Chaos, Complexity, and Inference (36-462)

Lecture 12: Quantifying Self-Organization

Cosma Shalizi

19 February 2009

Further reading: Shalizi *et al.* (2004, 2006)

More and Better Notions of Complexity

Kolmogorov complexity randomness
but pure noise is easy to describe ("toss a coin")
also, uncomputable

Sophistication Split minimal algorithm for *x* into program and data, s.t. *x* is a typical output for the program (Gács *et al.*, 2001)
"the ultimate model of a cat is of course another cat" (Rosenblueth *et al.*, 1943)
still uncomputable

Logical depth How long does the minimal algorithm take to run? (Bennett, 1986, 1990)
still uncomputable

Thermodynamic entropy *S*, "improbability" randomness issue again
also, it's *very probable* that the BZ reaction makes spiral waves

$\frac{S}{S_{max}}(1 - \frac{S}{S_{max}})$ a “so what?” quantity (Feldman and Crutchfield, 1998; Crutchfield *et al.*, 2000)

Probabilistic description Take two-part idea but use probabilistic descriptions (Rissanen, 1989; Grünwald, 2005)
breaks DL into (program/model part) + (noisy data details part)
we care about first part
What exactly needs describing and how?

Three Kinds of Complexity of Prediction

1. Induction/learning
2. Description/estimation
3. Computational/calculating

These are distinct!

Inductive complexity

How hard is to *learn* the right predictive model
units of samples or data-points
Depends on our choice of representation, not just the system
Huge & wonderful theory: take 36-702, or read Kearns and Vazirani (1994); Vapnik (2000)
Tangential to us

Descriptive complexity

How hard is to to *estimate* the the right state of the right model?
units of bits
How much information about the past is *necessary* for optimal prediction of the future? (Grassberger, 1986; Crutchfield and Young, 1989)
Does not depend on our models
Does depend on level of description; can't be helped

Computational complexity

How hard is to *calculate* the prediction from the right state in the right model?

units of time-steps

Many fascinating results

Machta (2006); Griffeath and Moore (1996); Lindgren *et al.* (1998); Moore (1997); Moore and Nordahl (1997); Machta and Machta (2005); Machta and Greenlaw (1994); Machta and Li (2001); Moore and Machta (2000); Moore and Nilsson (1999); Tillberg and Machta (2004)

Won't go into this further

"Statistics"

Given: data, say past behavior $X^t_{-\infty}$
Wanted: guess unobserved, say future X^∞_{t+1}
A statistic: a calculable function of the data we use for our inference
$R_t = \rho(X^t_{-\infty})$
mean, variance, moving average over last three steps, Fourier amplitude, maximum likelihood estimate, ...

Information in Statistics; Sufficiency

How much does ρ tell us about the future? **Predictive information**

$$I[X_{t+1}^{\infty}; R_t] = H[X_{t+1}^{\infty}] - H[X_{t+1}^{\infty}|R_t]$$

Basic observation:

$$I[X_{t+1}^{\infty}; R_t] \leq I[X_{t+1}^{\infty}; X_{-\infty}^{t}]$$

R is **predictively sufficient** when $I[X_{t+1}^{\infty}; R_t] = I[X_{t+1}^{\infty}; X_{-\infty}^{t}]$
Fact: optimal prediction only needs a sufficient statistic
no matter how you measure "optimal"

Minimal Sufficiency

$S_t = \sigma(\rho(X^t_{-\infty})) = \sigma(R_t)$
S_t is a statistic but coarser, retains less, of data

$$I[X^{\infty}_{t+1}; S_t] \leq I[X^{\infty}_{t+1}; R_t] \leq I[X^{\infty}_{t+1}; X^t_{-\infty}]$$

but also

$$I[S_t; X^t_{-\infty}] \leq I[R_t; X^t_{-\infty}]$$

information about past needed to make prediction
Minimal sufficient: S_t is sufficient but is a function of every other sufficient statistic R_t
Minimal sufficient statistics give us the forecasting complexity

Another way to say this: Necessary statistics

A statistic is **necessary** if it can be calculated from *all* sufficient statistics

If you are going to do optimal prediction, you *have* to know every necesssary statistic

Maximum likelihood estimates are generally necessary statistics (but not always sufficient)

"Minimal sufficient" = "necessary and sufficient"

