Physica A 593 (2022) 126889



Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa



Dynamic importance of network nodes is poorly predicted by static structural features



Casper van Elteren [a,b,*], Rick Quax [a,b], Peter Sloot [b,c,d]

[a] Institute for Advanced Study, University of Amsterdam, Amsterdam, 1012 GC, The Netherlands
[b] Computational Science Lab, University of Amsterdam, Amsterdam, 1098 XH, The Netherlands
[c] Complexity Science Hub Vienna, Vienna, Austria
[d] National Center for Cognitive Research, ITMO University, Saint Petersburg, Russia

ARTICLE INFO

Article history:
Received 26 October 2020
Received in revised form 25 November 2021
Available online 21 January 2022

Keywords:
Complex system
Information theory
Driver-node identification

ABSTRACT

One of the most central questions in network science is: which nodes are most important? Often this question is answered using structural properties such as high connectedness or centrality in the network. However, static structural connectedness does not necessarily translate to dynamical importance. To demonstrate this, we simulate the kinetic Ising spin model on generated networks and one real-world weighted network. The dynamic impact of nodes is assessed by causally intervening on node state probabilities and measuring the effect on the systemic dynamics. The results show that structural features such as network centrality or connectedness are actually poor predictors of the dynamical impact of a node on the rest of the network. A solution is offered in the form of an information theoretical measure named integrated mutual information. The metric is able to accurately predict the dynamically most important node ("driver" node) in networks based on observational data of non-intervened dynamics. We conclude that the driver node(s) in networks are not necessarily the most well-connected or central nodes. Indeed, the common assumption of network structural features being proportional to dynamical importance is false. Consequently, great care should be taken when deriving dynamical importance from network data alone. These results highlight the need for novel inference methods that take both structure and dynamics into account.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Understanding dynamical systems is a fundamental problem for the 21st century [1]. Despite the apparent differences and purposes of many real-world networked complex systems, previous research shows several universal characteristics in networks properties such as the small-world phenomenon [2], fat-tail degree [3], and feedback loops [4]. This has lead to the common but often implicit assumption that the connectedness of a node in the network is proportional to its dynamic importance [5]. For example in epidemic research, high degree nodes or "super-spreaders" are associated to dominant epidemic risk and therefore deserve special attention [6]. Yet, prior research shows that the shared network characteristics is not shared in the dynamic or functional properties that are exerted on these networks [7,8]. In particular, the dynamic importance of a node varies as a function of both the dynamics that exist on the network in addition to its structural connectedness. This effect was rigorously shown by Harush and colleagues [8]. In their study, the dynamic processes were varied while keeping the network structure the same. Both random generated networks as well as real-world networks

* Corresponding author.
E-mail addresses: c.vanelteren@uva.nl (C. van Elteren), r.quax@uva.nl (R. Quax), p.m.a.sloot@uva.nl (P. Sloot).

https://doi.org/10.1016/j.physa.2022.126889
0378-4371/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

were studied. Nodal importance was computed based on the "information flow" through a node. The flow through a node had a non-linear relation between its structural connectedness and the type of dynamics present in the system. These results highlight the non-intuitive and non-trivial interplay of the structure of a system and dynamics between the nodes of a system. However, the results required full knowledge of the system, i.e. both the network structure and the dynamics of the system are required to estimate the node with the highest dynamic importance. For a real-world system having both the dynamics available and the underlying network structure may be difficult. In addition, in their study the network structure was deemed constant. It remains unclear whether the proposed flow would generalize to systems with varying network structure. In addition, the results were obtained from steady-state. For dynamical processes an out-of-equilibrium approach highlight *how* trajectories generate systemic behavior. What is needed are methods that can detect the highest dynamic importance of a node from observations directly without knowing or assuming the underlying network or causal structure among variables.

A node's dynamic (or causal) importance, is traditionally inferred by means of interventions and counterfactuals [9,10]. Through external interventions the behavior of the system may change. That is, an intervention on a part of the system causes a divergence of the system behavior proportional to its dynamic importance; high causal importance relates to a large change in the system behavior. It is common for studies using causal interventions apply so-called hard intervention strategies [11]. In hard interventions a node is pinned to a state. This effectively removes all incoming connections to a node, removing all influences the system has on this node. The use of hard interventions is common in various disciplines such as gene-knockout experiments, epidemic spreading, network analysis and driver node identification.

Applying external interventions to identify nodes with high dynamical importance, referred to as driver nodes, is challenging for three reasons. First, hard interventions may cause a change in systemic behavior that does not occur in the non-intervened behavior (e.g. see Fig. 1B) or may be impossible in practice (e.g. requiring infinite resources [12]). This complicates theory forming on what mechanisms underlie observed systemic behavior. Second, most approaches for causal inference assume that the causal interactions follow a particular structure without (local) loops [13–15]. This assumption simplifies theoretical analysis and is in many cases justified under the assumption that causes precede their effects in time. However, this approach assumes that the underlying causal structure is known and or accessible. If the underlying causal structure is not known, it has to be inferred from data which prompts problems in terms of temporal resolution and scale. For example in a causal process where $A \rightarrow B \rightarrow C \rightarrow A$, the cycle may be unrolled over time. This effectively removes cycles. However, this requires the data to support time resolution such that it reflects unrolled (non-cyclical) causal processes [13]. Cyclical causal structures do occur in real-world systems such as ecosystems, biological systems, gene-regulatory systems and so on. Importantly, for many of these systems determining the correct temporal scale for causal interactions is non-trivial. Consequently, it remains an open question on how to perform causal inference applying these in general for cyclical causal events. Thirdly, the underlying causal structure may not be known or difficult to determine [16]. In addition, many dynamical systems are prohibited from analytical approaches to decompose each nodal dynamical importance directly due to the polyadic, often non-linear, interactions [17].

For closed systems, it is possible to avoid these challenges to deduce the driver node by considering cross-sectional time series and computing the correlation of each node with the entire system out-of-equilibrium and without lossy compression [18]. Here, a driver node is the element of the system that is correlated the most with the entire system as a function of time. It has the maximum dynamic (causal) impact when intervened upon across all nodes. By taking the entire system into account, any internal confounding information is implicitly accounted for; there cannot exist any other node with more dynamic importance than the node with the highest correlation as the system is closed and external confounding is excluded.

Shannon information theory offers profound advantages over previous approaches for defining a measure of dynamical importance of driver node on the behavior of all other nodes [19]. Firstly and most prominently, Shannon mutual information can quantify statistical associations among variables without bias to specific forms of association [20]. In particular, it equals zero if and only if the full probability distribution of the system state remains exactly unchanged regardless of the state of the node. Second, mutual information does not require *a priori* knowledge of the representational base of the system. That is, it allows for direct comparison among different systems that may have different units of measurement such as currency, density of animals, voltage per surface area and so on. Finally, it is defined for both discrete-valued and real-valued state variables.

The concept of measuring the dynamic importance of a node through information flow is not new. Colloquially, information flows from process X to process Y represents the existence of statistical coherence between the present information in Y and the past of X not accounted for by the past of Y. Various methods have been proposed in the past such as transfer entropy [21] and its derivatives [22–25]. Although originally intended as a predictive measure, the notions of information and information flow can be extended to causal influence or dynamic importance. Previous research developed several measures and methods for determining how much information flow between two processes is truly causal; examples include (but not limited to) conditional mutual information under causal intervention [26], causation entropy [27], permutation conditional mutual information [28], time-delayed Shannon mutual information [29].

These measures are commonly used to infer the information transfer between sets of nodes by possibly correcting for a third confounding variable [11,30]. That is, informational flows are used to determine how information is transferred or shared between pair or sets of variables. However, in polyadic settings most measures of information flow are prone to underestimate or overestimate nodal importance [31]. Determining how much information flow is causal between source and sink variables in polyadic settings remains difficult due to the so-called synergetic and redundant information [31,32].

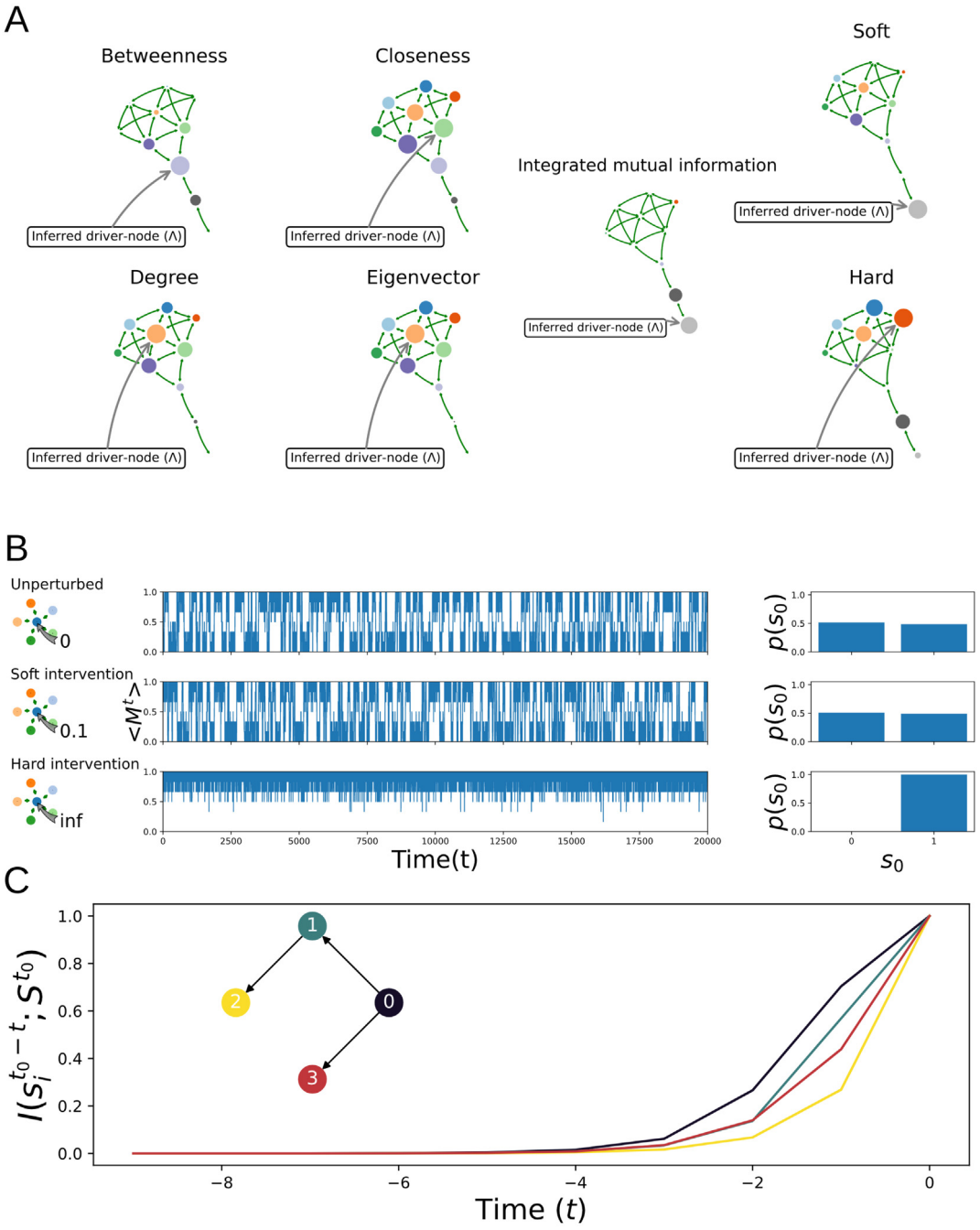


Fig. 1. (A) Driver node inference. Dynamics are simulated using kinetic Ising spin dynamics. Causal interventions (soft and hard) are shown in the left column. Structural metrics (betweenness, closeness, degree, eigenvector centrality) each produce different driver node estimates. Integrated mutual information predicts the driver node for soft causal interventions. High causal interventions produce different system dynamics (see B for an example) and different driver nodes. This figure show that dynamics interact with structure to produce non-trivial driver node estimates. In addition, causal intervention size (hard or soft) influence the observed system dynamics. (B) Effect of intervention size on system magnetization $\langle M^t \rangle = \frac{1}{n}\sum_i s_i^t$ with kinetic Ising spin dynamics (see 2.1). Energy is added (gray arrow) to the nodal Hamiltonian of the blue node (s_0). (top) Non-Intervened system dynamics are shown with the distribution of the blue node s_0 (right). (middle) Soft intervention on node s_0 keeps similar dynamics as the non-intervened system dynamics. (bottom) Hard interventions yield profound different system dynamics compared to non-intervened dynamics. (C) Example of non-causal inflation of mutual information decay. The network structure is given by the graph inset; consisting of a system with 4 nodes in which each node is binary variable. Node 0 has a 50/50 distribution and all other nodes copy the state of its predecessor. Node 0 has the largest causal influence as it can influence the state of all other nodes in the system (over time). The information content of node 3 is biased; it stores the information from 0, but has no downstream causal effects. Yet, its information decay is similar to 1 which has one downstream node. The true driver node (0) has the largest information decay over time and is not confounded.

Instead of focusing on the open problem of full information decomposition among variables [11,21,26,27,30], we focus here on the amount of information that a node shares with the entire system. A driver node is expected to have a corresponding high information flow from it to the system. We introduce a novel metric named integrated mutual information (IMI) based on time-delayed Shannon mutual information with a node and the entire system over time that captures driver nodes with the highest *causal* impact in ergodic systems. Additionally, we avoid the problem of confounding information synergy with redundancy by avoiding conditional mutual information. The consequence of this is that we quantify only *how much* causal impact *one single node* has on the *entire* network. In other words, we will not be able to infer which parts of the network are impacted more than others; nor how exactly the causal influence percolates through the network. This is nevertheless sufficient for the purpose of this study.

Using this approach it was previously shown analytically that the number of connections of a node does necessarily not scale monotonically with its dynamical importance in infinite-sized, locally-tree-like networks (i.e. networks without loops) [33]. For those networks, nodes with high dynamic importance were not nodes with high degree (so-called hubs). Instead, the nodes with intermediate connectedness were identified with the highest dynamical importance. This study extends this prior research by numerically computing information flow in finite-sized, random networks, i.e. where small feedback loops cannot be ignored, and studies the relation between connectivity and the dynamical importance of nodes in such networks.

The aim of this paper is to test the common (often implicit) hypothesis that the connectedness of a node is proportional to its dynamical importance. A node's dynamic importance is determined by simulating the (out-of-equilibrium) dynamics of a system under causal external interventions. The distance between the system state probability distributions with *and* without performing the intervention is used as a ground truth for a node's causal influence over time (see 2.2). The resulting causal impact score over time is compared with common network centrality metrics (**appendix** A.6) as well as our proposed information-based metric.

In contrast to other studies, the proposed metrics in this study (Section 2) does not make assumptions on network structure or type of dynamics. Small networks are used (up to 12 nodes) for which each node has an associated discrete state whose dynamics is governed by a stochastic update rule in discrete time. Smaller network sizes have the of studying "network motifs" that are embedded in larger network structure [34]; understanding causal flows of smaller structures may provide insights into larger systems consisting of a combination of smaller network structures. In addition, the use of small network sizes offers a numerical advantage for accurate estimation of causal and information measures. In total 16 Erdös–Rényi networks are generated (10 nodes), and one real-world weighted network (12 nodes) obtained from [35]. Temporal dynamics of the nodes are simulated using kinetic Ising spin dynamics with Glauber updating for the purpose of demonstrating a case of non-trivial relation between network connectivity and dynamic importance, but our approach easily generalizes to any other dynamics or networks generation model.

The results show that nodes with generally nodes with high structural connectedness as measured by closeness, betweenness, eigenvector and degree centrality are not the *driver node(s)*. Our novel metric, IMI, achieved significantly better performance in predicting the driver node for non-intervened dynamical systems. In addition and most importantly, hard causal interventions lead to causal flows that differ from the non-intervened dynamics. That is, as a function of external intervention, systemic behavior may result in "unnatural" system behavior that do not occur in the unperturbed system. The proposed metric IMI does not rely on the assumptions on dynamics, nor on assumptions on structural properties of the network. Therefore, the results of this study provide scientists of all fields a novel, reliable and accurate metric for the identification of driver nodes.

2. Theoretical background

2.1. Terminology

In this paper, we consider a complex system as a set of discrete random variables $S = \{s_1, s_2, \ldots, s_n\}$ with interaction structure $E = \{(s_i, s_k)|s_i, s_k \in S\}$, where each $s_i \in S$ has an alphabet A. This is also known as a (discrete) dynamical network [36]. Please note that we use the term node and variable interchangeably referring to an element $s_i \in S$. The system chooses its next state S^t in discrete time with probability:

$$p(S^t|S^{t-1}, \ldots, S^{t_0}) = p(S^t|S^{t-1}), \tag{1}$$

which is also known as a first-order Markov chain. More specifically, each discrete time step, a single variable $s_i \in S$ is chosen with uniform probability and updated. That is,

$$p(S^t|S^{t-1}, \ldots, S^{t_0}) = p(S^t|S^{t-1}) = \prod_j p(s_j^t|S^{t-1}) = p(s_i^t|S^{t-1}). \tag{2}$$

For temporal dynamics, we adopt here the Metropolis–Hasting algorithm [37]. By drawing a proposal state $S^{t+1} = X'$ from current state $S^t = X$ from a proposal distribution $g(X'|X)$ and accepting the new state X' with probability,

$$\begin{aligned} A(S^{t+1} = X', S^t = X) &= \min(1, \frac{p(S^{t+1} = X')g(S^{t+1} = X|S^t = X')}{p(S^t = X)g(S^{t+1} = X'|S^t = X)}) \\ &= \min(1, \frac{p(s_i^{t+1} = x')g(s_i^t = x|s_i^t = x')}{p(s_i^{t+1} = x)g(s_i^t = x'|s_i^t = x)}), \end{aligned} \tag{3}$$

where $x' \in A$. If the new state is not accepted, then the next state will be set to $S^{t+1} = X$. As the next state S^{t+1} is determined through considering updating a single variable $s_i \in S$, the next state X' will be generated through the proposal state x' drawn uniformly from the possible states A such that $g(x'|x) = g(x|x') = g(x') = \frac{1}{|A|}$. This means that $\frac{g(X'|X)}{g(X|X')} = \frac{g(x'|x)}{g(x|x')} = 1$.

For our experiments, we use the kinetic Ising model which is thought to fall in the same universality class as various other complex behaviors [38], such as (directed) percolation, diffusion, and many extensions of the model have been used as a base for opinion dynamics, modeling neural behavior and so on. For demonstrating our primary claim that high network connectivity does not necessarily lead to high dynamical importance, a single dynamics suffices as counterexample.

It is nevertheless important to emphasize that our proposed driver node inference method does not depend on the exact type of dynamics. That is, the choice for the kinetic Ising spin dynamics is arbitrary in that respect. Our methods only require that for a given dynamics the data-processing inequality is satisfied. More details on the methods and its assumption will follow in 2.3.

The kinetic Ising model consists of binary variables dictated by a Gibbs distribution that interact through nearest neighbor interactions. A prominent property of the Ising model in higher dimensions (two or more) is the phase transition from an ordered phase to a disordered phase by increasing the noise parameter β. For finite systems, the kinetic Ising model shows a continuous phase transition from ordered to unordered system regime (Fig. 2B). The Metropolis–Hastings update rule specifically for the kinetic Ising model equals:

$$p(\text{accept } X') = \frac{p(X')}{p(X)} = \begin{cases} 1 & \text{if } \mathcal{H}(X') - \mathcal{H}(X) < 0 \\ \exp(-\beta(\mathcal{H}(X')) - \mathcal{H}(X))) & \text{otherwise,} \end{cases} \tag{4}$$

where $\mathcal{H}(S)$ is the system Hamiltonian defined as

$$\mathcal{H}(S) = -\sum_{i,j} J_{ij} s_i s_j - h_i s_i. \tag{5}$$

Here β is the inverse temperature $\frac{1}{k_b T}$ with Boltzmann constant k_b, J_{ij} is the interaction strength between variables s_i and s_j; h_i represents external influence on node i. The matrix J effectively represents the network of the system: an edge between $s_i, s_j \in S$ exists if $|J_{ij}| > 0$. The networks considered in this study are undirected networks, which means that J is an $|S| \times |S|$ symmetric matrix. For the randomly generated networks the edges $J_{ij} \in \{0, 1\}$; for the real network data set the edge weights are positive and negative real numbers (Section 3).

The β parameter can be seen as the (inverse) noise parameter in the system. Low values of β will induce each node in the system to detach from the influence of its neighbors, i.e. the probability of finding a node in a state $p(s_i = a), a \in A$ will tend to uniform distribution as $\beta \to 0$. In contrast, high values of β increases the influence a neighbor of a node may have on determining the node's next state.

2.2. Causal interventions and dynamic importance

We call a node a *driver node* for a dynamical system when it has the largest causal impact on the system dynamics. In brief, we will determine a node's causal impact by simulating a transient intervention and subsequently quantifying the difference between the system dynamics under intervention and without intervention. We expect that the impact of the driver nodes will penetrate deeper into the system and remain present longer than for nodes with lower causal impact. Here, we define causal impact by means of external intervention on a node. The external causal intervention $\vec{\epsilon}$ on node $s_j \in S$ can be described as

$$p'_{s_j}(s_i^t | S^{t-1}) = p(s_i^t | S^{t-1}) + \vec{\epsilon}\delta_{ij}, \tag{6}$$

where δ_{ij} is the Kronecker-delta, and $\dim(\vec{\epsilon}) = |A|$. In addition, $\sum_{i=0}^{|A|} \epsilon_i = 0$ and $\sum_{i=0}^{|A|} |\epsilon_i| = c$ for some $c \in (0, 1]$. Note only those $\vec{\epsilon}$ are allowed that generate valid new probabilities, i.e. $0 \leq p' \leq 1$.

Relative to some equilibrium distribution $p(S^\tau)$, the effect of intervention $\vec{\epsilon}$ will result in a new system state equilibrium distribution $p'(S^\tau)$. Subsequently the intervention is removed at a random system state, after which the distribution of system states will gradually converge back to the original equilibrium. Nodes with higher dynamic importance will cause a larger difference in the system state probability distribution over time. Consequently, we quantify the causal impact of a node by integrating over time t the difference in system state distribution from the moment the intervention is released ($t = \tau$). Since our model is discrete in time, the integral becomes a summation. Thus, we define the causal impact of node $s_i \in S$ as

$$\begin{aligned} \Gamma(s_i) &= \sum_{t=\tau}^{\infty} \gamma(s_i^t) \Delta t \\ &= \sum_{t=\tau}^{\infty} \sum_{s_j \in S} D_{KL}(p'_{s_i}(s_j^t) \parallel p(s_j^t)) \Delta t \end{aligned} \tag{7}$$

where D_{KL} is the KL-divergence, and $\Delta t = 1$ throughout the paper. KL-divergence $D_{KL}(p \parallel q)$ quantifies the difference between probability distributions and is non-negative, invariant under affine parameter transformation and zero when $p = q$. For example, if $\Gamma(s_i) = 0$ the intervention on nodes i caused no difference in any state probabilities.

