The weighted spectral distribution; A graph metric with applications.

Dr. Damien Fay.

SRG group,
Computer Lab, University of Cambridge.

UNIVERSITY OF CAMBRIDGE

A graph metric: motivation.

Graph theory ↔ statistical modelling

- Data point.
- Statistical model,
  - ARMA model, neural network etc.
- error/residual
- Sum squared error

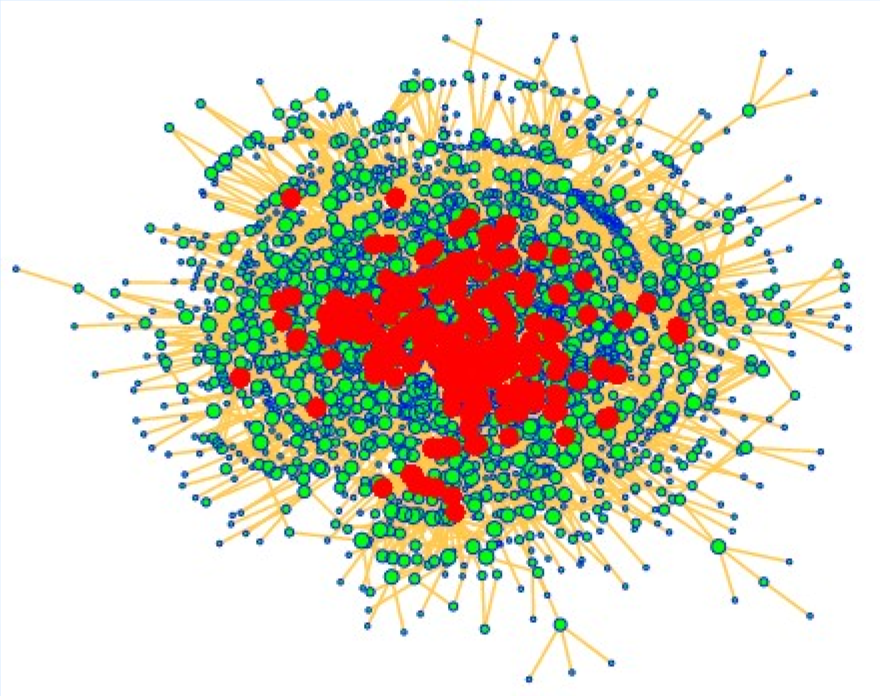
- Observed graph at time *t*
- Topology generator.
  - BA model.
  - INET model etc.
- Weighted spectral distribution.
- Quadratic norm between WSD's (Weighted Spectral Distributions.)

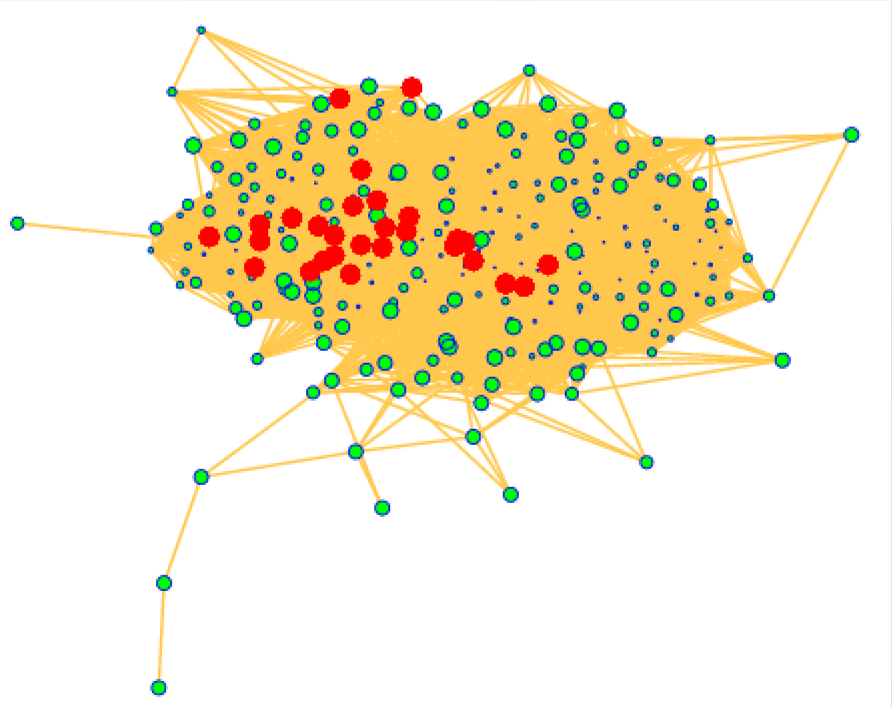
- Inference: has the process generating the network changed over time?
- Parameter estimation: what parameters best fit the observed graph/data points.
- Model validation: does the proposed topology generator represent the data well?
- Clustering: Can we separate classes of graphs into their respective clusters?

UNIVERSITY OF CAMBRIDGE

A metric for graph distance.

What is 'structure'?





- Both graphs share graph measures:
    - Clustering coefficient,
    - Degree distribution,
- There exists no other method for large networks.

Normalised Laplacian matrices.

Normalised Laplacian:

$$L(G)(u,v)=\begin{cases} 1 & if\ u=v, d_u \neq 0 \\ \frac{-1}{\sqrt{d_u d_v}} & u \neq v \\ 0 & otherwise \end{cases}$$

Alternatively using adjacency matrix and diagonal matrix D with degree of nodes:

$$L(G)=I-D^{-1/2}AD^{-1/2}$$

Expressing $L(G)$ using the eigenvalue decomposition:

$$L(G)=\sum_i \lambda_i e_i e_i^T$$

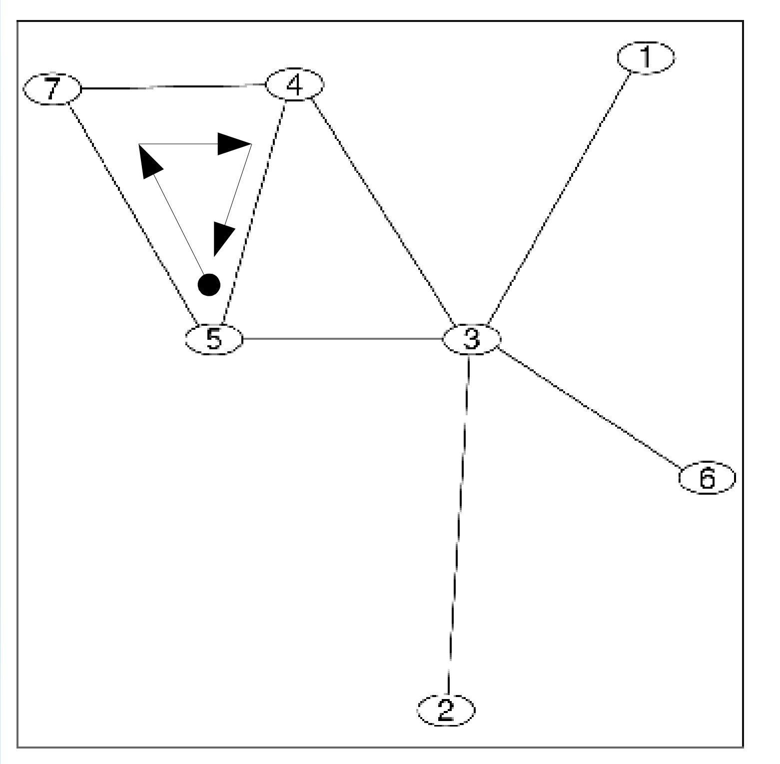
Note:

- $L(G)$ may be approximated using $\lambda_i\, e_i\, e_i^{\mathrm{T}}$ with approximation error proportion to $1\text{-}\lambda_i$
- e_i identifies i^{th} cluster in the data, assigning each node its importance to that cluster. (spectral clustering).
- Unlike spectral clustering we will use all eigenvalues.

Background theory.

- A *random walk cycle*: the probability of starting at a node taking a path of length N and returning to the original node.

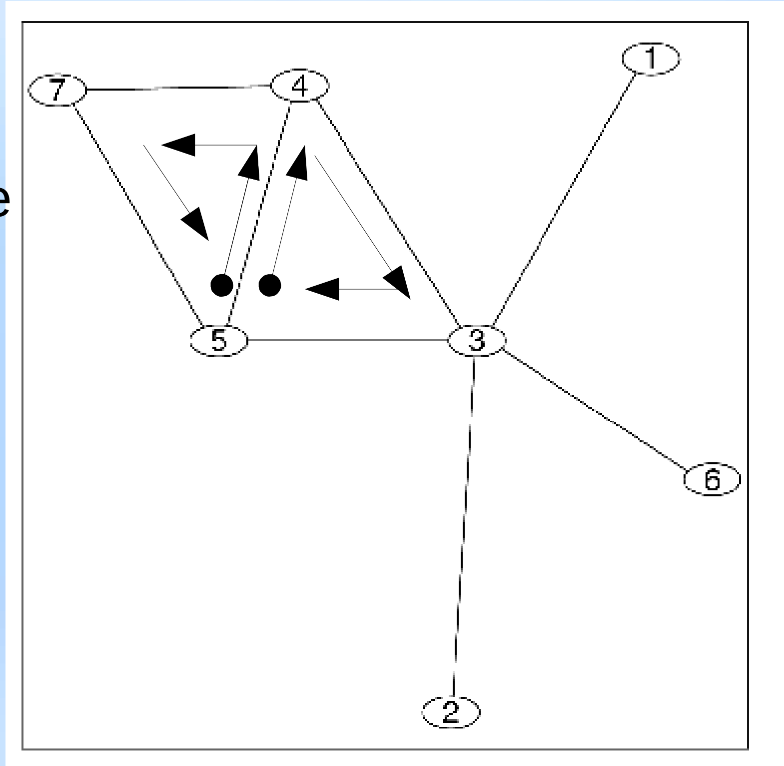
$\frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} = 0.055$



Background theory.

- A *random walk cycle*: the probability of starting at a node taking a path of length *N* and returning to the original node.

- Several alternative 3-cycles available.
- The *N*-cycles are a measure of the local connectivity of a node.



Theory: random walk cycles.

$$L(G) = I - D^{-1/2} A D^{-1/2}$$

Defining the matrix on the right as B:

$$B = D^{-1/2} A D^{-1/2}$$

The elements of B may be expressed in terms of degrees and the adjacency matrix as:

$$B_{i,j} = \frac{A_{i,j}}{\sqrt{d_i}\sqrt{d_j}}$$

$(B^N)_{ij}$ is the sum of products of all paths of length N starting at node i and ending at node j As $1\text{-}\lambda_i$ are the eigenvalues of B :

$$\sum_i (1-\lambda_i)^N = tr(B^N)$$

We defining a random walk cycle to be a random walk of length N that starts and end at the same node (repeats included). This may be expressed in terms of B by noting :

$$B_{i,j} B_{j,k} \ldots B_{l,i} = \frac{A_{i,j}}{\sqrt{d_i}\sqrt{d_j}} \frac{A_{j,k}}{\sqrt{d_j}\sqrt{d_k}} \ldots \frac{A_{l,i}}{\sqrt{d_l}\sqrt{d_i}} = \frac{1}{d_i d_j \ldots d_k}$$

UNIVERSITY OF CAMBRIDGE

Which simply results in the diagonal of B^N:

$$B_{i,j} B_{j,k} \dots B_{l,i} = \left(B^N\right)_{i,i}$$

Thus the eigenvalues of the normalised Laplacian can be related to random walk cycles in a graph via:

$$\sum_i \left(1-\lambda_i\right)^N = tr\left(B^N\right)$$

i.e. we may calculate the number of weighted *N*-cycles via:

$$\sum_i \left(1-\lambda_i\right)^N = tr\left(B^N\right) = \sum_C \frac{1}{d_i d_j \dots d_k}$$

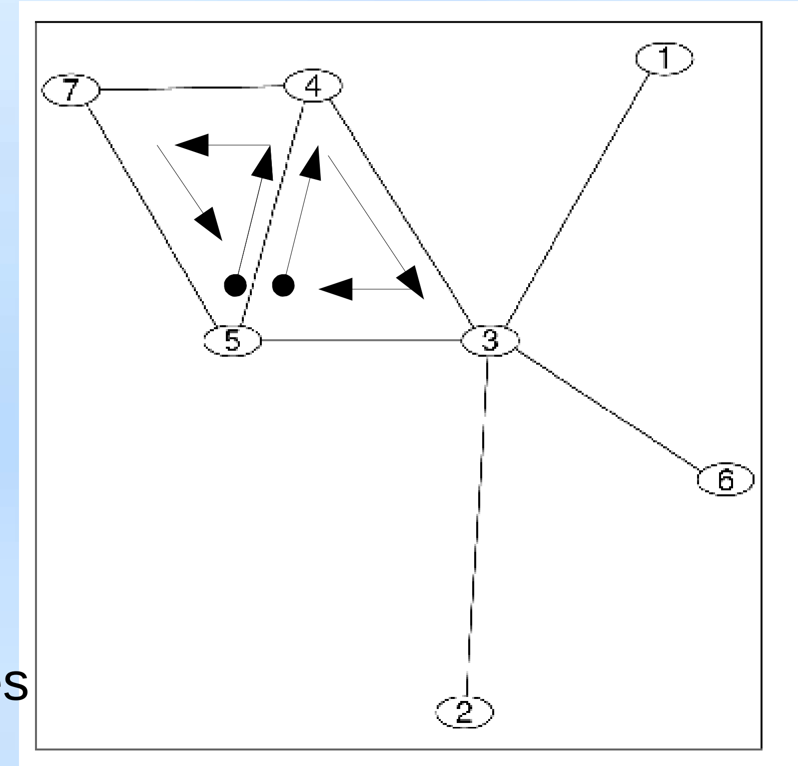
Where *C* is the set of *N*-cycles in the graph containing the *N* nodes *i*,*j*, ...*k*.

Background theory.

- A *random walk cycle*: the probability of starting at a node taking a path of length *N* and returning to the original node.

Theorem:

$$\sum_i \left(1-\lambda_i\right)^N = tr\left(B^N\right) = \sum_C \frac{1}{d_i d_j \ldots d_k}$$



- The right hand side is the sum of probabilities of all *N*-cycles in a graph.
- The left hand side relates this to the eigenvalues of the normalised Laplacian.
- We get a relationship between the eigenvalues of the normalised Laplacian (*global structure*) and the *N*-cycles (*local structure*) of a graph.

Theory: spectral distribution.

- Problem:
  - Estimating the eigenvalues is expensive and inexact, in addition we are really only interested in those near 0 or -2.
- Solution:
  - Using Sylvester's law of inertia and pivoting calculate the number of eigenvalues that fall in an interval => we are now looking at the distribution of eigenvalues, $f(\lambda)$.
  - The weighted spectral distribution can now be defined as:

$$WSD: G \rightarrow R^{|K|}\left\{k \in K : (1-k)^N f(\lambda = k)\right\}$$

Theory: metric definition.

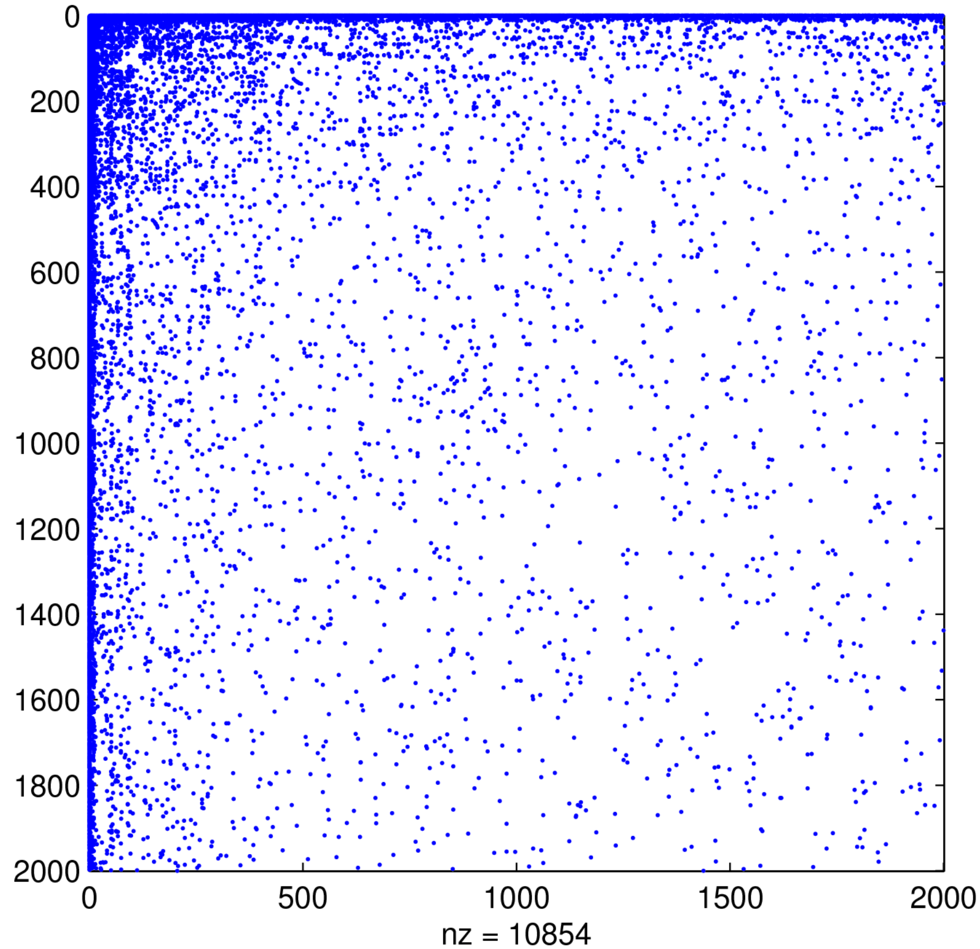
Finally we may define a metric based on the quadratic norm between the weighted spectral distributions of two graphs, G_1 and G_2, as:

$$\Delta(G_{1,}G_{2,}N)=\sum_{k\in K}(1-k)^N\left(f_1(\lambda=k)-f_2(\lambda=k)\right)^2$$

- Notes:
  - The number of components in a graph is equal to the number of eigenvalues at 0.
    - This is given the highest structural weighting.
  - Eigenvalues in the spectral gap (i.e. close to 1) are given very low weighting
    - as the spectral gap is expected to hold little structural information (it is important for other things!).
  - All the eigenvalues are considered not just the first *k.*
  - *Δ* is a metric in the strict sense except for the identity law which is true almost certainly.

