

REVIEW

Open Access

Information and phase transitions in socio-economic systems

Terry Bossomaier^{1*†}, Lionel Barnett^{2†} and Michael Harré^{3†}

*Correspondence:

tbossomaier@csu.edu.au

†Equal contributors

¹Centre for Research in Complex Systems, Charles Sturt University, Panorama Ave, Bathurst NSW 2795, Australia

Full list of author information is available at the end of the article

Abstract

We examine the role of information-based measures in detecting and analysing phase transitions. We contend that phase transitions have a general character, visible in transitions in systems as diverse as classical flocking models, human expertise, and social networks. Information-based measures such as mutual information and transfer entropy are particularly suited to detecting the change in scale and range of coupling in systems that herald a phase transition in progress, but their use is not necessarily straightforward, possessing difficulties in accurate estimation due to limited sample sizes and the complexities of analysing non-stationary time series. These difficulties are surmountable with careful experimental choices. Their effectiveness in revealing unexpected connections between diverse systems makes them a promising tool for future research.

Keywords: Phase transitions; Mutual information; Transfer entropy; Social networks; Stock markets; Expertise

Review

Diamonds are not a good very long term investment! They are steadily turning into graphite. It will take millions of years, but the most stable form of carbon at room temperature and pressure is graphite. Thus diamonds will undergo a phase transition to graphite, albeit over a very long timescale.

When we normally think of phase transitions we think of the states of matter, ice melting to water or water turning to steam. They are order/disorder transitions. In graphite the carbon atoms are linked together in layers. The layers can slide over one another giving graphite its excellent lubricant properties. In diamond the carbon atoms are linked together in a three dimensional structure with each carbon at the centre of a tetrahedron linked to carbons at all four corners. Thus carbon has to go through a major structural reorganization to change from diamond to graphite – the way the atoms are connected together changes dramatically.

We can easily see the outcomes of a phase transition of diamond to graphite or a solid turning into a liquid. But can we construct measures which go through a minimum or maximum at the transition? This turns out to be a surprisingly difficult question to answer. It gets even more difficult if we look for measures which can apply to systems in general, not just to the physical systems above. Organizations, societies, economies, stock markets all go through radical reorganization, but should these

changes be called phase transitions? This paper explores a metric based on information theory (Shannon 1948a) which is empirically a quite general indicator of a phase transition in progress. It then considers a much newer metric which, on examples to date, might be a predictor of impending transitions: the crashing of stock markets is an example of such a transition, early warning of which, would be highly desirable.

In this paper we explore two closely related phase transitions which appear in social economic systems. One is herding behaviour, where individuals behave more and more in the same way. The second is the connectivity avalanche, in which all of the elements of a system are becoming connected to one another.

To begin with we explore theoretical issues. Section “Overview of phase transitions and metrics” discusses the characteristics of phase transitions and the the peak in mutual information that usually accompanies them. It also introduces the idea of transfer entropy, a recent extension to mutual information, which in some cases is known to peak before a transition to synchronous behaviour. With this background, the first example, in section “Mutual information for phase transitions in a simple flocking model”, considers a computational example derived from physics. The computation of mutual information from continuous data is tricky, involving difficult decisions on bin sizes and other statistical issues, which are also considered in this section. The next two sections discuss phase transitions in two areas in the social/humanities domain. Section “Phase transitions in socio-economic systems” discusses how peaks in mutual information occur around stock market crashes. Section “Phase transitions in the acquisition of human expertise” discusses the reorganization of strategy in the human brain during the acquisition of expertise. Section “Inferring social networks with transfer entropy” discusses how transfer entropy is calculated in practice using the example of inferring social networks from time series data. Finally we conclude in section “Conclusions” with some opportunities for further work.

Overview of phase transitions and metrics

The simple physical notion of a phase transition, such as ice melting to water, is surprisingly hard to transfer to non-physical systems, such as society and organisations. This section will try to first look at the physical intuition behind the transition and then move on to look at some of the possible metrics.

The essential feature of a phase transition is a structural reordering of some kind, usually an order-disorder change. This usually involves some sort of long range order – everything gets connected so that things can be reconnected in a different way. In physical systems, we can define an *order parameter*. Transitions occur around particular values of the order parameter. We can make this idea more intuitive by looking at random graphs, section “Random graphs and phase transitions”.

A peak in *mutual information*, (section “Mutual information” is a widespread measure of a phase transition (Gu et al. 2006; Wilms et al. 2011). Mutual information theory is a very general indicator of critical transition, which incorporates all correlation effects and all nonlinear behaviour as well. Demonstrations for specific systems include the Vicsek model of section “Vicsek model background” and work by Matsuda et al. (1996), who obtained the mutual information for Ising systems, demonstrating that it was a good measure for critical behaviour in such systems that have second order phase transitions. There are, however, two distinct classes of phase transitions that are relevant to this discussion.

Second order phase transitions are indicated by peaks in mutual information and whereas *first order* phase transitions are indicated by discontinuities in the entropy, rather than the mutual information, see (Solé et al. 1996) for a brief discussion of the general differences in such systems. First order phase transitions are caused by discontinuities in the first derivative of a thermodynamic variable rather than the second derivative as is the case with second order phase-transitions, hence their names. For a discussion of these issues in terms of complex systems, and computation at the edge of chaos in particular, see (Langton 1990a).

There are other metrics. Recently, for example, *Fisher Information*, has been used in this context by Prokopenko et al. (2011). Another common indicator is *critical slowing down*, where the time taken to respond to perturbation increases near a transition (Dai et al. 2012), often accompanied by logarithmic oscillations. The latter have been extensively studied in stock markets by Sornette (2001).

Although the transitions referred to so far are of a singular nature, there are other mechanisms. In some cases a system may *flicker*, across the transition (Scheffer et al. 2012), spending increasing amounts of time in the alternative state. Thus flickering can also be an indicator of an impending irreversible system change.

The attraction of mutual information for this paper is its intuitive link to large scale order near the transition and its close relationship to *transfer entropy* (section “Transfer entropy and granger causality”). Although mutual information and related indicators co-occur with phase transitions across many systems, they have two shortcomings in terms of *prediction*: the precise timing depends on many factors and an exact prediction of when a transition will occur is fraught with difficulty; and the sign of a transition is not determined. Work remains to be done on how the peak in mutual information relates to the ordered and dis-ordered phases.

Random graphs and phase transitions

The idea of a random graph, introduced by Erdős and Rényi (1960), is to start with a set of N , nodes and add edges at random. At first the edges create small graph fragments. The total number of possible edges (without duplicates) is of the order of N^2 but for quite a small number of edges, of order N , large components form. Now as adding an edge may join two components together, the total connectivity rises very rapidly until every node in the graph becomes connected by some path to every other node. This rapid rise is referred to as the connectivity avalanche and represents a phase transition.

Before the connectivity avalanche, very few nodes are connected. Since the connected components are small the path lengths between nodes are also small. But as the components get rapidly bigger during the avalanche, more and more nodes become connected, but there is often only a single long path between them. As more nodes are added after the connectivity avalanche, they effectively provide short cuts and the average path length goes down again. This increase in path length is the analogue of the long range order seen during a phase transition.

As Green (Green and Newth 2005) showed, random graphs underlie many complex systems; thus a rigorous mapping to a random graph, demonstrating an isomorphism, guarantees that the system will have a phase transition. Although Green has done this for some example systems, the underlying principle is generally useful even where an exact isomorphism has not been demonstrated.

Mutual information

Shannon (1948a) defined *mutual information* in the course of obtaining the maximum information which could be transmitted across a channel. It is a functional, mapping from probability distributions of random variables to scalars. In Shannon's formulation, for a random variable X defined on a discrete alphabet with probability mass function $p(x)$, the information, q_i , sometimes called the *surprisal*, obtained from an event x_i is given, in bits, by Eq. 1.

$$q_i = -\log_2(p(x_i)) \quad (1)$$

If we average the information according to the probability of each event occurring, we end up with the Shannon *entropy*, Eq. 2.

$$H(X) \equiv - \sum_x p(x) \log_2 p(x) \quad (2)$$

With some caveats, this has the familiar interpretation as the average length of the shortest description of an observation from that variable, or how 'uncertain' a particular random variable is (Cover and Thomas 2006). *Mutual information* between two random variables, X and Y with joint probability mass function $p(x, y)$ is given

$$I(X; Y) \equiv H(X) + H(Y) - H(X, Y) \quad (3)$$

$$= - \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (4)$$

Mutual information is thus a functional of the *joint* distribution of X and Y . It is symmetric in X and Y , and can be given a natural interpretation as

the reduction in uncertainty in one variable from knowing the other, or the amount of information about X contained in Y .

Derivations and implications of these properties are given elsewhere (Cover and Thomas 2006)³. It forms a measure of interdependence of two variables. Unlike correlation, however, it is sensitive to nonlinear interactions, works on general nonparametric inference, and naturally performs well on discrete data. These qualities have led to an interest in the use of mutual information-based measures to automatically detect diverse classes of associations in data sets with few assumptions as to the functional form of the relationship (Reshef et al. 2011; Slonim et al. 2005).

Transfer entropy and granger causality

Mutual information (and its conditional variant), applied to stochastic processes, measure *contemporaneous* statistical dependencies between stochastic dynamical processes evolving in time; that is, given joint processes X_t, Y_t , the mutual information $I(X_t : Y_t)$ at a given time t might be read as:

The amount of information about the present of X resolved by the present of Y .

However, it would seem desirable to have an information-theoretic measure capable of capturing *time-directed* statistical dependencies between processes - *information flow*, if you will. To this end, the most obvious extension to contemporaneous mutual information is *time-delayed* (lagged) mutual information between processes. Thus one might

consider, say, $I(X_t : Y_{t-1}, Y_{t-2}, \dots)$ as a candidate measure for directed information flow from the past^b of Y to the present of X . This quantity might be read as:

The amount of information about the present of X resolved by the past of Y .

There is, however, a fundamental shortcoming to lagged mutual information as a measure of directed information flow: it fails to take into account *shared history* between processes. In his seminal paper (Schreiber 2000), Schreiber recognised that this could lead to spurious imputation of directed information flow and introduced a new measure (Kaiser and Schreiber 2002; Schreiber 2000) which explicitly takes account of a “shared past” between processes. Formally, the *transfer entropy* from process Y_t to process X_t may be defined as:

$$T_{Y \rightarrow X} \equiv I(X_t : Y_{t-1}, Y_{t-2}, \dots | X_{t-1}, X_{t-2}, \dots) \quad (5)$$

Thus in contrast to lagged mutual information, the past of the process X is conditioned out. Eq. (5) might be read as:

The amount of information about the present of X resolved by the past of Y *given the past of X* .

