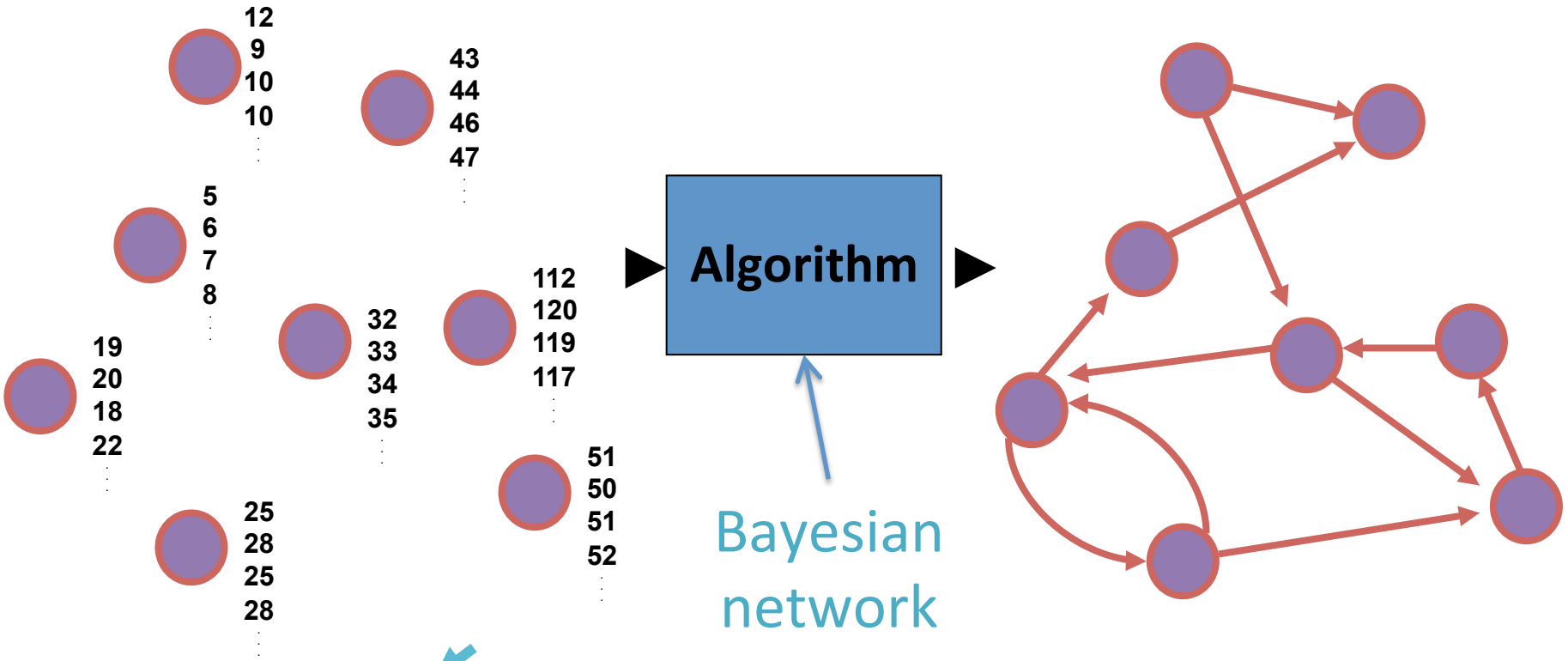
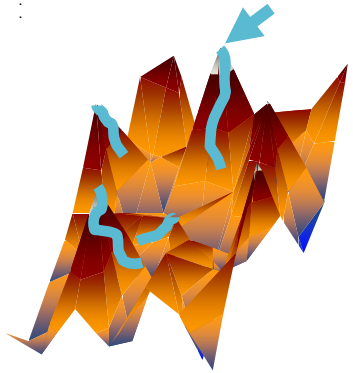
Bayesian Networks for Biological Discovery: Brains, Genes, and Ecosystems

V Anne Smith & Edwin Hui
Centre for Biological Diversity
School of Biology
University of St Andrews, Scotland

Biological Network Inference



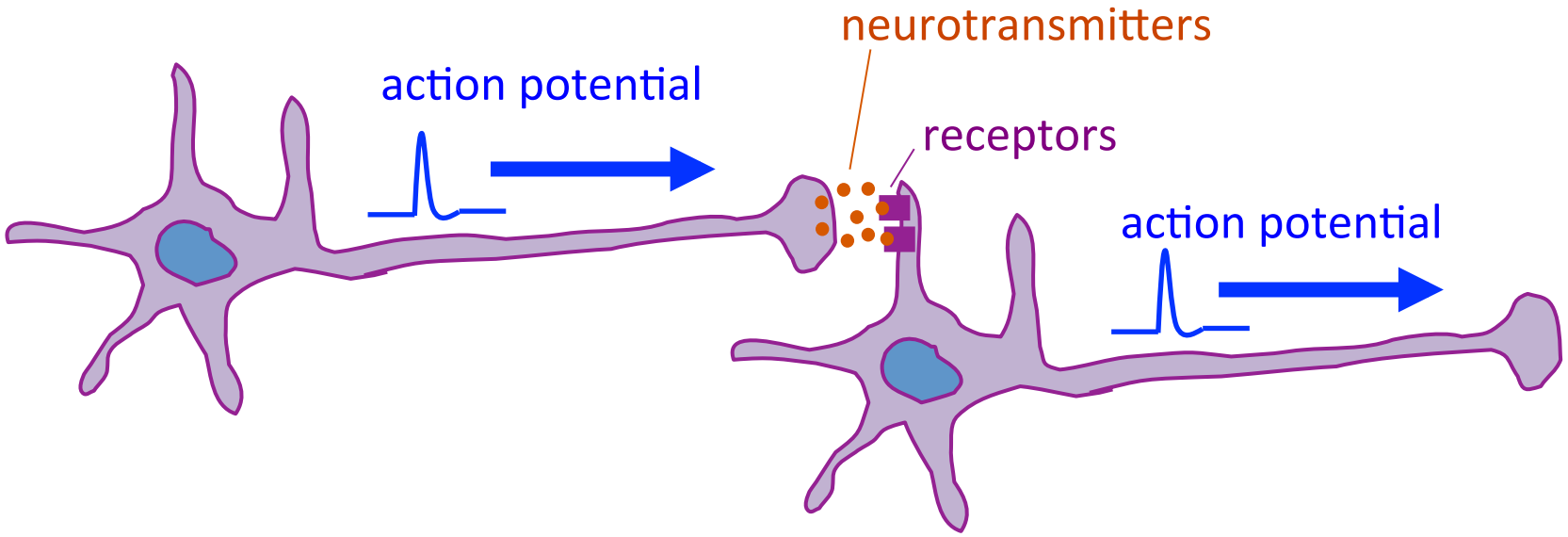


Heuristic search for the best Bayesian network to fit the data, under a given score.

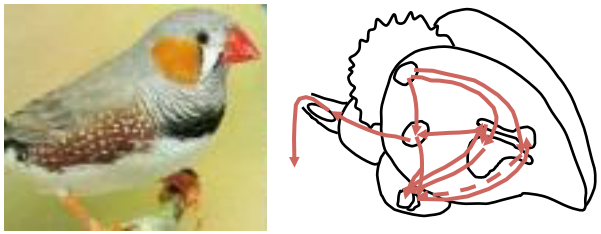
Biological Systems

- Neuronal systems
- Genetic systems
- Ecological systems
  - Rocky shore ecosystem & ecological resilience – Edwin Hui

Neuronal Information Flow Networks



excitation & inhibition



Neuronal information flow network

Neuronal Networks

- Well-suited to Bayesian networks

- Large amount of data (1,000s − 10,000s+ samples)
- Time series − enables dynamic Bayesian networks
- Neurons hypothesised to interact like a Bayesian network

Validated in Songbird Auditory System

Jing Yu

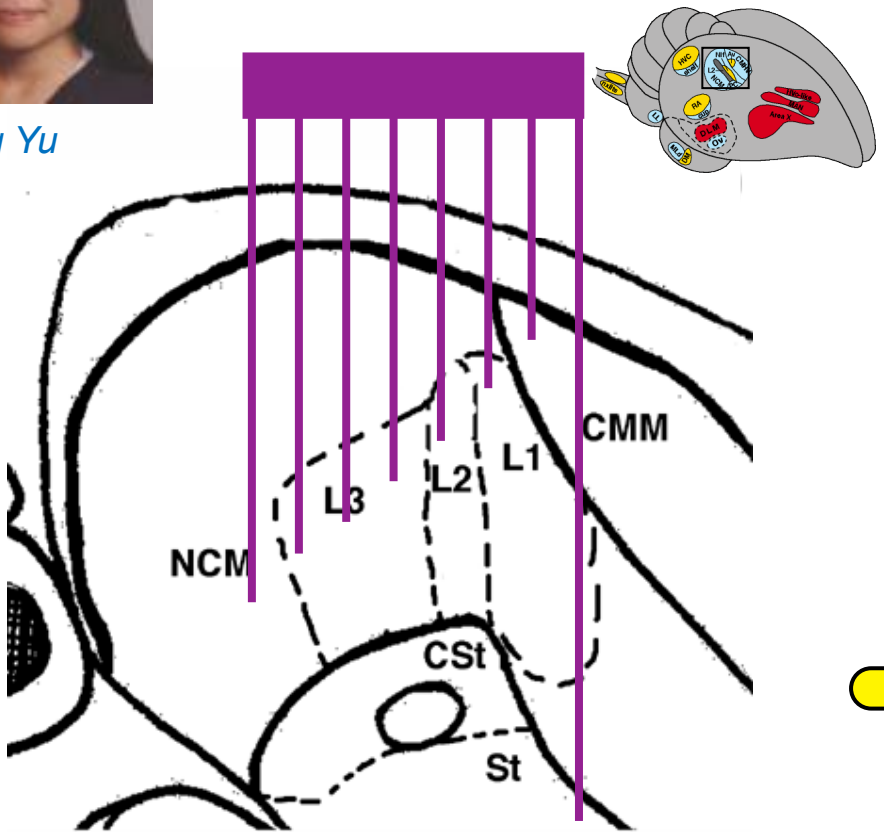
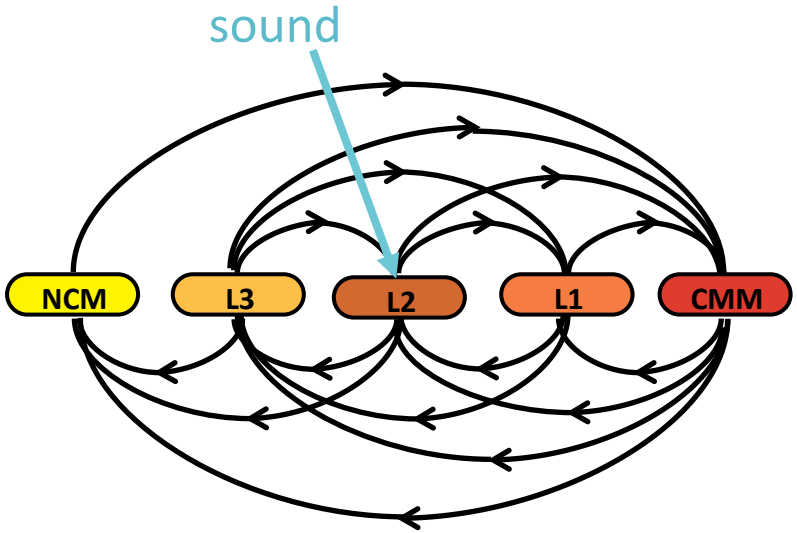


Photo by Tom Smulders



Data collection by Dr Tom Smulders, Newcastle University

Sound Stimuli

6 birds, 4 recording sessions, 4 stimuli

Photo by Haru Horita

(2) zebra finch song

(1) white noise

(1) amplitude modulated white noise

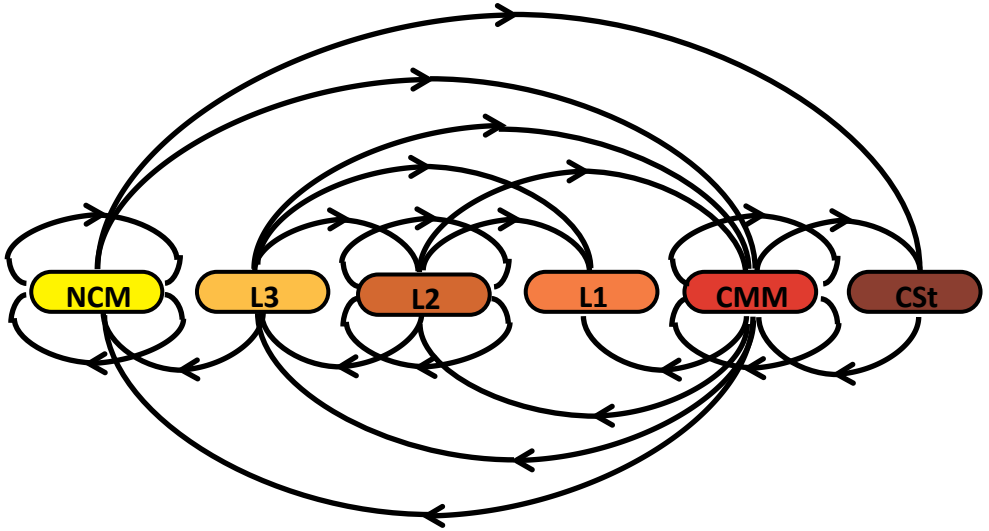
stimuli ~6 seconds long, 1/3 silence, 1/3 sound, 1/3 silence