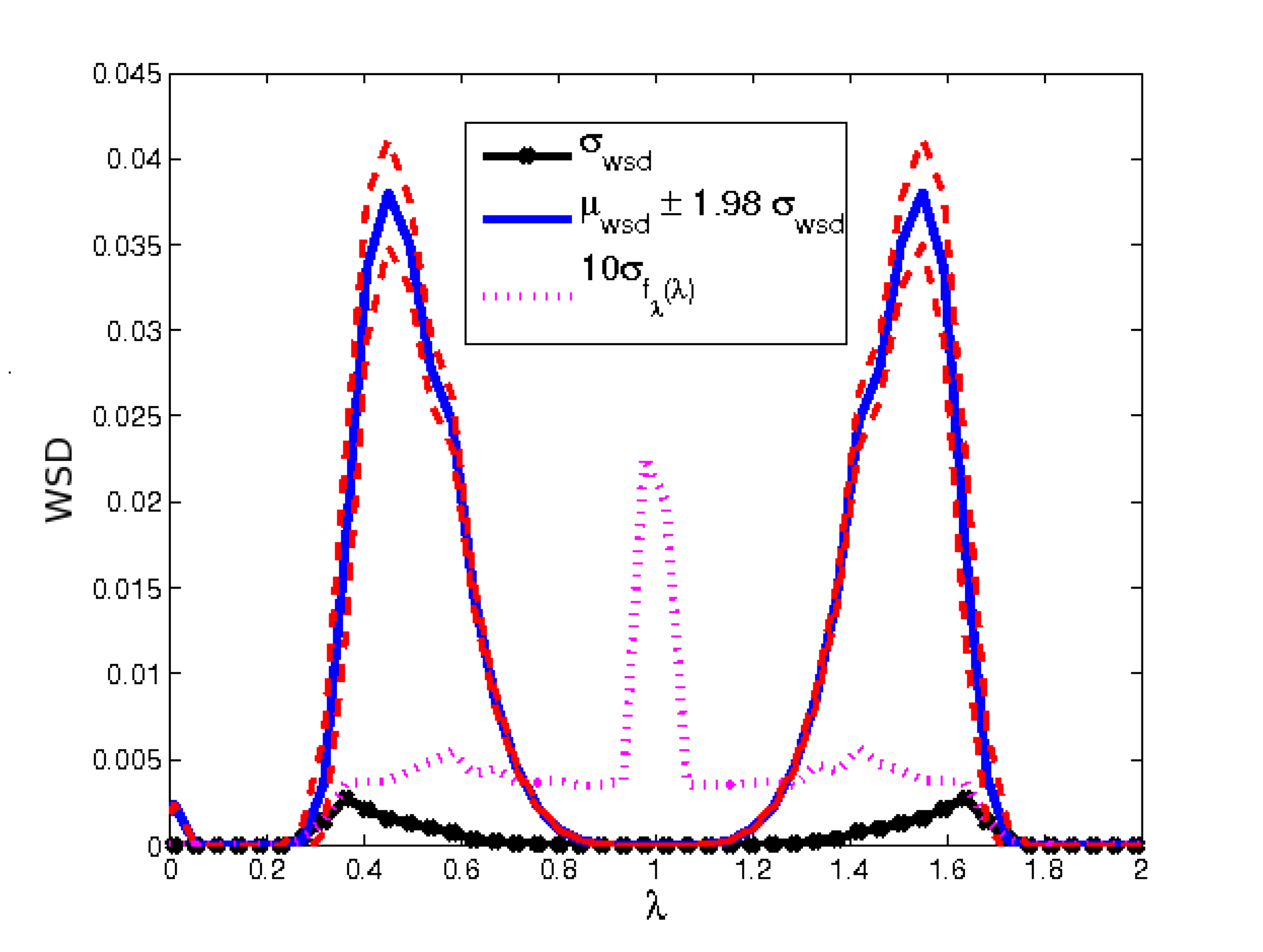
UNIVERSITY OF CAMBRIDGE

WSD example



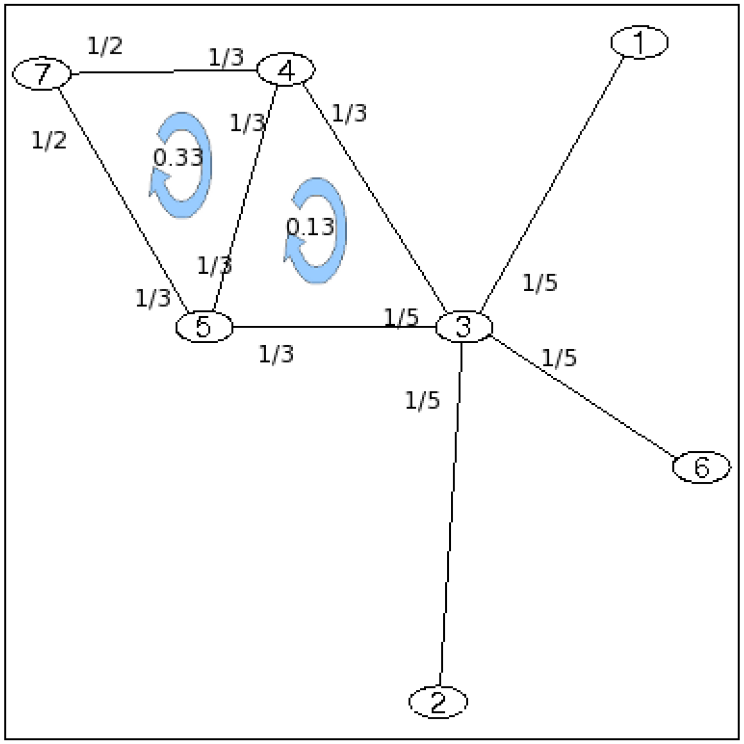
Adjacency matrix of an AB graph, 2000 nodes.

WSD example



WSD taken over 51 bins.

Simple example



Examine the number of 3-cycles in this graph.

- There are two 3-cycles
  - $\frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} \times 6 = 0.333$
  - $\frac{1}{3} \times \frac{1}{3} \times 1/5 \times 6 = \underline{0.133}$

    0.466

(note: 6 from the 6 directional cycles in each loop).

Normalised Laplacian eigenvectors.

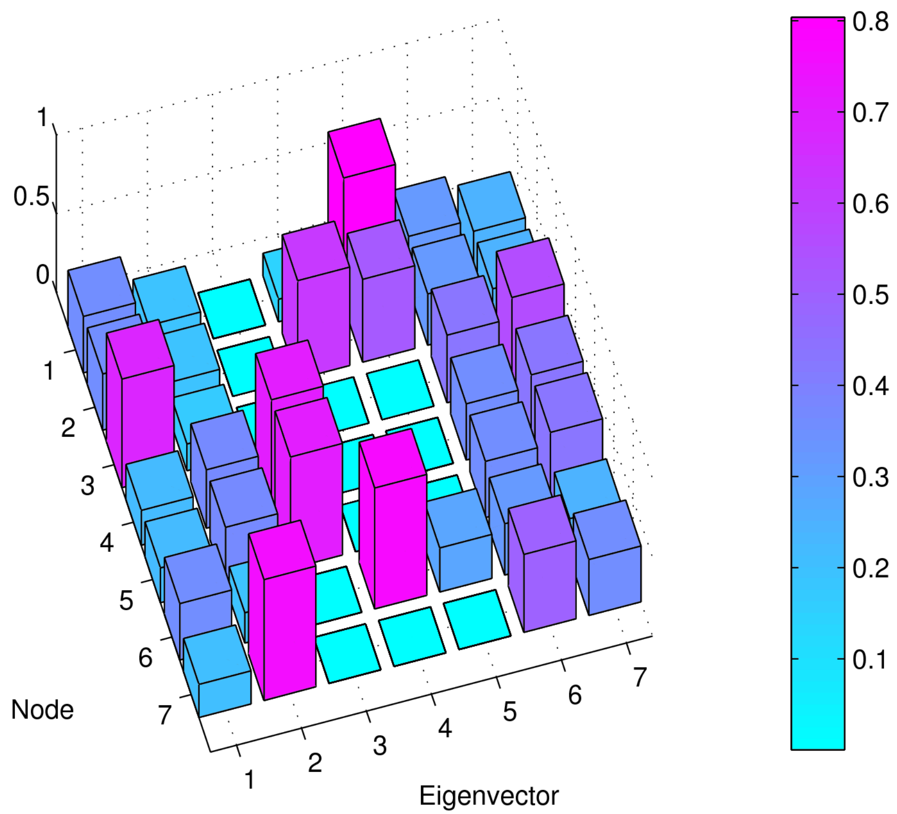
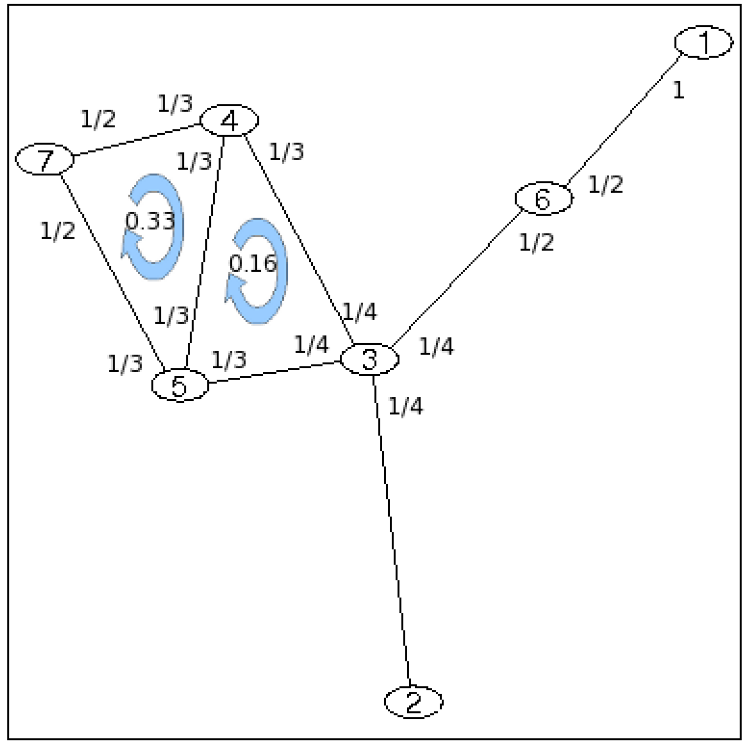


TABLE I

EIGENVALUES, WSD AND DOMINANT NODES OF EXAMPLE NETWORK.

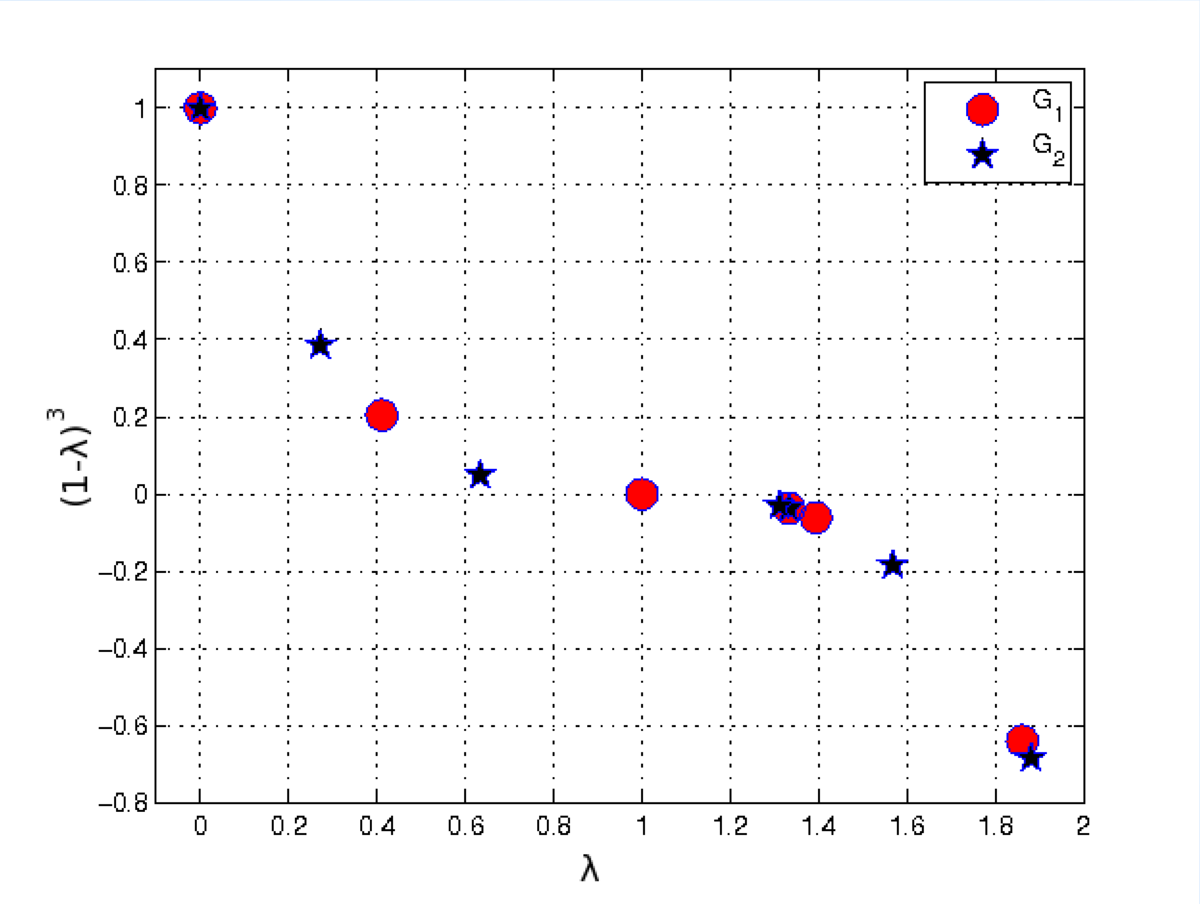
e_7	Eigenvector	λ	$1-\lambda$	$(1-\lambda)^3$	Dominant nodes
0.2500	1	1.8615	-0.8615	-0.6394	3,1,2,6
0.2500	2	1.3942	-0.3942	-0.0612	7,4,5
0.5590	3	1.3333	-0.3333	-0.0370	4,5
0.4330	4	1.0000	0.0000	0.0000	6,2
0.4330	5	1.0000	0.0000	0.0000	1,2,6
0.2500	6	0.4110	0.5890	0.2043	7,3
0.3536	7	0.0000	1.0000	1.0000	3,4,5,7
	$\sum_{i=1}^{7}(1-\lambda_i)^3$			0.4667	

Adjusting the network



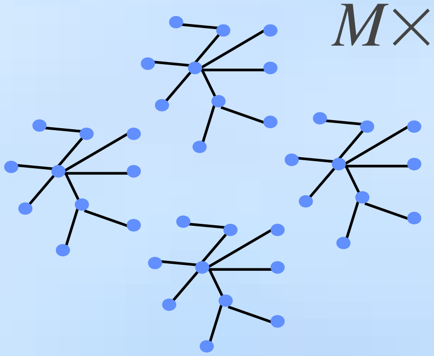
- Node 1 has been rewired from node 3 to link to node 6.
    - The loops are unchanged.
    - However, the random walk probabilities have changed.

WSD example



- The effect is to move the eigenvalues and thus the random walk cycle probabilities.
- Note: this is not the case when using the adjacency matrix.

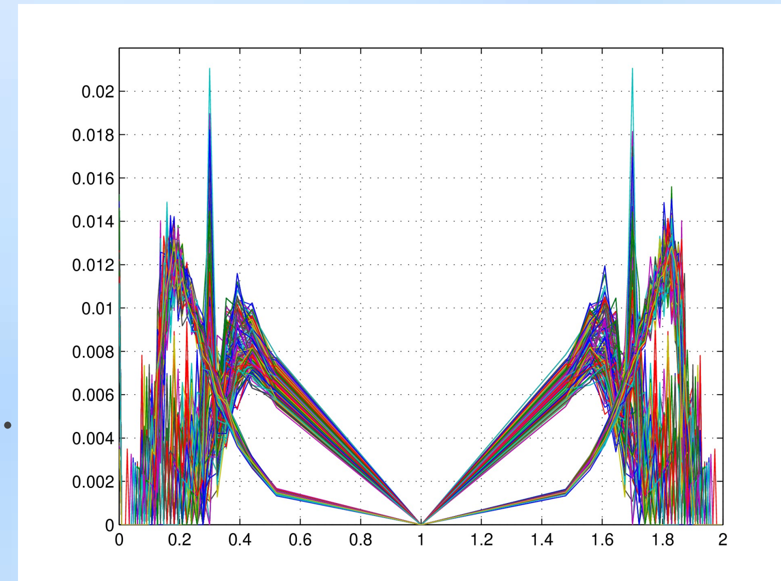
Clustering using the WSD.



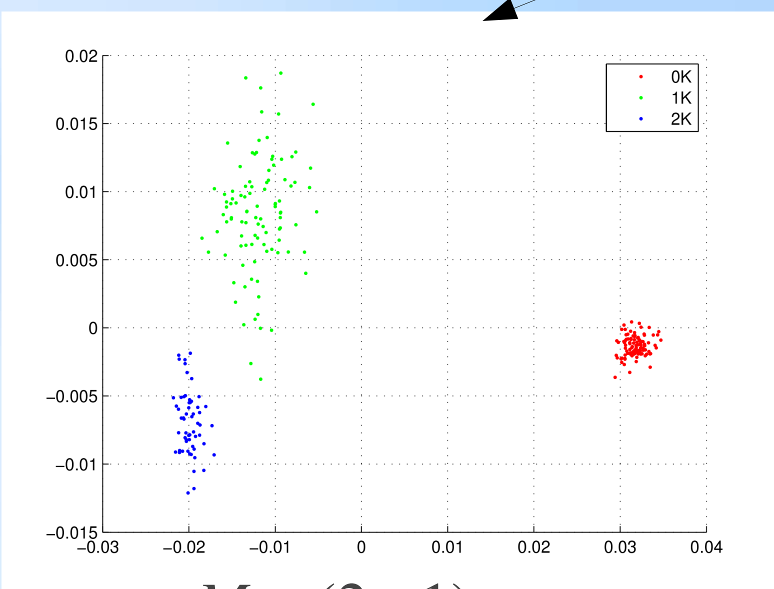
$M\times(N\times N)$

WSD

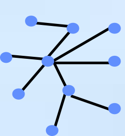
$M\times(K\times 1)$



Random Projection or Multidimensional scaling.



$M\times(2\times 1)$



M objects.
N- nodes.
K bins.
k co-ordinates

UNIVERSITY OF CAMBRIDGE

Random Projection.

Random projection is a technique for data compression often used in compressed sensing.

Basic idea is very simple:

Given a matrix $A = M \times K$ we wish to produce a matrix of reduced dimension $M \times k$ *where* $k << K$.

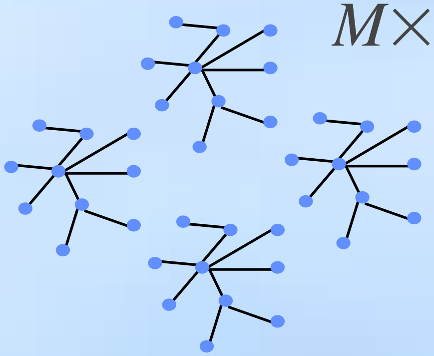
We can form an approximation to A in k dimensions by randomly projecting A onto an $M \times k$ projection matrix T where $T \sim N(0,1)$.

i.e. we simply multiply the data by a matrix of appropriate size containing random numbers!.

Note: $E[T_{i,j}\ T_{k,l}] = 0\ \forall\ i,j \neq k,l$ => inner product of two rows of T is zero in expectation => T is *(nearly)* **orthogonal**.

