Predicting the evolution of a large system of units using its structure of interaction is a fundamental problem in complex system theory. And so is the problem of reconstructing the structure of interaction from temporal observations. Here, we find an intricate relationship between predictability and reconstructability using an information-theoretical point of view. We use the mutual information between a random graph and a stochastic process evolving on this random graph to quantify their codependence. Then, we show how the uncertainty coefficients, which are intimately related to that mutual information, quantify our ability to reconstruct a graph from an observed time series, and our ability to predict the evolution of a process from the structure of its interactions. We provide analytical calculations of the uncertainty coefficients for many different systems, including continuous deterministic systems, and describe a numerical procedure when exact calculations are intractable. Interestingly, we find that predictability and reconstructability, even though closely connected by the mutual information, can behave differently, even in a dual manner. We prove how such duality universally emerges when changing the number of steps in the process. Finally, we provide evidence that predictability-reconstruction dualities may exist in dynamical processes on real networks close to criticality.

The relationship between structure and function is fundamental in complex systems, and important efforts have been invested in developing network models to better understand it. In particular, models of dynamics on networks have been proposed to assess the influence of network structure over the temporal evolution of the activity in the system. In turn, data-driven models, dimension-reduction techniques, and mean-field frameworks have deepened our predictive capabilities. Among other things, these theoretical approaches have shed light on the relationship between dynamics criticality and many network properties such as the degree distribution, the eigenvalue spectrum, and their community structure. Fundamentally, these contributions justify our inclination for measuring and using real-world networks as a proxy to predict the behavior of complex systems.

Models of dynamics on networks have also been used as reverse engineering tools for network reconstruction, when the networks of interactions are unavailable, noisy or faulty. The network reconstruction problem has stimulated many technical contributions: Thresholding matrices built from correlation or other more sophisticated measures of time series, Bayesian inference of graphical models, and models of dynamics on networks, among others. These techniques are widely used (e.g., in neuroscience, genetics, epidemiology, and finance) to reconstruct interaction networks on which network science tools can then be applied.
Interestingly, dynamics prediction and network reconstruction are usually considered separately, even though they are related to one another. The emergent field of network neuroscience is perhaps the most actively using both notions: Network reconstruction for building brain connectomes from functional time series, then dynamics prediction for inferring various brain disorders from these connectomes. Recent theoretical works have also taken advantage of these notions to suggest that dynamics may partly depend on the structure. In ref. 51, it was shown that time series generated by a deterministic dynamics evolving on a specific graph can be accurately predicted by a broad range of other graphs. These findings highlight how poor our intuition can be with regard to the relationship between predictability and reconstructability. Furthermore, recent breakthroughs in deep learning on graphs have benefited from proxy network substrates to enhance the predictive power of their models, with applications in epidemiology and pharmaceutics. However, the use of graph neural networks and those proxy network substrates is only supported by numerical evidence and lacks a rigorous theoretical justification. As a result, their enhanced predictability remains to be fully corroborated. There is therefore a need for a solid, theoretical foundation of reconstructability, predictability, and their relationship in networked systems.

In this work, we establish a rigorous framework that lays such a foundation based on information theory. Information theory has been regularly applied to networks and dynamics in the past. In network science, it has been used to characterize random graph ensembles—e.g., the configuration model and stochastic block models—to develop network null models and to perform community detection. In stochastic dynamical systems, information-theoretical measures have been proposed to quantify their predictability, complexity and causal emergence. In statistical mechanics, information transmission has been shown to reach a maximum value near the critical point of spin systems in equilibrium.

Our objective is to combine these ideas into a single framework, motivated by recent works involving spin dynamics on lattices and deterministic dynamics. Our contributions are fourfold. First, we use mutual information between structure and dynamics as a foundation for our general framework to quantify the structure-function relationship in complex systems. Second, this codependence naturally leads to the definition of measures of predictability and reconstructability. Doing so allows us to conceptually unify prediction and reconstruction problems, i.e., two classes of problems that are usually treated separately. Third, we design efficient numerical techniques for evaluating these measures on large systems. Finally, we identify a new phenomenon—a duality—where our prediction and reconstruction capabilities can vary in opposite directions. These findings further our understanding of the complexity of modeling networked complex systems, such as the brain, where both prediction and reconstruction techniques play critical roles.

Results

Information theory of dynamics on random graphs

Let us consider a random graph whose support, $\mathcal{G}$, consists in the set of all graphs of $N$ vertices, each of which has its respective nonzero probability $P(G | \mathcal{G})$ with $G \in \mathcal{G}$. In our framework, $P(G)$ can be any graph distribution and reflects, from a Bayesian perspective, our prior knowledge of the structure of the system. We also consider a general discrete-time stochastic process (also called dynamics hereafter) with $T$ time steps evolving on a realization of $G$ and representing the

![Figure 1](https://doi.org/10.1038/s41467-024-48020-x)

**Fig. 1** Information diagram of dynamics on random graphs. a Areas represent amounts of information: The entropies related to $G$ are shown on the left in blue and those related to $X$ are on the right in orange. Mutual information—the red intersection of $X$ and $G$—corresponds to the information shared by both $G$ and $X$. b The highly predictable / weakly reconstructable scenario, where $H(G) > H(X)$. c The reverse scenario, i.e., highly reconstructable / weakly predictable, where $H(G) > H(X)$ meaning that $H(X; G)$ contains most of the information related to the graph, but only a small fraction of the information related to the dynamics.
possible states of the system. More precisely, we denote \( P(X|G) \) the probability of a random and discrete-time state \( X = (X_{i,t})_{i,t} \) conditioned on \( G \), where \( X_i \) is the random state, with discrete support \( \Omega_i \) of vertex \( i \in \{1, \ldots, N\} \) at time \( t \in \{1, \ldots, T\} \). We stress that \( X \) is at this point any stochastic process be it Markovian or not. The initial condition of the process is \( X_1 = (X_{i,1})_i \). While we only exposed our framework in terms of discrete-time and discrete-state processes, it can be used for continuous-state deterministic dynamics (see Supplementary Note III) and in principle, it can also be generalized to continuous-state stochastic processes by considering a probability density function \( p(X|G) \).

The variables \( X \) and \( G \) form themselves a Bayesian network \( G \to X \), where the arrow indicates conditional dependence. From this model, we are interested in the mutual information between \( X \) and \( G \)—which is a symmetric measure that quantifies the codependence between the dynamics \( X \) and the structure \( G \), where \( I(X; G) = 0 \) when they are independent. It is equivalently given by

\[
I(X; G) = H(X) - H(X|G)
\]

where \( H(G) = -\langle \log P(G) \rangle \) and \( H(X) = -\langle \log P(X) \rangle \) are respectively the marginal entropies of \( G \) and \( X \), and \( H(G|X) = -\langle \log P(G|X) \rangle \) and \( H(X|G) = -\langle \log P(X|G) \rangle \) are their corresponding conditional entropies. In the previous equations, the marginal distribution for \( X \), the evidence, is defined as \( P(X) = \sum_{G \in \Omega} P(G) P(X|G) \). The posterior is obtained from Bayes’ theorem as \( P(G|X) = P(G) P(X|G) / P(X) \), using the given graph prior \( P(G) \) and the dynamics likelihood \( P(X|G) \).

\( I(X; G) \) is a non-negative measure bounded by \( 0 \leq I(X; G) \leq \min\{H(G), H(X)\} \). Figure 1a provides an illustration of Eq. (1) in terms of information diagrams.

The measures presented in Eq. (1) and above can all be interpreted in the context of information theory. Information is generally measured in bits which in turn is interpreted as a minimal number of binary—i.e., yes/no—questions needed to convey it. While entropy measures the uncertainty of random variables like \( X \) and \( G \), i.e., the minimal number of bits of information needed to determine their value, mutual information represents the reduction in uncertainty about one variable when the other is known. The fact that it is symmetric means that this reduction goes both ways: The reduction in the dynamics uncertainty when the structure is known is equal to that of the structure when the dynamics is known. Hence, mutual information measures the amount of information shared by both \( X \) and \( G \).

As an illustration, let us consider the physical example of a spin system that depends on \( G \) through a coupling parameter \( J \geq 0 \), where the spins are more (large \( J \)) or less (small \( J \)) likely to align with their first neighbors in \( G \). At \( J = 0 \), the spins are completely uncorrelated and flip with probability \( \frac{1}{2} \). In this case, \( H(X|G) = NT \) bits, corresponding to the maximum entropy of \( X \). We need precisely one binary question for each spin at each time for a given structure \( G \)—e.g., is the spin of vertex \( i \) at time \( t \) up? When \( J > 0 \), correlation is introduced between connected spins. As a result, a single question about the spin of vertex \( i \) at time \( t \) can provide additional information about the spins of other vertices at other times, and thus, \( H(X|G) < NT \). The interpretation of \( H(X) \) is analogous to that of \( H(X|G) \), as it measures the number of binary questions needed to determine \( X \) when the graph is unknown. From this perspective, the mutual information \( I(X; G) \), as expressed by the difference between \( H(X) \) and \( H(X|G) \), is the reduction in the number of questions needed to predict \( X \) ensuing from the knowledge of \( G \). Hence, \( I(X; G) \) measures to which extent the knowledge of the graph \( G \) improves our ability to forecast \( X \), i.e. its temporal predictability.

