780]$, intrinsic mutual information is within 0.008, or 2.4%, from the minimum of the other two classical metrics). As such, intrinsic mutual information equals time-delayed mutual information in the presence of a self-loop for the follower and absence of a self-loop for the leader or in the presence of both self-loops, provided the coupling gain is sufficiently weak. It is equal to transfer entropy in the absence of a self-loop for the follower or in the presence of both self-loops, provided the coupling gain is sufficiently strong. When both self-loops are absent, intrinsic mutual information is equal to time-delayed mutual information and transfer entropy. We stress that these conclusions are not affected by numerical artifacts in the estimation of the probability mass functions, whereby they rely on closed-form, asymptotic expressions of all the information-theoretic metrics.

2.4. Statistical inference

Here, we explore the feasibility of employing intrinsic mutual information for the inference of the directional coupling between the units and contrast its performance with time-delayed mutual information and transfer entropy. We utilize the time-series of the two units (leader and follower) to estimate all the joint probability distributions in the information-theoretic metrics (1)–(3). Without *a priori* knowledge of which is the leader and which is the follower, we calculate the information-theoretic metrics between the two units. These numerical values are contrasted with their corresponding null distribution in the absence of any interaction between the units (that is, $w = 0$), to decide whether a directional coupling exists or not, at a given confidence level. The null distributions are estimated by simulating model (6) for N repetitions each of length T. Should one not have access to the ground true mathematical model of the time-series, as in most of the practical applications, they could generate their null distributions through shuffling. We illustrate this possibility in the supplementary note 2.

Table 2. Stationary joint probability distribution of X^{F}_{t+1}, X^{L}_t, and $\overline{X}^{\mathrm{F}}_t$ for the computation of closed-form, asymptotic expressions of intrinsic mutual information for model (6).

$(x^{\mathrm{F}}_{t+1}, x^{\mathrm{L}}_t, \overline{x}^{\mathrm{F}}_t)$	$\mathrm{Pr}(X^{\mathrm{F}}_{t+1} = x^{\mathrm{F}}_{t+1}, X^{\mathrm{L}}_t = x^{\mathrm{L}}_t, \overline{X}^{\mathrm{F}}_t = \overline{x}^{\mathrm{F}}_t)$
$(0,0,0)$	$\alpha g_{1-\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,0) + \beta g_{\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,1)$
$(0,0,1)$	$(1-\alpha) g_{1-\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,0) + (1-\beta) g_{\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,1)$
$(0,1,0)$	$\alpha f_{1-\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,1) + \beta f_{\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,0)$
$(0,1,1)$	$(1-\alpha) f_{1-\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,1) + (1-\beta) f_{\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,0)$
$(1,0,0)$	$\alpha f_{\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,0) + \beta f_{1-\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,1)$
$(1,0,1)$	$(1-\alpha) f_{\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,0) + (1-\beta) f_{1-\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,1)$
$(1,1,0)$	$\alpha g_{\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,1) + \beta g_{1-\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,0)$
$(1,1,1)$	$(1-\alpha) g_{\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,1) + (1-\beta) g_{1-\eta_{\mathrm{F}}} \pi^{\mathrm{LF}}_{\infty}(0,0)$

Based on the theoretical predictions for time-delayed mutual information, transfer entropy, and intrinsic mutual information in figures 1(B)–(F), we focus the inference effort on the case considered in figures 1(B) and (F) ($\eta_{\mathrm{L}} = 0.95$ and $\eta_{\mathrm{F}} = 0.05$), which displays the richest dependence of intrinsic mutual information on the coupling gain. We consider three different values of w (0.1, 0.5, and 1); for each value, we run model (6) for $T = 2000$, compute information-theoretic metrics, and contrast their values with the null distributions—we denote as F_{IMI}, F_{TE}, and F_{MI} the cumulative null distributions of intrinsic mutual information, transfer entropy, and mutual information, respectively. We reject the hypothesis of $w = 0$ with a significance level of 0.05, which corresponds to cut-off values for time-delayed mutual information, transfer entropy, and intrinsic mutual information of $\mathrm{MI}_{95} = 1.58 \times 10^{-4}$, $\mathrm{TE}_{95} = 2.31 \times 10^{-3}$, and $\mathrm{IMI}_{95} = 1.55 \times 10^{-4}$, respectively, see figure 2(A). For each value of w, we perform $N = 1000$ simulations and we evaluate the false negatives (number of simulations in which we fail to reject the null hypothesis of absence of interaction of the leader on the follower) and false positives (number of simulations in which we reject the null hypothesis of absence of interaction of the follower on the leader). A more comprehensive analysis for different values of w from 0.1 to 1 in steps of 0.1 is reported in the supplementary note 3.

Simulation results indicate sensitivity—defined as the true positive rate—at the perfection level of all the information-theoretic metrics with respect to the inference of the directional interaction from the leader to the follower (false negative rate of zero for all values of w). Specificity—defined as the true negative rate—is more problematic and highly different among the information-theoretic metrics, as illustrated in figure 2(B). Independent of the value of w, transfer entropy yields the best inferences with a false positive rate of about 5%, a much better performance compared to intrinsic mutual information, which begets a rate of about 48%. For all values of w, time-delayed mutual information offers unacceptable results, where it erroneously misclassifies the entirety of the observations (similar results are found for different values of w, see supplementary note 2).

2.4.1. Explaining the excess of false positives

The inadequacy of time-delayed mutual information in identifying the directionality of the interaction between leader and the follower should have been anticipated, given that the dynamics of both units contains information about their past for the selected leader–follower configuration in figure 1(A). The asymptotic time-delayed mutual information is different from zero in both directions, thereby challenging statistical inference of a directional interaction. The higher false positive rate of intrinsic mutual information compared to transfer entropy is somewhat surprising, given that intrinsic mutual information was originally intended to better quantify information flow than the classical information-theoretic metrics.