UNIVERSITY OF CAMBRIDGE

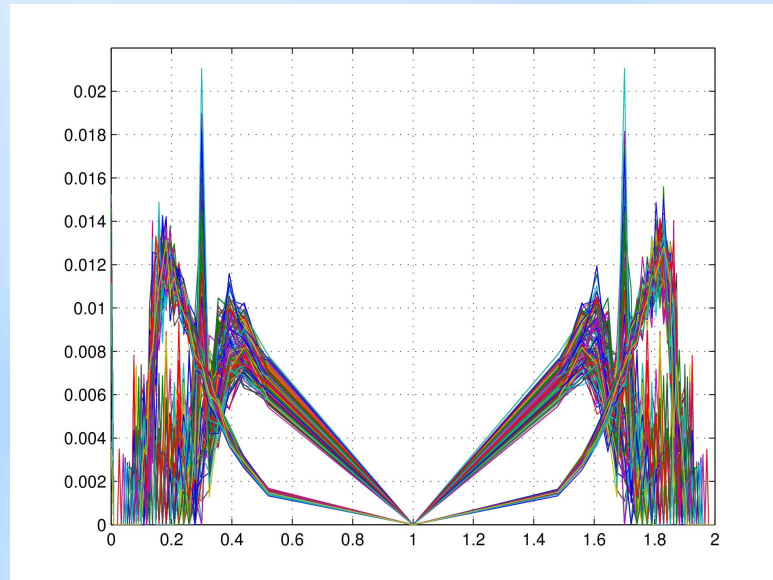
Random projection example.



$M \times (N \times N)$

WSD

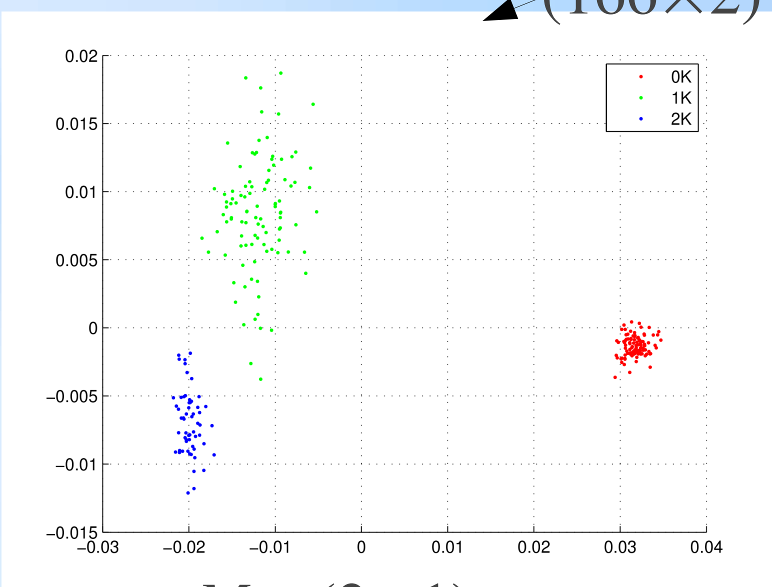
$M \times (K \times 1)$



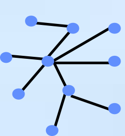
$\sim N(0,1)$

$(166 \times 71) \times (71 \times 2) =$

(166×2)



$M \times (2 \times 1)$



M objects.
N- nodes.
K bins.
k co-ordinates

UNIVERSITY OF CAMBRIDGE

Multi-dimensional Scaling.

Given

- matrix $A = M \times K$,
- a metric defining the distance between each row of A,

Aim:

- produce a matrix of reduced dimension $M \times k$ *where* $k << K$.

First we construct the dissimilarity matrix:

$$\Delta(i,j) = \Delta(G_i, G_j)$$

T construct the Gram matrix by double centring the distances as:

$$H = -1/2 J \Delta^{\circ 2} J \qquad J = I_N - 1_N 1_N^T / N$$

A projection into k dimensions may then be constructed using the first k eigenpairs of H:

$$Y = [V]_{1:k} [\Lambda]_{1:k}^{1/2} \qquad H = V \Lambda V^T$$

Aside (by coincidence current research involves):

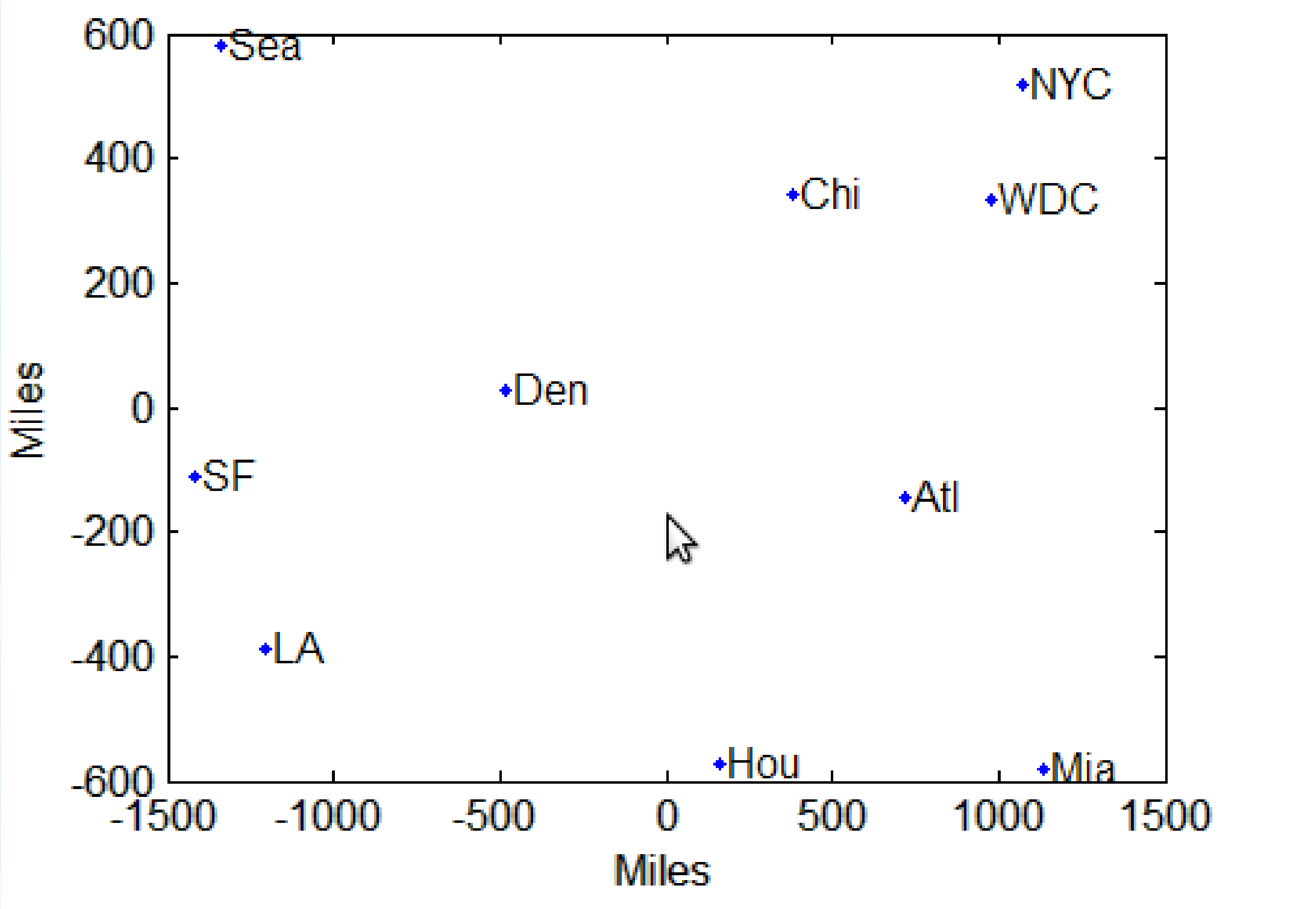
- MDS also forms the core of localisation and tracking techniques.
- If Δ is not complete several methods exist;
  - the Nystrom approximation for missing blocks;
  - Weighted MDS via SDP for missing elements.
  - Apply a particle filter to track movement and estimate distances and weights (error variance).

$\Delta(G_i, G_j)$ = quadratic norm between WSD's as defined earlier.

Example.

	Atlanta		Denver		LA					
Atlanta	0									
	587	0								
Denver	1212	920	0							
	701	940	879	0						
LA	1936	1745	831	1374	0					
	604	1188	1726	968	2339	0				
	748	713	1631	1420	2451	1092	0			
	2139	1858	949	1645	347	2594	2571	0		
	2182	1737	1021	1891	959	2734	2408	678	0	
	543	597	1494	1220	2300	923	205	2442	2329	0

example



600
400
200
0
-200
-400
-600
Miles
Sea
NYC
Chi
WDC
Den
SF
Atl
LA
Hou
Mia
-1500
-1000
-500
0
500
1000
1500
Miles

Example applications.

- Estimating the optimum parameters for a topology generator.
  - Comparing which topology generator produces a 'best' fit for the internet.
- Tracking evolution of the internet.

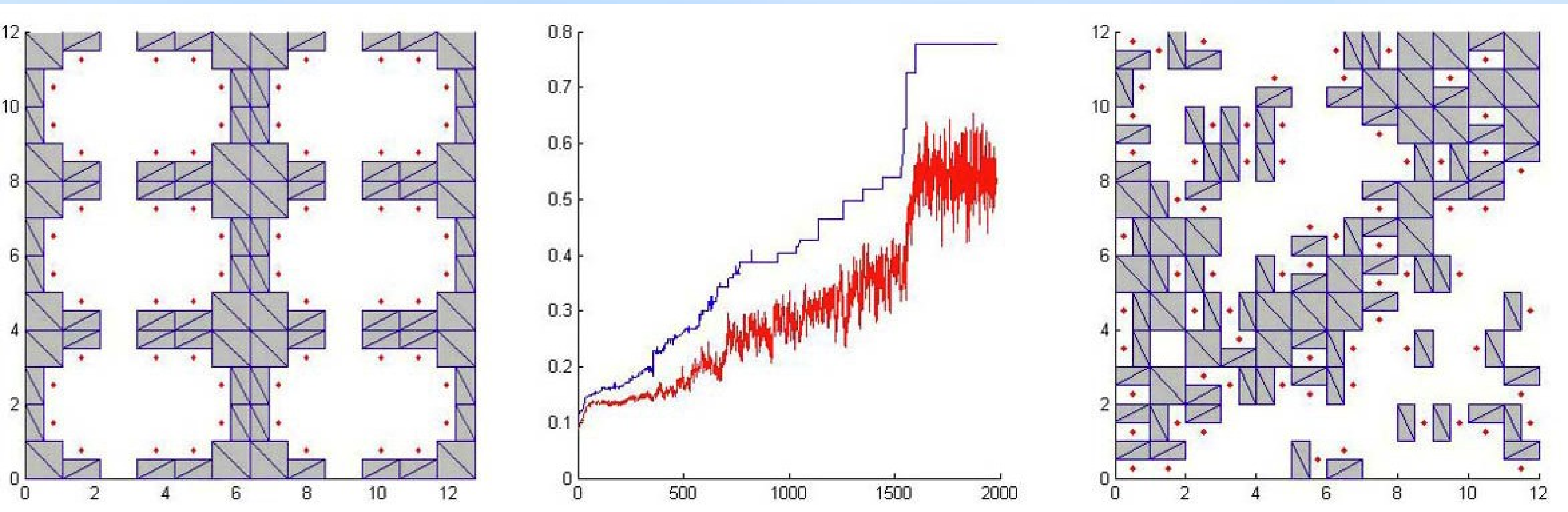
- Clustering applications:
  - Discriminating between topology generators.
  - Network application identification.
  - Orbis model analysis.

Internet AS topology models

- We compare 5 topology generators:
- The Waxman model
- The 2nd Barabasi and Albert Model (BA2)
- The Generalised Linear Preference model (GLP)
- The INET model
- Positive Feedback Preference model (PFP)
- To 2 data sets for the internet at AS level:
  - Skitter dataset (Traceroute based).
  - UCLA dataset (BGP looking glass server)

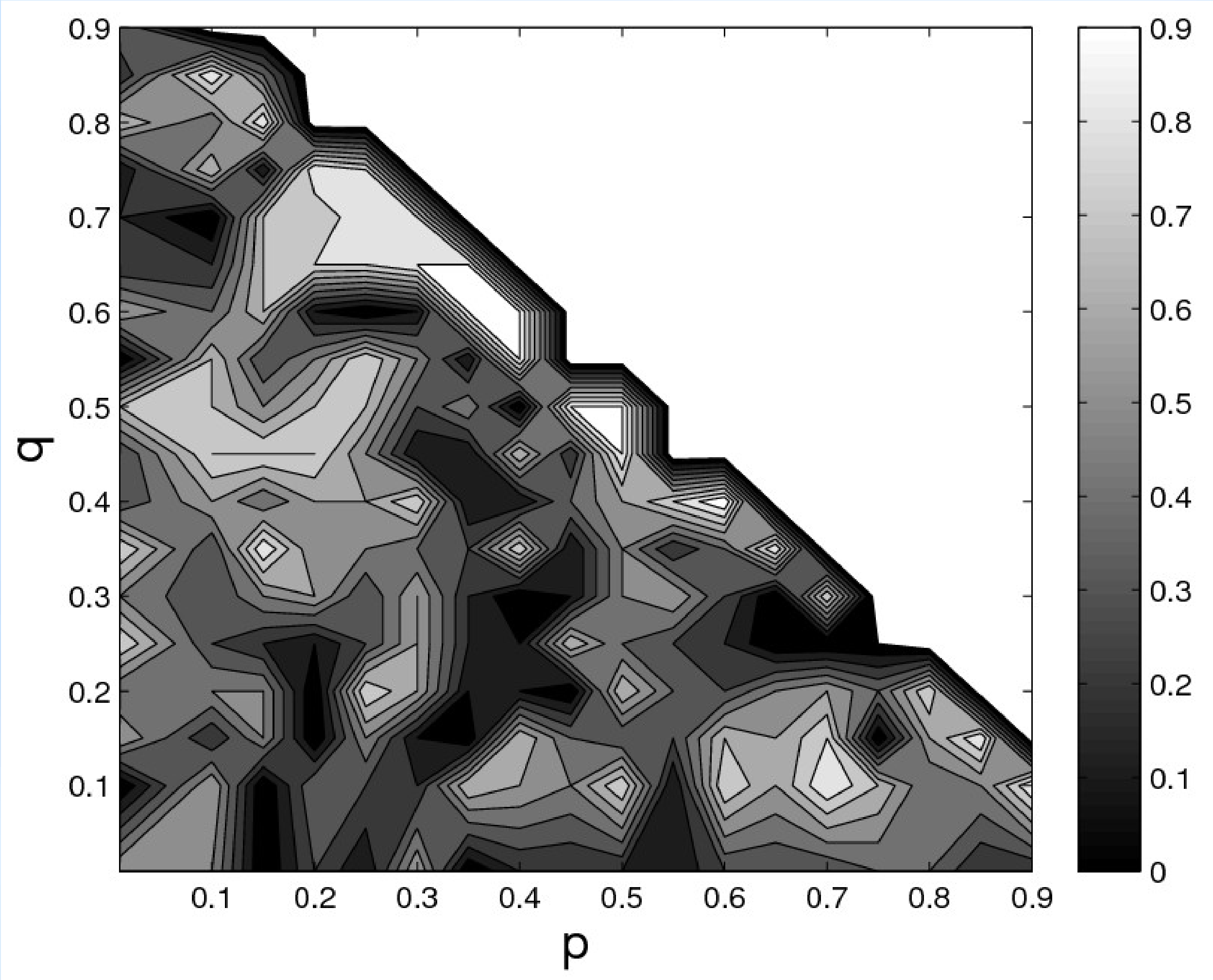
Related work

[3] S. Hanna, “Representation and generation of plans using graph spectra,” in *6th International Space Syntax Symposium*, Istanbul, (2007).



Application 1: Tuning topology generators.

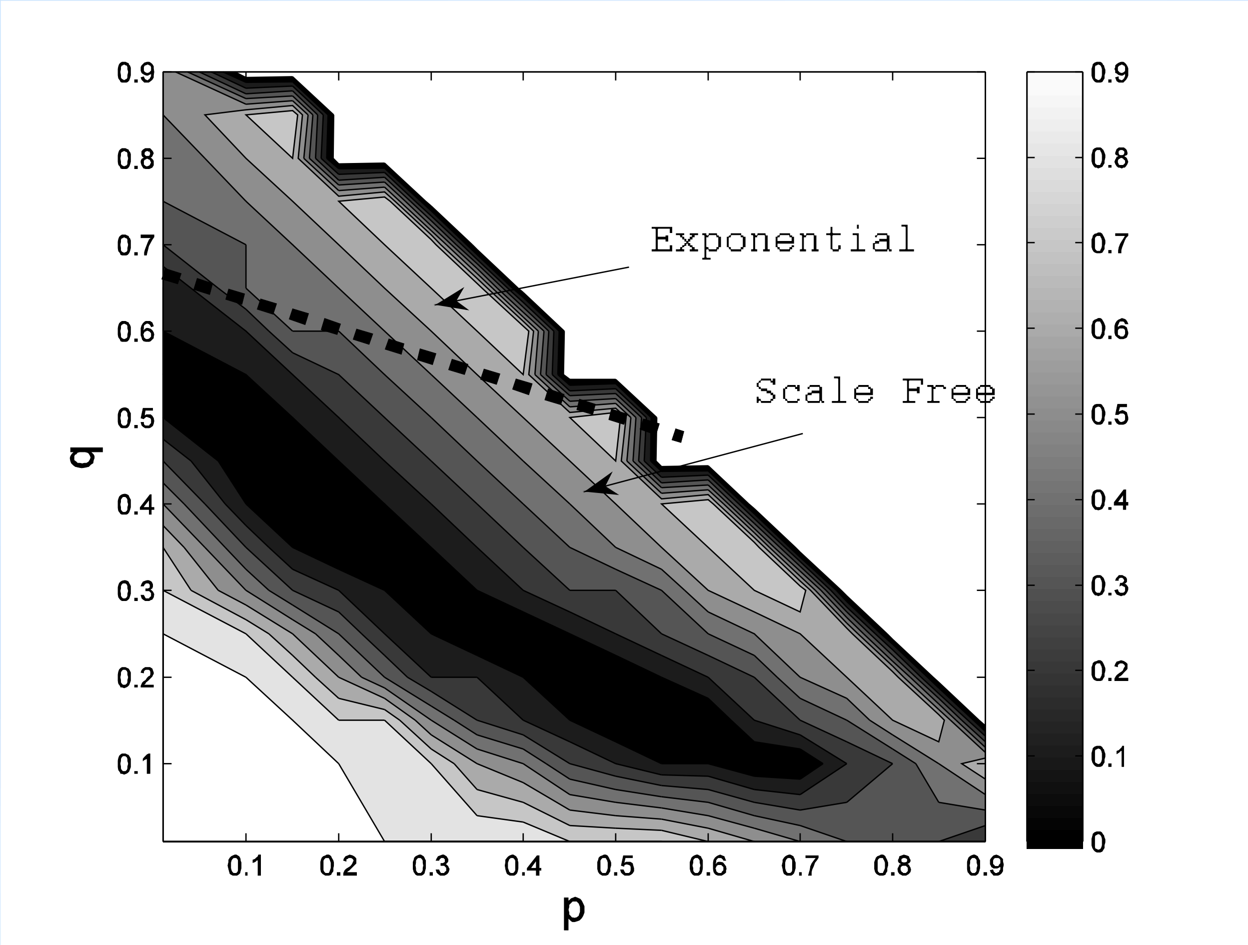
How NOT to select appropriate parameters for a topology generator.