An elegant, minimal example illustrating the necessity of conditioning out the past of the target process may be found in (Kaiser and Schreiber 2002). Note that the processes X_t, Y_t in (5) may be fully *multivariate*. Furthermore, given a third (possibly multivariate) jointly stochastic process Z_t , say, any common influence of Z_t on X_t and Y_t may be taken into account by conditioning, in addition, on the past of Z :

$$T_{Y \rightarrow X|Z} \equiv I(X_t : Y_{t-1}, Y_{t-2}, \dots | X_{t-1}, X_{t-2}, \dots, Z_{t-1}, Z_{t-2}, \dots) \quad (6)$$

This quantity is known as *conditional transfer entropy*, and is in particular used to define *pairwise-conditional* information flows

$$T_{X_i \rightarrow X_j | X_{[ij]}} \quad i \neq j \quad (7)$$

between component variables $X_i \rightarrow X_j$ of the system, conditioned on the remaining variables $X_{[ij]}$, where the notation denotes omission of the variables X_i, X_j . The set of pairwise-conditional information flows may be viewed as a weighted, directed graph, the i, j th entry quantifying information flow between individual elements X_i, X_j of the system.

Clive Granger won the Nobel Prize in economics for a closely related *parametric* measure in econometric theory (and, more recently, applied extensively to neural time series data (Ding et al. 2006)), the Wiener-Granger causality (Geweke 1982; Granger 1969; Wiener 1956). Here, rather than “information flow”, the measure is designed to reflect “causal” influence of one process on another, premised on a notion of causality whereby a causal effect temporally precedes and *helps predict* its influence. The measure is based on linear vector autoregressive (VAR) modelling: suppose that the (again, possibly multivariate) “predictee” process X_t is modelled as a vector linear regression on its own past, as well as on the past of the “predictor” process Y_t :

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \dots + B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + \varepsilon_t \quad (8)$$

We may consider that, given regression coefficients A_k, B_k , (8) furnishes a *prediction* for the present of the variable X , based on its own past and that of the variable Y . A suitable measure for the magnitude of the *prediction error* is given by the *generalised variance*

(Barett et al. 2010), defined as the determinant $|\text{cov}(\varepsilon_t)|$ of the covariance matrix of the residuals ε_t . Given a realisation of the processes, it may be shown that the generalised variance is proportional to the *likelihood* of the model (8); regression coefficients A_k, B_k may thus be calculated within a maximum likelihood framework so as to minimise the generalised variance^c.

We may now compare the regression (8) with a prediction model for X based only on its own past:

$$X_t = A'_1 X_{t-1} + A'_2 X_{t-2} + \dots + \varepsilon'_t \quad (9)$$

We then say that Y *Granger-causes* X iff the full model (8) furnishes a significantly better prediction than the reduced model (9). By “significantly better”, we mean that the null hypothesis of zero causality:

$$H_0 : B_1 = B_2 = \dots = 0 \quad (10)$$

should be rejected at a given significance level. Now the linear regression model (9) is *nested* in (i.e. is a special case of) the model (8), and standard theory (Hamilton 1994) tells us that the appropriate statistical test for the null hypothesis H_0 of (10) is a likelihood-ratio test. The Granger causality statistic is then formally the log-likelihood ratio

$$F_{Y \rightarrow X} \equiv \log \frac{|\text{cov}(\varepsilon'_t)|}{|\text{cov}(\varepsilon_t)|} \quad (11)$$

This quantity may be read as (*cf.* transfer entropy):

The degree to which the past of Y helps predict the present of X *over and above the degree to which X is already predicted by its own past.*

By another classical result (Wilks 1938) the asymptotic distribution of $F_{Y \rightarrow X}$ under the null hypothesis is χ^2 with number of degrees of freedom equal to the difference in the number of parameters between the full and reduced models; this enables significance testing of Granger causality^d. Parallel to the transfer entropy case, a third jointly distributed process Z_t may be conditioned out by appending its past to both the full and reduced regressions, yielding the *conditional Granger causality* (Geweke 1984) $F_{Y \rightarrow X|Z}$. As in (7), pairwise conditional causalities

$$F_{X_i \rightarrow X_j | X_{[ij]}} \quad i \neq j \quad (12)$$

may be calculated^e.

A powerful feature of Granger causality is that it admits a *spectral decomposition*; that is, Granger-causal influence may be measured at specific frequencies, or in specific frequency bands:

$$F_{Y \rightarrow X} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{Y \rightarrow X}(\omega) d\omega \quad (13)$$

where $f_{Y \rightarrow X}(\omega)$ is the Granger causality at frequency ω . We refer the reader to (Barnett and Seth 2011; Geweke 1982, 1984) for definitions and details.

For both transfer entropy and Granger causality, the issue of *stationarity* arises. Although formally both quantities are well-defined for *nonstationary* processes—the result then depends on the time t —empirically, estimation will generally require stationarity. The exception is where multiple synchronised realisations of the processes are available, but this is rarely the case in practice. Otherwise, nonstationarity must be dealt

with by *windowing* time series data; that is, dividing it into approximately stationary segments. Then a trade-off must be made between shorter window length, where estimation suffers through lack of data, and longer window length, where stationarity may be only approximate; in either case there is a risk of spurious inference. Pre-processing (e.g. detrending, differencing, filtering,...) may help improve stationarity.

Both measures, it may be noted, are invariant under a rather broad class of transformations. Barnett et al. (Barnett et al. 2011; Geweke 1982), showed that Granger causality is invariant under arbitrary stable, invertible digital filtering. Transfer entropy is invariant under a still wider class of nonlinear invertible transformations involving lags of the respective time series. In practice, though, even theoretically invariant transformations may impact causal inference (Barnett and Seth 2011).

Regarding the relationship between transfer entropy and Granger causality, it is proven in (Barnett et al. 2009a) that in the case where all processes have a jointly multivariate Gaussian distribution, the measures are entirely equivalent (and that, furthermore, a stationary Gaussian autoregressive process must be linear (Barnett et al. 2010)). Where the measures differ markedly is in the type of data to which they are naturally applicable, and the ease with which empirical estimation may be effected. Granger causality, based as it is on linear regression, is in general not suited to causal inference of *discrete-valued* data. On the other hand, for continuous-valued data, estimation of Granger causality is generally straightforward, as the comprehensive and well-understood machinery of linear regression analysis applies. There are, furthermore, mature and reliable software packages available for Granger casual estimation (Cui et al. 2008; Seth 2010). Further advantages of Granger causality are that (i) insofar as it is model-based with a known likelihood function, standard techniques of *model order estimation* (such as the Akaike or Bayesian Information Criteria (McQuarrie and Tsai 1998)) may be deployed, and (ii) asymptotic distributions for the sample statistic are known. For transfer entropy it is not clear how much history should be included in estimates, nor how the statistic may be easily significance tested, beyond standard non-parametric (but computationally intensive) methods such as permutation testing (Anderson and Robinson 2001; Edgington 1995). However, recent work in this area by Barnett and Bossomaier (2012), goes some way to resolving these issues. If a fairly general parametric model is assumed, then the transfer entropy again becomes a log-likelihood ratio, as for the Granger Causality. Moreover, its distribution is χ^2 , enabling straightforward statistical estimates.

It is also known (Kaiser and Schreiber 2002) that, for continuous variables, in-sample estimation of transfer entropy is problematic insofar as (in contrast to mutual information) the sample statistic in general fails to converge under naïve refining of state-space partitioning. Thus more sophisticated but less well-understood (and computationally expensive) techniques such as adaptive partitioning or kernel estimation must be used. Nonetheless, transfer entropy is attractive due to its non-parametric “model agnostic” nature, particularly when interactions are likely to be highly nonlinear and hence unsuited to linear modelling. Table 1 shows a summary comparison of the measures.

A promising application of transfer entropy is in the construction of information-theoretic *complexity* measures. In (Tononi et al. 1994) the authors introduce a “neural complexity” metric $C_N(X)$ designed to capture a notion of network complexity based on *integration/segregation balance* (Barnett et al. 2009b). The idea is that a multi-element dynamical system exhibiting “complex” behaviour will tend to lie somewhere between

Table 1 Comparison between transfer entropy and Granger causality

Feature	Transfer entropy	Granger causality
Parametric	No	Yes: linear VAR model
Predictive	No	Yes
Frequency decomposition	No	Yes
Transformation-invariance	Nonlinear filter	Linear filter
Estimation in sample	Hard	Easy
Model order estimation	Unknown	AIC, BIC, cross-validation...
Statistical distribution	Unknown	Known (asymptotic)
Software implementation	<i>Ad hoc</i>	Many available packages

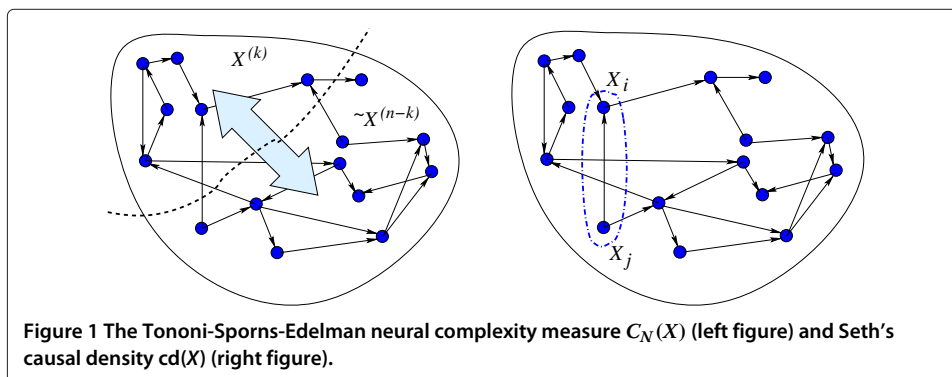
extremes of highly integrated, where every element tends to affect every other, and highly segregated, where elements behave almost independently. In the former case, a system will generally behave chaotically, while in the latter it will tend to decompose into simple independent processes (*cf.* “edge-of-chaos” phenomena (Langton 1990b)). These correspond to the random graph with few edges and the highly connected graph in section “Random graphs and phase transitions”. The complex behaviour lies at the phase transition, or connectivity avalanche.

The original Tononi-Sporns-Edelman measure $C_N(X)$ averages mutual information across bipartitions of the system (Figure 1, left-hand figure); it is, however, extremely unwieldy to calculate in practice and, moreover, fails to capture information *flow* as expounded above.

