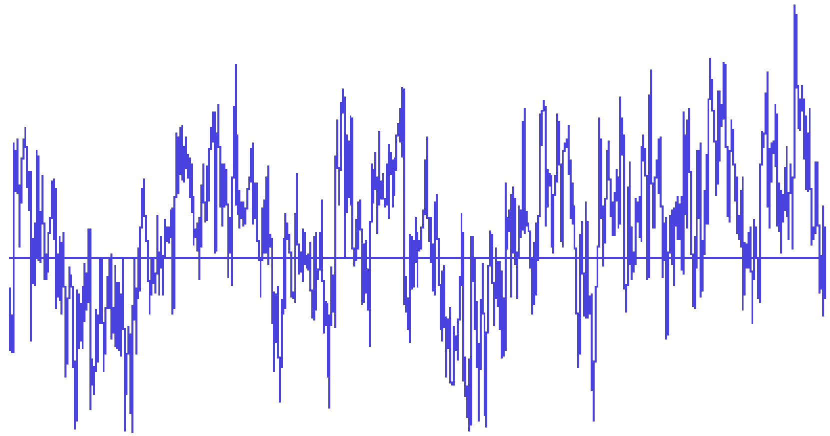
Systems and Signals

Department of Mathematics
Imperial College London

Overview

- **Cellular fluctuations and cellular energy variability**: Single cell signal processing, mitochondrial variability.
- **Principles of Natural networks**: Transport and optimization on, processing by and morphings along networks.
- **Structure of Natural data**: Highly comparative data analysis of time series, landscapes and networks. Inferring network structure from time series data, controlling network processes.

Highly Comparative Analysis



Ben Fulcher

Sumeet Agarwal

Jamie King

- What is the empirical structure of our signals and our methods?

What is Complex Systems for?

- *Perhaps* to detect patterns in disordered systems around us and…
- …to identify broad classes of mechanisms which can generate this behaviour.

- *So how's it doing?*

What is going on in Complex Systems?

- Extremely difficult to get an overview because:
  - Lots of models
  - Lots of data in…
  - Lots of journals in…
  - Lots of disciplines.

Troubles with data:

- Methods are partly socially selected
- …and sometimes inadequately compared
- Huge numbers of existing methods: innovating may even be bad
- Models are perhaps uncharted and hard to compare
- Data can be overfit and not compared
- Lots of disciplines reinvent and don’t communicate
- Complex systems is particularly vulnerable to the above

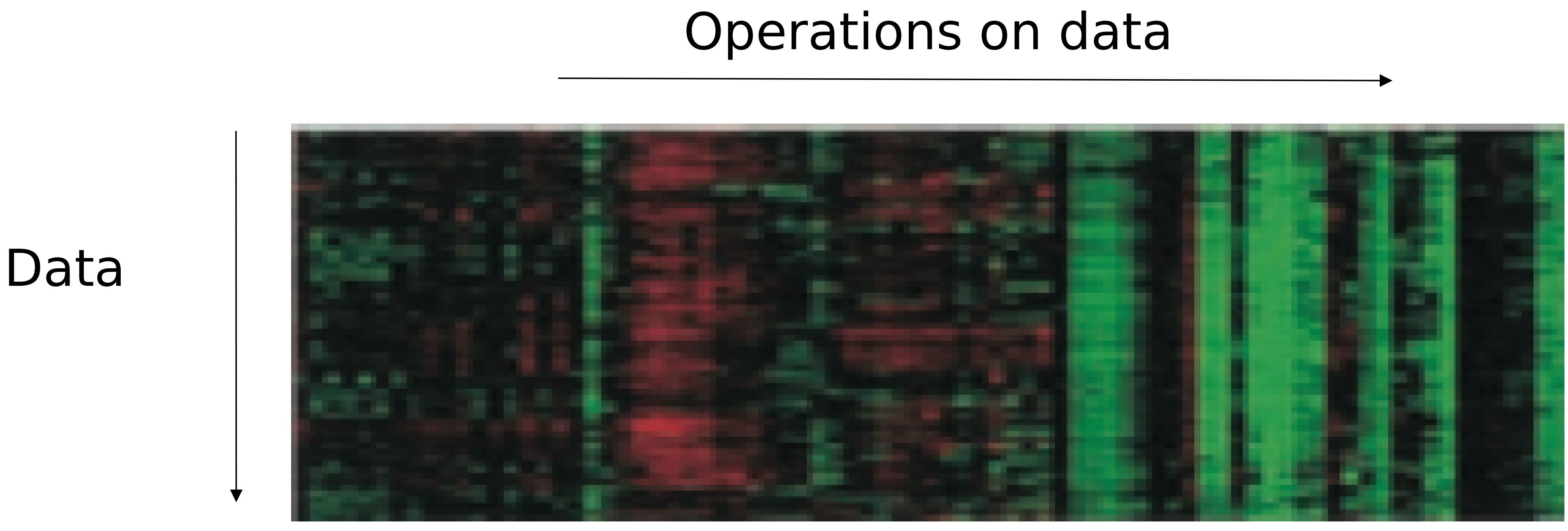
Is more theory needed?

- Yes. Theoretical approaches might supply a unification.