Tuning an AB2 model using the (unweighted) spectral difference.

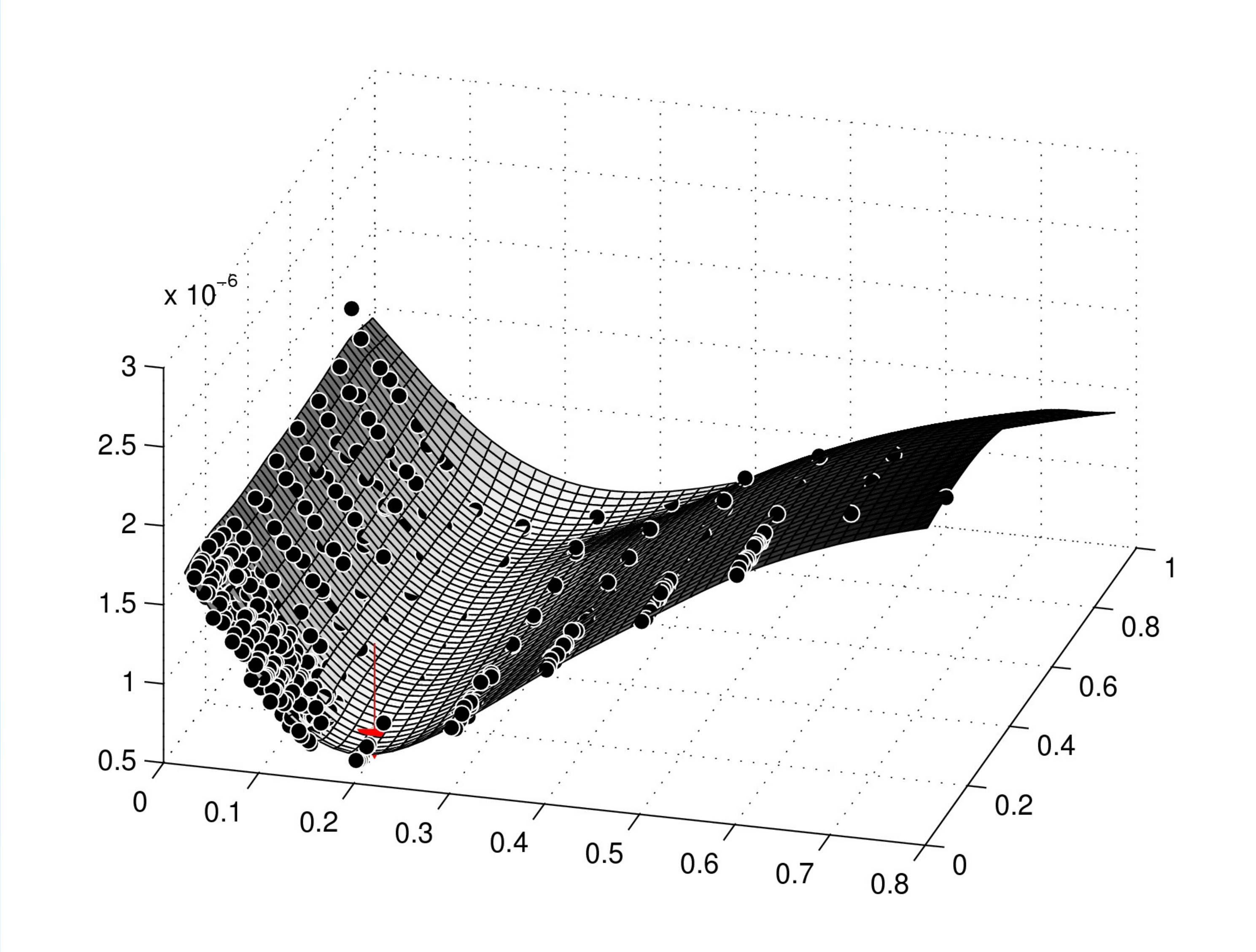
The WSD result.

How ~~NOT~~ to select appropriate parameters for a topology generator.



Tuning an AB model using a weighted spectral difference.

A 3-D view and example of speeding up the calculations



Tuning a GLP model using the WSD and active learning.

Application 1a: Comparing topology generators.

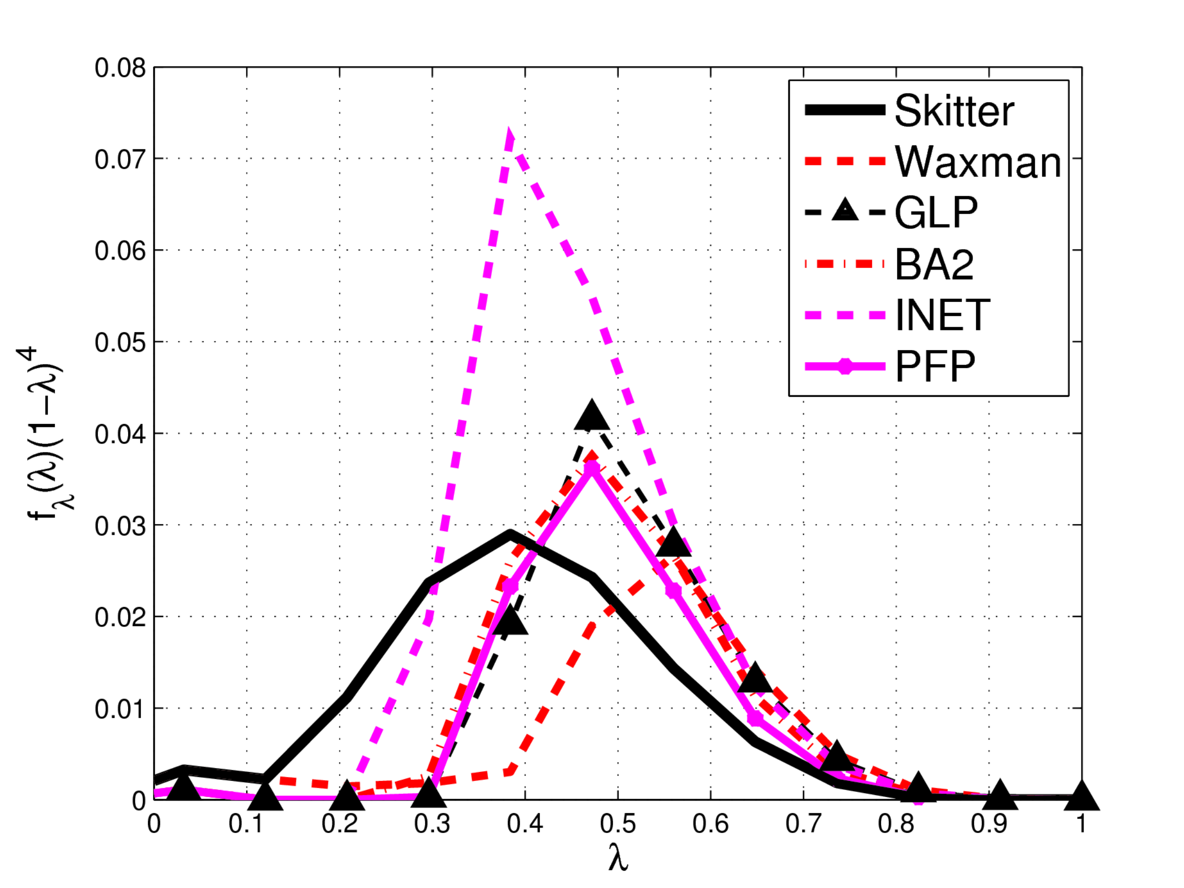


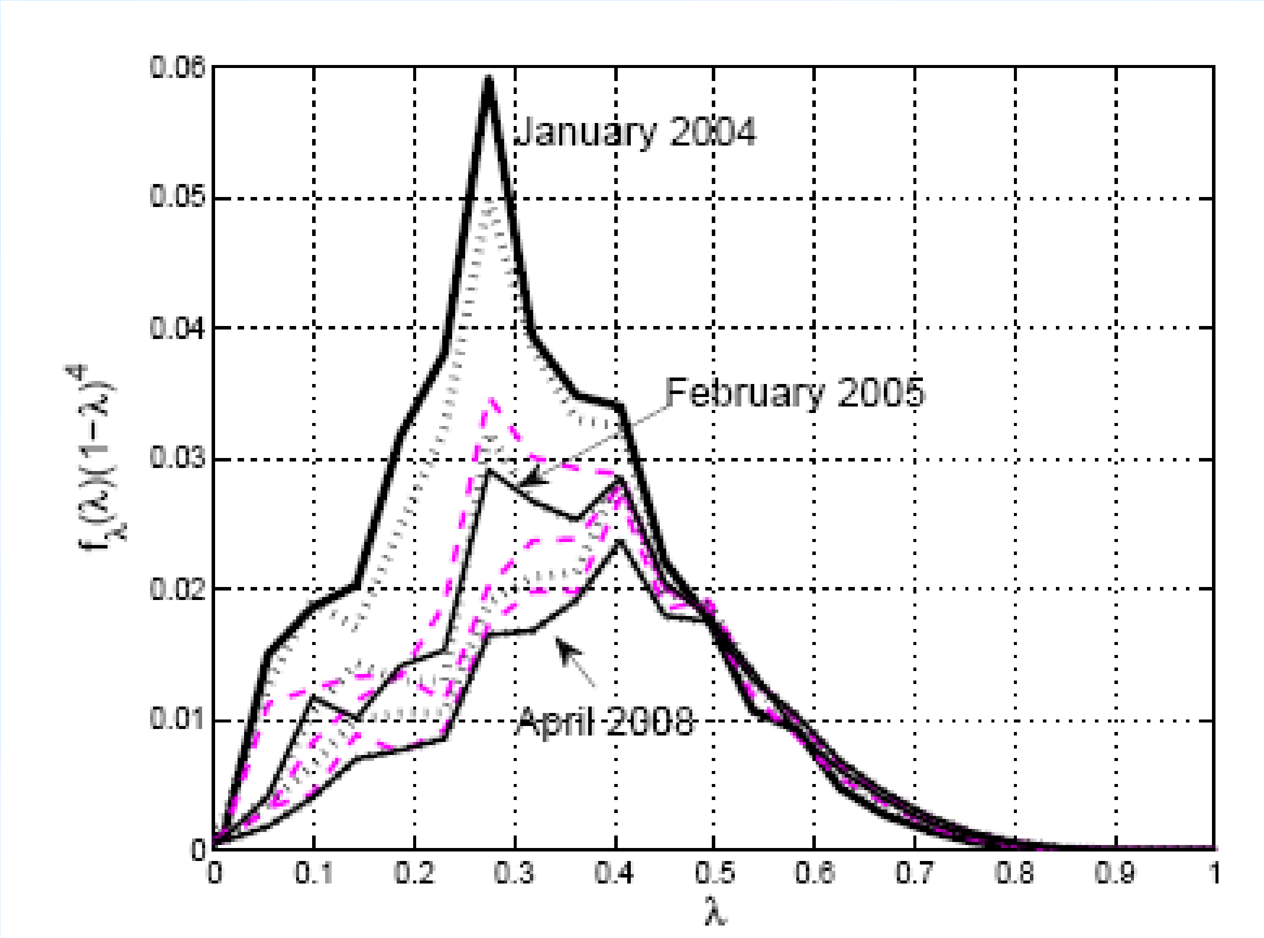
TABLE II

OPTIMUM PARAMETER VALUES FOR MATCHING SKITTER TOPOLOGY SAMPLED IN MARCH 2004.

Waxman	$\alpha = 0.08$ (default $= 0.15$)	$\beta = 0.08$ (default $= -0.2$)	$J(G_{Waxman}, G_{Skitter}) = 0.0026$
AB	$p = 0.2865$ (default $= 0.6$)	$q = 0.3145$ (default $= 0.3$)	$J(G_{AB}, G_{Skitter}) = 0.0014$
GLP	$p = 0.5972$ (default $= 0.45$)	$\beta = 0.1004$ (default $= 0.64$)	$J(G_{GLP}, G_{Skitter}) = 0.0021$
Inet	$\alpha = 0.1013$ (default $= 0.3$)	–	$J(G_{INET}, G_{Skitter}) = 0.0064$
PFP	–	–	$J(G_{PFP}, G_{Skitter}) = 0.0014$

Application 2: Tracking evolution of the internet.

We see quite clearly the change in the structure of the network.

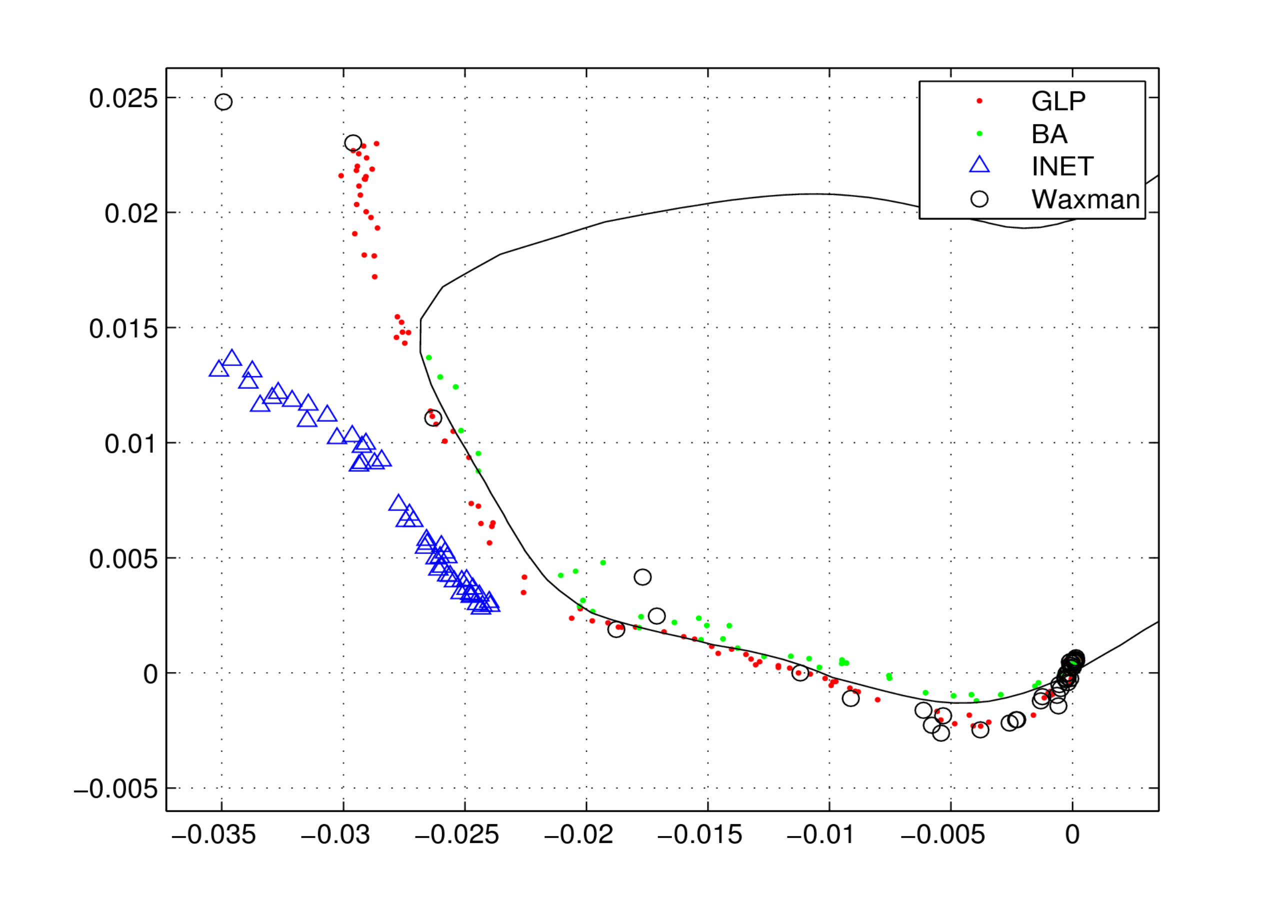


Example applications.

- ~~Estimating the optimum parameters for a topology generator.~~
  - ~~Comparing which topology generator produces a 'best' fit for the internet.~~
- ~~Tracking evolution of the internet.~~

- Clustering applications:
  - Discriminating between topology generators.
  - Network application identification.
  - Orbis model analysis.

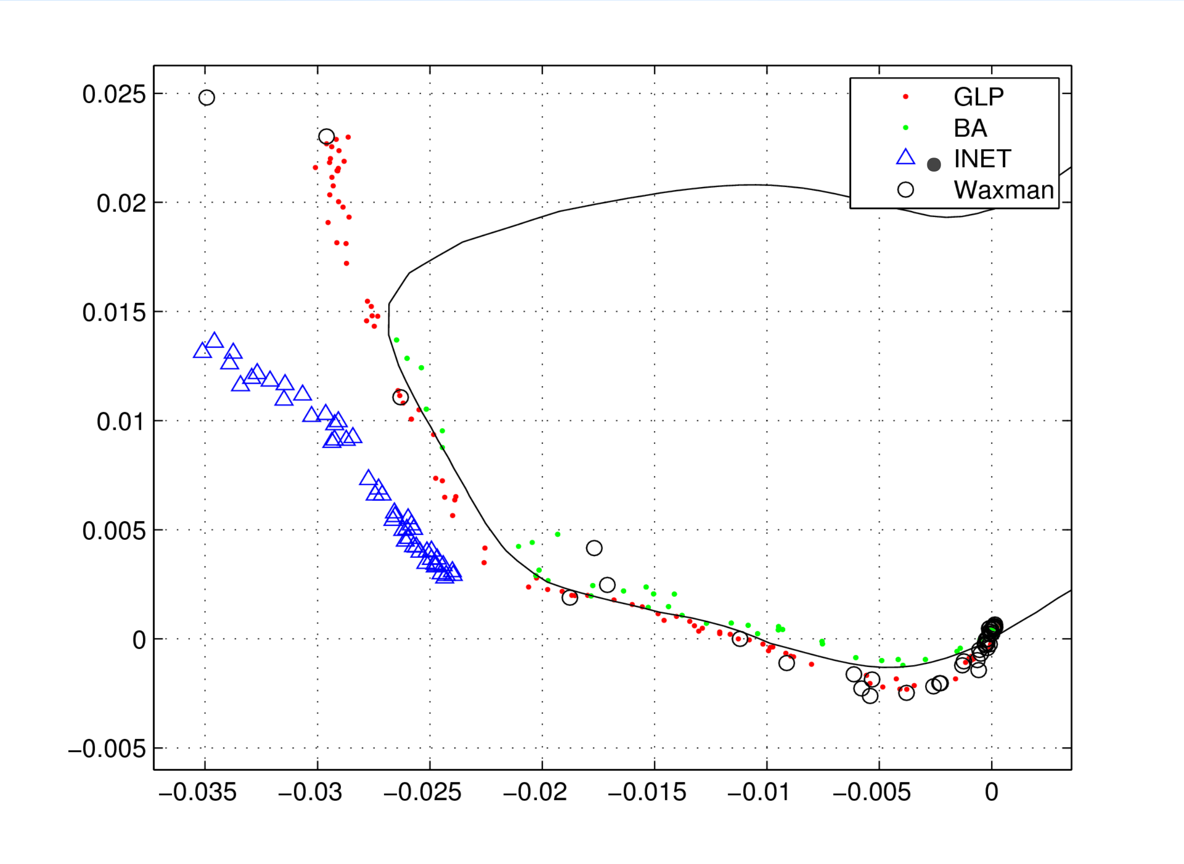
Application 3: Discriminating between topology generators.



