Short-lived correlations are essential for metastable transitions

Casper van Elteren, Rick Quax, Peter Sloot

April 6, 2022

Noise plays a fundamental role in many real-world systems from neural behavior, cellular regulation, physics and social systems. It is well established that even weak noise can result in large changes in system behavior such as transitions between or escape quasi-stable states. These transitions can correspond to critical events such as catastrophic failures or extinctions that make them essential to understand and quantify. Here, we propose two mechanism fundamental to metastable behavior. First, short-lived correlations are essential to initiate metastable transitions. Second, long-lived correlations are essential to stabilize after a metastable transition. Two new measures are introduced that can quantify the system stability and the initiator for the mestable transitions using tools from information theory. Metastable behavior was studied in a general a general class of systems are considered governed by the Boltzmann-Gibss distribution. In particular, the kinetic Ising model and Susceptible-Infected-Susceptible model are used. By means of measuring information flows, the contributions of short-lived and long-term correlations are observed before and during the metastable transition. The results show that short-lived correlations are essential to absorb noise and transmit the noise through the network, whereas long-term correlations are essential to retain stability outside the metastable transition. The results suggest that the choice of control variables for metastable transitions differs remarkable depending on the stability of the system. Information flows offers a unique way to study metastable transitions and may provide novel sensitivity markers to control and prevent metastable behavior in real-world complex systems.

1 Introduction

Complex dynamical systems such as cell-regulatory processes, epidemic spreading dynamics, opinion formation neural behavior and so on, often exhibit robust adaptive behavior. In the presence of noise, either external or internal, the system dynamics may rapidly adapt by moving from one stable behavior to another. These noise-induced events are often rare events and may be associated with undesirable outcome such as the extinction of an infectious disease or with adverse advents such as the clustering of cancerous cells or rapid spontaneous failure of a system. The mechanisms underlying these transitions are often not well understood. It is of vital importance to understand the underlying processes that cause metastable behavior to quantify the impact of noise on complex systems.

Here, we consider systems that follow the

Boltzmann-Gibbs distribution which has been used to describe a wide range of behaviors such as neural dynamics [**cite_needed**], opinion dynamics, ferromagnetic spins [1], organized criminal gang interactions [2] and so on [**cite_needed**]. The goal of this study the essential mechanism underlying metastable behavior by using statistical tools from information theory. It is proposed that short-lived correlations are essential to destabilize the system which allows long-term correlations to move the system state to another absorbing state.

Noise-induced transitions can be understood to be caused by either external or internal [3]. Traditionally, external noise is studied by replacing an external control variable with a stochastic process, whereas internal noise is caused by random interactions of the elements of the system. For example the transmission of COVID-19 from one individual to another random individual through homogeneous mixing. For both types of noise, the study of metastability results in finding the optimal path between metastable points either numerically or by analysis. The metastable trajectories are the analyzed in terms of the mean exit times or mean transition times, but yields no insights *how* the metastable transition occured. That is, what set of steps are essential for metastable behavior to occur.

From a statistical physics perspective, the system dynamics can be understood as a particle moving on an energy landscape (fig. 1). The system configuration for this "system particle" is represented by the current system state. Over time the system particle will traverse the energy landscape, evaluating its next move by evaluating the energy of the system at a new position in the energy landscape. New system states will preferable be chosen of new locations with lower energy than the current state. In this view metastability can be understood as the system particle moving between local energy minima by "climbing" out of these minima and crossing a so-called "tipping point" (the high-point of the energy barrier separating metastable states).

A tipping point is a point in the energy landscape that separates to metastable points (fig. 1 red circle). For deterministic dynamics the trajectory that the system particle makes is completely determined by the initial state and the system dynamics. Consequently, the system particle will always move towards to lowest reachable point given the current position of the particle. In contrast, with stochastic (noisy) system dynamics, the system is able to escape local minima.

Understanding the impact of noise induced transitions on the system dynamics of complex systems are often difficult due to the rarity of transitions. Consequently, analysis and computation often focus on the most important noise-induced events such as coexisting stable states, critical failures, and nucleation of coherent structures. These critical events are often studied by considering the optimal paths for the system to transition into the metastable state. Less is known, however, how the system *uses* the noise to traverse these paths [4]. For example, it is well known that metastable states can be traversed in stochastic systems by controlling a noise parameters (fig. 1). The increase in noise changes the energy landscape such that transitioning from one metastable state becomes energetically feasible. However, this requires access to an external control variable. The identification of this control-variable for real-world complex systems may be difficult identify or completely lacking. For example, some organisms such as the sponges, or octopi and other amphibians are able to self-organize damage to their bodies without any external control parameter. The system endogeneous dynamics drive the transition from one metastable state to the next.

More importantly, traditional forms of analysis focus on the system as a whole. Less is known how the internal elements of the system use the noise to iniate and traverse the the energy landscape. That is, is there a constructive mechanism that causes the system to cross between two metastable states. For many complex systems, there exists an inherently diversity in the elements that make up the system. Making it essential to design methods and metrics that gain insights *how* the system uses noise to traverse between metastable states.

