

# Networks in Biology: Part II

**Des Higham**

Department of Mathematics and Statistics  
University of Strathclyde

**SULSA event at IAS, Nov 2009**

# Background

Joel E. Cohen

**Mathematics is biology's next microscope, only better**

# Background

Joel E. Cohen

**Mathematics is biology's next microscope, only better**

Quantitative/computational work in biology may be **data driven** or may arise through **modelling**.

**Model**

**Quantitative, simplified description of a natural system**

Useful for

- **testing/comparing hypotheses**
- **making predictions**

This talk will focus on **networks**: **extracting useful information** and **modelling**.

# Typical Tasks

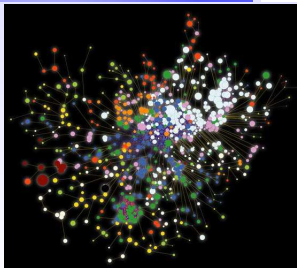
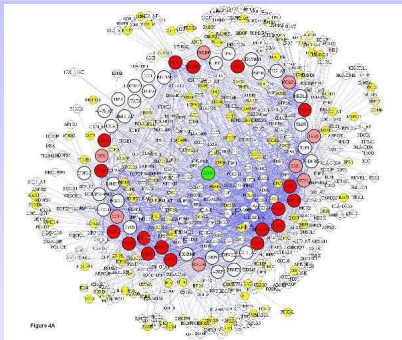
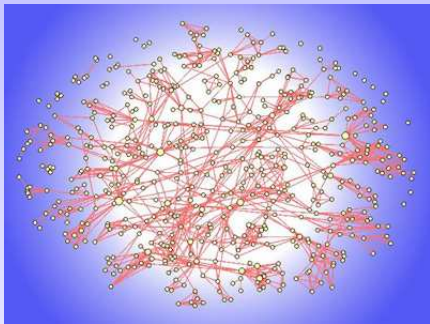
## Data Driven:

- find one or more well-connected **clusters**
- find specific connectivity **substructures**
- find '**important**' nodes or links
- **Compare** the properties of one network with another
- Distinguish between **two or more types of edges**

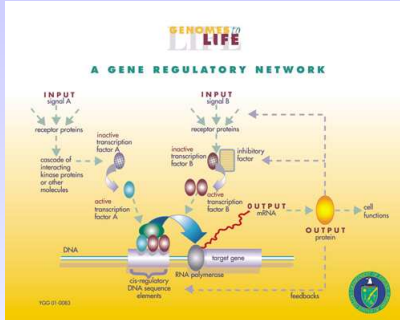
## Via Modelling Arguments:

- summarize a network in terms of a few **parameters**
- **explain** how the connectivity has arisen
- discover **missing** or **spurious** links
- predict the **future growth** of the network

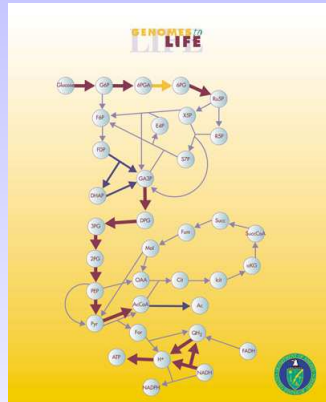
# Network Science: Connections are important



# Not doing ... Dynamics on a network



Gene regulatory



Metabolic

# Recall from Part I

**Network/Graph:**

**Adjacency Matrix:**

# Bi-clustering of Microarray Data

$M$  **genes** (e.g.  $M \approx 20,000$ )

$N$  **samples** (e.g.  $N \approx 20$ )