Seth (Seth et al. 2011, 2006) has developed an alternative Granger causality-based measure, *causal density*, which is both more computationally manageable and captures complexity as mediated by *time-directed* influences; it admits a transfer entropy-based analogue:

$$cd(X) \equiv \frac{1}{n(n-1)} \sum_{i \neq j} T_{X_i \rightarrow X_j | X_{[ij]}} \quad (14)$$

where n is the number of variables; i.e. causal density is the average of the pairwise-conditional information flows (7) (Figure 1, right-hand figure). Again, $cd(X)$ captures integration/segregation balance: for a highly integrated system the measure assumes a low value, since for each pair of variables X_i, X_j much of the information flow $X_i \rightarrow X_j$ is already resolved by the remaining variables $X_{[ij]}$, and conditioned out. For a highly segregated system the measure also assumes a low value, since the lack of coupling between variables results in comparatively few significant pairwise information flows.



Computational aspects of mutual information

There are numerical problems associated with the computation of mutual information. Of these there are two very obvious and related issues. The first is a consequence of the naive, i.e. fixed and equidistant, discretisation: some of the discrete bins (i) may contain no elements and therefore have a probability $p_i = 0$ while at the same time some $p_{i,j} \neq 0$ and in these cases the mutual information diverges, therefore both p_i and p_j need to be *absolutely continuous* (Lin 1991) with respect to each other to avoid such issues. Next we discuss how both of these issues can be addressed in practice, the first by defining a form of mutual information that is naturally continuous and therefore does not require an ad hoc discretisation step and the other addresses the case where, in a naturally discrete system numerical artefacts resulting in divergences can be avoided.

The choice of number of bins, bin boundaries and which estimator to use are all the subject of intensive research. Techniques such as adaptive partitioning (Cellucci et al. 2005; Darbellay and Vajda 1999) exist to optimise the bin selection. Bias-corrected estimators from histograms are also numerous, including shrinkage (Hausser and Strimmer 2009), Panzeri-Treves (1996), the Nemenmen-Shafee-Bialek (Nemenman et al. 2002; Wolpert and Wolf 1995) estimator, quadratic extrapolation (Strong et al. 1998).

For this experiment, where large data sets are to be handled, the most computationally rapid choices are paramount. To this end, Wicks et al. (2007) selected equal-width bins, using an intuitive length scale of $2R$ as the width of the bin (with number of bins equal to the $L2R$). Discretising the signal into bins, we calculate MI using an uncorrected *plugin estimator* between (presumed scalar) variables X and Y , based on occupancy values of the joint histogram. The term “plugin estimator” is used in the standard statistical sense: we simply substitute our estimated values for the distribution into the formula for mutual information, rather than construct an estimator by other means such as maximum likelihood.

Cellucci et al. (2005) show that equal bins can give very poor results, particularly when the datasets are small. They propose an adaptive binning algorithm in which the bins for X and Y are sized so that each contains the same expected number of values. The computational overhead of this procedure is at worst $n \log n$, corresponding to sorting each of the X and Y values.

Explicitly, for the histogram matrix Q of bin occupancy counts, we estimate the joint probability mass function p by normalising

$$\hat{p}_{X,Y}(x,y) = \frac{Q_{xy}}{\sum_i \sum_j Q_{ij}} \quad (15)$$

Then marginal probabilities can be estimated

$$\hat{p}_X(x) = \sum_j \hat{p}_{X,Y}(x,j) \hat{p}_Y(x,y) = \sum_i \hat{p}_{X,Y}(i,y) \quad (16)$$

from which the mutual information follows from Eq. 4.

For N_E adaptive bins along each axis the MI estimate simplifies to

$$\hat{I}(X,Y) = \sum_x \sum_y \hat{p}(x,y) \log_2 N_e^2 \hat{p}(x,y) \quad (17)$$

where N_e is the number of bins.

This estimator is subject to finite sample bias (Treves and Panzeri 2008), but for the purely comparative purposes of our experiment is empirically sufficient, and computationally very cheap.

The first is that while mutual information is well defined for continuous systems as shown at least as early as (Shannon 1948a), this is often difficult to implement in practice as it does not give rise to an obvious algorithm that does not first involve an initial discretisation step, thereby losing the information content of possibly important relationships. The second issue is that of systems that are naturally continuous. The continuous analogue of entropy and mutual information given in equations 2 and 3, the derivation of the entropy can be found in (Shannon 1948b):

$$H_{cont}(X) = - \int_X p(x) \log(p(x)) dx \quad (18)$$

$$I_{cont}(X; Y) = \int_X \int_Y p(x)p(y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy \quad (19)$$

where x and y are continuous variables. Such continuous systems occur in research on channel coding (MacKay 2003), chaotic dynamics (Granger and Lin 1994) and signal detection theory (Parra et al. 1996) amongst many others. A natural first estimate of 19 is to divide the continuous spaces X and Y into fixed bin sizes of δX and δY and uses the count of the occupancy of each bin in order to estimate the probability of bin occupancy. This has been used successfully in many different studies that have discretised continuous variables but it has been pointed out that it is often difficult to estimate the mutual information with such an approach (Kraskov et al. 2004). Other estimators include Gaussian-kernel inference, (Moon et al. 1995), spline estimators (Daub et al. 2004), and k -nearest neighbour estimators (Kraskov et al. 2004).

An alternative that retains much of the continuous properties of $I_{cont}(X; Y)$ is provided in (Kraskov et al. 2004). The key idea is to measure the distance between element e_i and its nearest neighbour e_j in the x -direction and e_k in the y -direction. For every element $e_i, i \in \{1, \dots, N\}$ located in the continuous space $Z = X \times Y$ a rank ordering of distances between i and all other elements can be constructed based on a distance $d(e_i(x), e_j(x))$ (distance between elements in the x -direction) and $d(e_i(y), e_k(y))$ (distance between elements in the y direction). The number of points that lie within the vertical strip defined by $d(e_i(x), e_j(x))$ are then counted: n_x and similarly for $d(e_i(y), e_k(y))$: n_y . This is called $k = 1$ clustering as it is based on the nearest neighbour, in the case of the second nearest neighbour being used to define $d(e_i(x), e_k(x))$ and $d(e_i(y), e_k(y))$ then $k = 2$ etc. For a system with N interacting particles in $Z = X \times Y$ the mutual information is approximated by (this is $I^{(2)}$ in (Kraskov et al. 2004)):

$$I_{cont}(X; Y) \approx \psi(N) + \psi(k) - 1/k - \langle \psi(n_x) + \psi(n_y) \rangle \quad (20)$$

where $\psi(n)$ is the Digamma function: $\psi(n) = \Gamma^{-1}(n)d\Gamma(n)/dn$, $\Gamma(n) = (n - 1)!$ (as n is a simple counting function and so always an integer) and $\langle \dots \rangle$ is the average over all $i \in \{1, \dots, N\}$ and across all samples.

As a numerical approximation this is an effective method that eliminates many of the difficulties and scales as $O(n)$. Kraskov et al. (2004) have shown that it is an effective estimate of the mutual information between two coupled Gaussians where the exact solution is known as well as for gene expression analysis, independent component analysis and data clustering (Kraskov et al. 2004). However, some data is naturally discrete in nature

and the development of accurate measures of mutual information to accommodate this is an important area of recent research.

Kraskov's algorithm described above can be thought of as an adaptive partitioning technique in the sense that the k^{th} nearest neighbour decides the width of a bin counting technique. An alternative is to start by discretising the space $Z = X \times Y$ into a grid and then adapting the bin sizes such that each 'vertical' or 'horizontal' strip in phase space has the same number of elements. The joint probability is then the occupancy of the rectangles formed by the intersection of these equally populated strips in X - and Y -space.

The calculations used in Cellucci et al's work (Cellucci et al. 2005) starts with N , the number of elements in Z -space, and the number of partitions that will be used to divide up the X - and Y -axes, labelled N_e . Each i^{th} partition of the X -axis contains, by definition, a set $S_X(i)$ of size N/N_e elements, equally each j^{th} partition of the Y -axis contains a set $S_Y(j)$ of size N/N_e . The joint probability $P_{X,Y}(i,j) = \text{count}(S_Y(j) \cap S_X(i))/N$, i.e. the count of the size of the intersecting set of the X and Y partitions divided by the total number of elements in the system. The mutual information is then:

$$I(X; Y) \approx \sum_{i=1}^{N_e} \sum_{j=1}^{N_e} P_{X,Y}(i,j) \log(N_e^2 P_{X,Y}(i,j)) \quad (21)$$

where the N_e^2 term accounts for the case of statistically independent distributions of $P_X(i)$ and $P_Y(j)$.

Mutual information for phase transitions in a simple flocking model

Herding or flocking behaviour has been the subject of many studies and it serves here as an illustration of the computation of mutual information and its attendant difficulties. Despite the apparent simplicity of mutual information measures, there is no one simple way which works in general for every data set. In the fields of bioinformatics and neuroinformatics in particular, much research has been done on the estimating of entropies from samples. Dozens of estimators and many code-bases (Daub et al. 2004; Hausser and Strimmer 2009; Ince et al. 2009; Slonim et al. 2005) are available for the task.

For observations generated by a simple *parametric model*, the mutual information functional may sometimes be calculated analytically from the underlying distribution. However, in the case that we wish to use a non-parametric model for our distributions, the procedure is rather more complicated.

Vicsek model background

To eliminate the complexities of real-world data, we use a synthetic data set, the well-studied model of self-propelled particles (SPP) of Vicsek (Vicsek et al. 1995). The SPP model is one of many accepted to undergo a phase transition with varying noise as a control parameter.

The SPP algorithm is a special case of Reynold's "Boids" flocking algorithm, (Reynolds 1987), remarkable for the small set of rules required to produce its rich behaviour. It has the virtues of trivial implementation, topologically simple configuration space, as there are no repulsive forces, and a small number of control parameters. Moreover, there is no known closed-form analytic relationship between system order and control parameters, much as with many experimental data sets. Details of the model and analysis of its

phases of behaviour are extensively studied elsewhere (Aldana et al. 2007; Chaté et al. 2008; Czirók and Vicsek 2000; Grégoire and Chaté 2004; Wicks et al. 2007).