Building the minimal sufficient statistic

Particular pasts $x, y, \ldots$
Equivalent: $x \sim y$ if and only if

$$\Pr\left(X_{t+1}^{\infty} | X_{-\infty}^{t} = x\right) = \Pr\left(X_{t+1}^{\infty} | X_{-\infty}^{t} = y\right)$$

Equivalence class: $[x] =$ all y such that $x \sim y$
Statistic:

$$\epsilon(x) = [x]$$

Maps particular histories to their distribution over future events, or to the set of histories with the same prediction
Value of statistic is **predictive state** or **causal state** S_t
Crutchfield and Young (1989); Shalizi and Crutchfield (2001)

Complexity

Grassberger/Crutchfield-Young statistical forecasting complexity:

$$C \equiv I[S_t; X^t_{-\infty}]$$

$= H[S_t]$ for discrete predictive states
How many (statistically) different things can the process do?

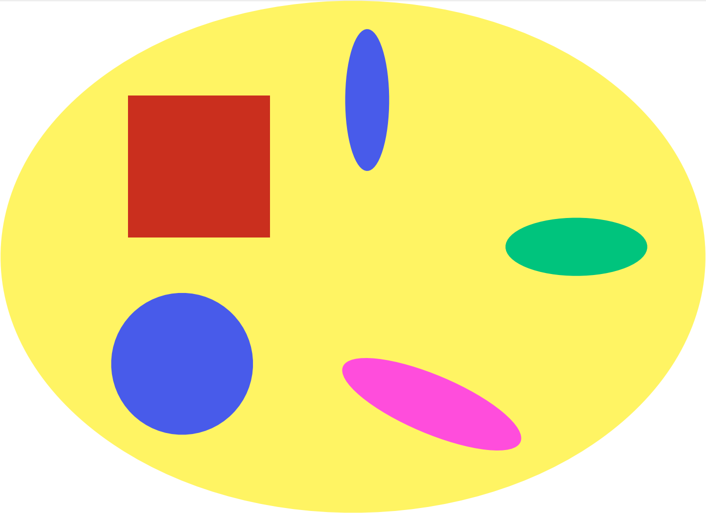
Some basic cases:

IID: one conditional distribution = unconditional distribution $\Rightarrow$ 1 state, $C = 0$

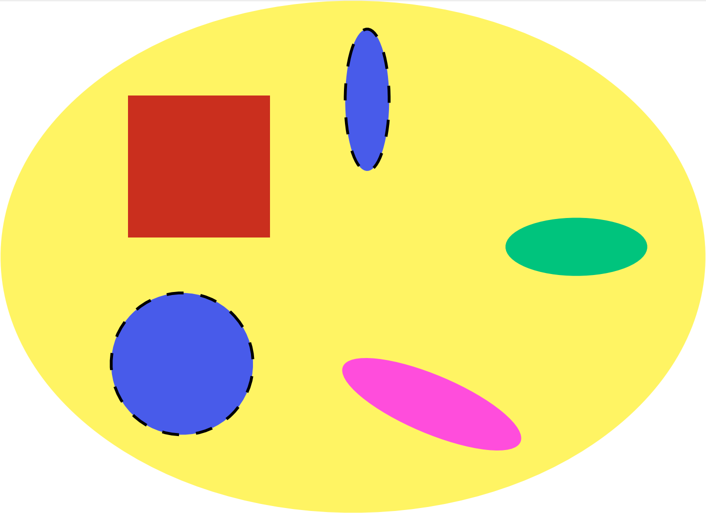
Periodic: p conditional distributions, 1 per phase $\Rightarrow C = \log_2 p$

Markov chains: states are their own predictive states

unless space is redundant



histories, colored by predictive distributions



partitioning into predictive states

Properties: Sufficiency, Markov

SUFFICIENCY:

$$\begin{aligned} \Pr\left(X_{t+1}^{\infty} | S_t = \epsilon(x_{-\infty}^{t})\right) &= \Pr\left(X_{t+1}^{\infty} | X_{-\infty}^{t} = x_{-\infty}^{t}\right) \\ I[X_{t+1}^{\infty}; S_t] &= I[X_{t+1}^{\infty}; X_{-\infty}^{t}] \end{aligned}$$