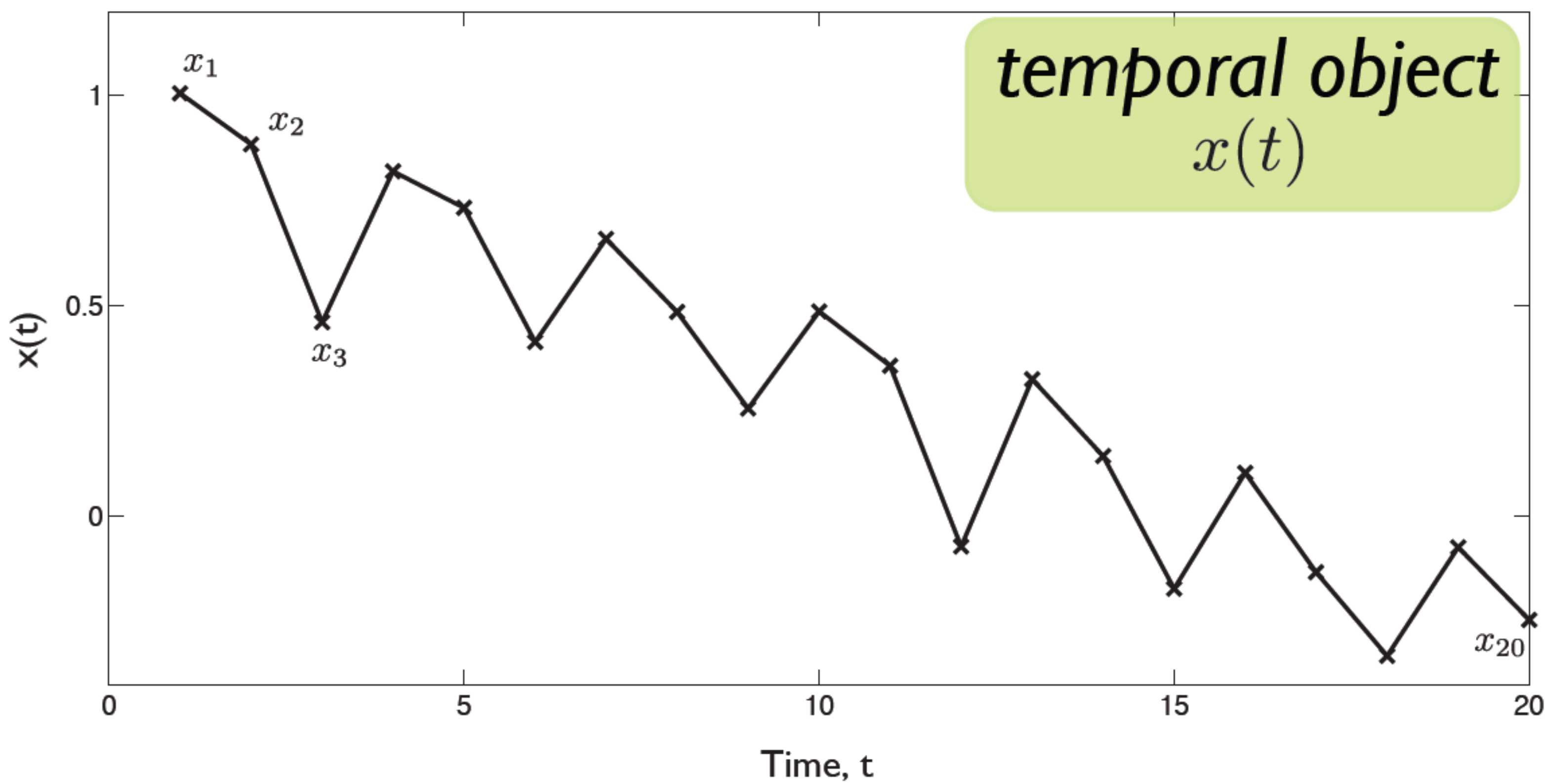
- But simplistically (or pragmatically) the number of possible relationships between methods, models and data is so rich (or contingent) that it is also worth considering crude alternatives.

A crude approach

- Allows fairer comparison of existing methods, models and data => (super)-families of data and methods.
  - Needed in complex systems
- Engages with the problem of collective overfitting.
- Helps creative development of new methods.
- Given a new piece of data, volunteers other similar data and models which could have generated it.
- Allows both improved, interpretable, and rapid supervised and unsupervised learning.



- (A bit like a DNA microarray)

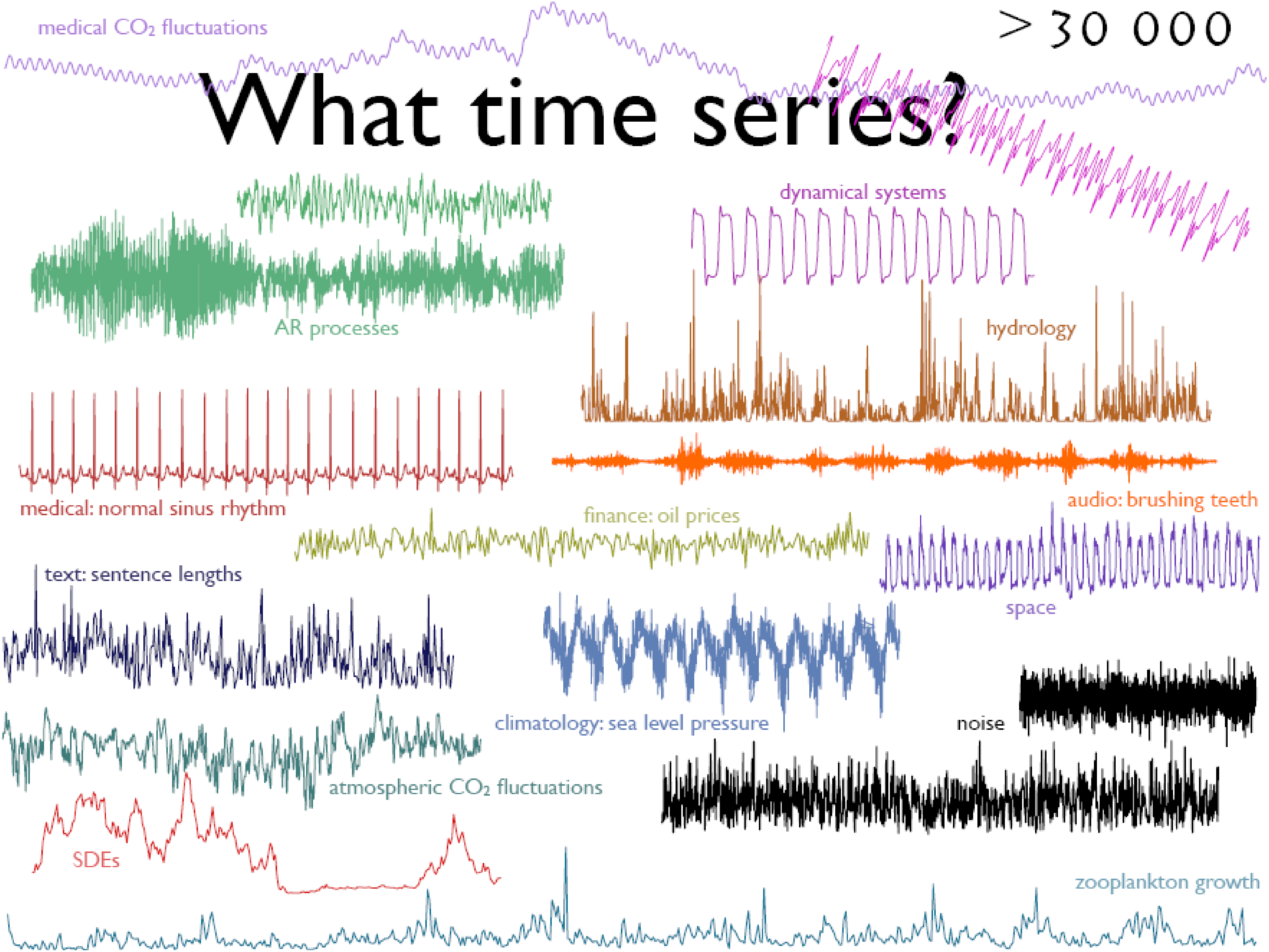


$$\mathbf{f} = (f_1, f_2, ..., f_N)$$

(unordered) set of informative features

What time series?

> 30 000



> 9 000

What operations?