$w_{ij} \geq 0$  represents the activity of **gene**  $i$  in **sample**  $j$

**Bi-clustering**: partition

- **genes** into groups **A** and **B**
- **samples** into groups **A** and **B**

so that

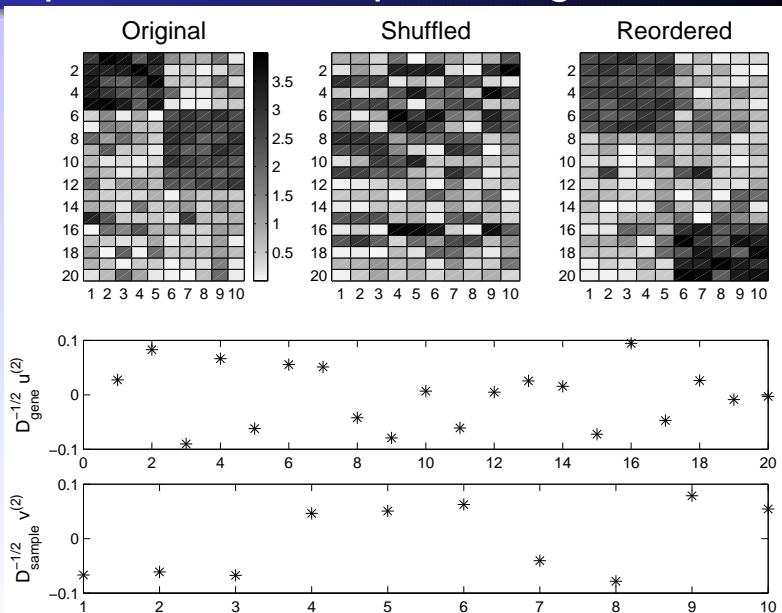
- **genes** in group **A** tend to be active in **sample** group **A** and inactive in **sample** group **B**
- **genes** in group **B** tend to be active in **sample** group **B** and inactive in **sample** group **A**

# Bi-clustering of Microarray Data

Biological motivation : **genes** involved in a common function should be active in a common set of **samples**

Justified on existing microarray datasets by Kluger at al., Genome Res., 2003, using an SVD approach

# Computational Example: using the SVD



# Two-signed Microarray Data

Some datasets have pos. and neg. data

- $w_{ij} > 0$  indicates **gene  $i$**  over-expressed in **sample  $j$**
- $w_{ij} < 0$  indicates **gene  $i$**  under-expressed in **sample  $j$**

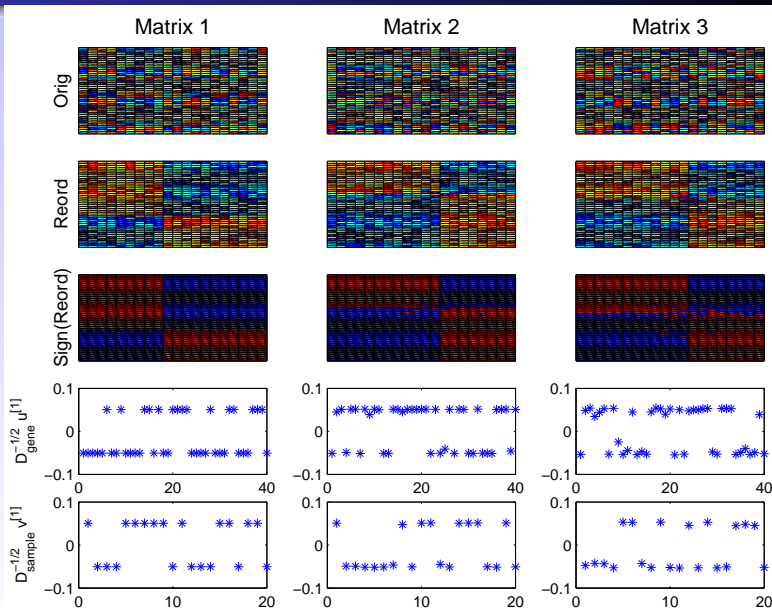
**Bi-clustering:** partition

- **genes** into groups **A** and **B**
- **samples** into groups **A** and **B**