RECURSIVE UPDATING: for some function T,

$$\begin{aligned} \epsilon(x_{-\infty}^{t+1}) &= T(\epsilon(x_{-\infty}^{t}), x_{t+1}) \\ S_{t+1} &= T(S_t, X_{t+1}) \end{aligned}$$

MARKOV:

$$S_{t+1}^{\infty} \perp S_{-\infty}^{t-1} | S_t$$

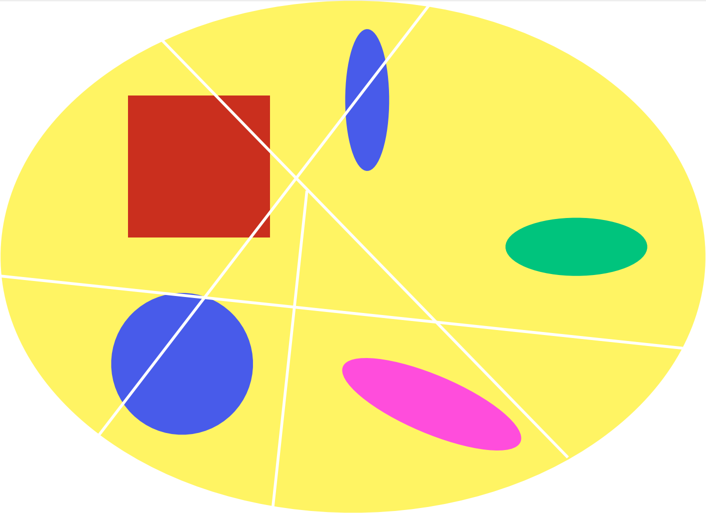
Predictive states give us a stochastic automaton

Properties: Minimality

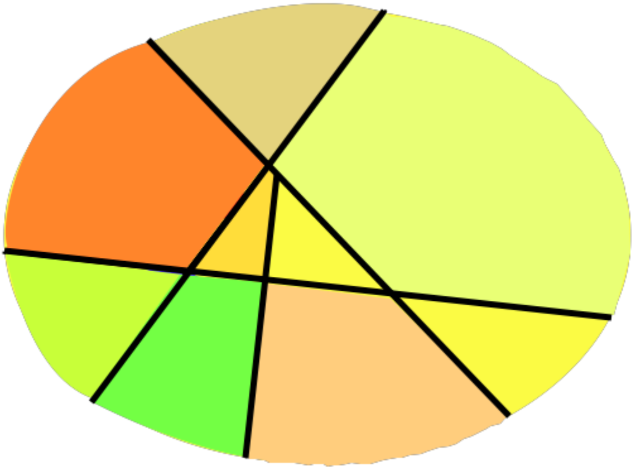
For any other sufficient statistic $R_t = \rho(X^t_{-\infty})$,

$$I[R_t; X^t_{-\infty}] \geq I[S_t; X^t_{-\infty}]$$

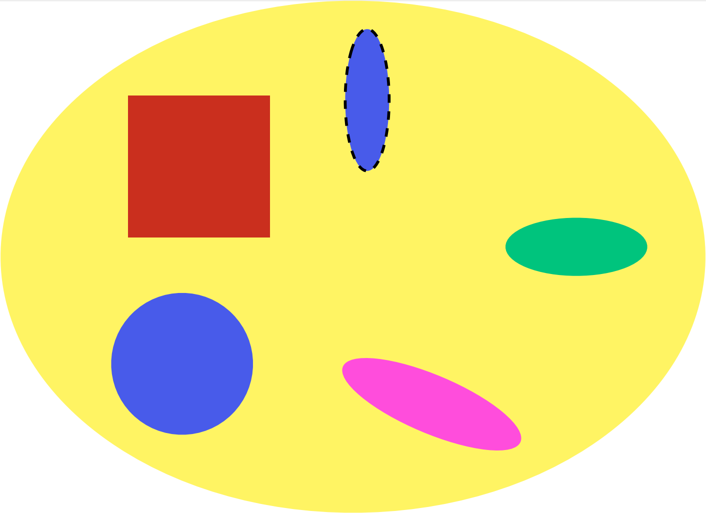
If $I[R_t; X^t_{-\infty}] < I[S_t; X^t_{-\infty}]$ then R_t is not sufficient



A partition which is not sufficient (cuts across predictive states)



