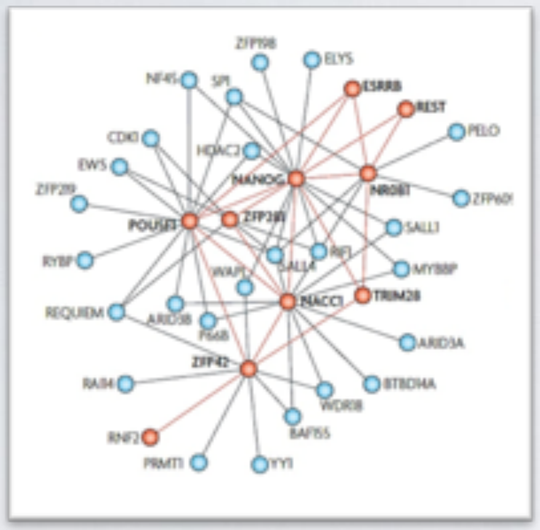
GEOMETRY AND TOPOLOGY OF NETWORKS AND DATA

Ruben Sanchez-Garcia
Mathematical Sciences
University of Southampton

Applied Algebraic Topology Network
Durham
24 September 2016

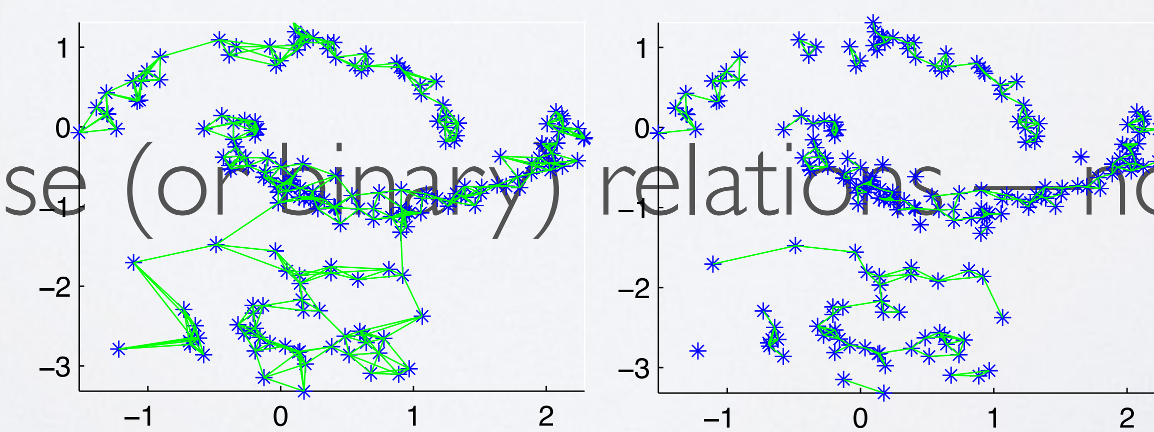
Network models

- Representation of a complex system using a network (graph)



Images: (L) MacArthur et al, Nature Reviews Molecular Cell Biology; (C) FEMA; (R) Flickr by Marc_Smith

- Networks, usually weighted, can also be used to represent *data sets*

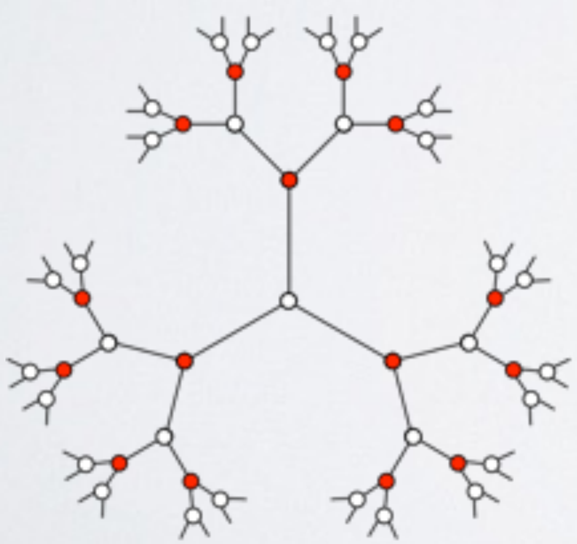


- Limitation: pairwise (or binary) relations — no higher-order structures

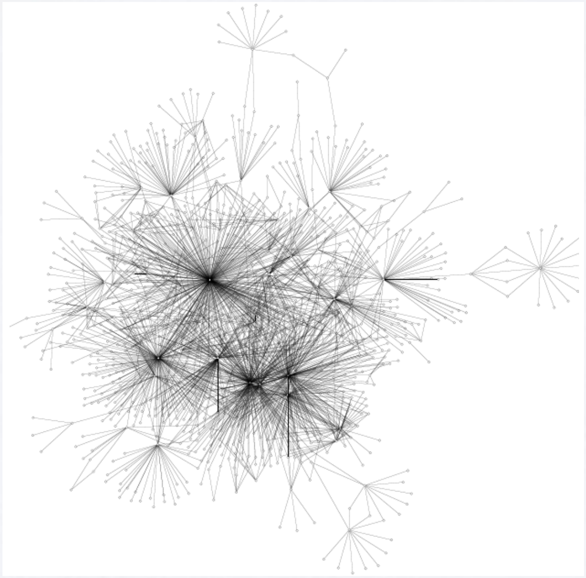
Symmetry in Complex Networks

Question: Are real-world networks symmetric?

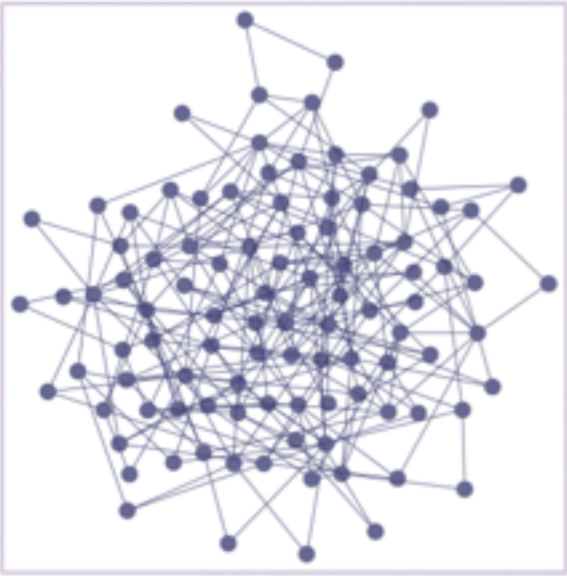
- **Aut(G)** permutations of the vertices preserving adjacency
- Symmetries relate to redundancy (structurally equivalent vertices)
  → system robustness, evolution from basic principles