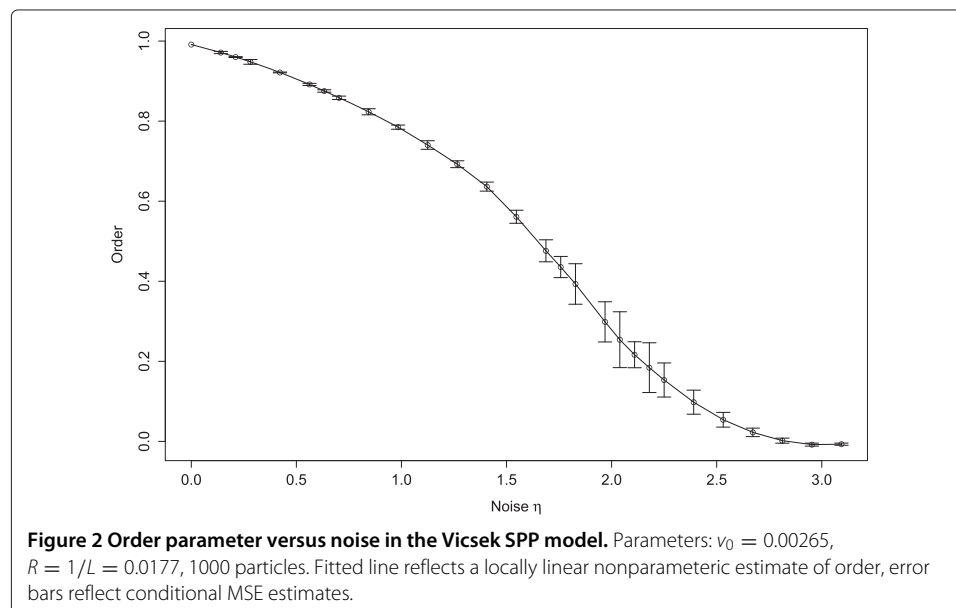
The SPP process is given as:

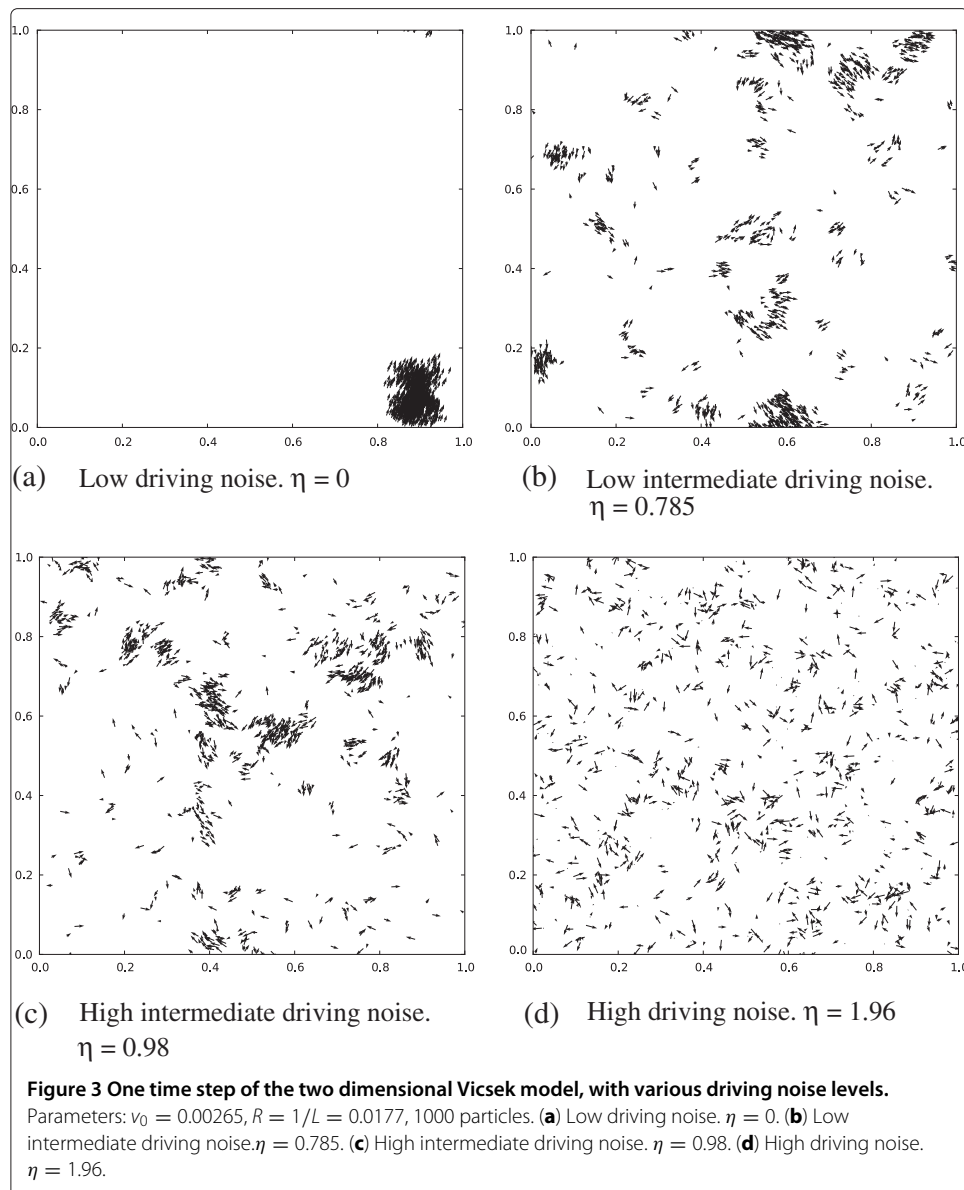
The only rule of the model is: at each time step a given particle driven with a constant absolute velocity assumes the average direction of motion of the particles in its neighborhood of radius r with some random perturbation added (Vicsek et al. 1995).

Vicsek's original specification is for an ensemble of particles moving in two spatial dimensions on the unit square with periodic (i.e. toroidal) boundary conditions. The model admits straightforward generalisation to other spatial dimensions, and alternate interaction topologies, but it is this original configuration that is used here. The system is parametrised by the temperature-analog noise parameter η , a fixed particle absolute speed v_0 , a number N of particles, an interaction radius, R , and a system side-length L . Particles travel at constant speed with random changes in direction as specified by the noise parameter. When two particles come within the interaction radius, their directions of movement are moved closer together. An *order parameter*, the magnitude of mean velocity (or "mean transport"), reflects the phase transition (Figure 2 with the side-length is set to 1 and particle interaction radius R to $1/L$).

Initial conditions assign to all particles positions uniformly at random on the unit square and velocities with a uniformly distributed angle. The simulation is run until the transient behaviours have died out, at which time the system reflects a degree of order dependent upon the noise – see Figure 3.

For different values of the noise control parameter, the system exhibits qualitatively different behaviour: for low noise parameters, the system exhibits spontaneous symmetry breaking, with particles tending to align in one specific direction. At high values of the noise parameter the particles approximate a random walk (and for maximal noise is precisely equal to a random walk). In between the system exhibitions transient regularity, and "clumpy" clustering.



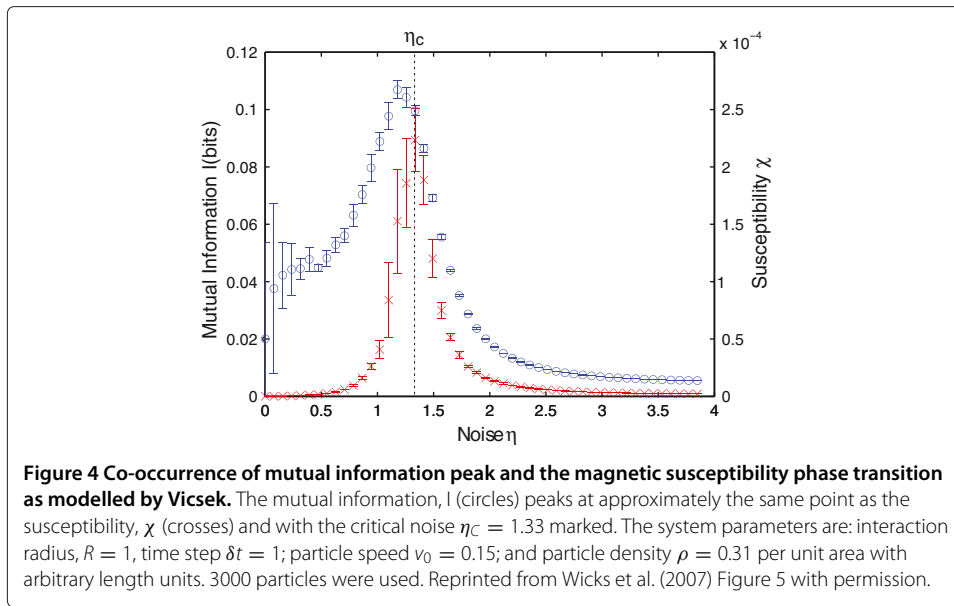


Wicks et al. (2007) then demonstrated under some simplifying assumptions that there was also a peak in mutual information around the same noise values as for the phase-transition. This is shown in Figure 4 reproduced from their paper.

This model, therefore, serves as a good illustration of the subtleties of calculating mutual information. It also serves as a possible heuristic for the economic phase transitions of section “Overview of phase transitions and metrics” whereby market traders can self-organise to ‘flock’ in their trading behaviour.

Phase transitions in socio-economic systems

The stock market is one complex system we would all like to understand and predict. Unfortunately it goes through bubbles and crashes, anticipated sometimes, but rarely precisely predicted. These phenomena *seem* like phase transitions: the market undergoes a

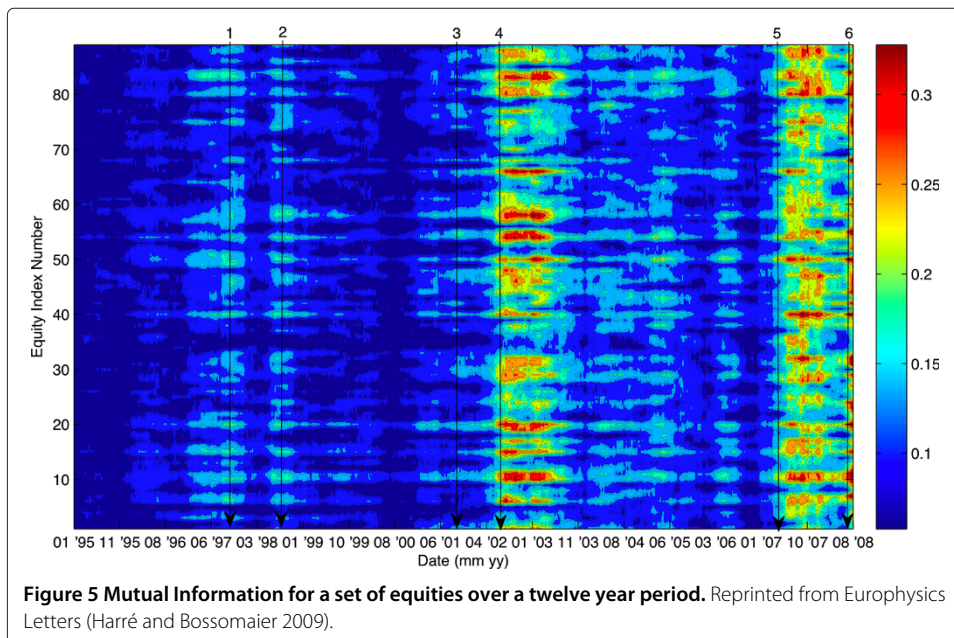


radical reorganisation. Harré and Bossomaier (2009) showed that they are indeed phase transitions, exhibiting a peak in mutual information.