Basic statistics

trimmed means
zero crossings
standard deviation
outliers
local extrema

Stationarity

StatAv
sliding windows
bootstraps
distribution comparisons

Static distribution

quantiles
moments
fits to standard distributions
hypothesis tests

Basis Functions

wavelet transform
power spectrum peaks
spectral measures
low frequency power

Correlation

linear autocorrelations
decay properties
automutual information
dependence on additive noise
nonlinear autocorrelations
time reversal asymmetry
generalized self-correlation function
recurrence structure
autocorrelation robustness
fluctuation analysis: scaling
randomization robustness
recurrence plots
seasonality testing

Model fits

primitive forecasting
Fourier fits
GARCH modeling
step-ahead dependence
exponential smoothing
AR models
state space models
hidden Markov models
piecewise splines
'walker' statistics
ARMA modeling
Gaussian Processes

Nonlinear

2D embedding structure
TSTOOL
TISEAN
fractal dimension
correlation dimension
Taken's estimator
Poincaré sections
surrogate methods
nonlinear prediction error
Lyapunov exponent
false nearest neighbours

Others

course-grained transition matrices
motif distributions
couple to dynamical systems
visibility graph
stick angle distribution
step detection algorithms
extreme events
drifting mean tests
PCA of embedded signal
domain-specific standard metrics

Information Theory

SampEn
distributional entropies
conditional entropies
binned entropies
kernel smoothed entropies
Tsallis entropies
ApEn