The driver node can then readily be defined as $\text{argmax}_{s_i \in S}(\Gamma(s_i))$. A numerical implementation of driver node identification is given in Section 3.2.6.

In this study we allow the intervention to evolve for some time period $\Delta t_{\text{nudge}} = \tau - t_0$ after which the nudge is removed (see 3.2.6 and Fig. 2 C). The intervention will transiently bring the system out of equilibrium. For ergodic systems, the effect of the intervention will be lost from the system over time. Namely, $P'_{s_i}(S)$, $\forall s_i$ will tend to equilibrium distribution $P(S)$ from τ on wards. The causal impact is computed relative to this τ (Fig. 2 C). As a consequence $\Gamma(s_i)$ will be finite for systems under study here, but may diverge for non-ergodic systems. The duration τ needs to be set appropriately that the intervention is allowed to percolate through the system. Here we used $\tau = 15$ for random networks consisting of 10 nodes and $\tau = 25$ for the real-world weighted network consisting of 12 nodes (see Section 3.1).

The time prior to τ is used for computing the causal impact on a node as the intervention could be disproportional affected by the act of changing the node distribution. It does not accurately reflect the causal impact of the node on the rest of the system. Only the decay after τ would be proportional to the causal impact a node has on all other nodes. The causal impact is therefore computed based on the decay of the causal impact relative to τ.

2.2.1. Intervention size

In experiments concerned with measuring causal flows in networks, often "hard" causal interventions are used to determine the causal impact of nodes [39–42]. Hard causal interventions are those that (effectively) remove all inputs from a node. In contrast, soft interventions, keep the existing causal inputs intact but add an additional causal effect to the dynamics of a variable. In general, the larger the added causal effect of the soft intervention, the more this intervention will 'overpower' the existing causal effects and hence the more the soft intervention converges to a hard intervention. In non-linear dynamical systems the intervention size is crucial, since even small interventions can have large effects, especially in the presence of bifurcations. In order to characterize the system without intervention as much as possible, one should thus aim to intervene with a minimal (but still measurable) effect. To illustrate this further, consider the system depicted in Fig. 1B, where each node is update with Eq. (4). The figure shows the effect of intervention size on the observed system dynamics. Namely, when a hard intervention is used (bottom) the system magnetizes. In contrast, in the non-intervened system, the system magnetizes periodically, i.e. there system dynamics evolve with time periods of magnetization and metastable switches to the other side of the magnetization. Contrasting the non-intervened system (top plot) with the bottom plot (hard) intervention, shows how the metastable behavior disappears. Soft interventions (middle plot) maintain the meta-stable behavior of the system.

There will always be a minimal intervention size for which there is a measurable (i.e. non-zero) causal effect (Eq. (7)), given finite amount of data. The causal impact will be maximal for hard interventions. As the intervention size is decreased so will causal impact. We hypothesize that there exists a lower bound for the soft interventions for which there will be a measurable causal impact. This minimal intervention size is dependent on the systems structure as well as the dynamics of the system and is difficult to determine *a priori*. Yet it is important to approach this minimum because if the intervention size is too large, then the intervened system dynamics will diverge from the non-intervened system dynamics, losing its representative capacity. Therefore, we estimate this minimum numerically, described further in Section 3.2.

2.2.2. Intervention in kinetic ising model

In order to measure the causal intervention, transition probability (6) needs to be adjusted. For the kinetic Ising model, the transition probability can be modified by adding *energy* to the node which results in the adjusted Hamiltonian (Eq. (5))

$$\mathcal{H}_{\text{nudge } s_k}(S) = -\sum_{i,j} J_{ij} s_i s_j - \sum_i h_i s_i - \eta s_i \delta_{ik}. \tag{8}$$

That is, an energy term is added to the intervened node s_k by letting it interact with an external 'spin state' η. Both soft and hard interventions were used (see Section 2.2.1). For hard interventions ($\eta \to \infty$) the nodal dynamics of the nudge node s_k are no longer influenced by nearest neighbor interaction. That is, the node state for the intervened node s_j will not change as a function of time. For soft interventions a grid-search was used with $\vec{\eta} := \{\eta : \eta = 0.55 + i, i \in \{1, \ldots, 10\}\}$ to determine a minimal yet measurable intervention strength. Once determined the same η was applied to each node in turn and this process is repeated for all networks in the experiments.

2.3. Measuring information flow

Each node in a dynamical system can be considered as an information storage unit [32,43]. For example in social networks gossip can be considered as information one person possesses. Similarly, disease can be present in one city while being absent in another. Over time through interaction, this information stored in a node will percolate throughout the system while at the same time decaying due to noise. The longer the information of a node stays in the system, the

longer it could affect the system dynamics. Therefore, dynamic impact of a node is upper-bounded by the amount of information a node shares with the entire system [32,33,44].

How does one measure information stored in a node? A node s_i dictated by stochastic and ergodic dynamics can be considered a random variable. In Shannon information theory information is quantified in bits, i.e. yes/no questions concerning the outcome of a random variable. The average information that a random variable can encode is called entropy and is defined as:

$$H(s_i) = -\sum_{s_i=x} p(x)\log p(x). \tag{9}$$

Note all log are base 2 in this paper unless specified otherwise.

Entropy can also be interpreted as the amount of uncertainty of a random variable. In the extremes the random variable either conveys no uncertainty (i.e. a node always assumes the same state), or is randomly chosen between all possible states (uniform distribution). For example consider a coin flip. One may ask how much information does a single coin flip encode? If the coin is fair, i.e. there is equal probability of the outcome being heads or tails, the amount of questions needed to determine the outcome is exactly 1. In other words, a fair coin encodes 1 bit of information. However, when the coin is unfair the information encoded is less than one. In the extreme case where the coin always turns up heads, the entropy is exactly 0.

The information shared between a node state s_i and a system state S can be quantified by mutual information [19,32,33,44,45]. Mutual information can be informally thought of as a non-linear correlation function which inherits its properties from the Kullback–Leibler divergence. Formally, mutual information quantifies the reduction in uncertainty of random variable X by knowing the outcome of random variable Y [19]:

$$\begin{aligned} I(X:Y) &= \sum_{x\in X, y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= H(X) - H(X|Y), \end{aligned} \tag{10}$$

where $p(x)$ and $p(y)$ are the marginals of $p(x,y)$ over X and Y respectively, and $H(X|Y)$ is the conditional entropy of X. The conditional entropy $H(X|Y)$ is similar to the entropy; it quantifies the reduction in uncertainty of the outcome X by knowing the outcome of Y. Please note that the yes/no question interpretation even applies to continuous variables; although it may take an infinite amount of questions to determine the outcome of a continuous random variable.

2.3.1. Mutual information and causality

Under some contexts, mutual information can be interpreted as being causal. For example for a Markovian, isolated system consisting of two nodes $S = \{s_i, s_j\}$ where s_j causally depends on the previous state of s_i, encoded by a conditional probability distribution. It can be readily shown that the causal impact (7) of s_i on s_j then reduces to mutual information [19], A.2. In this study, the entire system is considered as a *downstream* node. Importantly, the system is assumed to be measured in its entirety: no confounding variables are allowed. Within this context, the node with the highest integrated mutual information contains more causal information than correlation information than any other node in the system. Since there cannot exist any other node with more mutual information over time by construction of the system, this node must be the driver node. For any other node the measured mutual information can be (significantly) conflated (1C). Note that for directed graphs, these methods will not work as by construction the highest mutual information may be lead to wrongful identification of the driver node (see A.2).

Additionally, for directed graphs the arrow of time causes a different interpretation on what correlation with the driver node means in cases without unobserved variables (see A.2). For $t_0 + t$ where $t > 0$, integrated mutual information may correspond to correlation information of a driver node sender, whereas for $t < 0$, the mutual information relates to the correlation with a driver node sending information. The systems in this study, however, are systems without confounding possessing and undirected network structure. Consequently, information flows symmetrically across the interaction of two nodes and the driver node corresponds to the node with the highest integrated information.

2.3.2. Integrated mutual information

In a network of nodes each causal relation (edge) is obviously not isolated, so confounding variables exist. Therefore, the mutual information between a node state and a future state $I(S^{t_0-t}; s_i^{t_0})$ cannot be interpreted purely causal in general. Namely, this mutual information could in principle be created purely by another variable $s_j^{t_0-\omega}$ influencing both $s_i^{t_0}$ and S^{t_0-t}, even if there exists no causal influence from s^{t_0} to any other variable.

Assuming the network itself is isolated, the only node for which the mutual information with a future system state could not have been fully created by a confounding variable is the driver node. That is, the driver node has the largest mutual information with the future system state. If this were fully induced by a confounding variable, then a different node would have to have even larger mutual information with the same future system state. In addition, this must hold for all future system states. By definition of the driver node, this cannot be true under the condition that the system isolated.

This point is illustrated in 1 C. Note that for all other nodes, however, it is possible for its mutual information value to be inflated due to non-causal correlations. This may result to a non-zero mutual information $I(s_i^{t_0-t} : s_j^t)$ among the

two variables even if they do not depend on each other in causal manner (Fig. 1). Consequently, we define the integrated mutual information (IMI) for node (s_i) as the driver node, i.e. the node with the largest causal impact over time. The corresponding information for this driver node integrated over time ($\sum_{t=t_0}^{\infty} I(s_i^{t_0+t}; S^{t_0})\Delta t$) is causal in closed systems lacking confounding variables (Fig. 1C), where S^{t_0} is the system state at some time t_0, and $s_i^{t_0-t}$ is the state of a node t away from that system state. At time $t = 0$ the value equals $I(s_i^{t_0}; S^{t_0}) = H(s_i^t)$ for any node.

Here, undirected networks are considered. Due to detailed balance for undirected networks there exists a time symmetry in terms of variable dynamics. This means that for the systems considered in this study $I(s_i^{t_0-t}; S_0^t) = I(s_i^{t_0+t}; S_0^t)$ (see Appendix A.2.2). It is computationally easier to compute $I(s_i^{t_0+t}; S_0^t)$ rather than the reverse (see section. 3.2.2). For directed graphs, however, the meaning and interpretation of integrated mutual information changes depending on the direction in time it is computed (see Appendix A.2.2). This effect is outside the scope of the present study and will be the subject of future studies. For undirected graphs, the causal impact are time invariant and is equal forward and backward in time.

$$\mu(s_i) = \sum_{t=t_0}^{\infty} I(s_i^{t_0-t}; S^{t_0})\Delta t, \tag{11}$$

For all ergodic Markovian systems the delayed mutual information $I(s_i^{t_0-t}; S^{t_0})$ will *always* decay to zero as $t \to \infty$ [19,33]. This decay is monotonic, which follows from the data-processing inequality [19] and states that information can never increase in Markov chains without external information injection (Appendix A.1). The question is *how fast this decay takes place for each node* (Fig. 1C), and consequently how much *informational impact* the node will have on the system.

Taking the entire system as a downstream 'node', $I(S^{t+1}; s_i^t)$ represents pure causal information flow for the driver node as long as there are no confounding variables. For the rest of the study we will therefore use the notation $I(s_i^{t_0+t}; S^{t_0})$.

3. Methods and network data

3.1. Network data

3.1.1. Random networks

In total 16 Erdös–Rényi random networks were generated consisting of 10 nodes each. Each random network is generated by first drawing a random connection probability uniformly from [0,1], and then creating each possible undirected edge with probability r. Out of these 16 networks, ∼82 percent had a single connected component (Fig. 2). Each edge had unitary weight.

3.1.2. Real-world network: psychosymptoms

In addition to the generated random networks, we also test a small weighted network inferred from real data. This network differs from the random networks in that it is weighted and reflecting interactions among variables as a consequence of a real-world process, as well as reflecting inferred interactions from real data. The network data originates from the Changing Lives of Older Couples (CLOC) and compared depressive symptoms assessed via the 11-item Center for Epidemiologic Studies Depression Scale (CES-D) among those who lost their partner (N=241) with still-married control group (N=274) [35]. Each of the CES-D items were binarized with the aid of a causal search algorithm using Ising model developed by [46] and represented as a node with weighted connections (Fig. 5D). For more info on the procedure see [35,46,47]. The 11 CES-D items are (abbreviated names used in the remainder of this text in brackets): 'I felt depressed' (depr), 'I felt that everything I did was an effort' (effort), 'My sleep was restless' (sleep), 'I was happy' (happy)', 'I felt lonely' (lonely), 'People were unfriendly' (unfr), 'I enjoyed life' (enjoy), 'My appetite was poor' (appet), 'I felt sad' (sad), 'I felt that people disliked me' (dislike), and 'I could not get going' (getgo).

3.2. Numerical methods

3.2.1. Magnetization matching

A prominent feature of the (kinetic) Ising model is the phase change that occurs as a function of noise (Fig. 2B) [48]. In this paper we tested whether the amount of noise would (i) affect which nodes becomes the driver node in the system, and (ii) whether the correct driver node could be predicted using either IMI or network centrality metrics. We tested three levels of noise (Fig. 2B); a low noise level (80% of the maximum magnetization), a medium noise level (70%), and a high noise level (60%) were used. This magnetization matching was achieved by estimating the magnetization curve as a function of β^{-1} numerically.

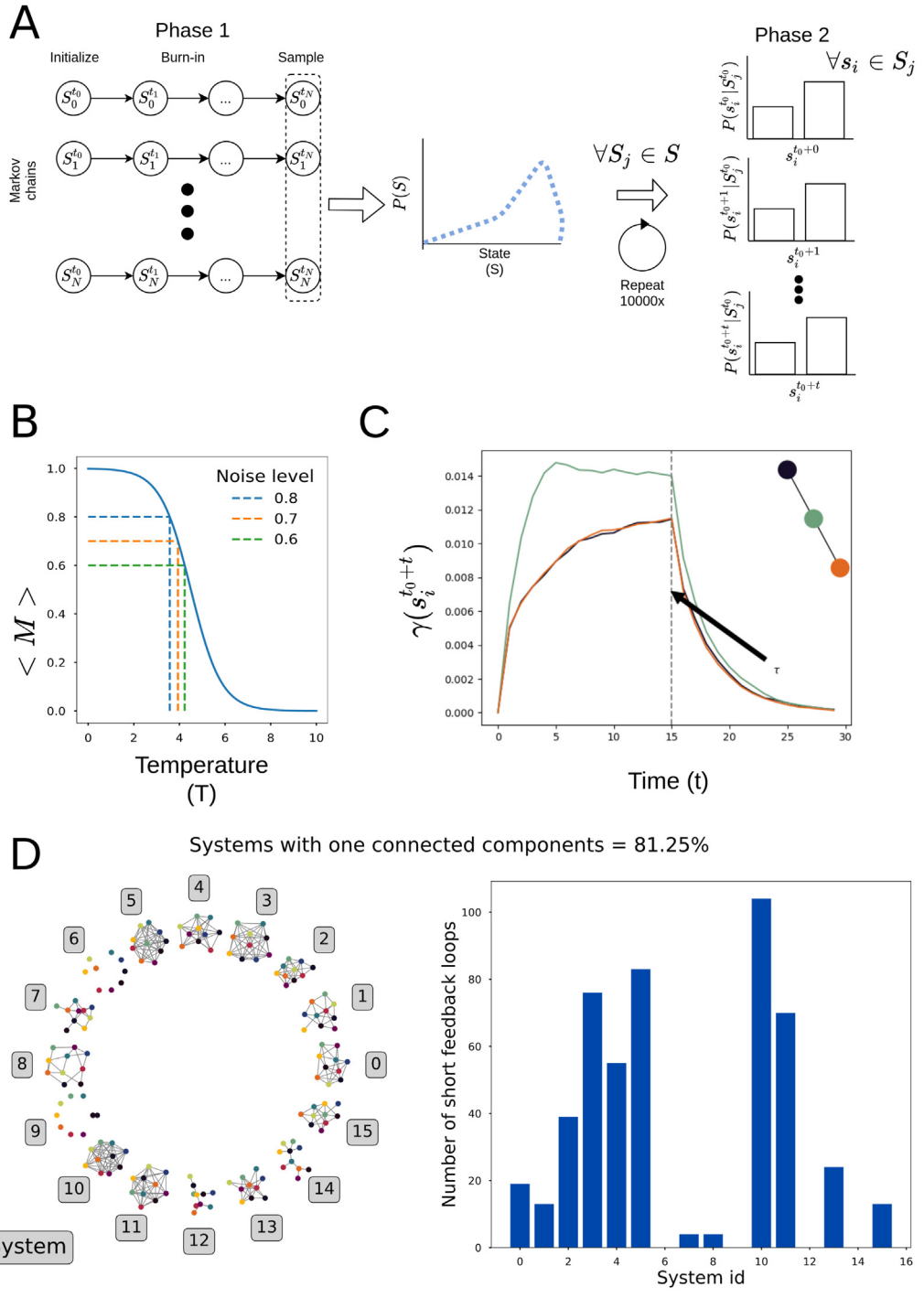


Fig. 2. (A) Phase 1 initializes 10000 independent random Markov chains and equilibrated the chains for $t_n = 1000$ time steps. The state distribution is estimated at $t_n = 1000$ over the 10000 chains. In phase 2 for system states conditional distributions are estimated for t time-steps from which mutual information is estimated $I(s_i^{t_0+t}; S^{t_0})$. (B) Illustration of temperature matching. The graph depicts the conceptual transition between an ordered state (aligned spins) to a disordered state for the kinetic Ising on complex networks. The noise level was matched to the magnetization ratio of the max magnetization. For increase in temperature the noise level increases. (C) Illustration of applied interventions in an undirected 3 state system (top right corner). Each node is nudged according to Eq. (7). The time prior to τ "pushes" the system dynamics out of equilibrium. From τ onwards the causal impact decays proportional to the causal impact a node has on the rest of the system. (D) Generated Erdös–Rényi networks. (left) Structure of each network, the number indicates the system id. (right)Number of triangles (feedback loops) for each system.

3.2.2. Estimating $p(S^{t_0})$ and $p(S^{t_0+t}|S^{t_0})$

For each noise level, $N = 1000$ independent Markov chains are run for simulation with 1000 steps (Fig. 2A). Each chain was initialized with random state distribution over nodes. We are interested in out-of-equilibrium dynamics. For the Ising model the node dynamics are symmetric in equilibrium, i.e. each node has maximum entropy. For any given trajectory, the time-scale for the system to move from a fully magnetized state, to the opposite magnetized state, will take longer than the dynamics that are exerted on local time scales. For short time-scales node with higher degree be more affected by large scale changes in their environment than lower degree nodes. For larger time-scales high-degree nodes will be frozen, and do not reflect short-time scale dynamics. In light of this, the simulations are conducted by sampling one side of the magnetization. This has the added benefit that it reduces the sign of the intervention, i.e. interventions are made to introduce higher entropy in each node which depends on the average magnetization in the system.

Each simulation step executes the following:

(1) Pick a node at random from the system with equal probability;
(2) Compute energy using Eq. (5);
(3) Flip the node state with probability Eq. (4).

From this set, the equilibrium distribution over states $p(S^{t_0})$ was constructed in the form of sample of N system states.

For this sample of system states, Monte-Carlo methods are similarly used to construct the conditional $p(s_i^{t_0+t}|S^{t_0})$. For each of the sample states $S_i \in S^{t_0}$, the procedure above was repeated 100000 times for 100 time steps in the psychosymptoms and 30 time steps for the random generated networks. All numerical experiments were repeated $n_{\text{trials}} = 20$ times to provide confidence intervals for the results in all nudge conditions and temperature settings (noise conditions).

3.2.3. Time symmetry and mutual information

Thusfar, the definition of causal impact and IMI is ambiguous to the whether t is positive or negative. Namely, if the node state $s_i^{t\pm1}$ is captured in forward in time or backward in time with respect to some state S^t. For undirected networks with Ising spin dynamics there exists time symmetry with respect to how causal influence flows through the network due to detailed balance (see Appendix A.2.2). However, for directed networks this is not the case. The results obtained here are obtained using forward simulation in time only. The detailed balance condition ensures that the results would be symmetric when simulating the system backwards in time.

3.2.4. Area under the curve estimation

The mutual information over time and KL-divergence over time were scaled for visual purposes in the range [0, 1] per trial set. A double exponential, $y = a\exp(-b(t-c)) + d\exp(-e(t-f))$, was fitted to estimate these curves and subsequently IMI (Eqs. (7) and (11)) using least squares regression (Fig. 3C, 5C). The kernel showed to be a good fit as indicated by the low fit error (Figs. 8, 9).

3.2.5. Sampling bias correction

Empirical estimates for mutual information are inherently contaminated due to sampling bias. In order to correct for this, Panzeri–Treves correction was applied [49]. This method offer a good performance in terms of signal-to-noise and computational complexity.

3.2.6. Driver node prediction and precision quantification

It is possible for two or more nodes to have exactly equal network structure features as well as node dynamics. For example consider a ring structure where each node has the exact same connectivity and all nodes have the same dynamics. In this case each node must have the same causal effect, and it would be impossible to disentangle these nodes causally from one another. Similarly graphs that are similar, e.g. show high degree of structural similarity but are not isomorphic, this causal separation may proof difficult for finite samples in stochastic settings as was discussed in 3.2.2. Consequently, we applied a parametric bootstrap procedure to estimate driver node sets (algorithm 1).

Driver set estimation. The area under curve values, e.g. integrated mutual information and causal impact, were resampled to generate bootstrap distributions (Appendix A.7). This creates a confidence interval for the integrated mutual information and causal impact. From these bootstrap distributions a driver node set is estimated (see algorithm 1 in Appendix A.7). The bootstrap procedure constructs driver node set Λ iteratively by comparing the overlap ϕ with of the bootstrap for each variable with the distribution of the estimated driver node. For all experiments $\phi = 0.5$. The driver node distribution was taken as the bootstrap distribution with the highest mean. Variables will be included to the driver node set Λ if the overlap between its bootstrap distribution i and driver node bootstrap distribution j was $\phi_{ij} > 0.5$. In total $N = 1e4$ bootstrap trial were constructed of size $M = n_{\text{trials}} = 20$; for each of the trials the average was computed. For each variables a Gaussian distribution was estimated over the N bootstrap. This distribution was used for computing the overlap with the driver node distribution.