They calculated mutual information between pairs of equities in five different industry and commercial sectors. Figure 5 shows the results for a period of over a decade.

For each equity the *daily deltas* were computed, the summary of the day's trading which indicates if the stock fell or rose overall. The distributions of these were then used to calculate the mutual information between stocks (Harré and Bossomaier 2009).

The vertical red band in late 2002 shows a peak in maximum mutual information for almost all equities. This corresponds to a significant correction in the Dow Jones index, but unlike other notable crashes does not have a name. In the bottom right hand corner



there is another red patch. The lowest numbered equities are the financial stocks. This fledgling tradition in late 2005 is undoubtedly linked to the subprime meltdown.

These are empirical results and as such there is no explicit order parameter. However, the Vicsek self-propelled particle model of section “Mutual information for phase transitions in a simple flocking model” provides an econophysical analogy of a stock market with two equities. If we consider each particle to be a trader and the axial position to be the perceived instantaneous investment value of a stock, then any trader has a view at any given time of the value of each stock. Her velocity is indicative of the rate of change of perception of value of each stock, and thus the trade likelihood. Since in most cases stock perceptions are cyclical, periodic boundary conditions are not too implausible.

Just as a change in order parameter can elicit behaviour change from the SPP model, so can an endogenous change in market optimism induce variation in stock market perceptions across all traders. As they approach the phase transition from the random side, stock movements become increasingly synchronised. In the ordered phase all traders are moving in the same direction and the market is surging or crashing.

The Vicsek phase transition is visible in the average *velocity* of the particles (traders). But empirically we observe the mutual information peak in the equity *prices*. We would expect this though. In the random phase there are no links between the trades of different equities and their prices are uncorrelated. In the ordered phase their prices changes are rigidly locked, but the entropy in their pricing is now very low. Thus the mutual information must peak somewhere in the intermediate state. Note that, in this interpretation, the phase transition would begin as the crash or bubble is forming: it is not the bubble or crash itself, but the need for sufficiently wide time windows makes this distinction moot.

The theoretical underpinnings of bifurcations and phase transitions in economics and finance have been around for many years. In the 1970's the mathematical framework of ‘catastrophe theory’ (Rosser 2007; Zeeman 1980) became a popular field of research as it provided one of the first formalisations of macro-economics that included a notion of both an equilibrium and non-linear state transitions (Zeeman 1974). This formalism provided a parsimonious description of bull and bear markets and market crash dynamics based on bullish bubbles using a small number of macro-economic parameters such as the proportion of chartists (traders who base their strategies on historical prices) and the proportion of fundamentalists (traders who base their strategies on the underlying business).

Such theoretical considerations have played an important role in socio-economic systems, but it was not until the onset of massive databases and high performance computing that it became possible to empirically study the ‘microscopic’ dynamics of the relationships between equities. Recent work has shown that there is an order parameter in financial markets (Plerou et al. 2003) (section “Overview of phase transitions and metrics”). This order parameter measures the net demand: before a phase transition, net market demand is zero; after the phase transition (when the market is no longer in equilibrium), the net demand either favours buying or selling.

Such *hysteresis* effects have been the basis of recent work in macro-economic models (Wolpert et al. 2012) as well. In this work a control parameter, mean tax rate, is varied in order to move from a low growth to a high growth game-theoretic equilibrium. Interestingly, the model applies to varying the parameter by either the market (free-market model) or a centralised government (socialist model).

Phase transitions in the acquisition of human expertise

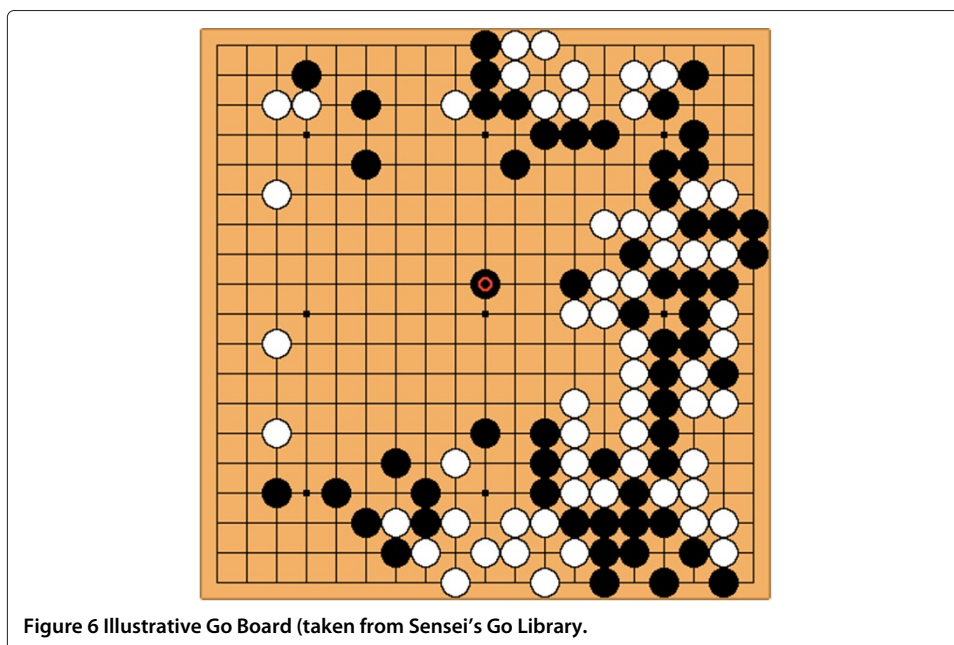
Thomas Kuhn in his famous book, *The Structure of Scientific Revolutions* (Kuhn 1962) discussed the idea of paradigm shifts in science or human knowledge, where everything is reorganised. Relativity and quantum mechanics were major paradigm shifts of the twentieth century. Much earlier Copernicus' idea, that planets travel around the sun, rather than everything around the earth, was a dramatic shift in thinking about the solar system.

Such shifts seem to occur in human thinking, where we learn to join the dots in different ways. Yet it is difficult to find ways to measure such changes. Since expertise requires a long time to develop, at least 10,000 hours according to Eriksson (1998), or the acquisition of 50,000 or more "chunks" according to Simon (Simon and Chase 1973), now thought to be as many as 300,000 (Gobet and Simon 2000). Thus any measurements on a single individual would have to take place over a long period.

Harré et al. found a solution to this using online data mining. Where decisions are recorded online, they can be analysed in large numbers, providing a quite different way of inferring shifts in human thinking and knowledge organisation. To do this they used the game of Go. This game is extraordinarily simple in structure and rules, but is as demanding for human players as chess. Moreover, the best computer programs do not come close to human experts at the time of writing in early 2012.

Figure 6 shows a sample Go board. Black and white stones are placed on the intersections, or *points* of a 19x19 grid. Black begins anywhere in the board, but typically in one of the corners. Stones do not move. They are removed from the board when they die. Stones require a contact along the grid, either directly, or via other stones of the same kind, with an unoccupied point, or *liberty*. This simple setup defines one of the oldest and most challenging of all board games.

Such developmental transitions have been implicated in the activation networks used to model human cognition as well as artificial intelligence systems (Shrager et al. 1987).



These models, using activation networks, emphasise the role of network topology in how information is accessed, implying that as topologies of associative networks change via learning there is the possibility of a network-wide phase transition occurring. Since this earlier work, many other theoretical models have argued that cognitive structures undergo significant reorganisation at intermediate stages of development. This has included models for epileptic fits (Percha et al. 2005) and clustering algorithms used to learn complex data relationships (Bakker and Heskes 2003).

Finding such transitions in cognitive data is more difficult though, although there has been some evidence of non-linear changes through work on the ‘inverted-U’ effect in novice-skilled-expert comparative studies (Rikers 2000). This effect is based on the observation that while increasing skill increases the quality of decisions (in medical practitioners, for example), other factors, such as the recall of individual case information and the ability to elaborate more extensively on such cases, peaks for the ‘skilled’ subjects but were equally low for both the ‘expert’ and the ‘novice’ groups. Such inverted-U observations have been made in chess (Gobet 1997), meditation practitioners (Brefczynski-Lewis et al. 2007) and emotional responses to art (Silva 2005) This work implies an intermediate point where cognitive organisation changes significantly, but as many studies only have a small number of skill levels, i.e. three: novice-skilled-expert, dramatic changes in a dependent variable such as depth of search or recall is often difficult to observe.

The use of entropy as an impicator of phase transitions in cognitive studies has also had some success in recent studies. The developmental transition of generalising a mechanical manipulation into a mathematical insight of the underlying logic, an ‘a-ha!’ moment has recently been reported using entropy and based on notions of self-organising criticality (Stephen et al. 2009). In this direction, some of the most exciting work has been carried out in transfer entropy (Dixon et al. 2010) applied to self-organising criticality and how it is the entropy that drives massive transitional changes in cognitive structures.

Finally, in a more basic experimental paradigm, Dutilh et al. (2010). have used the speed-accuracy trade-off and Catastrophe Theory in simple decision making to postulate that even some of our most primitive decision making processes might implicate phase transition-like behaviour. (See section “Overview of phase transitions and metrics”).

To find phase transitions in the development of expertise we use the same metric, a peak in mutual information. The order parameter is the level of expertise. In Go this is measured in Dan ranks, up to 8 Dan Amateur and 9 Dan professional. Up to 1 Dan Amateur has a separate set of ranks, 26 kyu, with 26 being the weakest and 1 the strongest.

For each rank Harré et al. studied a game tree – every possible move that can happen in a small region (7×7). The moves within the region are taken from actual games, thus they do not have to be sequential: a player may play somewhere else and come back to the region of analysis later.