20 repetitions, in 5 ms time bins = >20,000 data points

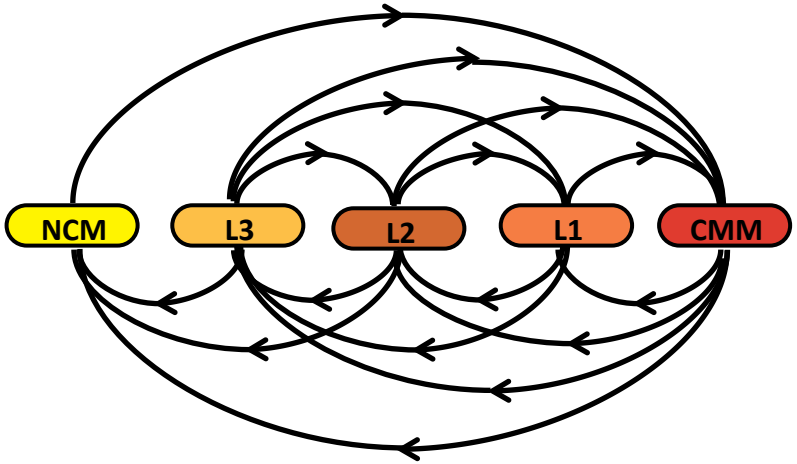
Data discretized; use BDe Score

Consensus network matches with known anatomy

consensus

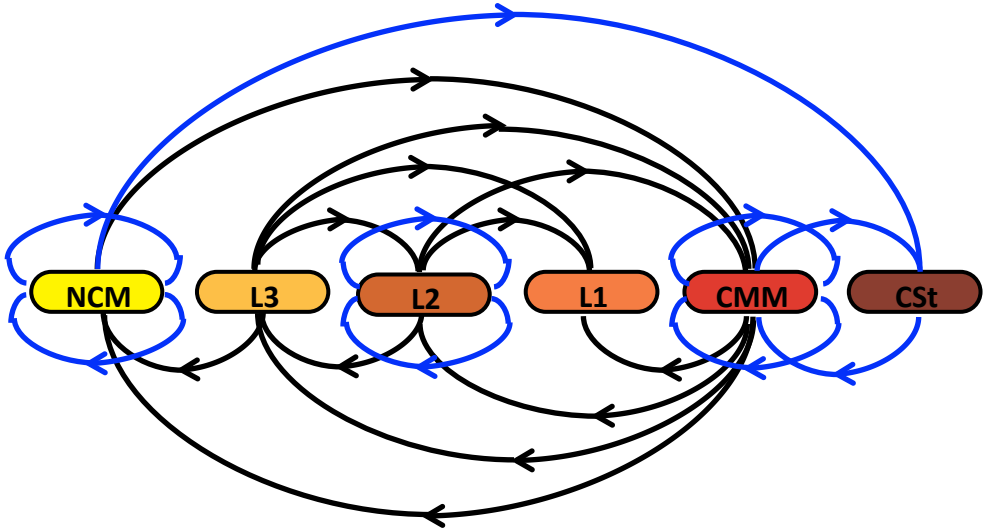


known anatomy

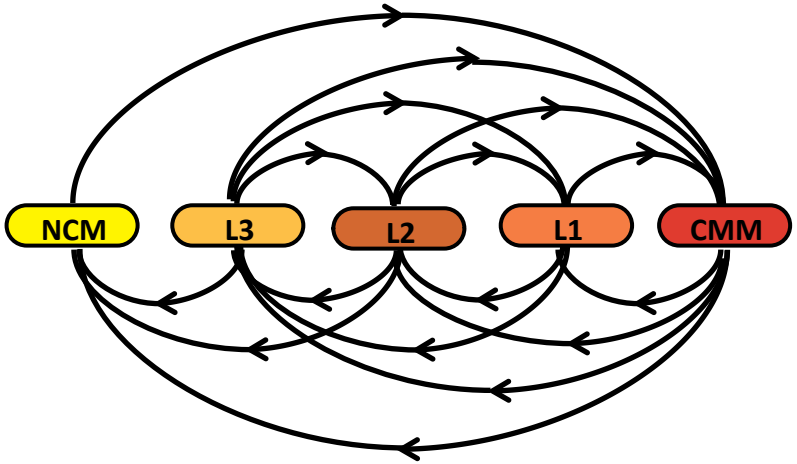


Consensus network matches with known anatomy

consensus



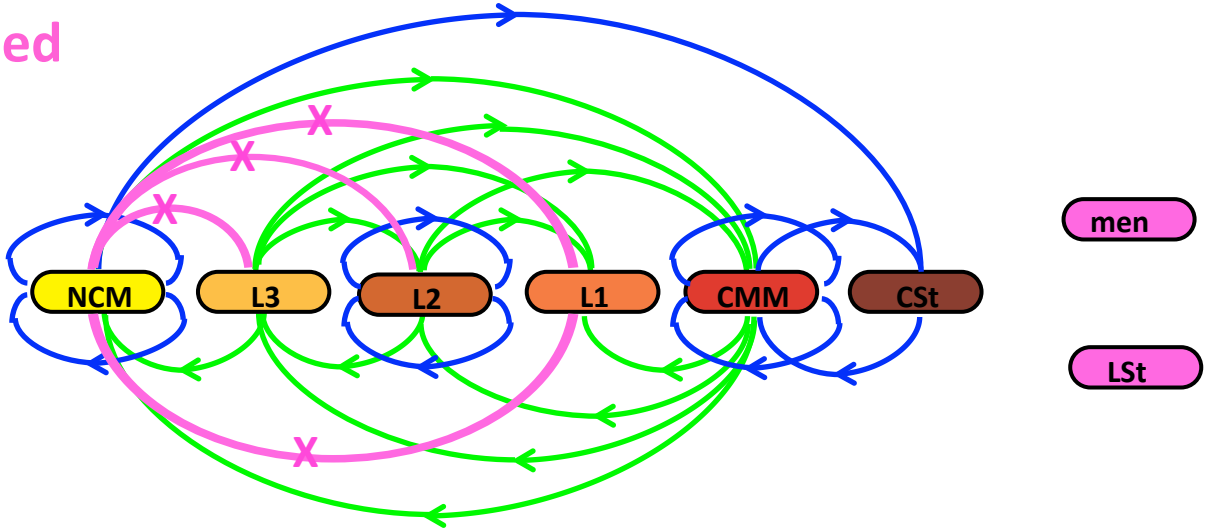
known anatomy



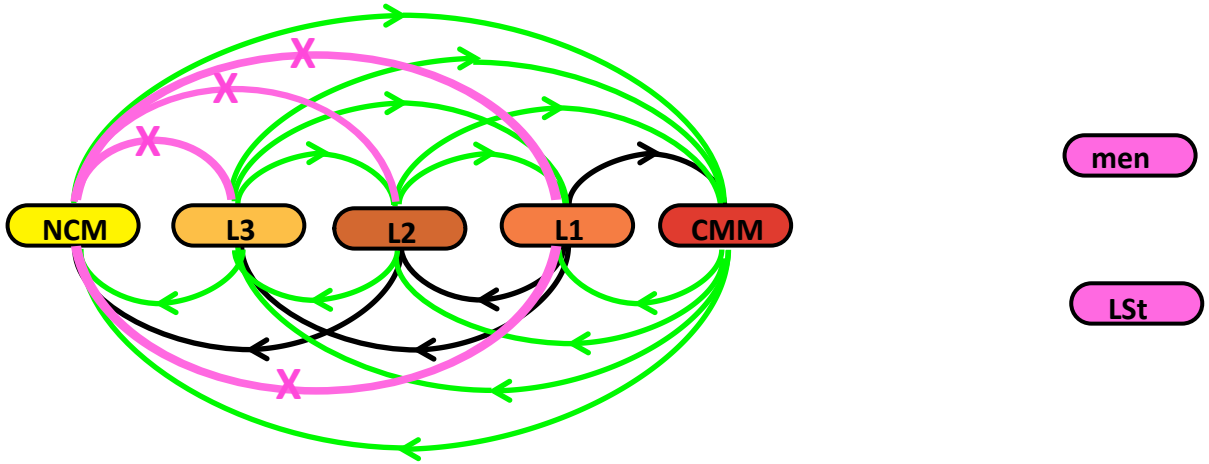
Consensus network matches with known anatomy

significantly avoided
($P=9.2 \times 10^{-6}$, hypergeometric test)

consensus

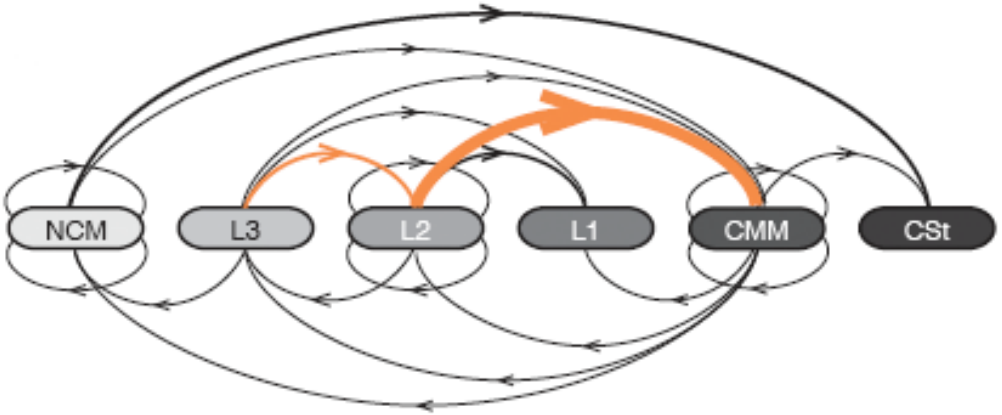


known anatomy

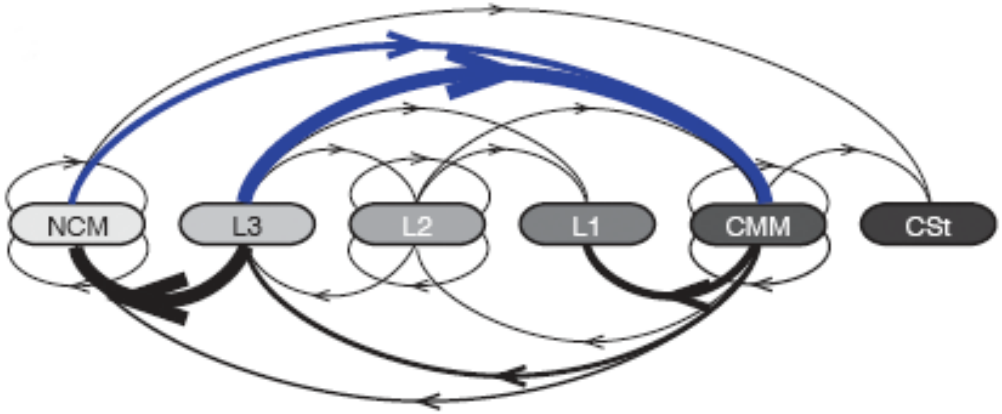


Noise & Song Differ

noise stimuli

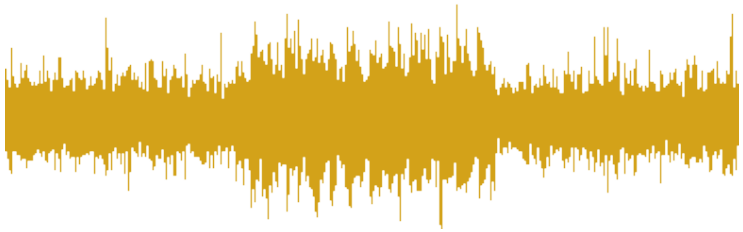


song stimuli



Spike Trains: Trouble for BDe!