The explanation for this counter-intuitive result largely has its roots in the fact that, for most parameter combinations, intrinsic mutual information corresponds to the minimum between transfer entropy and mutual information, as shown in figures 1(B)–(F). Put simply, intrinsic mutual information suffers from the sins of time-delayed mutual information. Under the assumption that intrinsic mutual information is the minimum of time-delayed mutual information and transfer entropy, we have

$$F_{\mathrm{IMI}}(x) = F_{\mathrm{TE}}(x) + \mathrm{Pr}(\mathrm{TE} > x, \mathrm{MI} \leqslant x), \tag{15}$$

where x is the generic value of intrinsic mutual information and we considered that the event $\{\min(\mathrm{TE}, \mathrm{MI}) \leqslant x\}$ is the union of the disjoint events $\{\mathrm{TE} \leqslant x\}$ and $\{\mathrm{TE} > x\} \cap \{\mathrm{MI} \leqslant x\}$. Likewise, by accounting for $\{\min(\mathrm{TE}, \mathrm{MI}) \leqslant x\}$ to be the union of the disjoint events $\{\mathrm{MI} \leqslant x\}$ and $\{\mathrm{MI} > x\} \cap \{\mathrm{TE} \leqslant x\}$, we establish

$$F_{\mathrm{IMI}}(x) = F_{\mathrm{MI}}(x) + \mathrm{Pr}(\mathrm{TE} \leqslant x, \mathrm{MI} > x). \tag{16}$$

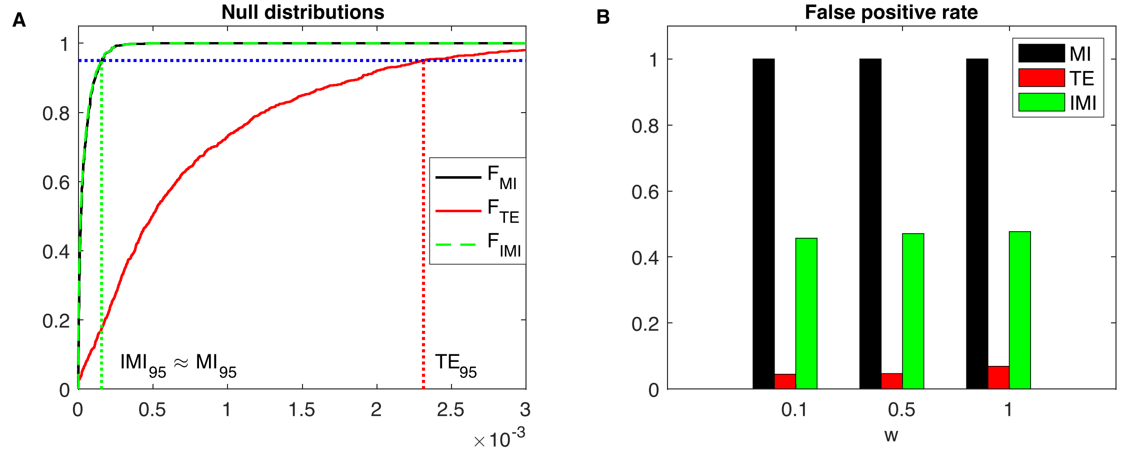

Figure 2. Statistical inference through information-theoretic metrics on time-series from model (6). Panel (A) depicts the null cumulative distributions of time-delayed mutual information (black solid line), transfer entropy (red solid line), and intrinsic mutual information (green dashed line) obtained by simulating $N = 1000$ iterations, each $T = 2000$ times-steps long, of the Boolean leader–follower model in (6) for $\eta_{\mathrm{L}} = 0.95$, $\eta_{\mathrm{F}} = 0.05$, and $w = 0$. The green and black lines are practically superimposed and visually indistinguishable. MI_{95}, TE_{95}, and IMI_{95} are the cut-off values for time-delayed mutual information, transfer entropy, and intrinsic mutual information, respectively, computed as the 95th percentile of their null cumulative distributions. Panel (B) shows the rate of false positives (detection of a link from follower to leader) as a function of the coupling gain w. Each value is obtained by classifying interactions in $N = 1000$ time-series of a leader–follower pair with $\eta_{\mathrm{L}} = 0.95$ and $\eta_{\mathrm{F}} = 0.05$, with each time-series being $T = 2000$ time-steps long. Black, red, and green bars correspond to time-delayed mutual information, transfer entropy, and intrinsic mutual information, respectively.

These two equalities together imply that $F_{\mathrm{IMI}}(x) \geqslant \max\{F_{\mathrm{TE}}(x), F_{\mathrm{MI}}(x)\}$. Hence, the cut-off value for intrinsic mutual information cannot be larger than those for mutual information and transfer entropy, that is, $\mathrm{IMI}_{95} \leqslant \min\{\mathrm{MI}_{95}, \mathrm{TE}_{95}\}$ as illustrated in figure 2(A). Noting that $F_{\mathrm{MI}}(x) > F_{\mathrm{TE}}(x)$ for all values of x (so that $\mathrm{IMI}_{95} < \mathrm{TE}_{95}$), we identify the following three modalities by which intrinsic mutual information would yield different inferences than those of transfer entropy:

(1) $\mathrm{IMI}^{\mathrm{F}\to\mathrm{L}} = \mathrm{TE}^{\mathrm{F}\to\mathrm{L}}$ and $\mathrm{IMI}_{95} < \mathrm{IMI}^{\mathrm{F}\to\mathrm{L}} \leqslant \mathrm{TE}_{95}$. In this case, intrinsic mutual information would reject the null hypothesis and it would yield a false positive, whereas transfer entropy would correctly infer the absence of a causal link from the follower to the leader, as illustrated in figure 3(A);

(2) $\mathrm{IMI}^{\mathrm{F}\to\mathrm{L}} = \mathrm{MI}^{\mathrm{F}\to\mathrm{L}}$, $\mathrm{IMI}^{\mathrm{F}\to\mathrm{L}} > \mathrm{IMI}_{95}$, and $\mathrm{TE}^{\mathrm{F}\to\mathrm{L}} \leqslant \mathrm{TE}_{95}$. Also in this case, intrinsic mutual information would yield a false positive, whereas transfer entropy would correctly identify a negative, see figure 3(B); and

(3) $\mathrm{IMI}^{\mathrm{F}\to\mathrm{L}} = \mathrm{MI}^{\mathrm{F}\to\mathrm{L}}$, $\mathrm{IMI}^{\mathrm{F}\to\mathrm{L}} \leqslant \mathrm{IMI}_{95}$, and $\mathrm{TE}^{\mathrm{F}\to\mathrm{L}} > \mathrm{TE}_{95}$. Different from Cases 1 and 2, intrinsic mutual information would correctly infer the absence of a causal link from the follower to the leader, as illustrated in figure 3(C).