We need three probability distributions to calculate the mutual information. Firstly there is the joint probability, $p(q, m)$, where q is a position and m is a move made at that position. Then we need the marginal probabilities, $p(q)$ and $p(m)$ of the position and move occurring respectively. For a 7×7 region there are approximately 3^7 possible positions. Some of these will be illegal, but, more importantly, many of them will never occur in actual play. Thus the analysis is tractable.

Figure 7 shows the mutual information as a function of rank. The MI peaks at around the transition from amateur to professional, agreeing with other evidence that a radical shift in strategic thinking occurs at this juncture.

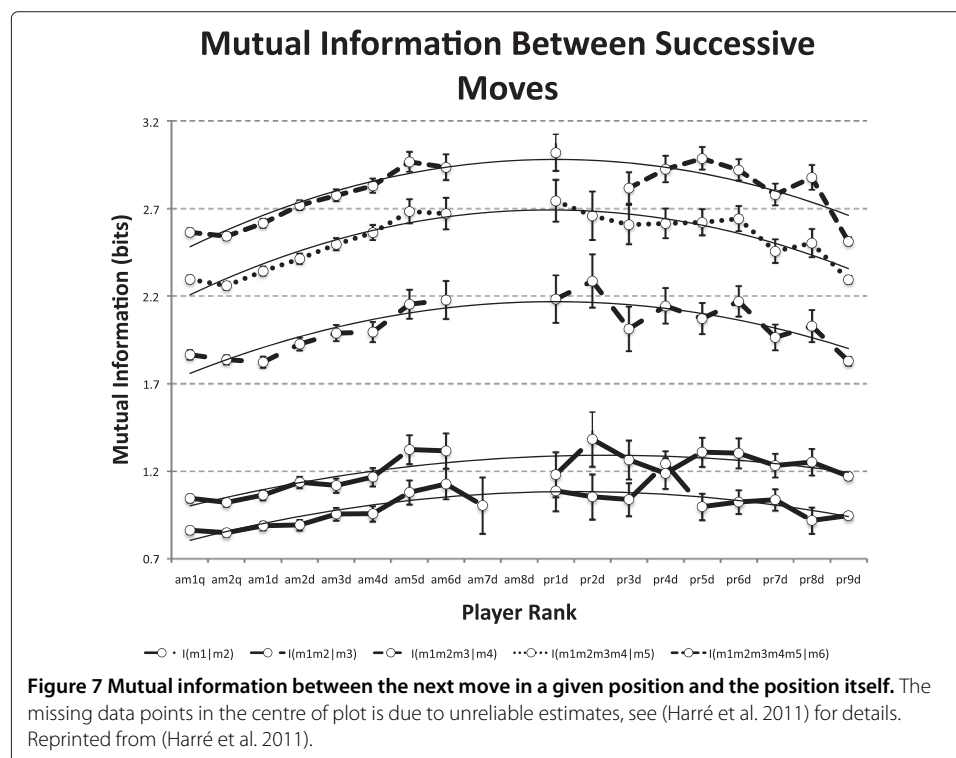
Inferring social networks with transfer entropy

Social networks have rapidly become one of the dominant features of today’s world. Understanding when such networks undergo phase transitions can be very important, such as in modelling the spread of disease or public opinion at election time. Numerous tools have been developed to measure information flowing between agents in a network, such as the analysis of email traffic (Kossinets and Watts 2006). But in some cases no direct measures of connectivity are available.

We encountered such a situation in customer records in the financial industry. Hence we developed techniques based on mutual information and transfer entropy of investment time series. This section discusses this methodology.

In (Bossomaier et al. 2010) a detailed data set of 42 million records describing the investment profiles of 1.5 million customers over a 24 month period was analysed with the aim of understanding the social networks among clients. In that study, pairwise (unconditional) mutual information between investment histories—lagged and unlagged—was calculated with the aim of identifying relationships between investment behaviour patterns that could be ascribed to social interaction.

Given that lagged mutual information is likely to be a misleading indicator of time-directed information flow (see Section “Transfer entropy and granger causality”), the study was recently repeated using transfer entropy. The exercise highlighted several features and difficulties with the practicalities of estimating and significance testing transfer



entropy in large datasets. Critically, while a large number of investment history time series were available, they were of short length; in practice only about 20 months' data was generally available per client record. This was largely due to the significant fraction of *missing data*, necessitating a principled approach to the handling of missing data in statistical estimation. Initially investment histories with more than 4 months' missing data were excluded; all subsequent analysis was performed on a per-product basis. It was found that (again due largely to the short length of histories) there were many duplicate time series within a product. These are statistically indistinguishable, so only *unique* time series were analysed, each corresponding to a specific group of customers within the given product. The final number of unique histories was of the order of 500–5000 per investment product.

As in the original study, it was deemed that the actual magnitude of monthly investments was not relevant; monthly investment states were thus classified simply into '+' (account credited), '-' (account debited) or '0' (no change). This discretisation of investment states (a practice commonplace in econometric analysis) also makes estimation of transfer entropies less problematic (*cf.* Section "Transfer entropy and granger causality"). The stance taken on remaining missing data was that it should be "conditioned out"; that is, statistics were estimated *conditional on all relevant investment states being valid (non-missing)*. Furthermore, as an attempt to control for common influences on all customers (within the given population of investment histories), transfer entropy was conditioned on a *consensus sequence*, C_t obtained by taking the most prevalent valid state in the population at each time step. Thus conditional transfer entropy was calculated as

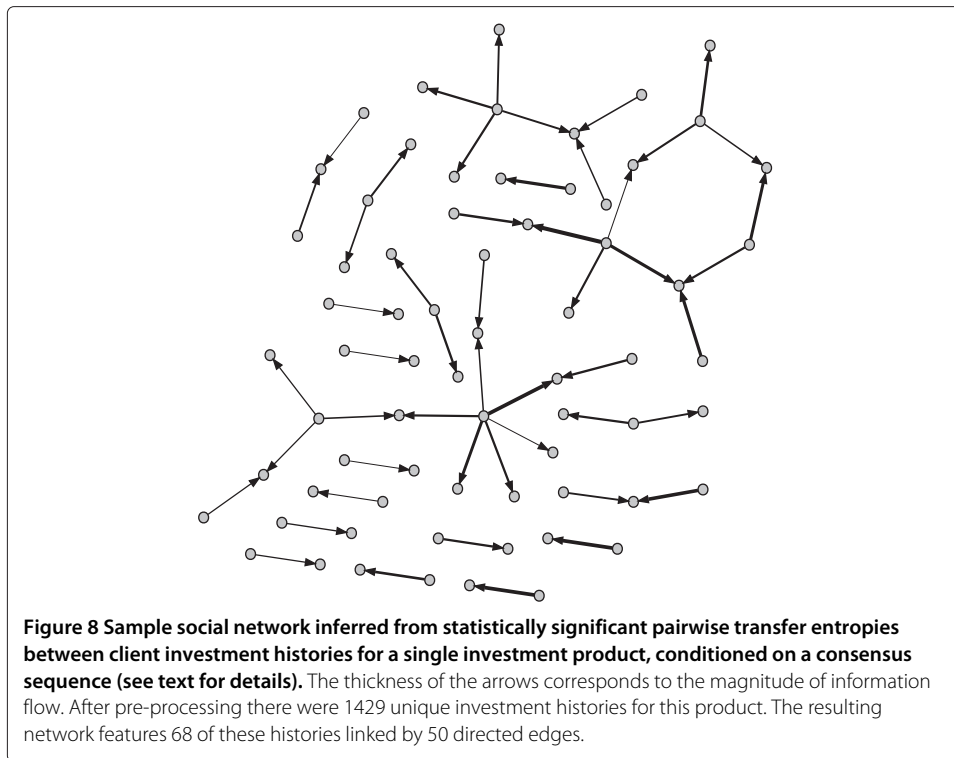
$$T_{Y \rightarrow X|C} \equiv I(X_t : Y_{t-1} | X_{t-1}, C_{t-1} \text{ and } X_t \neq *, X_{t-1} \neq *, Y_{t-1} \neq *) \quad (22)$$

where '*' denotes missing data. Due again to short history length, only one lag of history was taken into consideration. In sample, (22) was calculated by estimating the requisite probabilities as frequencies.

In order to derive networks of causal information flow, we calculated pairwise transfer entropies conditioned on the appropriate consensus sequence between all pairs of time-series X_i, X_j in the selected populations. These were then tested for statistical significance (see below) and significant information flows presented as directed graphs, weighted by the actual value of the transfer entropy statistic (Figure 8).

Note, however, that due to the large number of time series it was not possible to calculate pairwise-conditional statistics [Section "Transfer entropy and granger causality", eq. (7)]. Thus if there is e.g. significant information flow from $X_i \rightarrow X_j$ and also from $X_j \rightarrow X_k$ then it is likely that the information flow $X_i \rightarrow X_k$ will appear as significant too, even if there is no *direct* information flow from X_k to X_i ; i.e. the apparent information flow $X_i \rightarrow X_k$ is intermediated by X_j .

Pairwise information statistics were permutation tested to establish significance, against the null hypothesis of zero information flow. For each statistic, 1000 random permutations of the time series of the causal variable were generated (missing data is held in-place during permutation) to disrupt any possible causal effects. The unpermuted statistic was then tested against the resulting surrogate null distribution to derive a p-value, representing the probability that a value at least as large as the statistic being tested might be obtained by chance under the null hypothesis. A subtlety is that, while the number of (unique) values for a permuted statistic can be quite small—many different



permuted sequences will in general give rise to the same sample statistic—some values are very rare, and will frequently not be discovered by any of the 1000 permutations. It is thus possible that a p-value can come out as zero; this is patently unsatisfactory, as it would reject the null hypothesis at *any* significance level, thus giving rise to Type I errors (false positives). To address this issue, we use the fact that the *maximum* sample information flow statistic will be obtained when the (lagged) causal sequence is *identical* to the causee sequence. In general there will be only one possible permutation with this property, which thus occurs with probability^f

$$p_{I_{max}} \equiv \frac{n_0! n_+! n_-!}{(n_0 + n_+ + n_-)!} \quad (23)$$

where n_0 , n_+ and n_- are the number of 0, + and – states respectively in the investment history sequence being tested. The following procedure was implemented to mitigate the effects of spurious zero p-values: if a p-value was empirically calculated to be zero—i.e. the test statistic was larger than all permutation statistics—the resulting p-value was set to $p_{I_{max}}$ rather than zero. This does not preclude the possibility that the “true” p-value is actually larger than $p_{I_{max}}$, but can at least be expected to reduce substantially the number of false-positives that might otherwise arise due to zero p-values.