- Preponderance of non-spiking bins



songbird data ≈ normal

spike train ≈ Poisson

- Lack of *semantics*

$$BDe(G) = \log\left(\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})}\prod_{k=1}^{r_i}\frac{\Gamma(\alpha_{ijk}+N_{ijk})}{\Gamma(\alpha_{ijk})}\right)$$

high when parent and child both non-spiking; rare spiking events have little influence

Snap Shot Score

Christoph Feenders

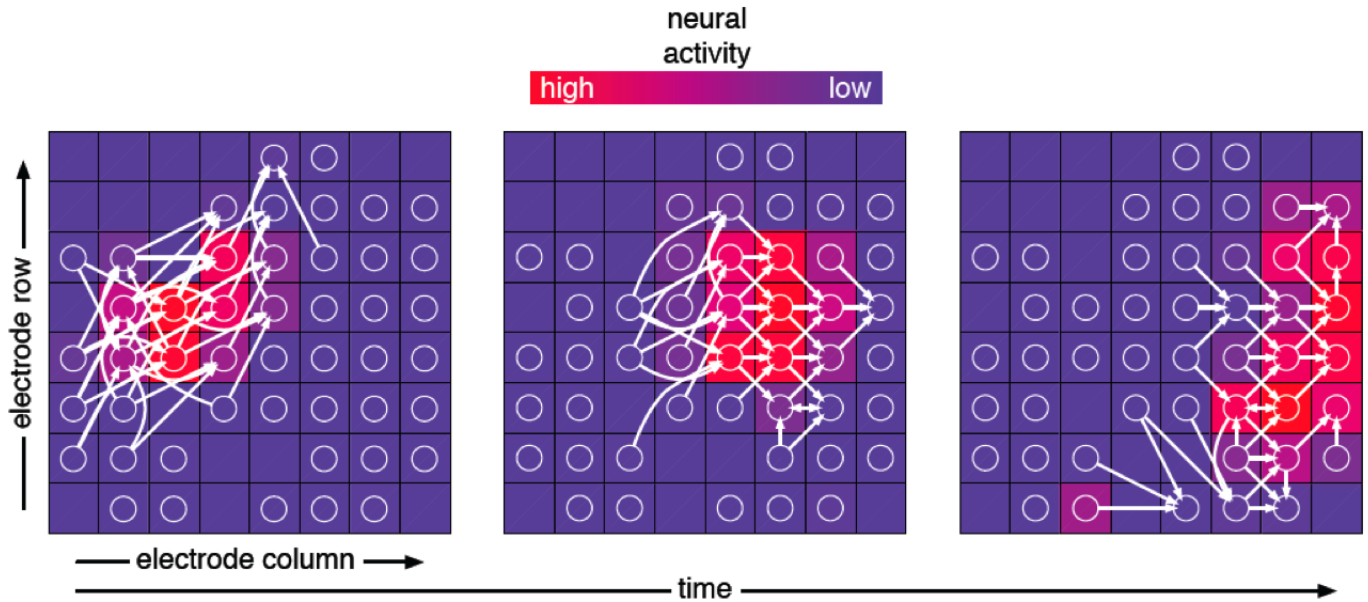
- Treat spiking and non-spiking differently
- Hybrid networks (discrete and continuous):
  - Combine precision of *spike train* with smoothing convolution, *activity level series*
- Likelihood score for excitatory interactions: conditional probability

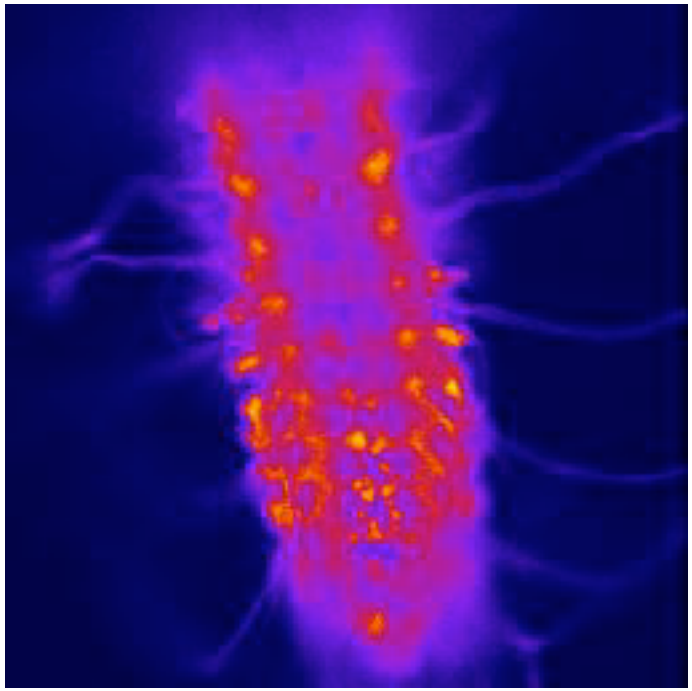
$$SSS(\Delta t) = \prod_{a,s} \frac{\sum_{t=1}^{T-\Delta t} a_t s_{t+\Delta t}}{\sum_{t=1}^{T-\Delta t} a_t}$$

$$P(\text{spike} \mid \text{parent activity}) = \frac{P(\text{spike and parent activity})}{P(\text{parent activity})}$$

Spike Train Data

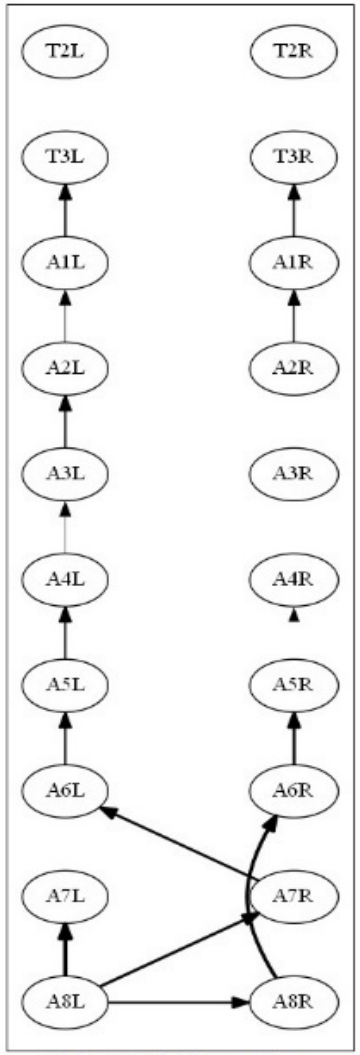
- Retinal activity waves
  - *Data from Dr Evelyne Sernagor, Newcastle University*
  - 5 sec recordings, multi-unit spike trains



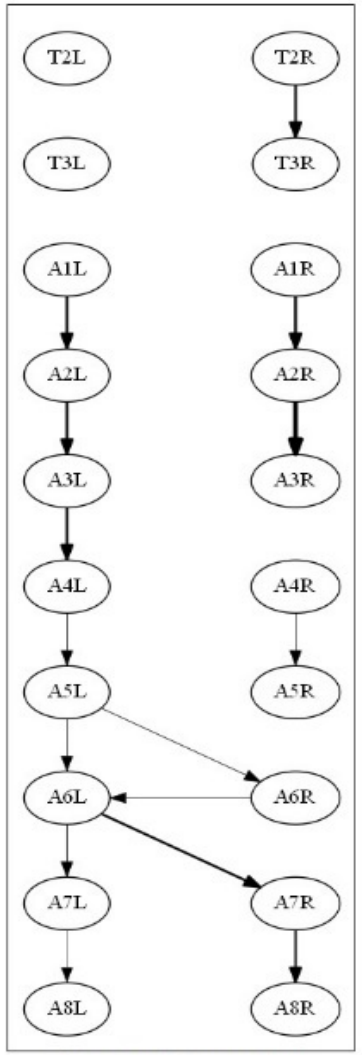


Data from Stefan Pulver,
St Andrews University

- Drosophila larvae: live imaging
- Identify ganglion as regions of interest



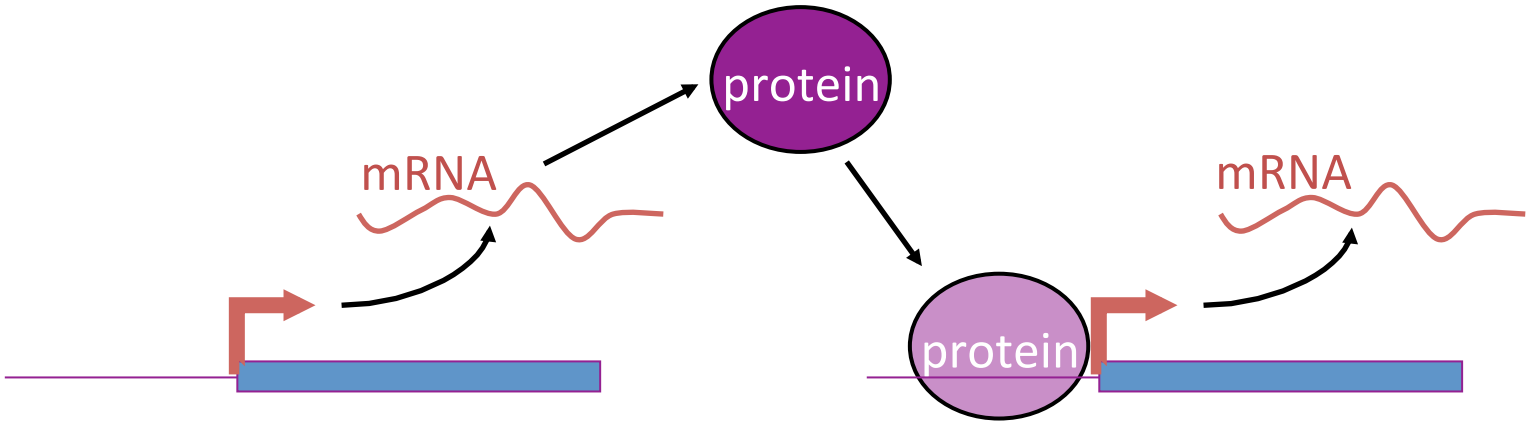
forward



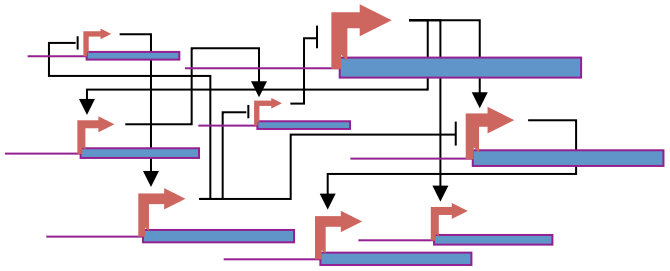
backward

Alistair Birse-Stewart-Bell, Staśa Tumpa, Jacob Francis

Gene Regulatory Networks

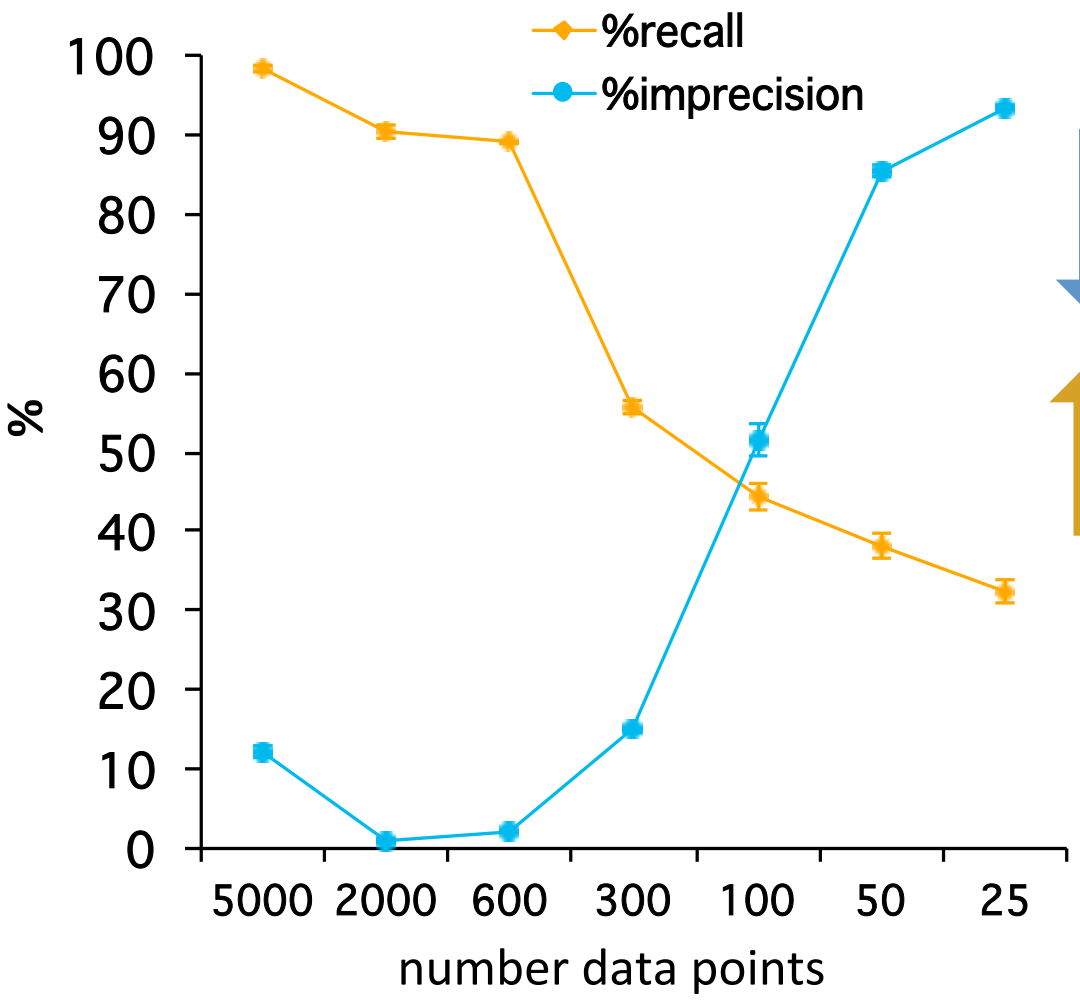


combinations of
activators & inhibitors



gene regulatory network

Simulation Results



Challenges:

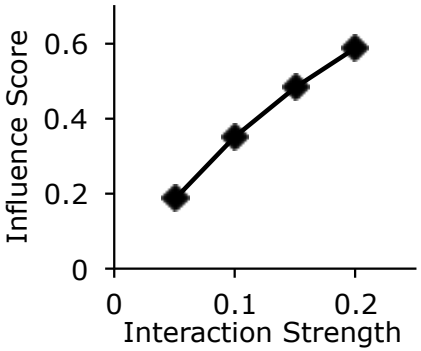
1. Decrease false positives

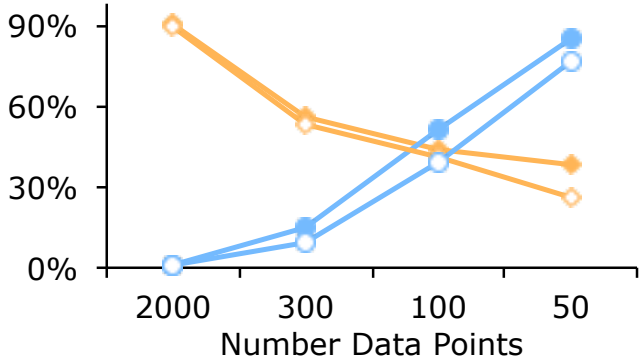
2. Increase true positives

Jing Yu

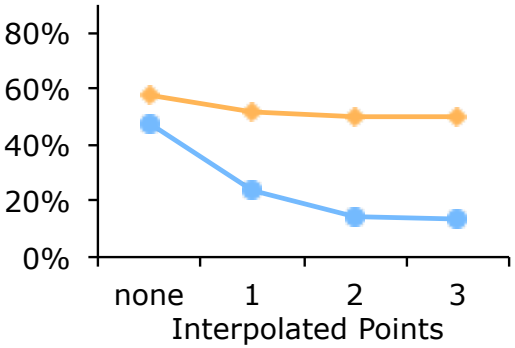
Decreasing False Positives

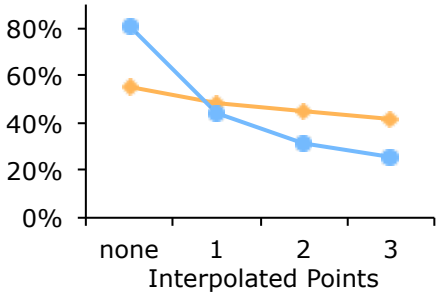
- Influence score
  - ↓ imprecision
  - ↓ recall





- Interpolation
  - ↓ imprecision
  - ↓ recall



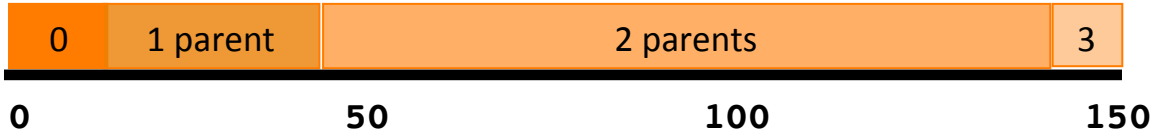


- Limit parent number
  - ↓ imprecision
  - ? recall

Limit on N to avoid FP

$$N \geq d^p(2d-1)$$

Given N, choose p to avoid FP

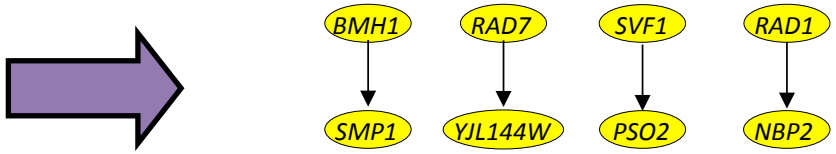


Dealing with Low Data Amount

Yeast stress response (36 data points,DBN)

Causton et al. 2001 Mol Biol Cell 12:323-337

- Influence score
- Interpolation
- Limit parent number



Biological Discovery in Genetic Data

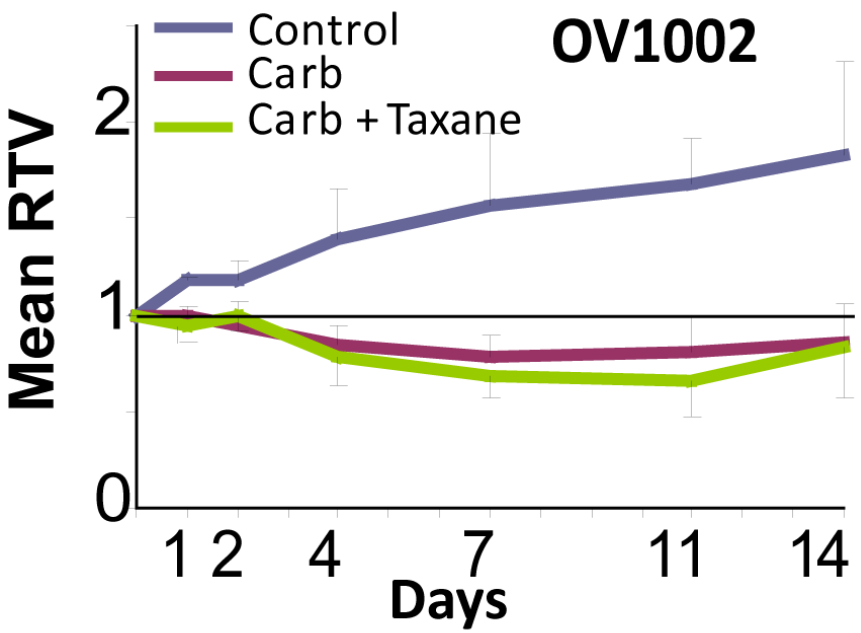
- Identifying gene/protein networks?
  - not as much
- Identifying genes/groups of genes of interest
  - including other variable types in the network
- Gene selection
  - differential expression
  - principal component analysis
  - naïve Bayes
  - clustering (of selected genes)

Ovarian Cancer Example

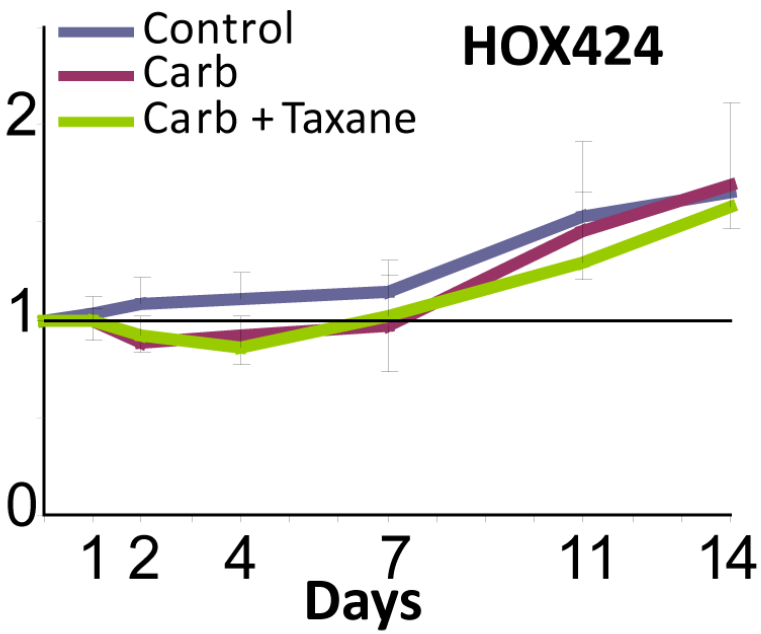
Antonis Koussounadis

- Treatment sensitive & resistant cell lines
  - OV1002 – sensitive
  - HOX424 – resistant
- Two treatment regimes
  - Carboplatin
  - Carboplatin + paclitaxel
- Differential gene expression between control and treatment