The goal of this study is to provide a novel, computational feasible method to study endogenous noise-induced metastability through the observations of the system only. The system considered follow are dictated by the Boltzmann-Gibbs distribution that has been shown to be a valid descriptor for many different kinds of systems ranging from neural behavior, to the interaction of gas particles, opinion dynamics, epidemic spreading and so on. Exact information flows are computed as function of the stability of the system (see 11.2) and compared with the structural features of the network. This exposes the mechanism *how* the system uses

noise to transition between metastable points. The results show that for out-of-equilibrium dynamics short-lived correlations are necessary for tipping behavior to absorb and transfer noise through the system. In contrast, low-term correlations are necessary for maintaining the system state in the next (metastable) state. Information theoretic measures are used to track the information flows as a function of the distance the system is to metastable transition. The results show that as a function of the stability of the system, the contribution to the information processing differs depending on the processes of short-lived and long-term correlations. This may form a new way of thinking for preventing rapid state transitions.

2 Results

In order to understand *how* the metastable transition occurs, the focus is first on the individual nodes of the system. Here the tipping point is defined as the point where the majority of the nodes switch from the "off" or 0 state to the "on" or 1 state. The tipping point is defined as the point that maximizes the system entropy. Each node is governed by kinetic Ising spin dynamics. The noise level (temperature) is set such that is maximizes the statistical complexity [5] (see 11.2).

In figure 2 the tendency for a node to flip from the majority to the minority state is computed as function of fraction of nodes possessing the majority states $+1$ in the system, denoted as N. Two things are observed. First, nodes with lower degree are more susceptible to noise than nodes with higher degree. For a given system stability nodes with lower degree tend to have have a higher tendency to flip. This is true for all distances of the system to the tipping point. In contrast, the higher the degree of the node, the closer the system has to be to a tipping point for the node to change its state. This can be explained by the fact that lower degree nodes, have less constraints compared to nodes with higher degree nodes. For Ising spin kinetics, the nodes with higher degree tend to be more "frozen" in their node dynamics than nodes with lower degree. Second, in order for a node to flip with probability with similar mass, i.e. $E[p(s_i)|N] = 0.2$ a node with higher degree needs to be closer to the tipping point than nodes with lower degree. In fact, the order of susceptibility is correlated with the degree; the susceptibility decreases with increasing degree and fixed fraction of nodes in state 1.

Figure 2 implies lower degree nodes generate noise in the system, causing the system stability to decrease. This then has a higher tendency to recruit higher degree nodes further destabilizing the system. For example, when a degree 2 nodes flips to the minority state, it creates a higher probability of degree 3 node to flip. Close to the tipping point, the highest degree node is "recruited" and the system tips.

However, In complex system, the node dynamics are not merely determined by the degree. For example, a node with degree 2 may be connected indirectly in a clique or connect to nodes that are not connected among themselves. The effect of the network downstream from the immediate neighbors may crease stability that cannot be captured in the local structural features of a node. For example, a node with degree may be part of a clique or connect two degree nodes. The former under Ising spin dynamics is more stable than the later example. The effect of network structure creates non-linear dependency on the flip probability of node as local network structure may create local stable clusters that defy the analysis performed in fig. 2. How the current node state correlates with the future system state is captured with information flows.

Informally, the information flows measures the statistical coherence between two random variables X and Y over time such that the present information in Y cannot be explained by the past of Y but rather by the past of X. Estimating information flow is inherently difficult due to the presence of confounding which potential traps the interpretation in the "correlation does not equal causation". Under some context, however, information flow can be interpreted as causal [6].

To track the influence of a node on the system dynamics, the mutual information between a nodes state and a future system state is computed $I(s_i : S^t)$. Two information features are explored: the adjusted integrated mutual information and asymptotic information (fig. 1). First, the integrated mutual information is computed [7] as the area under the curve. The adjusted mutual information is computed by substracting the information asymptote. For bistable systems, the switch between the metastable states are rare. For a local perturbation of a node switching its state from majority to minority, the system tends

Figure 1: *(a) A system consists of elements (circles) with a coupled interaction structure (edges). Each node has some intrinsic dynamics indicated by the energy lines (gray); a low configurational energy corresponds with a "stable" state. Metstability are often considered on a system level (b), where a macrscopic system state is decomposed by some control parameter. Here, the system consists of two metastable states. The system in a current metastable state (green) can only transition into another stable state by crossing the tipping point (red). Out of equilibrium the temporal dynamics of such a transition are depicted in (c). Over time the temporal dynamics cause may cause metastable transitions (red dot c). By studying the information flows as a function of tipping distance (e), The integrated mutual information represents the area under the curve for the information decay of a node with the system over time ($I(s_i : S^t)$): it is a measure of how much the current node state, predicts the future system state. Asymptotic information forms an approximation of long time scale dynamics. In contrast, the integrated mutual information captures the short time scale dynamics out-of-equilibrium. Through information features, the mechanism underlying metastable transitions can be understood (d): far away from the tipping point, information processing occurs in low degree nodes, as the system approaches the tipping point, the higher degree nodes are recruited. The information cascade unravels the mechanisms whereby short-lived correlations are essential for priming the system for the metastable transition. For more information on numerical approaches see 11.2.*