Similar observations can be made from the structural perspective. Suppose that \( X \) is the spin dynamics mentioned previously and \( G \) is a random graph, where each edge exists independently with probability \( p \). This yields \( H(G) = - \left( \frac{N}{2} \right) \log p + \left( 1 - p \right) \log \left( 1 - p \right) \), where \( \left( \frac{N}{2} \right) \) is the total number of possible undirected edges. When \( p \) is \( \frac{1}{2} \), we have \( H(G) = \left( \frac{N}{2} \right) \) bits, which is again the maximum entropy of \( G \). We therefore need precisely one binary question for each of the \( \left( \frac{N}{2} \right) \) edges in the graph—e.g., is there an edge between \( i \) and \( j \) to completely determine its state. When the dynamics \( X \) is known, \( H(G|X) \) is interpreted similarly to \( H(G) \), but also takes into account the observation of the spins \( X \) which introduces correlation between the edges of \( G \). As a result, each bit can provide information about more than one edge, even in the case \( p = \frac{1}{2} \) where we a priori need one bit per possible
edge to fully reconstruct $G$. Consequently, the knowledge of $X$ reduces uncertainty about $G$ (i.e., $H(G|X) \leq H(G)$, see ref. 52, Theorem 2.6.5), and therefore allows for its reconstruction; $I(X; G)$ thus measures the reconstructability of $G$, i.e. the extent to which information about $G$ can be revealed from $X$.

In practice, $I(X; G)$ can be used to explain the performance of both prediction and reconstruction algorithms (see “Performance of prediction and reconstruction algorithms” section for further detail). From the prediction perspective, it quantifies the sensitivity of the time series to the structure of interactions prescribed by $G$, i.e., the gain in predictability of including $G$ for the extrapolation of $X$. This can be measured by comparing the true transition probabilities of the process $X$ as given by the conditional model $P(X|G)$, with those predicted by models that do not include $G$ in their predictions. This experiment was performed in ref. 51 for deterministic dynamics on graphs, to show that high prediction accuracy of time series can sometimes be achieved without the knowledge of the true graph. In Fig. 2a, we use the mean absolute error—the same measure as in ref. 51—to perform the comparison. In turn, we associate the high predictive capabilities of the true conditional model where the error with the graph-independent model is high. Likewise, $I(X; G)$ > 0, we expect the system to be predictable and reconstructable to a certain degree. Otherwise, when $I(X; G) = 0$, the system is said both unpredictable and un reconstructable. Yet, $I(X; G)$ by itself is hardly comparable from one system to another. Indeed, a specific value of $I(X; G)$ may correspond to opposing scenarios when it comes to predictability and reconstructability, as shown in Fig. 1b, c. Thus, it is more convenient to use normalized quantities such as the uncertainty coefficients 

$$U(X|G) = \frac{I(X; G)}{H(X)},$$

$$U(G|X) = \frac{I(X; G)}{H(G)},$$

which are bounded between 0 and 1. Contrary to $I(X; G)$, $U(X|G)$ and $U(G|X)$ represent relative amount of information. For instance, $U(G|X) = 1$ implies that $I(X; G) = H(G)$, which in principle means that perfect reconstruction can be achieved as all the information of $G$ is contained in $X$. Likewise, $U(X|G) = 1$ means that $I(X; G) = H(X)$, which indicates that all the information in $X$ is determined by $G$: a perfectly accurate prediction of $X$ can be made with $G$ alone. This maximum value is guaranteed when $X$ is deterministic and there is only one initial condition (see Supplementary Note III). Having $U(X|G) = 0$ implies that $U(X|G) = U(G|X) = 0$, which again means that $G$ and $X$ are independent. Any value in-between of $U(X|G)$ and $U(G|X)$ represents different degrees of predictability and reconstructability, respectively.

The “Simple example” section will present simple concrete examples to provide a better intuition about these concepts. Before we get to these examples, we investigate the influence of the knowledge of the past of $X$ over the relationship between its future and its structure, as measured through reconstructability and predictability.

### Past-dependent mutual information

It is often the case that predictability measures the sensitivity to the initial conditions of a process $X$. For instance, refs. 66,68,83,84 used different versions of the mutual information between $X_t$ and $X_s$ as a direct measure of predictability. Then, a system is more predictable if the past allows it to better predict the future. In this spirit, we generalize our framework in such a way that the mutual information between the process $X$ and its structure $G$ includes some information about the past of $X$.

We define $X_{past}$ as the past of $X$ and $X_{future}$ as its future, such that $X = (X_{past}, X_{future})$, see Fig. 3a. We define $\tau$ as the length of $X_{past}$ and $T - \tau$ as the length of the future $X_{future}$. Our measure of interest in this case is $I(X_{future}; G|X_{past})$, and it is equal to

$$I(X_{future}; G|X_{past}) = I(X; G) - I(X_{past}; G),$$

which is a conditional mutual information—the green intersection in Fig. 3a. In turn, a small $\tau$ includes less contribution to the observed past, which leads to a scenario increasingly similar to that presented in “Information theory of dynamics on random graphs” section as shown by Fig. 3b. As $\tau$ gets larger, more contribution is left to $X_{past}$ resulting in a smaller $I(X_{future}; G|X_{past})$, even though the total mutual information $I(X; G)$—the union of the red, green, and gray sets—is large (see Fig. 3c). Similarly to “Information theory of dynamics on random graphs” section, we then define the partial uncertainty coefficients, bounded
between 0 and 1:

\[
U(X_{\text{future}}|G;X_{\text{past}}) = \frac{I(X_{\text{future}};G|X_{\text{past}})}{H(X_{\text{future}}|X_{\text{past}})},
\]

(4a)

\[
U(G|X_{\text{future}};X_{\text{past}}) = \frac{I(X_{\text{future}};G|X_{\text{past}})}{H(G|X_{\text{past}})},
\]

(4b)

measuring the partial predictability of \(X_{\text{future}}\) from \(G\) and partial reconstructability of \(G\) given \(X_{\text{future}}\), respectively. The above quantities can be expressed in terms of previously visited ones. In fact, in Eq. (3), \(I(X;G)\) and \(I(X_{\text{past}};G)\) can be expressed using Eq. (1). Likewise, the normalizing factor \(H(X_{\text{future}}|X_{\text{past}})\) is expressed in terms of state entropies using the joint entropy \(H(X) = H(X_{\text{past}},X_{\text{future}})\), i.e., \(H(X_{\text{future}}|X_{\text{past}}) = H(X_{\text{future}}) - H(X_{\text{past}})\). And finally, \(H(G|X_{\text{past}})\) is evaluated similarly to \(H(G|X)\).

Whereas the interpretation of the partial uncertainty coefficients is analogous to those presented in the previous section, they nevertheless measure conceptually different quantities. Indeed, by using \(I(X_{\text{future}};G|X_{\text{past}})\), it is implied that the information about the past has been removed from the total mutual information between \(X\) and \(G\). As a result, the partial predictability \(U(X_{\text{future}}|G;X_{\text{past}})\) measures the gain in predictability over \(X_{\text{future}}\), when including \(G\) in the prediction, compared to a model which only uses \(X_{\text{past}}\). Additionally, the removed information likely includes some information about \(G\), since \(I(X_{\text{past}};G) \geq 0\). Hence, the partial reconstructability, as defined by \(U(G|X_{\text{future}};X_{\text{past}})\), measures the reconstructability of the remaining information about \(G\) when observing \(X_{\text{future}}\), i.e., information which has not been unveiled from the observation of \(X_{\text{past}}\).

In essence, for some \(\xi > 0\), the case \(r = 1\) with \(T = \xi + 1\) is similar to the case \(T > \xi\) with \(T > \xi\) since \(X_{\text{future}}\) have the same length \(\xi\) in both cases. From a reconstruction perspective, they quantify the reconstructability of \(G\) from a process with \(\xi\) time steps. However, the reconstructed information is quite different in both cases, since with \(r = 1\) and \(T = \xi + 1\) no prior information is given—assuming that the initial conditions \(X_0\) are independent from \(G\)—while a lot of information has already been processed when \(T = T - \xi\). Furthermore, increasing \(r\) draws our attention away from the actual relationship between \(X\) and \(G\) of interest, since this relationship should exclude all information about \(X_{\text{past}}\). For this reason, we will mostly focus on the case \(r = 1\) in the remainder of the paper.

