

Weighted Spectra Distribution for Comparison of Internet Topologies

Hamed Haddadi
University College London
London, United Kingdom

Steve Uhlig, Almerima Jamakovic
Delft University of Technology
Delft, The Netherlands

Damien Fay
Computer Laboratory
University of Cambridge

Andrew Thomason
Mathematics department
University of Cambridge

Andrew W. Moore
Computer Laboratory
University of Cambridge

Richard Mortier
Vipadia Ltd
Cambridge, United Kingdom

Miguel Rio
University College London
London, United Kingdom

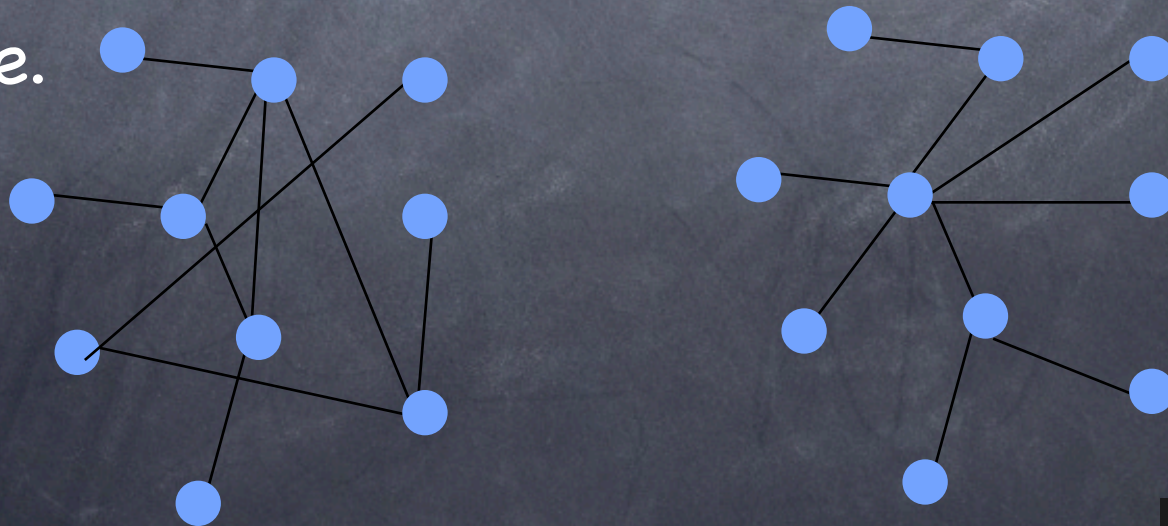
No Researcher is an island!

Presentation Outline

- Problem description
- Overview: the graph spectrum.
- Metric development.
- Applications.
 - Straw-man model.
 - Tuning topology generators.
- Conclusions.

Overall Goal

- How 'similar' are these two graphs/topologies?
- How do we define similarity:
 - Link count,
 - Node degree distribution,
 - Centrality,
 - Edit distance.