Figure 2: *Susceptibility of a node with degree k switching from the minority state 0 to the majority state 1 as a function of de neighborhood entropy for $\beta = 0.5$. The neighborhood entropy encodes how stable the environment of a spin is. As the system approaches the tipping point, the propensity of a node to flip from to the minority state increases faster for low degree nodes than for high degree nodes. Higher degree nodes require more change in their local environment to flip to the majority state. See for details 11.7.*

to relax back to the majority state. As the system approaches the tipping point, the system is more likely to relax to the other metastable point. When the next metastable point is chosen, however, the system fluctuates around this metastable point for a long time. This causes two phases to appear; one in which the nodes aligns it state to the local majority state, and the second, where the system relaxes to the equilibrium distribution over time. This relaxation can occur on the order of a million simulations steps. As an approximation, there exists an information offset. The adjusted integrated mutual information is computed as

$$\bar{\mu}_(s_i) = \sum_{t=0}^{\infty} (I(s_i : S^t) - \omega)\Delta t, \qquad (1)$$

where $I(s_i : S^t)$ is the time-delayed Shannon mutual information between a node s_i and the entire system S some time t away from equilibrium, and ω is the approximated offset (fig. 1). Intuitively, $I(s_i : S^t)$ represents how much a systems' future state "remembers" a nodes past state; mutual information can be seen as a non-linear correlation function. In a previous study it was shown that the driver-nodes in closed systems are those that maximizes $\mu(s_i)$ [7]. In this paper, the information flow are computed exactly by evaluating the system out-of-equilibrium as a function of system stability (see 11.2).

In figure 3, the exact adjusted integrated mutual information and exact information flows are shown for the Krackhardt kite graph. The noise was set such that it maximizes the statistical complexity (see 11.2). Far away from the tipping point, most information processing occurs in low degree nodes. As the system approaches the the tipping point, the local fluctuations are propagated and the nodes with the highest integrated mutual information moves to higher degrees. In addition, the asymptotic information becomes larger as the system approaches the tipping point. Note, that for $N = 0.1$, the relative entropy of the node with degree 1 is higher than for $N = 0.2$. This can be explained by the fact that for $N = 0.2$ two bits are effectively flipped when using Glauber dynamics. This stabilizes the degree 1 node relatively more than in the case where 1 bit is flipped ($N = 0.1$). Consequently, the shared information $I(s_i : S^0)$ is higher even though the system is considered to be more stable than for $N = 0.2$. The same pattern can be observed for epidemic spreading (fig. 5).

Finally, we note that on the tipping point, the information decays for all nodes more similar than further from the tipping point (fig. 3). The most distinguishing feature is the asymptotic information. The similarity in information decay can be explained by the shared similarity in the node dynamics. That is, at the tipping point each node has the same 50/50 distribution. In contrast, further away from the tipping point, nodes with lower degree are generate higher levels of entropy. As the system relaxes back into a metastable point, the relative impact that each node has on its neighbors are the same. That is, each node "moves" the neighbor's energy by a similar amount. The resulting node-node dynamics result due to this similarity in a similar decay curve and consequently similar adjusted integrated mutual information. The exact rates will diverge over time as the immediate degree and higher order network structure dissipates information differently.

The pattern where the nodes with the highest adjusted integrated mutual information moves to the "core" of the system is consistent for various other systems (fig. 7).

Figure 3: *(a) As the system approaches the tipping point the information processing moves from lower degree nodes to higher degree nodes. Each node is governed by kinetic Ising dynamics. The node size is proportional to the adjusted integrated mutual information. (b) Information flows as a function of system stability. Far from the tipping point the information processing is mainly in lower degree nodes. As the system approaches the tipping point, the information flows increases for all nodes. Higher degree nodes tend to have higher adjusted integrated mutual information and higher information offset. The information offset encodes the long-time scale correlation of the node with the system state. A higher asymptotic information implies that the system remembers the node state for longer than other nodes.*

Figure 4: *The tipping point is initiated from the bottom up. Each node is colored according to state 0 (black) and state 1 (yellow) Shown is a trajectory towards the the tipping point that maximizes* $\sum_{t=1}^{5} \log p(S^{t-1}|S^t, M(S^5) = 0.5)$. *As the system approaches the tipping point, low degree nodes flip first, and recruite "higher" degree nodes to further destablize the system and push it towards a tipping point. There are in total 30240 trajectories that reach the tipping point in 5 steps, and there are 10 trajectories that have the same maximized values as the trajectory shown in this figure.*

Figure 5: *(a) As the system approaches the tipping point the information processing moves from lower degree nodes to higher degree nodes. Each node is governed by Suseptible-Infective-Susceptible dynamics with infection rate = 0.1, and recovery rate = 0.1. The node size is proportional to the adjusted integrated mutual information. (b) Information flows as a function of system stability. Far from the tipping point the information processing is mainly in lower degree nodes. As the system approaches the tipping point, the information flows increases for all nodes. Higher degree nodes tend to have higher adjusted integrated mutual information and higher information offset. The information offset encodes the long-time scale correlation of the node with the system state. A higher asymptotic information implies that the system remembers the node state for longer than other nodes.*

