Topology of Networks and Data
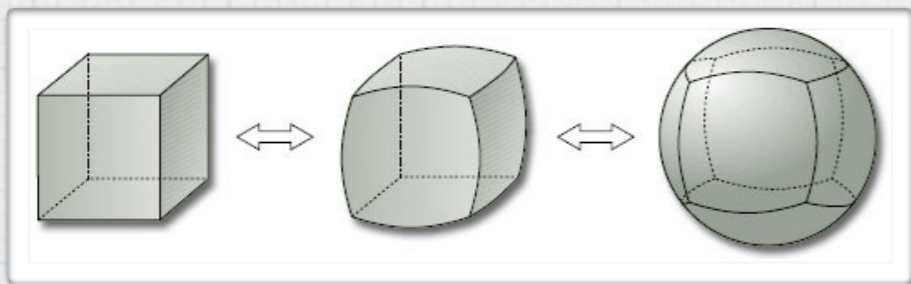
Ruben Sanchez-Garcia
Mathematical Sciences

WebScience MSc Guest Lecture
8 December 2016

1-slide summary
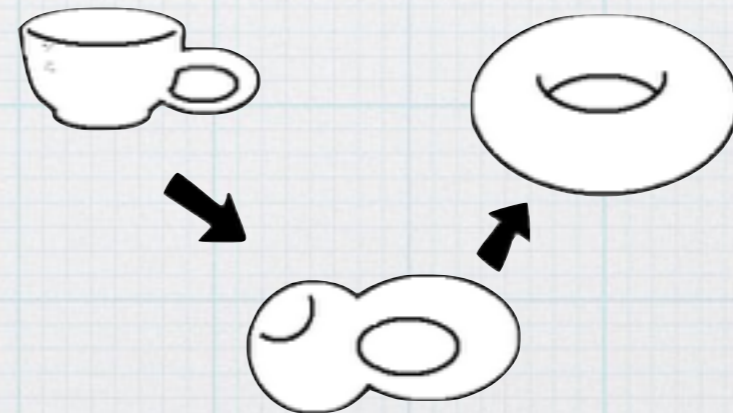
* I use TOPOLOGY to study complex systems, and data

* I use networks, and their higher-dimensional analogues, complexes, to represent 'shape'

* Data: mostly interested in unsupervised methods, particularly clustering, and ranking

Topology

* Mathematical study of 'shapes'