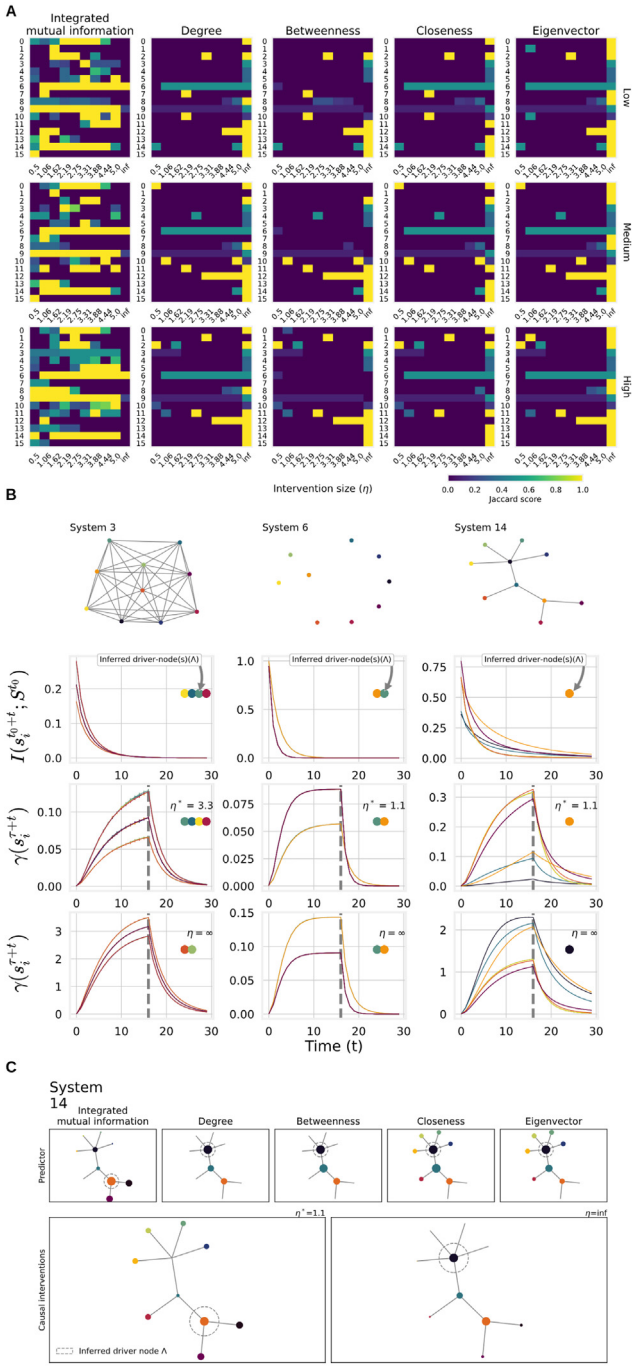


Fig. 3. (A) Jaccard score per system (see Fig. 2D for the various network structures) as a function of intervention size. Each row depicts an increase in noise indicated by the label on the right side of the plot. On average integrated mutual information is a better predictor for the driver node than the centrality metrics used in this study (see also Fig. 4). Additionally, centrality metric tend to become predictor for hard interventions ($\eta = \infty$), whereas this is not the (generally) so for integrated mutual information. Hard interventions can cause different causal dynamics not present in unperturbed system dynamics (see B/C). Integrated mutual information is based on observations of the system. The Jaccard score indicates that integrated mutual information can infer driver nodes for unperturbed dynamics. (B) Example of typical experimental results. The figures highlight a selection from the generated random networks form A for the low noise condition. (top) The graph structure of the system and the mutual information decay curves (second from top). (second bottom and bottom) Show the causal impact decay for soft intervention η^* and hard intervention $\eta = \infty$. The gray dotted line indicates $t = \tau$ where the nudge is removed from the system, the $t > \tau$ produces causal decay proportional to a node's causal importance. For soft causal interventions the inferred driver node based on integrated mutual information (second from top) is predictive for the true causal driver node (second from bottom) with soft causal interventions. Importantly, the causal driver node may change as a function of the intervention size (second bottom vs bottom plot). The centrality metrics tend to not correctly identify the driver node (see C for an example). (C) Mutual information decay ($I(s_i^{t_0+t}; S^{t_0+t})$) (top) and causal impact ($\gamma(s_i^{\tau+t})$ for minimal soft intervention η^* (middle) and hard intervention (bottom).

The bootstrap procedure cannot be applied to the network centrality metrics as there exists only one centrality rank assignment per network structure. Therefore, the driver nodes as inferred by maximum centrality metrics is the set of nodes (Λ) whose centrality metric (f) equals this maximum value, i.e.

$$\Lambda_{cent} = \text{argmax} f_{cent}(x) := \{x \in S : f_{cent}(s) \leq f(x) \forall s \in S\} \tag{12}$$

where f is the centrality function which assigns a real value to each node in the system. For degree centrality,

$$f_{deg}(a_i) = \sum_j a_{ij} \tag{13}$$

where a_{ij} is the weighted connectivity between node i and j in the adjacency matrix A of the network. If $a_{ij} > 0$ node i and j are connected. In this study degree, betweenness, closeness or eigenvector centrality were used. (see Appendix A.6 for the formal definitions for the centrality measures).

Ground truth comparison. The ground truth values are the driver node estimations for the causal impact bootstrap distribution. Each estimator also generated a driver node set estimation. That is for integrated mutual information, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, the bootstrap distribution generates an estimated driver set. To evaluate the performance of these estimators, an overlap score was computed with the ground-truth (causal impact estimation) using the Jaccard similarity metric:

$$J = \frac{A \cap B}{A \cup B}. \tag{14}$$

A Jaccard score of 1 means perfect overlap, i.e. the driver node set identified by causal impact and predictor set by IMI or one of the centrality metrics are identical. Conversely, a Jaccard score of 0 means completely disjoint driver-sets.

For every network, intervention size and temperature the similarity metric we computed the Jaccard score per predictor. Additionally, the relative performance ratio of the IMI predictor was evaluated by

$$R_{cent} = J_{IMI} - J_{cent} \tag{15}$$

where J_{cent} is the Jaccard score of the structural metrics (betweenness, closeness, betweenness, eigenvector centrality). This performance indicator falls within $[-1, 1]$ range: A value of -1 would indicate that the centrality metric correctly identified the driver node set; a score of 0 would indicate equal performance for the driver node identification between a centrality metric and IMI; a score of 1 would indicate a correct identification of the driver node for IMI, but a false identification of the centrality metric. The ratios were bootstrapped ($N = 100000$) and tested for significance at $\alpha = 0.01$. The driver node inferred performance is bound between $(0, \infty)$.

3.2.7. Software

A general toolbox was developed for analyzing any discrete systems using IMI, e.g. Susceptible–Infected–Recovered [50], Random Boolean networks [51]. The core engine is written in 3.07a with python 3.9.4 and offers C/C++ level performance,[1] for more information see https://github.com/cvanelteren/information_impact.

4. Results

4.1. Random network results

Driver node inference accuracy is depicted in Fig. 3A for both IMI and the network centrality measures. Three crucial observations can be made from the Jaccard scores. First, for nearly all systems there exists an intervention size for which IMI obtains a Jaccard score of 1 (perfect true driver node inference), whereas this is not always the case for the centrality metrics, see e.g. system 3, 7, 15. Secondly, IMI is predictive nearly only for soft intervention sizes, i.e. intervention sizes smaller or of similar order of magnitude as the existing forces acting on nodes(i.e. their degree in the J interaction matrix). In contrast, centrality metrics are mainly predictive for hard interventions. The statistical results reflect these observations (Fig. 4A and B).

In Fig. 3B average decay curves are shown for IMI (top) and different intervention sizes (middle) and hard interventions (bottom) for the systems 2, 5, 12. Their network structure is depicted in Fig. 3B (top). Comparing hard interventions (bottom) with a more moderate intervention strength (middle) in system 2, shows that the driver node can significantly differ depending on the intervention size applied to the system. The order of the causal importance is noticeably different for the minimal intervention η^* versus the strong intervention $\eta = \infty$ (middle and bottom plot in Fig. 3B).

[1] cvanelteren.github.io

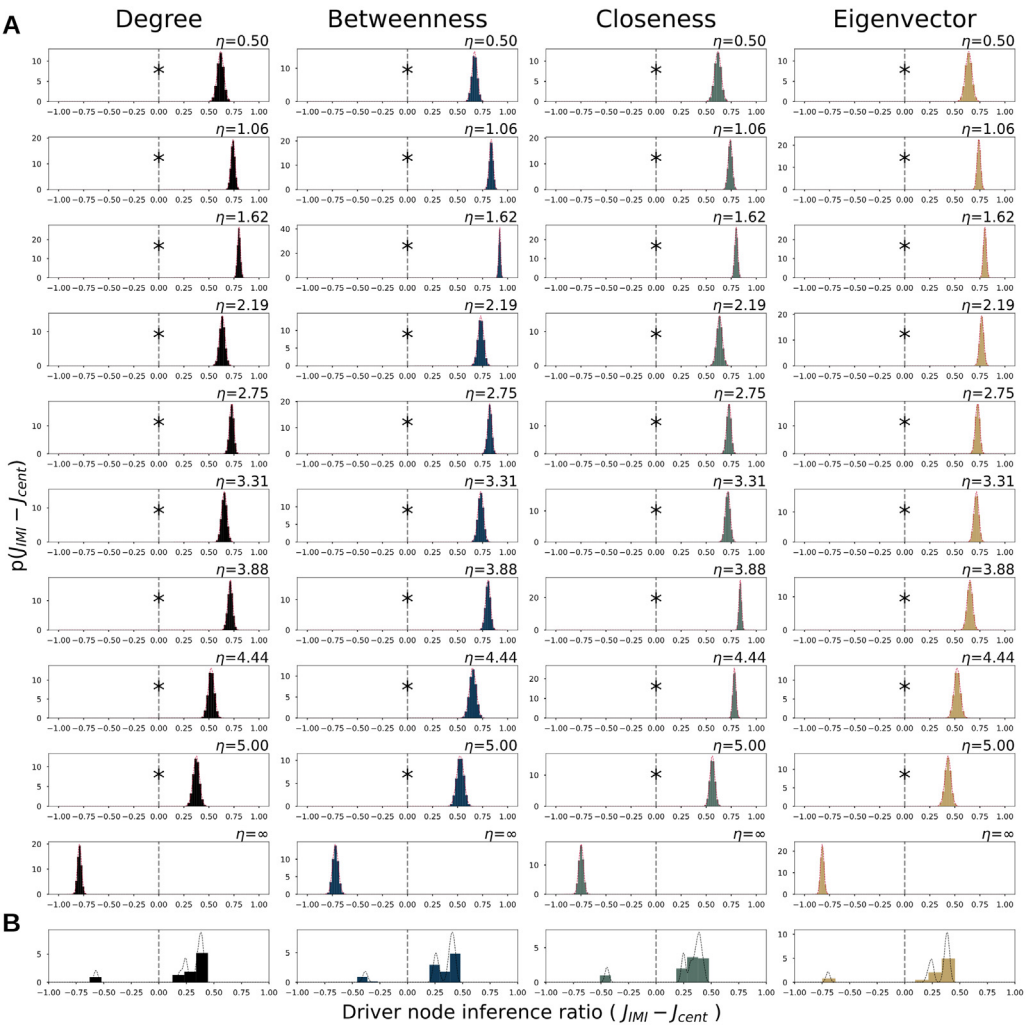


Fig. 4. (A) Bootstrap results for driver node inference score $R_{cent} = J_{IMI} - J_{cent}$ per intervention size and for all interventions and noise levels(B). A kernel density is fitted for each distribution and integrated over the interval $[-1, 0]$. An inference ratio of 1 indicates that IMI is better a predicting the driver node than the structural metric and vice versa for a score of 0. Significant intervals are indicated by a asterisk * ($\alpha = 0.01$). Except for the hard interventions ($\eta = \infty$) intervention size IMI is a better predictor than any of the network structural features.

4.2. Real-world network: psychosymptoms results

The psychosymptoms system reveals similar results to the random networks (Fig. 5). Namely, for medium to high noise level IMI yielded significantly higher Jaccard scores than centrality metrics for low causal intervention (Fig. 5C/D, $p \ll 0.01$), while not for hard interventions. In contrast, hard interventions yield different causal structures altogether which do not reflect non-intervened dynamics (Fig. 5A/B). For example the true driver node under hard interventions identify 'dislike' (medium and high noise) to be the driver node (Fig. 5A). Whereas for low causal intervention 'sad' is identified as driver node (Fig. 5A). This implies that intervention itself has impact on what causal structure is observed and that the intervention can show systemic behavior not present in the non-intervened system. The soft intervention of $\eta = 0.1$ was too low to provide proper resolution for identification of driver nodes in the low noise setting (Fig. 5A). Hard intervention in low noise condition did provide a different driver node than the soft intervention. The grid-search for optimal η^* was insufficient for the psychosymptom network and should be investigated in future studies.

In addition, the change in driver node observed in Fig. 5A highlights one major flaw in centrality metrics: they cannot account for a change in driver nodes due to a change in dynamics. The implicit assumption on dynamics that each centrality metric holds, provides only one estimate per network structure. In contrast, IMI does not depend on what

mechanisms generate system dynamics. Instead, it uses the distribution dictated by these system dynamics which match the true driver nodes.

Fried and colleagues postulated that 'lonely' was the gateway from which information spreads through the network, i.e. bereavement was embodied mainly by 'loneliness' which then percolated its effect to the other symptoms [35]. Since the data was cross-sectional, the comparison with the results from this study relies on the assumption that binary dynamics are representative of the absence and presence of psychological symptoms. If correct, the results from this study give a causal perspective on the associative results from [35]. The results from this study postulate that 'depr', 'lonely' and 'sad' have similar causal effect for moderate to high thermal noise.

It is important to emphasize that a quantification is given in terms of absolute effect size and not directed effects. This means that nudging for instance 'sleep' has some effect X on the psycho-symptom network, in what direction that effect is, or whether it has a positive or negative effect on the bereavement score / cognitive load of the patient is not clear, and should be the subject of future studies (see Appendix A.4).

5. Discussion

Structural metrics can be considered as implicitly assuming a particular dynamics [52,53]. For example, the eigenvector centrality has a clear analytical connection with linear dynamics, such as a simple diffusion process. Betweenness centrality on the other hand can be considered to assume that causation between nodes is transmitted mainly along the shortest paths between pairs of nodes (see Appendix A.6), such as in message or package routing. Finally, many more network centrality metrics have been developed based on specific dynamics, such as current and flow centrality measures. Thus, such network structural metrics could indeed be used, but a careful analysis of the dynamic equations is required to assess which centrality measure (if any) turns out appropriate.

The results from this study show that high degree nodes with Ising spin dynamics tend to actually *not* be a driver node for small interventions. In this model it can be understood since high degree nodes exhibit "frozen" behavior; their large(r) number of neighbors effectively sum up to a constant and strong force towards the majority state. Consequently, a constant (soft) intervention will be relatively ineffective with hubs. For non-intervened Ising dynamics, it makes intuitive sense that the dynamics are driven by the nodes which are neither "frozen" (hubs) nor are poorly connected (low-degree nodes), which is reflected by the soft intervention setting. This insight is specific to the Ising system but illustrates that the information flow and causal flow of a system is not merely determined by the connectedness of a node in the network. Rather, the connectedness of the entire system in addition to inter-node dynamics combined are crucially important for causal flows and driver node identification.

The results further imply that for unperturbed dynamics structural metrics may not be predictive for determining the causal importance of nodes. Soft interventions revealed different causal structure than hard interventions (Fig. 3 5). For hard interventions, structural metrics do become predictive for causal importance for the networks studied here. However, the dynamics of these systems are shown to deviate from the non-intervened dynamics, causing system dynamics that are not representative of the non-intervened system. This can be seen in the causal influence in Figs. 3 and 5 where the causal influence for hard interventions is opposite to soft causal interventions. Consequently, if the aim is to provide understanding to the information flows for non-intervened dynamics, then hard interventions are not preferred.

In addition, the results imply that in order to achieve maximum impact for a fixed 'intervention' budget (injected energy), choosing the high degree nodes is not necessarily optimal. The adjusted Hamiltonian (Eq. (8)) introduces a bias on low degree nodes. Here, a causal intervention was performed by adding fixed energy to the Hamiltonian. For fixed intervention size $\eta = c$, the causal impact on the nodal distribution will be relatively higher for nodes with low degree than nodes with higher degree. Higher degree nodes may have higher causal effect in principle if the same probability mass is changed. However, moving the same probability mass scales non-linearly in kinetic Ising (Eq. (5)). For a limited 'intervention budget' it is preferred to locate those elements of the system that reaches maximal causal effect.

For, some systems, however, the inferred driver node never matched the true driver node(s) (e.g. system 8 in Fig. 3). This could be due to two main reasons. First a grid-search was applied for the intervention size. It was argued that there would be a minimal intervention η^* which would lead to a measurable effect. It is possible that the parameter space used here missed the intervention size that provided enough resolution to accurately determined the driver node(s). Second, the numerical procedure for driver node inference could be optimized. The overlap of distribution was set to $\phi = 0.5$, to infer the driver node set Λ, to prevent false positives for driver node identification due to noise in observed system states. However, the choice of parameter was not optimized and could lead to ambiguous driver node inference. System 8 in particular was most nodes in the network had a similar connectivity pattern, and as such causal isomorphy occurs, i.e. the causal importance of a node is indistinguishable from any other node in the system. The bootstrap estimates led to high level but not perfect of overlap (see Appendix A.7). Consequently, if ϕ was set differently, the inferred driver nodes could be improved. In future studies, we aim to further look into what how this numerical procedure could be improved for inferring driver nodes in dynamical networks particularly for causal isomorphic nodes.

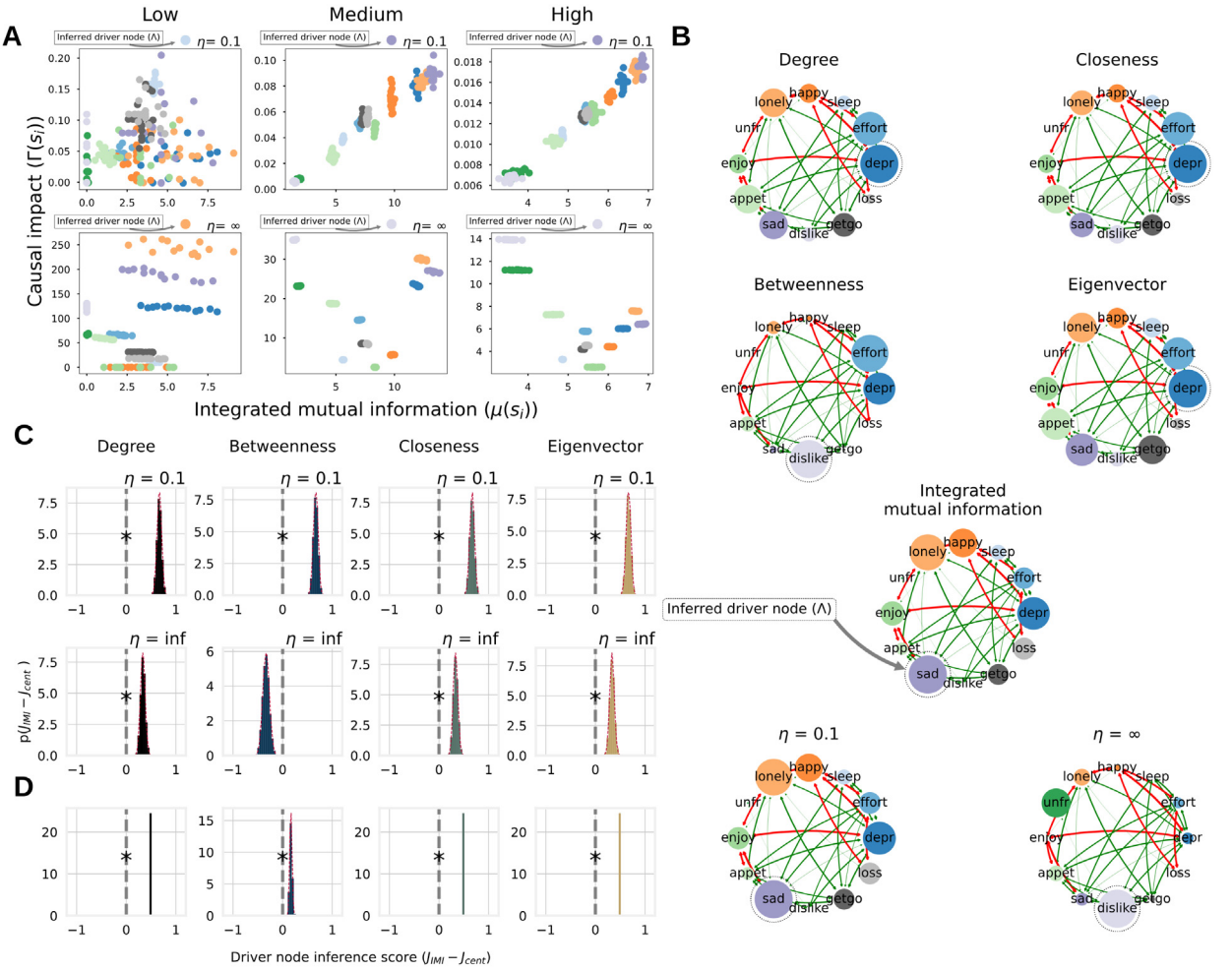


Fig. 5. (A) Causal impact versus integrated mutual information for different noise levels and intervention sizes. The inferred driver node is indicated above each figure with a dot. For low noise level the numerical estimates were too noise and driver estimation was inaccurate. For intermediate to high noise level, causal impact scale linearly with integrated mutual information for soft interventions. Causal driver node was wrongly predicted for hard interventions by integrated mutual information. Notice how the causal flows change as a function of intervention size (top versus bottom plot). (B) Driver node identification in real-world network of psychosymptoms for medium noise as a function network features (top 4 subplots) and integrated mutual information (middle), and true causal driver nodes per intervention size (bottom). Integrated mutual information correctly identified the driver node with soft interventions, but not for high interventions. Only betweenness centrality correctly identified the true driver node for high causal interventions. (C) Bootstrap distributions for driver node inference score as function of intervention size across noise levels. A score of 1 indicates that integrated mutual information identified the driver node correctly, but the structural metric did not, and vice versa for a score of -1. In red kernel density estimates are indicated and integrated over $[-1, 0]$. The asterisk (*) indicates a significant better inference score for integrated mutual information ($\alpha < .01$). Integrated mutual information was a better predictor than any of the network features for soft interventions. (D) Bootstrap distribution for driver node inference score as function of temperature and nudge size. The integrated mutual information was a significant better predictor across noise and intervention size ($\alpha < 0.05$ indicated by *).