Case 3 is then the only case in which intrinsic mutual information could outperform transfer entropy in filtering a spurious interaction. The possibility of this case to occur is related to time-delayed mutual information being able to filter the spurious link, which is never registered for any parameter combination. Cases 1 and 2 are prevalent in our study, due to the much fatter tail of the null distribution of transfer entropy compared to intrinsic mutual information, thereby explaining the excess of false positives when using intrinsic mutual information rather than transfer entropy (48% against 5%).

2.4.2. Extension to the modified Vicsek model

The proposed minimalistic Boolean model offers insight into the root causes for the superior performance of transfer entropy compared to intrinsic mutual information as a tool to infer leader–follower directional interactions. To support the generality of our findings, we now consider a leader–follower pair in the modified Vicsek model [27] as in Sattari *et al* [21]. Here, a leader particle L and a follower particle F move in a square domain of size $l \times l$ with periodic boundary conditions, and their planar positions at time $t \in \mathbb{Z}_{\geqslant 0}$ are described by the complex numbers r_t^{L} and r_t^{F}, respectively.

The two particles move at a constant speed, and the heading of the leader θ_t^{L} at time t influences the heading of the follower $\theta_{t+1}^{\mathrm{F}}$ at the next time-step when their distance is within a unitary interaction distance, according to the following equation:

$$\theta_{t+1}^{\mathrm{L}} = \arg\left(\exp(\imath\theta_t^{\mathrm{L}} w_{\mathrm{LL}})\right) + \psi_t^{\mathrm{L}}, \tag{17a}$$