* Similar to Geometry, but objects can be continuously deformed

http://www.renyi.mta.hu/~szilard/topology/CUBE=SPHERE.jpeg

http://atomsandvoid.files.wordpress.com/2011/02/topology1.png?w=570
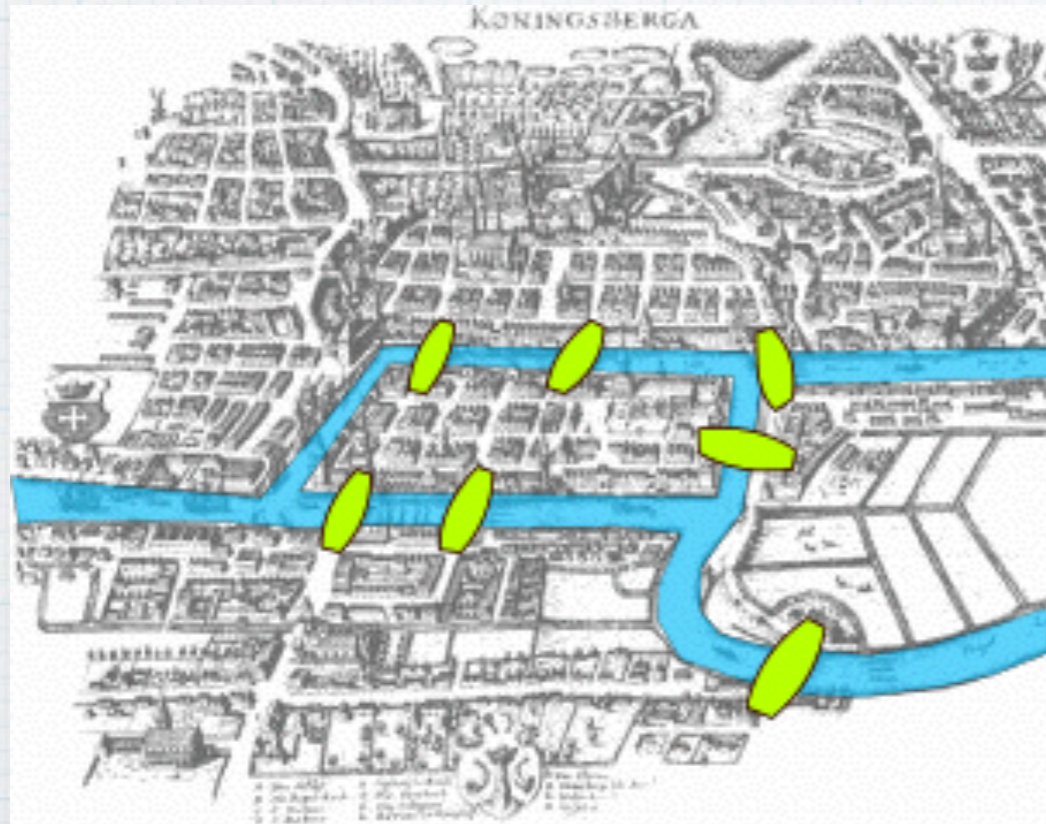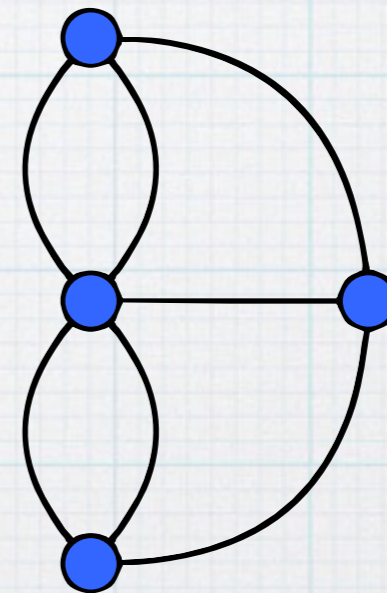
Seven bridges of Könisberg

Image by Bogdan Giuşcă (Wikipedia)

Euler (1735)

Discrete structures in Topology

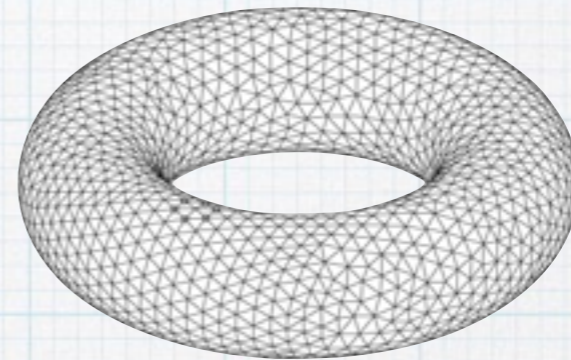* **Complexes: Discrete representations of continuous shapes**
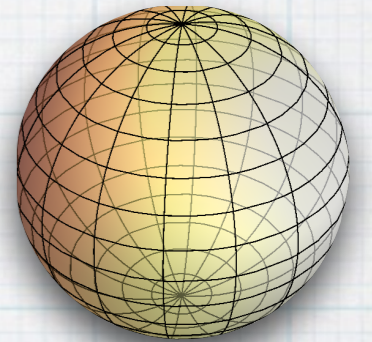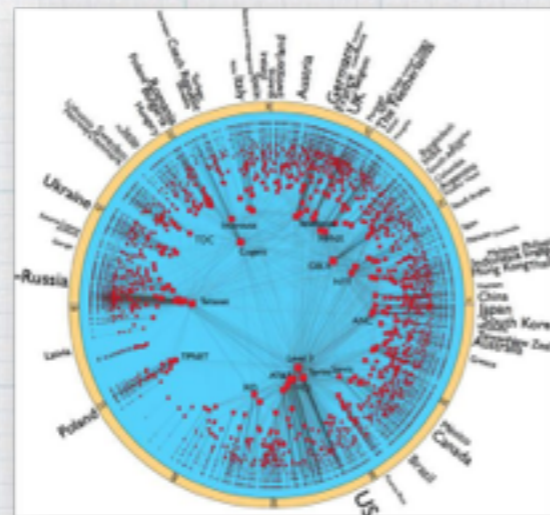