long-range scaling

power spectral density

linear models

stationarity

variance

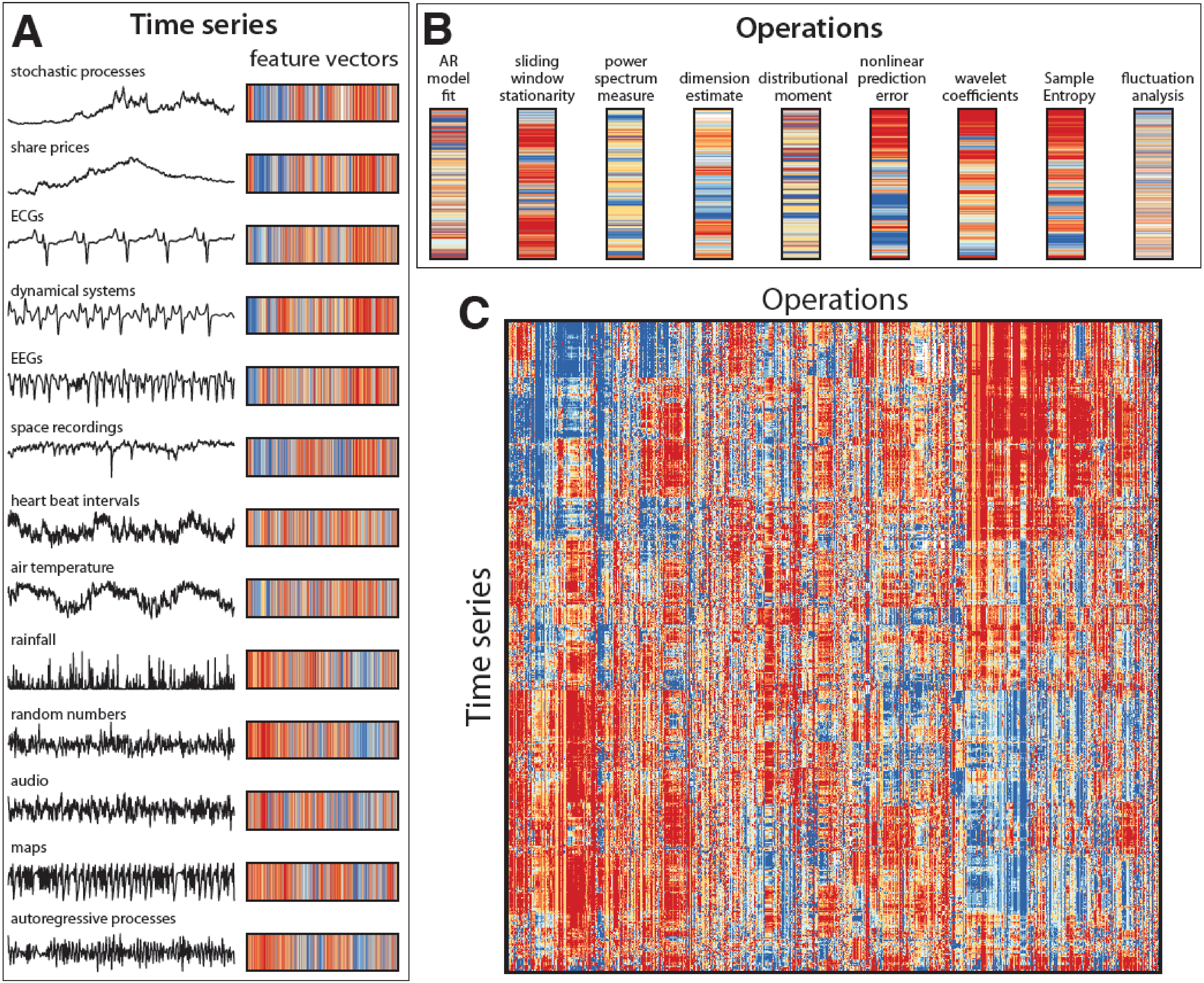
entropy

correlation dimension

complexity

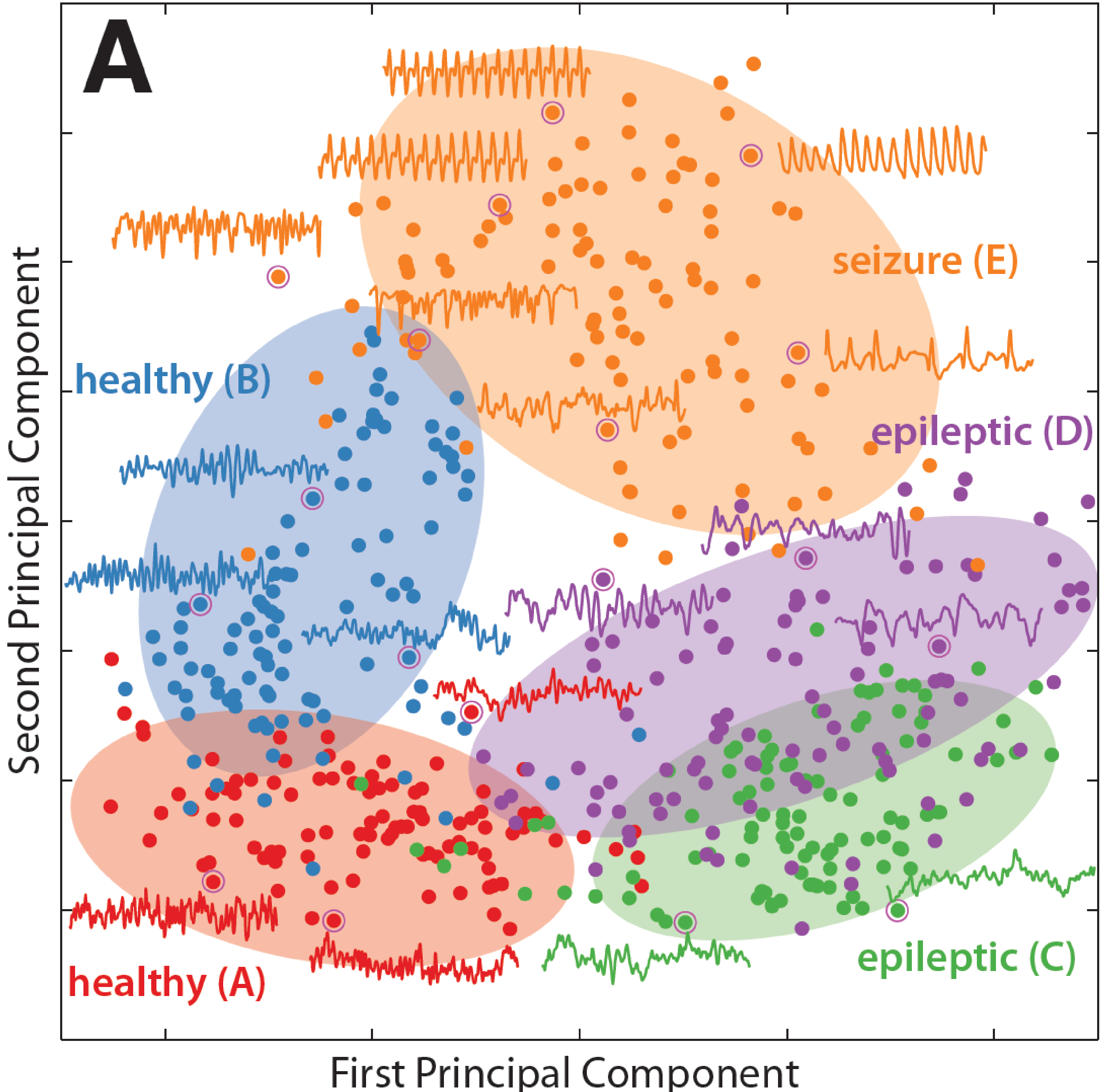
information theory

BIG PICTURE

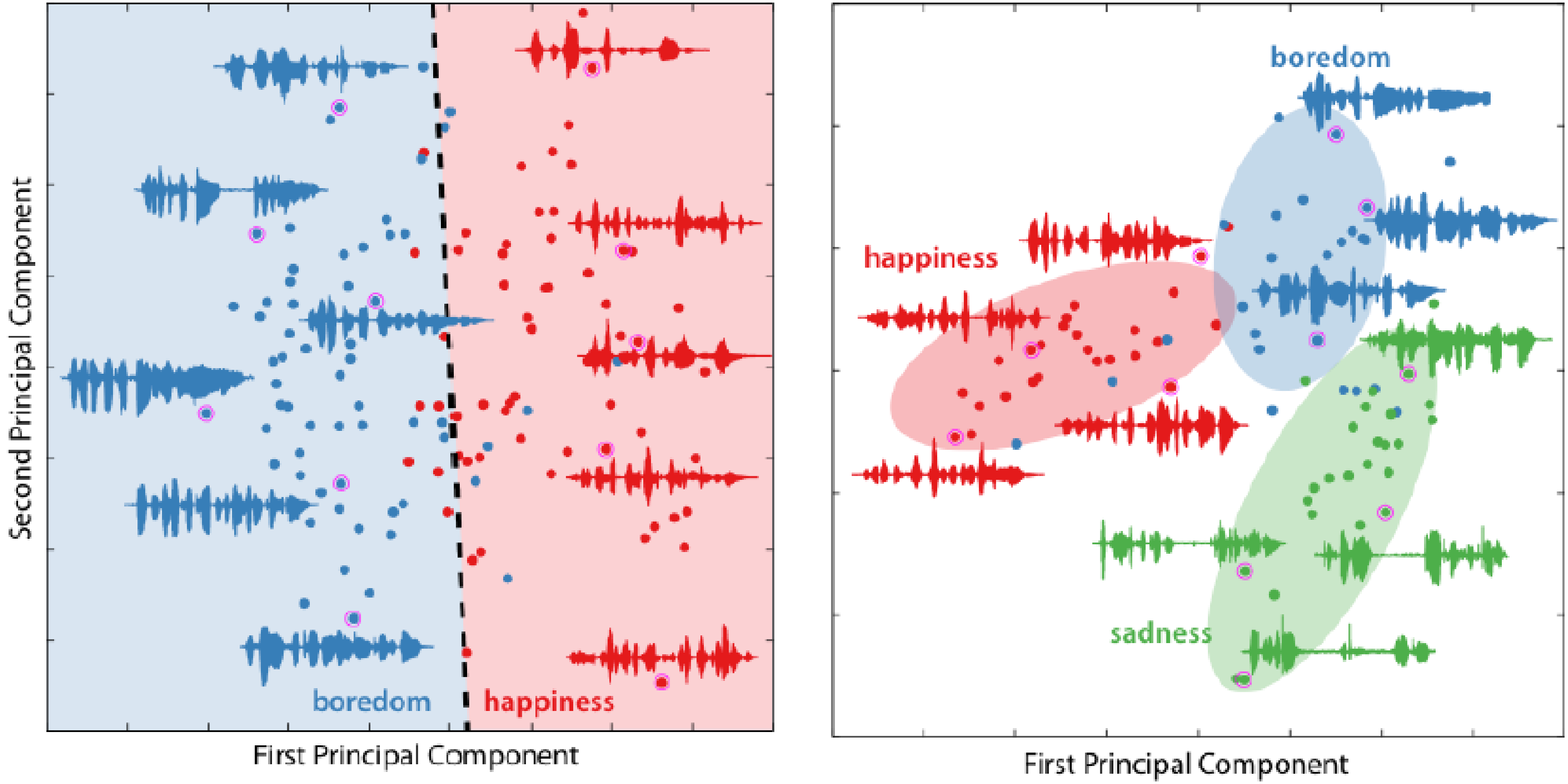

A
Time series
feature vectors
stochastic processes
share prices
ECGs
dynamical systems
EEGs
space recordings
heart beat intervals
air temperature
rainfall
random numbers
audio
maps
autoregressive processes
B
Operations
AR model fit
sliding window stationarity
power spectrum measure
dimension estimate
distributional moment
nonlinear prediction error
wavelet coefficients
Sample Entropy
fluctuation analysis
C
Operations
Time series

Doing (maybe) useful things

- Helps with identifying:
  - Epileptic’s EEG’s
  - Congestive heart failure
  - Parkinsonian speech
  - Emotional speech
  - Foetal Heart Rhythm

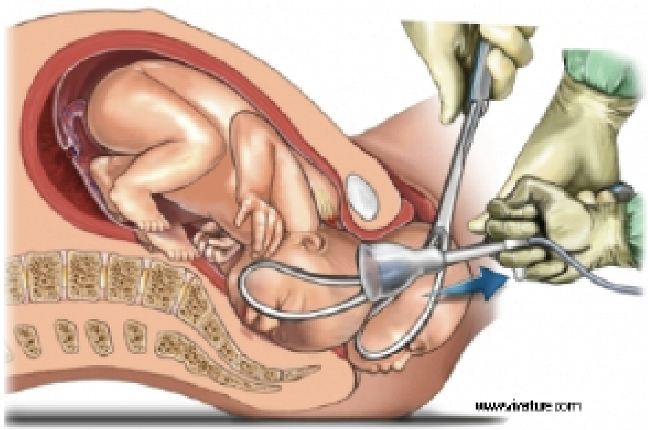

A
seizure (E)
healthy (B)
epileptic (D)
epileptic (C)
healthy (A)
Second Principal Component
First Principal Component