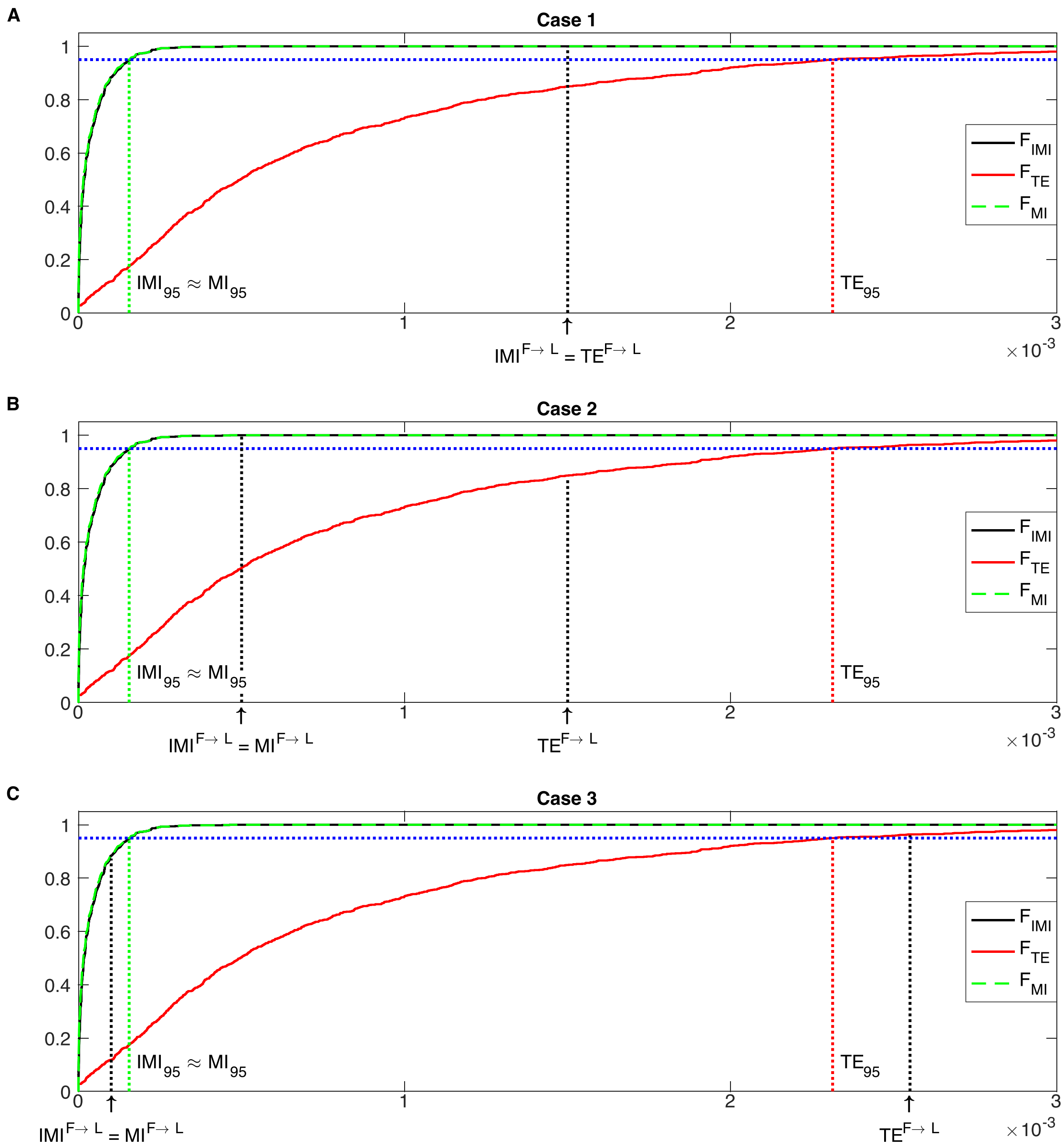

Figure 3. Explanation for the excess of false positives using intrinsic mutual information for the inference of directional interactions from time-series from model (6). In all panels, we present the null cumulative distributions for time-delayed mutual information (black solid line), transfer entropy (red solid line), and intrinsic mutual information (green dashed line) from 1000 simulations, each $T = 2000$ times-steps long, of the Boolean leader–follower model (6) for $\eta_{\mathrm{L}} = 0.95$, $\eta_{\mathrm{F}} = 0.05$, and $w = 0$. The green and black lines are practically superimposed and visually indistinguishable. MI_{95}, TE_{95}, and IMI_{95} are the cut-off values for time-delayed mutual information, transfer entropy and intrinsic mutual information, respectively, computed as the 95th percentile of their null cumulative distribution functions. Panels (A)–(C) correspond to Cases 1, 2, and 3, respectively, which identify the modalities by which intrinsic mutual information and transfer entropy yield different inferences for the possible link from follower to leader.

$$\theta^{\mathrm{F}}_{t+1} = \arg\left(\exp(\imath\theta^{\mathrm{F}}_t w_{\mathrm{FF}}) + w_{\mathrm{LF}}\mathcal{I}_t \exp(\imath\theta^{\mathrm{L}}_t)\right) + \psi^{\mathrm{F}}_t, \tag{17b}$$

$$r^i_{t+1} = r^i_t + s\exp(\imath\theta^i_t), \quad i \in \{\mathrm{L},\mathrm{F}\}. \tag{17c}$$

Here, $\mathcal{I}_t$ is an indicator function that is 1 if $|r^{\mathrm{L}}_t - r^{\mathrm{F}}_t| \leqslant 1$, and 0 otherwise; $\imath$ is the imaginary unit; ψ^{L}_t and ψ^{F}_t are the additive noises affecting the heading of the particles, chosen to be independent uniformly distributed random variables in $[-\eta/2, \eta/2]$, with $\eta > 0$; $s > 0$ is the common speed of all the particles, which corresponds to the number of interaction radii that are travelled in one time-step; $w_{\mathrm{LL}} \in \{-1, 1\}$ and $w_{\mathrm{FF}} \in \{-1, 1\}$ define the dynamics of the leader and follower, respectively (for example, in the absence of noise, if $w_{\mathrm{LL}} = 1$, the leader would not change its heading, whereas if $w_{\mathrm{LL}} = -1$ if it would flip its heading at every time-steps); and $w_{\mathrm{LF}} \geqslant 0$ weighs the influence the leader's current heading has on the next heading of the follower.