**Simple example**

The interpretation of reconstructability and predictability in terms of \(U(G|X)\) and \(U(X|G)\) can be grasped more firmly through an elementary example. We consider a system where only two graphs are possible, namely \(g_1\) and \(g_2\), such that \(P(g_1) = p\) and \(P(g_2) = 1 - p\). The entropy of this graph is therefore \(H(G) = -\xi(p)\), where \(\xi(p) = -p \log p - (1-p) \log (1-p)\) is the binary entropy. These two graphs can generate together three outcomes for \(X\), i.e., \(x_1, x_2,\) or \(x_3\). The graph \(g_1\) generates \(x_1\) and \(x_2\) with probabilities \(r\) and \(1-r\) respectively. Likewise, \(g_2\) generates \(x_3\) with respective probabilities \(s\) and \(1 - s\). As we can see, \(x_1\) can only be generated by \(g_1\) and \(x_3\) can only be the outcome of \(g_2\), while \(x_2\) can be generated by both graphs.

We now focus on the scenario with \(s = 0\)—the general expressions for \(I(X;G)\) and the other entropies are obtained in Section II of the Supplementary Information. In this case, only \(g_1\) can generate \(x_1\) and \(x_2\), while \(g_2\) can only generate \(x_3\). Therefore, we have perfect reconstructability of either graphs, meaning \(U(G|X) = 1\) for any \(p\) and \(r\), since the outcome of \(X\) tells us immediately which graph generated it. However, \(X\) is imperfectly predictable from \(G\) since \(U(X|G) = \frac{\log_2 p}{\log_2 C_0(1-p)} < 1\) when \(0 < p, r < 1\), even though \(x_3\) can be perfectly predicted from \(g_2\) as it is its only possible outcome. The remaining entropy, i.e., the second term of the denominator, \(p/(r)(1-p)\), corresponds to the uncertainty related to whether \(g_1\) generates \(x_1\) or \(x_2\).

When \(s = 1\), the system is both partially predictable and reconstructable, with \(U(X|G) = 1 - \frac{\log_2 r}{\log_2 C_0} < 1\) and \(U(G|X) = \frac{\log_2 s}{\log_2 (1-p)} < 1\) for all \(0 < p, r < 1\). Both \(g_1\) and \(g_2\) can generate \(x_3\), but the probability that \(g_1\) generates \(x_2\) decreases with \(r\). This results in a gradual increase of predictability and reconstructability as \(r\) approaches 1, where the systems tend to a one-to-one mapping between the outcomes of \(G\) and \(X\).

The intermediate cases when \(0 < s < 1\) are also interesting because they give rise to an interval in \(r\) where the system becomes less predictable but more reconstructable as \(r\) increases, as highlighted by the gray area in Fig. 4(c). This happens because, as \(r\) increases for a fixed \(s\), the growth of entropy of \(X\) dominates \(I(X;G)\), resulting in the dual behavior of \(U(G|X)\) and \(U(X|G)\).

**0-duality between predictability and reconstructability**

Predictability and reconstructability in dynamics on random graphs offer two perspectives of the same information shared by \(G\) and \(X\)—two sides of the same coin. However, as we have previously seen with simple examples, predictability, and reconstructability do not necessarily go hand in hand even though they are related: An increasing \(U(G|X)\) according to some parameter \(\theta\) of the system does not necessarily imply an increase of \(U(X|G)\) and vice versa. Furthermore, a high value of \(U(G|X)\) is not tied to a high value of \(U(X|G)\), and conversely, as illustrated in Fig. 1b, c. Indeed, \(U(G|X)\) and \(U(X|G)\) can take opposing values, depending on \(H(G)\) and \(H(X)\), for the same value of \(I(X;G)\). This phenomenon can also be observed in the performance of prediction and reconstruction (see Supplementary Note 1). In the literature, a hint of the existence of such dual behavior was recently corroborated in ref. 51 for continuous-state deterministic dynamics. The authors showed that high prediction accuracy can be achieved with graphs reconstructed from the very time series they want to predict, even if they are different from the original graph that
generated the time series. This phenomenon can be understood through our framework (see Section III of the Supplementary Information) and we now devote the rest of the section to precisely define and characterize the somewhat counterintuitive phenomenon of duality.

We identify a duality when \( U(X|G) \) and \( U(G|X) \) vary in opposite directions when a parameter, say \( \theta \), is changed. More specifically, we say that they are dual with respect to \( \theta \), or \( \theta \)-dual, in an interval \( \Theta \) if and only if the signs of their derivative with respect to \( \theta \) are different for every \( \theta \in \Theta \):

\[
\left. \frac{\partial U(G|X) \partial U(X|G)}{\partial \theta} \right|_{\theta=\theta^*} < 0.
\]

This criterion formally relies on the existence of regions \( \Theta \) where the variations of \( U(G|X) \) and \( U(X|G) \) with respect to \( \theta \) are opposite, regardless of their amplitude (see also “Formal definition of \( \theta \)-duality” section). We use this criterion to relate the existence of extrema of \( U(G|X) \) and \( U(X|G) \) with that of regions of \( \theta \)-duality (see Lemma 1 in “Formal definition of \( \theta \)-duality” section).

With our intuition being established from simple examples and our precise definition, we are finally ready to state one of the main results of the paper. Recalling that \( T \) is the length of process \( X \), we prove that reconstructability and predictability are \( T \)-dual for a vast class of Markov chains.

**Proof of the universality of the \( T \)-duality** section. It is a consequence of the fact that the mutual information is strictly increasing with \( T \), and so is \( U(G|X) \) since \( H(G) \) is independent of \( T \) whenever the entropy rate of \( X \) is positive. As a result, \( U(G|X) \) is non-deterministic. Moreover, suppose that the state spaces of \( X \) and \( G \) are finite, and that \( X \) has a finite nonzero entropy rate and that \( G \) has a nonzero entropy. Then there exists a positive constant \( \phi \) such that the uncertainty coefficients \( U(G|X) \) and \( U(X|G) \) are \( T \)-dual for all \( T > \phi \).

**Theorem 1.** Let \( X = (X_1, X_2, \ldots, X_T) \) be a Markov chain of length \( T \) whose transition probabilities are conditional to some discrete random variable \( G \) that is independent of \( T \) and such that \( H(X,T|G) > 0 \) for all \( t \in [1, \ldots, T-1] \) (i.e., \( X \) is non-deterministic). Moreover, suppose that the state spaces of \( X \) and \( G \) are finite, and that \( X \) has a finite nonzero entropy rate and that \( G \) has a nonzero entropy. Then there exists a positive constant \( \phi \) such that the uncertainty coefficients \( U(G|X) \) and \( U(X|G) \) are \( T \)-dual for all \( T > \phi \).

The proof of this theorem is in “Proof of the universality of the \( T \)-duality” section. It is a consequence of the fact that the mutual information is strictly increasing with \( T \), and so is \( U(G|X) \) since \( H(G) \) is independent of \( T \) whenever the entropy rate of \( X \) is positive. As a result, \( U(G|X) \) is non-deterministic. Moreover, suppose that the state spaces of \( X \) and \( G \) are finite, and that \( X \) has a finite nonzero entropy rate and that \( G \) has a nonzero entropy. Then there exists a positive constant \( \phi \) such that the uncertainty coefficients \( U(G|X) \) and \( U(X|G) \) are \( T \)-dual for all \( T > \phi \).

**Theorem 1.** Let \( X = (X_1, X_2, \ldots, X_T) \) be a Markov chain of length \( T \) whose transition probabilities are conditional to some discrete random variable \( G \) that is independent of \( T \) and such that \( H(X,T|G) > 0 \) for all \( t \in [1, \ldots, T-1] \) (i.e., \( X \) is non-deterministic). Moreover, suppose that the state spaces of \( X \) and \( G \) are finite, and that \( X \) has a finite nonzero entropy rate and that \( G \) has a nonzero entropy. Then there exists a positive constant \( \phi \) such that the uncertainty coefficients \( U(G|X) \) and \( U(X|G) \) are \( T \)-dual for all \( T > \phi \).

The proof of this theorem is in “Proof of the universality of the \( T \)-duality” section. It is a consequence of the fact that the mutual information is strictly increasing with \( T \), and so is \( U(G|X) \) since \( H(G) \) is independent of \( T \) whenever the entropy rate of \( X \) is positive. As a result, \( U(G|X) \) is non-deterministic. Moreover, suppose that the state spaces of \( X \) and \( G \) are finite, and that \( X \) has a finite nonzero entropy rate and that \( G \) has a nonzero entropy. Then there exists a positive constant \( \phi \) such that the uncertainty coefficients \( U(G|X) \) and \( U(X|G) \) are \( T \)-dual for all \( T > \phi \).