so that

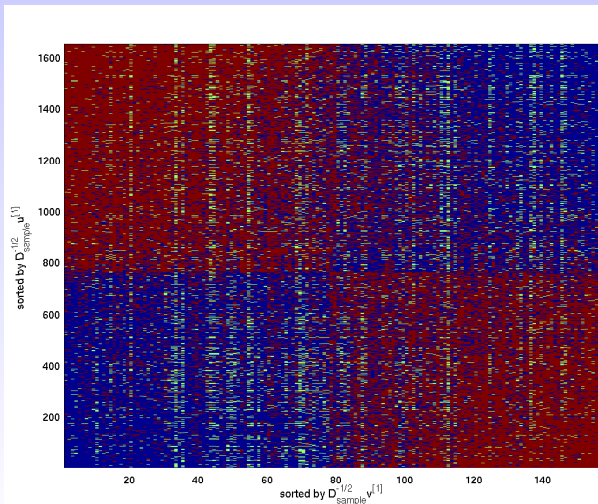
- **genes** in group **A** tend to be over-expressed in **sample** group **A** and underexpressed in **sample** group **B**
- **genes** in group **B** tend to be over-expressed in **sample** group **B** and underexpressed in **sample** group **A**

# Recovering a Checkerboard



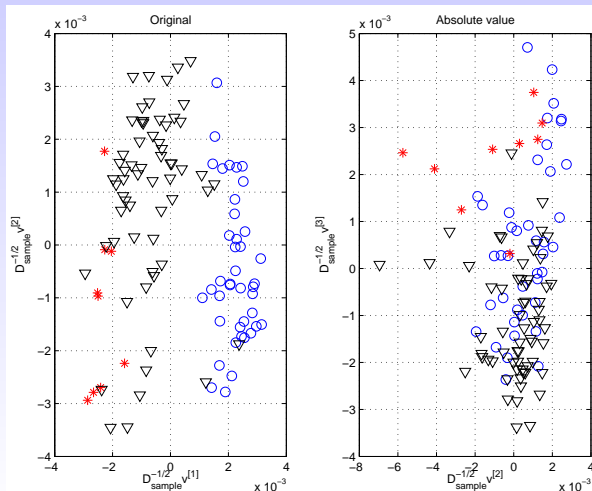
# Liver Cancer Data: Spectrally Reordered

**Data:** Chen et al. *Gene expression patterns in human liver cancers*, *Mol. Biol. Cell*, 2002. 1648 **genes**, 156 **samples**



# Prostate Cancer Data: Bi-clustered

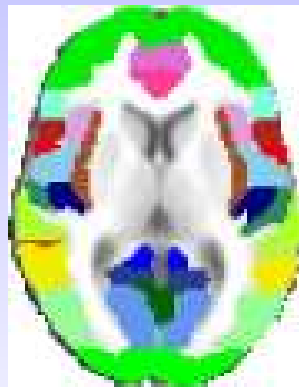
**Data:** Lapointe et al. *Gene expression profiling identifies clinically relevant subtypes of prostate cancer*, **PNAS**, 2004.  
5153 **genes**, 112 **samples**



# Anatomical Connectivity in the Human Brain

Crofts & Higham, *J. Roy. Soc. Interface*, 2009

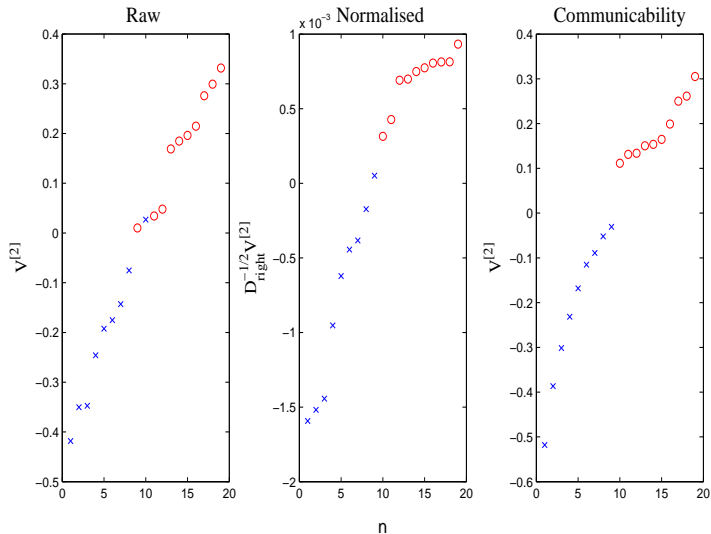
- 9 subjects - at least 6 months following first, left-hemisphere, subcortical stroke; and 10 age matched controls.
- **Diffusion Tensor Imaging** computes all connections between all voxels
- Connectivity network based on the **Harvard-Oxford cortical and subcortical structural atlas: 48 cortical regions and 8 subcortical regions.**