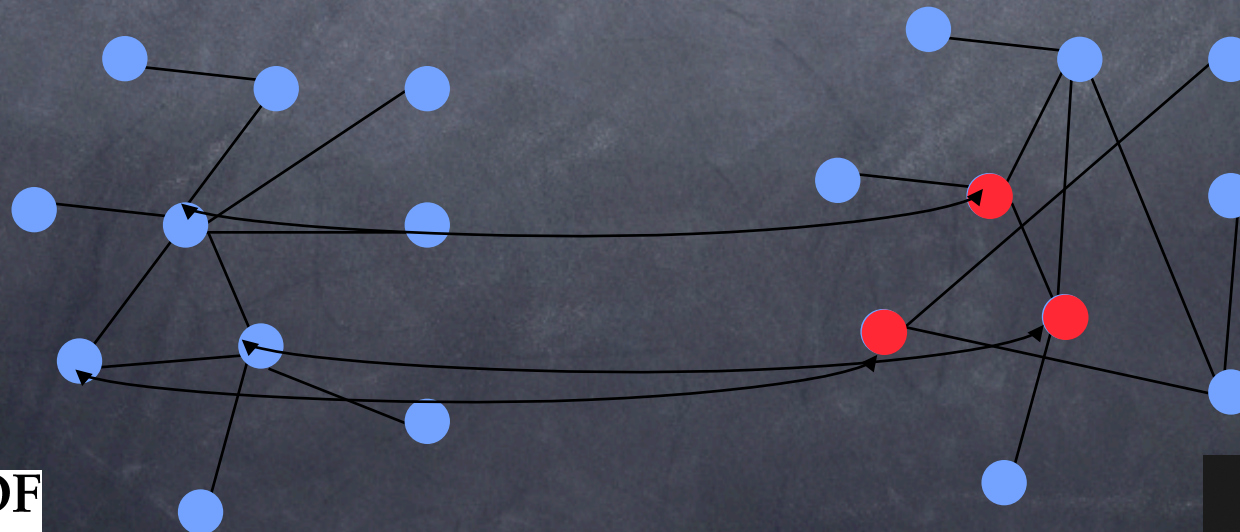


Overall goal

- The graphs we examine are different:
- we seek to generate 'internet like' topologies not internet topologies.
- Determine how the distribution of clusters are different in two graphs.
- The exact location of clusters is not important.
- Given a metric this can then be used to determine optimum parameters of a topology generator.
- Validation: it is difficult to validate a metric as this would require a metric!

Related work

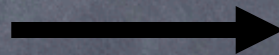
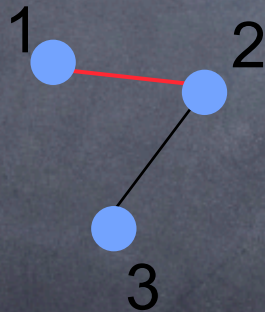
- Cluster analysis: A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*14. MIT Press, 2002
- Graph matching: B. Luo and E. Hancock. Structural graph matching using the em algorithm and singular value decomposition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(10):1120–1136, Oct 2001.
- S. Hanna. Representation and generation of plans using graph spectra. In 6th International Space Syntax Symposium, Istanbul, 2007.
- E. W. Zegura, K. L. Calvert, and M. J. Donahoo. A quantitative comparison of graph-based models for Internet topology. *IEEE/ACM Transactions on Networking (TON)*, 5(6):770–783, (1997).



Graph Spectra.

- The adjacency matrix:

$$A(G) = \begin{cases} 1 & u, v \text{ are connected} \\ 0 & u, v \text{ not connected} \end{cases}$$



$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- Symmetric. (bi-directional) Fully connected.

Graph Spectra.

- Associated matrices:
 - Computational Laplacian
 - $L(G) = D - A$
 - Normalised Laplacian
 - $(G) = D^{-1/2} L(G) D^{-1/2}$
 - Walk Laplacian
 - $rw(G) = D^{-1} L(G)$

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

Example

$$\begin{array}{ccc} \mathbf{A(G)} & \mathbf{D} & \mathbf{L(G) = D - A} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \end{array}$$

$$\mathfrak{L}(G) = \mathbf{D}^{-1/2} \mathbf{L(G)} \mathbf{D}^{-1/2}$$

$$\begin{bmatrix} 1 & -1 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & 1 & 1 \end{bmatrix}$$

The weighted spectrum

The weighted spectrum is defined as:

$$W(G, N) = \sum_i (1 - \lambda_i)^N \quad N \text{ an integer}$$

Calculating the spectrum is very computationally expensive:

The global structure of a graph is required.

Not all the eigenvalues are required.

Just their distribution $f\lambda(\lambda)$ (estimated in K bins)

Calculated using pivoting and Sylvester's Law of Inertia.

$$D(G_1, G_2) = \sum_{k=1}^K (1 - k)^N (f_1(\lambda = k) - f_2(\lambda = k))$$

Technical development

- Starting with the normalised Laplacian:

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

- and writing B for the matrix $D^{-1/2}L(G)D^{-1/2}$ we have from the definition of (G) that the entries of B are given by:

$$(D^{-1/2}AD^{-1/2})_{i,j} = \frac{A_{i,j}}{\sqrt{d_i}\sqrt{d_j}}$$

and so the M^{th} moment of B is just $\text{tr}B^N$.