regular graph



complex network



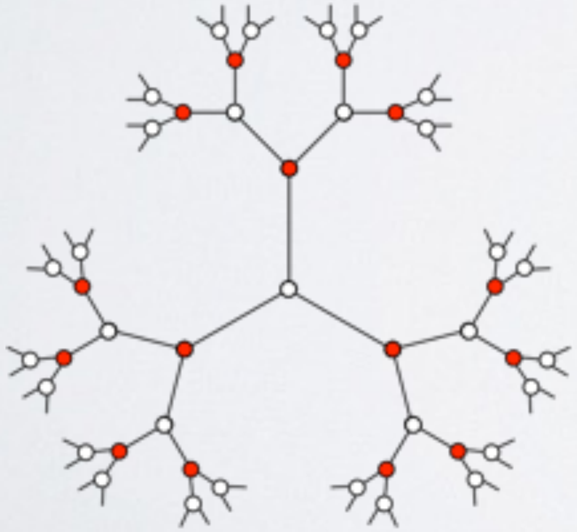
random graph

Symmetry in Complex Networks

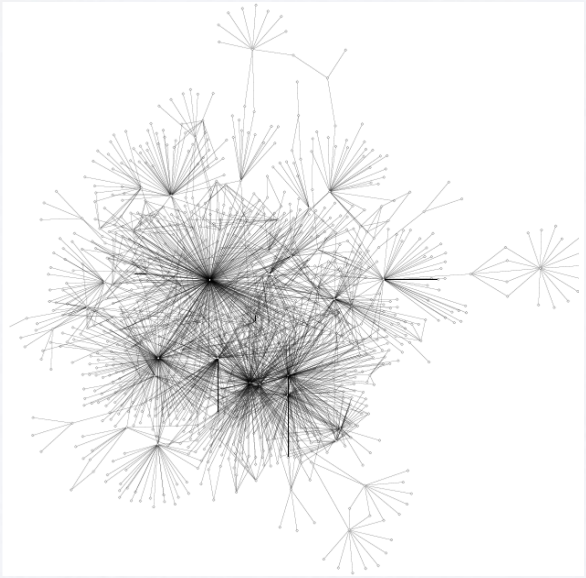
Question. Are real-world networks symmetric?

Regular graphs	large group of (global) symmetries
Random graphs	trivial group of symmetries
Complex networks	large (in absolute terms*) but localised

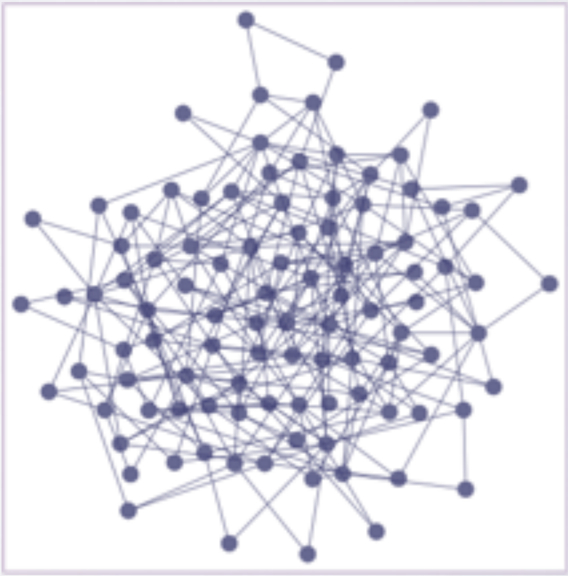
*10^9 - 10^{11298} in the networks studied



regular graph

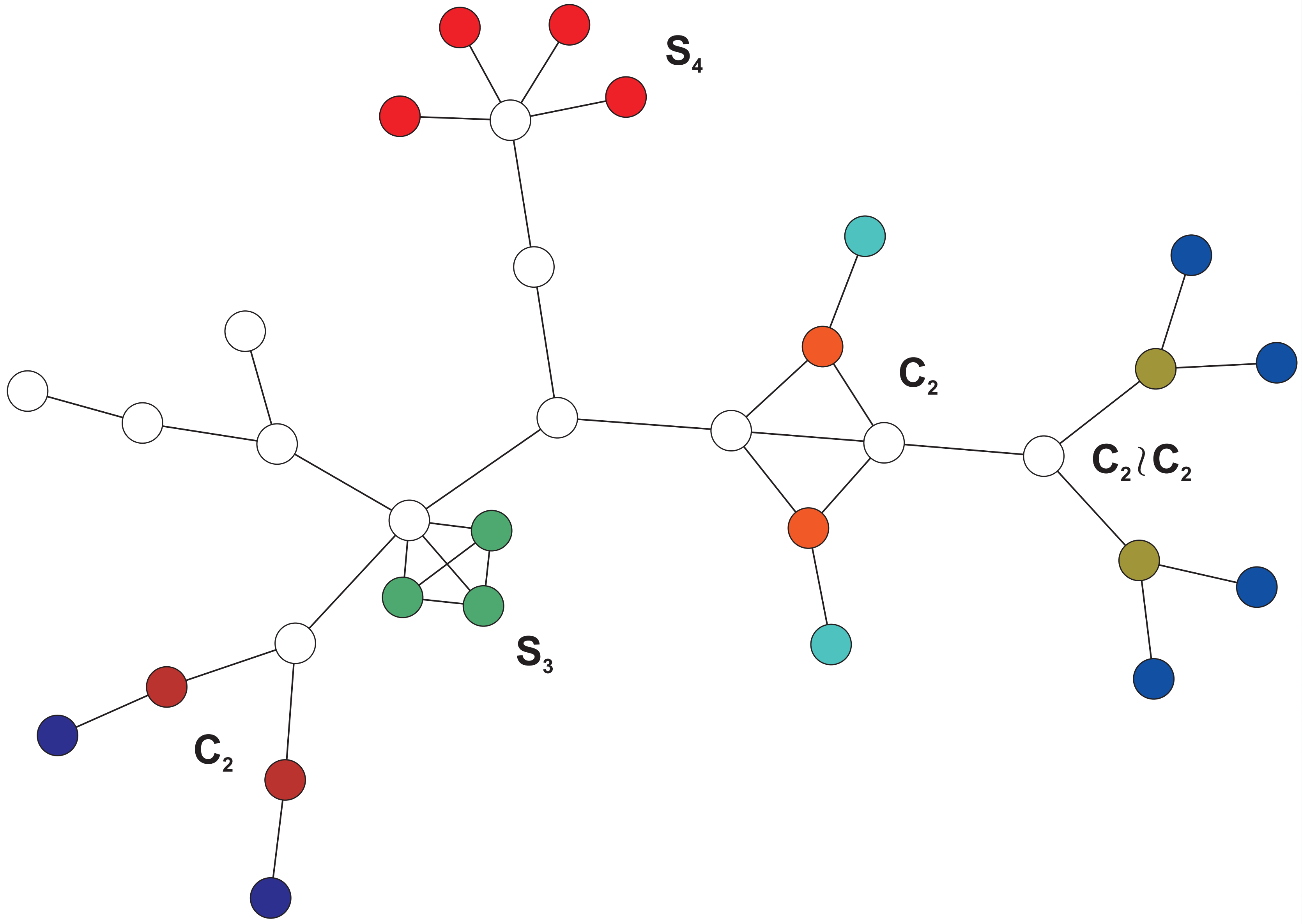


complex network



random graph

Toy example



$$
\begin{aligned}
\mathrm{Sym}(\mathcal{G}) &= S_2 \times S_3 \times S_4 \times S_2 \times (S_2 \wr S_2) \\
|\mathrm{Sym}(\mathcal{G})| &= 4608
\end{aligned}
$$