3 Discussion

Understanding how metastable transitions occur may help in understanding how, for example, a pandemic occurs, or a system undergoes critical failure. In this paper, the kinetic Ising model was used to study how endogenous information is used to traverse through a tipping point. Noise was fixed to maximize the statistical complexity of the system behavior (see 11.2). The results show that low degree nodes for kinetic Ising spin dynamics form sources of noise that are propagated as the system approaches the tipping point. Crucially, the node with the highest integrated mutual information corresponds to the driver-node for closed systems with ergodic dynamics [7]. The tipping point is caused an *information cascade* from the bottom-up where low degree nodes destabilize higher degree nodes, slowly allowing the system to climb out the energy barrier between metastable states. Closer to the tipping point, the asymptotic information represents the system "choosing" its new stable state. This leaves an approximated offset, encoding the remaining correlation the future system state has with the past node state as the system relaxes into the new metastable state. Together, the information flows, lay bare a separation of scales where a fast-time scale dynamics are captured by the adjusted mutual information and the approximated offset is captured by the information asymptote (fig. 3, 5). It is important to emphasize, that for the ergodic dynamics considered here, the information should decay back to zero due to the data-processing inequality. The asymptotic information approximates the decay as an offset as the slower phase occurs on many order of magnitude; that is after a the system transitions in to an new metastable states, it remains there for a relative long time compared to the fast-time scale dynamics (fig. 1 c).

4 Conclusions

The information theoretic approach offers an alternative view to understand metastable transitions. Adjusted integrated mutual information offers a novel way to understand how the system approaches, and crosses a tipping point. The driver

Figure 6: *White noise of the system macrostate outside the tipping point. Numerical simulations were performed using 6 different seeds. (a, b) White noise was estimated for the instantaneous system macrostate for the two stable point (a, b) (see 11.6). The intervention pinned the node at state +0. This causes the system to prefer the macrostate where the fractions of nodes are < 0.5 regardless of the node intervened on. Importantly, the figure shows that intervention on the lower degree nodes (e.g. 9 or 8) removes high frequency noise (c). Compared to the control condition (blue bands) the interventions on higher degree nodes (e.g. 4) produces more white noise for the system macrostate but less frequent tipping points. The high frequency noise is essential to initiate the metastable transition whereas higher degree nodes are essential to retain the stability when the tipping occured. Interventions on higher degree nodes prevents the tipping point from occurring as the higher degree nodes have to flip as the system crosses the tipping point. Interventions on higher degree nodes therefore produce higher levels of white noise for (a) but less for (b) as the system macrostate does not make the metastable state that often. (c) Shown are a system trajectory for the krackhardt kite graph with seed 1234. An intervention pins the node state at state +0. The figure shows that intervention on lower degree nodes remove high frequency noise (e.g. see node 9 or 8) when the system macrostate is below 0.5, but increased when the system is above 0.5. For lower degree nodes the system is more stable when the macrostate is below 0.5. In contrast, interventions on higher degree nodes (e.g. node 3), transitions less between metastable states but has increased noise when the system is <0.5.*

node far away from the tipping point is dominated by statistically more varied nodes (lower degree). As the systems approaches the tipping point, the driver node changes as more statistically stable nodes are destabilized by the lower degree nodes. On the tipping point, long-term correlations stabilizes the system inside the new metastable state. Importantly, the information perspective allows for estimating integrated mutual information directly directly estimated from data without knowing the mechanisms that drive the tipping behavior. The results highlight how short-lived correlations are essential to initiate the information cascade for crossing a tipping point.

5 Limitations

Adjusted integrated mutual information was computed based on exact information flows. This means that for binary systems it requires to compute a transfer matrix on the order of $2^{|S|} \times 2^{|S|}$. This reduced the present analysis to smaller graphs. It would be possible to use Monte-Carlo methods to estimate the information flows. However, $I(s_i : S^t)$ remains computationally expensive to compute.

In addition, the information approach will only work for systems that lack complete symmetry. Metastable transitions occur for finite-size kinetic Ising models. The current approach will not be able to discern node contributions due to the internal symmetries of the system (all nodes have the same degree). However, we speculate that the metastable transitions could be studied by not controlling the tipping point with the total fraction of nodes in a particular state. In contrast, one should fix the system state for a particular region in the grid-graph. In this sense, nodes with high variability will destabilize more stable nodes, creating an information cascade that forces the system to move between metastable states.

A general class of systems was studied governed by the Boltzmann-Gibbs distribution. For practical purposes the kinetic Ising model and SIS dynamics were only tested, but we speculate that the results should hold (in principle) for other systems dictated by the Boltzmann-Gibbs distribution. We leave the extension for other system Hamiltonians up to future work.