Writing b_{ij} as the elements of B , B^N is the sum of the products:

$$b_{i_0,i_1}b_{i_1,i_2}\dots b_{i_{N-1},i_N}$$

where $i_0 = i$ and $i_N = j$.

$$\sum_i (1 - \lambda_i)^N = \text{tr} B^N = \sum_C \frac{1}{d_{u_1} d_{u_2} \dots d_{u_N}}$$

Left hand side is the eigenvalue decomposition of B^N :

- First eigenvalue represents the strongest direction in the original matrix:

– Reconstruction error:
$$\sigma_i^2 \propto \frac{1}{|1 - \lambda_i|}$$

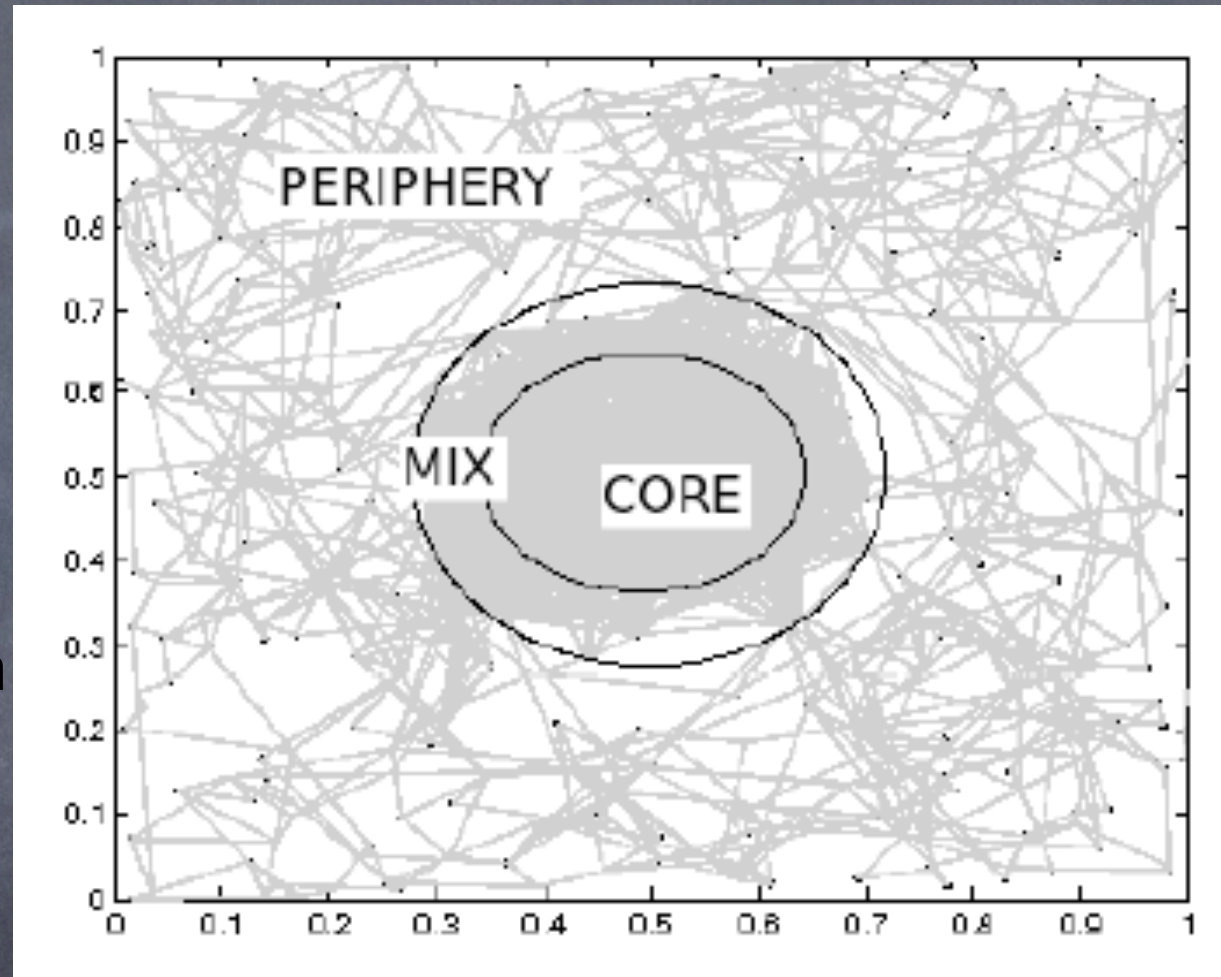
- Largest clusters identified by 1st eigenvector, 2nd largest, second eigenvector etc. random connections identified by the eigenvectors in the spectral gap.

- The left hand side is the structure of the graph.

Right hand side is the normalised number of N -cycles.