Real world examples

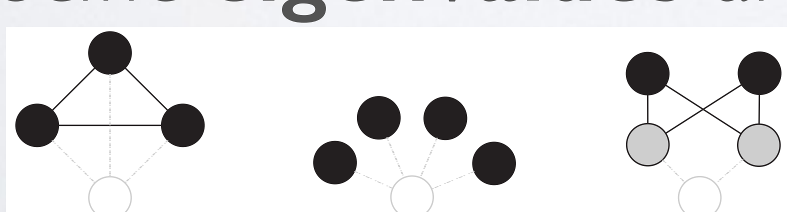
Yeast $S_2^{90} \times S_3^{26} \times S_4^{16} \times S_5^{8} \times S_6^{6} \times S_7^{5} \times S_8^{2} \times S_9^{2} \times S_{10}^{3} \times S_{11}^{2} \times S_{12} \times S_{13} \times S_{46} \times (S_2 \wr S_2)$

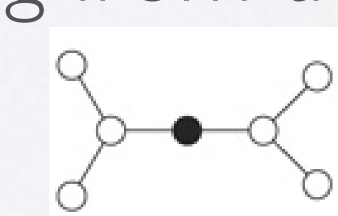
PhD $S_2^{43} \times S_3^{27} \times S_4^{16} \times S_5^{11} \times S_6^{10} \times S_7^{4} \times S_8^{5} \times S_9^{6} \times S_{10} \times S_{11}^{3} \times S_{12}^{2} \times S_{13}^{2} \times S_{35} \times (S_2 \wr S_2)^3 \times (S_5 \wr S_2)$

[**Computation:** *nauty* algorithm (McKay'81), GAP, decomposition theorem]

We found:

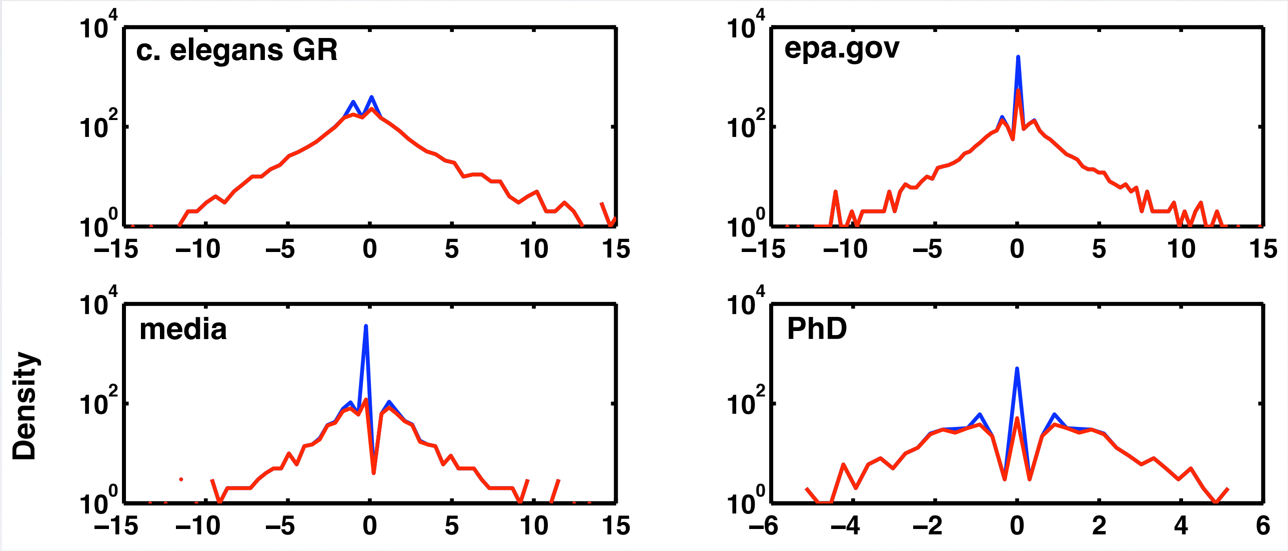
- **Aut(G)** direct product of symmetric groups and wreath products of symmetric groups (*tree-like*)
- Most factors (> 90%) S_n acting naturally on $k \geqq 1$ orbits (*basic symmetric motifs*) with a very constrained structure (Liebeck'88)
- Identify specific **eigenvalues** arising from the symmetry





Symmetric spectrum

- We studied how symmetries (automorphisms) affect network spectrum
- Symmetries give rise to high-multiplicity eigenvalues (peaks in spectral density)



- The network spectrum is the union of the *redundant* spectrum of the symmetric motifs, and the spectrum of the quotient network
- The redundant spectrum of the *basic* symmetric motifs is very constrained e.g.

$$\mathrm{RSpec}_1 = \{-1, 0\} \quad \mathrm{RSpec}_2 = \{-2, -\varphi, -1, 0, \varphi - 1, 1\}$$

$$\mathrm{RSpec}_3 = \{-3, -2, -1, 0, 1, \pm\sqrt{2}, \pm\sqrt{3}, -1 \pm \sqrt{2}, -1 \pm \sqrt{3}, \ldots\}$$

MacArthur, Sanchez-Garcia, Anderson *Symmetry in Complex Networks* **Discrete Appl. Math.** (2008)
MacArthur, Sanchez-Garcia *Spectral characteristics of network redundancy* **Phys. Rev. E** (2009)

Adaptive Networks

- Network topology and dynamics influence each other

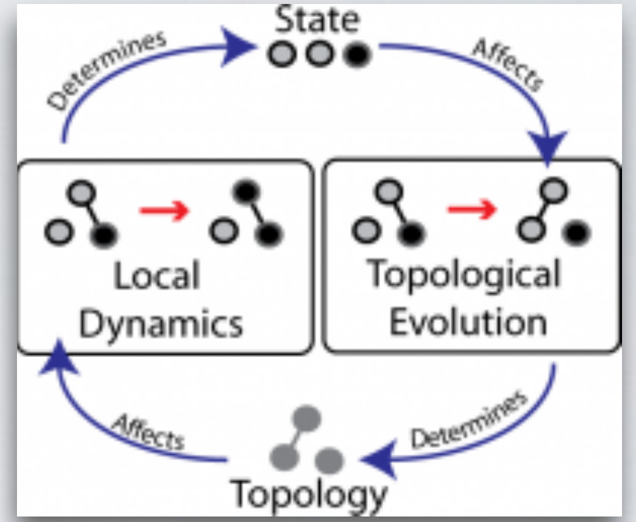
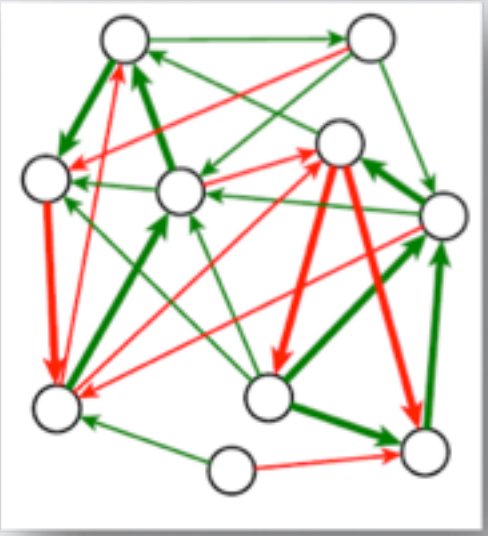