Figure 7: *Adjusted mutual information for a random tree (top), and Leder-Coxeter Fruchte graphs (middle, bottom). Each node is goverened by kinetic Ising spin dyanmics. Far away from the tipping point (fraction nodes +1 = 0.5) most information flows are concentrated on non-hub nodes. As the system approaches the tipping point (fraction = 0.5), the information flows move inwards, generating higher adjusted integrated mutual information for nodes with higher degree.*

6 Author contribution

Casper van Elteren was responsible for the initial and final draft, code, data and data analysis. Rick Quax was a core part of the conceptualization, and input for analysis, representing data. Peter Sloot was involved with review and feedback.

7 Acknowledgments

I would like to thank Fiona Lippert, and Jair Lenssen for providing insights and feedback in various ideas present in this paper. This research is supported by grant Hyperion 2454972 of the Dutch National Police.

8 Competing interests

The authors declare no competing interests.

9 Material and correspondence

For correspondence please contact c.vanelteren@uva.nl. Simulations were performed on a laptop with a AMD 4900HS 16 core CPU with a 2600RTX GPU.

10 References

1. Glauber, R. J. Time-Dependent Statistics of the Ising Model. *Journal of Mathematical Physics* **4,** 294–307. ISSN: 00222488 (1963).
2. D'Orsogna, M. R. & Perc, M. Statistical Physics of Crime: A Review. *Physics of Life Reviews* **12,** 1–21. ISSN: 15710645 (Mar. 2015).
3. Forgoston, E. & Moore, R. O. A Primer on Noise-Induced Transitions in Applied Dynamical Systems. *SIAM Review* **60,** 969–1009. ISSN: 0036-1445, 1095-7200. arXiv: 1712.03785 (Jan. 2018).
4. McDonnell, M. D. & Abbott, D. What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology. *PLoS Computational Biology* **5** (ed Friston, K. J.) e1000348. ISSN: 1553-7358 (May 2009).
5. López-Ruiz, R., Mancini, H. L. & Calbet, X. A Statistical Measure of Complexity. *Physics Letters A* **209,** 321–326. ISSN: 03759601. arXiv: 1009.1498v1 (1995).
6. van Elteren, C., Quax, R. & Sloot, P. Dynamic Importance of Network Nodes Is Poorly Predicted by Static Structural Features. *PHYSICA A,* 27 (2021).
7. van Elteren, C., Quax, R. & Sloot, P. Dynamic Importance of Network Nodes Is Poorly Predicted by Static Structural Features. *Physica A: Statistical Mechanics and its Applications,* 126889. ISSN: 03784371 (Jan. 2022).
8. Calim, A., Palabas, T. & Uzuntarla, M. Stochastic and Vibrational Resonance in Complex Networks of Neurons. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379,** rsta.2020.0236, 20200236. ISSN: 1364-503X, 1471-2962 (May 2021).
9. Czaplicka, A., Holyst, J. A. & Sloot, P. M. Noise Enhances Information Transfer in Hierarchical Networks. *Scientific Reports* **3.** ISSN: 20452322 (2013).
10. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* ISBN: 978-0-471-24195-9 (2005).
11. Nicolis, G. & Nicolis, C. Stochastic Resonance, Self-Organization and Information Dynamics in Multistable Systems. *Entropy* **18,** 172. ISSN: 1099-4300 (May 2016).
12. Lizier, J. T., Prokopenko, M. & Zomaya, A. Y. Information Modification and Particle Collisions in Distributed Computation. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **20,** 037109. ISSN: 1054-1500, 1089-7682 (Sept. 2010).
13. Schreiber, M. *Volume 1 Edited by K. Krickeberg· R. C. Lewontin . J. Neyman M. Schreiber* ISBN: 978-3-642-46246-7 ().
14. Ay, N. & Polani, D. Information Flows in Causal Networks. *Advances in Complex Systems* **11,** 17–41. ISSN: 0219-5259 (2008).
15. Runge, J. *et al.* Inferring Causation from Time Series in Earth System Sciences. *Nature Communications* **10,** 1–13. ISSN: 20411723 (2019).

16. Li, S., Xiao, Y., Zhou, D. & Cai, D. Causal Inference in Nonlinear Systems: Granger Causality versus Time-Delayed Mutual Information. *Physical Review E* **97,** 052216. ISSN: 2470-0045, 2470-0053 (May 2018).

17. James, R. G., Barnett, N. & Crutchfield, J. P. Information Flows? A Critique of Transfer Entropies. *Physical Review Letters* **116,** 1–6. ISSN: 10797114. arXiv: 1512.06479 (2016).

18. Janzing, D., Balduzzi, D., Grosse-Wentrup, M. & Schölkopf, B. Quantifying Causal Influences. *Annals of Statistics* **41,** 2324–2358. ISSN: 00905364. arXiv: 1203.6502v2 (2013).

19. Schamberg, G., Chapman, W., Xie, S.-P. & Coleman, T. P. Direct and Indirect Effects—An Information Theoretic Perspective. *Entropy* **22,** 854 (Aug. 2020).

20. Williams, P. L. & Beer, R. D. Nonnegative Decomposition of Multivariate Information. *arXiv:1004.2515 [math-ph, physics:physics, q-bio].* arXiv: 1004.2515 [math-ph, physics:physics, q-bio]. http://arxiv.org/abs/1004.2515 (2022) (Apr. 2010).