Effect of insufficiency on predictions



A sufficient, but not minimal, partition (2 blue states)

Properties: Uniqueness

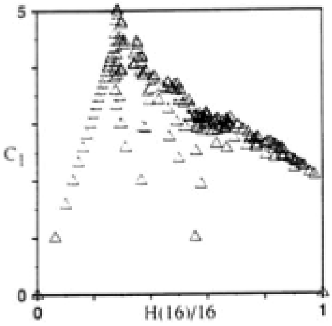
If R_t is sufficient and $I[R_t; X^t_{-\infty}] = I[S_t; X^t_{-\infty}]$, then or some function f,

$$R_t = f(S_t),\ S_t = f^{-1}(R_t)$$

Upshot

(GCY) Statistical complexity is well-defined and depends only on the process

Complexity is not entropy



Logistic map (193 parameter values between 0.75 and 1), from Crutchfield and Young (1989)

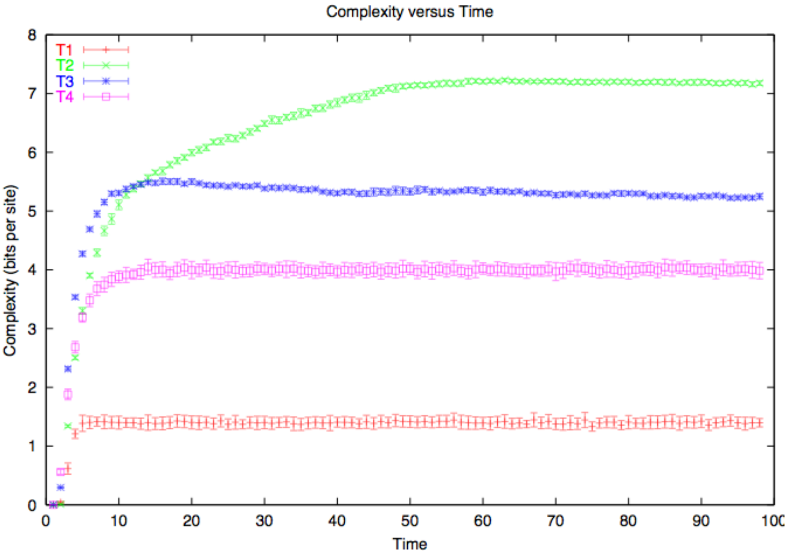
Vertical axis is complexity, horizontal axis is entropy rate

Peak occurs at period-doubling accumulation point

Going to higher dimensions

One of them has to be time (but see Young *et al.* (2005))
Idea: repeat the same construction, only use the local past around a point to predict the local future around that point (Shalizi, 2003)

Results for cyclic CA



Defining Self-Organization

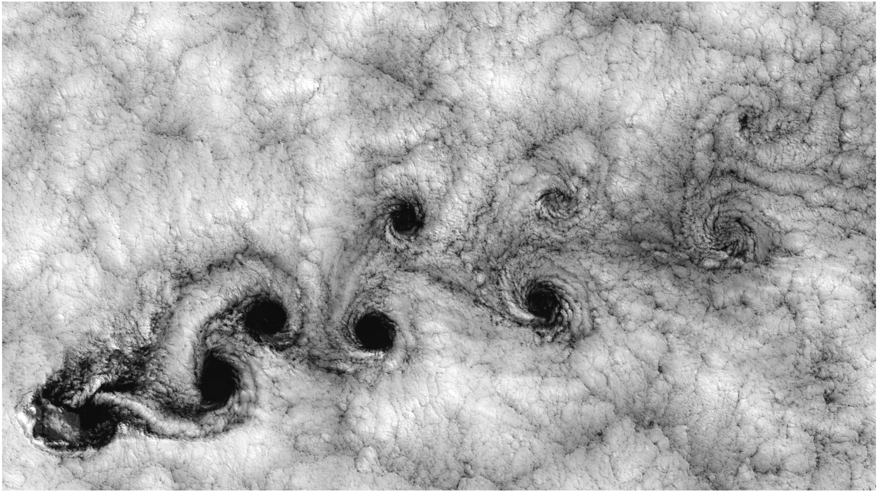
(Shalizi *et al.*, 2004)
An assemblage has self-organized between t_1 and t_2 if