The proof of this theorem is in “Proof of the universality of the \( T \)-duality” section. It is a consequence of the fact that the mutual information is strictly increasing with \( T \), and so is \( U(G|X) \) since \( H(G) \) is independent of \( T \) whenever the entropy rate of \( X \) is positive. As a result, \( U(G|X) \) is non-deterministic. Moreover, suppose that the state spaces of \( X \) and \( G \) are finite, and that \( X \) has a finite nonzero entropy rate and that \( G \) has a nonzero entropy. Then there exists a positive constant \( \phi \) such that the uncertainty coefficients \( U(G|X) \) and \( U(X|G) \) are \( T \)-dual for all \( T > \phi \).

Figure 5 illustrates the universality of the \( T \)-duality using the special case of binary Markov chains (i.e., \( \Omega = \{0,1\} \), see “Binary Markov chains on graphs” section). These systems are parametrized by their activation (0 → 1) and deactivation (1 → 0) probability functions, denoted \( a(n,\nu, m_i) \) and \( b(n,\nu, m_i) \), respectively. In general, the
activation and deactivation functions depend solely on $m_{1,2}$ and $m_{1,J}$, i.e., the number of active and inactive neighbors of vertex $i$ at time $t$. We present multiple examples of binary Markov processes with different origins in Table 1: The Glauber dynamics, the Susceptible-Infected-Susceptible (SIS) dynamics, and the Cowan dynamics.

The aforementioned Glauber dynamics, which have been used to describe the time-reversible evolution of magnetic spins aligning in a crystal, have been tremendously studied because of its critical behavior and its phase transition. Its stationary distribution is given by the Ising model which has found many applications in condensed-matter physics and statistical machine learning, to name a few. The SIS dynamics is a canonical model in network epidemiology, often used for modeling influenza-like disease, where periods of immunity after recovery are short. In this model, susceptible (or inactive) vertices can be infected by each of their infected (active) first neighbors, with a constant transmission probability, and recover from the disease with a constant recovery probability. The simplicity of the SIS model has allowed for deep mathematical analysis of its absorbing-state phase transition. Finally, the Cowan dynamics has been proposed to model the neuronal activity in the brain. In this model, quiescent neurons fire if their input current, coming from their firing neighbors, is above a given threshold. Its mean-field approximation reduces to the Wilson-Cowan dynamics, one of the most influential models in neuroscience. For each model, we can identify an inactive state—down, susceptible, or quiescent—and an active one—up, infectious, or firing. The corresponding activation and deactivation probabilities are given in Table 1.

Figure 5 numerically supports Theorem 1 and clearly illustrates the $T$-duality for each dynamics, with different values of their parameters and different past lengths $\tau$. We used the Erdős-Rényi model as the random graph on which these dynamics evolve. The support $\mathcal{G}$ is the set of all simple graphs of $N$ vertices with $E$ edges, and

$$P(G) = \binom{N}{2}^{-1}.$$

Note that, in this example, we consider the well-known Erdős-Rényi model for simplicity (Eq. (6)). Furthermore, we considered very small graphs of size $N=5$, because the exact evaluation of $I(X;G)$ is computationally intractable. For larger systems, biased estimators can be designed to bound $I(X;G)$ as we show in “Estimators of the mutual information” section. We demonstrate the flexibility of our framework with regard to the random graph models by using more sophisticated and data-driven graph models in the following section.

The $T$-duality persists for the past-dependent measures presented in “Past-dependent mutual information” section, as illustrated by the bottom row of Fig. 5, for $\tau = 7/2$. However, note that for sufficiently large $\tau$, the duality seems to disappear. We refer to Section VII of the Supplementary Information for further detail. One can only wonder how many different kinds of parameters can lead to $\theta$-dualities. Maybe some may control the general behavior of the dynamics, and others some aspect of the system structure which, in turn, may also impact the dynamics. In the next section, we investigate those that are related to critical phenomena in complex systems.

### Duality and criticality

Despite their different nature and range of applications, the three models presented in Table 1 share several properties of interest. For instance, each model has a coupling parameter that controls the influence of the state of the first neighbors on the transition probabilities. They all also feature a phase transition in the infinite size limit whose position is determined by the coupling parameter (see Section IX of the Supplementary Information). We now investigate the influence of criticality over the existence of $\theta$-dualities, where $\theta$ is a coupling parameter.

For the Glauber dynamics, this parameter is the coupling constant $\lambda$, which dictates the reduction (increase) in the total energy of a spin configuration when two neighboring spins are parallel (antiparallel). The Glauber dynamics features a continuous phase transition at a critical point $J_c$, where for $J < J_c$, the spins are disordered resulting in a vanishing magnetization, and for which this magnetization is nonzero when $J > J_c$.

For the SIS dynamics, it is the transmission rate $\mu$ that acts as a coupling parameter. Like the Glauber dynamics, the SIS dynamics possesses a continuous phase transition where, for $\lambda < \lambda_c$, the system reaches an absorbing—or inactive—state from which it cannot escape, and an active state, when $\lambda > \lambda_c$, where a nonzero fraction of the vertices remain active over time. It should be emphasized that in our considered version of the SIS dynamics, referring to the system reaching a true absorbing state is not strictly accurate due to the allowance for self-infection $\epsilon$, which enables escape from the completely inactive state. Instead, the system approaches a metastable state with most vertices becoming asymptotically inactive. However, it can be shown that the two-phase transitions are quite similar for small $\epsilon$.

The Cowan dynamics can both feature a continuous or a first-order phase transition between an inactive and an active phase depending on the value of slope $\nu$, for which the coupling parameter is $\nu$, i.e., the potential gain for each firing neighbors. The continuous and first-order phase transitions of the Cowan dynamics are quite different in that the latter is characterized by two thresholds, namely the forward and backward thresholds $\nu < \nu^c$ and $\nu^b$, respectively (see Section IX in the Supplementary Information). Hence, the Cowan dynamics has a first-order phase transition that exhibits a bistable region $\nu < \nu^c < \nu^b$, where both the inactive and active phases are reachable depending on the initial conditions.

To account for the heterogeneous network structure observed in a wide range of complex systems, we simulate the dynamics on the configuration model, a random graph whose—potentially heterogeneous—degree sequence $k$ is fixed and whose support $\mathcal{G}$ corresponds to the set of all loopy multigraphs of degree sequence $k$. The probability of a multigraph $g$ in this ensemble is

$$P(G = g) = \frac{(2E)!}{(2E)!} \frac{\prod [k_i]}{\prod [M_{ij}^0] M_{ij}^0},$$

where $M_0$ counts the number of edges connecting vertices $i$ and $j$ in the multigraph and $2E = \sum k_i$ is the number of half-edges in $g$. Like the Erdős-Rényi model, the configuration model fixes the number of edges, but also fixes the degree sequence $k = (k_1, \ldots, k_n)$.

Figure 6 shows the predictability and reconstructability, as estimated by the MF estimator, of the three dynamics evolving on instances drawn from the configuration model. The top row shows the...
results when using a synthetic degree sequence obtained from a geometric degree distribution, while for the bottom row, degree sequences from different real networks are used for each dynamics. These distributions are shown in Fig. 7. We used the Little Rock Lake food web\(^95\) (as in ref. 40) jointly with the Glauber dynamics to simulate the interaction between species. In the case of the SIS dynamics, we considered the European airline network\(^96\) to mimic the spread of an epidemic. Finally, to simulate the neural activity of the Cowan dynamics, we used the C. Elegans neural network\(^97\).

First, the results of Fig. 6 show a meaningful comparison between the dynamics for different types of structures. For example, on the one hand, the Glauber dynamics is globally less predictable than the other two, since its predictability coefficient is overall smaller. In other words, the knowledge of a graph provides less information about \(X\) in the Glauber dynamics in comparison with the others, relative to the total amount of information needed to predict \(X\). This is related to the time reversibility of the Glauber dynamics, which allows any vertex to transition from the inactive to the active state (and vice versa) with nonzero probability, at any time, effectively making the Glauber dynamics more random than the others—i.e. \(H(X)\) is greater for Glauber than the other processes. On the other hand, the SIS and Cowan dynamics are shown as practically unpredictable and unreconstructable when their coupling parameter is below their respective critical point. This precisely occurs in the inactive phase, where the system rapidly reaches the inactive state and no mutual information can be generated. By contrast, the Glauber dynamics does not reach an inactive state below its critical point, which explains the gradual increase in predictability and reconstructability in that region.