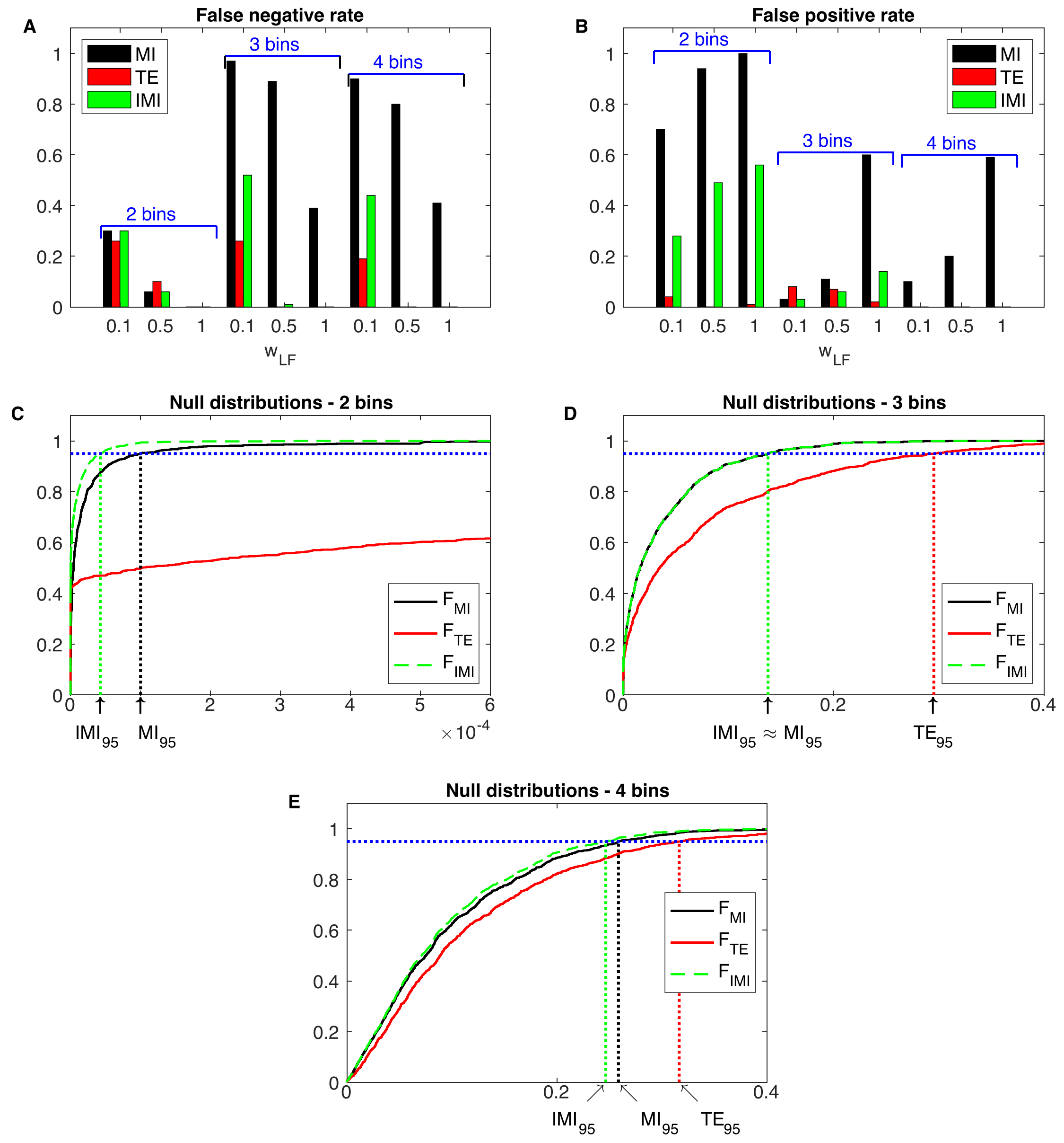

Figure 4. Inferring directional interactions in the modified Vicsek model (17) for $l = 2$, $s = 0.3$, $\eta = 0.1$, $w_{\mathrm{LL}} = -1$, and $w_{\mathrm{FF}} = 1$. Rate of false negatives (link from leader to follower undetected, panel (A) and false positives (detection of a link from follower to leader, panel (B) as functions of the coupling gain w_{LF} from the leader to the follower. Each point is obtained by classifying interactions in $N = 100$ time-series of a leader–follower pair, with each time-series being $T = 2000$ time-steps long. The time-series of the leader and follower are discretized with either two, three, or four equally-spaced bins to estimate joint probability distributions in the information-theoretic metrics (1)–(3). Black, red, and green bars correspond to time-delayed mutual information, transfer entropy, and intrinsic mutual information, respectively. Panels (C)–(E) depict the null cumulative distributions of time-delayed mutual information (black solid line), transfer entropy (red solid line), and intrinsic mutual information (green dashed line) obtained from 1000 simulations, each $T = 2000$ time-steps long, of the leader follower Vicsek model (17) for $w_{\mathrm{LF}} = 0$, and discretizing the time-series of the leader and follower into either two, three or four bins. MI_{95}, TE_{95}, and IMI_{95} are the cut-off values for time-delayed mutual information, transfer entropy, and intrinsic mutual information, respectively, computed as the 95th percentile of their null cumulative distributions. In panel (D), the green and black lines are practically superimposed and visually indistinguishable.

We report simulation results for $l = 2$, $s = 0.3$, $\eta = 0.1$, $w_{\mathrm{LL}} = -1$, and $w_{\mathrm{FF}} = 1$. In this case, the leader would tend to flip its heading at every time-step, while the follower would tend to maintain its heading, thereby mimicking the case of the Boolean model in figure 1(B). By varying the coupling w_{LF}, we modulate the influence of the heading on the follower. For $w_{\mathrm{LF}} = 0.1$, the follower is only marginally affected by the heading of the leader in its update process. For $w_{\mathrm{LF}} = 1$, the follower equally weights its heading and the heading of the leader in its update process.

For each value of w_{LF}, we perform $N = 100$ repetitions for initial headings randomly selected in $[-\pi, \pi]$, and initial position in the unit radius at the center of the square domain. For each repetition, we estimate TE and IMI from the leader to the follower and vice versa. The conditional probabilities in equations (1)–(3) are computed by discretizing the time-series of the heading so to obtain b equally-spaced bins, with b being equal to 2, 3 or 4. To detect interaction between the particles, we then compare the observed values of MI, TE, and IMI with the corresponding null distributions obtained by simulating the model for $N = 1000$ repetitions each of length $T = 2000$ for $w_{\mathrm{LF}} = 0$, that is, in the absence of coupling between the leader and the follower.

Our results on the modified Vicsek model confirm the inadequacy of time-delayed mutual information and the superiority of transfer entropy to intrinsic mutual information in identifying directional interactions, see figures 4(A) and (B). For all values of the coupling gain from the leader to the follower, and for both numbers of bins, we observe a rate of false positives and/or of false negatives higher than 50%. Although transfer entropy and intrinsic mutual information exhibit comparable levels of sensitivity for different values of the coupling gain from the leader to the follower and different number of bins, their specificity can be dramatically different. With respect to sensitivity, for the lowest coupling value ($w_{\mathrm{LF}} = 0.1$), intrinsic mutual information yields a slightly larger fraction of false negatives (30% against 26% for $b = 2$, 52% against 26% for $b = 3$, and 44% against 19% for $b = 4$). As the coupling increases, the accuracy of the inference obtained through intrinsic mutual information improves, reaching the same levels of transfer entropy, with no false negatives for $w_{\mathrm{LF}} = 1$. With respect to specificity, performance is highly related to the number of bins used in the discretization of the time-series. For coarse binning, the specificity of the inference considerably deteriorates when choosing intrinsic mutual information over transfer entropy (when $w_{\mathrm{LF}} = 1$, the false positive rate is 56% against 1% for $b = 2$, and 14% against 2% for $b = 3$). Predictably, reducing the coupling from the leader to the follower mitigates the difference in the specificity of the two inferences (when $w_{\mathrm{LF}} = 0.1$, the false positive rate is 28% against 4% for $b = 2$, and 3% against 8% for $b = 3$), due to the weaker interaction between the units. A finer binning reduces the gap between the two metrics, whereby we register perfect specificity of both transfer entropy and intrinsic mutual information for $b = 4$.

Similar to the Boolean model, the difference in specificity should be sought in the relationship between intrinsic mutual information, transfer entropy, and time-delayed mutual information. We confirm that intrinsic mutual information is also well-approximated by the minimum between transfer entropy and time-delayed mutual information, whereby numerical values of intrinsic mutual information and the minimum of the other two classical metrics are statistically indistinguishable at a confidence level of 0.05 across 99.7% of the 900 cases (3 parameter values × 3 numbers of bins × 100 repetitions) reported in figure 4, see section 4 and supplementary note 4. Likewise, the cumulative null distribution of time-delayed mutual information is always above that of transfer entropy, see figures 4(C) and (D). As a result, the same three cases identified for the Boolean model in figure 3 are possible, and the extent to which intrinsic mutual information under-performs transfer entropy relates to instances of Case 3 being outnumbered by instances of Cases 1 and 2. The lowest specificity of intrinsic mutual information is registered when the false positive rate of time-delayed mutual information is larger than 50%; this corresponds to the occurrence of only 0.25% instances of Case 3.