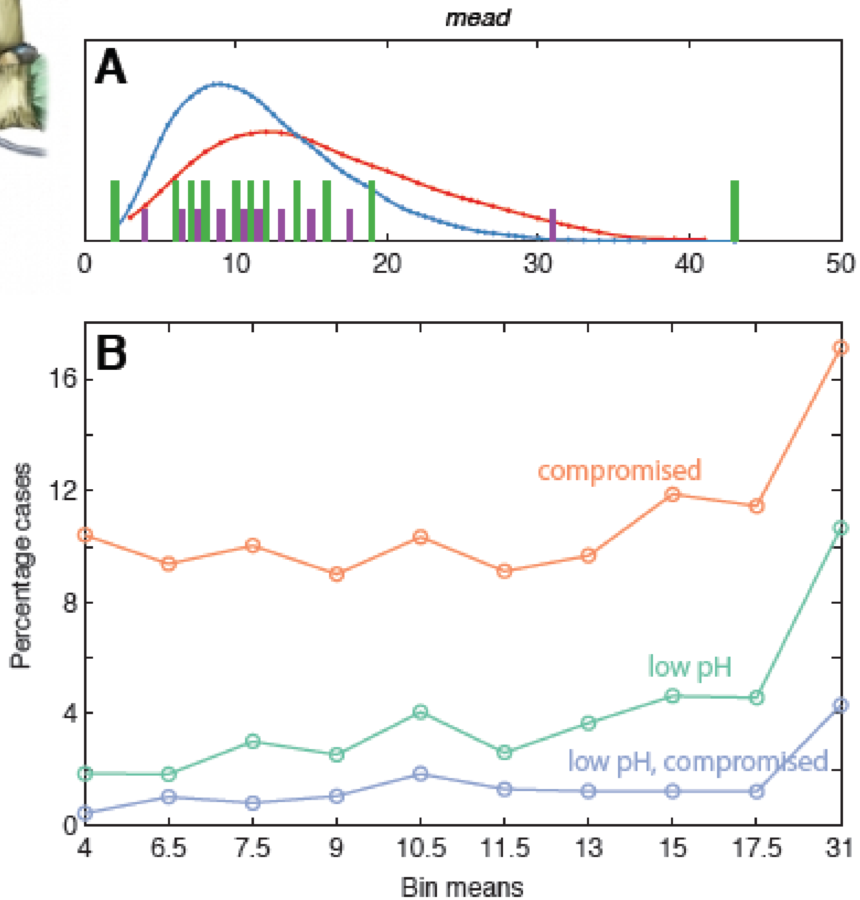
Emotional Speech



Fetal heart rate

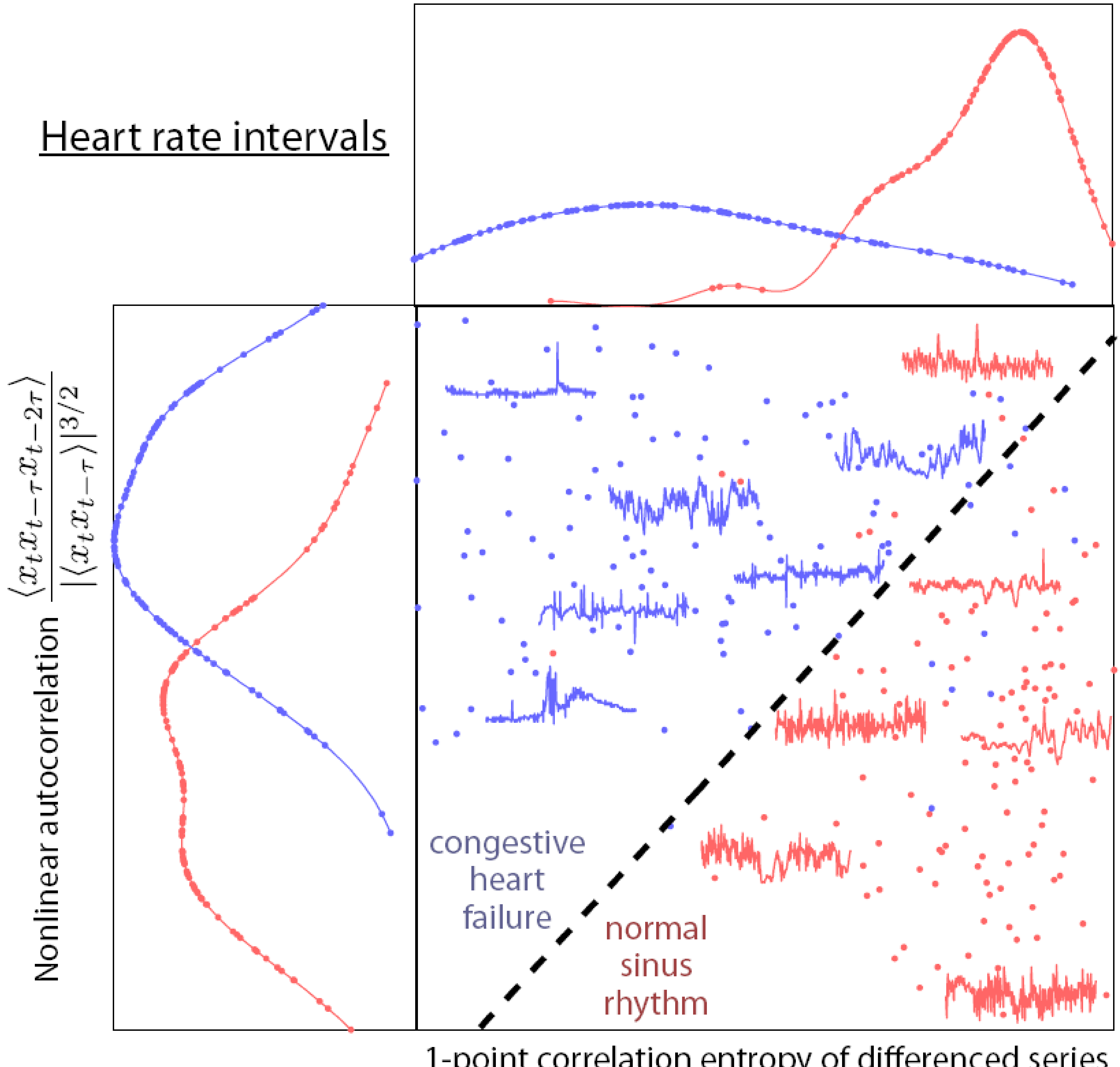
- Diagnose low umbilical pH





Arterial and venous blood samples

http://www.k2ms.com/


Heart rate intervals
Nonlinear autocorrelation $\frac{\langle x_t x_{t-\tau} x_{t-2\tau} \rangle}{|\langle x_t x_{t-\tau} \rangle|^{3/2}}$
congestive heart failure
normal sinus rhythm
1-point correlation entropy of differenced series

Summary I

- Helps us to get some first ideas about:
  - new data
  - new models (ones that can fit and ones that can simulate)
  - new methods
- Helps us to be critical about data/models/methods and innovate with some confidence in both models and methods.
- Might help with some of the goals of complex systems.
- Is useful for identifying the structure of sets of data and for identifying critical distinguishing features between them.