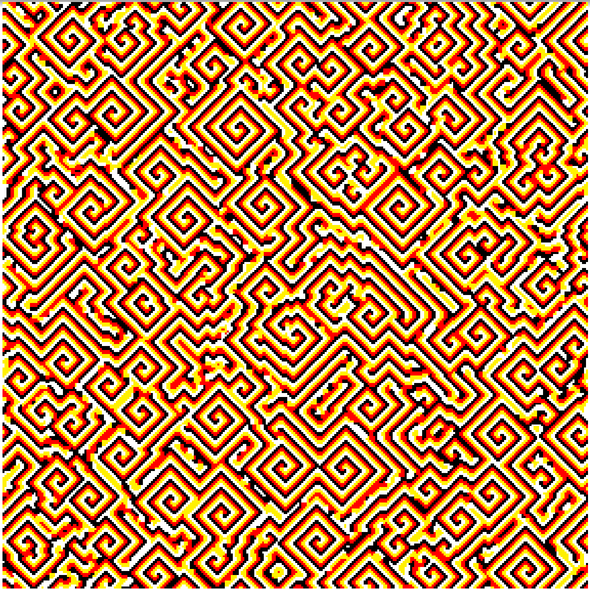
1. $C(t_1) > C(t_2)$
2. Not all of the increase is due to an external organizer

In CA we don't need to worry about item (2); in general need to "exorcise demons"
One approach: replace actual outside driver by noise with a similar distribution, see what difference it makes (Delgado and Solé, 1997)

Coherent Structures

Examples from last time: spirals, targets, ...





Order Parameters

Structure means: Symmetry is broken $\equiv$ picture does not look the same in everyt direction

But also: Symmetry is only *partially* broken

Identify a function which measures this symmetry breaking = **order parameter**

Guess how order parameter contributes to energy, check that it gives right distribution

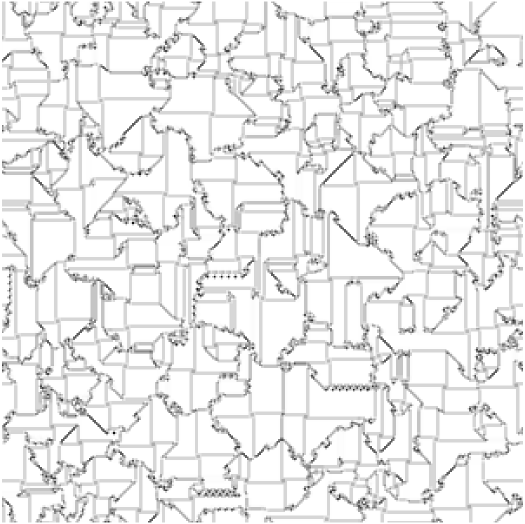
Find singularities in order parameter field = structures