500 graphs are generated of each:

- Waxman, GLP, AB, INET.
- Using random parameters.

Result



- The boundary between the AB and GLP models is pretty tight:

  A support vector machine is used (50-50 data split between training and test set); 11% misclassification error.

- INET occupies a different part of the projection space.
- Waxman maps to pretty much everywhere (not shown).
- Conclusion: The WSD+RP separates topologies with different structures.

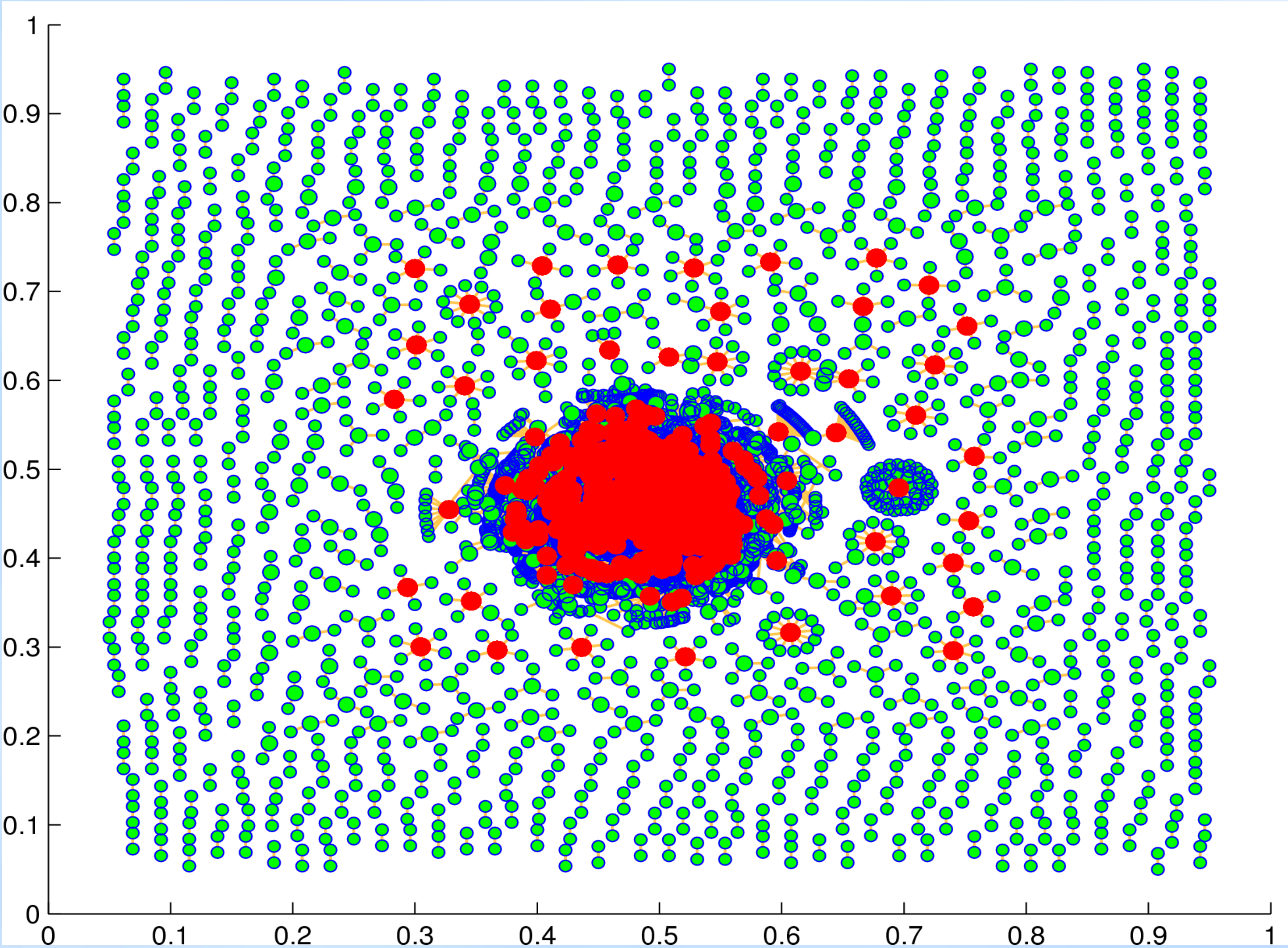
Application 4: Network application identification.

CAIDA data*

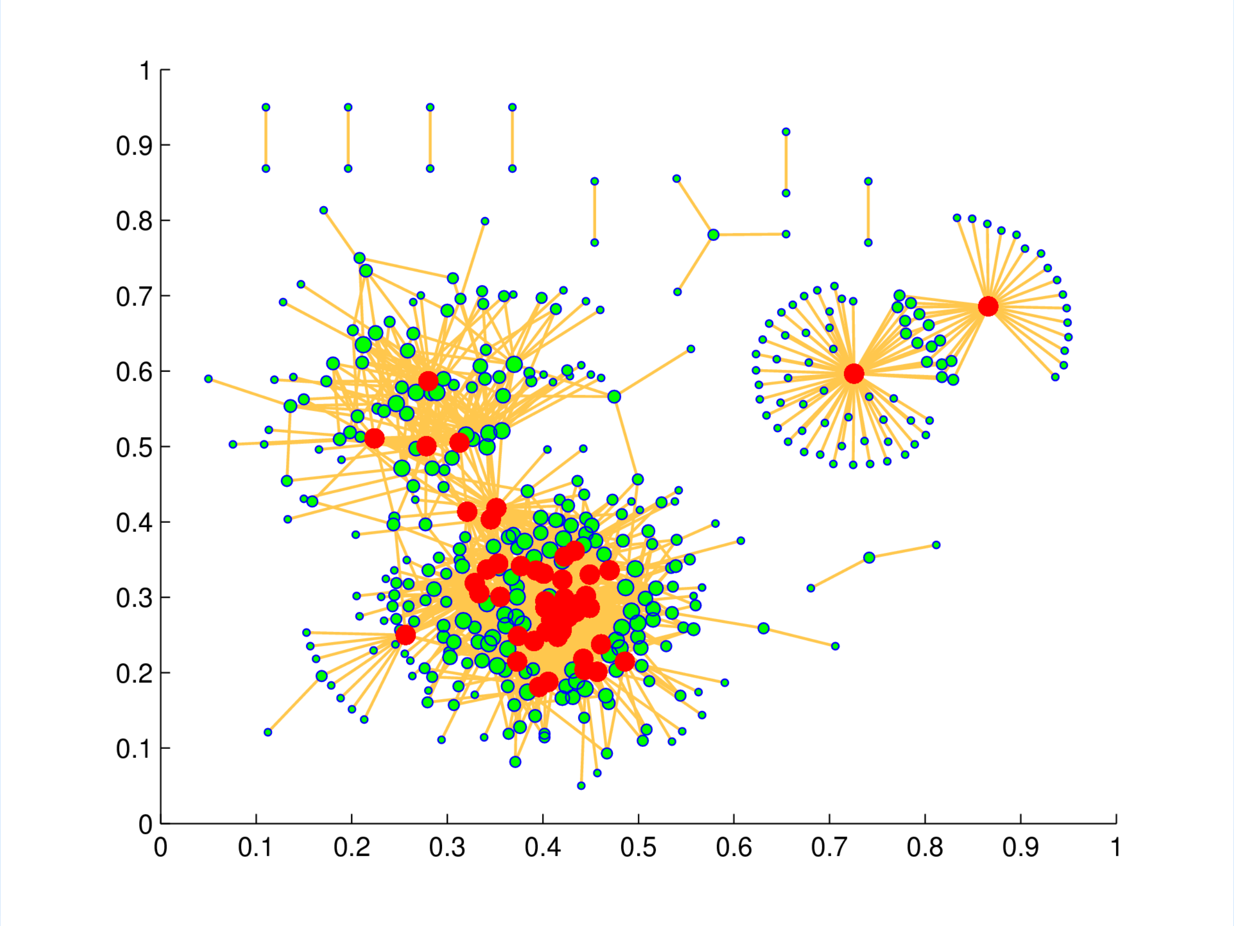
- Packet capture and deep packet inspection.
- 5 mins intervals.
- 10 Applications are tracked. Each forming a subgraph of interactions on the network (i.e. routers interacting with the same application).
- Aim: Given a graph can we estimate the application?.

*Exploiting Dynamicity in Graph-based Traffic Analysis: Techniques and Applications, Marios Iliofotou, M. Faloutsos, and M. Mitzenmacher, In ACM CoNEXT 2009, Dec.

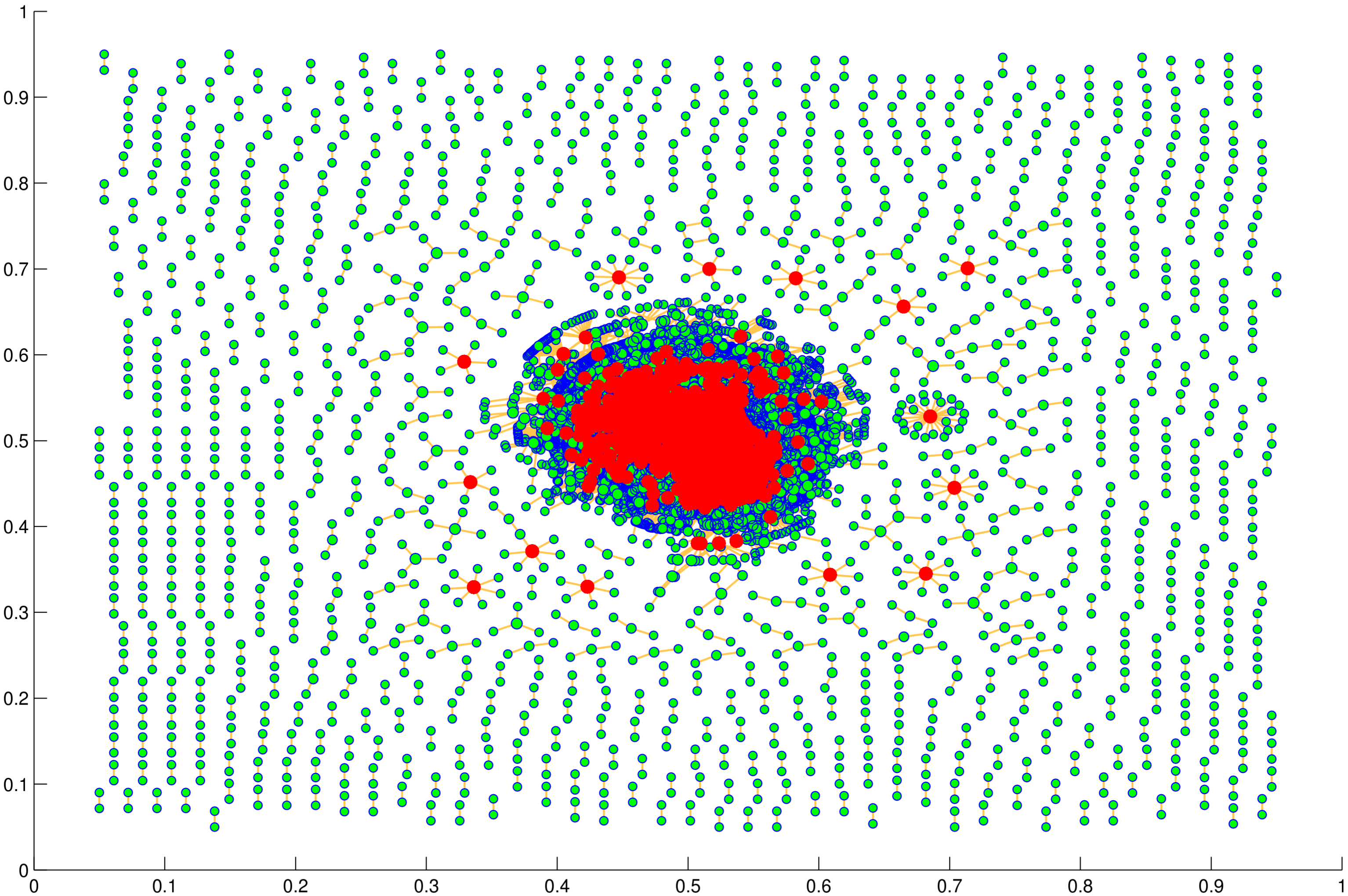
An application graph. (E-donkey)