A further issue to be addressed is that we are performing *multiple* hypothesis tests (Miller 1981); i.e. one for each pairwise information flow statistic within the population under consideration. Under this scenario, at any given significance level, α , we would expect approximately $\alpha \times$ (number of pairwise statistics) Type I errors (false positives). There are various established approaches to controlling for this effect, which generally require a stronger level of evidence for rejection of null hypotheses. Unfortunately, in our situation where the number of simultaneous hypothesis tests may be very large [it will be

Table 2 Number of significant information flow sample statistics by product at significance level $\alpha = 0.01$, with per-time-series family-wise error rate correction (see text)

Product	LMI uncond	LMI cond	TE uncond	TE cond
1	29449	2971	7540	1986
2	739	109	179	75
3	191	94	112	50
4	112	116	90	31
5	12	19	30	15
6	38	27	25	15
7	60	6	20	8
8	7	10	6	4
9	52	7	18	3

LMI = Lagged mutual information, TE = Transfer entropy; conditional/unconditional refers to conditioning on the population consensus sequence.

$n(n - 1)$ where n is the number of unique time series in the test population] common approaches such as controlling for *false discovery rate* or *family-wise error rate* (Benjamin and Hochberg 1995) are highly likely to yield *no* significant statistics at acceptable significance levels; essentially, Type I errors will have been traded off for an unacceptably high Type II (false negative) error rate. As a compromise, for each time-series X_j in the given population, we estimated significance for information flow statistics by controlling the family-wise error rate among *all time series that might potentially exhibit a significant causal influence on series X_j* - i.e. for all statistics $X_i \rightarrow X_j$. Family-wise error rates were calculated at a significance level $\alpha = 0.01$ using the well-known *Bonferroni correction* (Miller 1981).

Numbers of significant information flows under this scheme are displayed in Table 2. We see that conditioning out the common influence represented by the consensus sequence usually (but not always) reduces the number of significant statistics. Transfer entropy generally yields fewer significant information flows; this is consistent with the known deficiencies of lagged mutual information as a measure of directed information flow, whereby spurious information flows may be reported due to shared historical influences (Section “Transfer entropy and granger causality”).

Worthwhile future work on this study would include comparison of the *causal density* dynamic complexity measure (Section “Transfer entropy and granger causality”) between investor networks; however, this presents a technical challenge with regard to the difficulties mentioned above in obtaining true pairwise-conditional transfer entropies.

Conclusions

Mutual information is a useful indicator of phase transitions. It peaks in the same region as other indicators, such as the magnetic transition in the Vicsek model. We have shown that the calculation of mutual information is fraught with difficulty, but it can be used with care to find phase transitions in socio-economic and cognitive systems.

Transfer entropy, a conditioned extension of lagged mutual information, is closely related to mutual information and is a powerful new technique. It can be used to infer causal information flows within complex systems (Lizier 2008) and holds out the possibility of being able to predict phase transitions before they occur. Of particular interest for future study would be to investigate the behaviour of the information-theoretic dynamical complexity measures described in Section “Transfer entropy and granger causality”

with regard to phase transitions and the application to socio-economic systems from organisational change to the onset of recessions.

Endnotes

^aThese informational properties can be extended, with some qualifications, to continuous random variables as well — see, for example, (Gray 1991) — but as random variable considered herein are either discrete or discretised before analysis, these extensions will not be discussed here. Generalisations of mutual information to more than two variables also exist.

^bHere, and below, we leave ambiguous the number of lags to be included in expressions of this type; in principle one might include the *entire* past of a process, while for empirical estimation (or on domain-specific grounds) lags may be truncated at some finite number.

^cIt is a standard result in time series analysis that maximum likelihood estimation of the regression parameters in (8) is equivalent to minimisation of the “total variance” $\text{trace}(\text{cov}(\varepsilon_t))$, e.g. by a standard ordinary least squares (OLS) (Hamilton 1994). Other equivalent approaches are by solution of the Yule-Walker equations for the regression, e.g. via the LWR algorithm or one of its variants (Morettin 1984).

^dIn the case of a *univariate* predictee variable X , the Granger statistic is sometimes defined as the R^2 -like statistic $\exp(F_{Y \rightarrow X}) - 1$, which has an asymptotic F - rather than a χ^2 null distribution (Hamilton 1994).

^eIn fact this is probably the most commonly encountered variant of Granger causality, at least in the neuroscience literature; confusingly, it is frequently this quantity that is referred to as “multivariate” (conditional) Granger causality, as opposed to the case $F_{Y \rightarrow X|Z}$ where the individual variables X, Y, Z are themselves multivariate.

^fThis will not be precisely the case if e.g. there number of + states and – states in the sequence is the same; in this case the permutation derived by swapping + and – states will yield an additional maximum information sequence. We do not believe this affects significance test results unduly.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

All authors read and approved the final manuscript.

Acknowledgements

This work was supported by US Air Force grant AOARD 104116, Australian Research Council Grants LP0453657 and DP0881829. Lionel Barnett is supported by the Dr Mortimer and Theresa Sackler Foundation. Dan Mackinlay produced the Vicsek diagrams and contributed helpful discussions.

Author details

¹Centre for Research in Complex Systems, Charles Sturt University, Panorama Ave, Bathurst NSW 2795, Australia.

²Department of Informatics, University of Sussex, Falmer, Brighton BN1 9QH, UK. ³Department of Civil Engineering, University of Sydney, Sydney, NSW, Australia.

Received: 7 November 2012 Accepted: 28 January 2013

Published: 8 April 2013

References

- Aldana, M, Dossetti V, Huepe C, Kenkre VM, Larralde H: **Phase transitions in systems of self-propelled agents and related network models.** *Phys Rev Lett* 2007, **98**:095702.
- Anderson, MJ, Robinson J: **Permutation tests for linear models.** *Aust NZ J Stat* 2001, **43**:75–88.
- Bakker, B, Heskes T: **Clustering ensembles of neural network models.** *Neural Netw* 2003, **16**(2):261–269.
- Barnett, L, Bossomaier T: **Transfer entropy as a log-likelihood ratio.** *Phys Rev Lett* 2012, **109**:138105.
- Barnett, L, Seth AK: **Behaviour of Granger causality under filtering Theoretical invariance and practical application.** *J Neurosci Methods* 2011, **201**(2):404–419.

- Barnett, L, Barrett AB, Seth AK: **Granger causality and transfer entropy are equivalent for Gaussian variables.** *Phys Rev Lett* 2009a, **103**(23):238701.
- Barnett, L, Buckley CL, Bullock S: **Neural complexity and structural connectivity.** *Phys Rev E* 2009b, **79**(5):51914.
- Barrett, AB, Barnett L, Seth AK: **Multivariate Granger causality and generalized variance.** *Phys Rev E* 2010, **81**(4):41907.
- Benjamini, Y, Hochberg Y: **Controlling the false discovery rate: A, practical and powerful approach to multiple testing.** *J Roy Stat Soc B* 1995, **57**:289–300.
- Bossomaier, T, Standish RK, Harré M: **Simulation of trust in client-wealth management adviser relationships.** *Int J Simul Process Model* 2010, **6**:40–49.
- Brefczynski-Lewis, J, Lutz A, Schaefer H, Levinson D, Davidson R: **Neural correlates of attentional expertise in long-term meditation practitioners.** *Proc Natl Acad Sci* 2007, **104**(27):11483.
- Cellucci, CJ, Albano AM, Rapp PE: **Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms.** *Phys Rev E* 2005, **71**(6):66208.
- Chaté H, Ginelli F, Grégoire G, Raynaud F: **Collective motion of self-propelled particles interacting without cohesion.** *Phys Rev E* 2008, **77**:046113.
- Cover, TM, Thomas JA: *Elements of Information Theory.* Wiley Series in Telecommunications and Signal Processing. New York: Wiley-Interscience; 2006.
- Cui, J, Lei X, Bressler SL, Ding M, Liang H: **BSMART: a Matlab/C toolbox for analysis of multichannel neural time series.** *Neural Netw, Special Issue Neuroinformatics* 2008, **21**:1094–1104.
- Czirók, A, Vicsek T: **Collective behavior of interacting self-propelled particles.** *Physica A: Stat Theor Phys* 2000, **281**:17–29.
- Darbellay, GA, Vajda I: **Estimation of the information by an adaptive partitioning of the observation space.** *IEEE Trans Inf Theory* 1999, **45**:1315–1321. [http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=761290]
- Dai, L, Vorselen D, Korolev KS, Gore J: **Generic indicators for loss of resilience before a tipping point leading to population collapse.** *Science* 2012, **336**(6085):1175–1177.
- Daub, CO, Steuer R, Selbig J, Kloska S: **Estimating mutual information using B-spline functions - an improved similarity measure for analysing gene expression data.** *BMC Bioinformatics* 2004, **5**:118.
- Ding, M, Chen Y, Bressler S: **Granger causality: Basic theory and application to neuroscience.** In *Handbook of Time Series Analysis.* Edited by Schelter, S, Winterhalder M, Timmer J. Weinheim: Wiley; 2006:438–460.
- Dutilh, G, Wagenmakers E, Visser I, Van Der Maas H: **A phase transition model for the speed-accuracy trade-off in response time experiments.** *Cogn Sci* 2010.
- Dixon, J, Stephen D, Boncoddò R, Anastas J: **The self-organization of cognitive structure.** *Psychol Learn Motiv* 2010, **52**:343–384.
- Edgington, ES: *Randomization Tests.* New York: Marcel Dekker; 1995.
- Erdős, P, Rényi A: *On the Evolution of Random Graphs.* Budapest: The Hungarian Academy of Sciences; 1960.
- Ericsson, K: **The scientific study of expert levels of performance general implications for optimal learning and creativity 1.** *High Ability Studies* 1998, **9**:75–100.
- Geweke, J: **Measurement of linear dependence and feedback between multiple time series.** *J Am Stat Assoc* 1982, **77**(378):304–313.
- Geweke, J: **Measures of conditional linear dependence and feedback between time series.** *J Am Stat Assoc* 1984, **79**(388):907–915.
- Gobet, F: **A pattern-recognition theory of search in expert problem solving.** *Thinking Reasoning* 1997, **3**(4):291–313.
- Gobet, F, Simon H: **Five seconds or sixty? Presentation time in expert memory.** *Cogn Sci* 2000, **24**:651–682.
- Granger, CWJ: **Investigating causal relations by econometric models and cross-spectral methods.** *Econometrica* 1969, **37**:424–438.
- Granger, C, Lin J: **Using the mutual information coefficient to identify lags in nonlinear models.** *J Time Ser Anal* 1994, **15**(4):371–384.
- Gray, RM: *Entropy and Information Theory.* New York: Springer-Verlag; 1991. [<http://ee.stanford.edu/~gray/it.html>]
- Green, D, Newth D: **Towards a theory of everything: grand challenges in complexity and informatics.** *Complexity Int* 2005, **8**. [<http://www.complexity.org.au/ci/vol08/index.html>]
- Grégoire, G, Chaté H: **Onset of collective and cohesive motion.** *Phys Rev Lett* 2004, **92**(2):025702.
- Gu, SJJ, Sun CPP, Lin HQQ: **Universal role of correlation entropy in critical phenomena.** *J Phys A: Math Gen* 2006. [<http://arxiv.org/pdf/quant-ph/0605164>]
- Hamilton, JD: *Time Series Analysis.* Princeton: Princeton University Press; 1994.
- Hausser, J, Strimmer K: **Entropy inference and the James-Stein estimator with application to nonlinear gene association networks.** *J Mach Learn Res* 2009, **10**:1469–1484.
- Harré, M, Bossomaier T: **Phase-transition – behaviour of information measures in financial markets.** *Europhysics Lett ERA A* 2009, **87**:18009.
- Harré, M, Bossomaier T, Gillett A, Snyder A: **The aggregate complexity of decisions in the game of go.** *Eur Phys J B ERA A* 2011, **80**:555–563.
- Ince, RA, Petersen RS, Swan DC, Panzeri S: **Python for information theoretic analysis of neural data.** *Front Neuroinformatics* 2009, **3**. [<http://www.hubmed.org/display.cgi?uids=19242557>]
- Kaiser, A, Schreiber T: **Information transfer in continuous processes.** *Physica D* 2002, **166**:43–62.
- Kraskov, A, Stögbauer H, Grassberger P: **Estimating mutual information.** *Phys Rev E* 2004, **69**:066138–066153.
- Kossinets, G, Watts DJ: **Empirical analysis of an evolving social network.** *Science* 2006, **311**(5757):88–90.
- Kuhn, T: *The Structure of, Scientific Revolutions:* University of Chicago Press; 1962.
- Langton, C: **Computation at the edge of chaos: Phase transitions and emergent computation.** *Physica D: Nonlinear Phenomena* 1990a, **42**:12–37.
- Langton, CG: **Computation at the edge of chaos.** *Physica D* 1990b, **42**:12–37.
- Lin, J: **Divergence measures based on the Shannon entropy.** *Inf Theory, IEEE Trans* 1991, **37**:145–151.