21. Quax, R., Apolloni, A. & a Sloot, P. M. The Diminishing Role of Hubs in Dynamical Processes on Complex Networks. *Journal of the Royal Society, Interface / the Royal Society* **10Q,** 20130568. ISSN: 1742-5662. arXiv: 1111.5483 (2013).

22. López-Ruiz, R., Mancini, H. L. & Calbet, X. A Statistical Measure of Complexity. *Physics Letters A* **209,** 321–326. ISSN: 03759601 (1995).

23. Virtanen, P. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17,** 15 (2020).

11 Appendix

11.1 Background, scope & innovation

Noise induced transitions produces may produce metastable behavior that is fundamental for the functioning of complex dynamical systems. For example in neural systems, the presence of noise increase information processing. Similarly, the relation between glacial ice ages and earth eccentricity has been shown to have a strong correlation. Metastability manifests itself by means of noise that can be of two kinds [3]. External noise originates form events outside the internal system dynamics [8, 9]. Examples include the influence of climate effects, population growth or a random noise source on a transmission line. External noise is commonly modeled by replacing an external control or order parameter by a stochastic process. Internal noise, in contrast, is inherent to the system itself and is caused by random interactions of elements of the system, e.g. individuals in a population, or molecules in chemical processes. Both types of noise can generate metastable transitions between one metastable state to another. In this paper, the metastable behavior is studied of internal noise in complex dynamical networks governed by the kinetic Ising dynamics.

In this work a novel approach using information theory is explored to study metastable behavior. It offers profound benefits over traditional methods used in metastable analysis as it is model-free, can be used for both discrete and continuous variables, and can be estimated directly from data [10]. Shannon information measures such as mutual information and Fisher information can be used to study how much information the system dynamics share with the control parameter [11, 12]. These approaches allow to measure when, for example, a phase transition occurs. However, for many complex systems an external control may not be accessible or be absent all together. In addition, knowing about the order parameter does not gain additional insight *how* the system uses noise to transition between stable points(e.g. see 1).

Information flows may be used to study how a system transitions between metastable points. Informally, information flow refers to the statistical coherence between two random processes X and Y such that the present information in Y cannot be better explained by the past of X than the past of Y. Various methods exist to study information flow such as transfer entropy [13], conditional mutual information under causal intervention [14], causation entropy [15], time-delayed shannon mutual information [16] and so on. These methods are used to infer the transfer of information between sets of nodes by possible correcting for a third variable. In a multivariate setting most of these methods are prone to overestimate or under-

estimate the causal flows [17]. In past work, the authors developed an novel method that reliable estimates the driver nodes in complex systems using information theory. Using integrated mutual information in closed ergodic systems, the most causal node is exempt from any spurious statistical correlations. Consequently for driver nodes the information flows in these systems is proportional to its causal influence out-of-equilibrium. Instead of focusing on a (full) decomposition of statistical variance of source and sinc variables [18–20], the focus here is on understanding *how* the metastable behavior of the system occurs.

The present study innovates on prior research on information flow and causal node identification by applying integrated mutual information (IMI) directly to metastable transitions applies to determine how metastable transitions arise in complex systems [6, 21]. As complex systems are defined by a wide variety of different types or classes systems (e.g. open or closed) and types of dynamics (e.g. equilibrium or out of equilibrium), we restrict this work to systems that have probability distributions of the form $P(S) \propto \exp -\beta(S)$, where $\mathbb{H}(S)$ the energy of the system. In particular, the bistable behavior of magnetic spins on networks are studied dictated by kinetic Ising spin dynamics. The kinetic Ising model is considered to be one of the simplest models that shows which shows bistability at finite size. It is important to emphasize that the proposed information theoretic measures have more implications than merely the kinetic Ising model. The measures can be computed based on observations from the systems and can therefore be directly estimated from data independent on the underlying process or model. The use of kinetic Ising model serves a convenience to show the value of the proposed method. Additionally, it is hypothesized that for metastable transitions, short-time scales can be approximated using ergodic system dynamics.

11.2 Methods & definitions

11.2.1 Model

To study metastable behavior, we consider a system as a collection of random variables $S = \{s_1, \ldots, s_n\}$ governed by the Boltzmann-Gibbs distribution

$$P(S) = \frac{1}{Z} \exp(-\beta \mathcal{H}(S)), \tag{2}$$

where is the inverse temperature $\beta = \frac{1}{T}$ which control the noise in the system, $\mathcal{H}(S)$ is the system Hamiltonian which encodes the node-node dynamics. The choice of the energy function dictates what kind of system behavior we observe. Here, we focus on arguable the simplest models that shows metastable behavior: the kinetic Ising model, and the Susceptible-Infected-Susceptible model.

Temporal dynamics are simulated using Glauber dynamics sampling. In each discrete time step a spin is randomly chosen and a new state $X' \in S$ is accepted with probability

$$p(\text{accept} X') = \frac{1}{1 + \exp(-\beta \Delta E)}, \tag{3}$$

where $\Delta E = \mathcal{H}(X') - \mathcal{H}(X)$ is the energy difference between the current state X and the proposed state X'.