5.1. Limitations

The systems considered are discrete and ergodic. IMI assumes that the data-processing inequality holds for the system (Section 2.3.2). The data-processing inequality in ergodic systems ensures that $I(s_i^{t_0+t}; S^{t_0})$ monotonically approaches zeros as $t \to \infty$ (see Appendix A.1). As a consequence IMI will always be finite for ergodic systems. For non-ergodic systems, the data-processing inequality can only guarantee that $I(s_i^{t_0+t}; S^{t_0})$ never increases as a function of t. Namely, the data-processing inequality ensures that no local manipulation of information may increase the information content of a signal. This implies that as $t \to \infty$, IMI may not converge for nodes with non-zero baselines. For these non-ergodic systems, however, it may be possible determine the driver nodes by considering finite time-scales, or subtracting the asymptotic γ value and reporting it separately.

Furthermore, this study only considered information flows undirected systems. For directed systems, the largest integrated mutual information may in fact not be the driver node due to correlation. In Appendix A.2 an integrator is node is considered, where the integrator stores the system information but lacks any causal outgoing connections. The information storage capacity in this node will be higher than any other node due to its increased alphabet. If this integrator node encodes a copy operator, its information decay will result in the highest integrated mutual information while lacking any causal potential. In the systems considered here, this situation cannot occur. That is, for undirected systems information flow has equal potential along an edge. Any causal effect a node has in undirected dynamical systems, will percolate to all neighbors of a node. For directed graphs the information flow is not time symmetric as detailed balance cannot be ensured. Consequently, the use of forward mutual information $I(s_i^{t_0+t} : S^{t_0})$ may lead to false estimates of driver nodes in these cases. For undirected graphs, if there would exist a node with higher integrated mutual information, its causal impact will also be higher and as such it would be the driver node. The asymmetry of information flow, and the compound nature of both correlation and causation restricts the use of integrated information undirected graphs.

In addition, the focus on short-time scales biases information processing to less constrained nodes. By measuring the system out-of-equilibrium, the system dynamics for the kinetic Ising model has the tendency to "freeze" the behavior of nodes as a function of increasing degree. For undirected systems, this assumption seems sensible as the higher causal potential for hubs forms at the same time a higher constraint for switching its state. The flip probability of node s_i in the kinetic Ising model as a function of degree will tend to zero given that the neighborhood is congruent with the node's current state. Measuring the system in equilibrium, will therefore result in nodes with higher degree to have lower information decay. In contrast, out-of-equilibrium a node state will have lower variability and consequently less causal effect on the instantaneous system states. Therefore, the driver node identification through integrated mutual information are only valid for non-equilibrium kinetic Ising spin dynamics with undirect network structures. Since the Ising spin dynamics falls within a larger universality class, such as (directed) percolation. This implies that the results from this study may apply to different dynamical systems sharing the same core assumptions.

Systems of size at most $n = 12$ were used. The size of the system was chosen due in order to provide high reliability of the probability distributions. In addition, larger graphs can be decomposed in various different network motifs [34]. It is believed that these motifs form the "computational" buildings blocks for larger complex systems under the assumption of nearest neighbor interactions. That is, understanding the motifs would gain insights in how a macroscopic property emerges from local interactions. The aim of this study was to relate structural connectedness to dynamic importance; the exact nature or occurrence of motifs were not the focal point. In real-world systems, however, it is exactly the composition and interaction of these motifs that are vital to complex systems. Here, the motifs were implicit on the real-world network and the generated structures. We leave it up to future work to map out the driver nodes of common network motifs in different dynamics and relate the structural importance to the dynamical importance.

6. Conclusions

Our results indicate that dynamic importance cannot necessarily be reliably inferred from network structural features alone, demonstrated here using kinetic Ising spin dynamics. The goal of this paper was to show that structural methods can provide unreliable estimates of the driver node in dynamical systems. The results from this study show that the common assumption of structurally central or well-connected nodes being simultaneously dynamically most important is not necessarily true. This implies that we cannot abstract away the dynamics of a dynamic system before inferring driver nodes. The proposed information theoretic metric, integrated mutual information (IMI), was better able identify the driver node for non-intervened dynamics in systems. Importantly, IMI does not require knowing the dynamics equations and/or the network structure of the system. It is instead calculated directly from a cross-section of time-series of the system without interventions. The proposed metric could potentially be useful in applications with rich data sets and where performing interventions are infeasible or impractical.

CRediT authorship contribution statement

Casper van Elteren: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Rick Quax:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Peter Sloot:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by grant Hyperion 2454972 of the Dutch National Police. In addition, Rick Quax acknowledges funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 848146 (ToAition).

We thank M.Sc. Fiona Lippert, Dr. Paul Duijn and Dr. Thijs Vis for fruitful discussions in support of this paper. We declare that no conflict of interests are present in the conduction of this study.

Appendix A

A.1. Data-processing inequality

The data-processing inequality can be used to show that no clever manipulation of the data can improve the inferences made from that data (see for more detail [19]).

Definition 1. Random variables $X \to Y \to Z$ are said to form a Markov chain if the conditional distribution of Z depends only on Y and is conditionally independent of X. Specifically, X, Y, Z form a Markov chain if the joint probability can be written as:

$$p(x, y, z) = p(x)p(y|x)p(z|y) \tag{16}$$

Theorem 1 (*Data-Processing Inequality*)**.** *If $X \to Y \to Z$, then $I(X;Y) \geq I(X;Z)$.*

Proof. By the chain rule, the mutual information can be expanded in two different ways:

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + I(X;Z|Y) \end{aligned} \tag{17}$$

Since X and Z are conditionally independent given Y, we have $I(X;Z|Y) = 0$. Conversely, if $I(X;Y|Z) \geq 0$, this would give

$$I(X;Y) \geq I(X;Z). \tag{18}$$

Thus we only have equality if and only if $I(X;Y|Z) = 0$ for Markov chains. Similarly, one can prove that $I(Y;Z) \geq I(X;Z)$■.

Corollary 1. *If $Z = g(Y) \to I(X;Y) \geq I(X;g(Y))$*

Proof. $X \to Y \to g(Y)$ forms a Markov chain. ■

This result implies that no function $g(Y)$ can increase the information about X.

Corollary 2. *If $X \to Y \to Z$, then $I(X;Y|Z) \leq I(X;Y)$*

From Eq. (17) it is noted that $I(X;Z|Y) = 0$ due to the definition the Markov chain and $I(X;Z) \geq 0$. Therefore:

$$I(X;Y|Z) \leq I(X;Y)\blacksquare \tag{19}$$

The dependence of X and Y is decreased or remains unchanged by the observation of a "downstream" random variable Z. For any complex system in which the state distribution follows a Markov chain, i.e. $X^{t_0} \to X^{t_0+1} \to \cdots \to X^{t_0+\infty}$. The mutual information $I(X^{t_0}; X^{t_0+t})$ will always decay to zero as $t \to \infty$.

A.2. A note on causality and mutual information

A.2.1. Mutual information and causal bivariate interaction

Mutual information decays to causal interactions for simple bivariate interactions. For Markovian systems, the future state of the system is independent of its past given its present (Eq. (2)). Therefore, we can write [19]

$$p(s_j^{t+1}|S^t) = p(s_j^{t+1}|s_i^t). \tag{20}$$

For any two nodes $s_i, s_j \in S$, the KL-divergence $D_{KL}(p(s_j^{t+1}|s_i^t) \parallel p(s_j^t))$ reduces to $I(s_i^{t+1}; s_j^t)$.

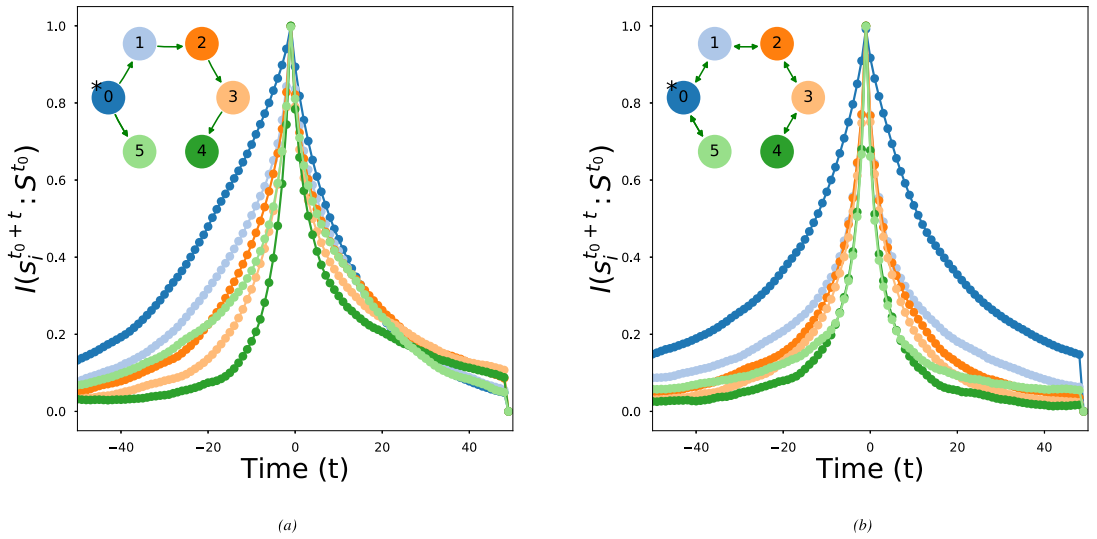


Fig. 6. Example of time symmetry in directed and undirected networks. Fig. 6a shows the asymmetry that occurs when information flow is directed. The time before the system state $t < t_0$ can be interpreted as information sending. Namely, nodes that have the most impact on the current system state S^{t_0}. In contrast, information for $t > t_0$ as information receiving; nodes that receive information from S^{t_0}. The most striking example is node 4 which has a sharp decay for $t < t_0$ but a relatively fat tail for $t > t_0$. This change is due to the difference in meaning of the IMI, e.g. sending vs receiving.
Fig. 6b shows that for undirected networks there is no difference between node importance before or after t_0; information flows both directions..

Theorem 2. *$E_{s_i}[D_{KL}(p(s_j^{t+1}|s_i^t) \parallel p(s_j^t))] = I(s_i^{t+1}; s_j^t)$ when s_i and s_j have no common neighbors.*

Proof.

$$
\begin{aligned}
E_{s_i^t}[D_{KL}(p(s_j^{t+1}|s_i^t) \parallel p(s_j^t)] &= E_{s_i^t}\left[E_{s_j^{t+1}|s_i^t}\left[\log \frac{p(s_j^{t+1} \mid s_i^t)}{p(s_j^{t+1})}\right]\right] \\
&= \sum_{s_i^t} p(s_i^t) \sum_{s_j^{t+1}} p(s_j^{t+1} \mid s_i^t) \log \frac{p(s_j^{t+1} \mid s_i^t)}{p(s_j^{t+1})} \\
&= \sum_{s_i^t} p(s_i^t) \sum_{s_j} \log p(s_j^{t+1} \mid s_i^t) \log p(s_j^{t+1} \mid s_i^t) - \sum_{s_j^{t+1}} p(s_j^{t+1}) \log p(s_j^{t+1}) \\
&= H(s_j^{t+1}) - H(s_j^{t+1} \mid s_i^t) \\
&= I(s_j^{t+1} : s_i^t) \blacksquare.
\end{aligned}
\tag{21}
$$

A.2.2. Mutual information and time symmetry

The methods applied in the main text imply that the metric can be used symmetrically. In this study time-delayed mutual information was performed in a 'forward' manner for practical purposes. Namely, the system state was simulated for positive t from some t_0. For undirected networks there is a symmetry with regard to where information flows. Information is not bounded by any directionality of edges (Fig. 6(a)).

It is important to emphasize that this (generally) is not the case for directed networks. For kinetic Ising models detailed balance is not guaranteed and as such mutual information may be conflated for integrated nodes (see A.2.3). If information is constricted to flow in one direction, the mutual direction of time simulation is crucial. Additionally, directed networks show that the metric can be applied for different purposed. This can be seen in Fig. 6, where forward simulations gives 'information sinks' and backward simulation provides 'information sources'. IMI in directed networks will provide information about what nodes receive the most information over time, i.e. the correlation of a node with "sending" information. In contrast, simulating backwards shows what nodes have most impact on the instantaneous state of the system. This dual-use of information will be the focus of future studies.

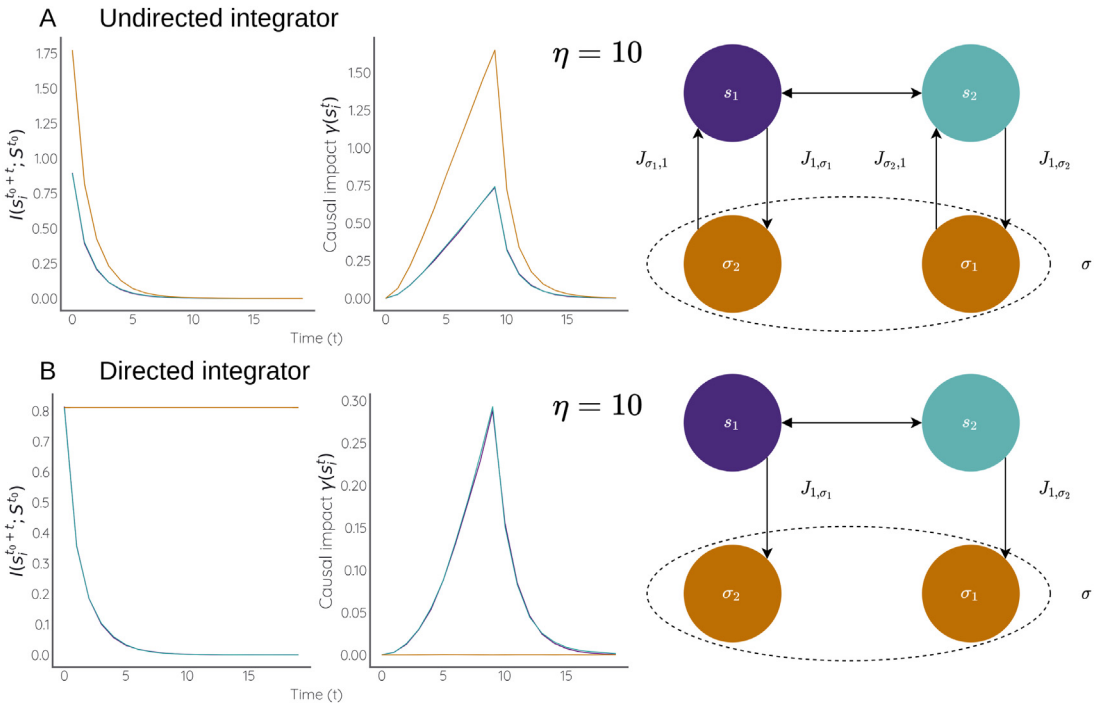


Fig. 7. Integrated mutual information for "integrators" may fail to identify driver node for directed graphs (B) but not for undirected (A). Each node in the system is governed by kinetic Ising spin with $\sigma(t) = \{\sigma_1(t), \sigma_2(t)\}$. In B, the integrator node copies the system state ($J_{1,\sigma_2} = J_{\sigma_2,2} \to \infty$). As the integrator has no causal outgoing connections, integrated mutual information does not reflect the driver node. For directed graphs, detailed balance cannot be ensured violates one of the assumptions of integrated mutual information. In A, detailed balance is ensured and the integrator node reflects the driver node.

A.2.3. Integrated mutual information and integrator nodes

As noted in A.2.2 the driver node can only be determined for ergodic systems with symmetric edges, that is the interaction $J_{s_i,s_j} = J_{s_j,s_i} \forall s_i, s_j \in S$. For directed graphs (i.e. non-symmetric edge weights) integrated mutual information may be conflated and produce a wrong estimate for the driver node. These cases are, however, outside the scope of the manuscript, and undirected integrators can readily be identified by integrated mutual information.

Consider the system $S(t) = \{s_1(t), s_2(t), \sigma(t)\}$ where integrator $\sigma(t) = \{\sigma_1(t-1), \sigma_1(t-1)\}$ with $\mathbb{H}(\sigma) = -J_{s_1,\sigma_1} s_1 \sigma_1 - J_{s_2,\sigma_2} \sigma_2 s_2$, and each node is governed by kinetic Ising spin dynamics (see Fig. 7). The node $\sigma(t)$ is the integrator node that consists of two "sub-nodes" that connects to node s_1, s_2 separately. Note that the interaction weight J can be adjusted. The traditional kinetic Ising spin dynamics are obtained for $J_{i,j} = 1$ we obtain the traditional kinetic Ising spin dynamics, and as $J_{i,j} \to \infty$ the dynamics tend to a copy operator. In addition, the weights may be set such that a directed graph is obtained (see Fig. 7B).

We consider two cases. First the case where σ contains no outgoing causal connections, that is $J_{\sigma_1,1} = J_{\sigma_2,2} = 0$. In the extreme case where the integrator is a copy operator ($J_{\sigma_1,1} = J_{\sigma_2,2} \to \infty$), the integrator σ stores information the system, but has itself no causal impact. Integrated mutual information in this setting cannot be used as a driver node identification as the shared mutual information with the system is merely correlational.

For undirected edges, however, e.g. $J_{\sigma_1,1} = J_{1,\sigma_1} = 1, J_{\sigma_2,2} = J_{2,\sigma_2} = 1$ (Fig. 7A), integrator node will have the highest causal potential as the impact of an intervention on the integrator can percolate through the system. Integrated mutual information correctly identifies the driver node. Note that the size of the connection weights matters. As $J_{\sigma_i,i} = J_{i,\sigma_i} \to \infty$ the interaction the integrator node will tend to a copy operator. For a copy operator the dynamics of the system will degenerate. In this setting, the out-of-equilibrium dynamics will be non-existing as the dynamics of s_i and s_j will be frozen through the edge with the integrator. This will results that out-of-equilibrium, the system state will merely be copied and no transient dynamics will be possible. Information flows will not occur as the relative entropy will be zero. This setting violates detailed balance. Relaxing the interaction J, enables detailed balance and we see under these conditions the driver node is correctly identified (Fig. 7A). The integrator node's causal impact can flow due to the detailed balance.

In conclusion, directed edges violates detailed balance and consequently integrated mutual information will not be able to identify the driver node as the node with the highest mutual information through forward inference, i.e. $I(s_i^{t_0+t} : S^{t_0})$.

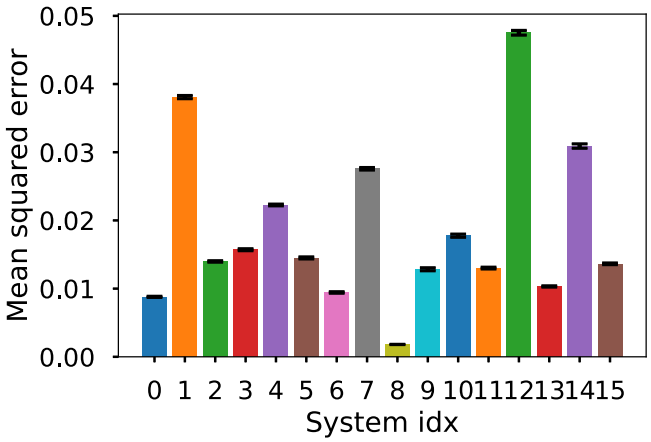


Fig. 8. Average mean squared error $\pm 2SEM$ per system.

Nodes with the highest causal impact are correctly identified through integrated mutual information for kinetic Ising spin dynamics on undirected network structures where detailed balance is ensured.

A.2.4. Time reversibility and detailed balance

In order to show that $I(s_i^{t+1}; S^t) = I(s_i^{t-1}; S^t)$, we have to show that $p(S_*^{t+1}|S_*^t) = p(S_*^{t-1}|S^t)$.

Two-sided Markov chains. For a positive recurring Markov chain $\{S^t : t \in \mathbb{N}\}$ with transition matrix P and stationary distribution π, let $\{S_*^t : t \in \mathbb{N}\}$ a stationary version of this chain, i.e. $S^0 \sim \pi$. We can construct a two-sided extension of S^t by defining a shift $k \geq 1$, extending the process S backwards in time: $S_*^t(k) = S_*^{t-k}$ with $0 \leq k < \infty$. It is true that by stationarity S_*^{t-k} has the same stationary distribution as S^t: we arrive at $\{S_*^t : t \in \mathbb{Z}\}$.

Detailed balance. Let $\{S_*^t : t \in \mathbb{Z}\}$ be a two-sided extension of a positive recurrent Markov chain with transition matrix P and stationary distribution π. A transition from state i to j is denoted with notation $i \to j$.

$$\begin{aligned} p(S_*^1 = j|S^0 = i) &= p(S_*^1(k=1) = j|p(S_*^0(k=1)) = i) \\ &= p(S_*^0 = j|S_*^{-1} = i) \\ &= \frac{p(S_*^{-1} = i|S_*^0 = j)p(S_*^0 = j)}{p(S_*^{-1} = i)} \\ &= \frac{\pi_j}{\pi_i} P_{i\to j} \end{aligned} \tag{22}$$

In other words the time-reverse Markov chain is a Markov chain with transition probabilities:

$$P_{i\to j}\pi_i = \pi_j P_{j\to i} \tag{23}$$

This is also known as detailed balance. In the manuscript, ergodic systems are used and therefore satisfy the time-symmetry. This results that $I(s_i^{t+1}; S^t) = I(s_i^{t-1}; S^t)$.

A.3. Data correction and fit errors

A.4. Validation of psychosymptoms

The results from this study imply that for low thermal noise not enough resolution was possible to reliably estimate the driver node. For medium to high noise levels, the 'sad' emerged as the driver node.

In the original study, the bereavement score was most affected by 'lonely', and showed weak negative associations with 'happy' and 'effort' (Fig. 10 adopted from [35]). Consequently, it seems that medium to high thermal noise is most congruent with the original study. Fried and colleagues postulated that 'lonely' was the gateway from which information spreads through the network, i.e. bereavement was embodied mainly by 'loneliness' which then percolated its effect to the other symptoms. Since the nature of the data was cross-sectional, the comparison with the results from this study relies on the assumption that binary dynamics are representative of the absence and presence of psychological symptoms.

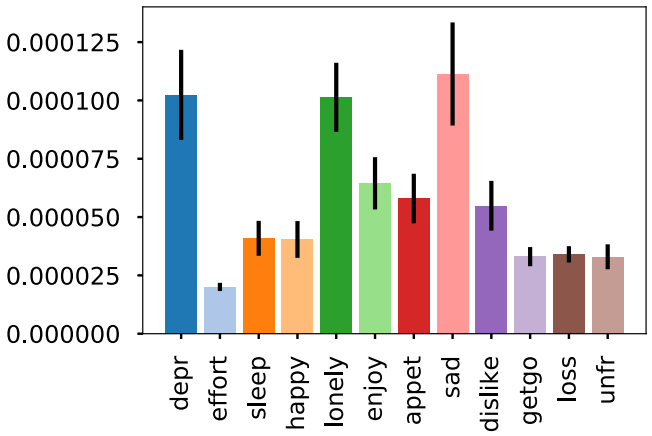


Fig. 9. Average mean squared error $\pm 2SEM$ for psychosymptom system.