3. Discussion

Intrinsic mutual information has been recently proposed as a precise measure of information flow in complex systems [19], bearing important insight into collective behavior [21]. Rephrasing the words of James *et al* [19], given two stochastic processes X and Y, there is an intrinsic information flow from X to Y when the past of X is individually predictive of the future of Y. Such an intrinsic information flow is not exactly quantified by transfer entropy, which also incorporates synergistic information flow, that is, the reduction of uncertainty about the future of Y by the simultaneous knowledge of the present state of X and Y. Likewise, it is not captured by time-delayed mutual information, which will also incorporate shared information flow, that is, when the past of X is predictive of the present of Y in the same manner as the past of Y. Intrinsic mutual information is an easy-to-compute upper bound for intrinsic information flow, which, different from transfer entropy, is free from contributions related to synergistic information. Whether intrinsic mutual information can be used for hypothesis-testing and inference of directional interactions between units from their time-series has never been attempted. In this study, we provide an answer to this question through closed-form results on a minimalistic Boolean model that captures salient features of leader–follower dynamics and simulation results on the modified Vicsek model by Sattari *et al* [21].

Our theoretical and computational results do not point at a practical advantage of intrinsic mutual information versus transfer entropy in the inference of pairwise interactions. Surprisingly, we observe that the precise quantification of information flow through intrinsic information does not bestow any advantage with respect to transfer entropy in both sensitivity and specificity. None of the considered scenarios, let them be simulations of the Boolean model or of the modified Vicsek model, offers evidence in favor of a performance improvement attained through the use of intrinsic mutual information. As such, care should be placed when employing intrinsic mutual information in the discovery of causal relationships. The simultaneous consideration of synergistic and intrinsic information flows by transfer entropy seems to offer a more reliable basis to minimize false positives and negatives compared to intrinsic mutual information.

While the Vicsek model is the most widespread choice for the study of collective dynamics from biology to swarm robotics [28], its general mathematical treatment is difficult, if not impossible. In its basic incarnation, the model leads to state-dependent, switched, nonlinear, stochastic dynamics that preclude the exact quantification of any information-theoretic quantity. Working with a Boolean model helps clarify two main aspects that would remain opaque from a mere computational endeavor. First, we determine under which condition (intrinsic noises and coupling gain) intrinsic mutual information reduces to any of the classical information-theoretic metrics (time-delayed mutual information and transfer entropy), offering further backing to the critique of transfer entropy by James *et al* [18] and reinforcing numerical predictions by Sattari *et al* [21]. We highlight a complex dependence of intrinsic mutual information on the system dynamics, whereby we demonstrate that intrinsic mutual information depends on added noise and on the strength of the coupling gain in a complex, nonlinear fashion. As a first approximation, intrinsic mutual information equals the minimum between time-delayed mutual information (compounding intrinsic and shared information flows) and transfer entropy (compounding intrinsic and synergistic information flows). This result suggests that shared and synergistic information flows do not coexist for the considered Boolean model, except for a narrow window of coupling gains.

Second, we pinpoint at the modalities by which intrinsic mutual information offers reduced performance in the inference of directional interactions compared to transfer entropy. While intrinsic mutual information and transfer entropy display similarly high sensitivity, intrinsic mutual information has considerably lower specificity. Low specificity of intrinsic mutual information can be traced back to the same sins of time-delayed mutual information, whose null distribution has a slimmer tail compared with that of transfer entropy, thus favoring the rejection of the null-hypothesis. Since intrinsic mutual information can be approximated as the minimum between time-delayed mutual information and transfer entropy, the tail of its null distribution will be at least as slim as that of time-delayed mutual information. When intrinsic mutual information coincides with transfer entropy (null synergistic information flow), it may happen that intrinsic mutual information would score a false positive, despite transfer entropy being capable of filtering a spurious interaction from the leader to the follower. When intrinsic mutual information is equal to time-delayed mutual information (null shared information flow), one cannot exclude the possibility that intrinsic mutual information would outperform transfer entropy. However, this would rely on mutual information exhibiting adequate specificity, a rare possibility throughout our statistical analysis. As a result, we warn prudence with the use of intrinsic mutual information as a tool for the discovery of directional interactions.

Several prior studies have pointed at the merit of exact results on information-theoretic metrics [25, 29–34]. The use of exact theoretical values rather than their statistical estimates alleviates the dependence of any claim on the statistical methods adopted for estimation and brings to light the specific role of model parameters on any information-theoretic metric. For example, Smirnov [30] computed closed-form results of transfer entropy over a class of benchmark systems (autoregressive processes and Markov chains), demonstrating typical factors that may lead to spurious couplings in real-world applications. Hahs and Pethel [31] have established closed-form results for transfer entropy for autoregressive processes with multiple timetags. Novelli *et al* [34] and Goodman and Porfiri [33] independently demonstrated the dependence of transfer entropy on topological properties of network nodes within theoretical studies of a linearly coupled Gaussian model and a Boolean system, respectively. Boolean models have been further investigated in a sequence of studies by some of these authors and others [25, 29, 32].

The study is not free of limitations. First, we presently lack of a general form for the cumulative distribution of conditionally independent variables for hypothesis-testing. As such, claims regarding superiority of transfer entropy against intrinsic mutual information in terms of specificity are based on numerical estimations of null distribution, conducted for specific parameter choices. Some work has been conducted in this direction [35], but available approximations are based on low-order Taylor expansions that do not consider the temporal structure of the time-series, thereby hindering their application to the problem of leader–follower interactions between systems with memory; see section 4 and supplementary note 5. Such a drawback is also at the core of the second, main limitation of this study: the lack of a comparison between the inferences of transfer entropy and intrinsic mutual information beyond coarse-grained dynamics for the modified Vicsek model. In fact, the present comparison is limited to discretizing the heading of the particles with at most four bins. Such a computation required about one hundred hours on a state-of-the-art machine, and computational time would scale exponentially with the number of bins. Access to a closed-form approximation for the null distributions of all the salient information-theoretic quantities for coarse- and fine-grained dynamics would address this issue.