Sethna (1991) is great on this, can be read with minimal knowledge of physics

Sethna (2006) is not quite so easy to read

Finding Order Parameters Is Hard

Usual approach: trial and error, tradition, analogy

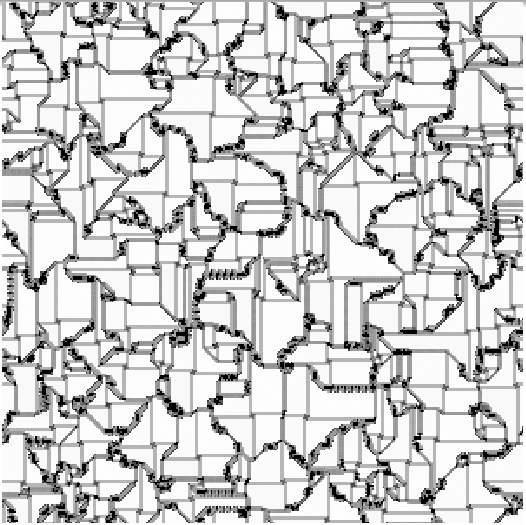


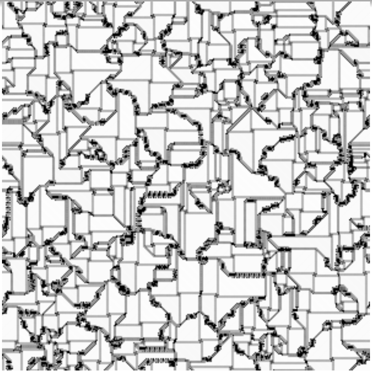
Local Statistical Complexity

Define

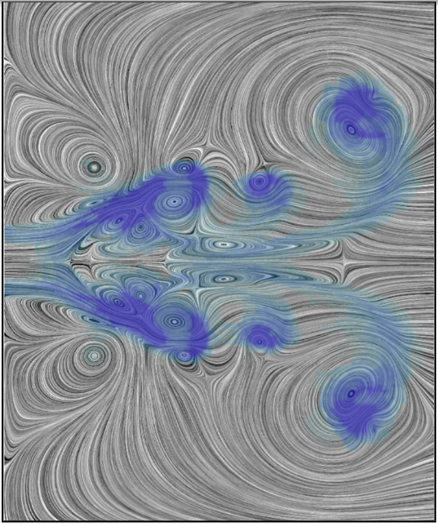
$$c(r,t) = -\log_2 \Pr\left(S(r,t) = \epsilon(x(r,t))\right)$$

shows *where* complexity is localized, as opposed to just how much





Hand-found order parameter (much insight needed) *vs.* automatically-calculated local complexity (no insight needed)



Numerical simulation of fluid flow past a cylinder (flow-lines in grey) + local complexity (blue) Jänicke *et al.* (2007)

Bennett, Charles H. (1986). "On the Nature and Origin of Complexity in Discrete, Homogeneous Locally-Interacting Systems." *Foundations of Physics*, **16**: 585–592.

— (1990). "How to Define Complexity in Physics, and Why." In *Complexity, Entropy, and the Physics of Information* (Wojciech H. Zurek, ed.), pp. 137–148. Reading, Massachusetts: Addison-Wesley.

Crutchfield, James P., David P. Feldman and Cosma Rohilla Shalizi (2000). "Comment I on 'Simple Measure for Complexity'." *Physical Review E*, **62**: 2996–2997. URL `http://arxiv.org/abs/chao-dyn/9907001`.

Crutchfield, James P. and Karl Young (1989). "Inferring Statistical Complexity." *Physical Review Letters*, **63**: 105–108. URL `http://www.santafe.edu/~cmg/compmech/pubs/ISCTitlePage.htm`.

Delgado, Jordi and Ricard V. Solé (1997). "Collective-induced Computation." *Physical Review E*, **55**: 2338–2344. URL http://www.lsi.upc.es/~jdelgado/PREcollective.pdf. doi:10.1103/PhysRevE.55.2338.

Feldman, David P. and James P. Crutchfield (1998). "Measures of Statistical Complexity: Why?" *Physics Letters A*, **238**: 244–252. URL http://hornacek.coa.edu/dave/Publications/MSCW.html.

Gács, Péter, John T. Tromp and Paul M. B. Vitanyi (2001). "Algorithmic Statistics." *IEEE Transactions on Information Theory*, **47**: 2443–2463. URL http://arxiv.org/abs/math.PR/0006233.

Grassberger, Peter (1986). "Toward a Quantitative Theory of Self-Generated Complexity." *International Journal of Theoretical Physics*, **25**: 907–938.

Griffeath, David and Cristopher Moore (1996). "Life Without Death is P-Complete." *Complex Systems*, **10**: 437–447.

Grünwald, Peter (2005). "A Tutorial Introduction to the Minimum Description Length Principle." In *Advances in Minimum*

Description Length: Theory and Applications (P. Grünwald and I. J. Myung and M. Pitt, eds.). Cambridge, Massachusetts: MIT Press. URL http://arxiv.org/abs/math.ST/0406077.

Jänicke, Heike, Alexander Wiebel, Gerik Scheuermann and Wolfgang Kollmann (2007). "Multifield Visualization Using Local Statistical Complexity." *IEEE Transactions on Visualization and Computer Graphics*, **13**: 1384–1391. URL http://www.informatik.uni-leipzig.de/bsv/Jaenicke/Papers/vis07.pdf. doi:10.1109/TVCG.2007.70615.

Kearns, Michael J. and Umesh V. Vazirani (1994). *An Introduction to Computational Learning Theory*. Cambridge, Massachusetts: MIT Press.

Lindgren, Kristian, Cristopher Moore and Mats Nordahl (1998). "Complexity of Two-Dimensional Patterns." *Journal of Statistical Physics*, **91**: 909–951. URL http://arxiv.org/abs/cond-mat/9804071.