Several additional observations are worth making. All dynamics exhibit maxima for \(U(G | X)\) and \(U(X | G)\), which delineate a region of duality illustrated by the shaded areas (two for Cowan, that is one for each branch). These regions are close to, but systematically above, their respective phase transition thresholds, regardless of type of degree sequence. A similar phenomenon in spin dynamics on non-random lattices has been reported by previous works\(^79,80\), in which the information transmission rate between spins—a measure akin to \(I(X; G)\)—is maximized above the critical point. Our numerical results are consistent with theirs, and suggest that their findings regarding near-critical systems even apply beyond spin dynamics on fixed lattices, to other types of processes on more heterogeneous and random structures.

**Discussion**

In this work, we used information theory to characterize the structure-function relationship with mutual information. We showed how...
From a practical perspective, the existence of such a $\theta$-duality can be critical to network modeling applications, since it also suggests a predictability-reconstructability trade-off. On one hand, by choosing the parameter $\theta$, we can minimize the uncertainty of the reconstructed structure, but this may result in a structure that is less informative regarding the dynamics. On the other hand, we can consider the reverse case, where the process is maximally influenced by the inferred structure, whose uncertainty is nevertheless not minimized. Analogous to the position-momentum duality in the Heisenberg uncertainty principle of quantum mechanics, the predictability-reconstructability duality must be accounted for in our network models if we are to disentangle complex systems.

**Methods**

**Binary Markov chains on graphs**

The models used throughout the paper are for the most part Markov chains $X = (X_1, X_2, \ldots, X_T)$, that are governed by a conditional probability $P(X|G)$ that can be factored as follows:

$$P(X | G) = P(X_1) \prod_{t=1}^{T-1} P(X_{t+1} | X_t, G).$$

(8)

The probability $P(X_{t+1}|X_t, G)$ is the global transition probability from state $X_t$ to state $X_{t+1}$, and $P(X_t)$ represents the probability distribution of the initial conditions, which is independent of $G$ in our case. More specifically, we assume that $X_t$ is a random binary vector of size $N$, and that the global transition probability can be factored in terms of local transition probabilities as follows:

$$P(X_{t+1} | X_t, G) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \left[ a(n_{i, t+1} - n_{i, t}) \right] \left[ 1 - \alpha(n_{i, t+1} - n_{i, t}) \right] \left[ 1 - \beta(n_{i, t+1} - n_{i, t}) \right] \left[ 1 - \gamma(n_{i, t+1} - n_{i, t}) \right] \right\}.$$

(9)

As mentioned in “$\theta$-duality between predictability and reconstructability” section, the functions $a$ and $\beta$ correspond to the activation and deactivation probabilities. In the general case, they are dependent on the number of active neighbors $m_i$ and inactive neighbors $n_i$ of a node $i$ such that $m_i + n_i = k_i$, where $k_i$ is the degree of this node.

**Performance of prediction and reconstruction algorithms**

To substantiate our claim about the interpretation of $I(X; G)$, we used different prediction and reconstruction algorithms and compared in Fig. 2 their performance with $I(X; G)$. In this section, we elaborate on this analysis.

**Prediction algorithms.** The prediction algorithms used in Fig. 2 correspond to Markov models that predicts a transition--activation and deactivation--probability matrix $P_t$, where $P_{ij}$ corresponds to the probability that node $i$ at time $t$ transition to the active state in the next time step. To make the comparison with $I(X; G)$, we compare the transition probability matrix $P^*$ of the true model—in the case of Fig. 2, the Glauber dynamics where the entries of $P^*$ are given by the activation $a$ and deactivation $\beta$ probabilities (see Table 1)—with those predicted by models learned from time series generated by the Glauber dynamics. These models are trained with 100 concatenated time series, each generated using a different graph sampled from the Erdős-Rényi model. The models are then trained to predict the time series without the knowledge of the structure. The input of these models is the complete state of the system at time $t$, i.e., $X_t$, and the output is a vector $P_t = (P_{1t}, \ldots, P_{Nt})$, where $P_{it}$ is the predicted probability that node $i$ transition to the active state at time $t$. We use the mean absolute

![Fig. 7](image-url) | Degree distributions of the graphs used in Fig. 6. (a) Multigraphs with geometric degree distribution $\mathcal{P}(k) = (1-p)p^k$ where $p = 5/6$, with $N = 1000$ nodes and $M = 2500$ edges. (b) Little Rock Lake food web. (c) European airline route network. (d) C. Elegans neural network. See “Real networks” section for further details about the graphs.
error (MAE) between \( P' \) and \( P \) to compare them, i.e.,

\[
\text{MAE}(P', P) = \frac{1}{N^T} \sum_{t=1}^{N} \sum_{i=1}^{T} |P'_{it} - P_{it}|.
\]

(10)

In doing so, the MAE quantifies the difference between a graph-dependent model and a graph-independent one, which highlights the importance of \( G \) over the prediction of \( X \), which is a proxy of \( f(X; G) \).

We consider two graph-independent prediction models: a logistic regression model and a multilayer perceptron (MLP). In both models, the predicted transition probabilities at time \( t \) are given by

\[
\hat{P}_t = \frac{1}{e^{f(X_t)} + 1},
\]

where \( f(X_t) \) is a learnable function, that is linear for the logistic regression model, i.e.

\[
f_{\text{logistic}}(X_t) = AX_t \cdot b,
\]

and non-linear for the MLP:

\[
f_{\text{MLP}}(X_t) = \text{ReLU}[W_1\text{ReLU}(W_2X_t + b_1) + b_2],
\]

such that

\[
\text{ReLU}(X) = \begin{cases} X & \text{if } X > 0, \\ 0 & \text{otherwise} \end{cases}
\]

The weight matrices \( A, W_1, \) and \( W_2, \) and bias vectors \( b, b_1, \) and \( b_2, \) are learned via stochastic gradient descent using a cross-entropy loss.

**Reconstruction algorithms.** In Fig. 2, we also illustrated the relationship between the performance of reconstruction algorithms and \( f(X; G) \). These algorithms are given the time series and they compute a score matrix \( S \), such that \( S_{ij} \) for each pair of nodes \((i, j)\) correlates with a probability that an edge exists between them. For the correlation matrix method \(^{12} \), this score is simply the correlation coefficient:

\[
S_{ij} = \frac{\Sigma_{ij}}{\Sigma_1},
\]

where \( \Sigma_1 \) is the error variance of the auto-regressive model of \( i \) and \( \Sigma_{ij} \) is the error variance of the other model that also includes \( j \). Finally, in the transfer entropy method \(^{13} \), the score is given by the entropy transfer from the time series of \( j \) to the time series of \( i \):

\[
S_{ij} = T_{X_j \rightarrow X_i},
\]

where

\[
T_{X_j \rightarrow X_i} = H(X_{t\mid t-1}X_{t-1}) - H(X_{t\mid X_{t-1}X_{t-1}}).
\]

The entropies involved in the computation of \( T_{X_j \rightarrow X_i} \) are evaluated using the maximum likelihood estimators of the probabilities \( P(X_{t\mid X_{t-1}}) \) and \( P(X_{t\mid X_{t-1}, X_{t-1}}) \), estimated from the time series itself.

We quantify the accuracy of the reconstruction using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. This curve is obtained by comparing the true positive rate with the false positive rate, for different thresholds \( \theta \in [\min(S), \max(S)] \). The AUC, being the integral of that curve, therefore represents the probability that the score matrix \( S \) classifies correctly a node pair connected by an edge.

**Formal definition of \( \theta \)-duality**

In what follows, we define the duality between predictability and reconstructability by taking a more general stance: Instead of considering a stochastic process \( X \) evolving on a random graph \( G \), we let \( G \) be any discrete random variable conditioning the probability of \( X \). First, we define the local duality of the uncertainty coefficients. The latter are considered as continuously differentiable functions with respect to a parameter \( \theta \) whose domain is some non-empty interval of the real line.

\[\text{Definition 1. (Local duality). The uncertainty coefficients } U \mid X \mid G \text{ and } U \mid G \mid X \text{ are locally dual with respect to } \theta \text{ if and only if } \frac{\partial U(X \mid G)}{\partial \theta} \frac{\partial U(G \mid X)}{\partial \theta} < 0. \]

(19)

The definition of the \( \theta \)-duality, a global property, follows that of the local duality.