Despite these two main limitations, our work brings forward important insight into the use of the novel concept of intrinsic mutual information as an inference tool of pairwise interactions underpinning collective dynamics. Transfer entropy has been, rightfully, criticized for its inability to detail information flow between coupled units [18, 19, 21]—a task that is seamlessly accomplished through the use of intrinsic mutual information. Yet, accomplishing this task may not translate into an improved statistical inference, especially with respect to specificity. Perhaps, this is one of the few cases in which Voltaire's famous aphorism applies: 'perfect is the enemy of good.'

4. Materials and methods

4.1. Intrinsic mutual information

For a discrete random variable Y, the uncertainty associated with Y is quantified by its (Shannon) entropy [1]

$$H(Y) = -\sum_{y\in\mathcal{Y}} \Pr(Y=y)\log_2 \Pr(Y=y). \tag{18}$$

Given another discrete random variable Z, the joint entropy of the pair (Y,Z) is

$$H(Y,Z) = -\sum_{y\in\mathcal{Y},z\in\mathcal{Z}} \Pr(Y=y,Z=z)\log_2 \Pr(Y=y,Z=z), \tag{19}$$

whereas the entropy of Y conditional to Z is

$$H(Y|Z) = -\sum_{y\in\mathcal{Y},z\in\mathcal{Z}} \Pr(Y=y,Z=z)\log_2 \Pr(Y=y|Z=z). \tag{20}$$

Note that the above definitions imply that $H(Y|Z) = H(Y,Z) - H(Z)$.

Mutual information between Y and Z is defined as

$$I(Y;Z) = H(Z) - H(Z|Y) = -\sum_{\substack{y\in\mathcal{Y}\\ z\in\mathcal{Z}}} \Pr(Y=y,Z=z)\log_2 \frac{\Pr(Y=y|Z=z)}{\Pr(Y=y)}. \tag{21}$$

By definition, mutual information is symmetric, whereby from the definition of conditional probability $\Pr(Y=y|Z=z) = \Pr(Y=y,Z=z)/\Pr(Z=z)$, so that one obtains that the right-hand-side of (21) is equal to $I(Z;Y)$. Furthermore, both entropy and mutual information are non-negative from Jensen inequality [1].

Next, given a third random variable W, we introduce conditional mutual information $I(Y;Z|W)$ as the mutual information between Y and Z conditional to W. This quantity is expressed as

$$I(Y;Z|W) = \sum_{\substack{y\in\mathcal{Y},z\in\mathcal{Z}\\ w\in\mathcal{W}}} \Pr(Y=y,Z=z,W=w)\log_2 \frac{\Pr(Z=z|Y=y,W=w)}{\Pr(Z=z|W=w)}. \tag{22}$$

Surprisingly, conditioning is not a subtractive operation so that conditioning on a third variable can increase information shared: it is possible that $I(Y;Z|W) > I(Y;Z)$. This phenomenon is known as conditional dependence [36] and a specific example based on exclusive OR logic has been proposed by Sattari *et al* [21] (therein, Y and Z are independent, but given W they become related in a deterministic manner).

In other words, conditional mutual information 'is sensitive to both intrinsic dependencies between Y and Z , as well as dependencies induced by W' [19]. A way to filter dependencies induced by W is to utilize the notion of intrinsic (conditional) mutual information between Y and Z when given W by Maurer and Wolf [20],

$$I(Y;Z\downarrow W) = \inf\left\{ I(Y;Z|\overline{W}) \text{ such that } \Pr(Y,Z,\overline{W}) = \sum_{w\in\mathcal{W}} \Pr(Y,Z,W=w)\Pr(\overline{W}|W=w)\right\}. \tag{23}$$

Here, $\overline{W}$ is an auxiliary variable taking values in $\mathcal{W}$ and related to W by means of the conditional probability $\Pr(\overline{W}|W)$—taking the form of an unknown (finite or infinite) $|\mathcal{W}|\times|\mathcal{W}|$ matrix. Intrinsic mutual information is the infimum of $I(Y;Z|\overline{W})$ over all possible random variables that can be generated from W through $\Pr(\overline{W}|W)$. When $\overline{W}$ is a constant ($\Pr(\overline{W}|W)$ corresponding to a matrix with all zeros but a column of ones) and $\overline{W}$ is identical to W ($\Pr(\overline{W}|W)$ corresponding to the identity matrix), intrinsic mutual information reduces to mutual information and conditional mutual information, respectively [21].

Intrinsic conditional mutual information has been used in cryptography as an upper bound for the secret key rate of transmission between a pair of sender/receiver having access to Y and Z against an adversary having access to W. In other words, the secret key is the maximum rate at which the sender/receiver can agree on a secret S so that the information that can be obtained on S from W is arbitrarily small. The definition of intrinsic mutual information begets the following, intuitive, inequalities:

$$0 \leqslant I(Y;Z\downarrow W) \leqslant I(Y;Z) \tag{24a}$$

$$I(Y;Z\downarrow W) \leqslant I(Y;Z|W). \tag{24b}$$

4.2. Relationship between information-theoretic metrics in the modified Vicsek model

The closed-form, asymptotic expressions of time-delayed mutual information, transfer entropy, and intrinsic mutual information for the Boolean model (6) indicate that for a wide range of parameters, intrinsic mutual information coincides with the minimum of time-delayed mutual information and transfer entropy. Such a claim is at the core of our explanation for reduced specificity of intrinsic mutual information when compared to transfer entropy.

We numerically verified whether this claim would also hold true for the modified Vicsek model (17). For low ($w_{\mathrm{LF}} = 0.1$), medium ($w_{\mathrm{LF}} = 0.5$), and high ($w_{\mathrm{LF}} = 1$) values of the coupling from the leader to the follower, we numerically estimated time-delayed mutual information, transfer entropy, and intrinsic mutual information, which, in turn, required the estimation of the probability density functions in equations (1)–(3), respectively. For each parameter value, these estimations were performed on $N = 1000$ time-series of leader and follower, each $T = 2000$ time-steps long. To account for the finiteness of the time-series, we associated with each point-estimate an interval at a confidence level of 0.05. Specifically, for each of the three information-theoretic measures, the width of the interval was selected as the 95th percentile of the cumulative null distribution obtained from simulating the case $w_{\mathrm{LF}} = 0$.