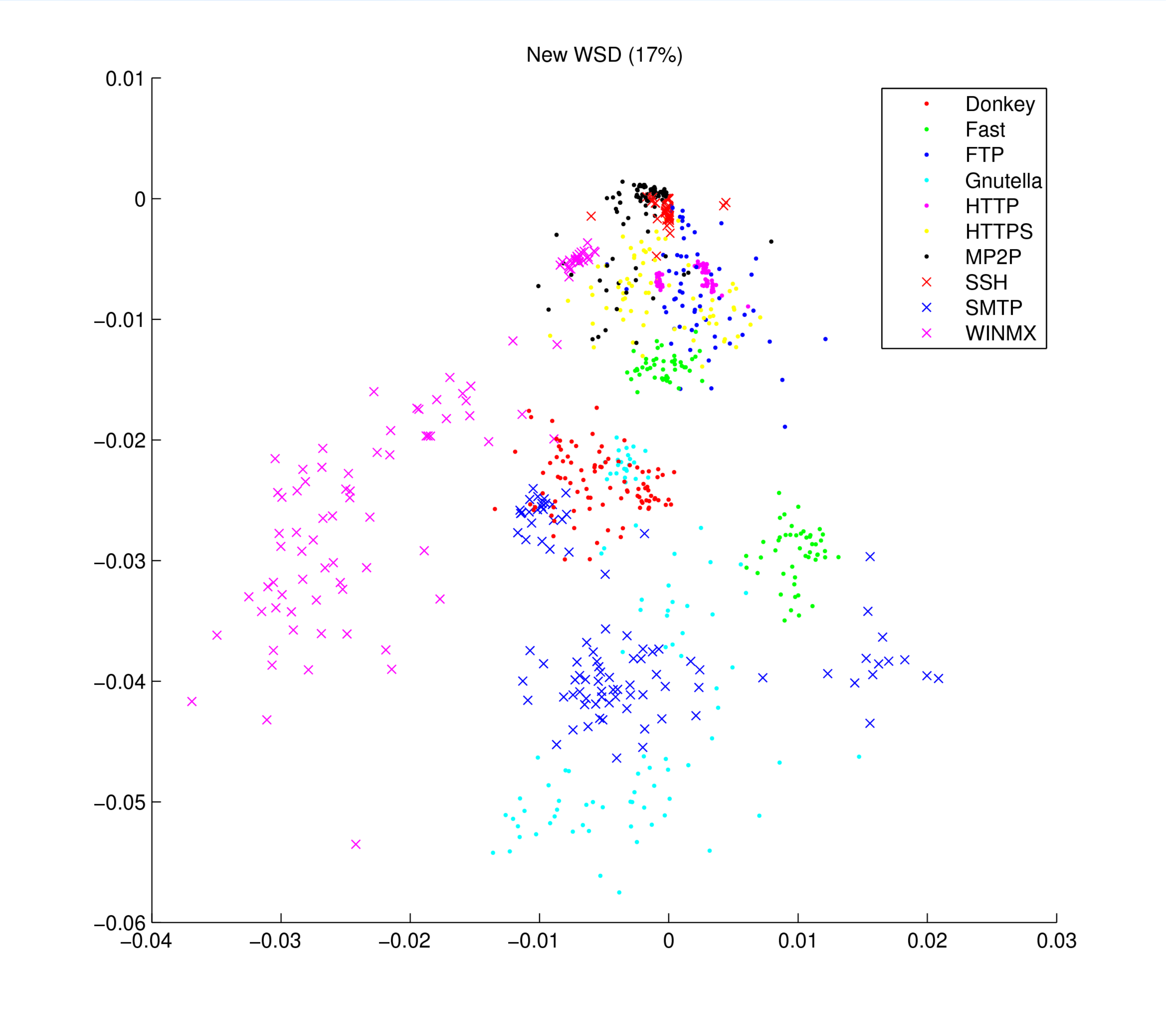
Example: MP2P



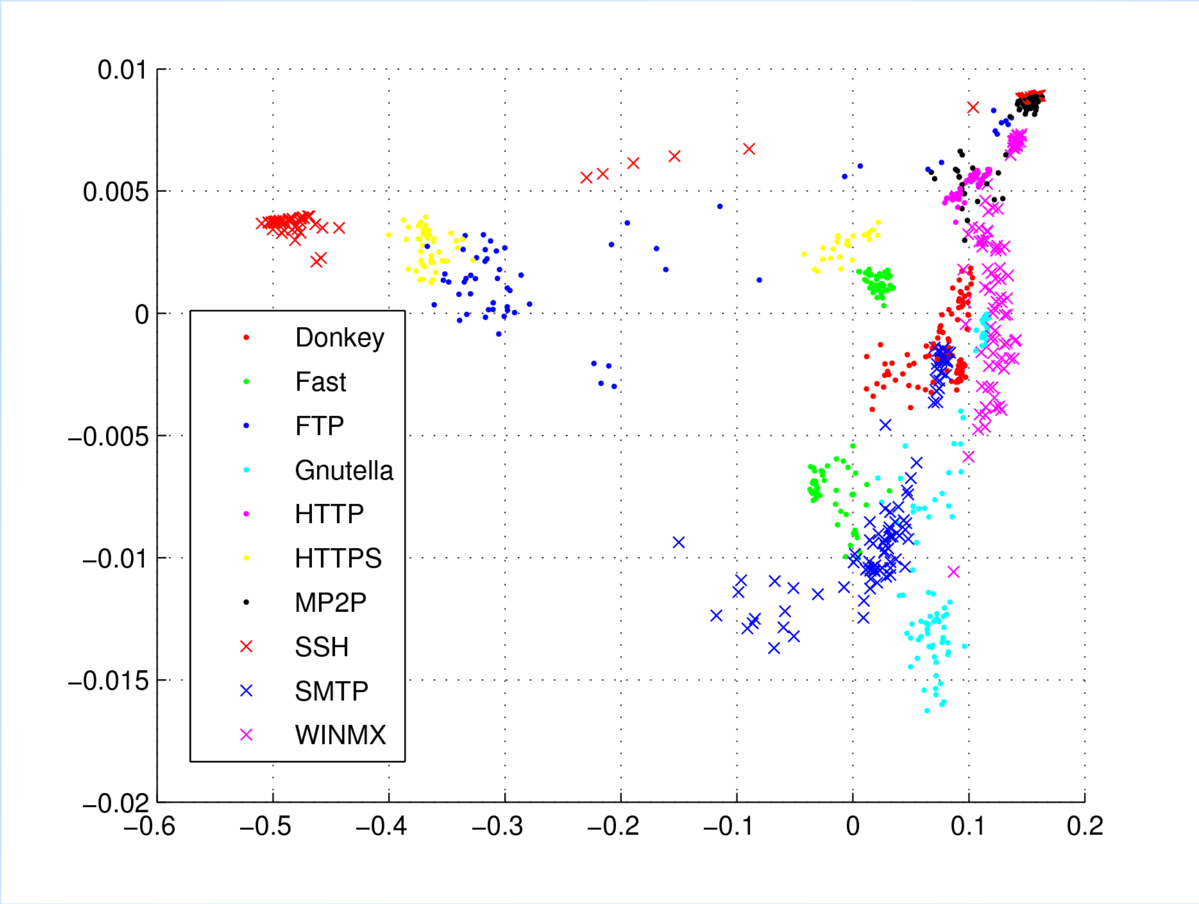
Example: SMTP



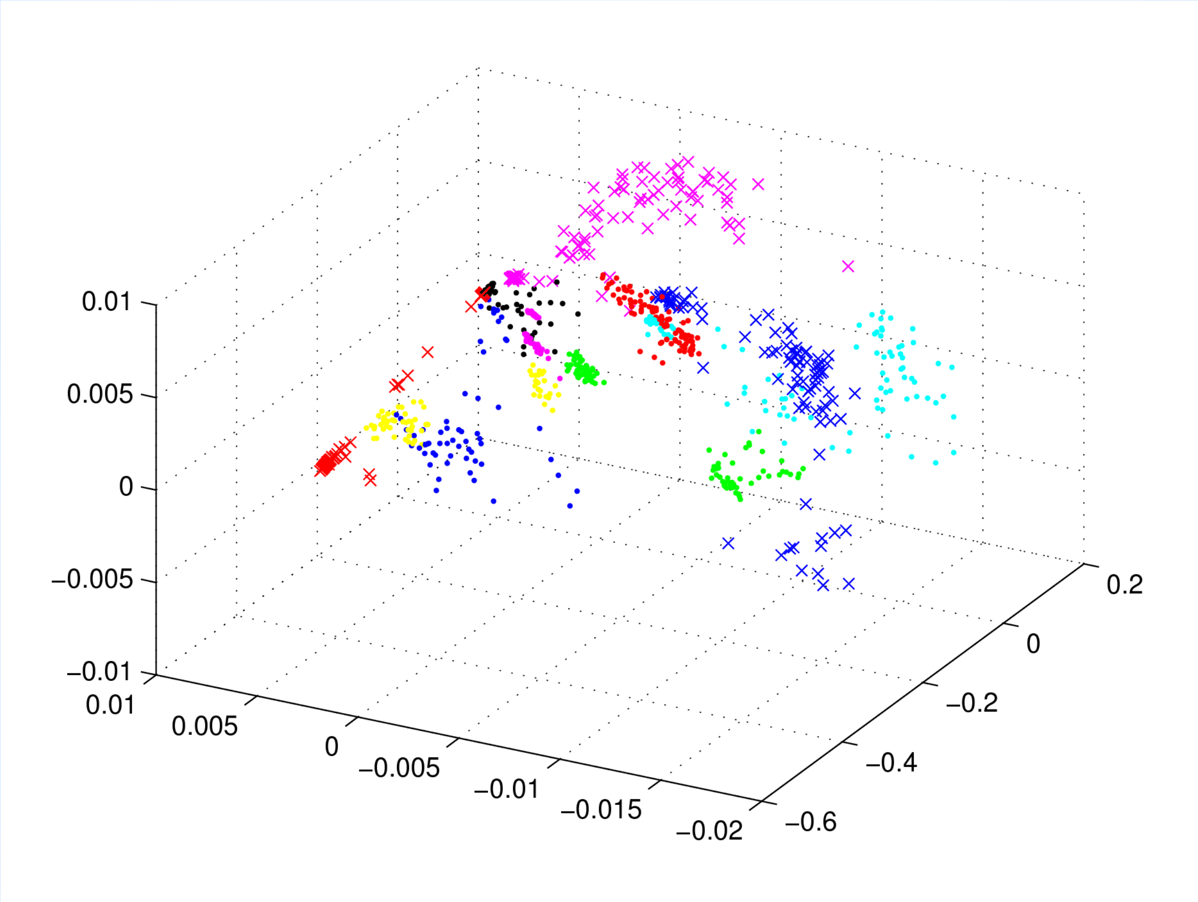
Result using random projection.



Result using MDS.



From a different angle.



Clustering results.

TABLE 1
Confusion matrix for Internet applications.

	Donkey	Fast	FTP	Gnutella	HTTP	HTTPS	MP2P	SSH	SMTP	WINMX
Donkey	**0.96**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00
Fast	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FTP	0.00	0.00	**0.67**	0.00	0.00	**0.20**	0.13	0.00	0.00	0.00
Gnutella	**0.22**	0.02	0.00	**0.74**	0.00	0.00	0.00	0.00	0.02	0.00
HTTP	0.00	0.00	0.06	0.00	**0.91**	0.03	0.00	0.00	0.00	0.00
HTTPS	0.00	0.03	0.03	0.00	0.00	**0.94**	0.00	0.00	0.00	0.00
MP2P	0.00	0.00	0.04	0.00	0.07	0.00	**0.89**	0.00	0.00	0.00
SSH	0.00	0.00	0.03	0.00	0.00	0.06	**0.31**	**0.60**	0.00	0.00
SMTP	0.17	0.04	0.00	0.00	0.00	0.00	0.00	0.00	**0.78**	0.00
WINMX	0.00	0.00	0.00	0.00	0.00	0.00	**0.22**	0.00	0.04	**0.74**

Application 5: The Orbis topology generator.

•

Orbis is a topology generator based on the *configuration model*.

- Paper: Sigcomm '06.
- Given a particular degree distribution we form link stubs:

The Orbis topology generator.

•

Connect Stubs at random until all connections are complete.

- This is known as the **1k** model and is defined solely by a degree distribution.
- A **0K** model is defined solely by the average degree.

The Orbis topology generator.

•

Connect Stubs at random until all connections are complete.

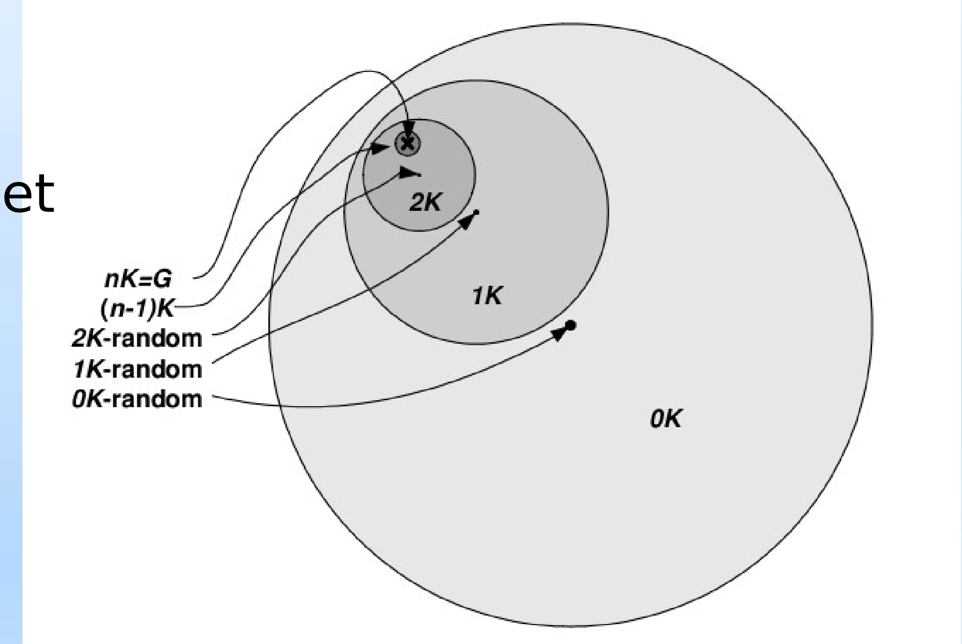
- A **2k** model and is defined by the joint-degree distribution and connects stubs based on the samples from this distribution:
- i.e. a node with degree 3 may connect to a node with degree 2 and so on.

The dK series.

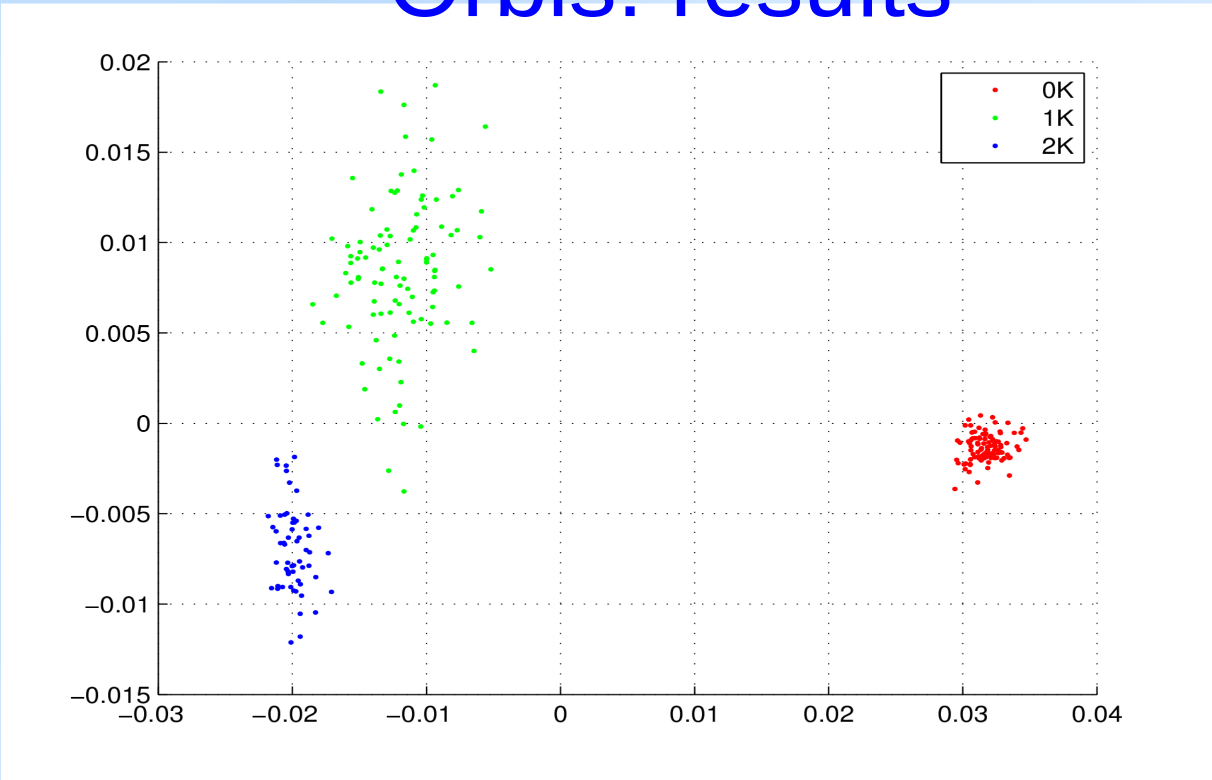
- The **0k, 1k, 2k** form a series called the **dK** series which specifies a subset of random graphs in which

  nk ⊂ ...⊂ **2k** ⊂ **1k** ⊂ **0k**

- This was shown diagrammatically in the Sigcomm '06 paper.
- **Aim**: To validate this subdivision of graphs using WSD and MDS.

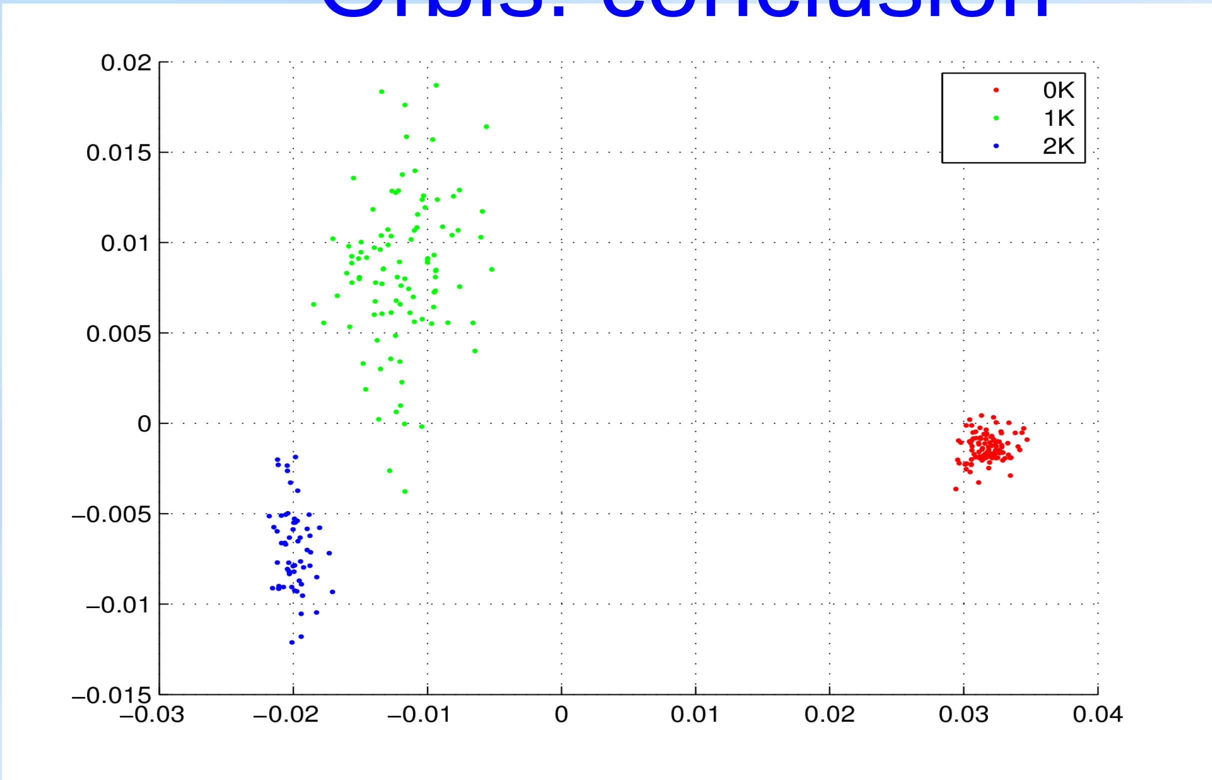


Orbis: results



- An AB model with 5000 nodes and average degree 5.6 is used as the base model for extracting an average degree, a degree distribution and a joint degree distribution.
- 100 graphs from the following ensembles are sampled:
- **0k** ensemble (average degree 5.6)
- **1k** ensemble (power law distribution)
- **2k** ensemble (no name).

Orbis: conclusion



Result:

- **0k –** very narrow set of graphs generated (ER – narrow degree distribution)
- **1k** ensemble (larger set of graphs)
- **2k** ensemble – the joint degree distribution of an AB model is not even close to the average of a **1K** specified by a degree distribution.

Application 6: Facial expression recognition.

•

Yale face dataset.

- 15 subjects performing 11 facial expressions.
- Image:

→ resize 256x256

→ quadtree decomposition

→ graph

→ WSD

→ MDS

→ clustering.

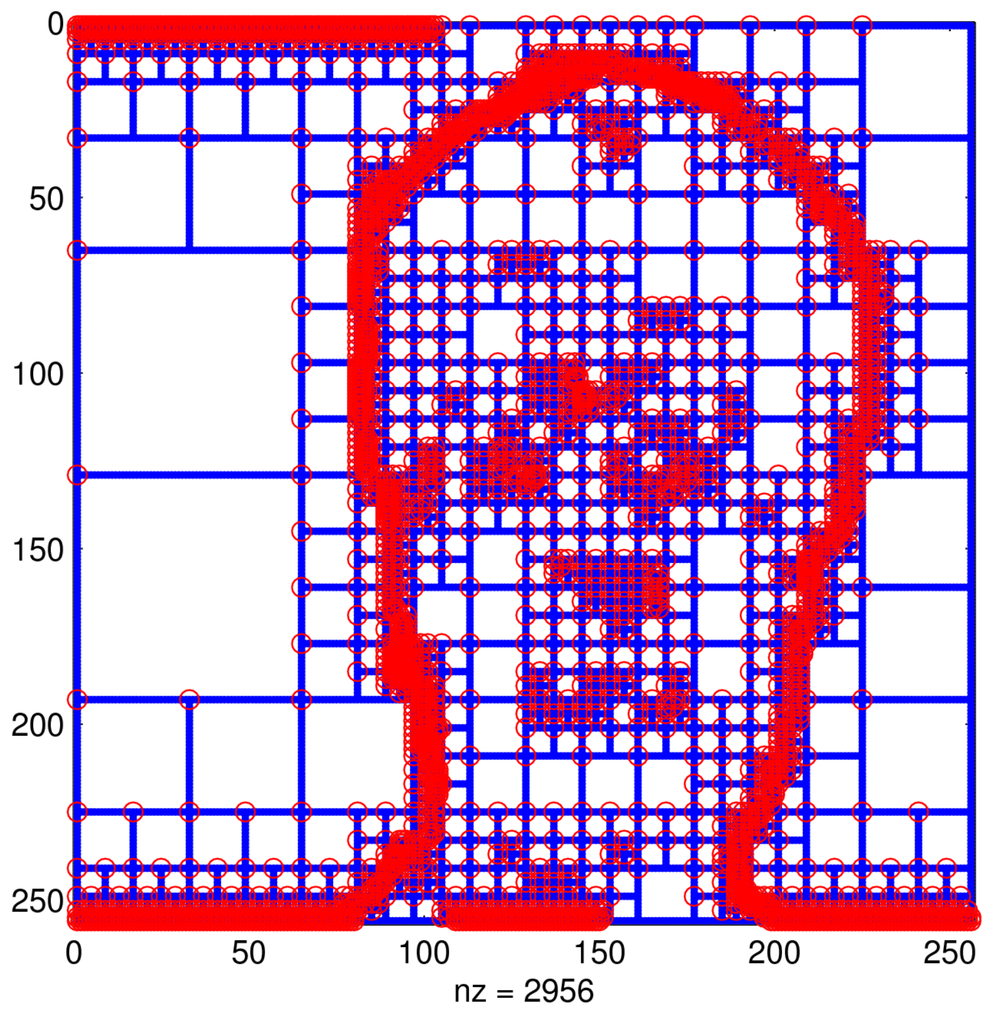
Original image.