**Definition 2. (\( \theta \)-Duality). The uncertainty coefficients } U \mid X \mid G \text{ and } U \mid G \mid X \text{ are dual with respect to } \theta \text{ or } \theta \text{-dual, in the interval } \Theta \text{ if and only if they are locally dual for all values of } \theta \text{ in } \Theta. \text{ From these definitions, we relate the presence of extrema of } U \mid X \mid G \text{ and } U \mid G \mid X \text{ with the existence of a } \theta \text{-duality.} \text{ Lemma 1. (\( \theta \)-duality between extrema). Let } \Theta \text{ be a non-empty sub-interval of the variable } \theta \text{ whose one endpoint is a local extremum of } U \mid X \mid G \text{ and the other, a local extremum of } U \mid G \mid X. \text{ Moreover, suppose that } U \mid X \mid G \text{ and } U \mid G \mid X \text{ do not have critical points in } \Theta. \text{ Then the extrema points delineate a region of } \theta \text{-duality if and only if they are both maxima (or both minima).} \text{ The proof of this lemma is available in Supplementary Note IV.} \text{ Proof of the universality of the } \theta \text{-duality} \text{ In what follows, we prove Theorem 1, that shows the universality of the } \theta \text{-duality, where } T \text{ is the number of steps in the process } X. \text{ We make use of the two following lemmas, that are proved in Supplementary Information (Notes V and VI), regarding the monotonicity of } R \mid X \mid G \text{ with respect to } T \text{ and the existence of continuous extensions of } U \mid X \mid G \text{ and } U \mid G \mid X, \text{ that will allow us to apply the Definition 1 involving derivatives.} \text{ Lemma 2. (Monotonicity of mutual information with } T\text{). Let } X = (X_1, X_2, \ldots, X_T) \text{ be a Markov chain of length } T \text{ whose transition probabilities are conditional to some discrete random variable } G \text{ that is independent of } T \text{ and such that } H(X_{t+1} \mid X_t) > 0 \text{ for all } t \in [1, \ldots, T-1]. \text{ Suppose moreover that the state spaces of } X \text{ and } G \text{ are finite. Then the mutual information } I(X; G) \text{ is nonzero and monotonically increasing with } T \in Z++. \text{ Lemma 3. (Continuous extension of uncertainty coefficients with } T\text{). Let } X = (X_1, X_2, \ldots, X_T) \text{ and } G \text{ respectively be a Markov chain and a discrete random variable as in Lemma 2. Then the uncertainty coefficients } U \mid G \mid X \text{ and } U \mid X \mid G, \text{ interpreted as functions of } T \in Z++, \text{ can be uniquely generalized to functions, respectively } f(T) \text{ and } g(T), \text{ that are} \]
holomorphic for all $t \in \mathbb{C}$, and thus real analytic for all $t \in \mathbb{R}$. Moreover, $H(X)$ can be extended to a function $h(T)$ that is analytic for all $t \in \mathbb{R}$, except where $f(T) = 0$.

Next, we prove Theorem 1.

**Proof.** According to Lemma 3, the quantities $U(X|G)$, $U(G|X)$, and $H(X|G)$, which were originally defined as real functions of $T \in \mathbb{R}$, have unique analytic extensions on the positive real axis, i.e., $T \in \mathbb{R}_+$. This allows us to treat $U(X|G)$, $U(G|X)$, and $H(X)$ as continuously differentiable functions with respect to $T$, where $U(G|X) = \frac{xG}{t\rho}$ and $H(X)$ are also holomorphic.

Now, by hypothesis, the entropy rate of the Markov chain $X$, $R := \lim_{n \to \infty} \frac{H(X^n)}{n}$, is well defined and nonzero. Hence, $H(X) \cdot RT$, i.e., $H(X)$ is positive and asymptotically linearly increasing with $T$. Moreover, since $G$ is independent of $T$ and $I(X;G) > 0$, it follows that $(X,G)$ is monotonically increasing with respect to $T$ by Lemma 2. As a result, $U(G|X) = \frac{xG}{\rho}$ is also monotonically increasing, since its denominator is independent of $T$, by assumption. This translates to the strict inequality $\frac{\partial U(G|X)}{\partial T} > 0$. If there exists a $T$-duality, i.e., there is a domain of $T$ where Eq. (19) is true, then $U(X|G)$ must be monotonically decreasing with $T$ or $\frac{\partial U(G|X)}{\partial T} < 0$ in that domain. To prove this, note that we can relate the two uncertainty coefficients using

$$H(X) = \frac{H(G)U(G|X)}{U(X|G)}.$$  \hspace{1cm} (20)

This leads to the following differential equation

$$\frac{\partial}{\partial T} \left[ \log U(X|G) \right] = \frac{\partial}{\partial T} \left[ \log U(G|X) \right] - \frac{\partial}{\partial T} \left[ \log H(X) \right],$$  \hspace{1cm} (21)

where we used the fact that $\frac{\partial U(G|X)}{\partial T} = 0$. Hence, to show that $U(X|G)$ is monotonically decreasing with $T$, the following inequality must hold

$$\frac{\partial}{\partial T} \left[ \log U(G|X) \right] < \frac{\partial}{\partial T} \left[ \log H(X) \right].$$  \hspace{1cm} (22)

Suppose for a moment that $U(X|G)$ is in fact increasing, such that Eq. (22) is false. This will eventually give rise to a contradiction. Let $g(T) = U(G|X)$ and $h(T) = H(X)$ be continuous functions of $T$ such that their derivative with respect to $T$ are respectively given by $g'(T) = \frac{\partial g(T)}{\partial T}$ and $h'(T) = \frac{\partial h(T)}{\partial T}$. Note that $0 < f(t) \leq 1$ and $h(t) > 0$ for all $t \in \mathbb{R}_+$. If Eq. (22) is false, then

$$\log g(T) \geq \log h(T).$$  \hspace{1cm} (23)

Using Grönwall’s inequality, Theorem 1.2.1, we get

$$\frac{g(T)}{h(a)} \geq \frac{h(T)}{h(a)}, 0 < a < T.$$  \hspace{1cm} (24)

So far, we have established that $h(T) = H(X) - RT$ and that $U(G|X)$ is monotonically increasing. We have also proved that if $U(X|G)$ is not monotonically decreasing with $T$, then inequality (24) is satisfied. However, the latter inequality and $h(T) = RT$ readily imply that $g(T)$ belongs to the class $\Omega(T)$, which is the set of all $g(T)$ such that there exist positive constants, $S$ and $T'$, for which $g(T) \geq ST$ for all $T \in \mathbb{R}_+$.

Two cases must be considered. First, if $ST > 1$, then $g(T) \geq ST > 1$, which is in direct contradiction with $g(T) \leq 1$ whenever $T \geq T$. Second, if $ST \leq 1$, then choose $T' > S^1 \geq T'$, so that $g(T) \geq ST' > 1$ for all $T \geq T'$. This again contradicts the inequality $g(T) \leq 1$ whenever $T \geq T'$. As a result, inequality (24) cannot be satisfied when $T \geq \varphi$, with $\varphi = \max(T', T^*)$. We thus conclude that $U(X|G)$ is monotonically decreasing for all $T \geq \varphi$. Therefore, $U(G|X)$ and $U(X|G)$ are $T$-dual in the interval $[\varphi, \infty)$. $\square$

**Estimators of the mutual information**

The mutual information $I(X;G)$ is generally intractable. Its intractability stems from the evaluation of the evidence probability, which is defined by the following equation:

$$P(X = x) = \sum_{G \in G} P(G \in X)P(X = x|G = g).$$  \hspace{1cm} (25)

Indeed, this sum potentially counts a number of terms which grows exponentially with the number of vertices $N$ in the random graph. More specifically, the evidence probability appears in two entropy terms needed to compute the mutual information, namely the marginal entropy $H(X) = - \log P(X)$ and the reconstruction entropy $H(G|X) = - \log \left( \frac{P(G|X,G)}{P(X)} \right)$, where $\left( f(Y) \right)$ denotes the expectation of $f(Y)$. Fortunately, the evidence probability, and in turn the mutual information, can be estimated efficiently using Monte Carlo techniques, which we present in this section.

**Graph enumeration approach.** For sufficiently small random graphs $(N \leq 5)$, the evidence probability can be efficiently computed by enumerating all graphs of $G$ and by adding explicitly each term of Eq. (25). Then, we can estimate the mutual information by sampling $M$ graphs $\{g^{(m)}\}_{m=1,M}$, followed by $M$ time series $\{x^{(m)}\}_{m=1,M}$—such that $x^{(m)}$ is generated with $g^{(m)}$, and by computing the following arithmetic average:

$$I(X,G) \approx \frac{1}{M} \sum_{m=1}^{M} \log P(X = x^{(m)}|G = g^{(m)}) - \log P(X = x^{(m)}).$$  \hspace{1cm} (26)

The variance of this estimator scales with the inverse of $\sqrt{M}$. In Fig. 5, we used this estimator to compute the mutual information, where $M = 1000$.