Scenario: A Crisis Has Occurred

???

Key Questions

- How long does it take for information to spread through the population?

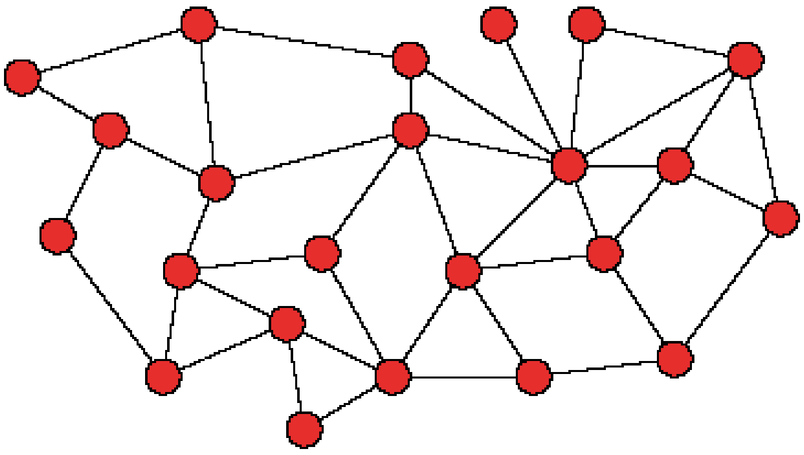
- In what ways do people become informed?

Once Informed, Always Informed

- Each person begins in the “uninformed” state
- Once alerted to news of the crisis, a person quickly verifies the news, entering the “informed” state and staying there forever.
- In other words, these are SI models.

How does information spread?

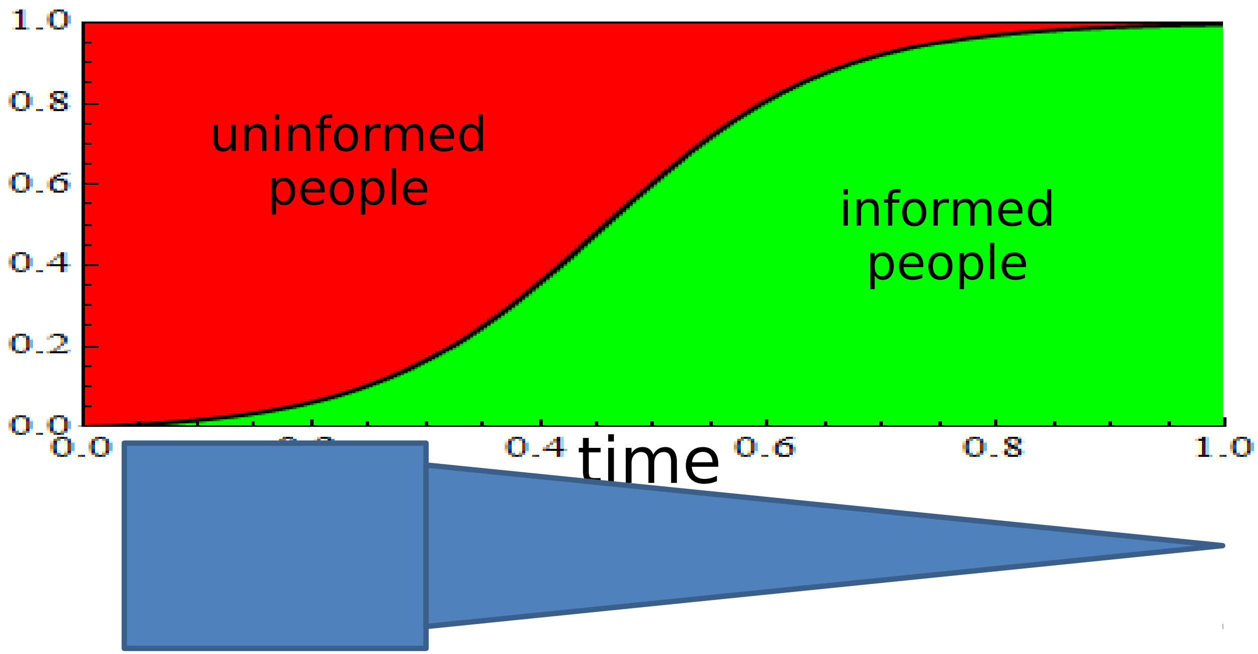
- Broadcast
  - Television, radio, etc.
- Sharing
  - Phone calls, speaking in person, etc.