Machta, Benjamin and Jonathan Machta (2005). "Parallel dynamics and computational complexity of network growth models." *Physical Review E*, **71**: 026704. URL http://arxiv.org/abs/cond-mat/0408372.

Machta, Jonathan (2006). "Complexity, parallel computation and statistical physics." *Complexity*, **11(5)**: 46–64. URL http://arxiv.org/abs/cond-mat/0510809.

Machta, Jonathan and Raymond Greenlaw (1994). "The Parallel Complexity of Growth Models." *Journal of Statistical Physics*, **77**: 755–781. URL http://arxiv.org/abs/cond-mat/9403006.

Machta, Jonathan and Xuenan Li (2001). "Parallel dynamics and computational complexity of the Bak-Sneppen model." *Physica A*, **300**: 245–270. URL http://arxiv.org/abs/cond-mat/0105296.

Moore, Cristopher (1997). "Majority-Vote Cellular Automata, Ising Dynamics, and P-Completeness." *Journal of Statistical*

Physics, **88**: 795–805. URL
http://arxiv.org/abs/cond-mat/9701118.

Moore, Cristopher and Jonathan Machta (2000). "Internal Diffusion-Limited Aggregation: Parallel Algorithms and Complexity." *Journal of Statistical Physics*, **99**: 661–690. URL http://arxiv.org/abs/cond-mat/9909233.

Moore, Cristopher and Martin Nilsson (1999). "The Computational Complexity of Sandpiles." *Journal of Statistical Physics*, **96**: 205–224. URL http://arxiv.org/abs/cond-mat/9808183.

Moore, Cristopher and Mats G. Nordahl (1997). "Lattice Gas Prediction is P-Complete." Electronic preprint. URL http://arxiv.org/abs/comp-gas/9704001.

Rissanen, Jorma (1989). *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific.

Rosenblueth, Arturo, Norbert Wiener and Julian Bigelow (1943). "Behavior, Purpose and Teleology." *Philosophy of*

Science, **10**: 18–24. URL http://www.jstor.org/pss/184878.

Sethna, James P. (1991). “Order Parameters, Broken Symmetry, and Topology.” In *1990 Lectures in Complex Systems* (Lynn Nadel and Daniel L. Stein, eds.), vol. 3 of *Santa Fe Institute Studies in the Sciences of Complexity (Lectures)*, pp. 243–265. Redwood City, California: Addison-Wesley.

— (2006). *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. Oxford: Oxford University Press.

Shalizi, Cosma Rohilla (2003). “Optimal Nonlinear Prediction of Random Fields on Networks.” *Discrete Mathematics and Theoretical Computer Science*, **AB(DMCS)**: 11–30. URL http://arxiv.org/abs/math.PR/0305160.

Shalizi, Cosma Rohilla and James P. Crutchfield (2001). “Computational Mechanics: Pattern and Prediction, Structure and Simplicity.” *Journal of Statistical Physics*, **104**: 817–879. URL http://arxiv.org/abs/cond-mat/9907176.

Shalizi, Cosma Rohilla, Robert Haslinger, Jean-Baptiste Rouquier, Kristina Lisa Klinkner and Cristopher Moore (2006). "Automatic Filters for the Detection of Coherent Structure in Spatiotemporal Systems." *Physical Review E*, **73**: 036104. URL http://arxiv.org/abs/nlin.CG/0508001.

Shalizi, Cosma Rohilla, Kristina Lisa Klinkner and Robert Haslinger (2004). "Quantifying Self-Organization with Optimal Predictors." *Physical Review Letters*, **93**: 118701. URL http://arxiv.org/abs/nlin.AO/0409024.

Tillberg, Dan and Jonathan Machta (2004). "Structural and computational depth of diffusion limited aggregation." *Physical Review E*, **69**: 051403. URL http://arxiv.org/abs/cond-mat/0309519.

Vapnik, Vladimir N. (2000). *The Nature of Statistical Learning Theory*. Berlin: Springer-Verlag, 2nd edn.

Young, Karl, Yue Chen, John Kornak, Gerald B. Matson and

Norbert Schuff (2005). “Summarizing Complexity in High Dimensions.” *Physical Review Letters*, **94**: 098701.