**Variational mean-field approximation.** In this approach, we estimate the posterior probability instead of the evidence probability. According to Bayes’ theorem, the posterior probability is

$$P(G|X) = \frac{P(G)P(X|G)}{P(X)}.$$  \hspace{1cm} (27)

Behind this estimator is a variational mean-field (MF) approximation that assumes the conditional independence of the edges. For simple graphs, the MF posterior is

$$P_{MF}(G|X) = \prod_{i,j} \left[ \pi_{ij}(X)^{A_j} (1 - \pi_{ij}(X))^{1-A_j} \right],$$  \hspace{1cm} (28)

where $\pi_{ij}(X) = P(A_{ij} = 1|X)$ is the marginal conditional probability of the existence of the edge $(i,j)$ given $X$. For multigraphs, a similar expression can be obtained, but instead involves a probability $\pi_{ij}(m|X) = P(M_{ij} = m|X)$ that there are $m$ multiedges between $i$ and $j$. In this case, the MF posterior becomes

$$P_{MF}(G|X) = \prod_{i,j} \left[ \pi_{ij}(m|X)^{\delta_{m,0}} \right],$$  \hspace{1cm} (29)

where $\delta_{m,0}$ is the Kronecker delta. The MF approximation allows to compute a lower bound of the true posterior entropy, such that

$$H(G|X) \geq - \langle \log P_{MF}(G|X) \rangle,$$  \hspace{1cm} (30)
as a consequence of the conditional independence between the edges\(^{82}\), Theorem 2.6.5. Using the MF approximation and a strategy similar to the exact estimator, we compute the MF estimator of the mutual information as follows:

\[
I(G \mid X) \geq \frac{1}{M} \sum_{m=1}^{M} \left[ \log P_{MB}(G = g^{(m)} \mid X = x^{(m)}) - \log P(G = g^{(m)}) \right].
\]

(31)

To compute \(P_{MB}(G = g^{(m)} \mid X = x^{(m)})\), we sample a set \(Q^{(m)} := \{g_1^{(m)}, \ldots, g_\ell^{(m)}\}\) of \(\ell\) graphs from the posterior distribution \(P(G \mid X = x^{(m)})\). Then, we estimate the probabilities \(\pi_{ij}(X) \approx n_{ij}^{(m)} / \sum j n_{ij}^{(m)}\), using their corresponding maximum likelihood estimate, where \(n_{ij}^{(m)}\) is the number of times the edge \((i,j)\) is seen in \(Q^{(m)}\). An analogous maximum likelihood estimate is made in the multigraph case, where \(\pi_{ij}(ω) \approx n_{ij}^{ω(m)} / \sum ω n_{ij}^{ω(m)}\) counts the number of times there were \(ω\) multiedges between \(i\) and \(j\) in \(Q^{(m)}\). This estimator is a lower bound of the mutual information—a consequence of Eq. (30). Hence, it is biased, and the extent of this bias is dependent on the quality of the conditional independence assumption with respect to the true random graph. Note that the MF estimator can yield negative estimates of the mutual information (see Section VIII of the Supplementary Information).

In Fig. 6, we fix the number of graphs sampled from the posterior distribution to \(Q = 1000\), and propose \(5\) moves between each sample (see also “Markov chain Monte Carlo algorithm” section for more detail).

Markov chain Monte Carlo algorithm

To sample from the posterior distribution, we use a Markov chain Monte Carlo (MCMC) algorithm where, starting from a graph \(g\), we propose a move to graph \(g'\), according to a proposition probability \(P(G = g' \mid G = g)\), and accept it with the Metropolis-Hastings probability:

\[
\min \left( 1, e^{-\Delta} \frac{P(G = g')G = g)}{P(G = g' \mid G = g)} \right).
\]

(32)

where \(\Delta = P(G = g' \mid X = x \mid G = g) / P(G = g' \mid X = x)\) is the ratio between the joint probability of the two graphs with the time series \(X\). This ratio can be computed efficiently in \(O(T)\), by keeping in memory \(n_{\text{in}}\), the number of inactive neighbors, and \(n_{\text{out}}\), a number of active neighbors, for each vertex \(i\) at each time \(t\) (see ref. 40). Equation (32) allows to sample from the posterior distribution \(P(G \mid X)\) without the requirement to compute the intractable normalization constant \(P(X)\). We collect graph samples at every \(N\) move, where we fix \(δ = 5\) in all experiments.

We consider two types of random graphs with different constraints: The Erdős-Rényi model and the configuration model. Hence, we need two different sampling algorithms to apply our MCMC algorithm, that is one for each model. We assume that the support of the Erdős-Rényi model is the set of all simple graphs of \(N\) vertices with \(E\) edges. In this case, we consider a hinge flip move, where an edge \((i,j)\) is sampled uniformly from the edge set of the graph \(G\) and a vertex \(k\) is sampled uniformly from its vertex set. Then, with probability \(\frac{1}{2}\) we rewire edge \((i,j)\) by either selecting \(i\) or \(j\) to connect with \(k\). Note that, because we consider the support \(G\) of \(G\) to be a space of simple graphs, all moves resulting in the addition of a self-loop or a multiedges are rejected with probability 1. As a result, the proposition probability is the same for any move from \(g\) to \(g'\):

\[
P(G = g' \mid G = g) = \frac{1}{2N} \Rightarrow \frac{P(G = g' \mid G = g)}{P(G = g' \mid G = g')} = 1.
\]

(33)

For the configuration model, we assume that the support is the set of all loopy multigraphs of \(N\) vertices whose degree sequence is \(k\). In this case, we propose double-edge swap moves according to the prescription of ref. 110. We refer to it for further details.

Real networks

In this section, we present the real networks used in the bottom panels of Fig. 6. The networks have been downloaded from the Netzschleuder network catalog\(^{111}\).

Little Rock Lake food web. The Little Rock Lake food web\(^{95}\) is composed of \(N = 183\) nodes and \(M = 2,494\) edges, where nodes represent taxa (like species) found in Little Rock Lake in Wisconsin, and edges represent feeding patterns between two taxa. As presented in ref. 95, this network is directed, but for the purpose of our paper we reciprocated all edges. Also, note that the Glauber dynamics, which we used jointly with the Little Rock Lake food web in Fig. 6, was also used in ref. 40 to simulate a simplified interaction between the taxa.

European airline route network. The European airline route network\(^{96}\) is a multiplex network composed of \(N = 450\) and \(M = 3,588\) edges, where nodes represent airports and edges are routes between them. These edges have different types, encoding the different airlines. In our paper, we do not make any distinction between the edge types for simplicity.

C. Elegans neural network. The C. Elegans neural network\(^{97}\) used in Fig. 6 is an undirected network of \(N = 514\) and \(M = 2,363\) edges representing the neural network of male C. Elegans worms. The nodes are neurons and edges represent when there are gap junctions between neurons.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The real network data used in the paper were downloaded from the network repository Netzschleuder\(^{111}\). The details are given in “Real networks” section.

Code availability

The Python code used to generate the results of the paper is available on GitHub\(^{112}\).

References

1. Barabási, A.-L. Network science. Phil. Trans. R. Soc. A 371, 20120375 (2013).
2. Latora, V., Nicosia, V. and Russo, G. Complex Networks: Principles, Methods and Applications (Cambridge Univ. Press, 2017).
3. Newman, M. E. J. Networks 2nd edn (Oxford Univ. Press, 2018).
4. Barzel, B. & Barabási, A.-L. Universality in network dynamics. Nat. Phys. 9, 673–681 (2013).
5. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. Rev. Mod. Phys. 87, 925 (2015).
6. Boccaletti, S. et al. Explosive transitions in complex networks’ structure and dynamics: percolation and synchronization. Phys. Rep. 660, 1–94 (2016).
7. Iacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. Nat. Commun. 10, 2485 (2019).
8. Hébert-Dufresne, L., Scarpino, S. V. & Young, J.-G. Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement. Nat. Phys. 16, 426–431 (2020).
9. Murphy, C., Laurence, E. & Allard, A. Deep learning of contagion dynamics on complex networks. Nat. Commun. 12, 4720 (2021).
29. Laurence, E., Murphy, C., St-Onge, G., Roy-Pomerleau, X. and McCabe, S. et al. netrd: A library for network reconstruction and analysis.

30. Pietras, B. and Daffertshofer, A. Network dynamics of coupled oscillators: B. Phase-reduction techniques. Phys. Rep. 819, 1–109 (2019).

31. Thibeault, V., St-Onge, G., Dubé, L. J. and Desrosiers, P. Threefold way to the dimension reduction of dynamics on networks: an application to synchronization. Phys. Rev. Res. 2, 043215 (2020).

32. Thibeault, V., Allard, A. and Desrosiers, P. The low-rank hypothesis of complex systems. Nat. Phys. https://doi.org/10.1038/s41567-023-02303-0 (2024).

33. Pastor-Satorras, R. and Vespignani, A. Epidemic spreading in scale-free networks. Phys. Rev. Lett. 86, 3200 (2001).

34. Hébert-Dufresne, L. and Althouse, B. M. Complex dynamics of synergistic coinfections on realistically clustered networks. Proc. Natl. Acad. Sci. USA 112, 10551–10556 (2015).