Overall, we found that the confidence interval for intrinsic mutual information overlaps with (at least one of) that of transfer entropy and mutual information in 99.6% of the cases, whereby intrinsic mutual information is statistically indistinguishable from the minimum between transfer entropy and mutual information. This result is robust to different choices of the coupling gains w_{LF} and number of bins b, see supplementary material.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

P De Lellis was supported by the Research Project PRIN 2017 'Advanced Network Control of Future Smart Grids' funded by the Italian Ministry of University and Research (2020–2023). M Ruiz Marín was supported by Ministerio de Ciencia, Innovación y Universidades under Grant Number PID2019-107800GB-I00/AEI/10.13039/501100011033. M Porfiri was supported by the National Science Foundation under Grant Numbers ECCS 1928614, CMMI 1901697, and CMMI 1953135. This study was part of collaborative activities carried out under the program of the region of Murcia (Spain): 'Groups of Excellence of the region of Murcia, the Fundación Séneca, Science and Technology Agency' Project 19884/GERM/15.

ORCID iDs

Pietro De Lellis https://orcid.org/0000-0002-2656-6454
Manuel Ruiz Marín https://orcid.org/0000-0001-9228-6410
Maurizio Porfiri https://orcid.org/0000-0002-1480-3539

References

[1] Cover T M and Thomas J A 1991 *Elements of Information Theory* vol 1 (New York: John Wiley & Sons, Inc.) pp 279–335
[2] Vicente R, Wibral M, Lindner M and Pipa G 2011 *J. Comput. Neurosci.* **30** 45–67
[3] Bullmore E and Sporns O 2009 *Nat. Rev. Neurosci.* **10** 186–98
[4] Porfiri M, Sattanapalle R R, Nakayama S, Macinko J and Sipahi R 2019 *Nat. Hum. Behav.* **3** 913–21
[5] Grabow C, Macinko J, Silver D and Porfiri M 2016 *Chaos* **26** 083113
[6] Hlinka J, Hartman D, Vejmelka M, Runge J, Marwan N, Kurths J and Paluš M 2013 *Entropy* **15** 2023–45
[7] Runge J *et al* 2019 *Nat. Commun.* **10** 1–13
[8] Orange N and Abaid N 2015 *Eur. Phys. J. Spec. Top.* **224** 3279–93
[9] Valentini G, Pavlic T P, Walker S I, Pratt S C, Biro D and Sasaki T 2021 *Elife* **10** e68653
[10] Pilkiewicz K *et al* 2020 *J. R. Soc. Interface* **17** 20190563

[11] Schreiber T 2000 *Phys. Rev. Lett.* **85** 461
[12] Bossomaier T, Barnett L, Harré M and Lizier J T 2016 Transfer entropy *An Introduction to Transfer Entropy* (Berlin: Springer) pp 65–95
[13] Runge J 2018 *Chaos* **28** 075310
[14] Papana A, Papana-Dagiasis A and Siggiridou E 2020 *Int. J. Bifurcation Chaos* **30** 2050250
[15] Sun J and Bollt E M 2014 *Physica* D **267** 49–57
[16] Runge J, Heitzig J, Marwan N and Kurths J 2012 *Phys. Rev.* E **86** 061121
[17] Staniek M and Lehnertz K 2008 *Phys. Rev. Lett.* **100** 158101
[18] James R G, Barnett N and Crutchfield J P 2016 *Phys. Rev. Lett.* **116** 238701
[19] James R G, Ayala B D M, Zakirov B and Crutchfield J P 2018 arXiv:1808.06723
[20] Maurer U M and Wolf S 1999 *IEEE Trans. Inf. Theory* **45** 499–514
[21] Sattari S, Basak U S, James R G, Perrin L W, Crutchfield J P and Komatsuzaki T 2022 *Sci. Adv.* **8** 1–13
[22] Vicsek T, Czirók A, Ben-Jacob E, Cohen I and Shochet O 1995 *Phys. Rev. Lett.* **75** 1226
[23] Lizier J T, Pritam S and Prokopenko M 2011 *Artif. Life* **17** 293–314
[24] Hartnett A T, Schertzer E, Levin S A and Couzin I D 2016 *Phys. Rev. Lett.* **116** 038701
[25] Porfiri M 2017 *IEEE Trans. Control Netw. Syst.* **5** 1864–74
[26] Cassandras C G and Lafortune S 2009 *Introduction to Discrete Event Systems* (Berlin: Springer Science & Business Media)
[27] DeLellis P, Porfiri M and Bollt E M 2013 *Phys. Rev.* E **87** 022818
[28] Vicsek T and Zafeiris A 2012 *Phys. Rep.* **517** 71–140
[29] Mori F and Okada T 2020 *Phys. Rev. Res.* **2** 043432
[30] Smirnov D A 2013 *Phys. Rev.* E **87** 042917
[31] Hahs D W and Pethel S D 2013 *Entropy* **15** 767–88
[32] Porfiri M and Ruiz Marín M 2018 *IEEE Trans. Netw. Sci. Eng.* **5** 42–54
[33] Goodman R H and Porfiri M 2020 *Math. Eng.* **2** 34–54
[34] Novelli L, Atay F M, Jost J and Lizier J T 2020 *Proc. R. Soc.* A **476** 20190779
[35] Goebel B, Dawy Z, Hagenauer J and Mueller J C 2005 An approximation to the distribution of finite sample size mutual information estimates *IEEE Int. Conf. on Communications* vol 2 pp 1102–6
[36] Husmeier D 2005 Introduction to learning Bayesian networks from data *Probabilistic Modeling in Bioinformatics and Medical Informatics* (Berlin: Springer) pp 17–57