•

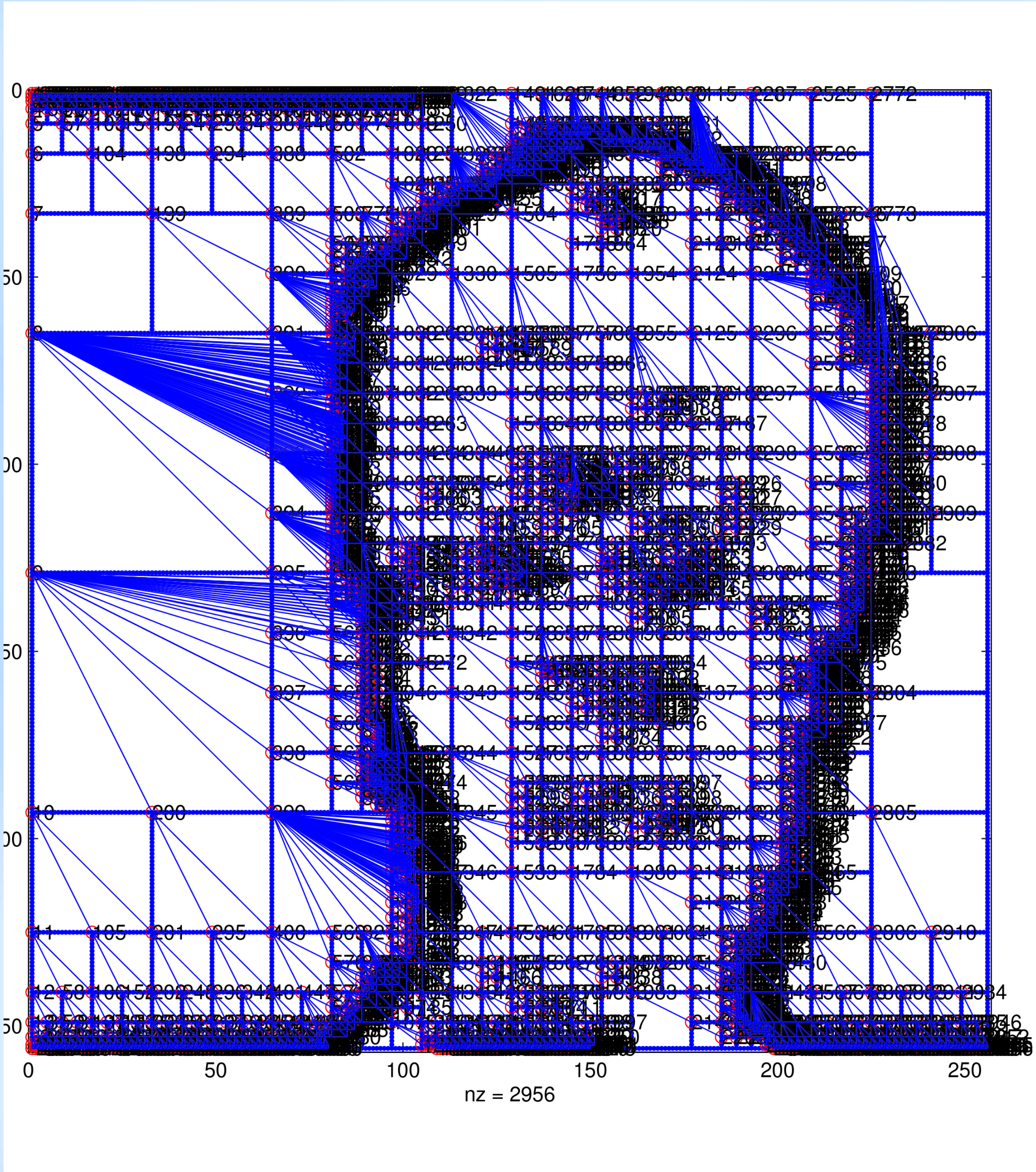
Resized image.

Quadtree decompostion.

•

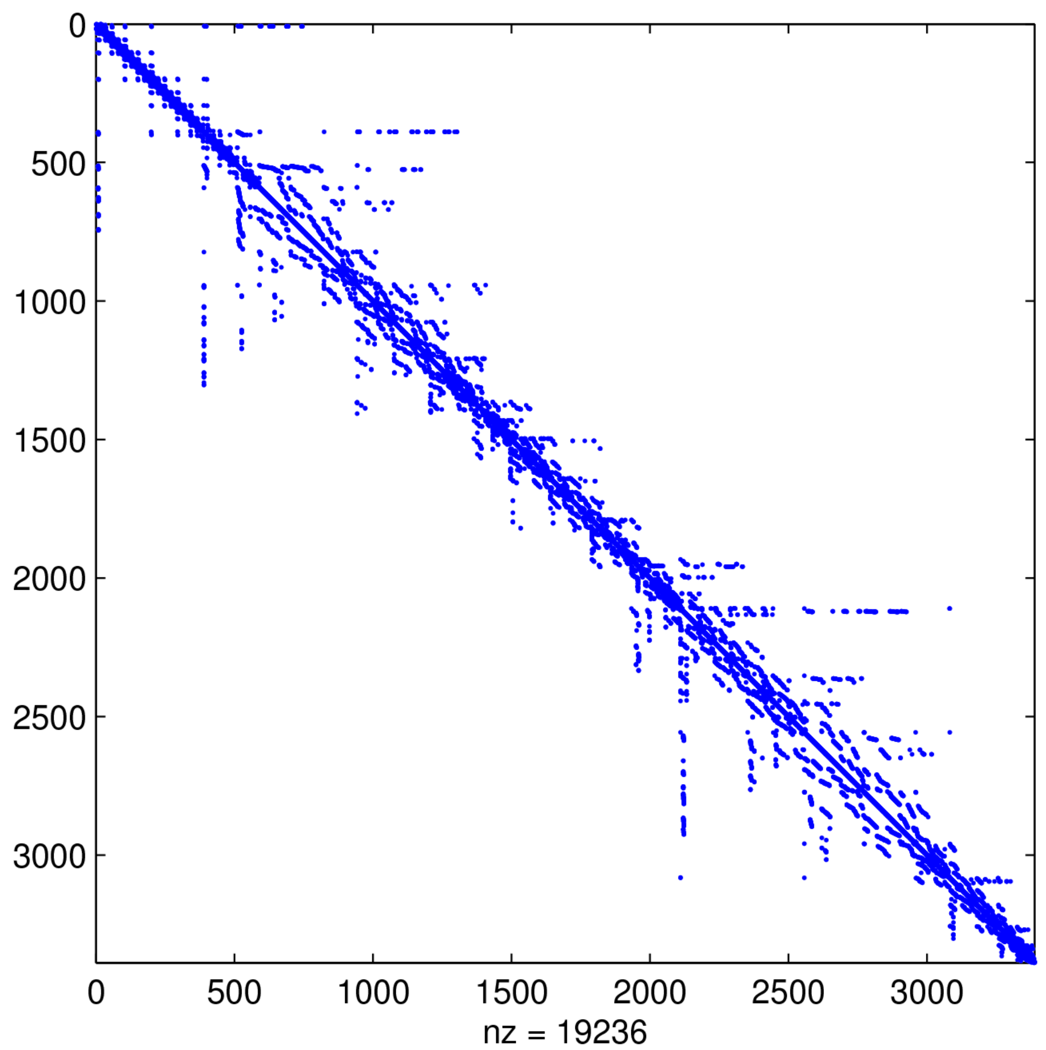


Graph construction.



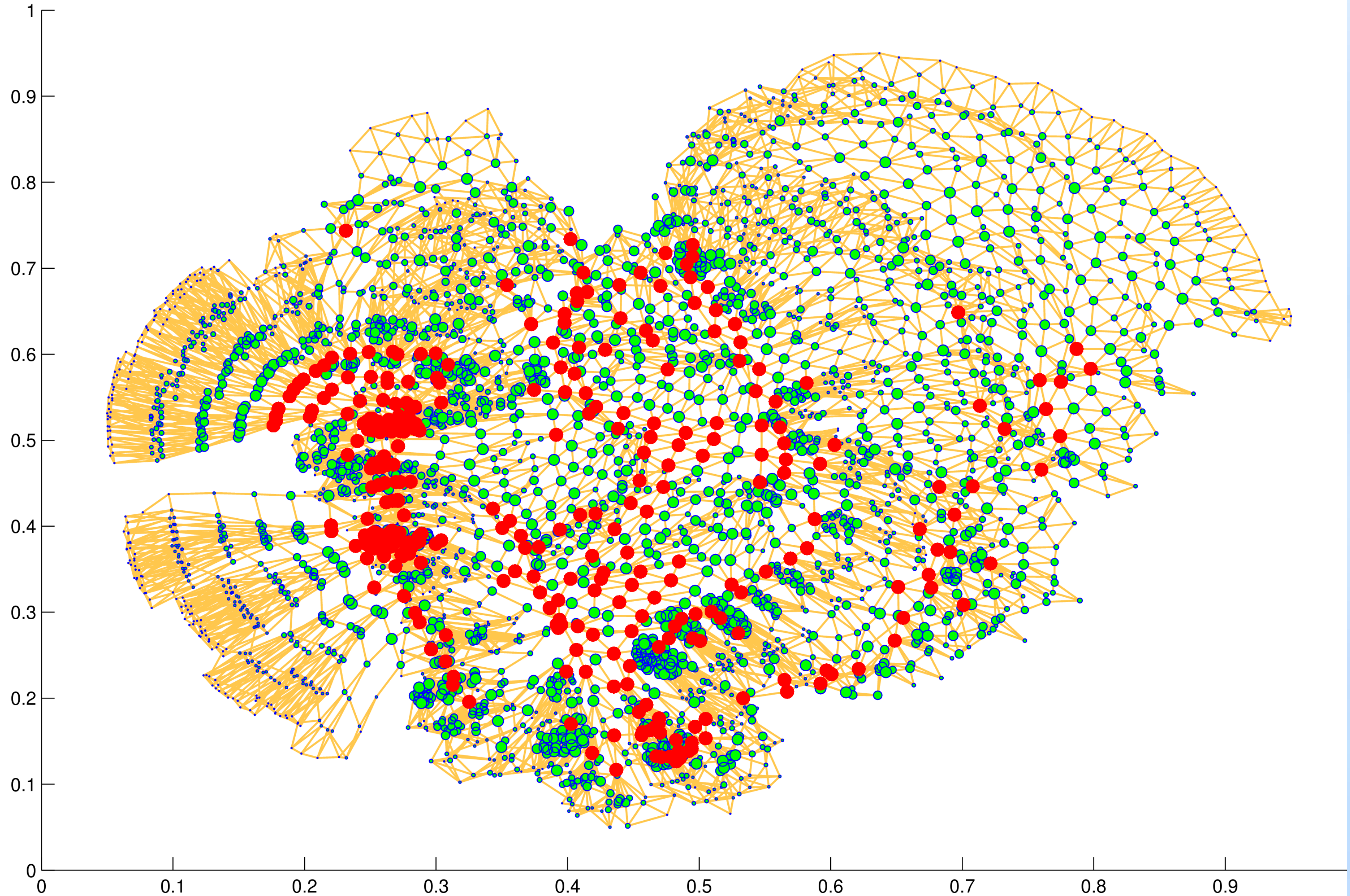
UNIVERSITY OF CAMBRIDGE

(weighted) Adjacency matrix.



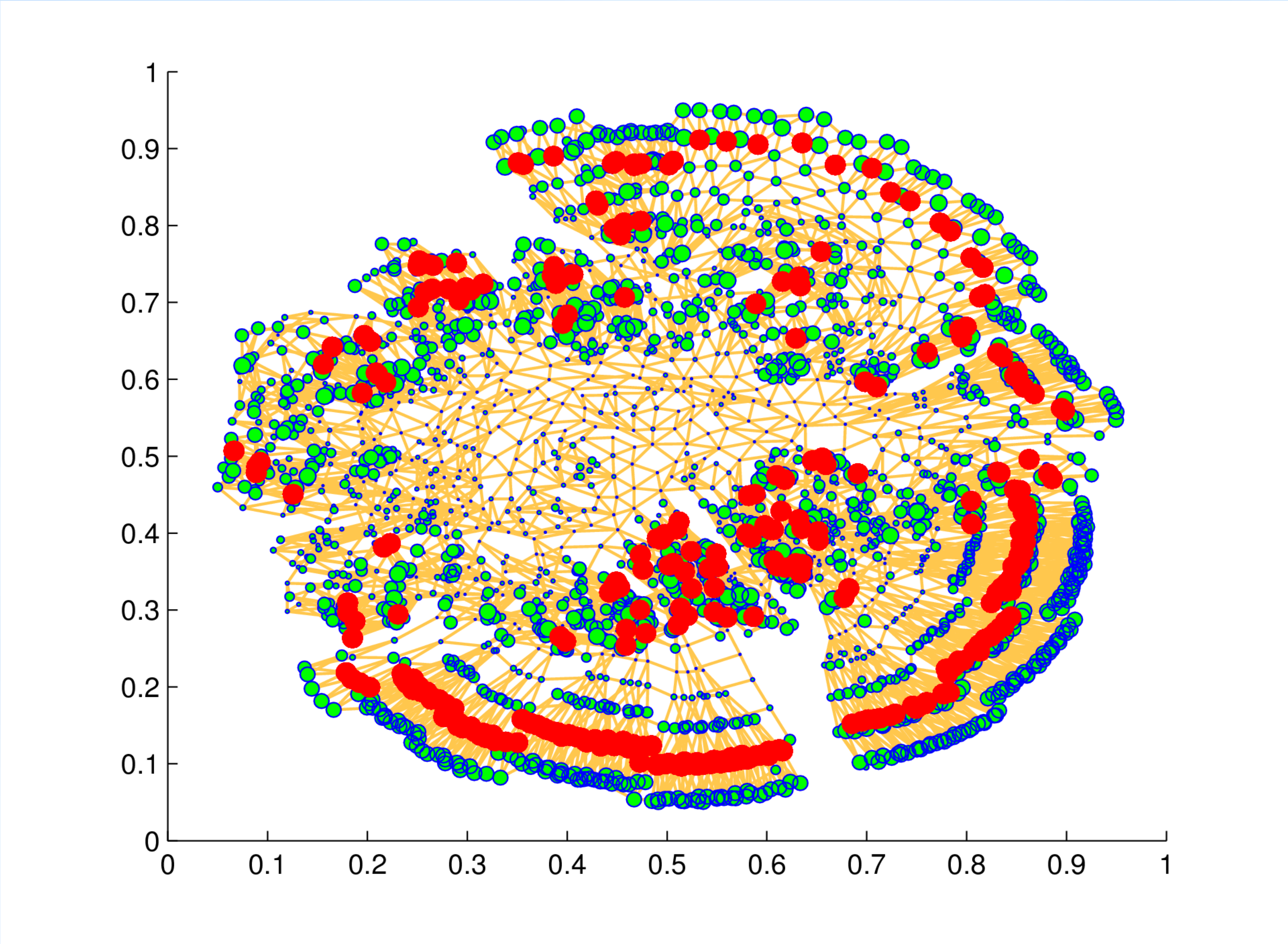
UNIVERSITY OF CAMBRIDGE

Weighted graph (face 1).



Weighted graph (face 10).