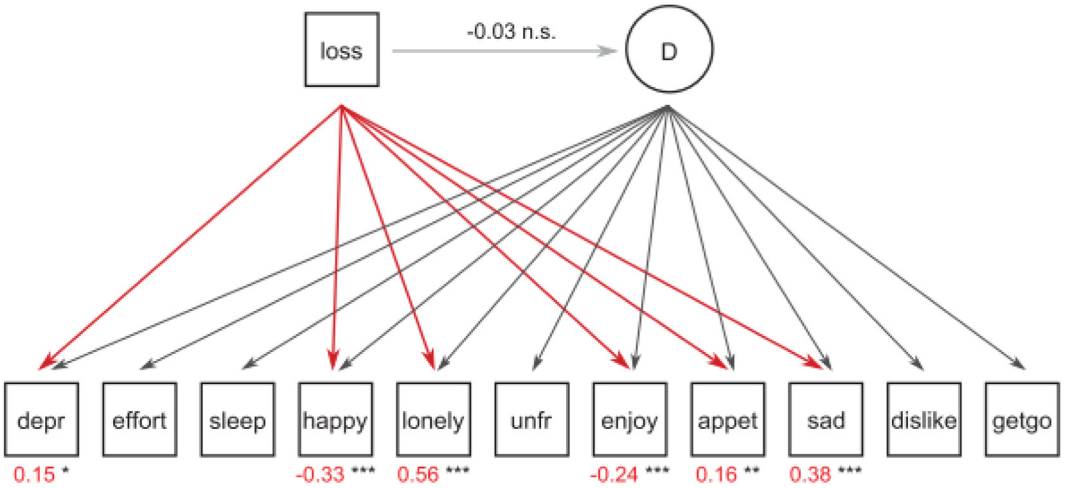


Fig. 10. Main results from Friend and colleagues [35]. The network represents the output from a Multiple Indicators Multiple Causes (MIMIC) model. The red lines indicate significant direct effects of spousal loss on Center for Epidemiological Studies Depression Scale(CES-D items); standardized estimates of these affects are represented in red below the symptoms. There was no significant loading of loss on the latent factor *D*. For more info see [35].

If correct, the results from this study give a causal perspective on the associative results from [35]. The results from this study postulate that 'depr', 'lonely' and 'sad' have similar causal effect for moderate to high thermal noise.

It is important to emphasize that a quantification is given in terms of absolute effect size and not directed effects. This means that nudging for instance 'sleep' has some effect X on the psycho-symptom network, in what direction that effect is, or whether it has a positive or negative effect on the bereavement score / cognitive load of the patient is not clear, and should be the subject of future studies.

As a final note, the field of psychometrics is concerned with relating how observables (e.g. behavior, responses on questionnaires, *etc*) relate to theoretical cognitive constructs such as intelligence or mental disorders. A common approach in understanding high level phenomena such as depression is to use a latent variable model, i.e. assuming some high abstract feature to be the cause of the observables (or vice versa). Only recently has this paradigm shifted from a latent variable model to a network based approach [16,54,55]. Marsman et al. recently reconciled these two approached by showing statistical equivalence between the Ising model and canonically used latent variable models in psychometrics [56]. The two approaches thus highlight different aspects in theory building; measurement invariance and correlation structure may be interesting from a common cause approach but not from a network perspective which is more interested

in dynamical aspects of the system. Both approaches, however, aid in highlighting different aspects of the psychological constructs.

A.5. Code manual

Accompanying this paper, I developed a general framework for analyzing discrete systems using IMI. The code is written in python 3.7.2 and uses cython 0.28.2 for c/c++ level performance. The code is freely available on cvanelteren.github.ioincludes the latest build instructions.hat follows here is a brief overview of the framework.

A.6. Structural methods

Network analysis has traditionally resulted in analyzing the structure of the network. A fundamental concept within network science is centrality, and how to measure the centrality of nodes has become an essential part of understanding networked systems such as social networks, the internet, biological networks, traffic and ecological networks. At its core, a centrality measure quantifies the 'importance' of a node based on some structural property. It allows ranking nodes based on a real-valued function.

There is, however, a long-standing debate concerning what centrality metrics actually measure for networked systems [6,52,53,57] . From a network theoretical perspective most centrality measures, e.g. betweenness, closeness, eigenvector and degree centrality, essentially classify the 'walk structure' of a network [52,53]. A walk from node i to node j is a sequence of adjacent nodes that begins with i and ends with j. The structure of walks can be divided along different criteria. For example a trail is a walk in which no edge (i.e. adjacent pair of nodes) is repeated. In contrast, a path is a trail in which no node is visited more than once. Similarly, one could define a walk structure by only using the shortest path from one node to another, or by using random movements between nodes (random walks).

Alternatively, from a complex systems perspective, centrality metrics *implicitly assume* dynamics on the network structure. Betweenness centrality for example, computes centrality based on how often a node acts as a bridge along the shortest path between two other nodes. If one assumes that the network has dynamics D where information between nodes follows the shortest path, this metric may be a valid description to use and identify dynamically important nodes.

In the best case, a centrality metric is fully predictive for identifying important nodes a complex system. Consequently, the centrality metric can be used to understand the system. However, an issue with the use of centrality metrics is determining which centrality metric to use. Consider for example Fig. 1A; different centrality metrics can identify different nodes as most central. This has lead to the common observation that some centrality measures can 'get it wrong' when the aim is to predict dynamical important structure in networked systems. Additionally, the ranking produced through some centrality metric does not quantify inter-rank differences. This potentially leads to underestimation of nodal influence when used in dynamic context [6].

We will show how centrality measures have no meaningful prediction power of the most causal node in nodes dictated by the Gibbs measure. We are aware that centrality measures do not embody the full extent of what structural methods embody, or what network science in particular has to offer. However, many structural methods share the common characteristics listed above, i.e. they quantify the walk structure of a network. For our analysis, we used the weighted variants of degree centrality, betweenness centrality, information centrality, and eigenvector centrality. What follows is a brief description of commonly used centrality metrics.

A.6.1. Degree centrality

Degree centrality is the best-known measure of all the centrality measures. It is often thought that degree centrality is indicative for the dynamic importance of a node. This intuition is based on the concept of flow: the more connection a node has, the more interaction potential that node has and therefore the more important a node must be. Freeman defined centrality measure as the count of the number of edges incident upon a given node [58]:

$$c_i^{\text{deg}} = \sum_j a_{ij} \tag{24}$$

where a_{ij} is the row/column of node i in the adjacency matrix A of the network. Please note that the entries a_{ij} are weighted and not binary.

A.6.2. Betweenness centrality

Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. It was introduced as a measure for quantifying the control of communication among humans in social

networks by Freeman [58]. Nodes that have a high probability to occur on a randomly chosen shortest path between two randomly chosen vertices have a high betweenness. Formally, this can be written as:

$$c_i^{\text{betw}} = \sum_{j,k} \frac{\sigma(j,k|i)}{\sigma(j,k)} \tag{25}$$

where $\sigma(j,k)$ represents the number of shortest paths between node j and k, and $\sigma(j,k|i)$ is the subset that goes through node i. We use the normalized version of betweenness that divides the betweenness score by the number of pairs of vertices (not including node i);

$$\begin{aligned} c_i^{\text{betw}} &= \frac{1}{Z} \sum_{j,k} \frac{\sigma(j,k|i)}{\sigma(j,k)} \\ Z &= \frac{((n-1)(n-2))}{2} \end{aligned} \tag{26}$$

A.6.3. Closeness centrality

Closeness centrality is defined as the reciprocal sum of the length of the shortest paths between the node i and all other nodes in the network $j \in N$:

$$c_i^{\text{close}} = \frac{1}{\sum_j^N d(i,j)}. \tag{27}$$

From a complex system perspective it assumes that information is transferred along its shortest paths. A node with short distance to many other nodes will be able to quickly transfer its information to other nodes in the network.

A.6.4. Eigenvector centrality

Eigenvector centrality is the most difficult centrality measure to give an intuitive feeling for. Where A is the adjacency matrix of the system, eigenvector centrality of node i is defined as:

$$c_i^{\text{ev}} = \frac{1}{\lambda} \sum_j a_{ij} x_j \leftrightarrow Ax = \lambda x \tag{28}$$

For any square matrix of rank n, the matrix will have at most n eigenvector–eigenvalues pairs. A common choice for eigenvector centrality is motivated by The Perron–Frobenius theorem, and involves choosing the eigenvector x with the largest eigenvalue λ [59,60]. This has the desired property that if A is irreducible, or equivalently if the network is strongly connected, that the eigenvector x is both unique and positive.

The sign and size of the eigenvalue are important for the relation between the value and importance of a node. In linear differential equations negative eigenvalues correspond to non-oscillatory exponentially stable solutions. In contrast, in difference equations it indicates an oscillatory behavior. Geometrically speaking, negative eigenvector embodies a linear transformation across some axis.

Intuitively speaking, eigenvector centrality quantifies the influence of a node in the network. It assigns relatives scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. A high eigenvector score implies that the node is connected to many other nodes that themselves have high scores. Google PageRank and Katz centrality are variants of eigenvector centrality [61]. A node with high eigenvector centrality is not necessarily a node that has many connection (incoming or outgoing). For example a node may have a high eigenvector centrality if it has few connections, but those connections are connected to nodes that are of high importance.

A.7. Bootstrap distributions

A total of 10000 bootstrap samples were conducted with replacement. For each nodal bootstrap distribution a gaussian kernel density was estimated. The node with the highest causal impact was chosen as the initial driver set Λ. Then the overlap ϕ between this driver node and the remaining nodes in the system was computed iteratively. We considered the overlap $\phi = 0.5$ or higher was sufficient for the node was to be considered causally similar to the driver node. Therefore, the proposed node will be included in the driver node set if $\phi \geq 0.5$.

A.7.1. Numerical procedure

```python
def bootstrapDrivers(D, N, M, phi=0.5) -> dict:
    """
    Determines driver-nodes through the  boostrap distribution It consists as an
    iterative procedure  that takes  the max  value in the  data, then  a normal
    distribution is  fitted an tested  with overlap for  the other nodes  in the
    data at `phi`

    Parameters
    ----------
    D: np.array
        2d matrix consisting of size n_features x n_observations
    N: int
        Total number of samples used for bootstrap
    M: int,
    number of draws for each of each bootstrap sample
    phi : float
        Overlap between distribution.

    Returns
    ----------
        dict of driver-nodes under overlap `phi`
    """
    # part of standard library python
    from statistics import NormalDist
    import numpy as np

    # generate subsamples
    n_features, n_obs = D.shape
    D_bar = np.zeros((n_features, N))
    # sample from observations
    for var in range(n_features):
        D_bar[var] = np.random.choice(D[var], size=(N, M)).mean(1)

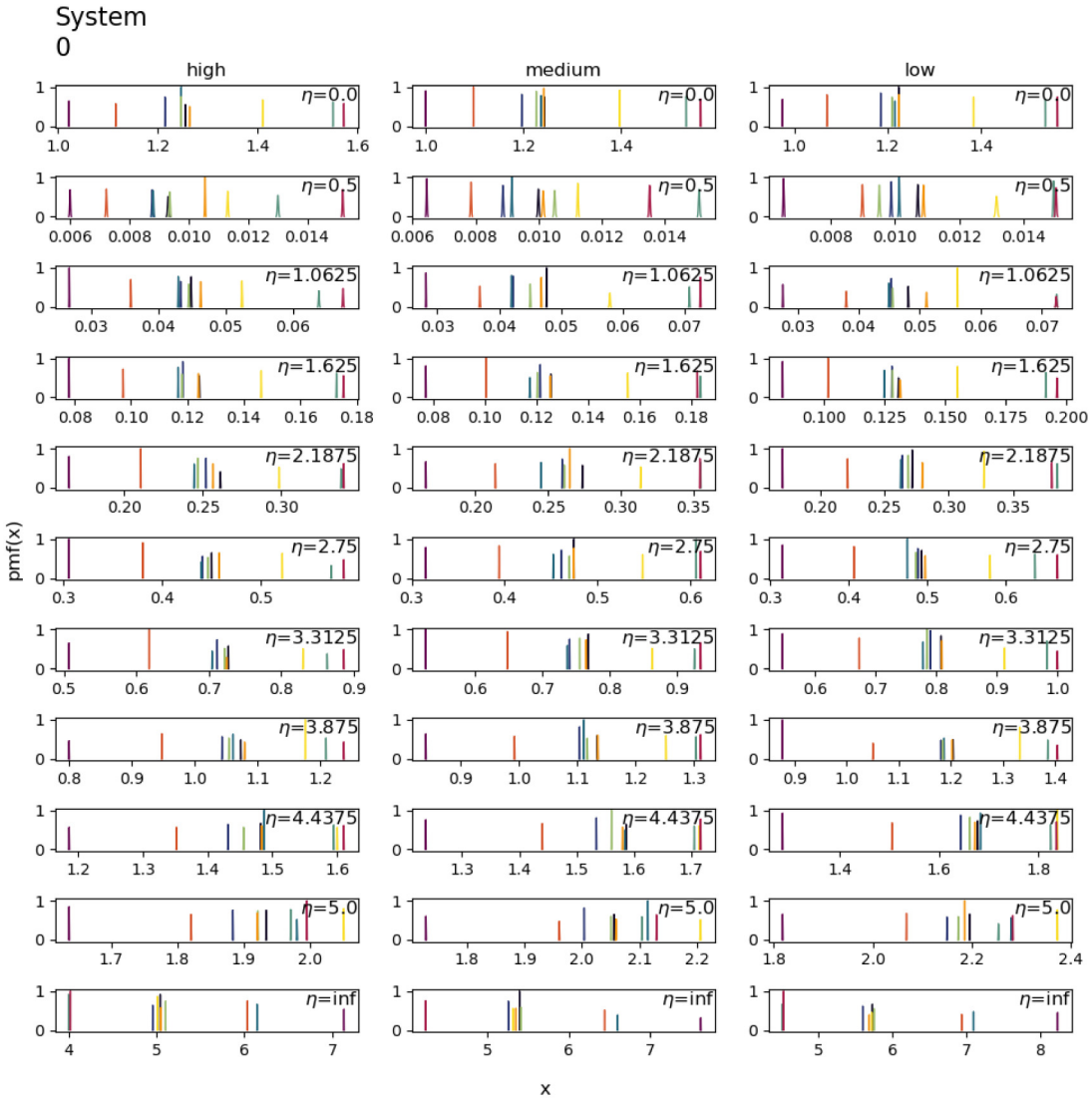
    # create bootstrap distribution
    driver = D_bar.mean(1).argmax()
    driverDist = NormalDist().from_samples(D_bar[driver])

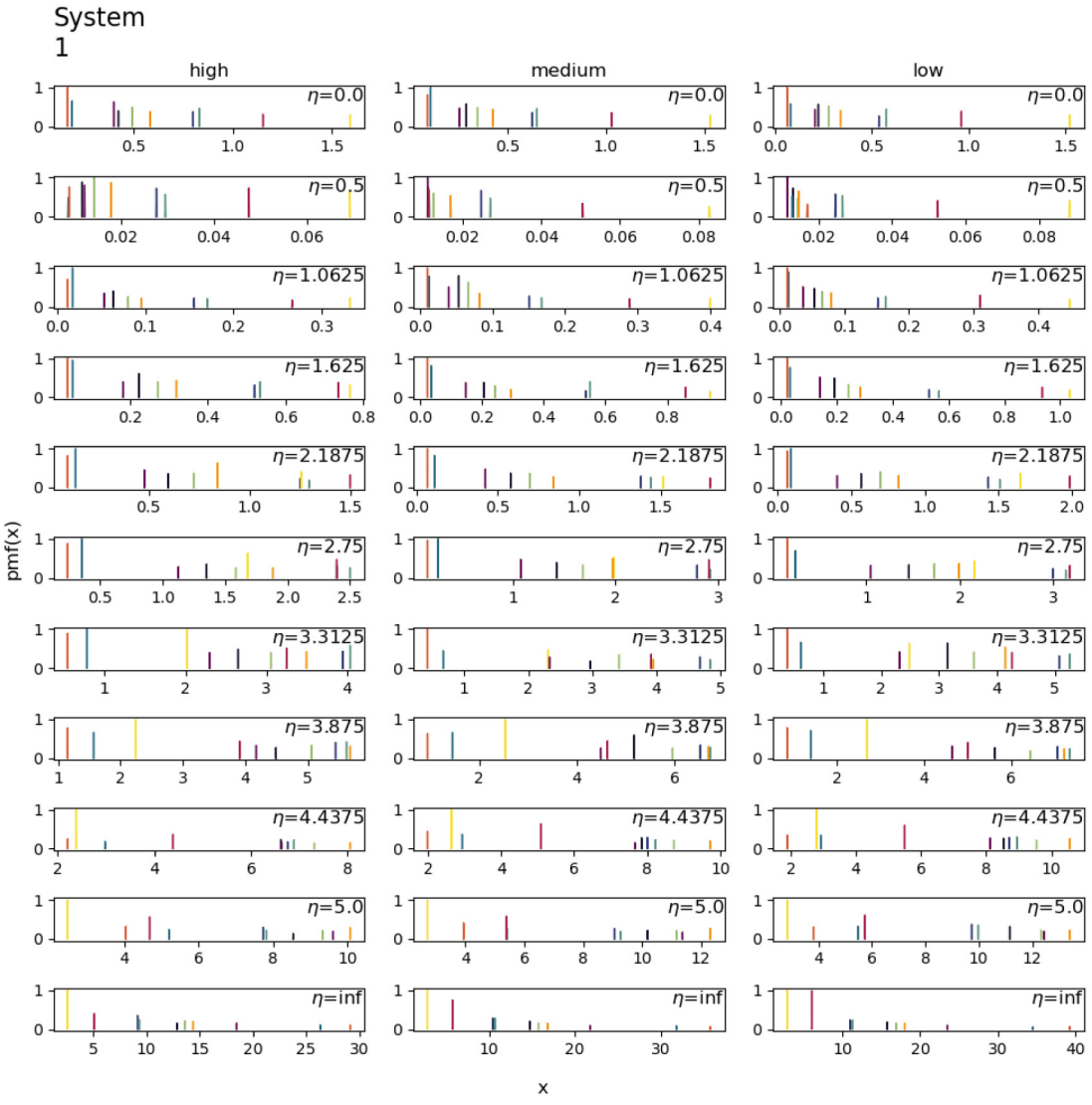
    # other nodes to consider
    drivers = {} # Lambda
    options = np.arange(n_features)
    for var in options:
        # fit distribution
        otherDist = NormalDist().from_samples(D_bar[var])
        # compute overlap
        overlap = driverDist.overlap(otherDist)
        if overlap > phi:
            drivers[var] = (otherDist.mean, otherDist.variance)
    return drivers
```

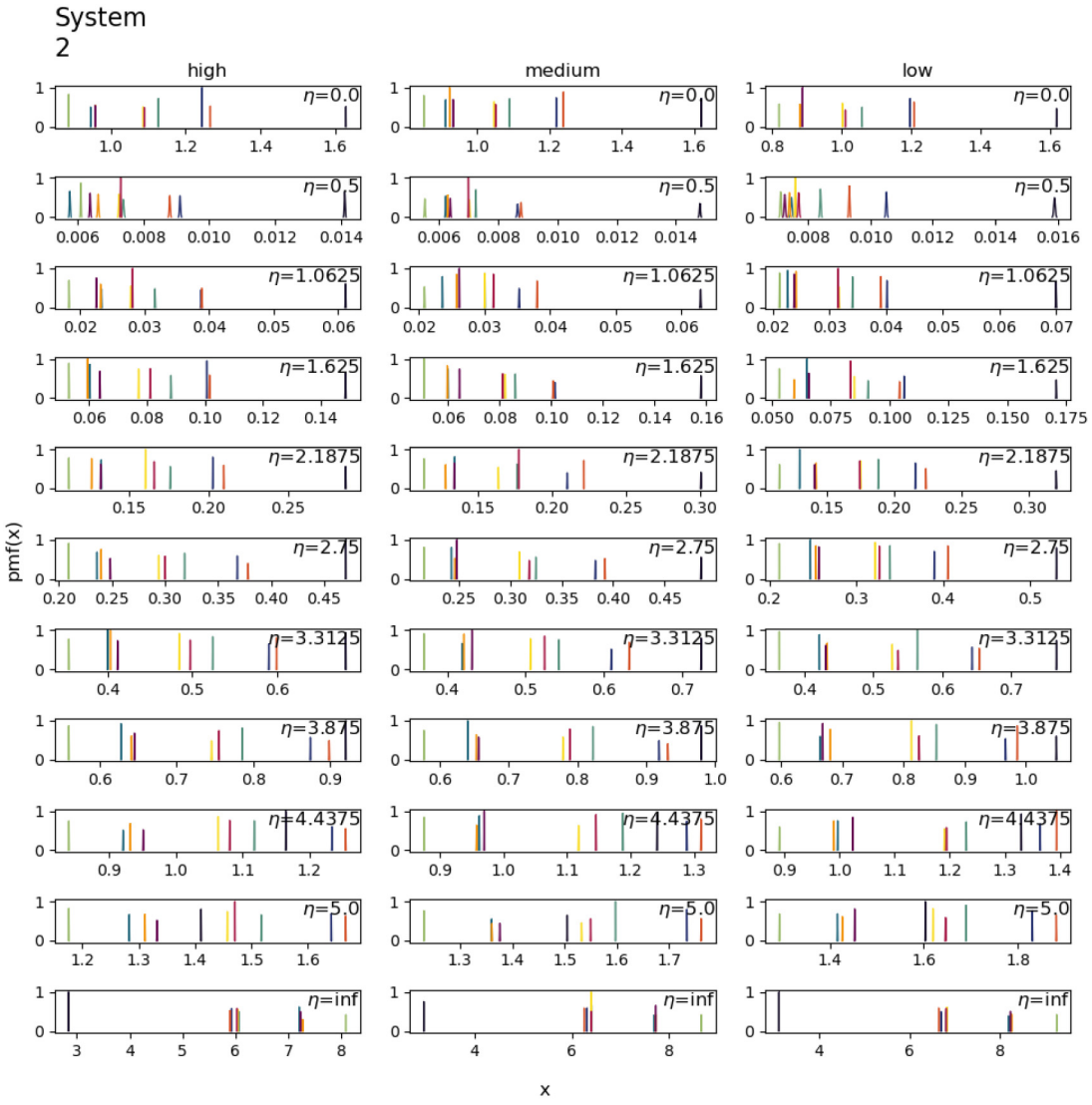
Listing 1: Driver node detection algorithm in python.