Tumour Growth - Relative Tumour Volume

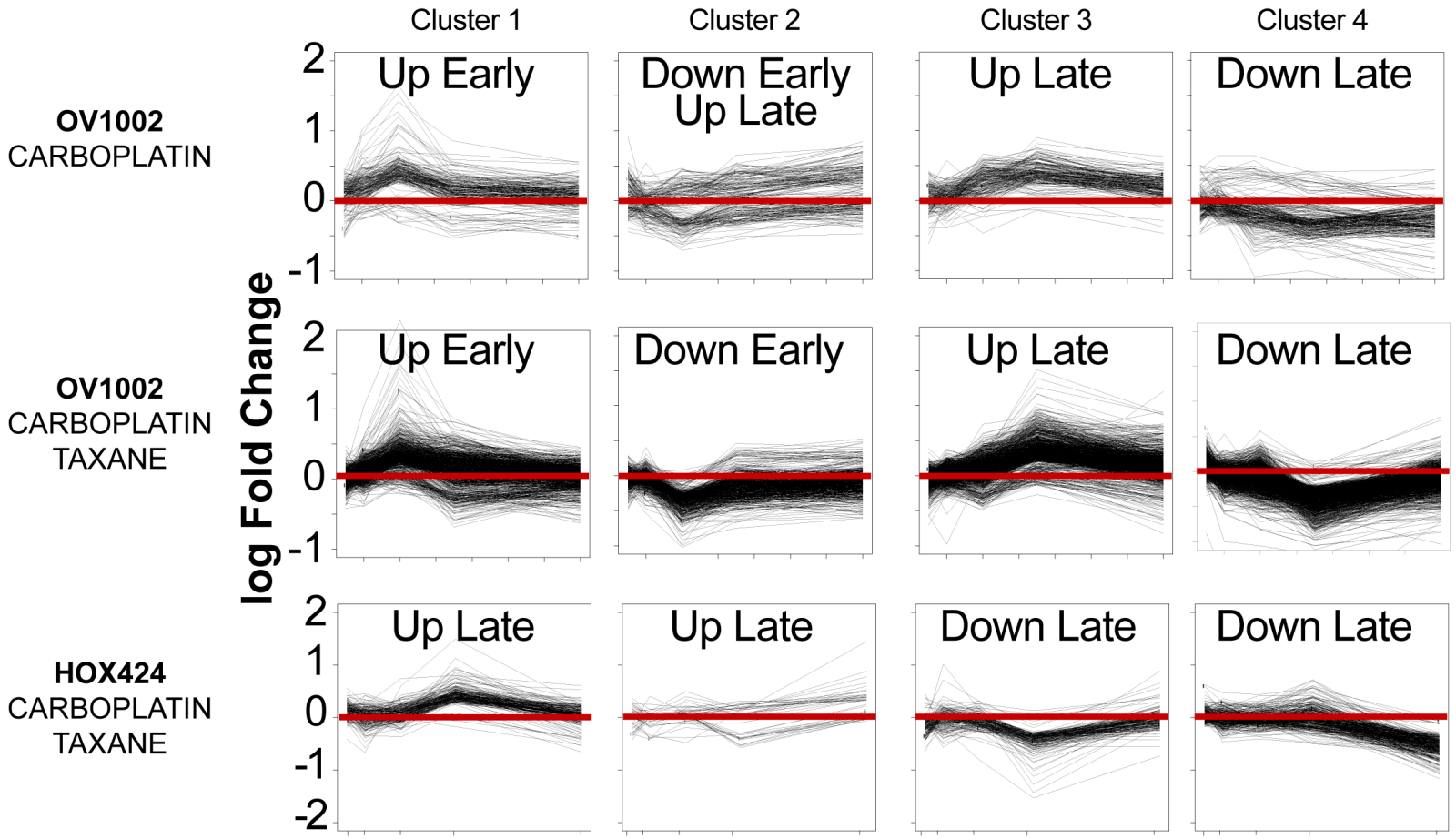


Sensitive - responding

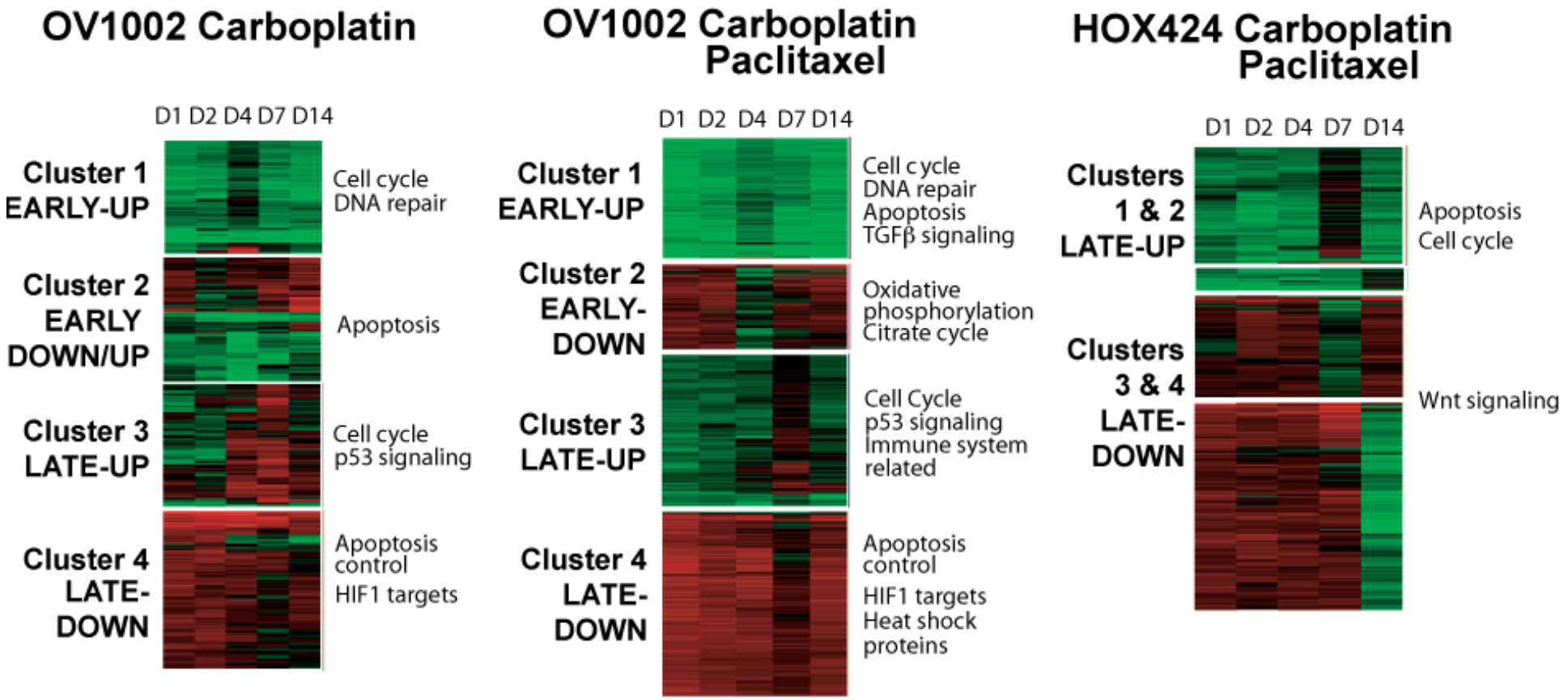


Resistant – not responding

Clustering Differentially Expressed Genes

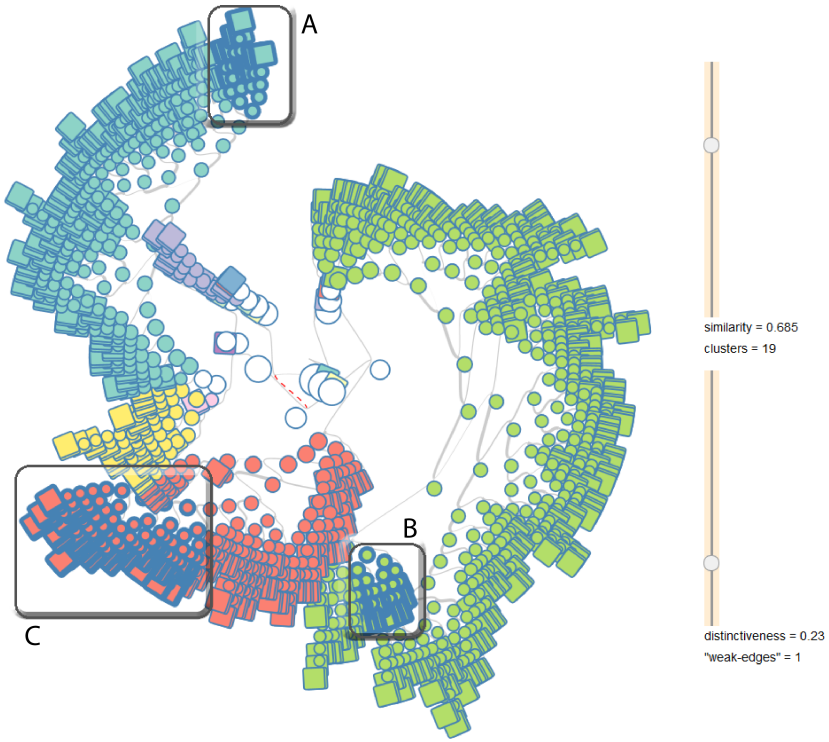


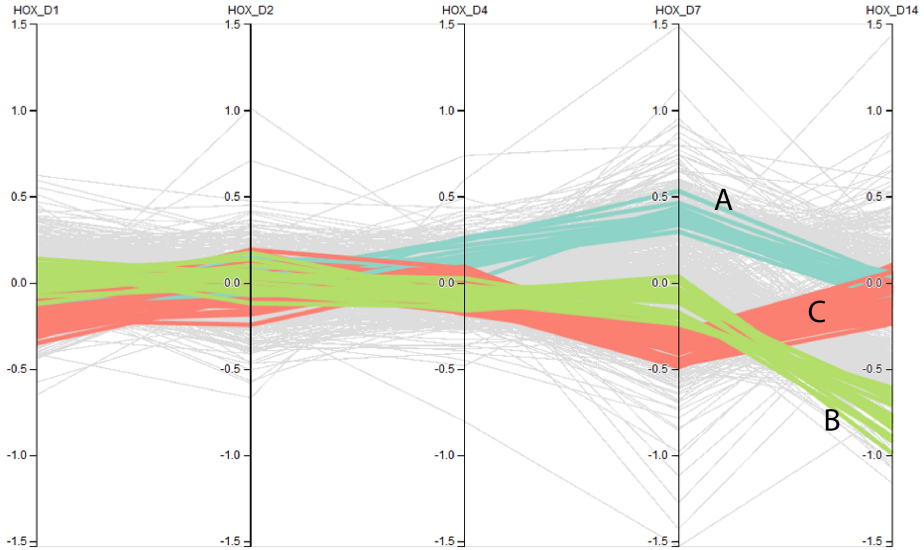
Pathway Overexpression Analysis



MLCut: Multi-level cuts for clustering

Thanasis Vogogias

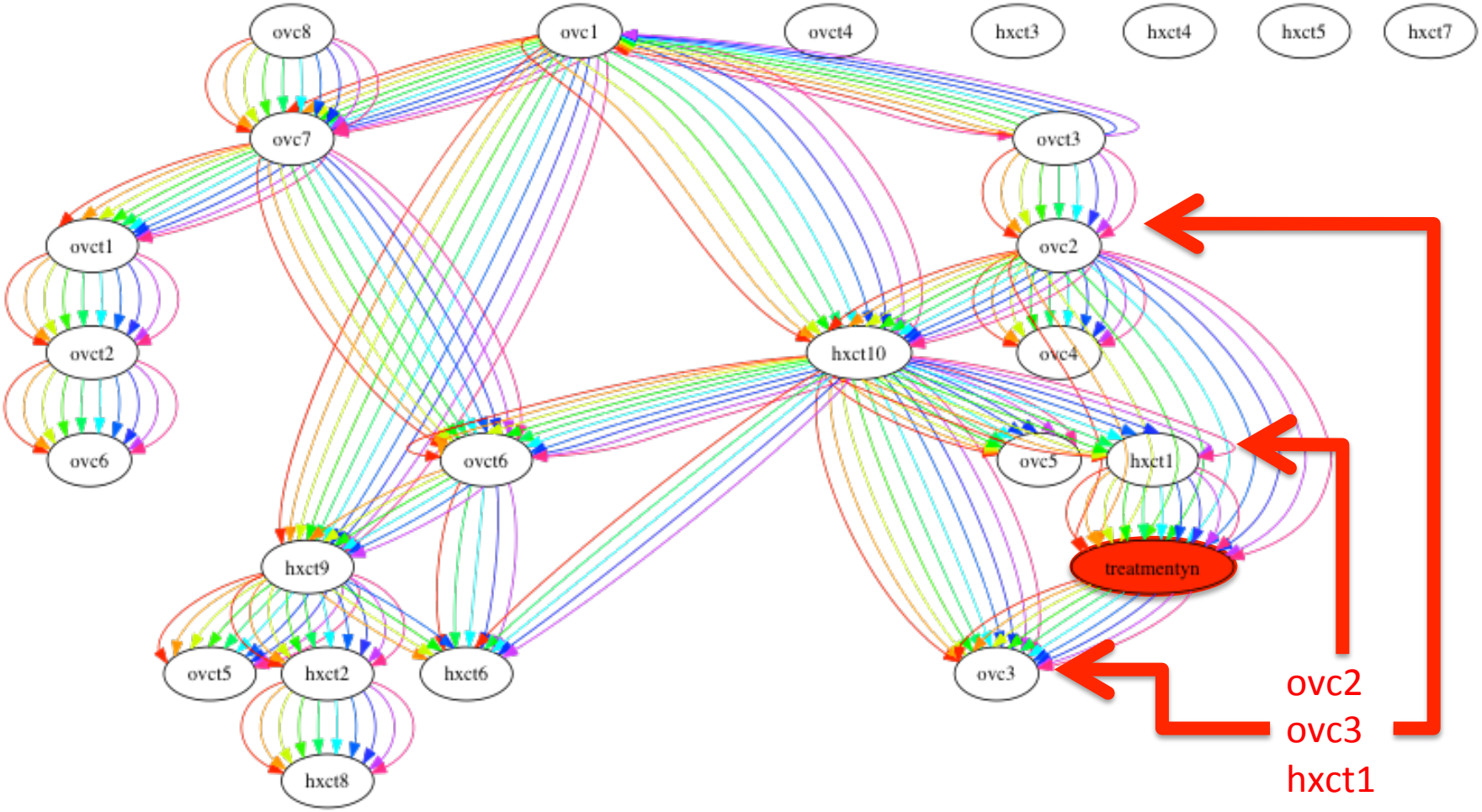




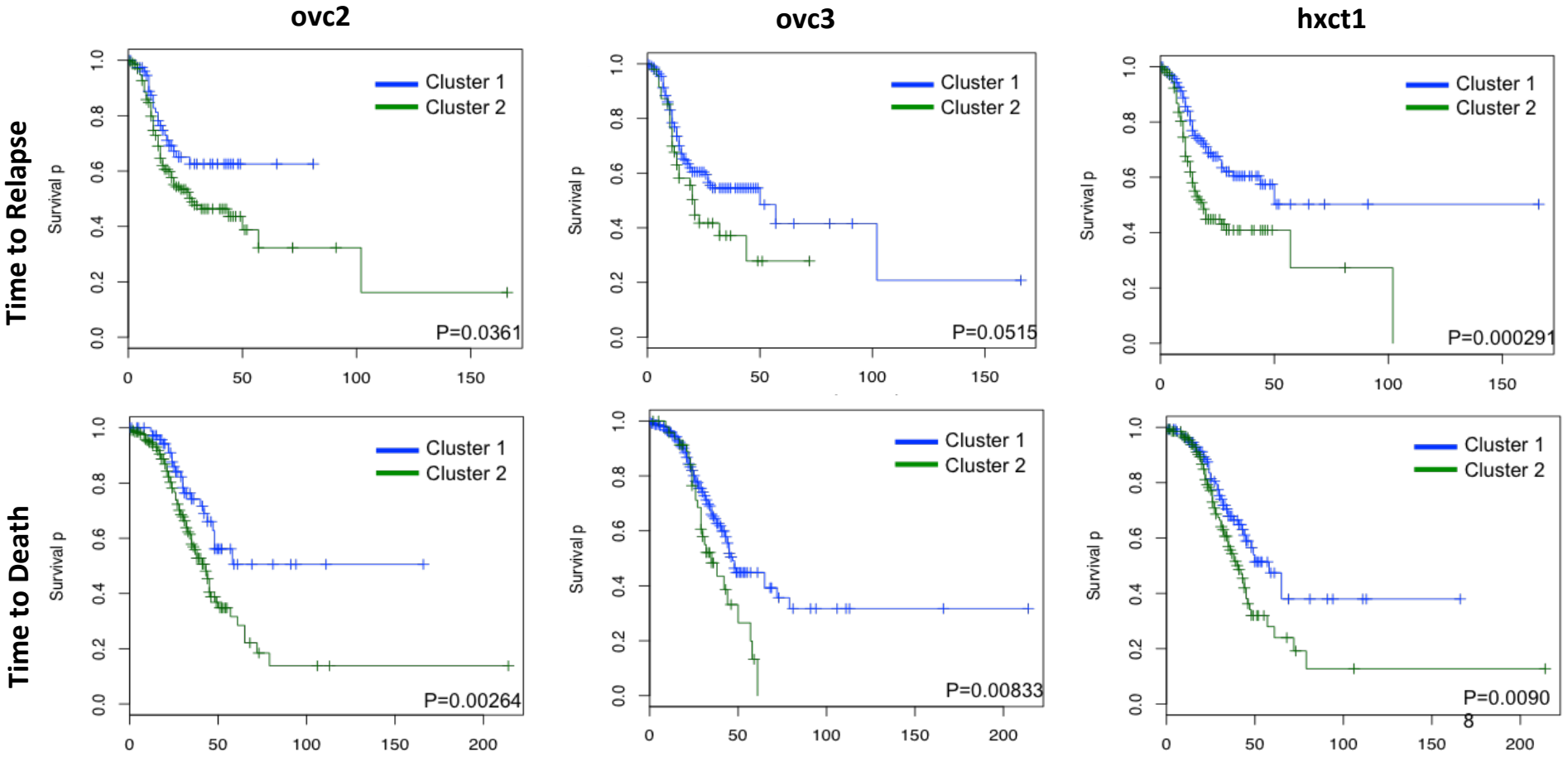
Bayesian networks from clusters

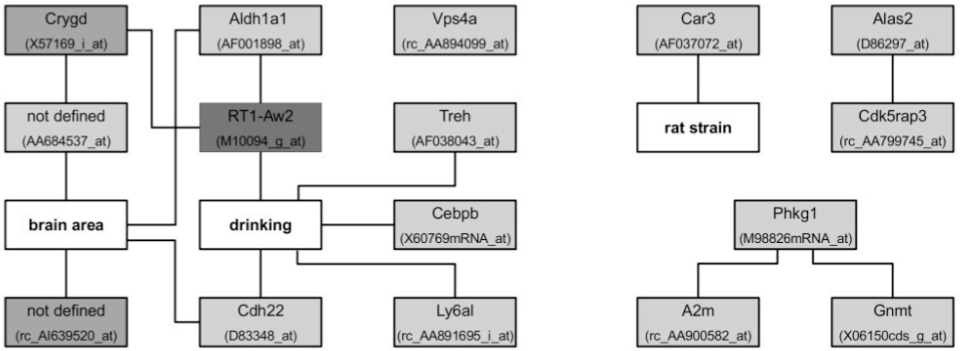
Hannah Currant

- MLCut to define clusters
- Get original data (not differential expression) and build Bayesian network
  - variable for treatment vs control
- Bayesian networks: probabilistic relationships, most direct influences