Image by Ag2gaeh (Wikipedia)
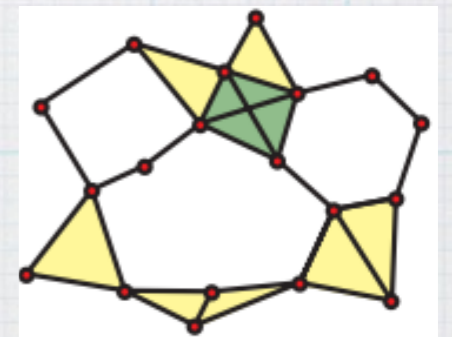
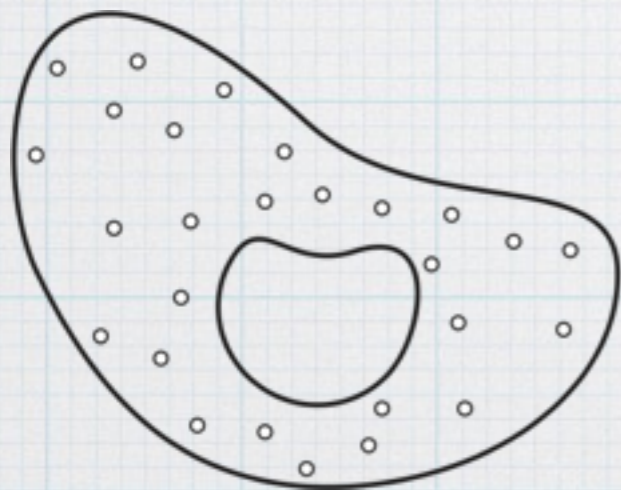Image: Mathematica

* **Example: Networks**

biological

technological

socio-technological

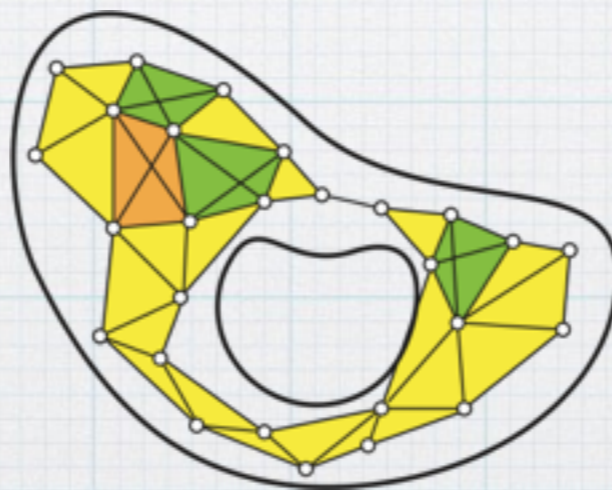
abstract complex

From Data to Complexes

* DATA = data points x_i with pairwise distances (or similarities) s_{ij}

* Example: point clouds in high-dimensional space
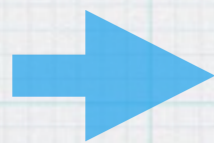
point cloud → data complex → TOPOLOGICAL FEATURES

Images: Jacek Brodzki
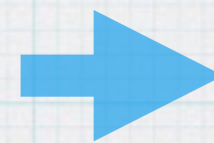
Topological Features

* **Example:** Euler characteristics of a complex

$$\chi = \#nodes - \#links + \#faces$$

 $\chi($ $) = 2$ $\chi($ $) = -1$

$\chi($ $) = 0$ Etc.