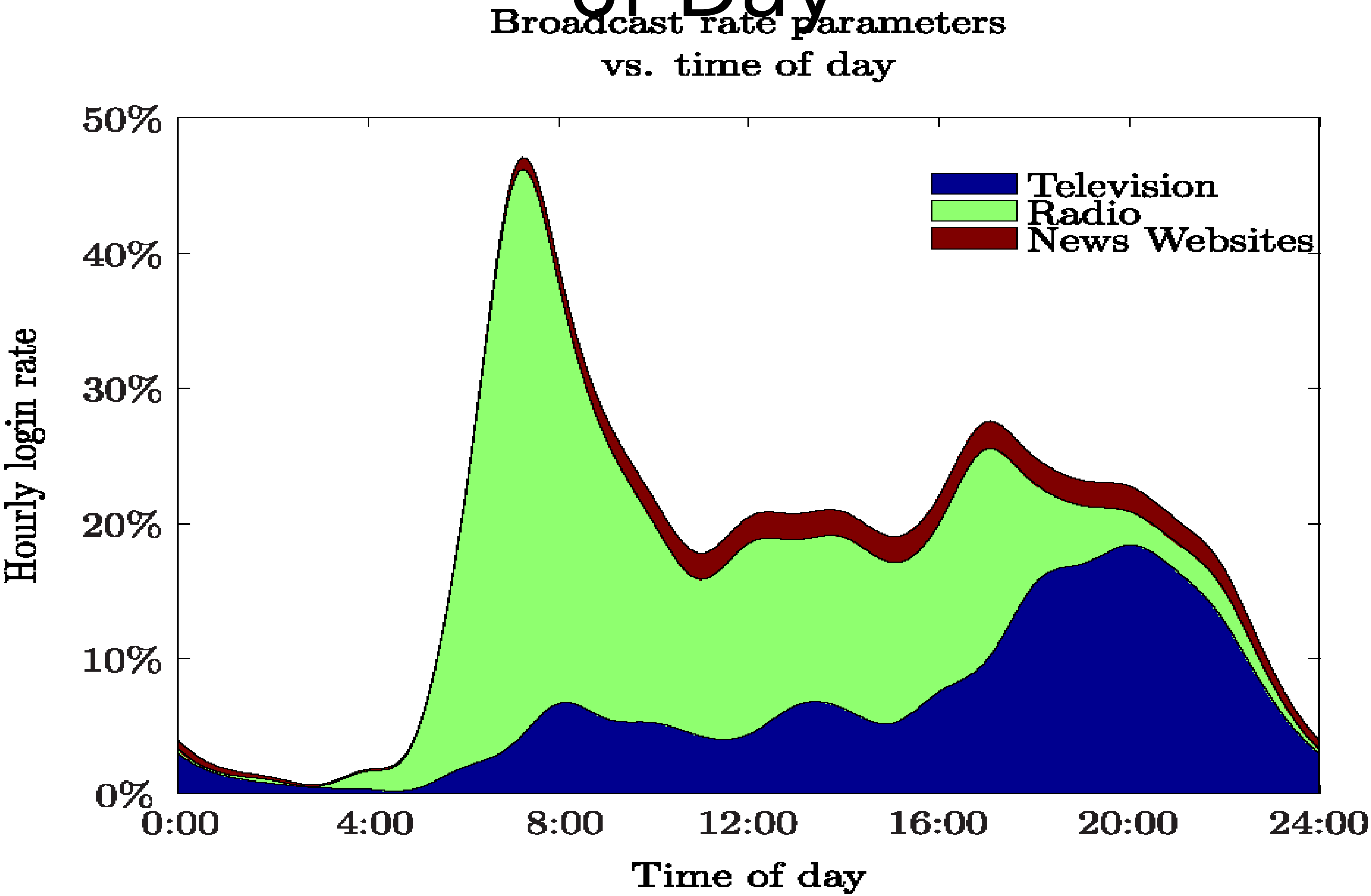
Mean-Field Model

- Assumes homogeneous population

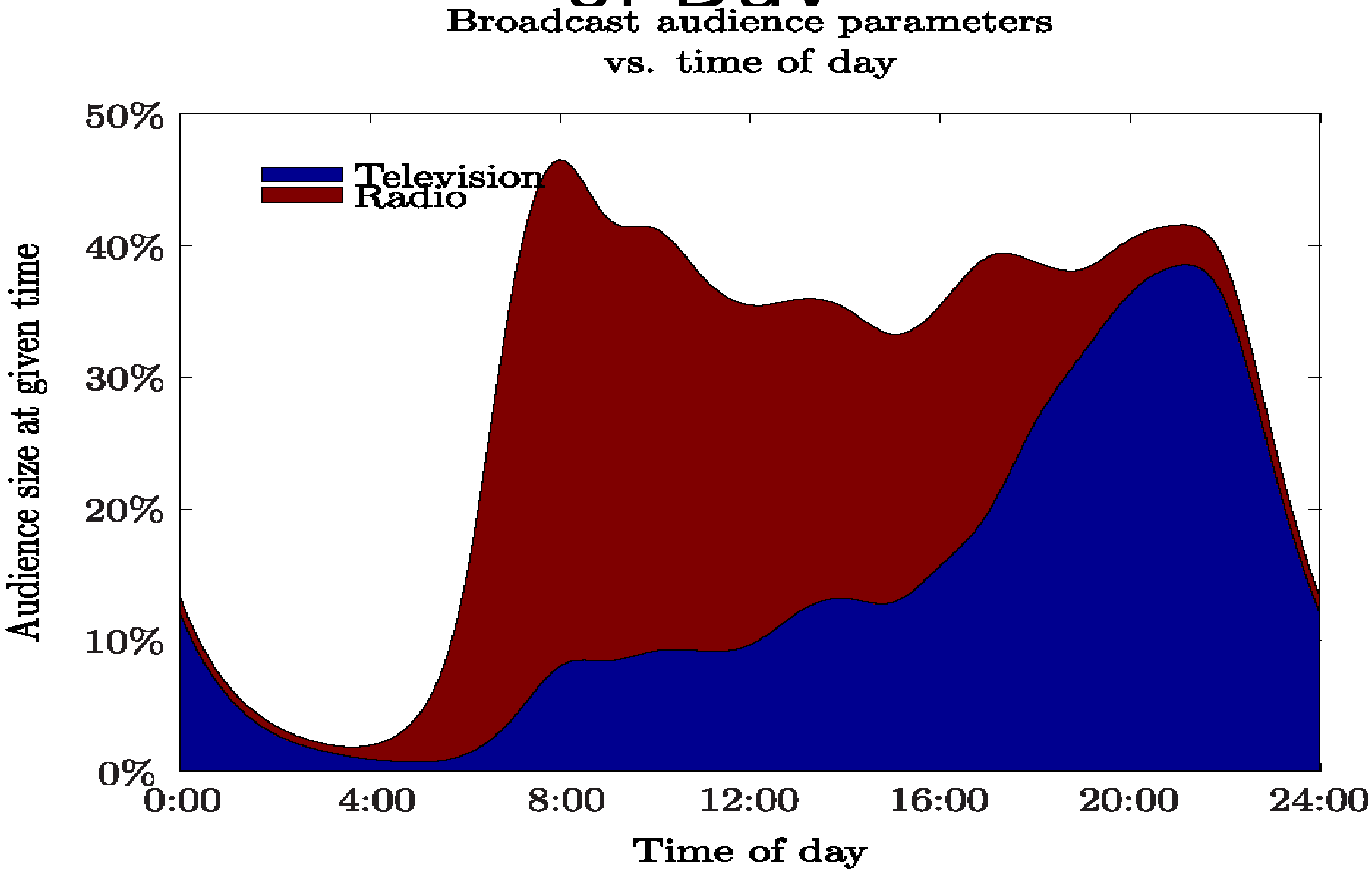
rate of rumour spread = influence of broadcast + influence of social spread