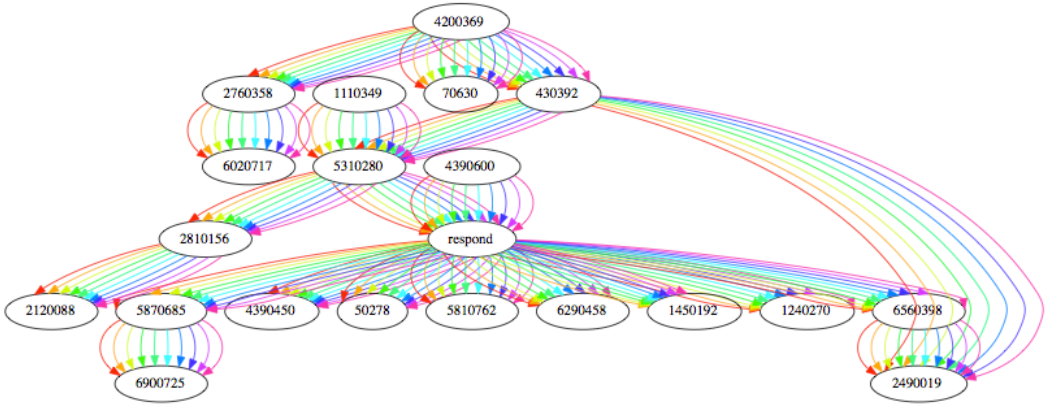

ovc8
ovc1
ovct4
hxct3
hxct4
hxct5
hxct7
ovc7
ovct3
ovct1
ovc2
ovct2
hxct10
ovc4
ovc6
ovct6
ovc5
hxct1
hxct9
treatmentyn
ovct5
hxct2
hxct6
ovc3
hxct8
ovc2
ovc3
hxct1

Independent Clinical Dataset

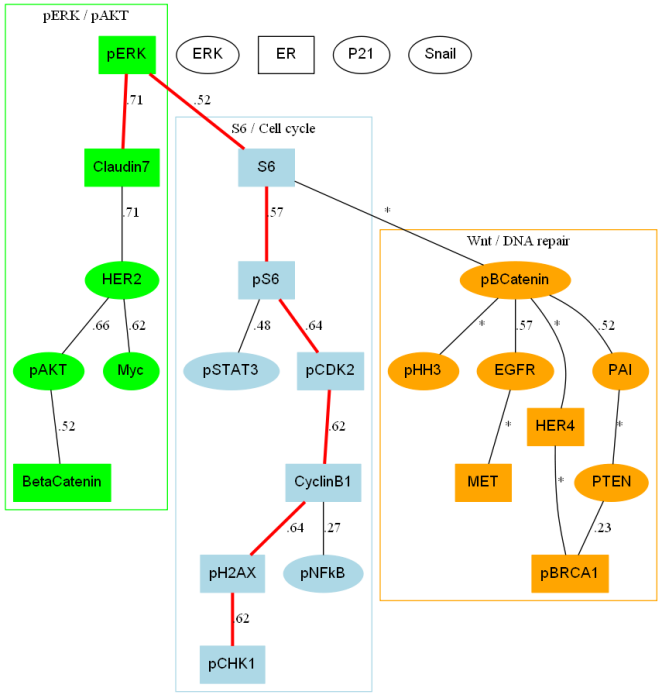




Genes related to alcoholism in rats

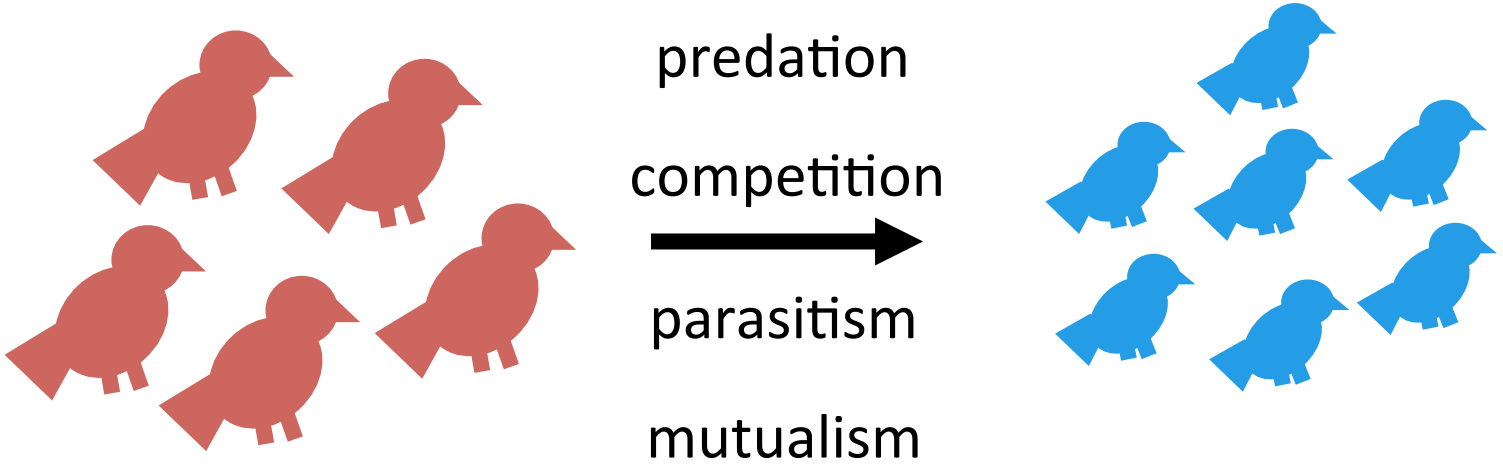


Genes related to patient response in melanoma

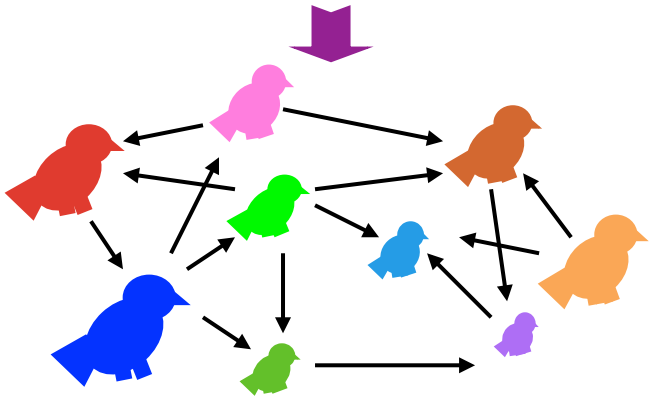


Protein pathways in ovarian cancer

Ecological Interaction Networks

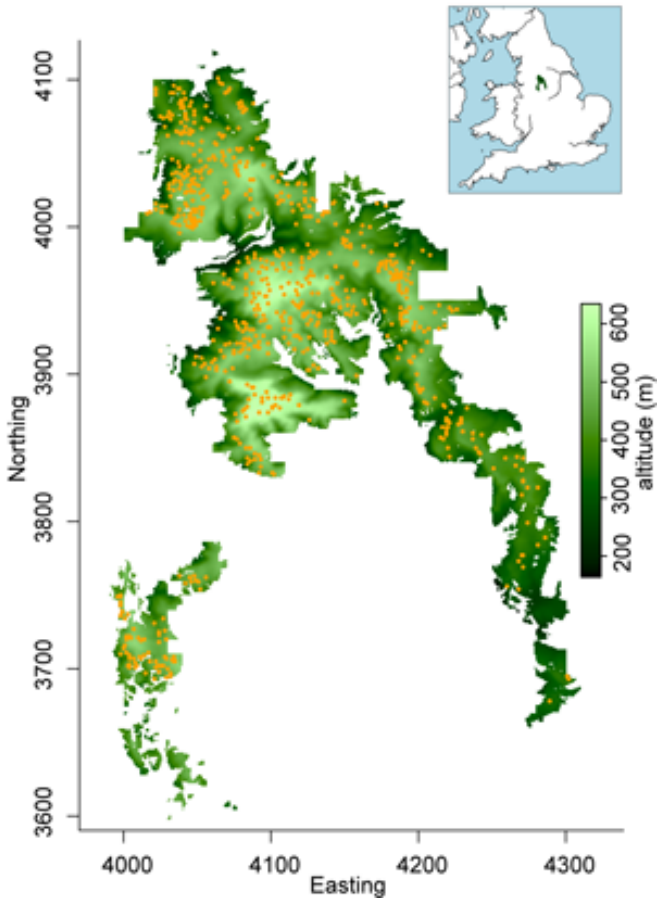


combination of interactions



ecological interaction network

Peak District National Park

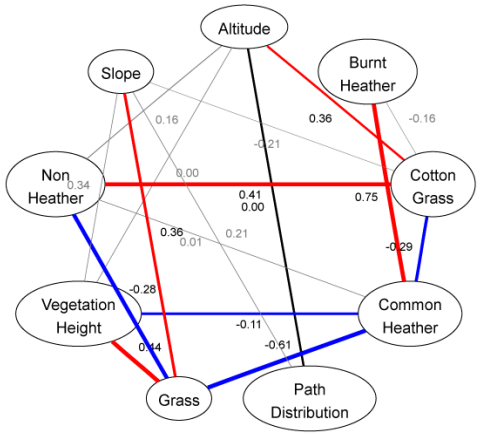


- 37 birds species and 9 habitat variables measured in 1990 and 2004
  - *data from RSPB & Dr Colin Beale, York*
- Extracted at 0.5, 1, 2, and 5 km^2
  - 2099, 610, 176, and 39 data points, respectively
- Static Bayesian network

Peak District Bird Populations

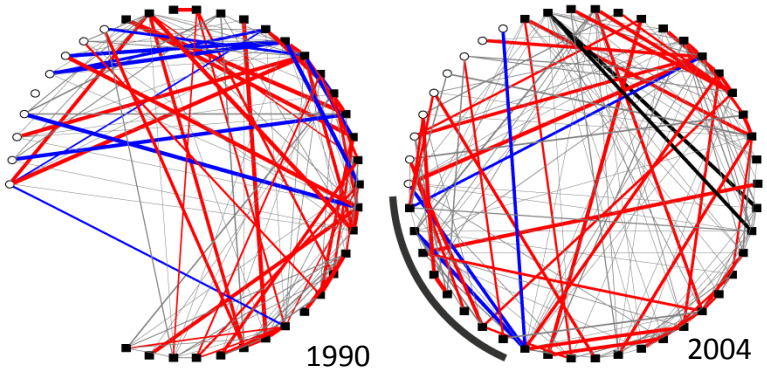
Habitat

- Reveal known interactions



Species & species-habitat

- Interspecific interactions positive
- New species interacting
- ID highly connected species (e.g., Golden Plover)