Image: www.biond.org

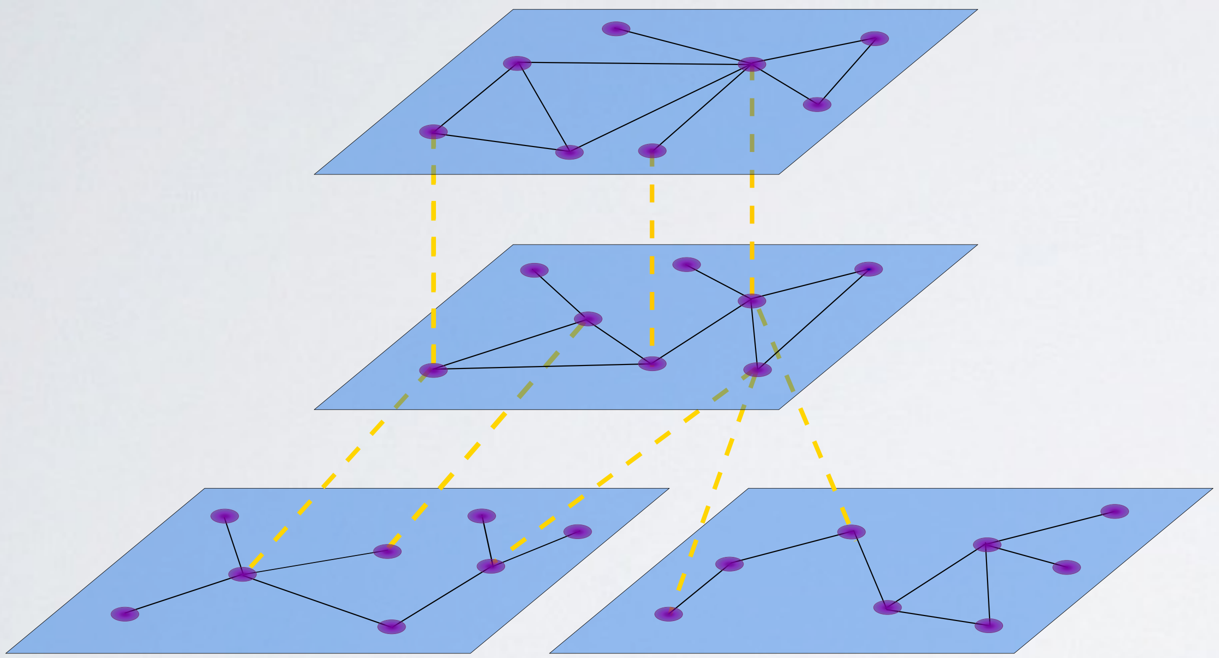
- We studied a biologically motivated model of an adaptive regulatory network
  - The system self-organises to a critical state
  - We analytically related stability to cycle structure (via graph eigenvalues and Rouché's theorem in complex analysis)

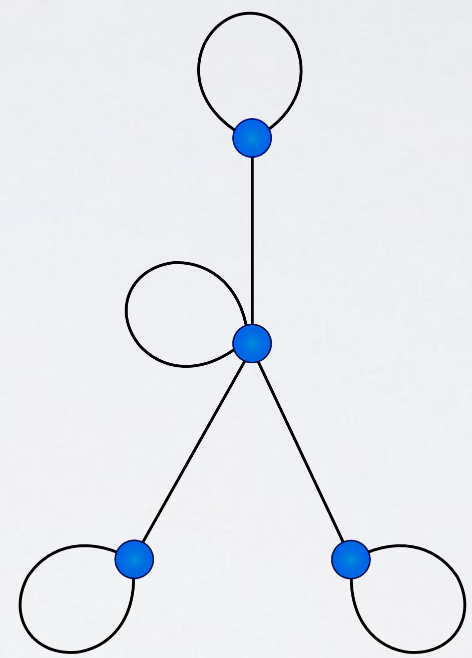


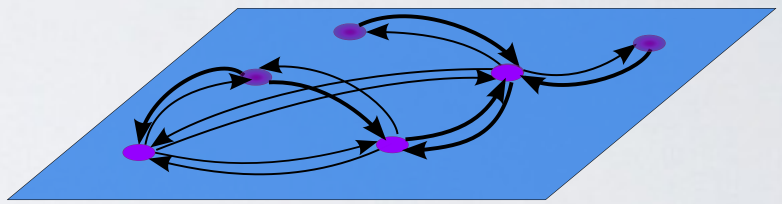
MacArthur, Sanchez-Garcia, Ma'ayan *Microdynamics and criticality of adaptive regulatory networks* **Phys. Rev. Lett.** (2010)

Multilayer Networks

- We studied two natural quotients associated to a *multilayer network*:







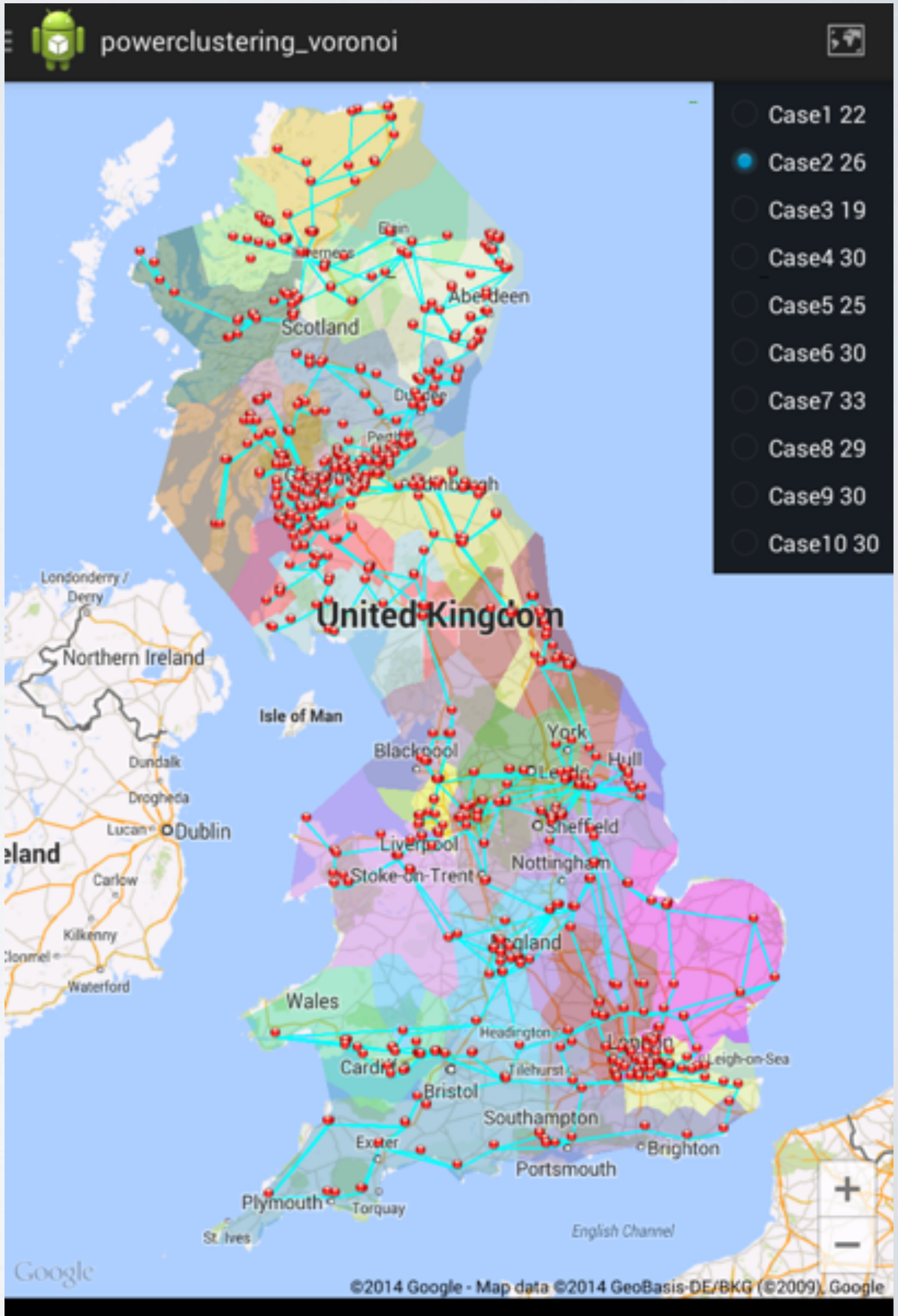
- The eigenvalues of the quotient *interlace* those of the parent graph:

$$\lambda_i \leq \mu_i \leq \lambda_{i+(n-m)}$$

Sanchez-Garcia, Cozzo, Moreno *Dimensionality reduction and spectral properties of multilayer network* **Phys. Rev. E** (2014)

UK Power Grid

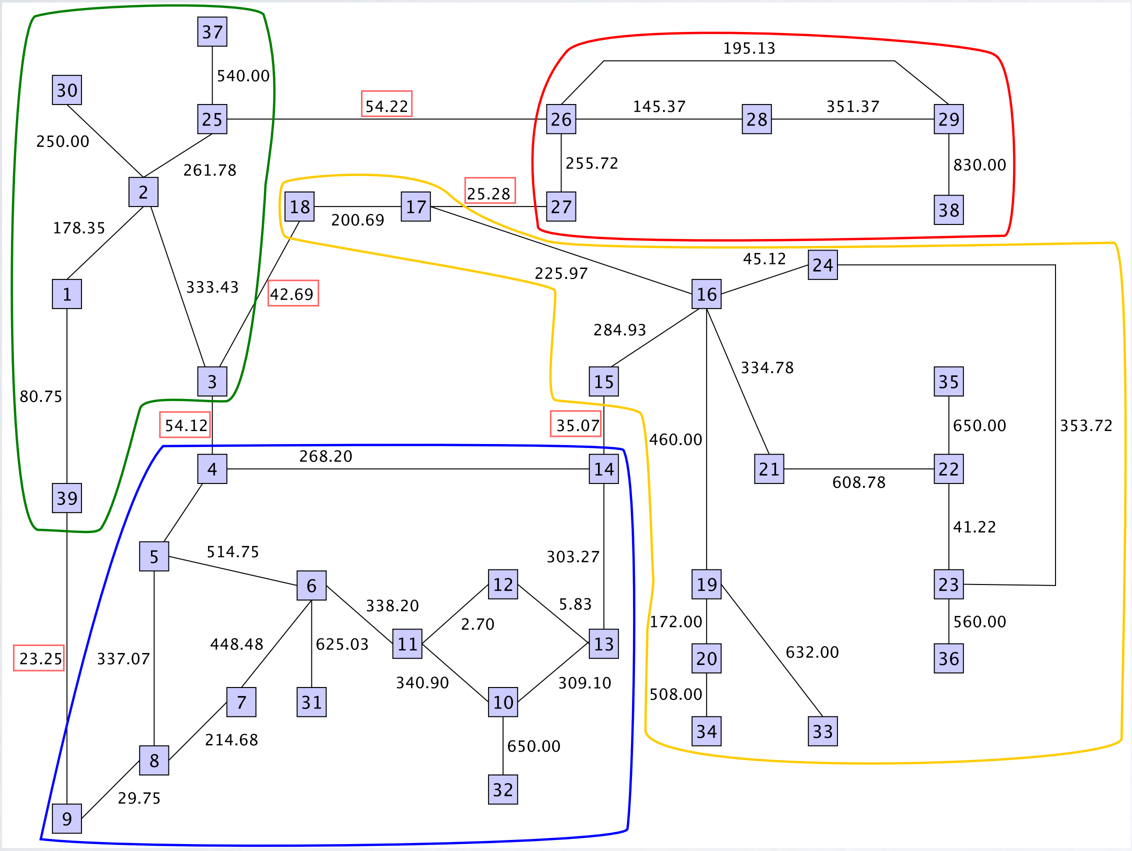
- EPSRC-funded project *Preventing wide-area blackouts through adaptive islanding* (2010-14) PI: Jacek Brodzki
- Collaboration between power engineers and mathematicians
- Network modelling approach
- Graph Laplacian as main analytical tool (spectral clustering)
- Android App (Yuki Ikuno)