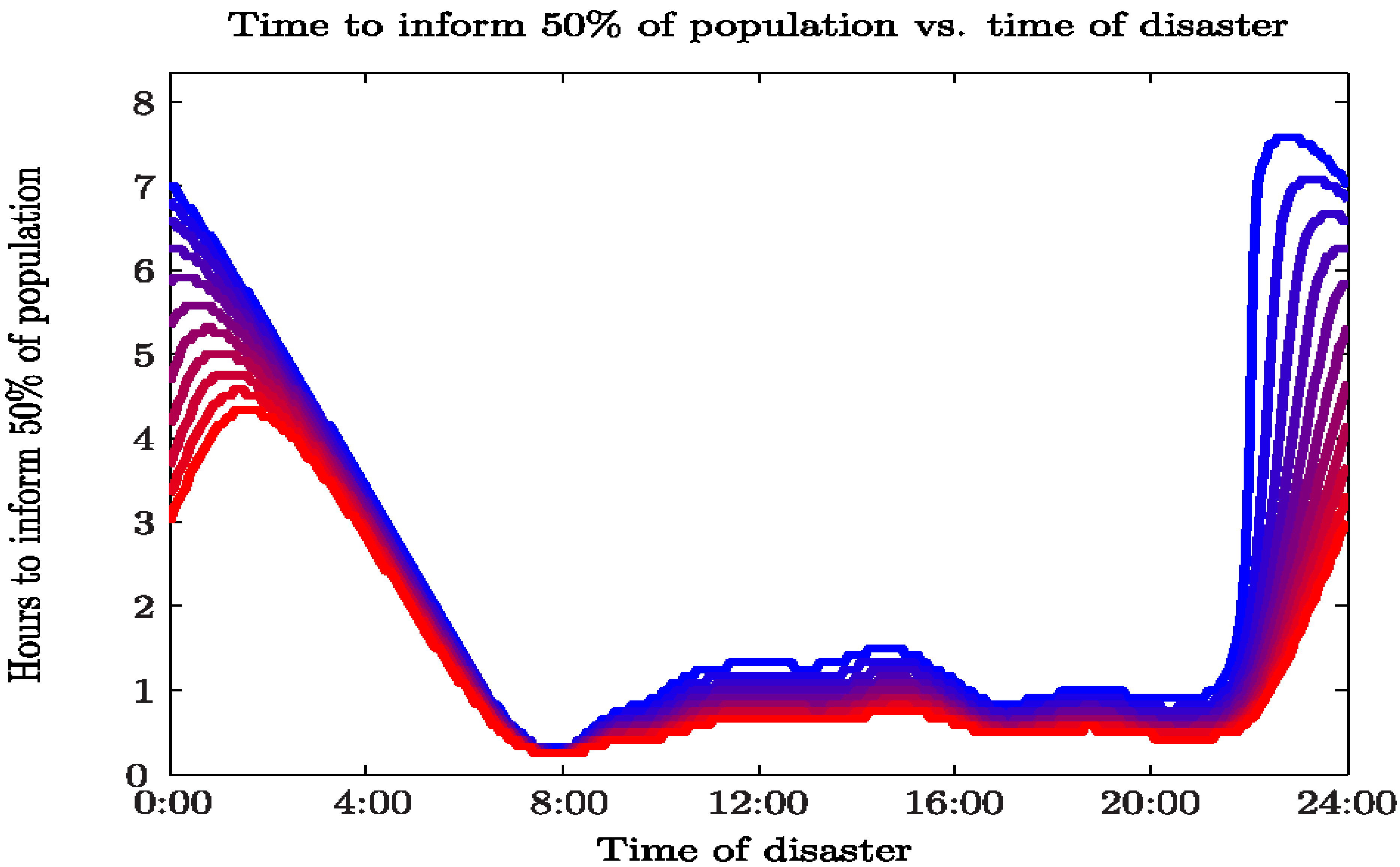
Spreading Parameters vs. Time of Day



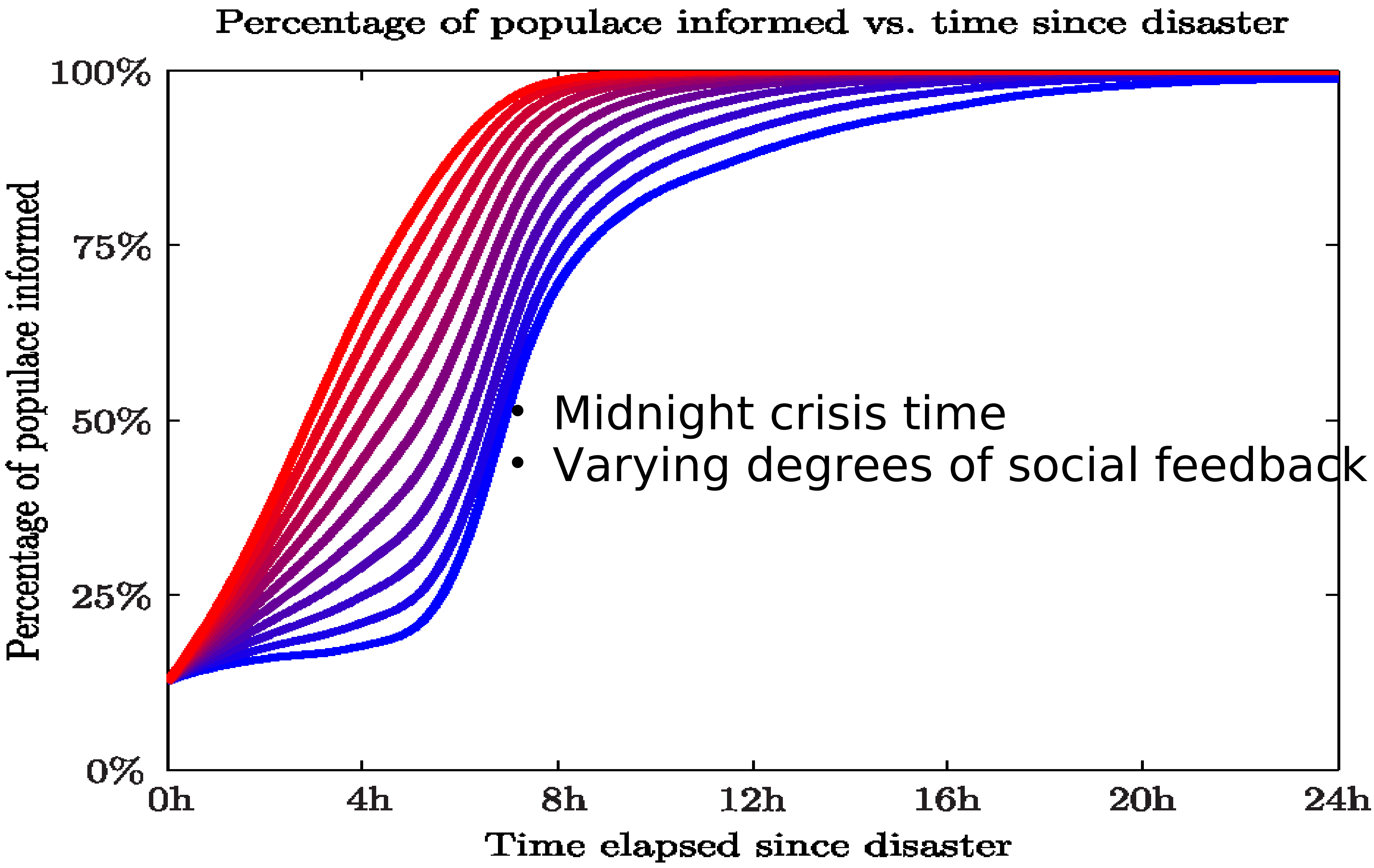
Immediately Informed vs. Time of Day



Speed of Information Spread



Infection Curve Examples



- Midnight crisis time
- Varying degrees of social feedback

The Vaccine Confidence Project

monitoring public confidence in immunisation programmes

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Objective: to detect episodes of loss in vaccine confidence, by monitoring media reports, along with other information like survey data

Can we come up with an index/formula that combines various factors and gives us an indication of the level of confidence in a particular vaccine at a given place and time?

Would also like to be able to predict future trends, and ideally model the spread of news/rumours on undrelying social networks

Working with Heidi Larson and David Smith at LSHTM, we would like to focus on attempting this in the Indian context

Twitter data has been used, e.g. to track the influenza A H1N1 pandemic in the U.S. [Signorini et al., PLoS ONE 2011]; but such data less available for developing countries

Directions for Development

Automate data collection and processing, like sentiment analysis of online media reports

Use vaccinators themselves as sensors for concerns? Convenient mechanism for vaccinators and others to convey sentiment on the ground: applications on mobile devices?

Such applications could also be used for broadcast and control purposes, to pre-empt or combat loss in confidence

Mathematically model relationship between various indicators and vaccine coverage levels; attempt to train a model that has some predictive power

Reconstruct social network of people/regions from observed information flows