- Lizier, J, Prokopenko M, Zomaya A: **A framework for the local information dynamics of distributed computation in complex systems.** *Phys Rev E* 2008, **77**:026110.
- MacKay, D: *Information Theory, Inference, and, Learning Algorithms.* Cambridge: Cambridge University Press; 2003.
- Matsuda, H, Kudo K, Nakamura R, Yamakawa O, Murata T: **Mutual information of Ising systems.** *Int J Theor Phys* 1996, **35**(4):839–845.
- McQuarrie, ADR, Tsai CL: *Regression and, Time Series Model Selection.* Singapore: World Scientific Publishing; 1998.
- Miller, RG: *Simultaneous Statistical, Inference.* New York: Springer-Verlag; 1981.
- Moon, Yi, Rajagopalan B, Lall U: **Estimation of mutual information using kernel density estimators.** *Phys Rev E* 1995, **52**:2318–2321.
- Moretting, PA: **The Levinson algorithm and its applications in time series analysis.** *Int Stat Rev* 1984, **52**:83–92.
- Nemenman, I, Shafee F, Bialek W: **Entropy and inference, revisited.** In *Advances in Neural Information Processing Systems 14 Volume 14.* Edited by Dietterich, TG, Becker S, Ghahramani Z, Cambridge, The MIT Press; 2002.
- Parra, L, Deco G, Miesbach S: **Statistical independence and novelty detection with information preserving nonlinear maps.** *Neural Comput* 1996, **8**(2):260–269.
- Panzeri, S, Treves A: **Analytical estimates of limited sampling biases in different information measures.** *Netw Comput Neural Syst* 1996, **7**:87–107. [http://www.ingentaconnect.com/content/apl/network/1996/00000007/00000001/art00006]
- Percha, B, Dzakpasu R, Zochowski M, Parent J: **Transition from local to global phase synchrony in small world neural network and its possible implications for epilepsy.** *Phys Rev E* 2005, **72**(3):031909.
- Plerou, V, Gopikrishnan P, Stanley H: **Two-phase behaviour of financial markets.** *Nature* 2003, **421**(6919).
- Prokopenko, M, Lizier JT, Obst O, Wang XR: **Relating fisher information to order parameters.** *Phys Rev E* 2011, **84**(4):041116. [http://lizier.me/joseph/publications/2011-Prokopenko-RelatingFisherInfoToOrderParams.pdf]
- Reshef, DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC: **Detecting novel associations in large data sets.** *Science* 2011, **334**(6062):1518–1524.
- Reynolds, CW: **Flocks, herds and schools: A distributed behavioral model.** In *SIGGRAPH '87 Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, Volume 21.* New York: ACM; 1987:25–34.
- Rikers, R, Schmidt H, Boshuizen H: **Knowledge encapsulation and the intermediate effect.** *Contemp Educ Psychol* 2000, **25**(2):150–166.
- Rosser Jr, J: **The rise and fall of catastrophe theory applications in economics: Was the baby thrown out with the bathwater?.** *J Econ Dyn Control* 2007, **31**(10):3255–3280.
- Scheffer, M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, van de Koppel J, van de Leemput IA, Levin SA, van Nes EH, Pascual M, Vandermeer J: **Anticipating critical transitions.** *Science* 2012, **338**(6105):344–348.
- Schreiber, T: **Measuring information transfer.** *Phys Rev Lett* 2000, **85**(2):461–464.
- Seth, AK: **A MATLAB toolbox for Granger causal connectivity analysis.** *J Neurosci Methods* 2010, **186**:262–273.
- Seth, AK, Barrett AB, Barnett L: **Causal density and integrated information as measures of conscious level.** *Phil Trans R Soc A* 2011, **369**:3748–3767.
- Seth, AK, Izhikevich E, Reeke GN, Edelman GM: **Theories and measures of consciousness: an extended framework.** *Proc Natl Acad Sci USA* 2006, **103**(28):10799–10804.
- Shannon, CE: **A mathematical theory of communication.** *Bell Syst Tech J* 1948a, **27**:379–423.
- Shannon, C: **The bell system technical journal 27.** *Math Theory Commun* 1948b:379–423.
- Shrager, J, Hogg T, Huberman B: **Observation of phase transitions in spreading activation networks.** *Science* 1987, **236**(4805):1092.
- Silvia, P: **Emotional responses to art: From collation and arousal to cognition and emotion.** *Rev Gen Psychol* 2005, **9**(4):342.
- Simon, H, Chase W: **Skill in Chess: Experiments with chess-playing tasks and computer simulation of skilled performance throw light on some human perceptual and memory processes.** *Am Sci* 1973, **61**(4):394–403.
- Slonim, N, Atwal GS, Tkačik G, Bialek W: **Information-based clustering.** *Proc Natl Acad Sci* 2005, **102**:18297–18302. [http://www.hubmed.org/display.cgi?uids=16352721]
- Solé R, Manrubia S, Luque B, Delgado J, Bascompte J: **Phase transitions and complex systems.** *Complexity* 1996, **1**(4):13–26.
- Sornette, D, Johansen A: **Significance of log-periodic precursors to financial crashes.** *eprint arXiv:cond-mat/0106520* 2001.
- Stephen, D, Boncoddio R, Magnuson J, Dixon J: **The dynamics of insight Mathematical discovery as a phase transition.** *Mem Cogn* 2009, **37**(8):1132–1149.
- Strong, SP, Koberle R, de Ruyter van Steveninck RR, Bialek W: **Entropy and information in neural spike trains.** *Phys Rev Lett* 1998, **80**:197–200. [http://arxiv.org/abs/cond-mat/9603127]
- Tononi, G, Sporns O, Edelman GM: **A measure for brain complexity Relating functional segregation and integration in the nervous system.** *Proc Natl Acad Sci USA* 1994, **91**:5033–5037.
- Treves, A, Panzeri S: **The upward bias in measures of information derived from limited data samples.** *Neural Comput* 2008, **7**(2):399–407.
- Vicsek, T, Czirók A, Ben-Jacob E, Cohen I, Shochet O: **Novel type of phase transition in a system of self-driven particles.** *Phys Rev Lett* 1995, **75**:1226–1229. [http://arxiv.org/abs/cond-mat/0611743]
- Wicks, RT, Chapman SC, Dendy R: **Mutual information as a tool for identifying phase transitions in dynamical complex systems with limited data.** *Phys Rev E* 2007, **75**: [http://arxiv.org/pdf/physics/0612198]
- Wiener, N: **The theory of prediction.** In *Modern Mathematics for Engineers.* Edited by Beckenbach, EF. New York: McGraw Hill; 1956:165–190.
- Wilks, SS: **The large-sample distribution of the likelihood ratio for testing composite hypotheses.** *Ann Math Stat* 1938, **6**:60–62.
- Wilms, J, Troyer M, Verstraete F: **Mutual information in classical spin models.** *J Stat Mech: Theory Exp* 2011, **2011**:P10011. [http://arxiv.org/abs/1011.4421]

- Wolpert, DH, Wolf DR: **Estimating functions of probability distributions from a finite set of samples.** *Phys Rev E* 1995, **52**(6):6841–6854. [http://arxiv.org/abs/comp-gas/9403001]
- Wolpert, DH, Harré M, Olbrich E, Bertschinger N, Jost J: **Hysteresis effects of changing the parameters of noncooperative games.** *Phys Rev E* 2012, **85**:036102. [http://link.aps.org/doi/10.1103/PhysRevE.85.036102]
- Zeeman, E: **On the unstable behaviour of stock exchanges.** *J Math Econ* 1974, **1**:39–49.
- Zeeman, E: *Catastrophe Theory*: Reading: Addison-Wesley Pub. Co; 1980.

doi:10.1186/2194-3206-1-9

Cite this article as: Bossomaier et al.: Information and phase transitions in socio-economic systems. *Complex Adaptive Systems Modeling* 2013 **1**:9.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