Images: Mathematica

Topological Data Analysis
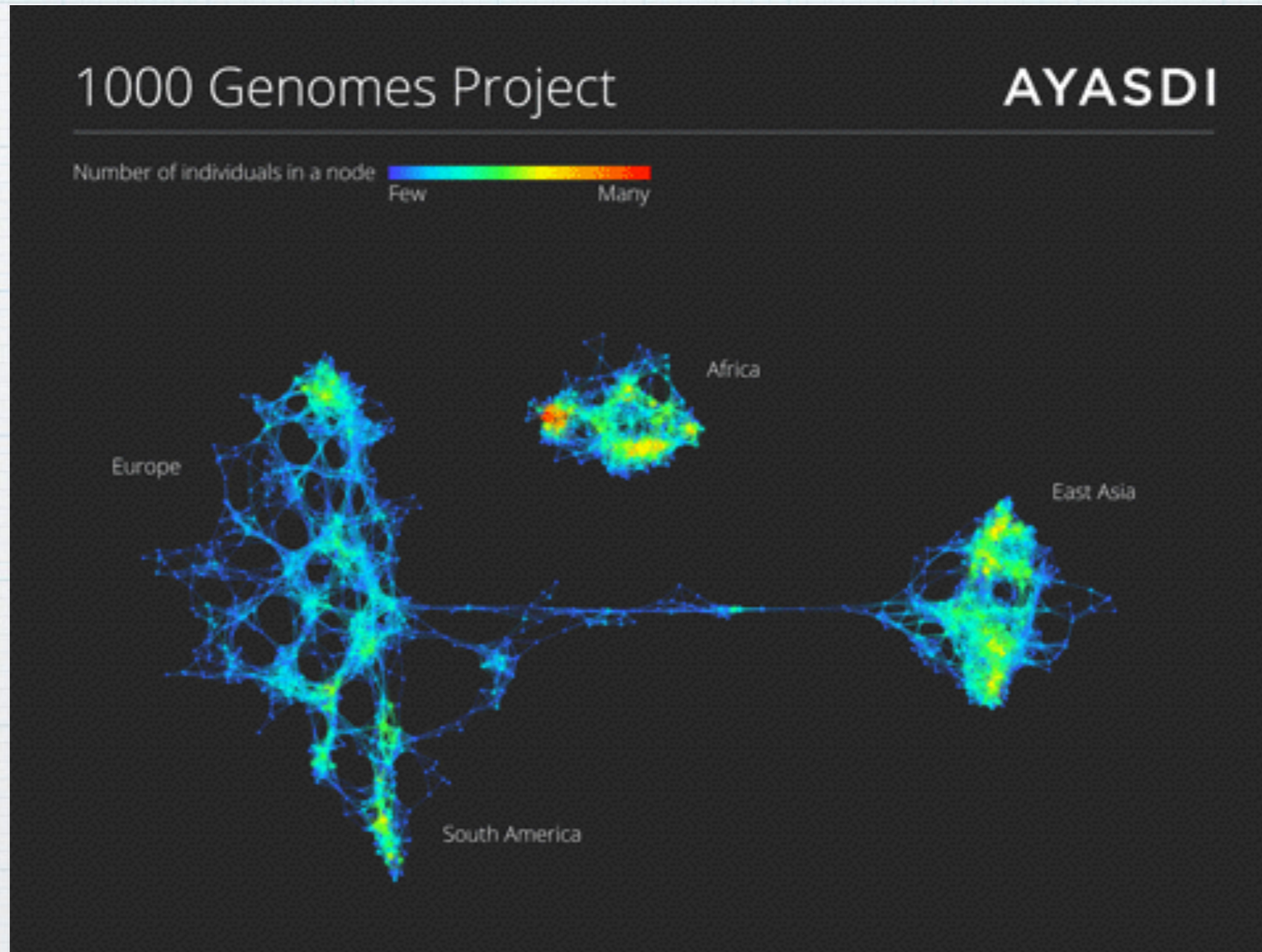