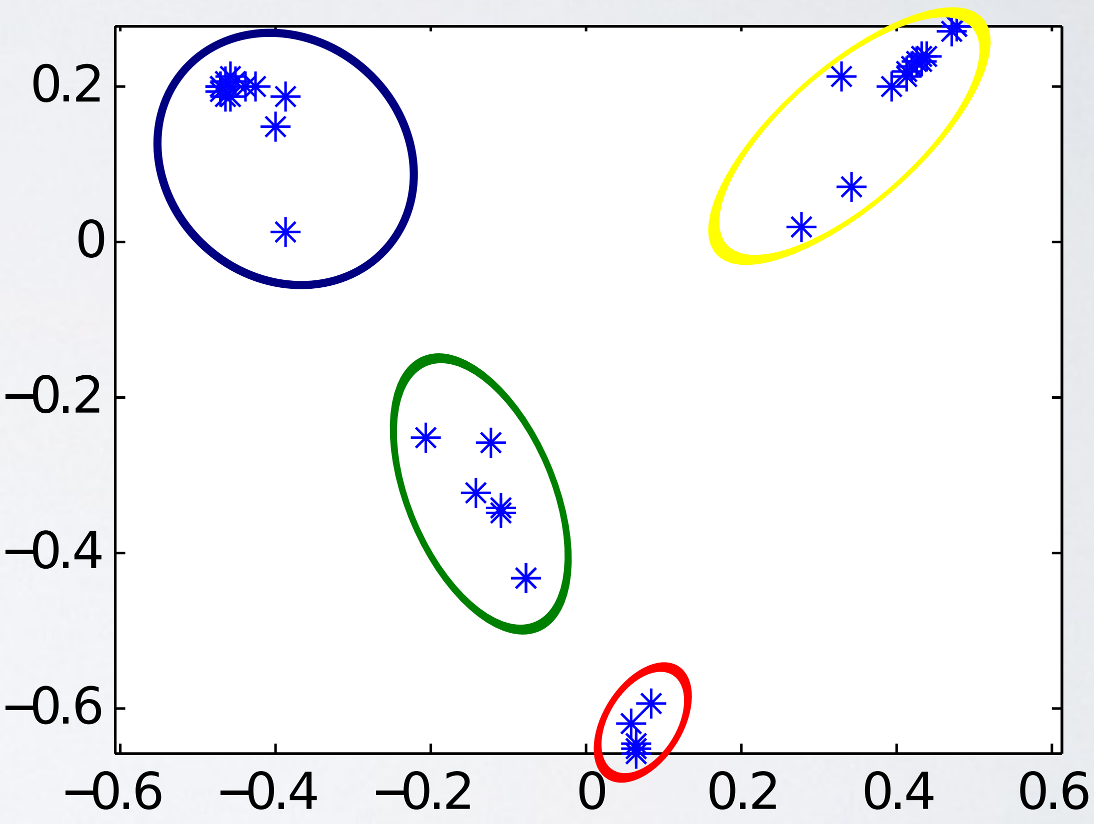
Sanchez-Garcia et al *Hierarchical spectral clustering of power grids* **IEEE Transactions of Power Systems** (2014)

Spectral Clustering

- The Laplacian eigenvectors give geometric coordinates to the vertices (*spectral embedding*) revealing clusters

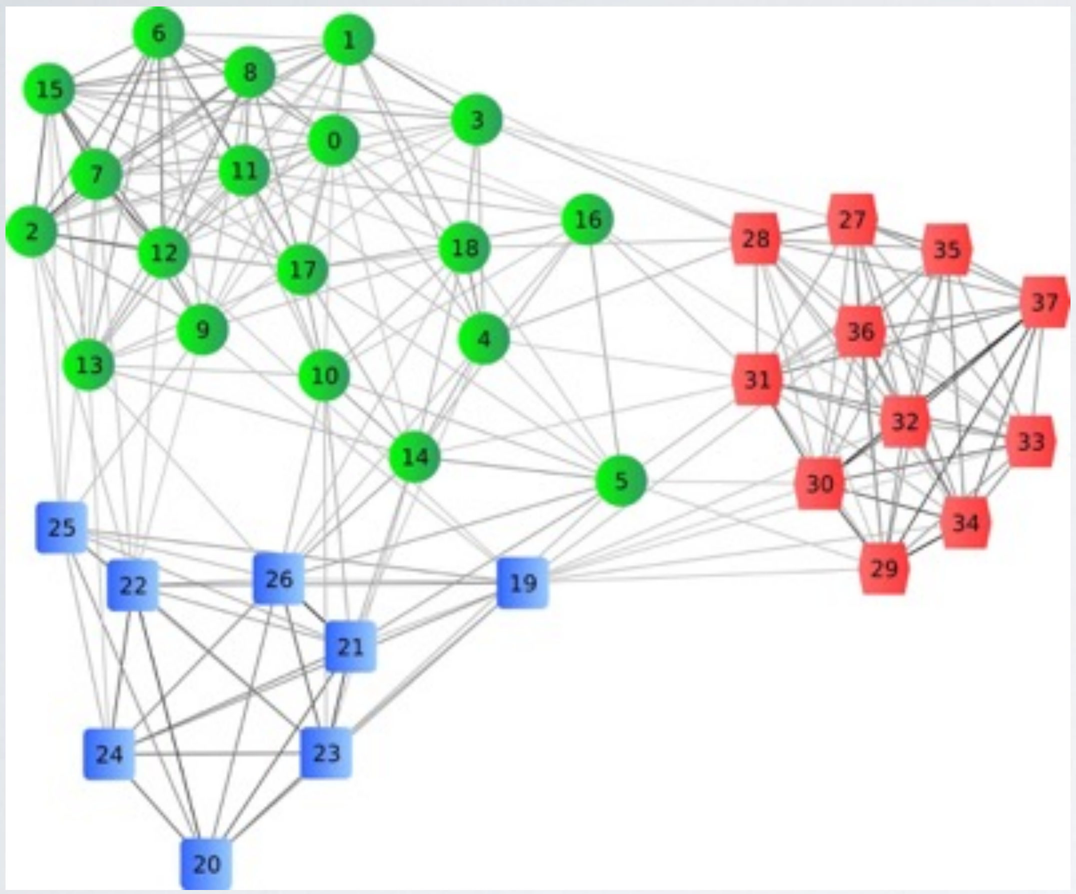


IEEE 39-bus test system



3-dimensional spectral embedding

Spectral Clustering



- **Cluster:** 'Almost connected component'
- The *isoperimetric ratio* of $\emptyset \neq S \subseteq V$ is

$$\phi(S) = \frac{\sum_{i \in S, j \notin S} a_{ij}}{\sum_{i,j \in S} a_{ij}} = \frac{\text{boundary of } S}{\text{volume of } S}$$

- The 'best' *k*-partition (minimising the worst ratio) is

$$h_G(k) = \min_{S_1,\ldots,S_k} \left(\max_{1 \leq i \leq k} \phi(S_i) \right)$$

the *k*-way Cheeger constant of the graph

The Laplacian spectrum $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ is closely related to clustering:

- $\lambda_k = 0$ iff the graph has (at least) k connected components
- Higher-order Cheeger inequalities (Lee et al, 2012) $\frac{\lambda_k}{2} \leq \phi_G(k) \leq Ck^2\sqrt{\lambda_k}$