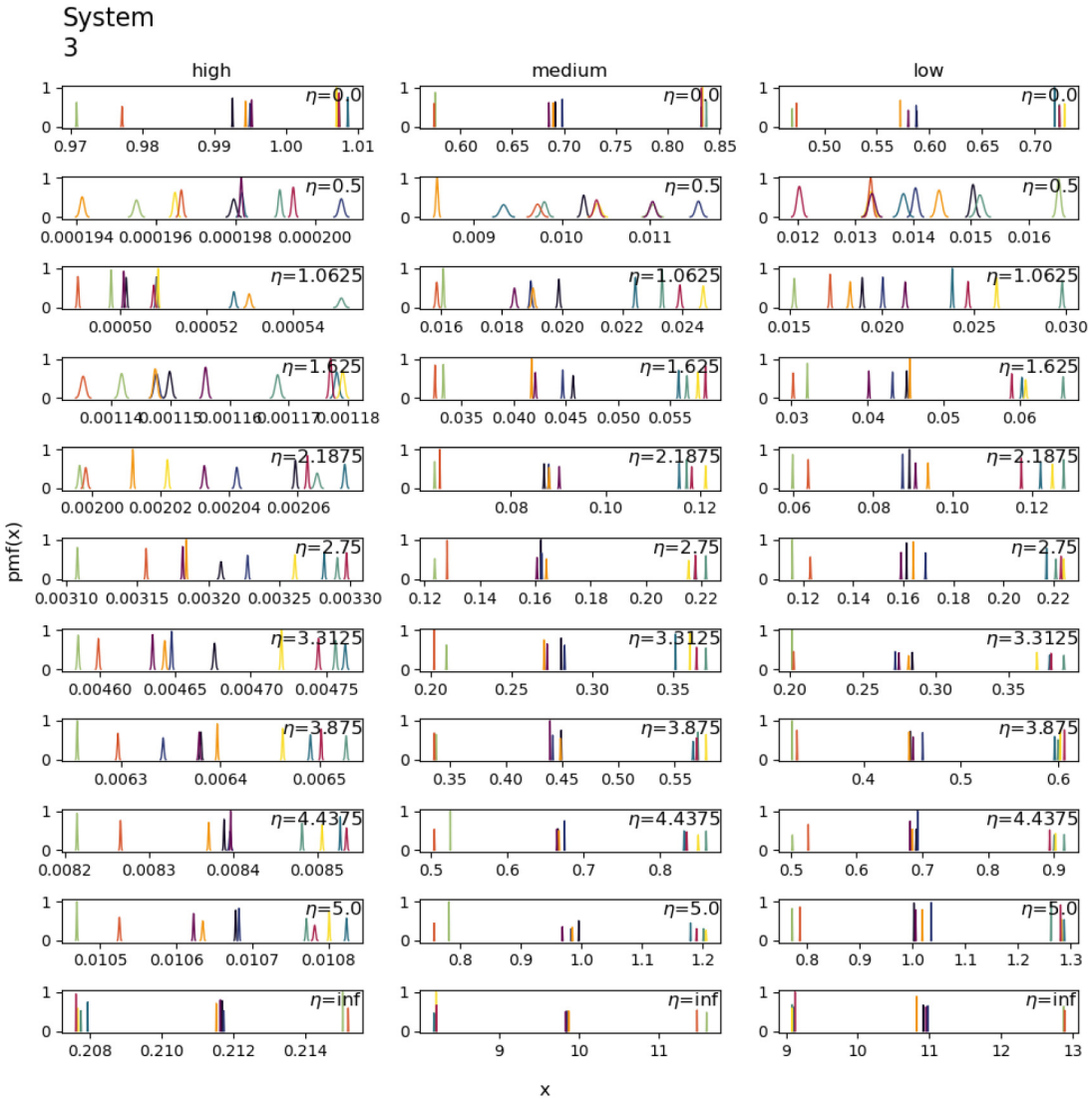
A.7.2. Erdös-Rényi networks

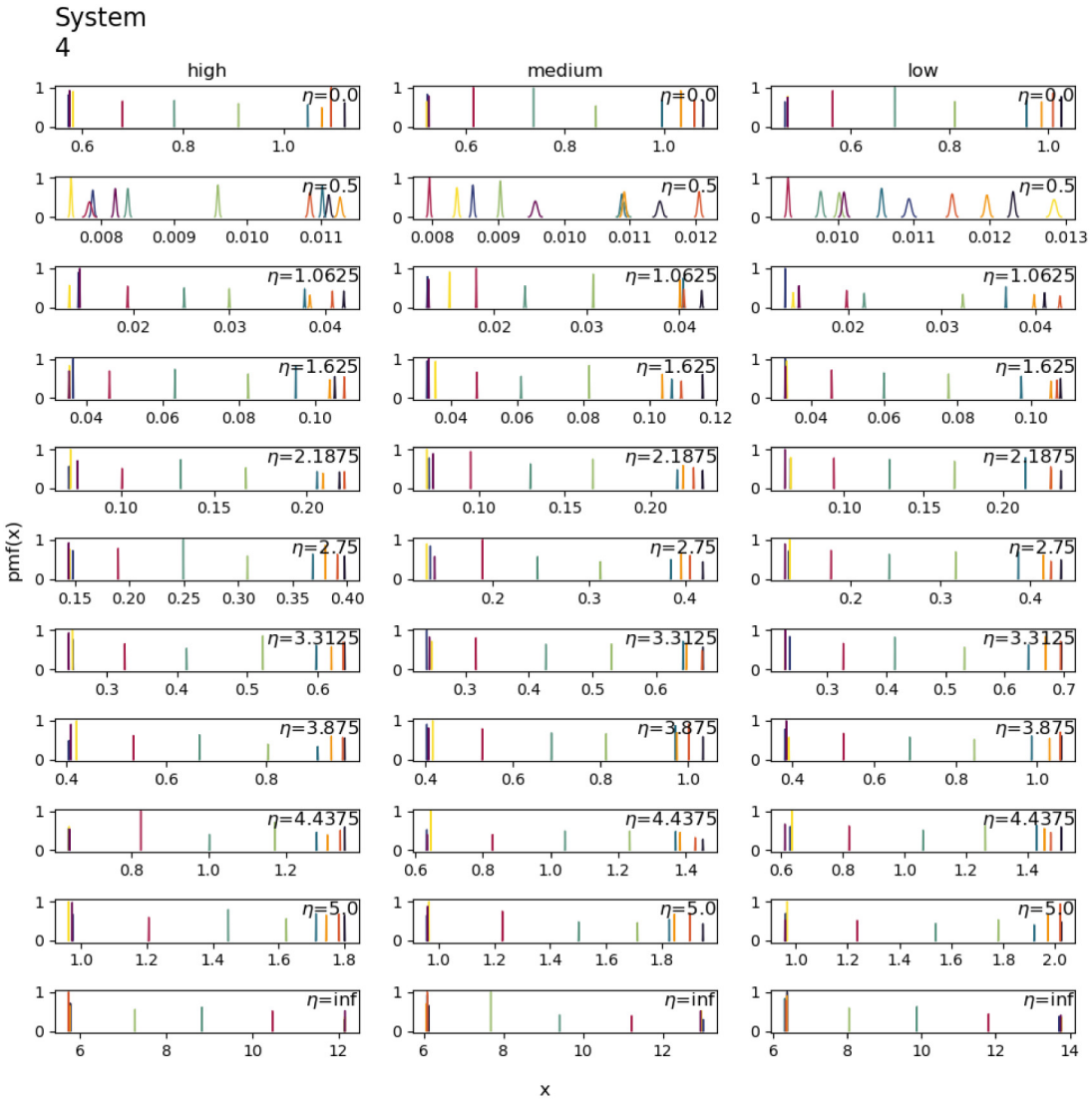
Bootstrap distribution kernel density estimates for Erdös-Rényi networks. X indicates area under the curve for mutual information or causal impact respectively. Note *x* here refers to the input variable which are the area under curves either for integrated mutual information ($\eta = 0$) or for causal impact $\eta > 0$.