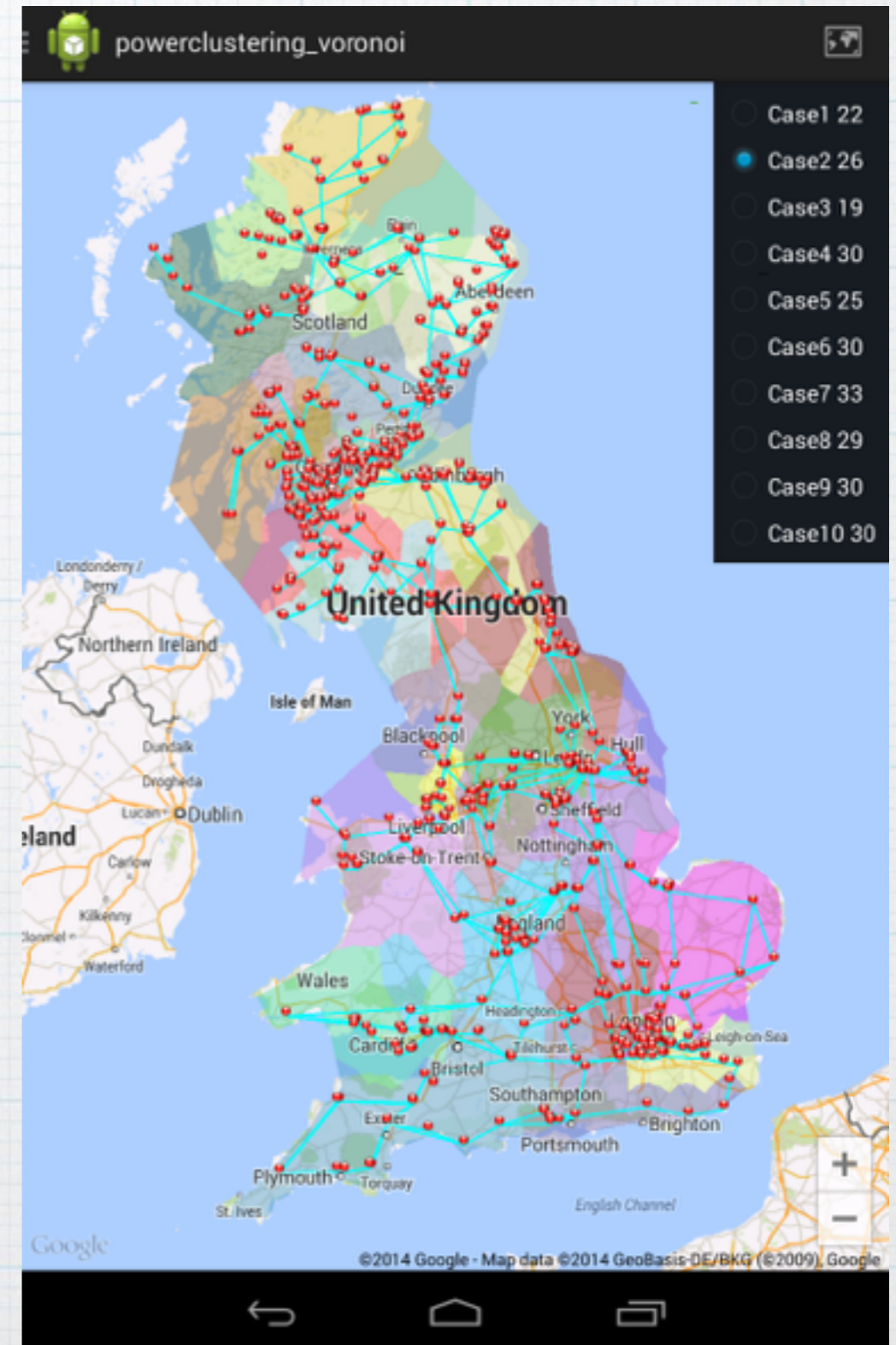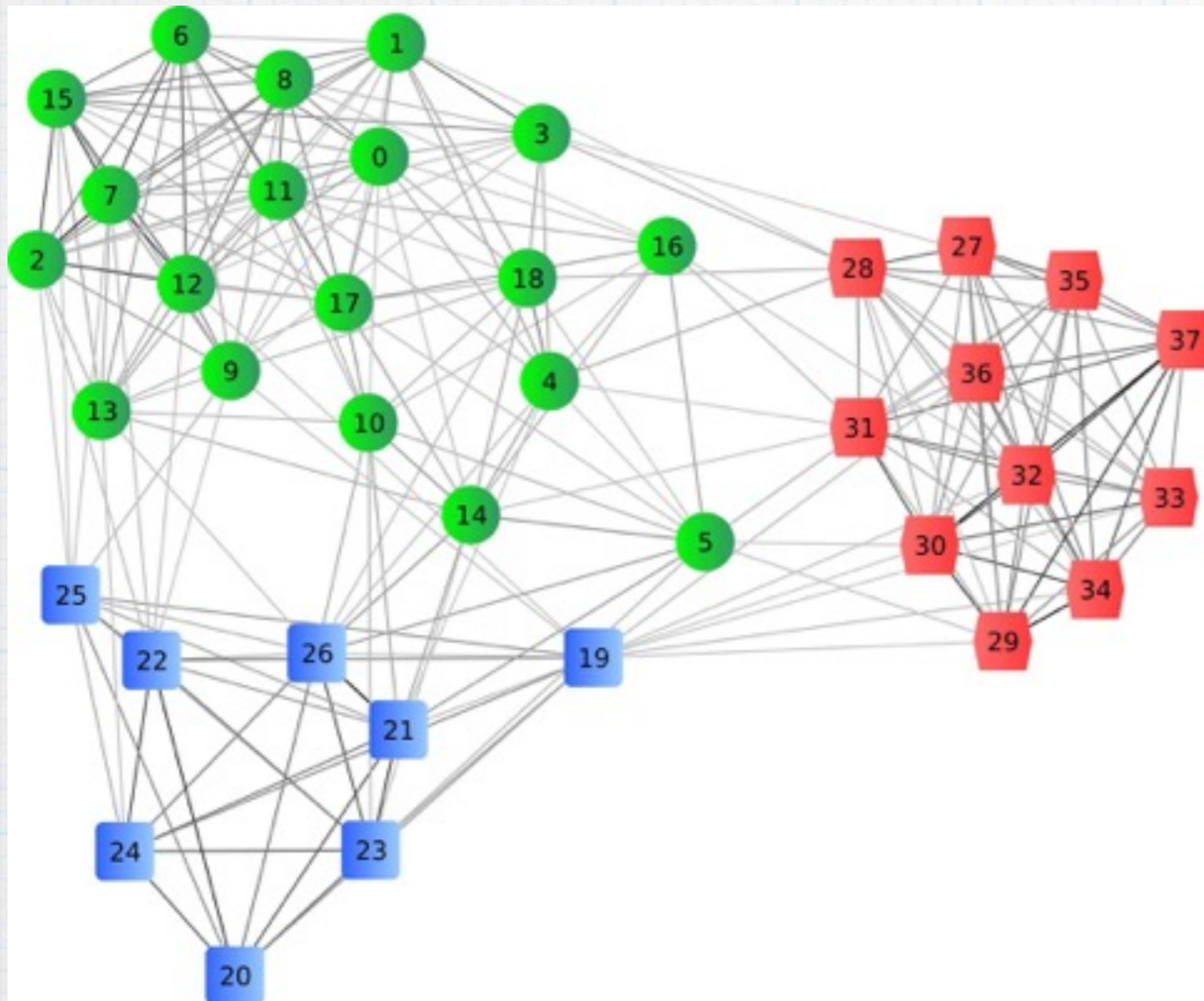
Clustering/Community detection