Algebraic Graph Theory

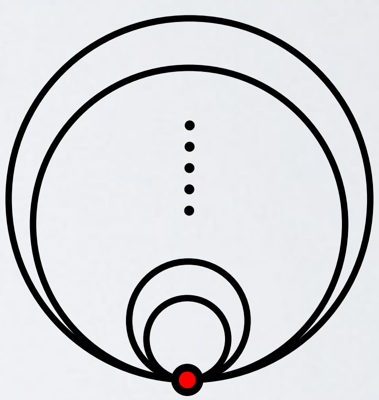
- **Adjacency matrix** $n_0 \times n_0$ matrix $[A]_{ij} = 1$ if *i* and *j* adjacent
- **Incidence matrix** $n_0 \times n_1$ matrix $[B]_{ik} = \pm 1$ if vertex *i* final/initial vertex of edge *k* (after orienting edges)
- **Laplacian matrix** $n_0 \times n_0$ matrix $L = BB^t = D - A$
- There are weighted and normalised versions of the Laplacian
- Eigenvalues & eigenvectors of these matrices reflect the *structure* and *dynamics* of a network

From networks to complexes

- A network is a 1-dimensional model of a complex system, or data set
- Networks generalise to higher-dimensional *topological complexes*
- There are adjacency, incidence, and Laplacian matrices, at each dimension
- Complexes can be studied combinatorially, topologically or geometrically
- Complexes can be constructed from a network (e.g. clique complex) or from a data set (e.g. point cloud)

Topological Data Analysis

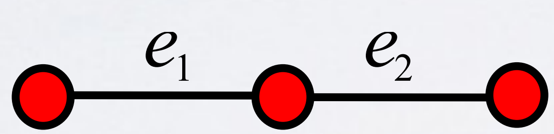
Data → Complex (filtration) → Topological features (persistent)

- **Towards Geometrical Data Analysis…**
  - Incorporate aspects of the geometry
  - Topologically every network is equivalent to 
  - A good candidate (motivated by network analysis): *Combinatorial (discrete, Hodge) Laplacian*

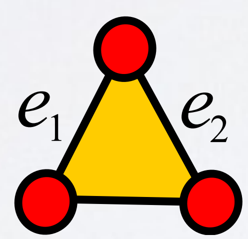
Adjacency matrices

- Complexes can be seen at three levels: combinatorial, topological or geometrical

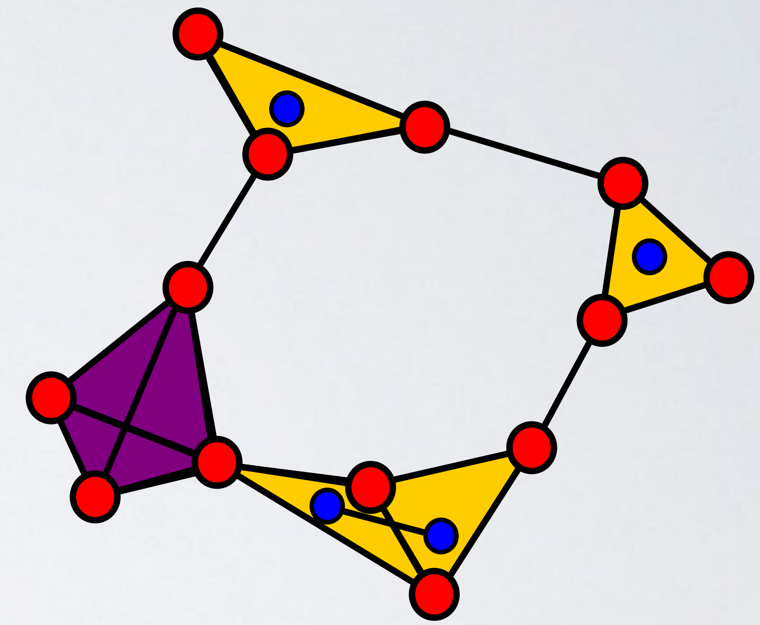
- Two k-simplices are **lower adjacent** if they intersect at a common (k-1)-simplex, and **upper adjacent** if they belong to the same (k+1)-simplex



lower



upper



- We have lower/upper $n_k \times n_k$ adjacency matrices A_k^{lower}, A_k^{upper} encoding the combinatorial structure at dimension k

- They can be seen as adjacency matrices of (k-dual) graphs, revealing the combinatorial structure at dimension k

Boundary matrices

- Organise topological information as a chain complex
- Algebraically: $C_{i+1} \xrightarrow{B_i} C_i \xrightarrow{B_{i-1}} C_{i-1}$ chain complex

  where C_i is the real vector space of dimension n_i the number of i-simplices, B_i is the boundary matrix
- B_i is a $n_i \times n_{i-1}$ matrix with entries $0, \pm 1$ given by the signs of the (i-1)-simplices in the boundary of the i-simplices
- Homology: $H_i = \ker(B_{i-1})/\mathrm{im}(B_i)$

Weighted complexes

- One way to encode (aspects of) the geometry is by using *weighted complexes:* associate a positive weight to each simplex (a choice of inner product $\langle e_i, e_j \rangle = \delta_{ij} w_i$)

- Algebraically:

$$\begin{tikzcd} C_{i+1} \arrow[loop above, "W_{i+1}"] \arrow[r, "B_i"] & C_i \arrow[loop above, "W_i"] \arrow[r, "B_{i-1}"] & C_{i-1} \arrow[loop above, "W_{i-1}"] \end{tikzcd}$$

where C_i is the real vector space of dimension n_i the number of i-simplices, B_i is the boundary matrix, and W_i is the diagonal matrix of weights

Laplacian matrices

- There are Laplacians at each dimension of a simplicial complex

$$\begin{tikzcd} C_{i+1} \arrow[r, "B_i"] \arrow[loop above, "W_{i+1}"] & C_i \arrow[r, "B_{i-1}"] \arrow[loop above, "W_i"] & C_{i-1} \arrow[loop above, "W_{i-1}"] \end{tikzcd}$$

$$L_i^{up} = W_i^{-1}\, B_i^T\, W_{i+1}\, B_i$$

$$L_i^{down} = B_{i-1} W_{i-1}^{-1} B_{i-1}^T W_i$$

$$L_i = L_i^{up} + L_i^{down}$$

  where C_i is the real vector space of dimension n_i the number of i-simplices, B_i is the boundary map, and W_i represents a choice of inner product

- The kernel of L_i is isomorphic to $H_i(X;\mathbb{R})$, a topological invariant

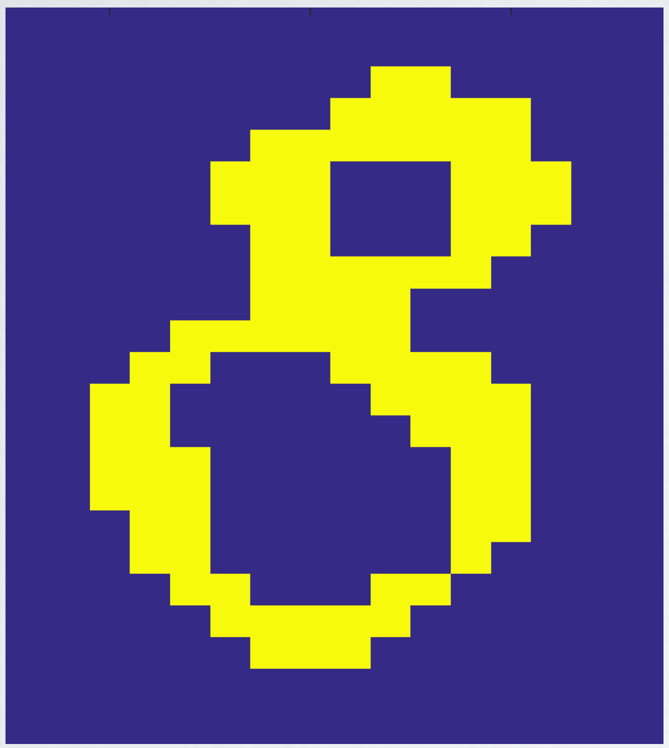
- the non-zero spectrum of L_i is the union of the non-zero spectrum of L^{up} and L^{down}, and it encodes the geometry with respect to the inner product

- This general framework developed by Horak & Jost (2013)
  "…we wish to propose this Laplacian spectrum as a new tool in data analysis."