11.2.2 Kinetic Ising model

The traditional Ising model was originally developed to study ferromagnetism, and is considered one of the simplest models that generate complex behavior. It consists of a set of binary distributed spins $S = \{s_1, \ldots s_n\}$. Each spin contains energy given by the Hamiltonian

$$\mathcal{H}(S) = -\sum_{i,j} J_{ij} s_i s_j - h_i s_i. \tag{4}$$

where J_{ij} is the interaction energy of the spins s_i, s_j. The interaction energy effectively encodes the underlying network structure of the system. Different network structures are used in this study to provide a comprehensive numerical overview of the relation between network structure and information flows (see 11.2). The interaction energy J_{ij} is set to 1 if a connection exists in the network.

For sufficiently low noise (temperature), the Ising model shows metastable behavior (fig. 1 c). Here, we aim to study *how* the system goes through a tipping point by tracking the information flow per node with the entire system state.

11.2.3 SIS model

The SIS model is arguable the simplest model to study epidemic spreading. Each agent can either be susceptible (0) or infected (1). The agents

can transition from susceptible to infected proportional to the number of infected people it is in contact with. In addition, each agent has a base rate of becoming infectious. One can describe the SIS dynamics using the Hamiltonian as:

$$\mathcal{H}(S)_{SIS} = \sum_i (2s_i - 1)(1-\eta)^{\sum_j A_{ij}s_j} - \mu s_i, \quad (5)$$

where η is the infection rate, μ is the recovery rate, and A_{ij} is 1 if s_i and s_j have an interaction, 0 otherwise.

11.3 Information flow on complex networks

Informally, the information flows measures the statistical coherence between two random variables X and Y over time such that the present information in Y cannot be explained by the past of Y but rather by the past of X. Estimating information flow is inherently difficult due to the presence of confounding which potential traps the interpretation in the "correlation does not equal causation". Under some context, however, information flow can be interpreted as causal [6]. Let $S = \{s_1, \ldots, s_n\}$ be a random process, and S^t represent the state of the random process at some time t. The information present in S is given as the Shannon entropy

$$H(S) = \sum_{x \in S} p(x) \log p(x) \quad (6)$$

where $\log$ is base 2 unless otherwise stated, and $p(x)$ is used as a short-hand for $p(S = x)$. Shannon entropy captures the uncertainty of a random variable; it can be understood as the number of yes/no questions needed to determine the state of S. This measure of uncertainty naturally extends to two variables with Shannon mutual information. Let s_i be an element of the state of S, then the Shannon mutual information $I(S; s_i)$ is given as

$$\begin{aligned} I(X;Y) &= \sum_{x \in S, y \in s_i} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= H(S) - H(S|s_i) \end{aligned} \quad (7)$$

Shannon mutual information can be interpreted as the uncertainty reduction of S after knowing the state of s_i. Consequently, it encodes how much statistical coherence s_i and S share. Over time Shannon mutual information can be extended to encode how much *information* (in bits) flows from state s_i to S^t

$$I(S^t; s_i) = H(S^t) - H(S^t|s_i). \quad (8)$$

Prior results showed that the nodes with the highest causal importance are those nodes that have the highest information flow (i.e. maximize 8) [7]. Intuitively, the nodes for which the future system "remembers" information from a node in the past, is the one that "drives" the system dynamics. Formally, these driver nodes can be identified by computing the total information flow between S^t and s_i can be captured with the integrated mutual information [6]

$$\mu(s_i) = \sum_{\tau=0}^{\infty} I(s_i^{t-\tau}; S^t). \quad (9)$$

The driver nodes are the nodes that maximize 9. Note that in [7] $I(S : s_i^t)$ was considered. Here, information flows are computed out-of-equilibrium with symmetry breaking. That is, the system dynamics are evolved by starting the system at a distance from the tipping point and evolving it out-of-equilibrium. This causes $I(s_i^t : S)$ to not follow the data processing inequality as information may flow back into a node. The choice for computing $I(s_i^t : S)$ over $I(s_i : S^t)$ was done for computational feasibility in [7] . Furthermore, the data processing inequality was not violated when considered the system without symmetry breaking. For 8 the data processing inequality is guaranteed, however it is computationally more challenging to compute (see 5).

11.4 Noise matching procedure

The Boltzmann-Gibbs distribution is parameterized by noise factor $\beta = \frac{1}{kT}$ where T is the temperature and k is the Boltzmann constant. For high β values metastable behavior occurs in the kinetic Ising model. The temperature was chosen such that the statistical complexity [22] was maximized. The statistical complexity C is computed as

$$C = \bar{H}(S)D(S), \quad (10)$$

where $\bar{H}(S) = \frac{H(s)}{-\log_2(|S|)}$ is the system entropy, and $D(S)$ measures the distance to disequilibrium

$$D(S) = \sum_i (p(S_i) - \frac{1}{|S|})^2. \tag{11}$$

A typical statistical complexity curve is seen in 8. The noise parameter β is set such that it maximizes the statistical complexity using numerical optimization (COBYLA method in scipy's `optimize.minimize` module) [23].