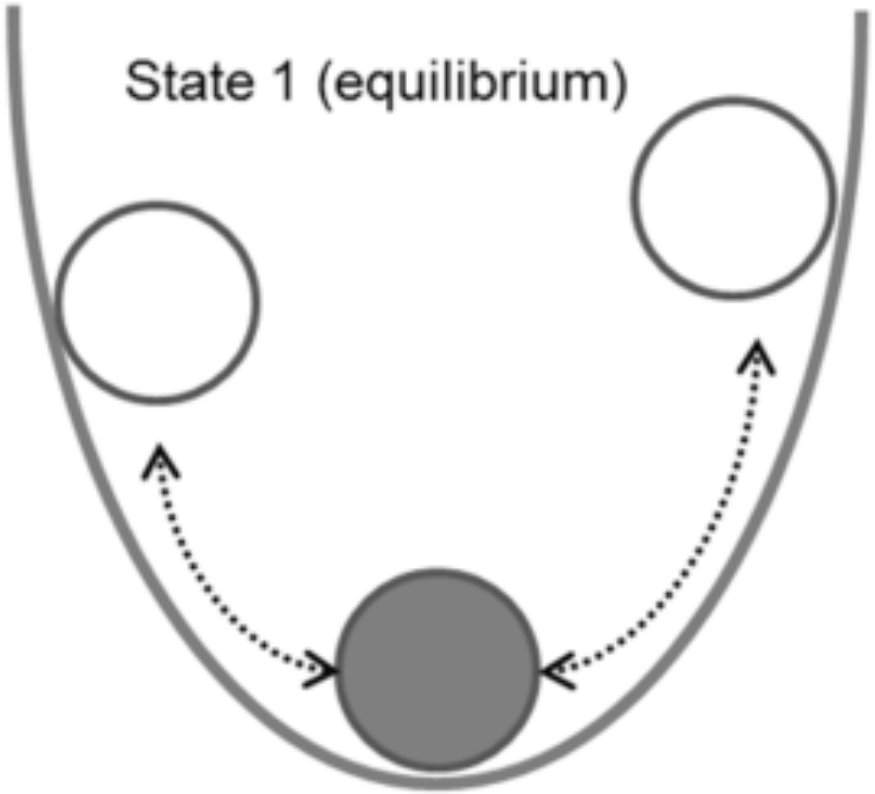


- Characteristic spatial scale for interactions

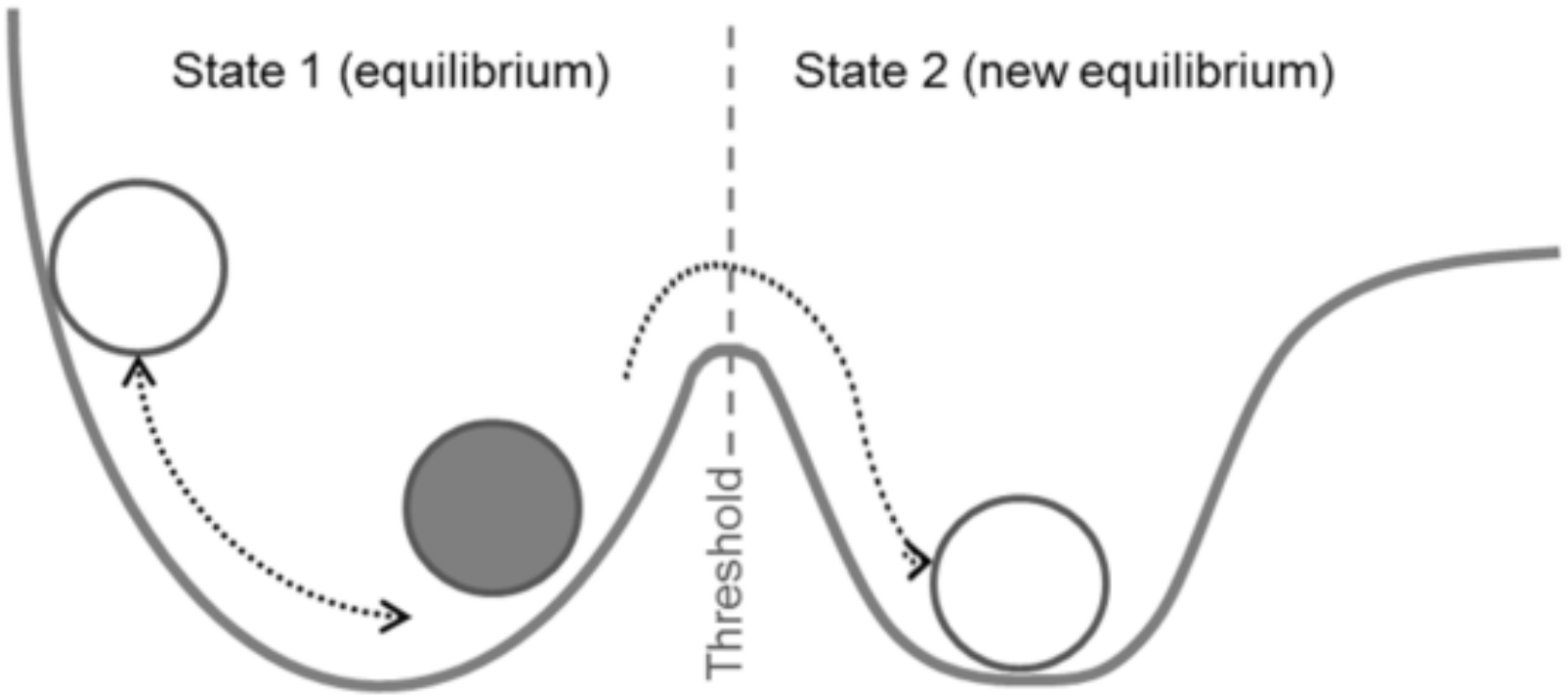
Bayesian Networks in Ecology

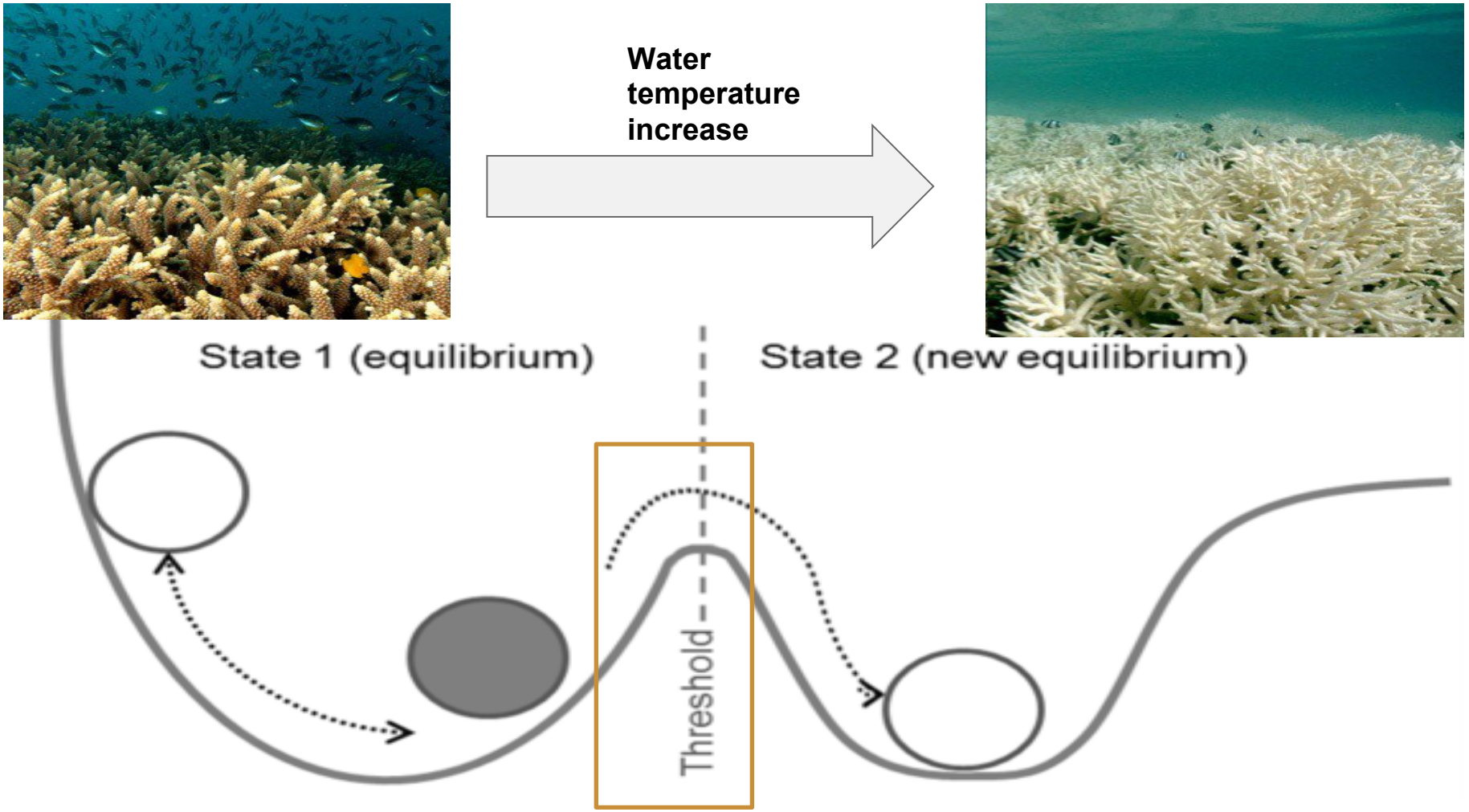
Edwin Hui

Ecological resilience



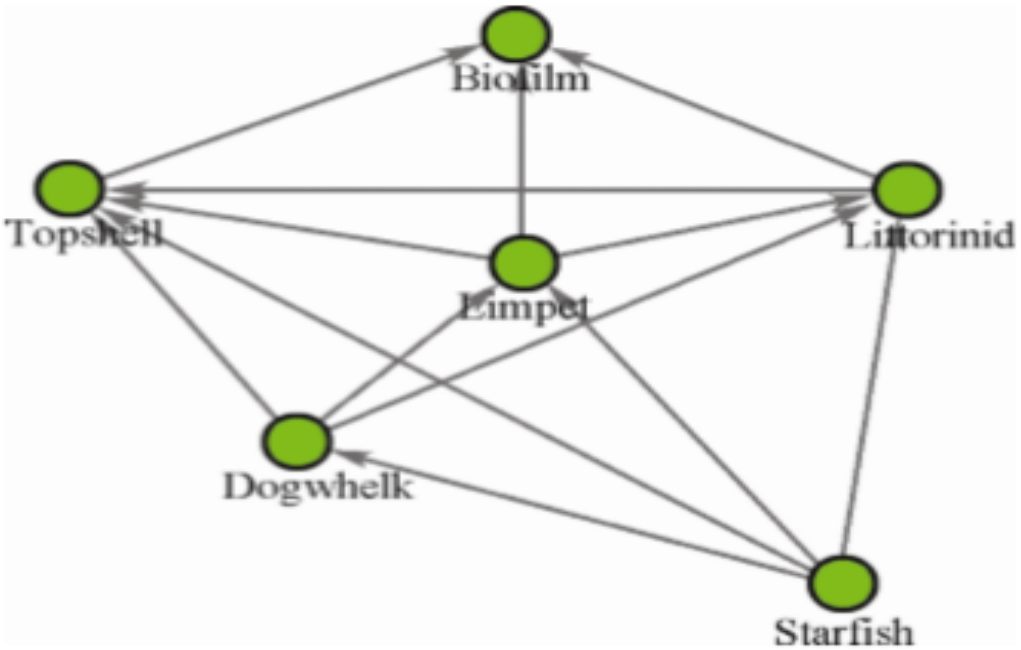
Ecological resilience




Water
temperature
increase
State 1 (equilibrium)
State 2 (new equilibrium)
Threshold

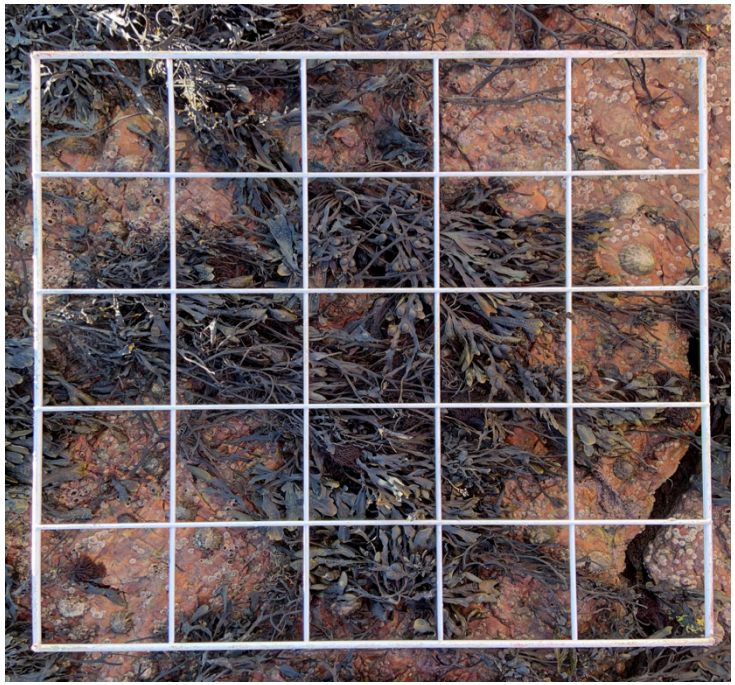
Rocky Shore Ecology

Bayesian Networks

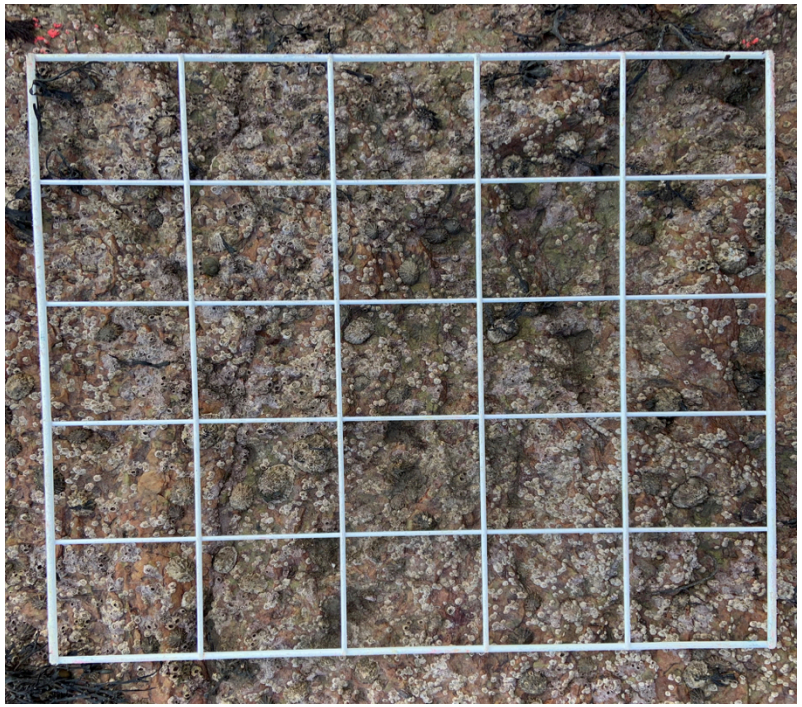


- Conditional dependencies translate well into ecological relationships within a community