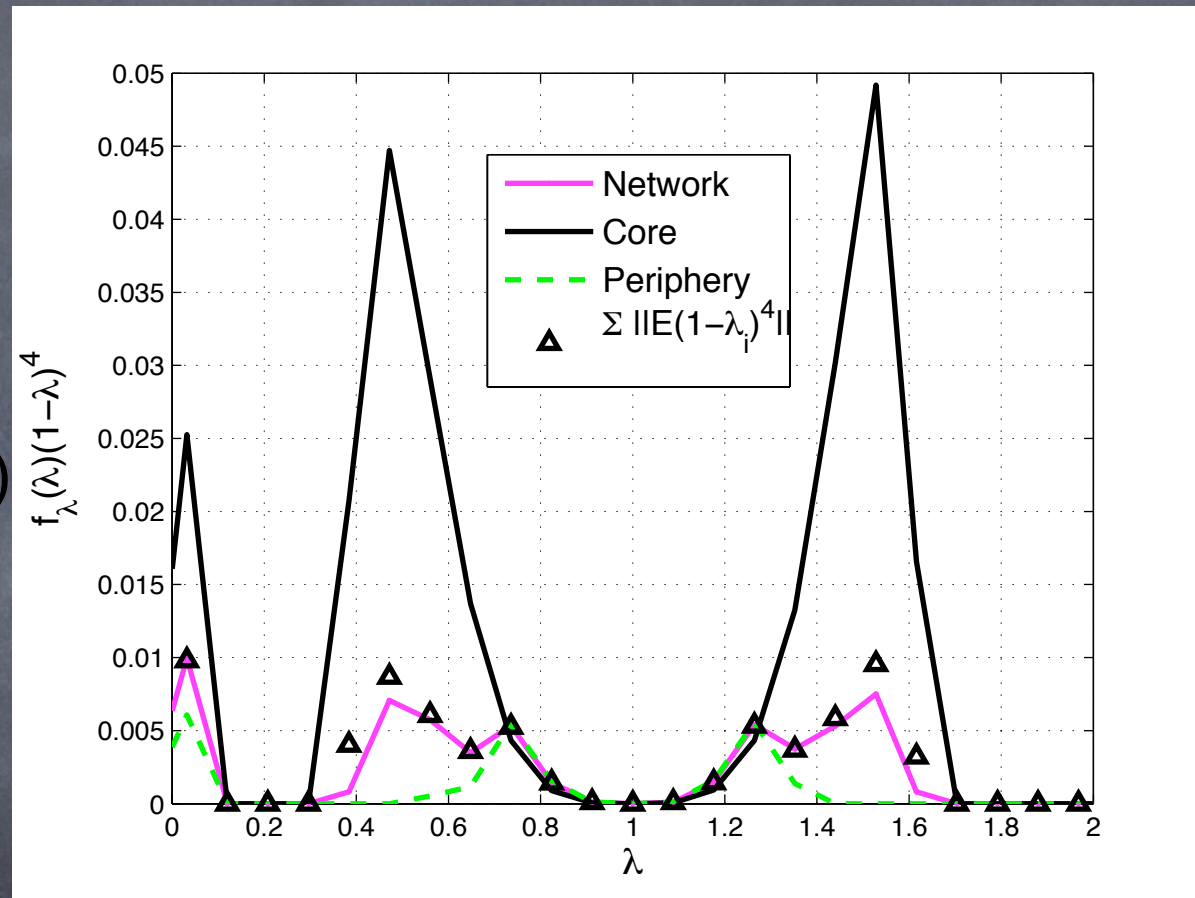
Application 1: The Straw-man topology model

- Strawman Model
- Demonstration only
 - Core
 - Periphery
 - Waxman connections
- Aim:
 - To demonstrate spectrum