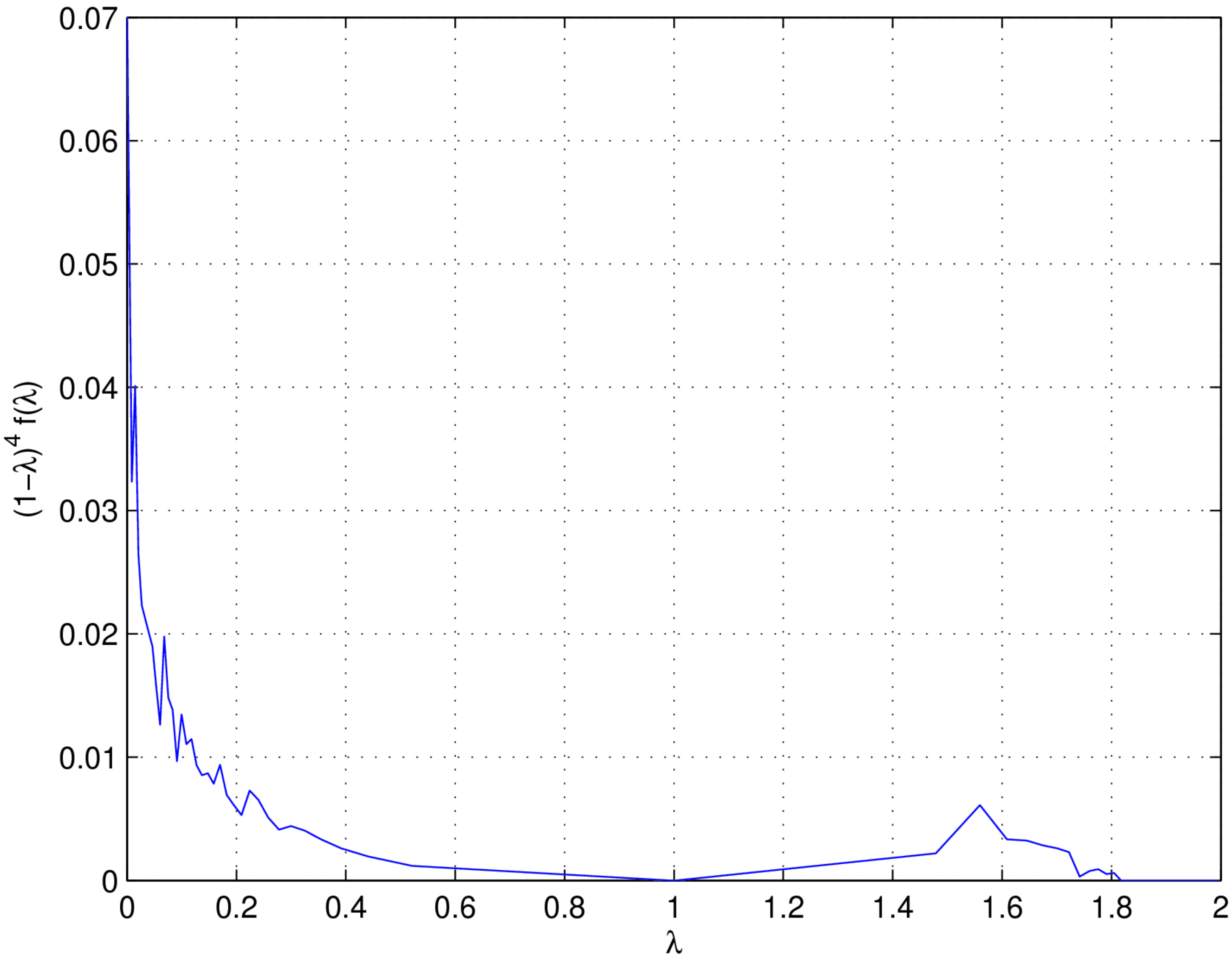
•



UNIVERSITY OF CAMBRIDGE

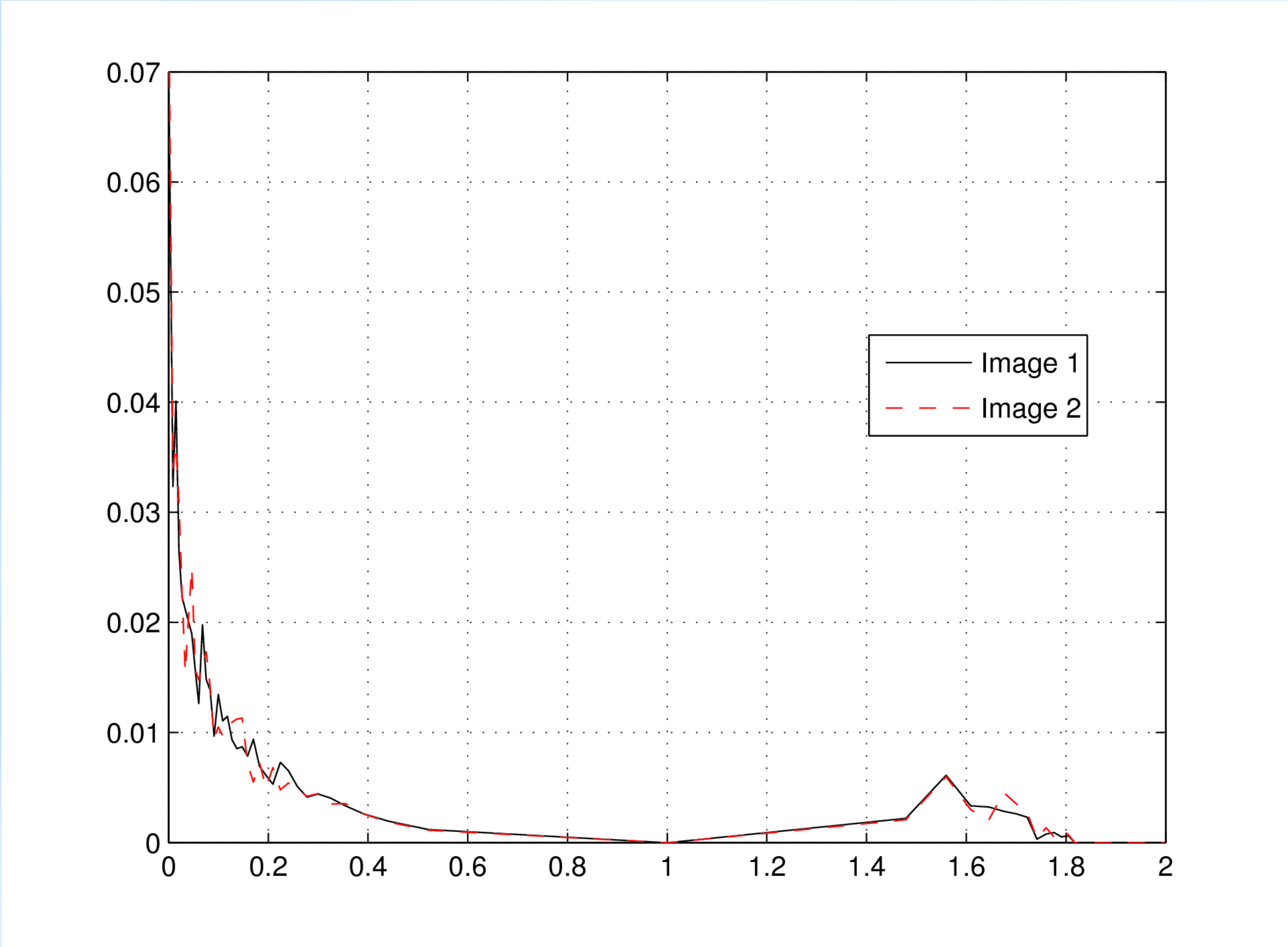
WSD.

•

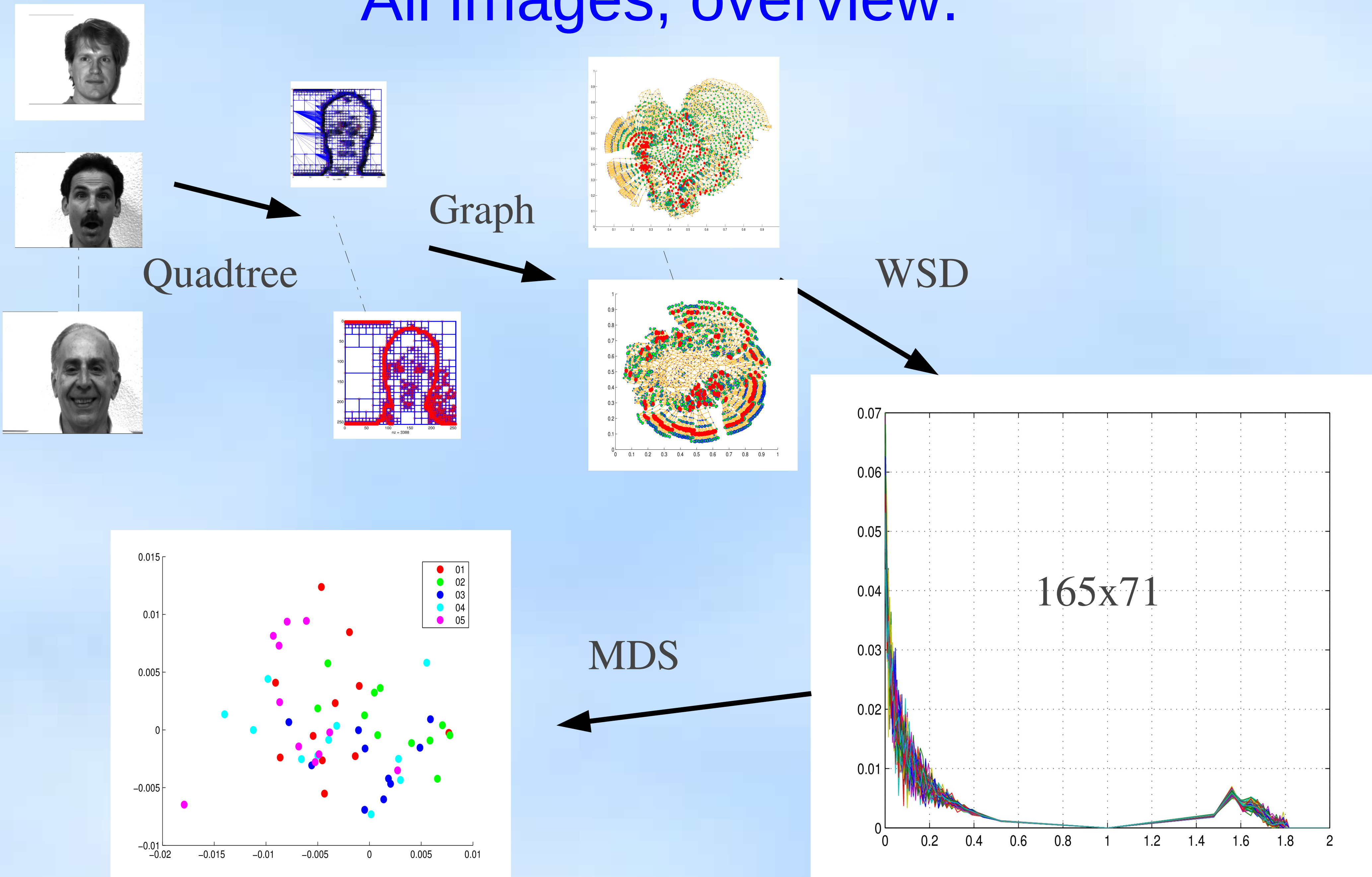


0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
$(1-\lambda)^4 f(\lambda)$
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
λ

WSD's of two images.



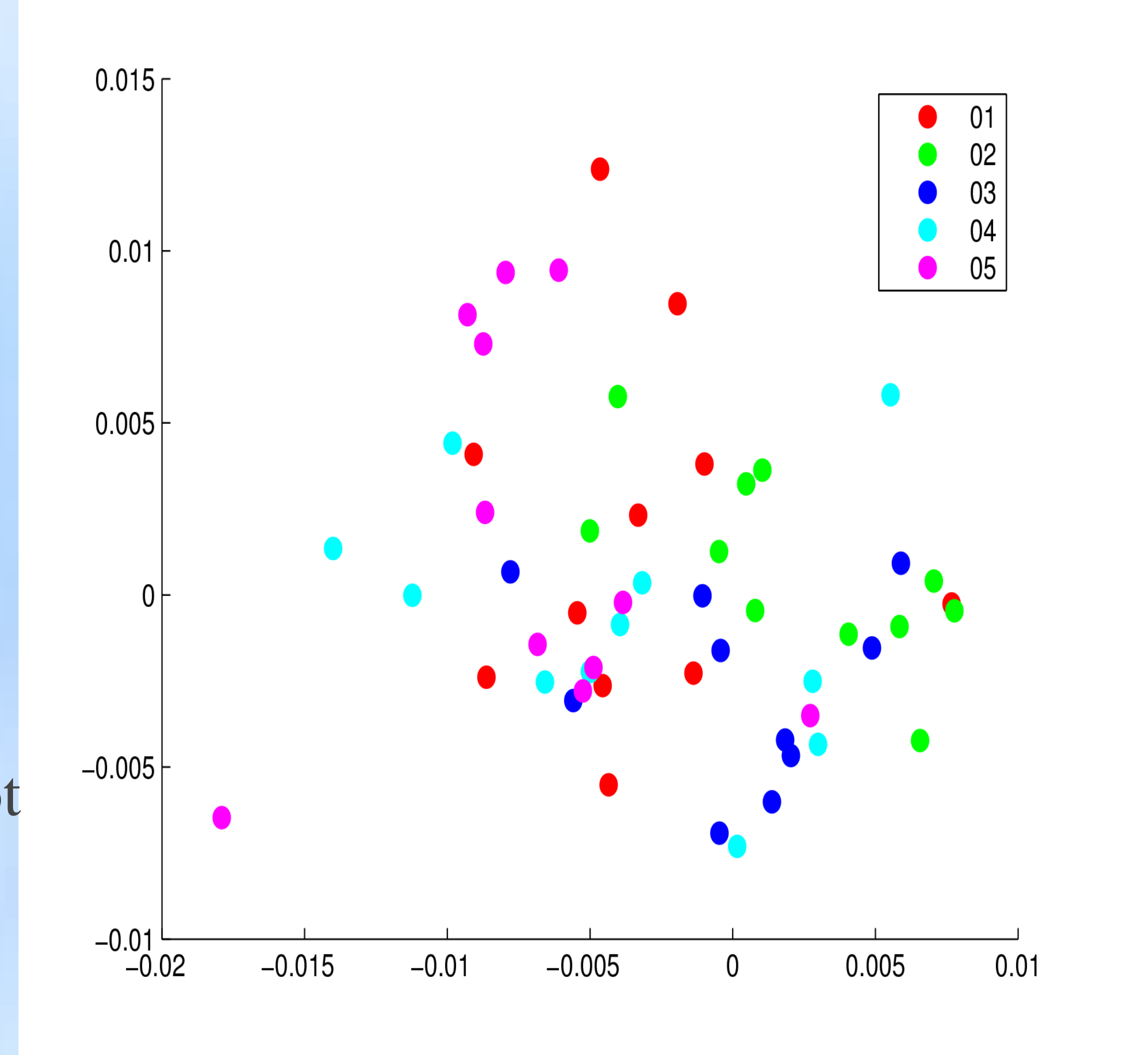
All images, overview.



UNIVERSITY OF CAMBRIDGE

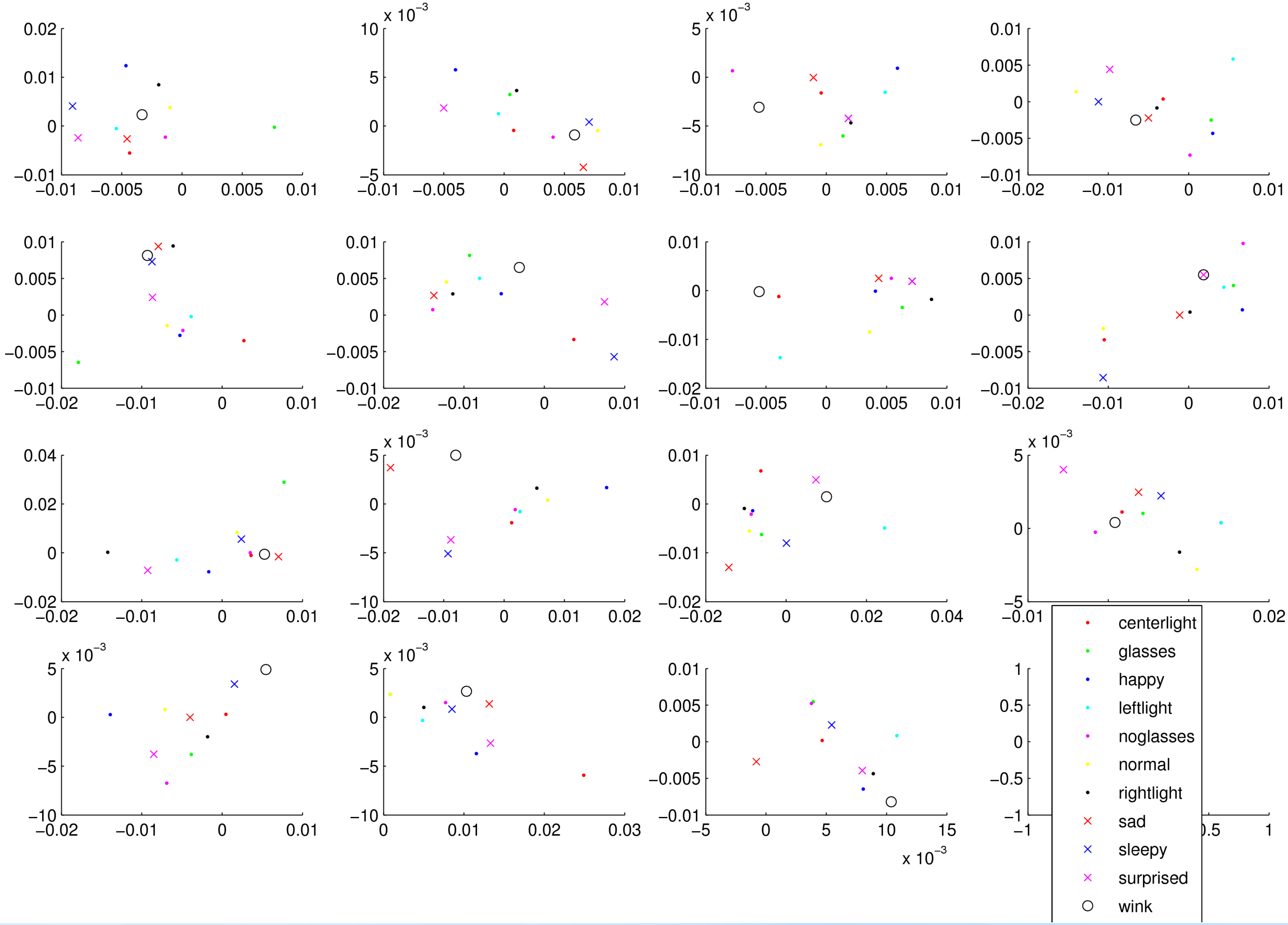
5 subjects all facial expressions.

- Only 5 subjects shown for clarity.
- The separation between subjects is not very good.
- Clusters do exist.
- Semantic information is missing.
- Shadow effects introduce a lot of noise.



UNIVERSITY OF CAMBRIDGE

Breakdown by expression.



Summary.

- A graph metric, the WSD was introduced.
  - Based on N-cycles
  - Normalised Laplacian eigenvalue distribution.
  - Weighted distribution.
  - Quadratic norm between weighted distributions forms a graph metric.
- *Model* fitting applications:

  Optimum parameters for a topology generator.
  - Tracking evolution of the internet.
- Clustering applications:
  - Discriminating between topology generators.
  - Network application identification.
  - Orbis model analysis.
  - Face recognition*

UNIVERSITY OF CAMBRIDGE

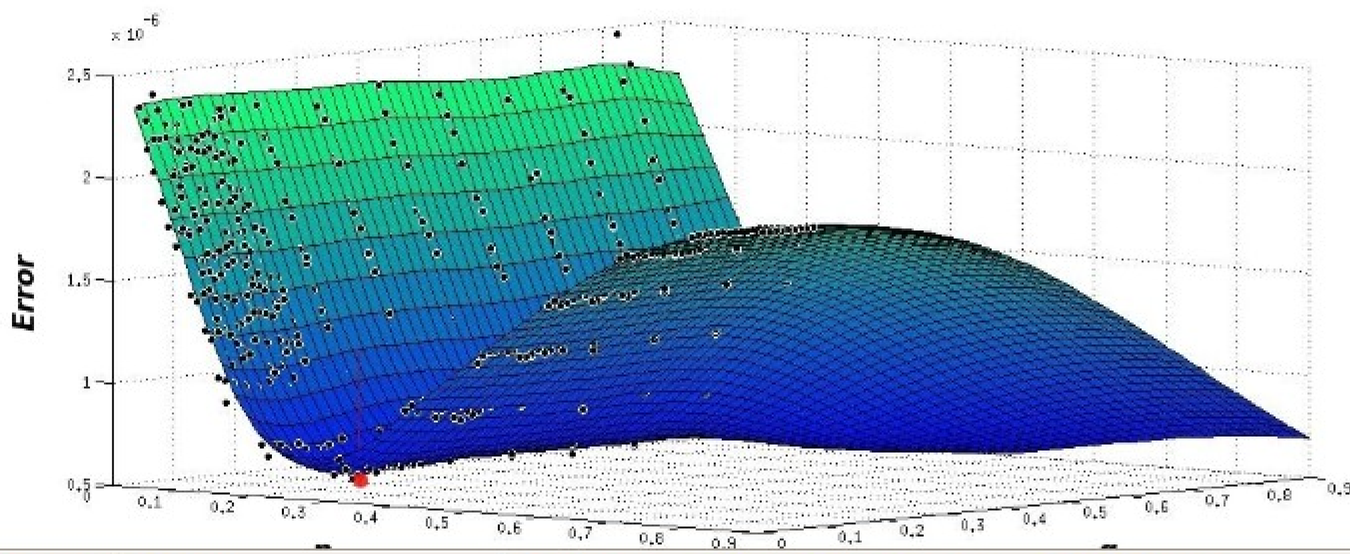
Questions ?

UNIVERSITY OF CAMBRIDGE

Computer Laboratory
Weighted Spectral Distribution Toolbox

Computer Laboratory > Research > Systems Research Group > NetOS > MASTS > Weighted Spectral Distribution Toolbox

Weighted Spectral Distribution Toolkit



Fay, D. , Haddadi, H., Moore, A. W., Uhlig, S., Tomason, A., Mortier, R., A weighted spectral distribution: theory and applications, *IEEE trans on Networking*, to appear.

Fay, D. Haddadi, H., Moore, A., Mortier, R., Uhlig, S., Jamakovic, A. "A Weighted Spectrum Metric for Comparison of Internet Topologies", *ACM SIGMETRICS PER Performance Evaluation Review*, December 2009.