Control

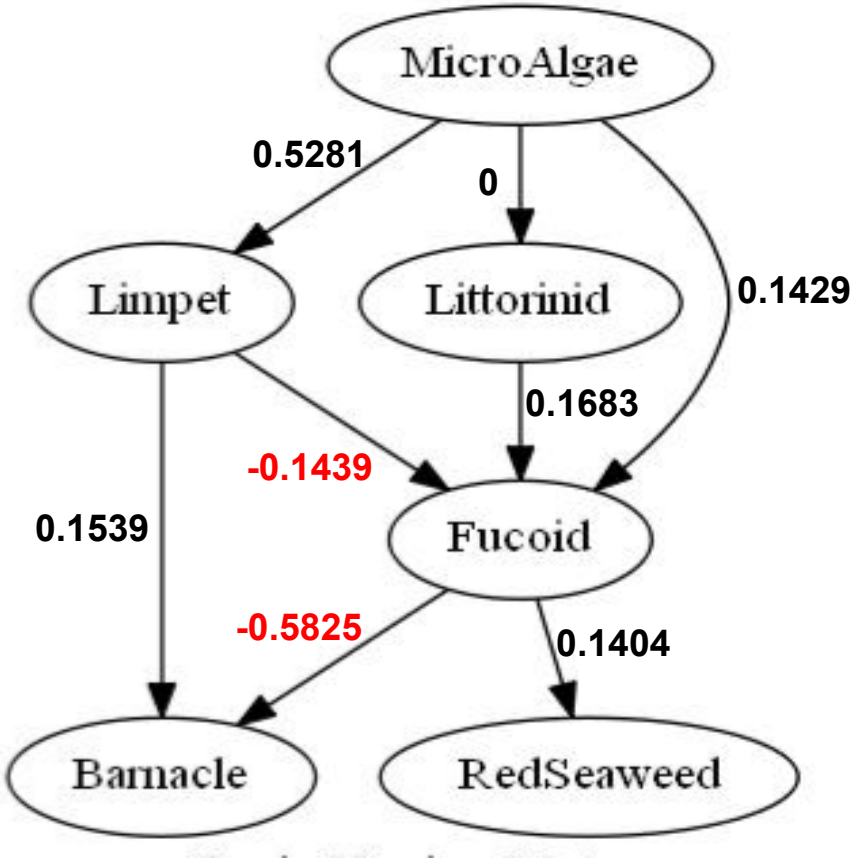


Fucoid removal



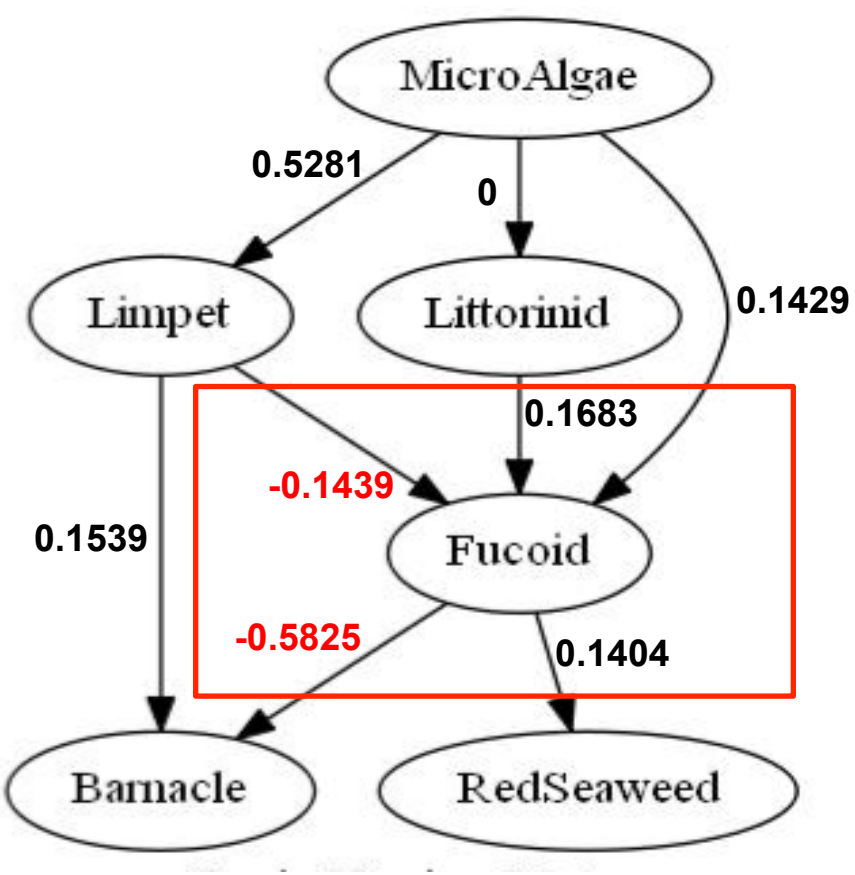
AIM: Reveal dynamic bayesian network of post disturbance changes to community structure

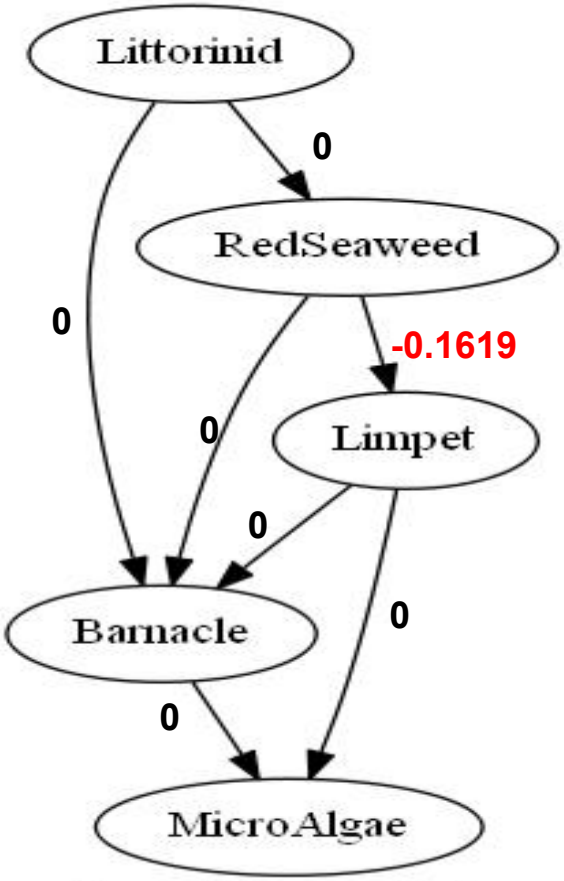
Control network



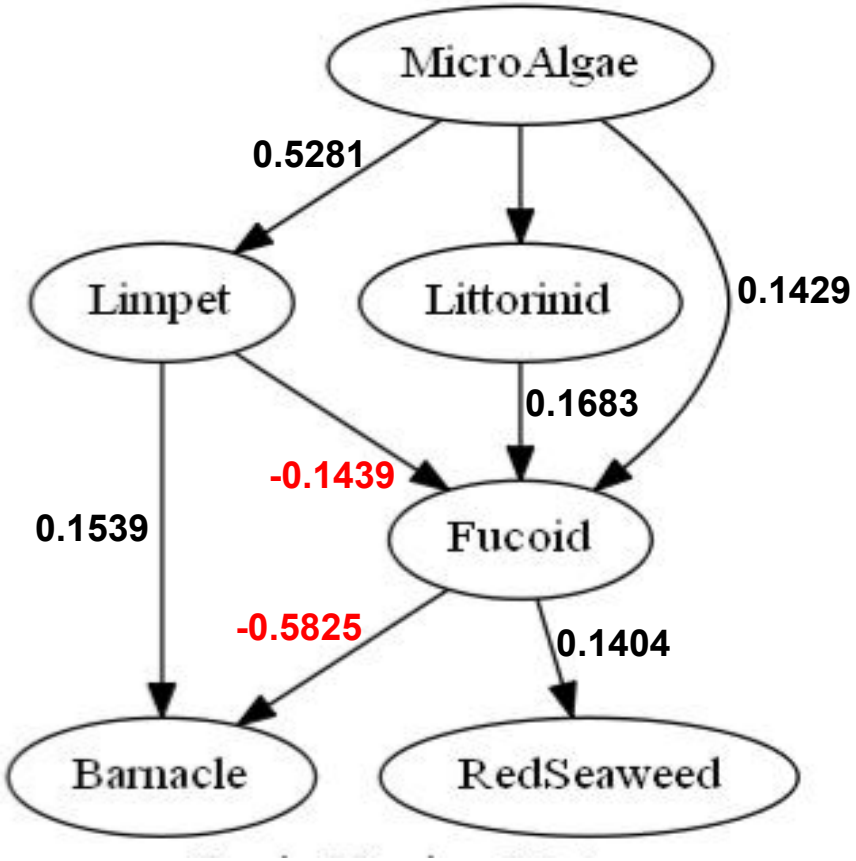
- Fucoid vs Barnacle competitive interactions
- Grazing relationships between Limpet and Littorinid and Microalgae
- All consistent with what we would expect from literature

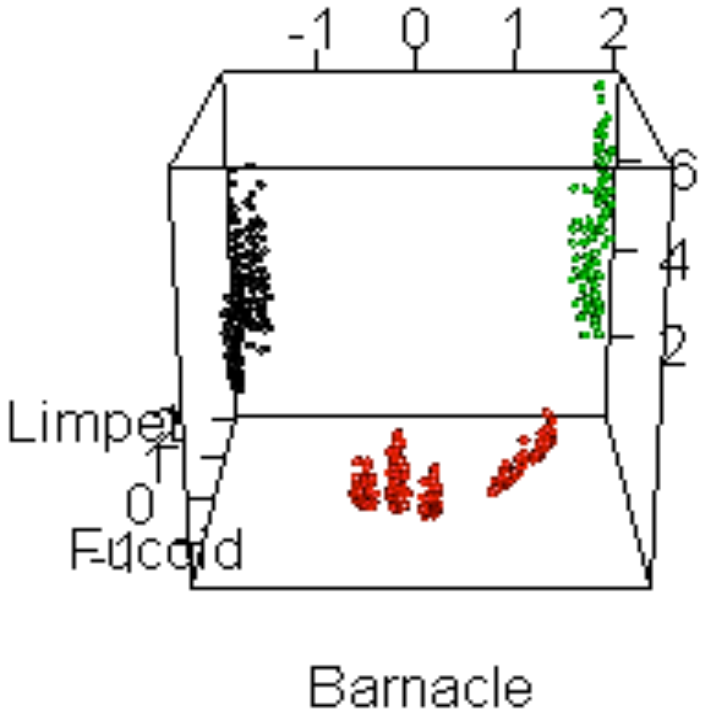
Control VS Fucoid removal





Blueprints for future study?





Acknowledgements

Gene Networks:
Jing Yu (PhD Duke)
Alexander J Hartemink (Duke)
Erich D Jarvis (Duke)
Amit Arora (Postdoc)
Franziska Matthäus (Goethe)
Simon Langdon (Edinburgh)
David Harrison (St Andrews)
Antonis Koussounadis (Postdoc)
Jessie Kennedy (Napier)
Daniel Archambault (Swansea)
Athanasios Vogogias (PhD Napier)
Hannah Currant (BSc)

Neuronal Networks:
Jing Yu (PhD Duke)
Alexander J Hartemink (Duke)
Erich D Jarvis (Duke)
Shih-Chieh Lin (Duke)
Luis Baccalá (São Paulo)
Koishi Sameshima (São Paulo)
Tom V Smulders (Newcastle)
Evelyne Sernagor (Newcaslte)
Christoph Feenders né Echtermeyer (PhD)
Stefan Pulver (St Andrews)
Alistair Birse-Stewart-Bell (BSc)
Staśa Tumpa (BSc)
Francis Jacob (PhD)

Ecological Networks:
Isobel Milns (BSc)
Colin Beale (York)
Dirk Husmeier (Glasgow)
Jack Lennon (James Hutton)
Andrej Aderhold (PhD)
Emily King (MRes)
Margaret Wallace (Postdoc)
Andrew Brierley (St Andrews)
Richard Stafford (Bournemouth)
Iain Matthews (St Andrews)
Edwin Hui (MSc-Res)

Funding:
NIH, St Andrews School of Biology, Royal Society, SULSA, BBSRC, EPSRC, NERC, MRS

V Anne Smith *http://biology.st-andrews.ac.uk/vannesmithlab/*