11.5 Exact information flows $I(s_i; S^t)$

In order to compute $I(s_i : S^t)$, the conditional distribution $p(S^t|s_i)$ and $p(S^t)$ needs to be computed. For Glauber dynamics, the system S transitions into S' by considering to flips by randomly choosing node s_i. The transition matrix $P(S^t|s_i) = \mathbf{P}$ can be constructed by computing each entry p_{ij} as

$$\begin{aligned} p_{ij,i\neq j} &= \frac{1}{|S|}\frac{1}{1+\exp(-\Delta E)} \\ p_{ii} &= 1 - \sum_{j,j\neq i} P_{ij}, \end{aligned} \tag{12}$$

where $\Delta E = \mathcal{H}(S_j) - \mathcal{H}(S_j)$ encodes the energy difference of moving from S_i to S_j. The state to state transition $\mathbf{P}$ matrix will be of size $2^{|S|} \times 2^{|S|} \times |\mathcal{A}_{s_i}|$, were $|\mathcal{A}_{s_i}|$ is the size of the alphabet of s_i, which becomes computationally intractable due to its exponential growth with the system size $|S|$. The exact information flows can then be computed by evaluating $p(S^t|s_i)$ out of equilibrium by evaluating all S^t for all possible node states s_i where $p(S^t)$ is computed as

$$p(S^t) = \sum_{s_i} p(S^t|s_i)p(s_i). \tag{13}$$

11.6 White noise estimation procedure

Let $M(S^t)$ represent the instantaneous system macrostate compute as the system average

$$M(S^t) = \frac{1}{|S|}\sum_i s_i. \tag{14}$$

The metastable behavior is characterized for the Ising model by the system fluctuating around two stable points $T1$ with $M(S) \approx 0$ and $T2$ with $M(S) \approx 1$ for most of the time. By pinning intervention the node is pinned to the $+0$ state, effectively biasing the macrostate $M(S^t)$ towards

Figure 8: *(a) Statistical complexity (C), normalized system entropy ($H(S)$) and disequilibrium ($D(S)$) as a function of the temperature ($T = \frac{1}{\beta}$) for Krackhardt kite graph. The noise parameter was set such that it maximizes the statistical complexity (vertical black line). The values are normalized between [0,1] for aesthetic purposes. (b) State distribution $P(S)$ for temperature that maximizes the statistical complexity in (a) as a function of nodes in state $+1$.*

$T1$. For any particular trajectory the fluctuations around the stable points contributed differently for nodes depending on the nodes embedness in the system; lower degree nodes tend to produce higher fluctuations than higher degree nodes (see main text). We define the fluctuations as "white noise" and characterize the white noise as function of the pinning intervention on different nodes. The white noise is characterized by first computing the set of windows $W = \{w_i | i \in 1, \ldots n\}$ where each window $w_i \in W$ represents the duration in a trajectory that the system state stayed in either $T1$ or $T1$. Next the white noise was characterized as

$$\eta = \frac{1}{|W|}\sqrt{\sum_{j}^{|w_i|}(w_{ij} - \overline{w_i})^2}, \tag{15}$$

where $\overline{w_i}$ is the average of window w_i of the instantaneous macrostate and w_{ij} is a particular instantaneous state. The white noise was estimated for $T1$ and $T2$ separately and displayed in fig 6.

11.7 Switch susceptibility as a function of degree

First, we investigate the susceptibility of a spin as a function of its degree. The susceptibility of a spin switching its state is a function both of the system temperature T and the system dynamics. The system dynamics would contribute to the susceptibility through the underlying network structure either directly or indirectly. The network structure produces local correlations which affects the switch probability for a given spin.

As an initial approximation, we consider the susceptibility of a target spin s_i to flip from a majority state to a minority state given the state of its neighbors where the neighbors are not connected among themselves. Further, the assumption is that for the instantaneous update of s_i the configuration of the neighborhood of s_i can be considered as the outcome of a binomial trial. Let, N be a random variable with state space $\{0,1\}^{|N|}$, and let $n_j \in N$ represent a neighbor of s_i. We assume that all neighbors of s_i are i.i.d. distributed given the instantaneous system magnetization

$$M(S^t) = \frac{1}{|S^t|}\sum_i s_i^t. \tag{16}$$

Let the minority state be 1 and the majority state be 0, the expectation of s_i flipping from the majority state to the minority state is given as:

$$\begin{aligned} E[p(s_i = 1|N)]_{p(N)} &= \sum_{N_i \in N} p(N_i)p(s_i = 1|N_i) \\ &= \sum_{N_i \in N}\prod_{j}^{|N_i|} p(n_j)p(s_i = 1|N_i) \\ &= \sum_{N_i \in N}\binom{n}{k} f^k(1-f)^{n-k}p(s_i = 1|f), \end{aligned} \tag{17}$$

where f is the fraction of nodes in the majority states, n is the number of neighbors, k is the number of nodes in state 0. In figure 2. this is computed as a function of the degree of spin s_i. As the degree increases, the susceptibility for a spin decreases relatively to the same spin with a lower degree. This implies that the susceptibility of change to random fluctuations are more likely to occur in nodes with less external constraints as measured by degree.