<http://www.fmrib.ox.ac.uk/fsl/fslview/atlas-descriptions.html>

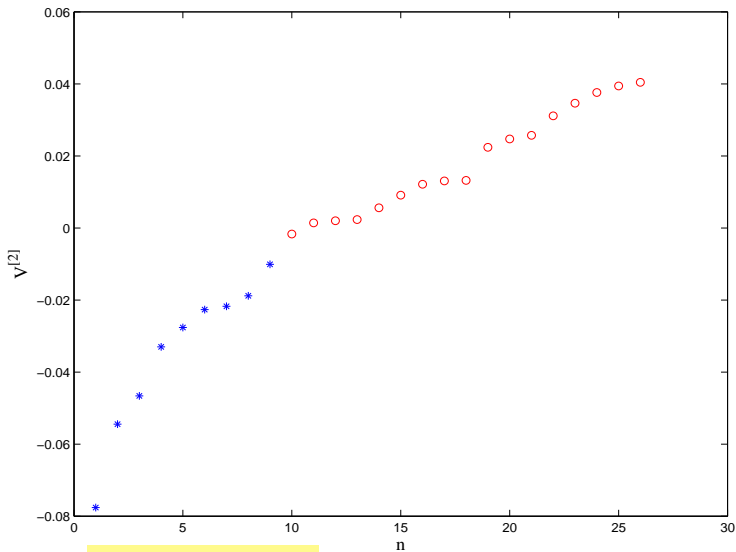
Methodology from FMRIB Oxford: **Heidi Johanson-Berg, Tim Behrens, Saad Jbabdi.**  
Stroke data supplied by **Rose Bosnell.**

# Unsupervised Clustering of Patients



New **weighted communicability** measure gives a statistically significant improvement.

# Connections within the RHS only:



Evidence for compensatory rewiring of RHS for stroke victims.

# Random Graph Models

**Erdos-Renyi/Gilbert**

**Watt-Strogatz/Small World**

**Geometric**

**Range-Dependent**

**Stickiness**

# Random Graph Models

**Erdos-Renyi/Gilbert**

**Watt-Strogatz/Small World**

**Geometric**

**Range-Dependent**

**Stickiness**

**Uses?**

- understand how the network evolved into its current state
- predict future changes
- summarize in terms of a few parameters
- compare different networks
- identify missing and spurious links

Lots of challenging computational issues.

# Y2H Protein-Protein Interaction Networks

**Noisy:** 50–90% false positive, 50–90% false negative

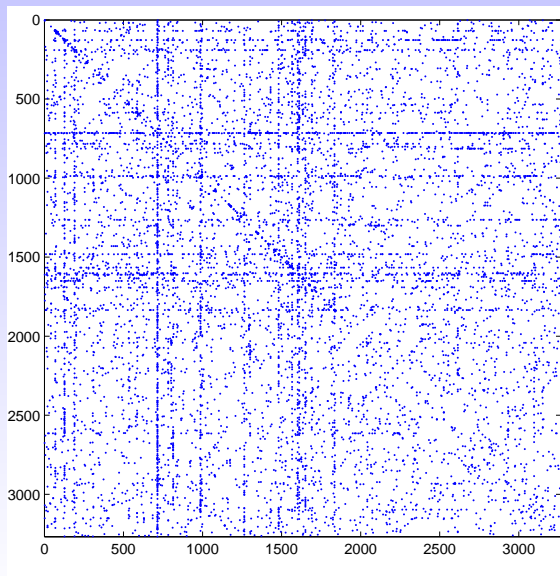
## Two types of false positive

- **Technical:** experimental limitations
- **Biological:** don't occur in vivo
  - not expressed at same time
  - not in same sub-cellular compartment, or same tissue

Interactions may also depend on the **environment**

How can we use this data . . . . . ?