* Spectral clustering: network eigenvectors reveal clustering structure

Topological Ranking

* Current WebScience student Conrad D'Souza

* Ranking is a key activity that permeates the Web

* Challenges: missing data, inconsistencies

* Topological ranking (HodgeRank): global ranking & explains residual error

* Case study: horse racing data

1-slide summary (again!)

* I use TOPOLOGY to study complex systems, and data

* I use networks, and their higher-dimensional analogues, complexes, to represent 'shape'

* Data: mostly interested in unsupervised methods, particularly clustering, and ranking

Want to know more?

[1] G. Carlsson, Topology and data, Bulletin of the American Mathematical Society, 2009.

[2] JavaPlex: Persistent homology and topological data analysis library http://javaplex.github.io/javaplex/

[3] Mapper in Python (Daniel Müllner) http://danifold.net/mapper/index.html

[4] Ayasdi http://www.ayasdi.com/

[5] I. J. Good, The Philosophy of Exploratory Data Analysis, Philosophy of Science, 1983.

[6] M. Nicolau, A. J. Levine, and G. Carlsson, Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival, PNAS, 2011.

[7] J. M. Chan, G. Carlsson, and R. Rabadan, Topology of viral evolution, PNAS, 2013.

[8] Sayan Mukherjee: https://stat.duke.edu/~sayan/Publications.html

or talk to me

R.Sanchez-Garcia@soton.ac.uk

Office 8023 in Building 54

Thank you