35. St-Onge, G., Young, J.-G., Laurence, E., Murphy, C. and Dubé, L. Phase transition of the susceptible-infected-susceptible dynamics on time-varying configuration model networks. Phys. Rev. E 97, 022305 (2018).

36. St-Onge, G., Thibeault, V., Allard, A., Dubé, L. J. and Hébert-Dufresne, L. Master equation analysis of mesoscopic localization in contagion dynamics on higher-order networks. Phys. Rev. E 103, 032301 (2021).

37. St-Onge, G., Sun, H., Allard, A., Hébert-Dufresne, L. and Bianconi, G. Universal nonlinear infection kernel from heterogeneous exposure on higher-order networks. Phys. Rev. Lett. 127, 158301 (2021).

38. Ferreira, S. C., Castellano, C. and Pastor-Satorras, R. Epidemic thresholds of the susceptible-infected-susceptible model on networks: a comparison of numerical and theoretical results. Phys. Rev. E 86, 041125 (2012).

39. Castellano, C. and Pastor-Satorras, R. Relating topological determinants of complex networks to their spectral properties: structural and dynamical effects. Phys. Rev. X 7, 041024 (2017).

40. Pastor-Satorras, R. and Castellano, C. Eigenvector localization in real networks and its implications for epidemic spreading. J. Stat. Phys. 173, 1110–1123 (2018).

41. Hébert-Dufresne, L., Noël, P.-A., Marceau, V., Allard, A. and Dubé, L. J. Propagation dynamics on networks featuring complex topologies. Phys. Rev. E 82, 036115 (2010).

42. St-Onge, G., Thibeault, V., Allard, A., Dubé, L. J. and Hébert-Dufresne, L. Social confinement and mesoscopic localization of epidemics on networks. Phys. Rev. Lett. 126, 098301 (2021).

43. Brugere, I., Gallagher, B. and Berger-Wolf, T. Y. Network structure inference, a survey: motivations, methods, and applications. ACM Comput. Surv. 51, 1–39 (2018).

44. Peixoto, T. P. Reconstructing networks with unknown and heterogeneous errors. Phys. Rev. X 8, 041011 (2018).

45. Young, J.-G., Cantwell, G. T. and Newman, M. E. J. Bayesian inference of network structure from unreliable data. J. Complex Netw. 8, cnaa046 (2020).

46. Young, J.-G., Valdovinos, F. S. and Newman, M. E. J. Reconstruction of plant–pollinator networks from observational data. Nat. Commun. 12, 3911 (2021).

47. Laurence, E., Murphy, C., St-Onge, G., Roy-Pomerleau, X. and Thibeault, V. Detecting structural perturbations from time series using deep learning. http://arxiv.org/abs/2006.05232 (2020).

48. McCabe, S. et al. netrd: A library for network reconstruction and graph distances. J. Open Source Softw. 6, 2990 (2021).

49. Kramer, M. A., Eden, U. T., Cash, S. S. and Kolaczyk, E. D. Network inference with confidence from multivariate time series. Phys. Rev. E 79, 061916 (2009).

50. Schreiber, T. Measuring information transfer. Phys. Rev. Lett. 85, 461–464 (2000).

51. Seth, A. K. Causal connectivity of evolved neural networks during behavior. Netw. Comput. Neural Syst. 16, 35–54 (2005).

52. Abbeel, P., Koller, D. and Ng, A. Y. Learning factor graphs in polynomial time and sample complexity. J. Mach. Learn. Res. 7, 1743–1788 (2006).

53. Salakhutdinov, R. and Murray, I. On the quantitative analysis of deep belief networks. In Proc. 25th International Conference on Machine Learning 872–879 (Association for Computing Machinery, 2008).

54. Bento, J. and Montanari, A. Which graphical models are difficult to learn? In Advances in Neural Information Processing Systems 1033–1031 (MIT Press, 2009).

55. Salakhutdinov, R. and Larochelle, H. Efficient learning of deep Boltzmann machines. In Proc. Thirteenth International Conference on Artificial Intelligence and Statistics 693–700 (Proceedings of Machine Learning Research, 2010).

56. Bresler, G., Mossel, E. and Sly, A. Reconstruction of Markov random fields from samples: some observations and algorithms. SIAM J. Comput. 42, 563–578 (2013).

57. Amin, M. H., Andriyash, E., Rolfe, J., Kulchytskyy, B. and Melko, R. Quantum Boltzmann machine. Phys. Rev. X 8, 021050 (2018).

58. Peixoto, T. P. Network reconstruction and community detection from dynamics. Phys. Rev. Lett. 123, 128301 (2019).

59. Hinne, M., Heskes, T., Beckmann, C. F. and Van Gerven, M. A. J. Bayesian inference of structural brain networks. NeuronImage 66, S43–552 (2013).

60. Breakspear, M. Dynamic models of large-scale brain activity. Nat. Neurosci. 20, 340–352 (2017).

61. Bassett, D. S., Zurn, P. and Gold, J. I. On the nature and use of models in network neuroscience. Nat. Rev. Neurosci. 19, 566 (2018).

62. Wang, Y., Joshi, T., Zhang, X.-S., Xu, D. and Chen, L. Inferring gene regulatory networks from multiple microarray datasets. Bioinformatics 22, 2434–2440 (2006).

63. Prasse, B., Achterberg, M. A., Ma, L. and Van Mieghem, P. Network-inference-based prediction of the COVID-19 epidemic outbreak in the Chinese province Hubei. Appl. Netw. Sci. 5, 35 (2020).

64. Musmeci, N., Battiston, S., Caldarelli, G., Pulliga, M. and Gabrielli, A. Bootstrapping topological properties and systemic risk of complex networks using the fitness model. J. Stat. Phys. 151, 720–734 (2013).

65. Bassett, D. S. and Sporns, O. Network neuroscience. Nat. Neurosci. 20, 353 (2017).

66. Sporns, O. Structure and function of complex brain networks. Dialogues Clin. Neurosci. 15, 247–262 (2013).

67. Fornito, A., Zalesky, A. and Breakspear, M. The connectomics of brain disorders. Nat. Rev. Neurosci. 16, 159–172 (2015).

68. Van den Heuvel, M. P. and Sporns, O. A cross-disorder connectome landscape of brain dysconnectivity. Nat. Rev. Neurosci. 20, 435–446 (2019).

69. Prasse, B. and Van Mieghem, P. Predicting network dynamics without requiring the knowledge of the interaction graph. Proc. Natl. Acad. Sci. USA 119, e2205517119 (2022).

70. Zhang, Z., Cui, P. and Zhu, W. Deep learning on graphs: A survey. AI Open 1, 57–81 (2020).

71. Fout, A., Byrd, J., Shariat, B. and Ben-Hur, A. Protein interface prediction using graph convolutional networks. In Advances in Neural Information Processing Systems, Vol. 32, 6530–6539 (MIT Press, 2017).

72. Zitnik, M., Agrawal, M. and Leskovec, J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics 34, i457–i466 (2018).
Acknowledgements
We are grateful to Guillaume St-Onge and Vincent Painchaud for useful
comments, and to Simon Lizotte and François Thibault for their help in
designing the software. This work was supported by the Fonds de
recherche du Québec – Nature et technologies (V.T.), the Conseil de
recherches en sciences naturelles et en génie du Canada (C.M., A.A.,
P.D.), the Sentinelie Nord program of Université Laval, funded by the
Fonds d’excellence en recherche Apogée Canada (C.M., A.A., P.D.), and
the Fonds d’accélération des collaboration en santé du Québec – Alli-
ance Neuro-CERVO (A.A., P.D.). We acknowledge Calcul Québec and
Digital Research Alliance of Canada for their technical support and
computing infrastructures.

Author contributions
C.M., V.T., A.A. and P.D. developed the framework and wrote the paper.
C.M., V.T. and P.D. wrote the mathematical proofs. C.M. performed the
numerical analysis.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains
supplementary material available at
https://doi.org/10.1038/s41467-024-48020-x.

Correspondence and requests for materials should be addressed to
Charles Murphy or Patrick Desrosiers.

Peer review information Nature Communications thanks Filippo Radic-
chi, and the other, anonymous, reviewers for their contribution to the
peer review of this work. A peer review file is available.

Reprints and permissions information is available at
http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jur-
isdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons
Attribution 4.0 International License, which permits use, sharing,
adaptation, distribution and reproduction in any medium or format, as
long as you give appropriate credit to the original author(s) and the
source, provide a link to the Creative Commons licence, and indicate if
changes were made. The images or other third party material in this
article are included in the article’s Creative Commons licence, unless
indicated otherwise in a credit line to the material. If material is not
included in the article’s Creative Commons licence and your intended
use is not permitted by statutory regulation or exceeds the permitted
use, you will need to obtain permission directly from the copyright
holder. To view a copy of this licence, visit http://creativecommons.org/
licenses/by/4.0/.

© The Author(s) 2024