# Adjacency Matrix: Ito et al. 2001, Yeast PPI



## Typical questions:

- Are there any other proteins like protein Y?
- What is the biological function of protein X?
- Which proteins act together?
- What happens if protein Z is removed?

Also: which are the false pos/negs ?

# Big picture...

PPI networks are not regular

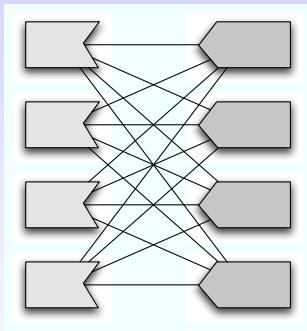
Describe them by a random graph model?

- capture many PPI networks with a **small number of parameters**:
  - **distinguish between different organisms**
  - **get evolutionary insights**
- generate **synthetic data sets** to test algorithms

# Lock-and-Key Model

J. L. Morrison, R. Breitling, D. J. Higham and D. R. Gilbert, **Bioinformatics**, 2006

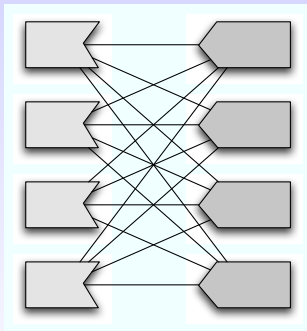
**Idea:** two proteins interact because they 'fit together'  
⇒ complementary domains, i.e. **locks** and **keys**



# Lock-and-Key Model

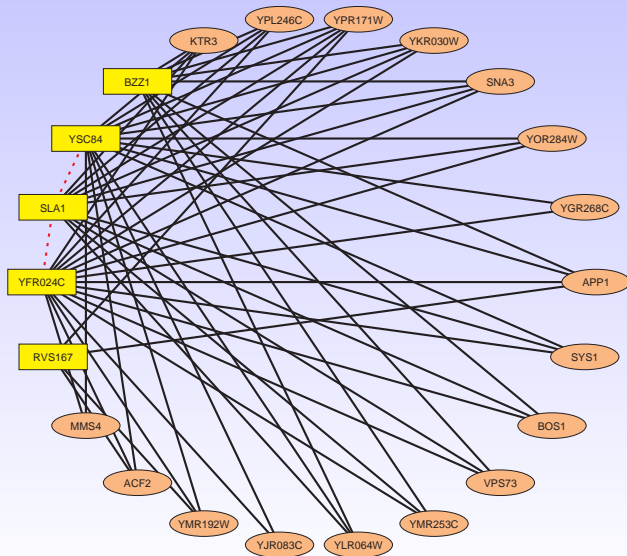
J. L. Morrison, R. Breitling, D. J. Higham and D. R. Gilbert, **Bioinformatics**, 2006

**Idea:** two proteins interact because they 'fit together'  
⇒ complementary domains, i.e. **locks** and **keys**



Develop an **algorithm** to discover locks and keys . . .

# Result for Uetz et al. (2000) yeast data



# Further Investigation . . .

Other biological data shows that  
all five proteins in one group possess the SH3 domain

⇒ we have identified the key!

Recent experiments (Kessels & Qualmann 2004, Friesen et al. 2005) show that the SH3 domain is involved in trafficking of **vesicles**

All proteins in the other group are part of the actin cortical patch assembly mechanism of **vesicle** endocytosis (Drees et al. 2001)

[**vesicle**: *small, enclosed compartment within a cell*]

# Some Resources

## Articles/Books:

- *Complex networks and simple models in biology*, de Silva & Stumpf, R. Soc. Interface, 2005,
- *Statistical mechanics of complex networks*, Albert & Barabási, Rev. Mod. Phys., 2002
- *An Introduction to Systems Biology*, Alon, Chapman & Hall, 2006

## Software Tools: (find them via Google)

- *GraphCrunch*
- *UCINET*
- *Pajek*
- *CONTEST*
- *NESSIE*