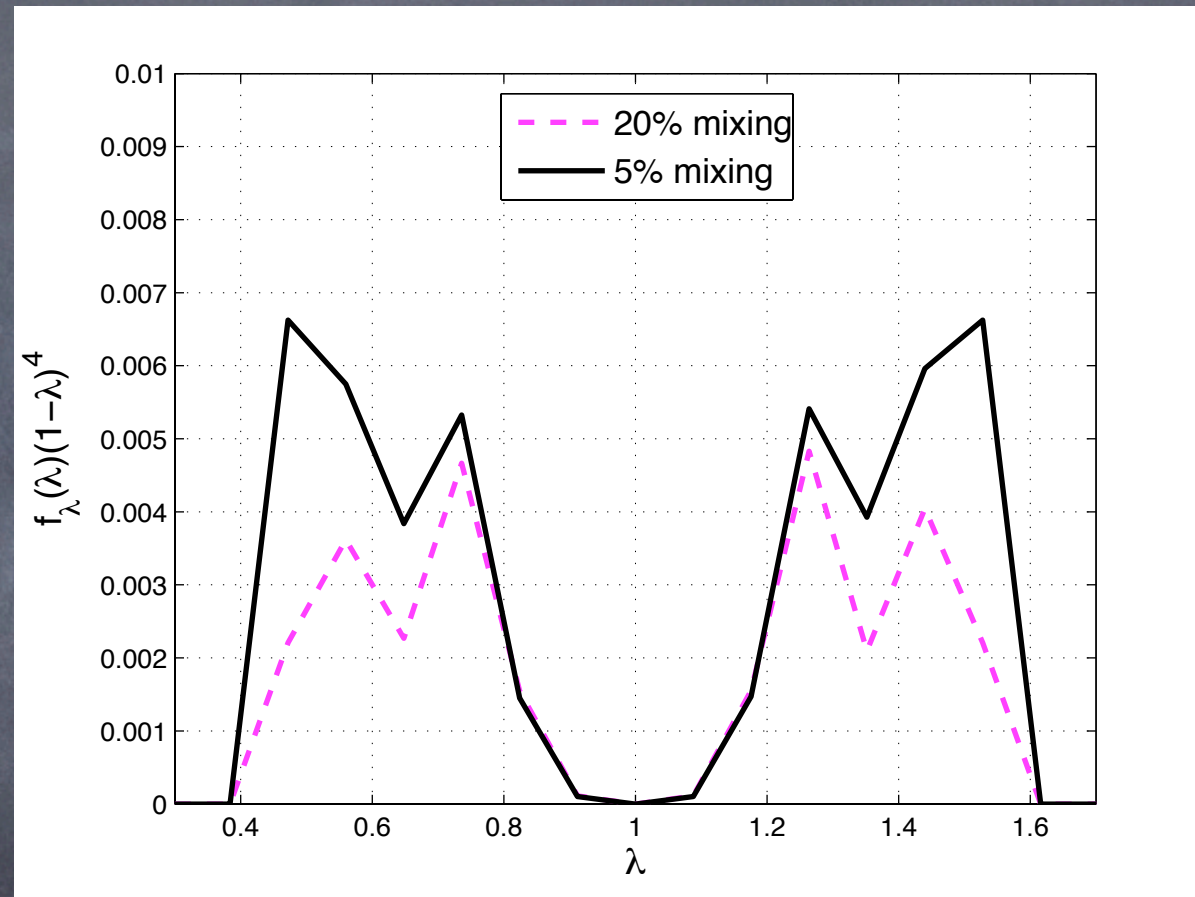
Application 1: The Straw-man topology model

Core and periphery add to form the weighted spectrum
= the (weighted)
combination of these spectra
(i.e. an estimate of the network)
is exact for no overlap.