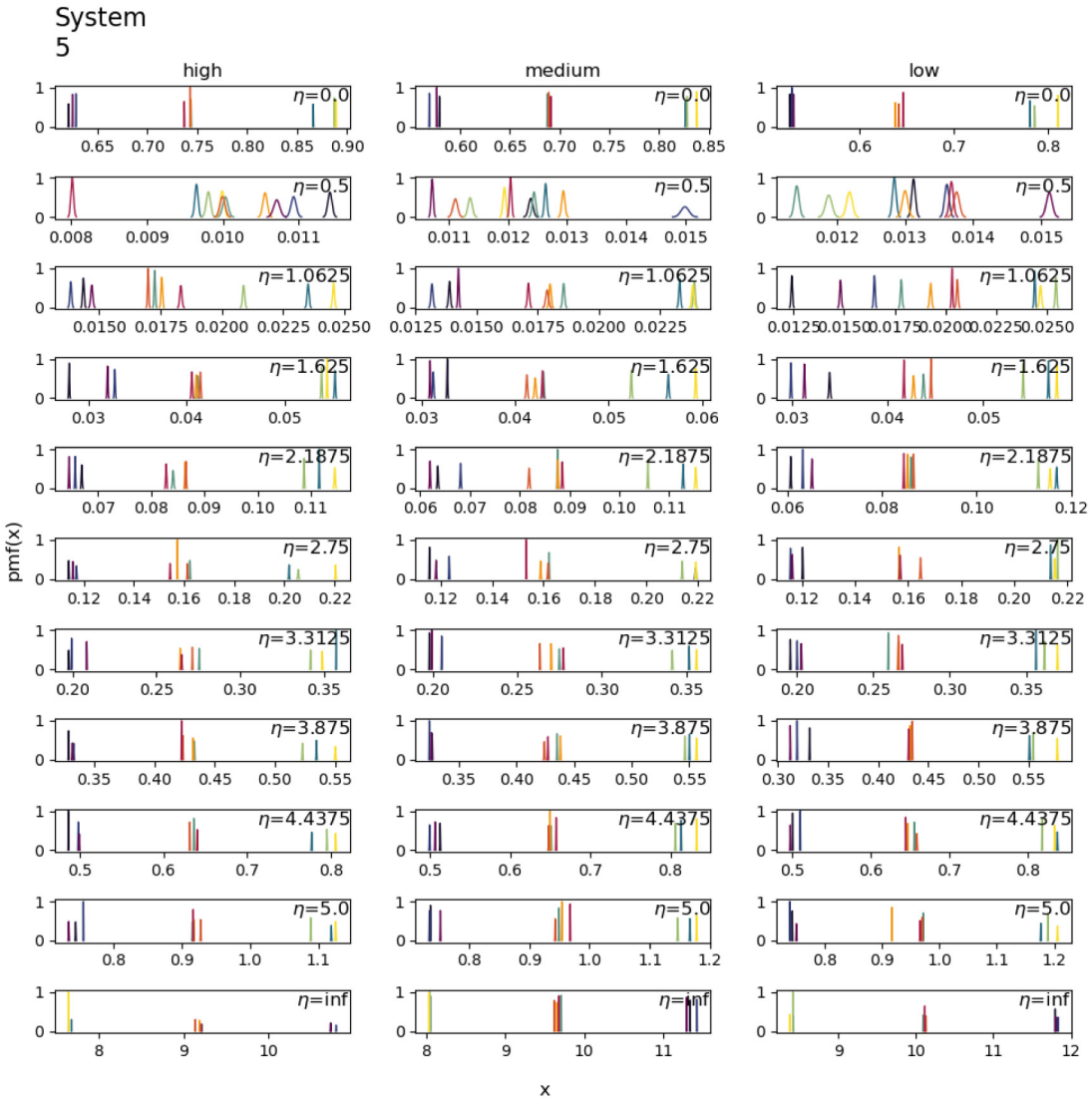

System
0
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

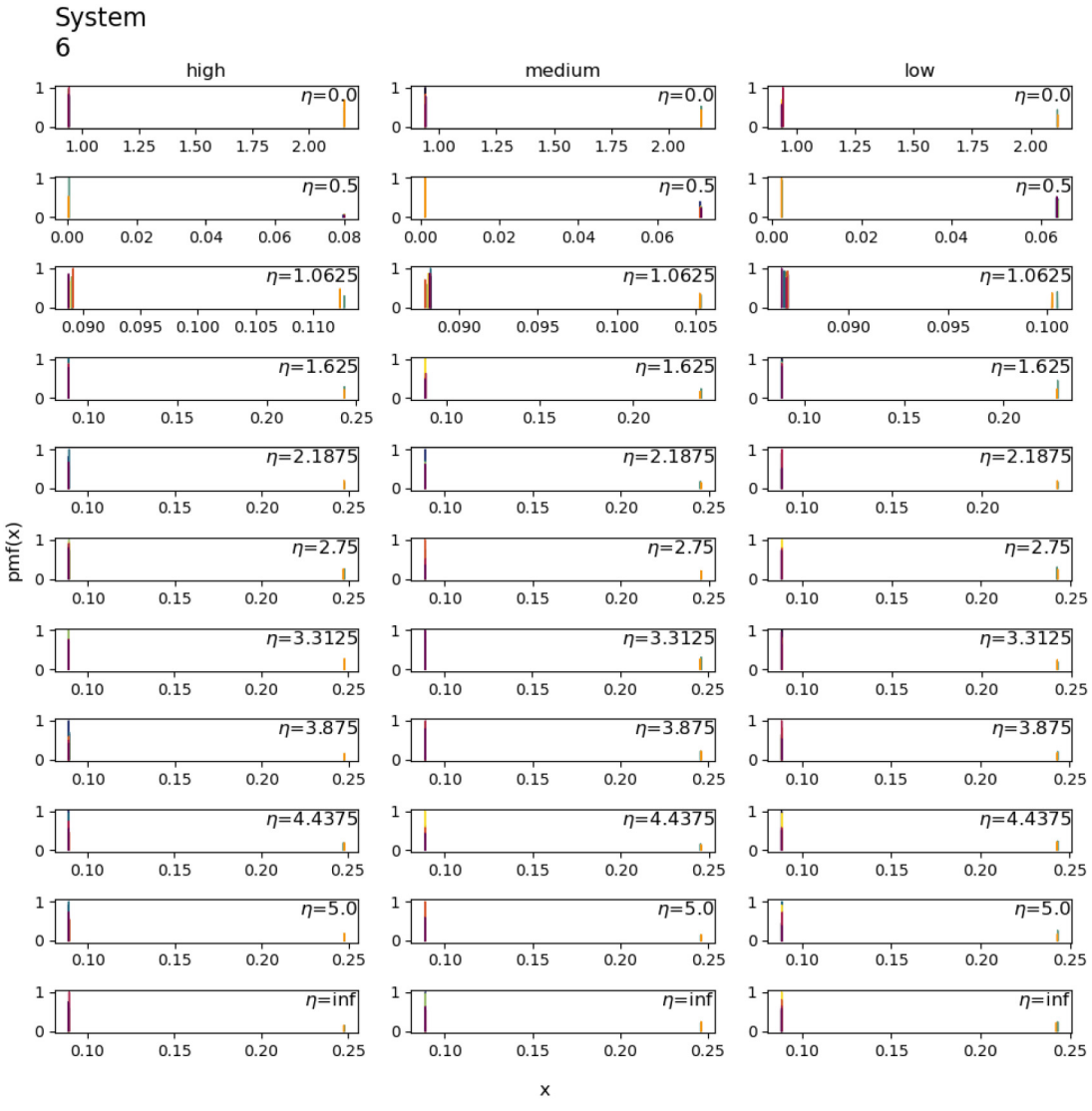

System
1
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

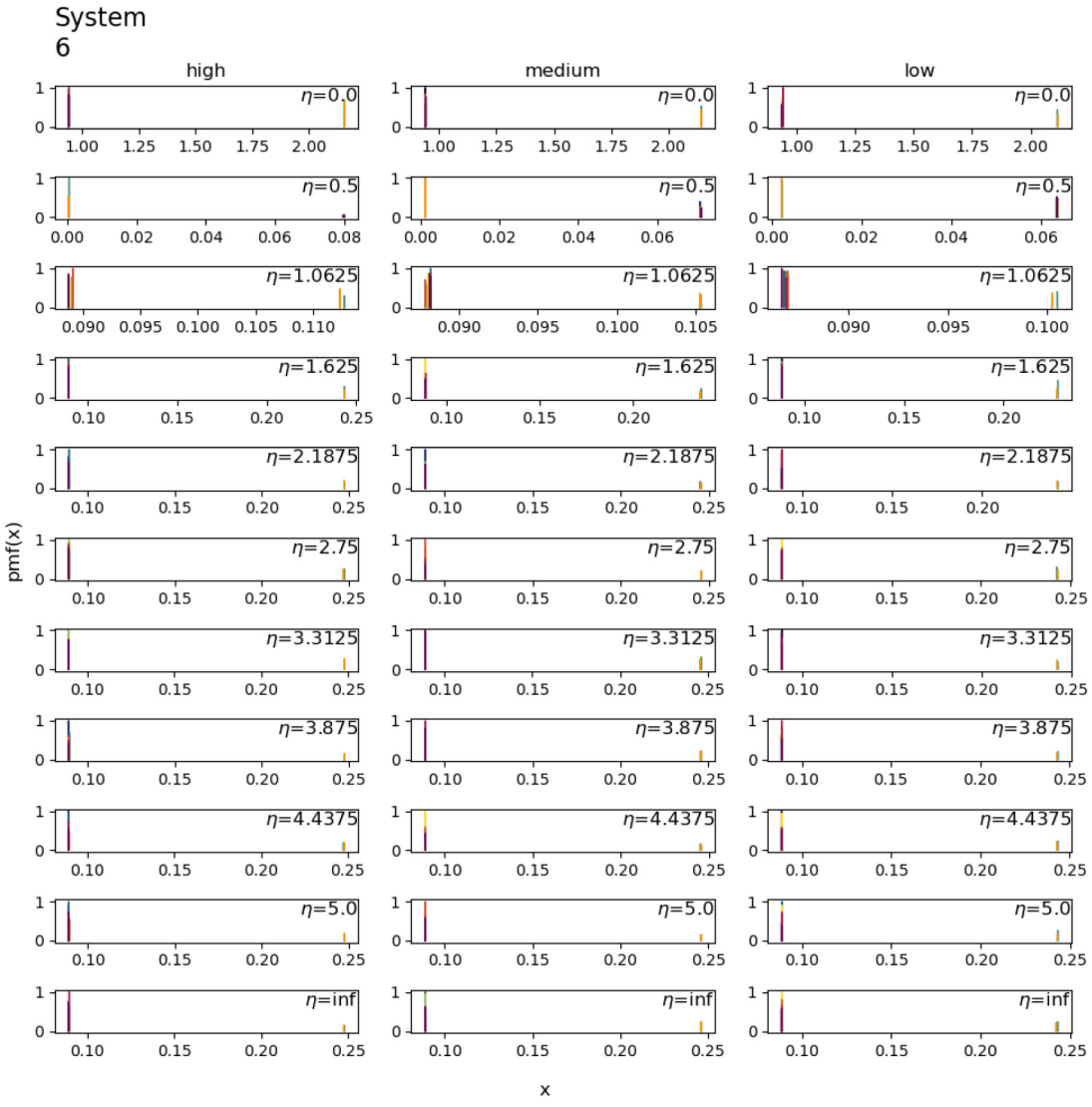

System 2
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

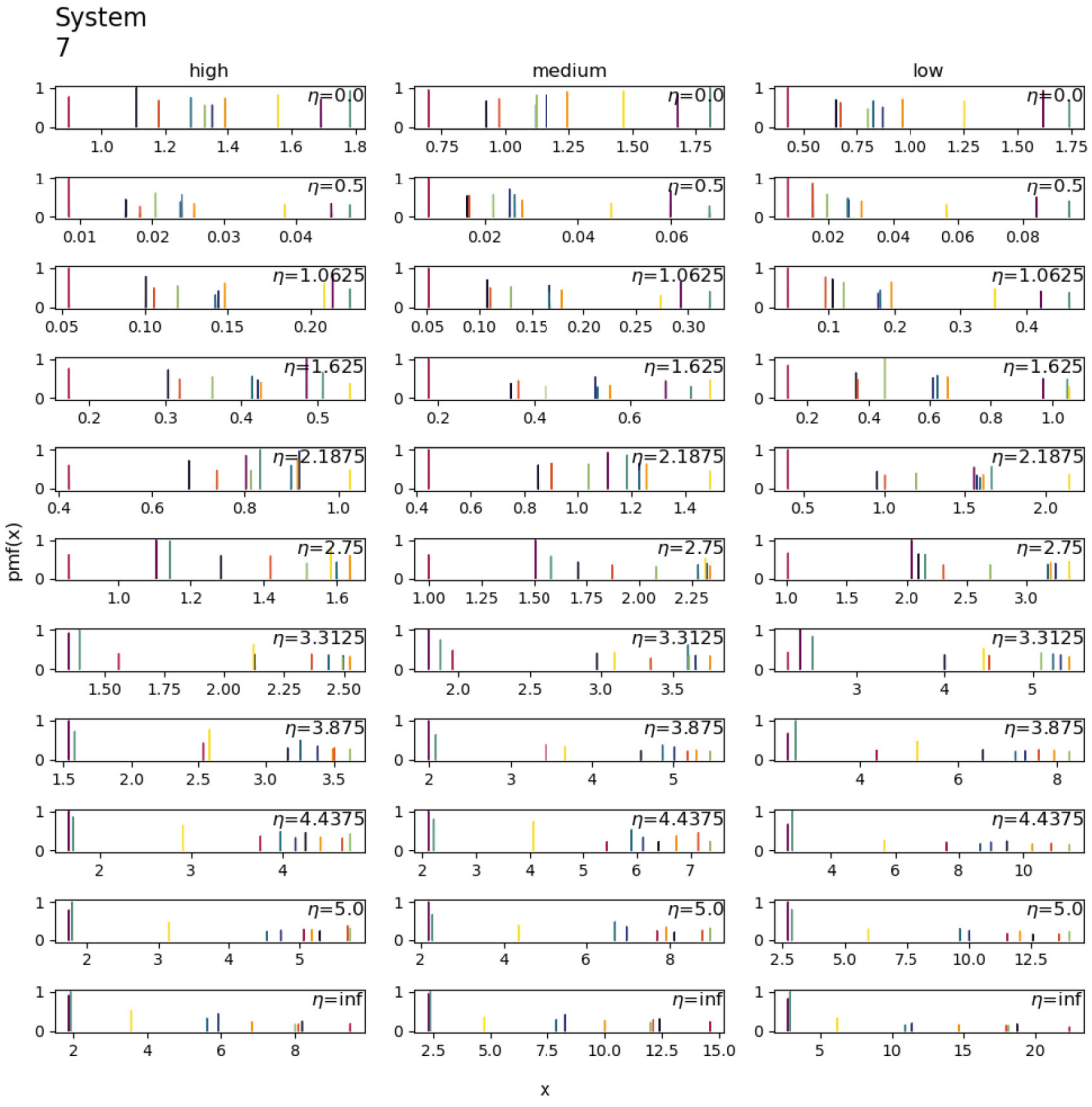

System
3
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

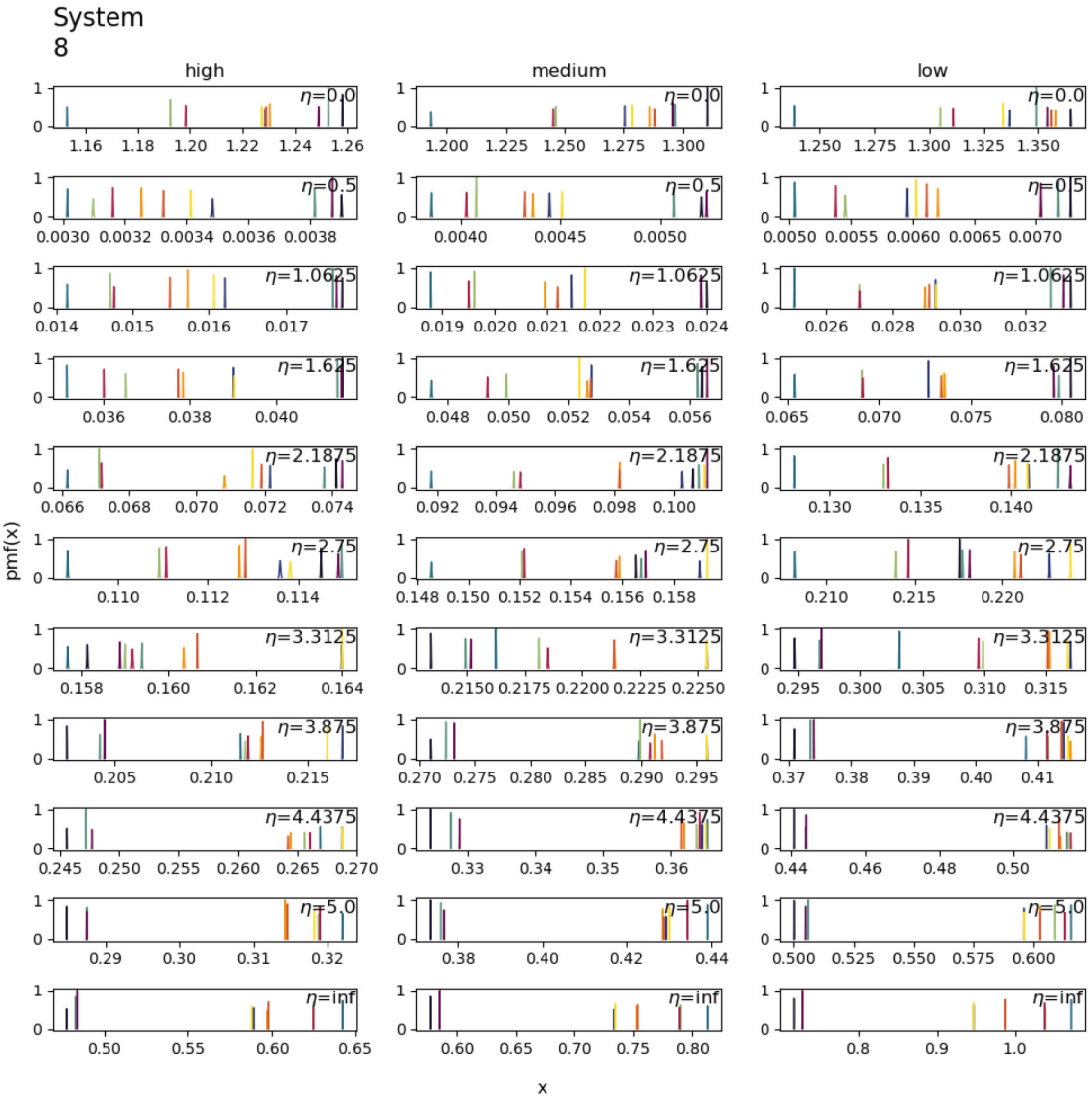

System
4
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

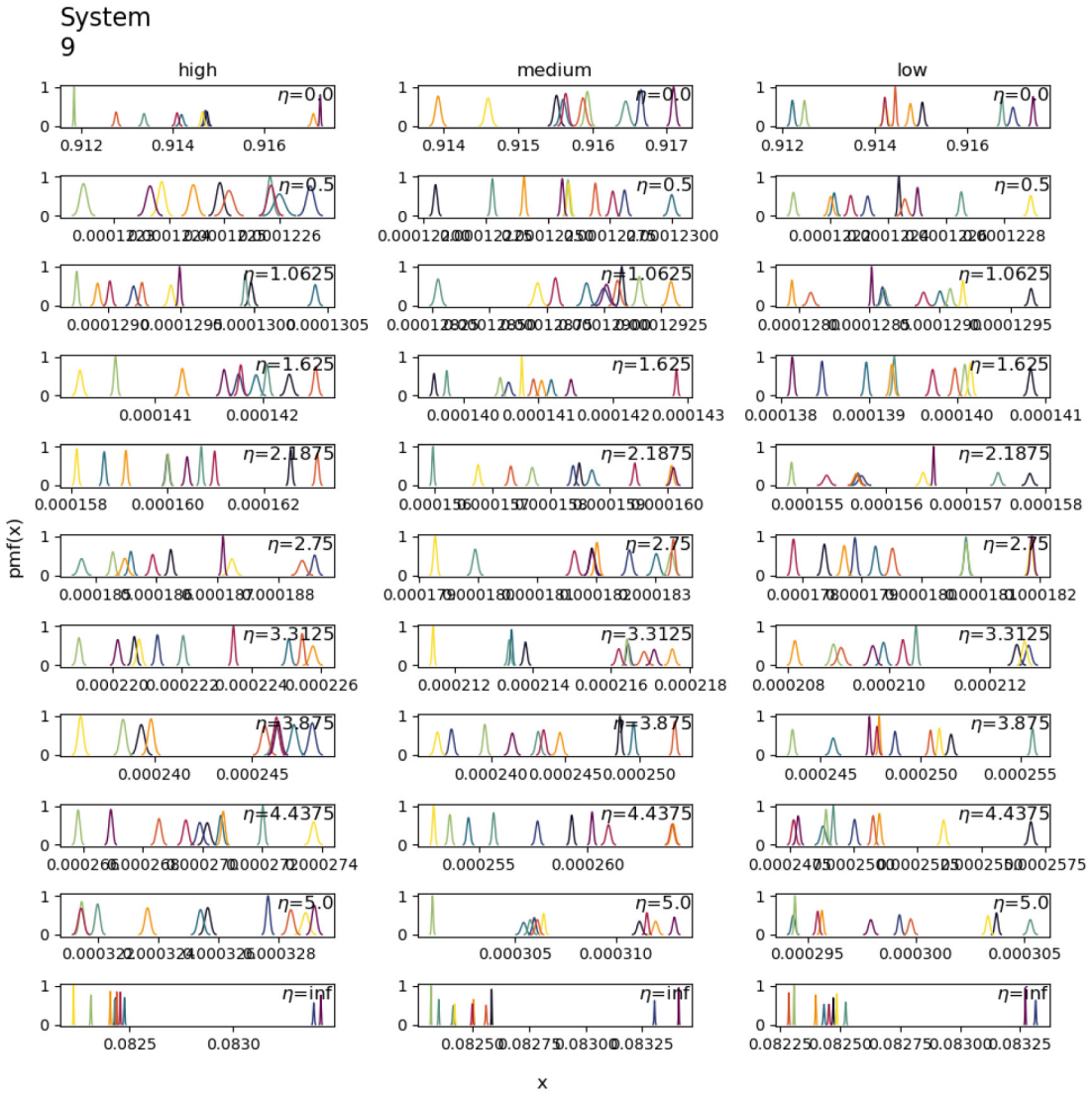

System
5
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

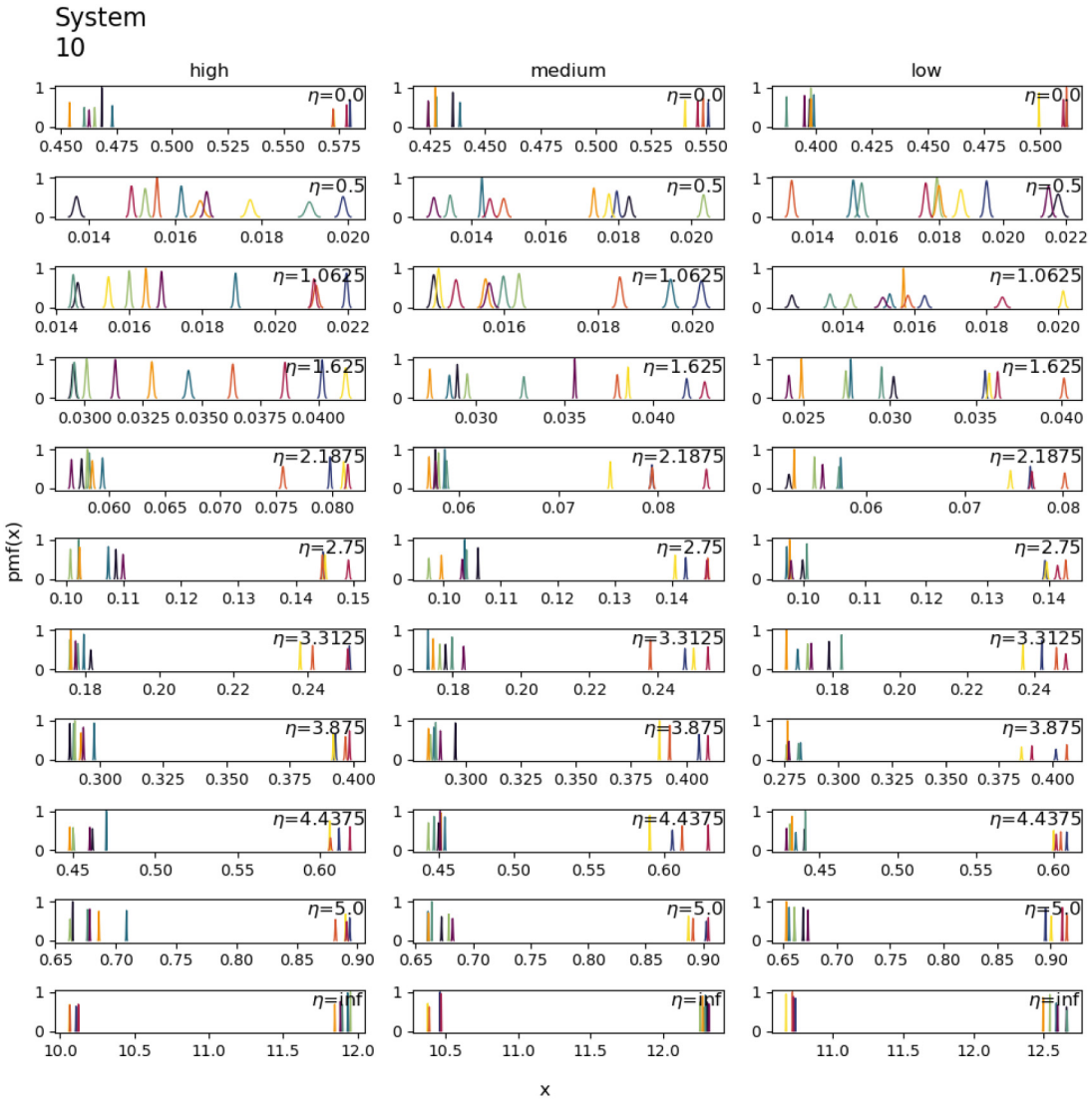

System
6
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

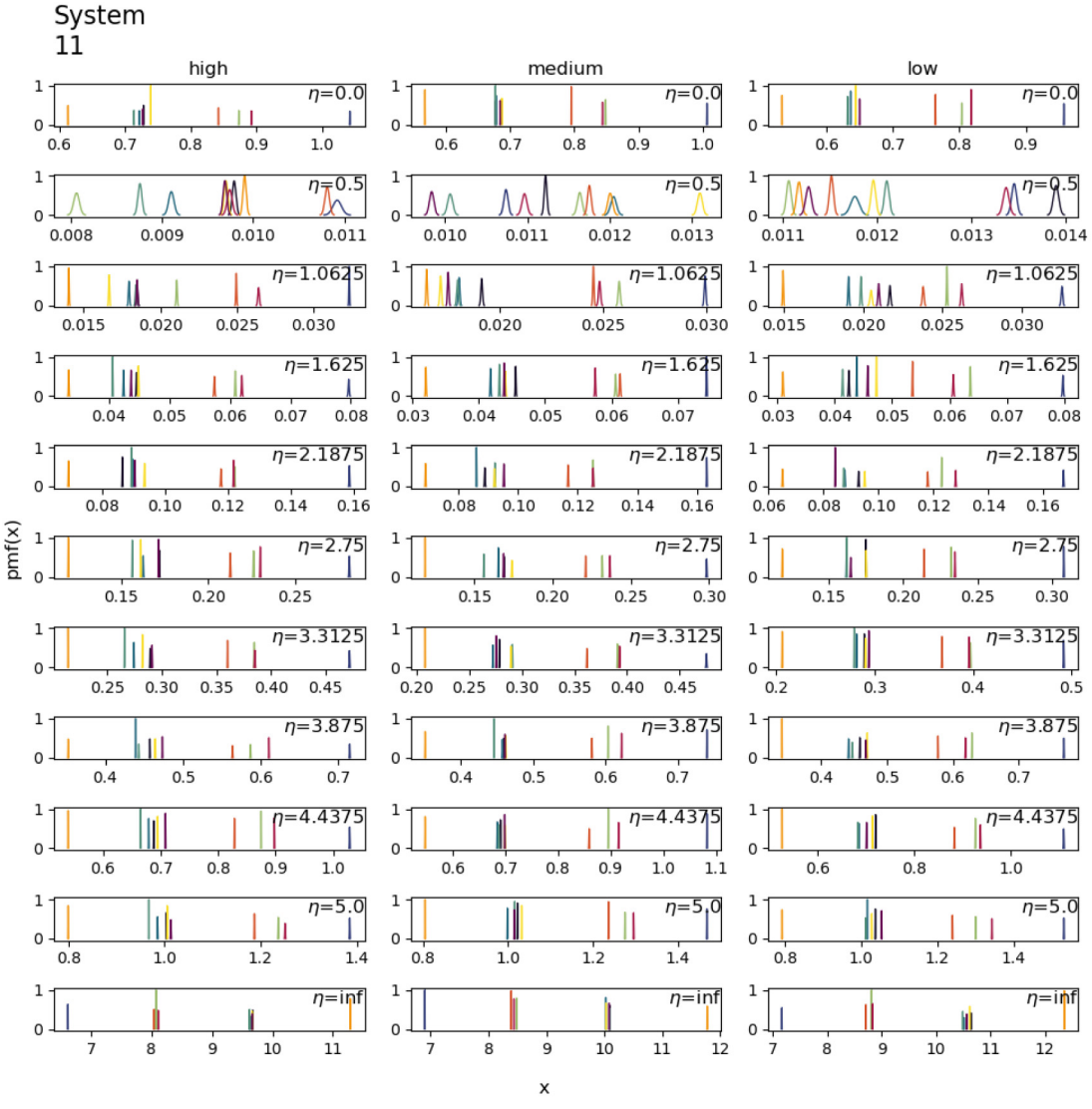

System
6
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

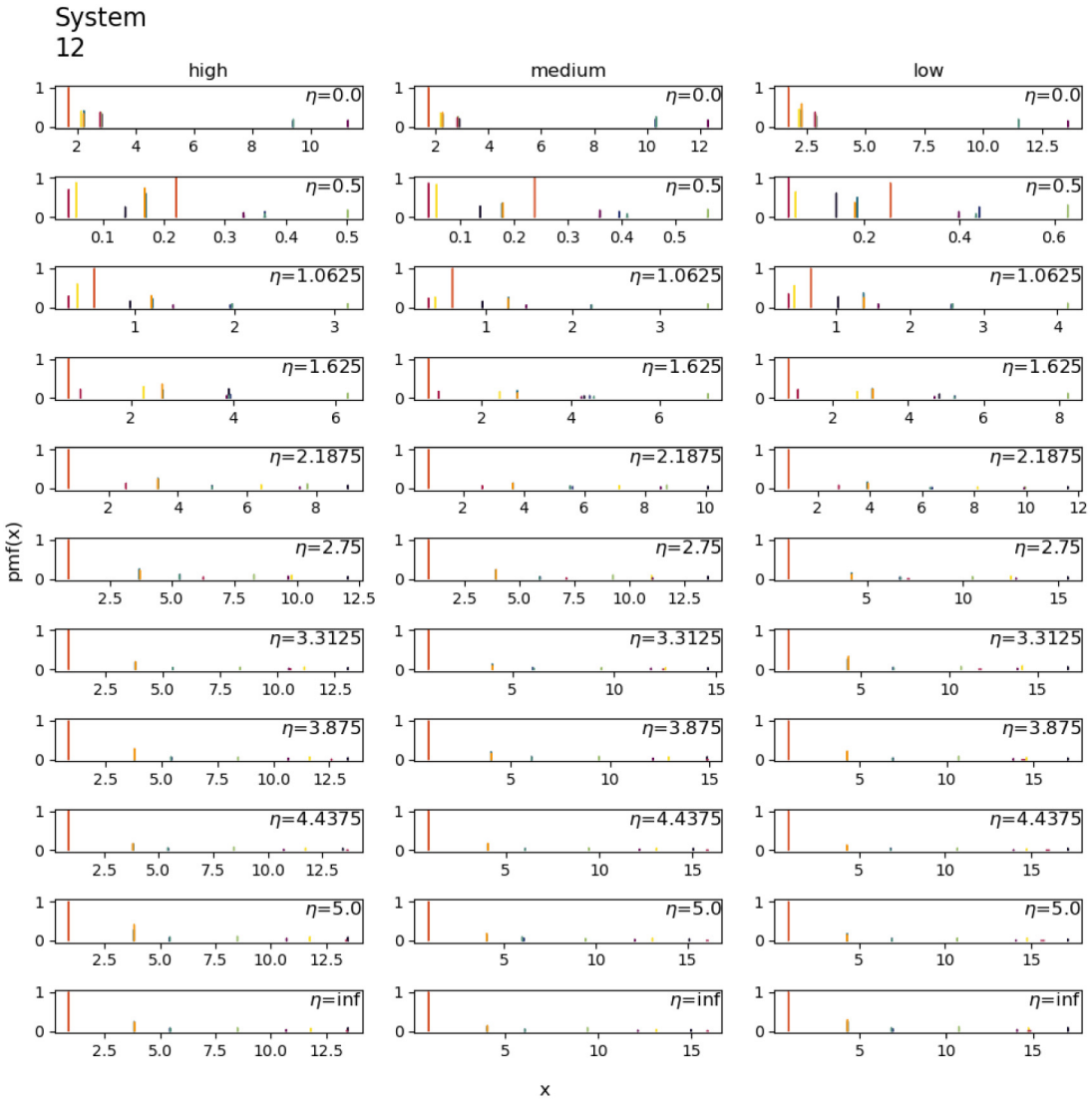

System
7
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

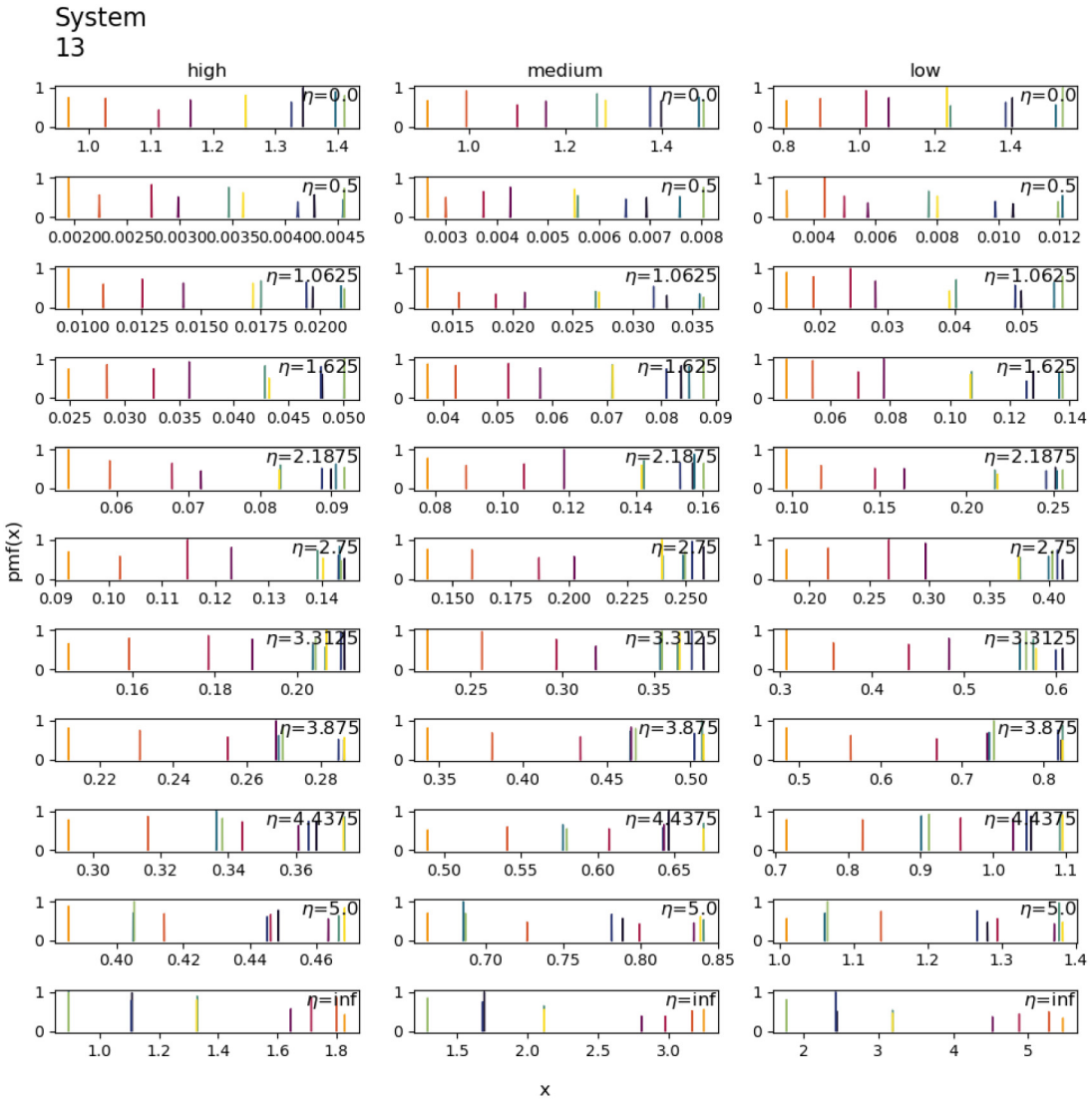

System
8
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

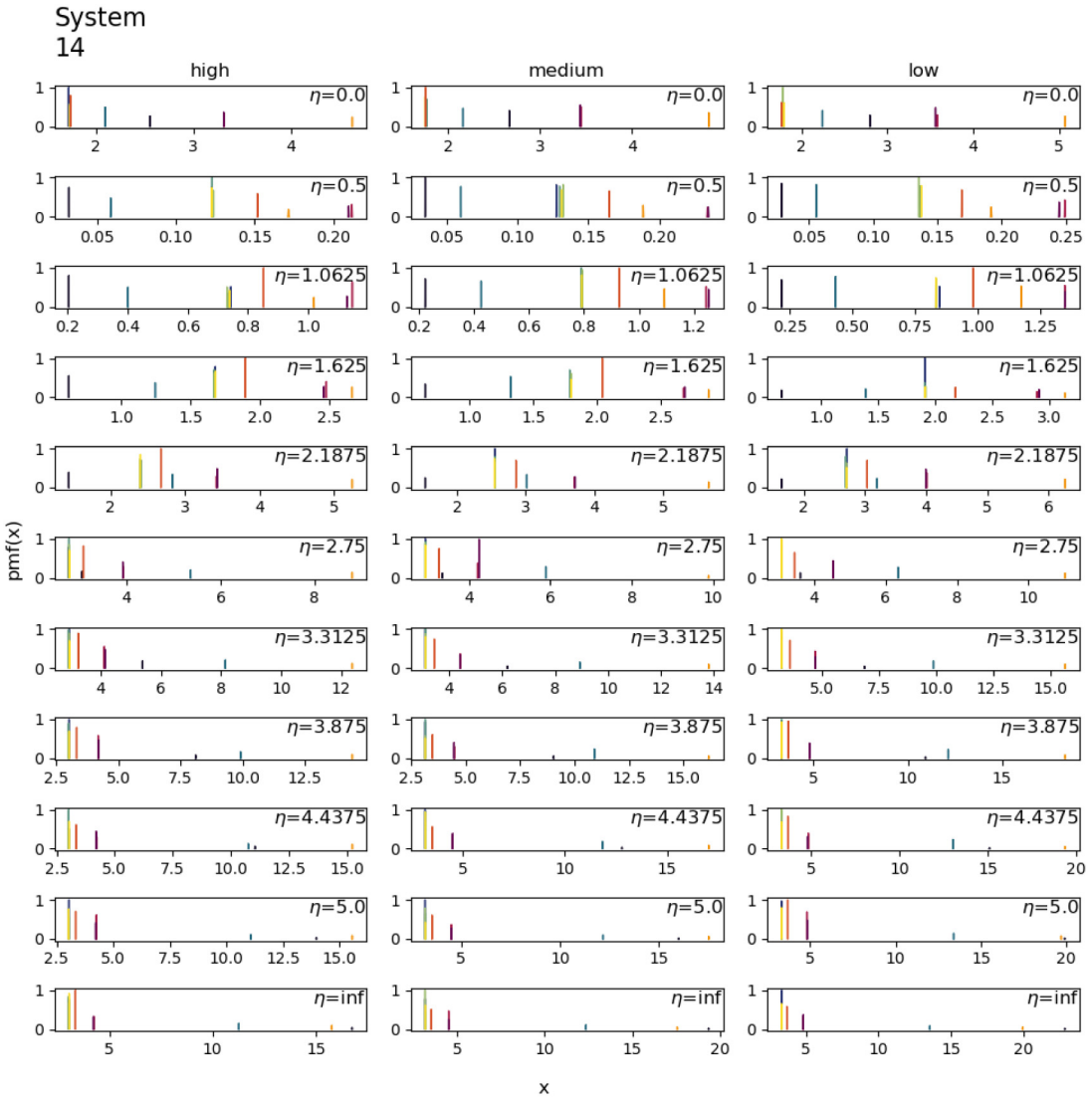

System
9
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

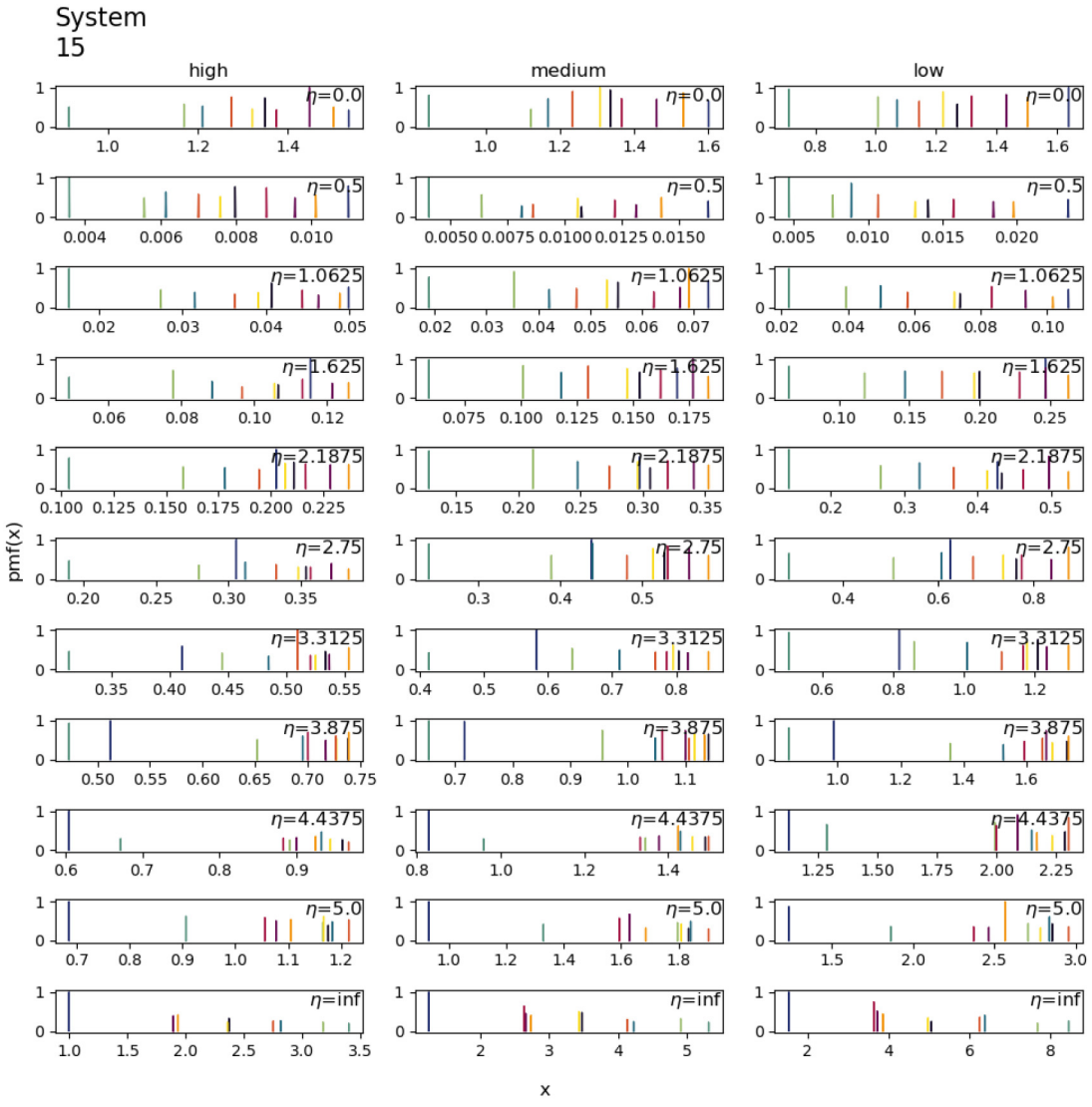

System
10
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x


System
11
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x


System
12
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x


System
13
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x


System
14
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x


System
15
high
medium
low
η=0.0
η=0.5
η=1.0625
η=1.625
η=2.1875
η=2.75
η=3.3125
η=3.875
η=4.4375
η=5.0
η=inf
pmf(x)
x

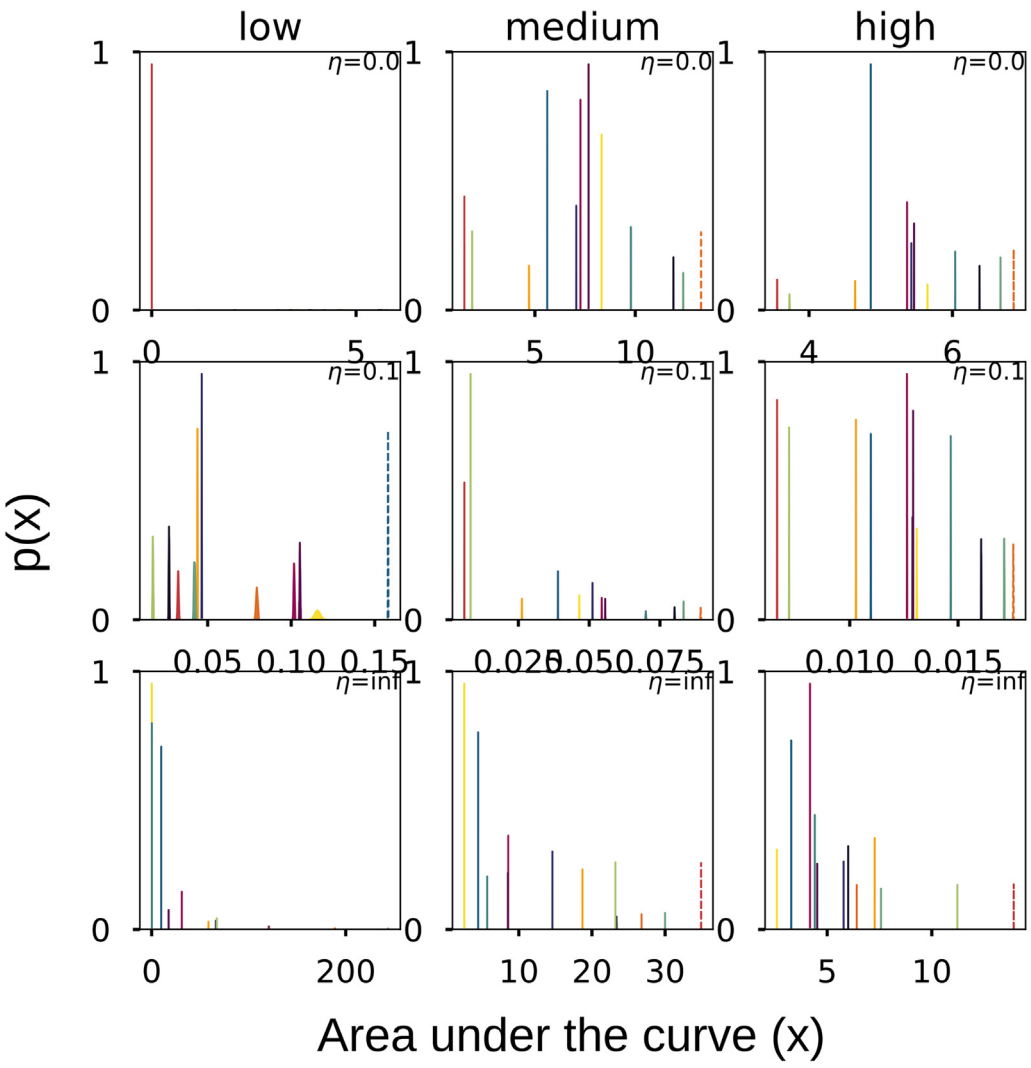


Fig. 11. Bootstrap distribution for psychosymptom network. Driver node is indicated by a dotted line.

A.7.3. Psycho networks

Note x here refers to the input variable which are the area under curves either for integrated mutual information ($\eta = 0$) or for causal impact $\eta > 0$ (see Fig. 11).

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physa.2022.126889.

Extra information containing the background of centrality metrics and its implicit assumptions; from a complex systems perspective, each centrality metric can be viewed to implicitly assume a dynamic on a graph. This has profound implications for analyzing the output of many metrics in network science. The document also contains background information on the real-world network of psychosymptoms; from where it was obtained and how the data was collected. Importantly, the supplementary information contains additional analysis of time assymmetries in information flows for directed graphs. Lastly, it contains bootstrap distributions for the estimates used in the study.

References

[1] S. Hawking, Unified theory is getting closer, hawking predicts, San Jose Mercury News (2000).
[2] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, 393, 1998, p. 3.
[3] R. Pastor-Satorras, J. Rubí, A. Diaz-Guilera (Eds.), Statistical mechanics of complex networks, Lecture Notes in Physics, Springer, New York, 2003.

[4] R. Thomas, D. Thieffry, M. Kaufman, S. de Chimie-Physique, Dynamical behaviour of biological regulatory networks–I. biological role of feedback loops and practical use of the concept of the loop-characteristic state. en, 30.
[5] P. Ianishi, O.A. Gonzattoa Junior, M.J. Henriques, D.C. do Nascimento, G.K. Mattar, P.L. Ramos, A. Ara, F. Louzada, Probability on graphical structure: a knowledge-based agricultural case, Ann. Data Sci. (2020).
[6] M. Šikić, A. Lančić, N. Antulov-Fantulin, H. Štefančić, Epidemic centrality - is there an underestimated epidemic impact of network peripheral nodes?, Eur. Phys. J. B 86 (2013) 1–23, arXiv:1110.2558.
[7] B. Barzel, A.-L. Barabási, Universality in network dynamics, Nat. Phys. 9 (2013) 673–681.
[8] U. Harush, B. Barzel, Dynamic patterns of information flow in complex networks, Nature Commun. 8 (2017) 1–11.
[9] J. Pearl, Causality, second ed., 2000, pp. 1–386.
[10] J. Woodward, Interventionism and causal exclusion, Philos. Phenomenol. Res. 91 (2015) 303–347.
[11] G. Schamberg, W. Chapman, S.-P. Xie, T.P. Coleman, Direct and indirect effects—an information theoretic perspective, Entropy 22 (2020) 854.
[12] W.X. Wang, Y.C. Lai, C. Grebogi, Data based identification and prediction of nonlinear and complex dynamical systems, Phys. Rep. 644 (2016) 1–76, arXiv:1704.08764.
[13] J.M. Mooij, T. Heskes, D. Janzing, B. Scholkopf, On Causal Discovery with Cyclic Additive Noise Models, 10.
[14] P. Forré, J.M. Mooij, Causal Calculus in the Presence of Cycles, Latent Confounders and Selection Bias, 10.
[15] F. Dablander, M. Hinne, Node centrality measures are a poor substitute for causal inference, Sci. Rep. 9 (2019) 1–13.
[16] S. Epskamp, D. Borsboom, E.I. Fried, Estimating psychological networks and their accuracy: A tutorial paper, Behav. Res. Methods 50 (2018) 195–212.
[17] J. Ladyman, J. Lambert, K. Wiesner, What is a complex system?, Eur. J. Philos. Sci. 3 (2013) 33–67.
[18] P.A. Stokes, P.L. Purdon, A study of problems encountered in granger causality analysis from a neuroscience perspective, Proc. Natl. Acad. Sci. 114 (2017) E7063–E7072.
[19] T.M. Cover, J.A. Thomas, Elements of information theory, Elements Inf. Theory (2005).
[20] J.B. Kinney, G.S. Atwal, Equitability, mutual information, and the maximal information coefficient, Proc. Natl. Acad. Sci. 111 (2014) 3354–3359.
[21] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85 (2) (2000) 461–464, arXiv: nlin/0001042v1.
[22] J.T. Lizier, B. Flecker, P.L. Williams, Towards a synergy-based approach to measuring information modification, in: IEEE Symposium on Artificial Life (ALIFE), 2013, pp. 43–51 arXiv:1303.3440.