[Horak, Jost *Spectra of combinatorial Laplace operators on simplicial complexes* **Adv. in Math.** 244 (2013)]

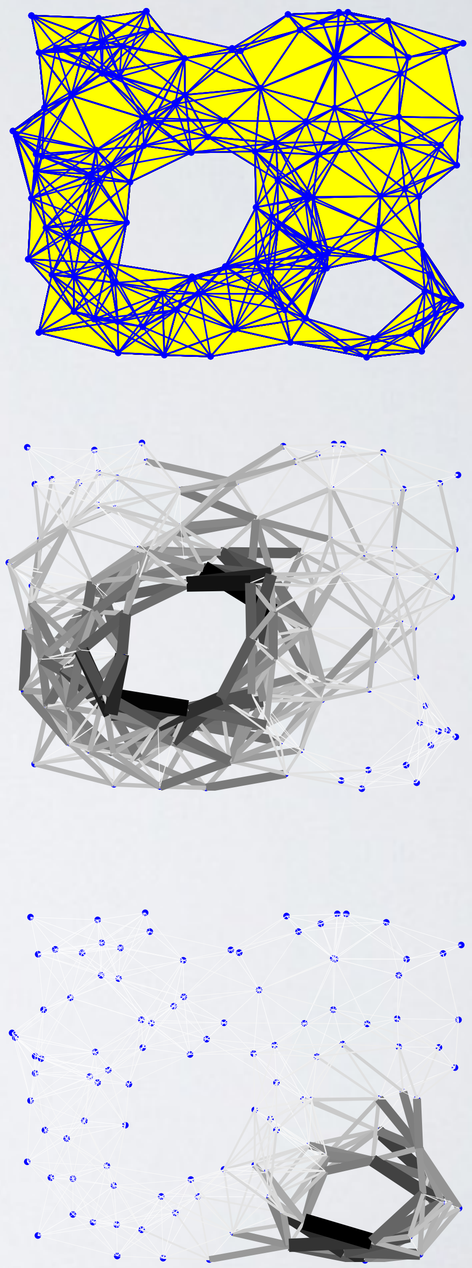
Examples

(Digit recognition) '8' encoded as simplicial complex
(92 vertices, 236 edges, 179 triangles)



$$L_i = L_i^{up} + L_i^{down}$$

$\mathrm{eig}(L_1)$	
0.0000	β_1
0.0000	β_1
0.0353	L_1^{down}
0.0920	L_1^{down}
0.1431	L_1^{down}
0.2589	L_1^{down}
0.3898	L_1^{down}
0.4810	L_1^{down}
0.5241	L_1^{up}
0.6236	L_1^{up}



Muhammad, Egerstedt, *Control Using Higher Order Laplacians in Network Topologies*

Application: Ranking in horse racing

- Current work by Conrad D'Souza (PhD Maths & Management)
  Co-supervisors: T. Ma, V. Sung and J. Johnson
- Topologically-inspired global ranking from pairwise comparison between alternatives [Jiang et al. *Statistical ranking and combinatorial Hodge theory* **Math. Progr.** 127 (2011)]
- Rank horses and predict race winner based on past performance data
- Particularly useful for *incomplete* and *inconsistent* data
- Represent horse racing data using a weighted 2-dim. simplicial complex:
  nodes = horses   edges weighted by pairwise preferences $\alpha_{ij} = -\alpha_{ji}$
  triangles weighted by boundary $\alpha_{ij} + \alpha_{jk} + \alpha_{ki}$(local inconsistencies)
- **Problem:** Find global ranking $r \in C^0$ with induced pairwise ranking $\delta_0(r)$ close to observed pairwise ranking $Y = (\alpha_{ij}) \in C^1$
- Minimum norm solution given by a Laplacian matrix $r = \Delta_0^\dagger \, \delta_0^* \, Y$
- Use Hodge theory to study inconsistencies (e.g. global vs local)

HodgeRank

- Scores: Y_{ij}^{α} preference i over j by voter [race] α [beaten lengths]
- Aggregate scores: $\overline{Y}_{ij} = \dfrac{\sum_{\alpha \in \Lambda_{ij}} \omega^{\alpha} Y_{ij}^{\alpha}}{\sum_{\alpha \in \Lambda_{ij}} \omega^{\alpha}}$ $[\omega^{\alpha} = e^{-\frac{t_\alpha}{h}}]$
- Optimisation: $\min_{s \in C^0} \|\delta_0 s - \overline{Y}\|_W^2 = \min_{s \in C^0} \sum_{i,j \in V} w_{ij} \left(X_{ij} - \overline{Y}_{ij}\right)^2$

$$[\, w_{ij} = \frac{1}{1 + \Delta_1^{\mathrm{up}}(i,j)} \,]$$

- Minimum norm solution: $s^* = \Delta_0^{\dagger} \delta_0 \overline{Y}$
- Residual: $R^* = \overline{Y} - \delta_0 s^* = \mathrm{proj}_{\mathrm{im}(\delta_1^*)} \overline{Y} + \mathrm{proj}_{\ker(\Delta_1)} \overline{Y}$

local inconsistencies | global inconsistencies

Horse ranking: Results

- Data set: UK races 2008-2012 (5 years) about 36k races and 38k horses
- Computation: 4-5h in IRIDIS (12k cores)
- Validation: conditional logit model ($\widetilde{R}^2$, p-value, LLR), Kelly betting algorithm
- Results: improves predictions, significance, 'makes money'

CONCLUSIONS

- Use of weighted complexes & discrete Laplacian in complex systems modelling, and exploratory data analysis (*Geometrical Data Analysis*)
- **Advantages**
  - Natural generalisation of algebraic (& spectral) graph theory
  - Rich mathematical theory underpinning this approach
- **Challenges:**
  - Validity of modelling approach (e.g. meaningful, computationally feasible)
  - Relation to Topological Data Analysis (e.g. persistent homology)
  - Spectral signatures
  - Metrics, curvature, (hidden) geometries
  - (Quasi-)Symmetry   Etc.
- Upcoming review paper *Geometry and Topology of Networks* (with Ben MacArthur)

THANK YOU