Application 1: The Straw-man topology model

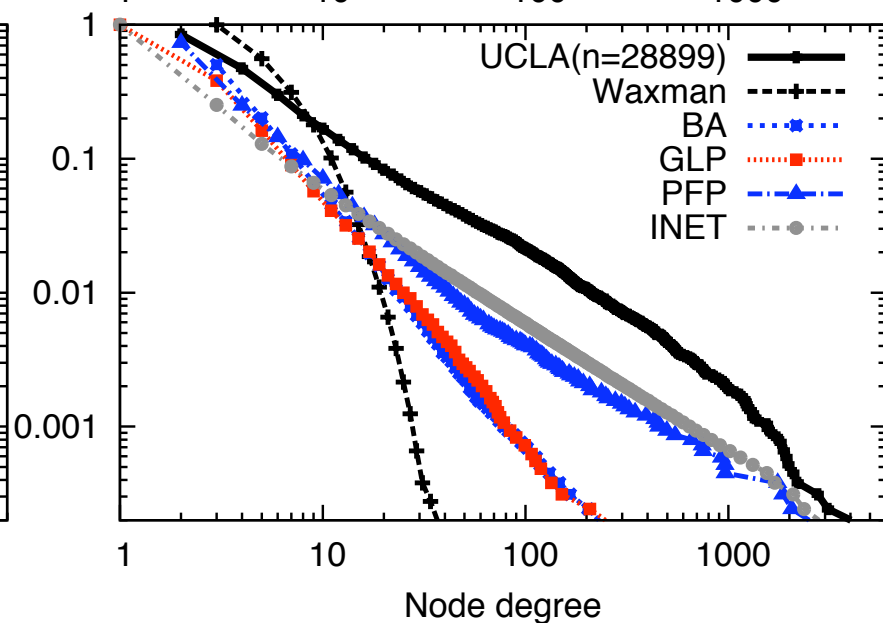
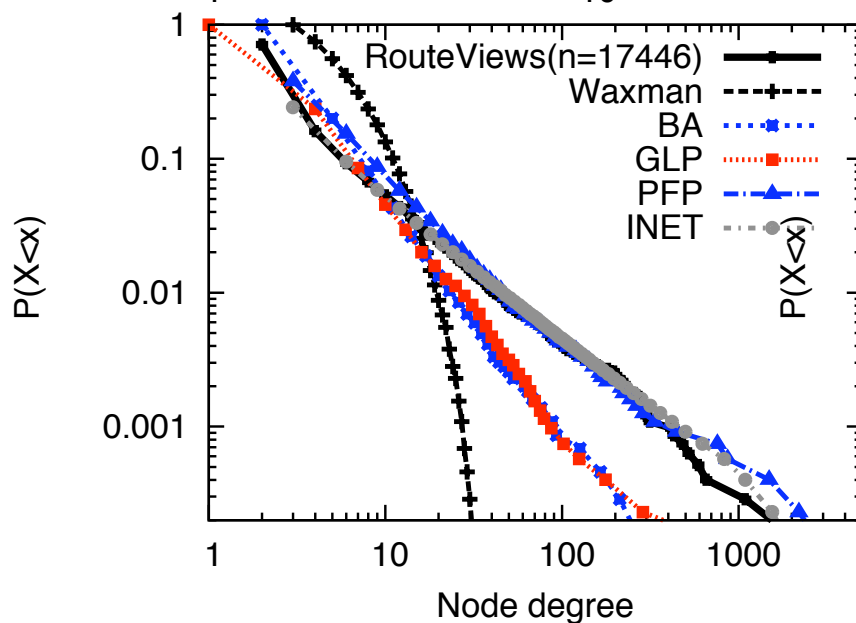
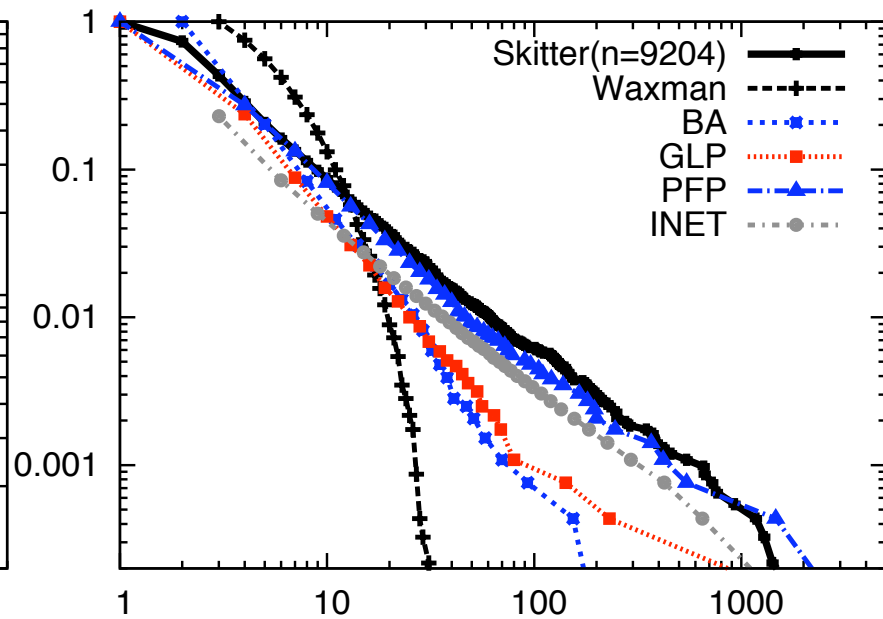
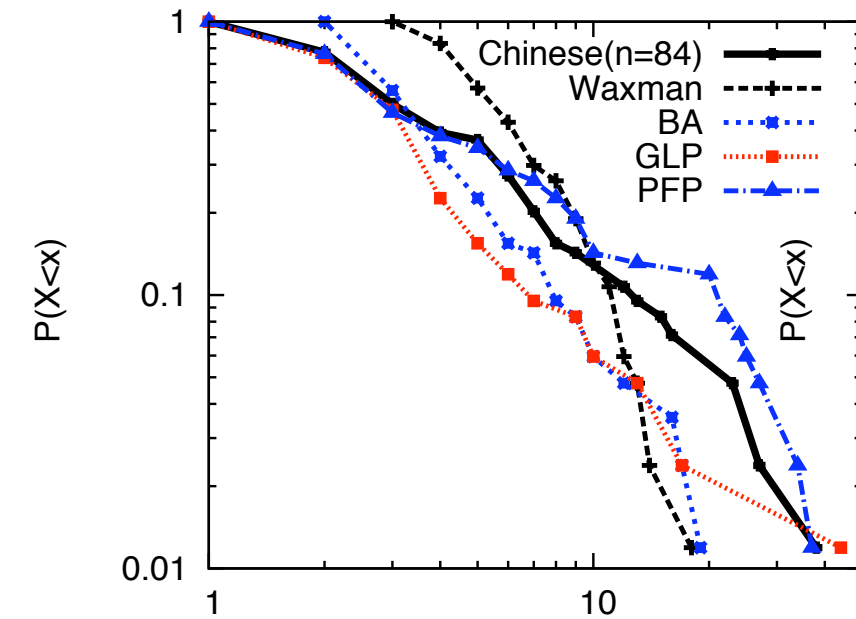
- Increase core and periphery mix.
- Core becomes less dominant.
- Core peak in spectrum reduced.