[23] A. Sensoy, C. Sobaci, S. Sensoy, F. Alali, Effective transfer entropy approach to information flow between exchange rates and stock markets, Chaos Solitons Fractals 68 (2014) 180–185.
[24] M. Songhorzadeh, K. Ansari-Asl, A. Mahmoudi, Two step transfer entropy – an estimator of delayed directional couplings between multivariate EEG time series, Comput. Biol. Med. 79 (2016) 110–122.
[25] M. Wibral, J.T.L. Editors, S. Kelso, Directed Information Measures in Neuroscience, 2014.
[26] N. Ay, D. Polani, Information flows in causal networks, Adv. Complex Syst. 11 (2008) 17–41.
[27] J. Sun, E.M. Bollt, Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings, Physica D 267 (2014) 49–57, arXiv:1504.03769.
[28] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M.D. Mahecha, J. Muñoz-Marí, E.H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schö, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, J. Zscheischler, Inferring causation from time series in earth system sciences, Nature Commun. 10 (2019) 1–13.
[29] S. Li, Y. Xiao, D. Zhou, D. Cai, Causal inference in nonlinear systems: Granger causality versus time-delayed mutual information, Phys. Rev. E 97 (2018) 052216.
[30] D. Janzing, D. Balduzzi, M. Grosse-Wentrup, B. Schö, Quantifying causal influences, Ann. Statist. 41 (2013) 2324–2358, arXiv:1203.6502v2.
[31] R.G. James, N. Barnett, J.P. Crutchfield, Information flows? a critique of transfer entropies, Phys. Rev. Lett. 116 (2016) 1–6, arXiv:1512.06479.
[32] R. Quax, O. Har-Shemesh, P.M. Sloot, Quantifying synergistic information using intermediate stochastic variables, Entropy 19 (2017) 7–10, arXiv:1602.01265.
[33] R. Quax, A. Apolloni, P.M. a Sloot, The diminishing role of hubs in dynamical processes on complex networks., J. R. Soc. Interface / R. Soc. 10Q (88) (2013) 20130568, arXiv:1111.5483.
[34] U. Alon, Network motifs: Theory and experimental approaches, Nature Rev. Genet. 8 (2007) 450–461.
[35] E.I. Fried, C. Bockting, R. Arjadi, D. Borsboom, M. Amshoff, A. Cramer, S. Epskamp, F. Tuerlinckx, D. Carr, M. Stroebe, From loss to loneliness: the relationship between bereavement and depressive symptoms, J. Abnormal Psychol. 124 (2015) 256–265.
[36] J. Feng, J. Jost, M.-P. Qian (Eds.), Networks: From Biology to Theory, Springer, London, 2007.
[37] W.K. Hastings, Monte Carlo Sampling Methods Using Markov Chains and Their Applications, 1970, p. 13.
[38] G. Ódor, Universality classes in nonequilibrium lattice systems, Rev. Modern Phys. 76 (2004) 663–724.
[39] X. Zhang, J. Han, W. Zhang, An efficient algorithm for finding all possible input nodes for controlling complex networks, Sci. Rep. 7 (1) (2017) 1–8.
[40] Y.Y. Liu, A.L. Barabási, Control principles of complex systems, Rev. Modern Phys. 88 (3) (2016) 1–61, arXiv:1508.05384.
[41] A.J. Gates, L.M. Rocha, Control of complex networks requires both structure and dynamics, Sci. Rep. 6 (2016) 1–11, arXiv:1509.08409.
[42] G. Yan, P.E. Vé, E.K. Towlson, Y.L. Chew, D.S. Walker, W.R. Schafer, A.L. Barabási, Network control principles predict neuron function in the caenorhabditis elegans connectome, Nature 550 (2017) 519–523.
[43] R. Quax, O. Har-shemesh, S. Thurner, P.M.A. Sloot, Stripping syntax from complexity : an information-theoretical perspective on complex systems, march, 2016, ArXiv Preprint, arXiv:1603.03552.
[44] R. Quax, D. Kandhai, P.M. Sloot, Information dissipation as an early-warning signal for the lehman brothers collapse in financial time series, Sci. Rep. 3 (2013) 1898.
[45] R.G. James, J.P. Crutchfield, Multivariate dependence beyond Shannon information, Entropy 19 (2017) arXiv:1609.01233.
[46] C.D. Vana Borkulo, D. Borsboom, S. Epskamp, T.F. Blanken, L. Boschloo, R.A. Schoevers, L.J. Waldorp, A new method for constructing networks from binary data, Sci. Rep. 4 (2014) 1–10, arXiv:1112.5635.
[47] S. Epskamp, J. Kruis, M. Marsman, Estimating psychopathological networks: Be careful what you wish for, PLoS ONE 12 (2017) 1–13, arXiv:1604.08045.
[48] R.J. Glauber, Time-dependent statistics of the Ising model, J. Math. Phys. 4 (1963) 294–307.
[49] S. Panzeri, R. Senatore, M. Montemurro, R. Petersen, Correcting for the sampling bias problem in spike train information measures, J. Neurophysiol. 98 (2007) 1064–1072.
[50] H. Matsuda, A. Sasaki, Global stability of an SIR epidemic model with time delays, 1994, pp. 251–268.
[51] I. Harvey, T. Bossomaier, Time out of joint: attractors in asynchronous random boolean networks, in: Proceedings of the Fourth European Conference on Artificial Life, 1997, pp. 67–75 arXiv:1104.0592.
[52] S.P. Borgatti, M.G. Everett, A graph-theoretic perspective on centrality, Social Networks 28 (2006) 466–484.
[53] S.P. Borgatti, Centrality and network flow, Social Networks 27 (2005) 55–71.

[54] L.J. Waldorp, V.D. Schmittmann, D. Borsboom, A. Cramer, R.A. Kievit, S. Epskamp, Deconstructing the construct: A network perspective on psychological phenomena, New Ideas Psychol. 31 (2011) 43–53.
[55] D. Borsboom, A.O. Cramer, V.D. Schmittmann, S. Epskamp, L.J. Waldorp, The small world of psychopathology, PLoS ONE 6 (2011).
[56] Nature guide to authors: First paragraphs for letters, How to construct a nature summary paragraph, Nature 435 (2005) 114–1148.
[57] L.F. Bringmann, T. Elmer, S. Epskamp, W. Krause, D. Schoch, M. Wichers, E. Snippe, What do centrality measures measure in psychological networks ?, 2018, pp. 1–34.
[58] L. Freeman, Centrality in social networks, Social Networks 1 (1979) 215–239, arXiv:cond-mat/0112110.
[59] P. Debye, P. Scherrer, Werk Übergeordnetes werk, Nachr. Ges. Wiss. Göttingen, Math.-Physik. Klasse 2 (1918) 101–120.
[60] G.F. Frobenius, Über matrizen aus nicht negativen elementen, Sitzungsberichte Der Preussischen Akademie Der Wissenschaften Zu Berlin (1912) 456–477.
[61] A.N. Langville, C.D. Meyer, A survey of eigenvector methods for web information retrieval, SIAM Rev. 47 (2005) 135–161.