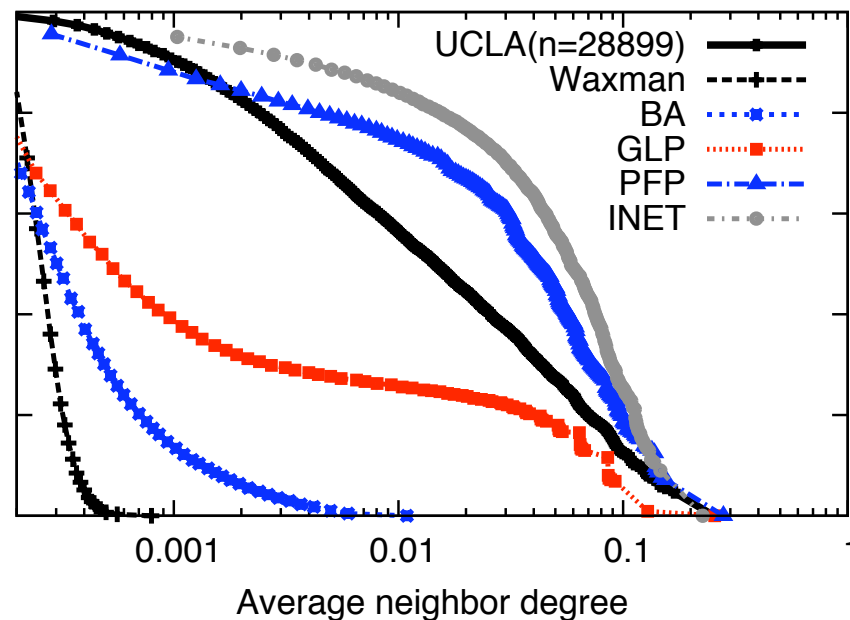
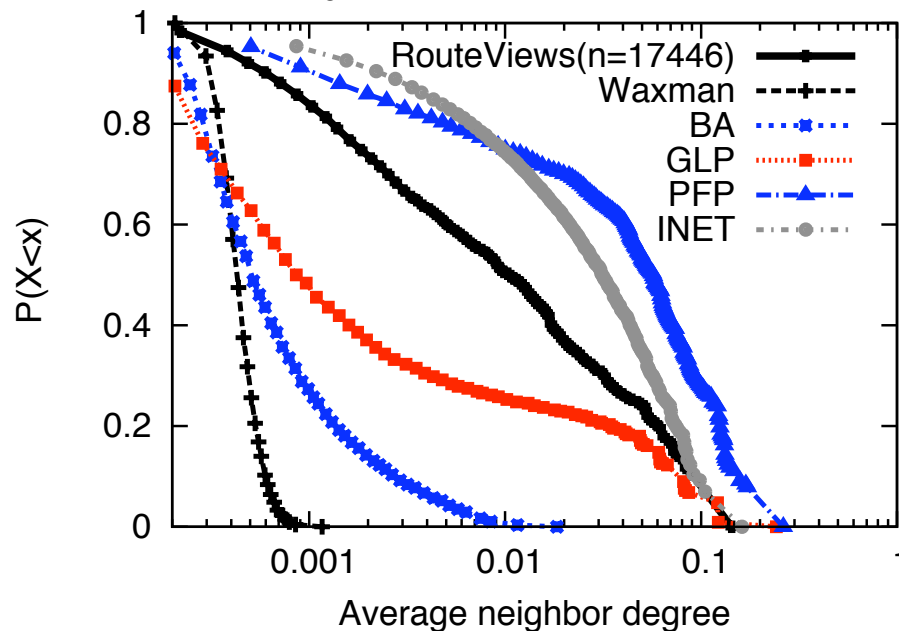
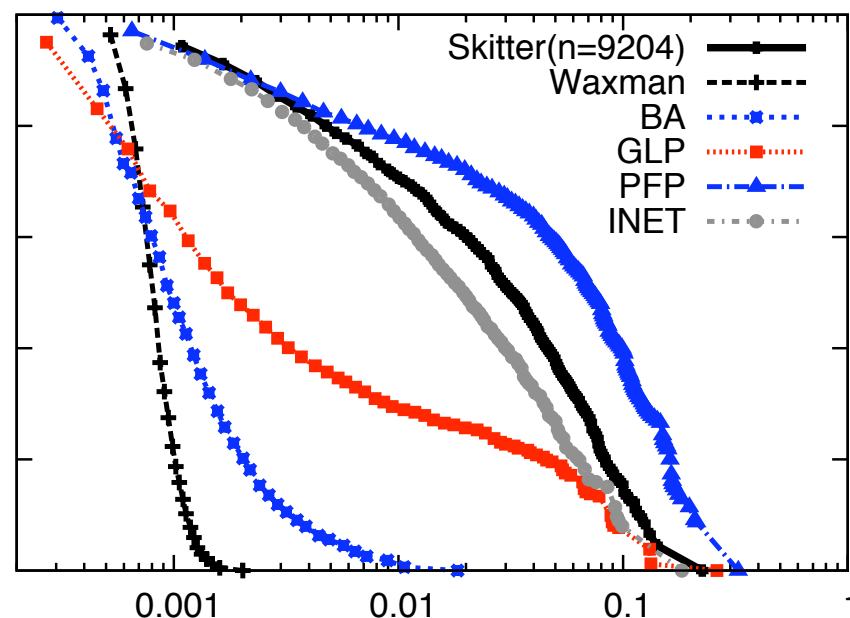
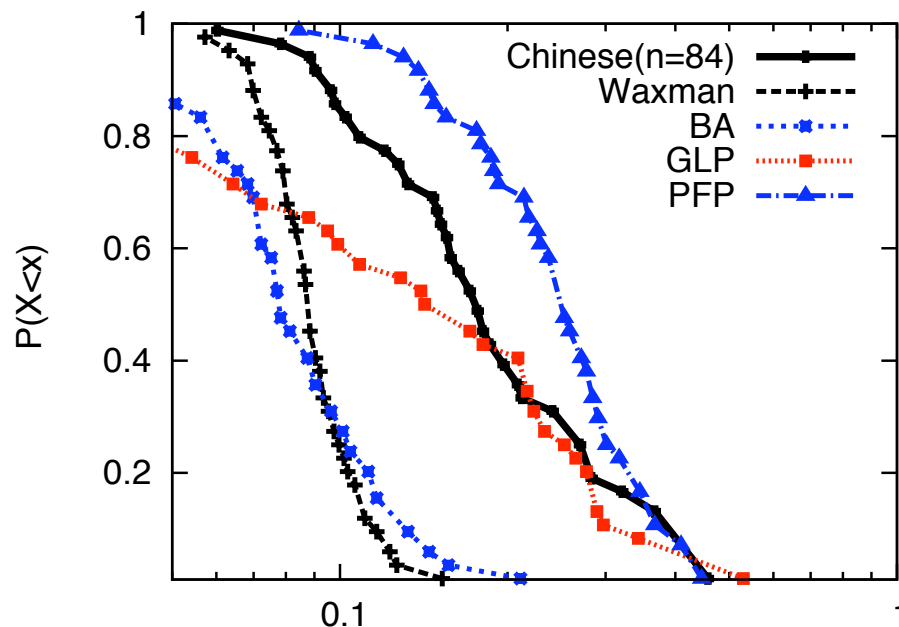
Application 2: topology

- 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 one data set for the internet at AS level:
- The skitter data set.

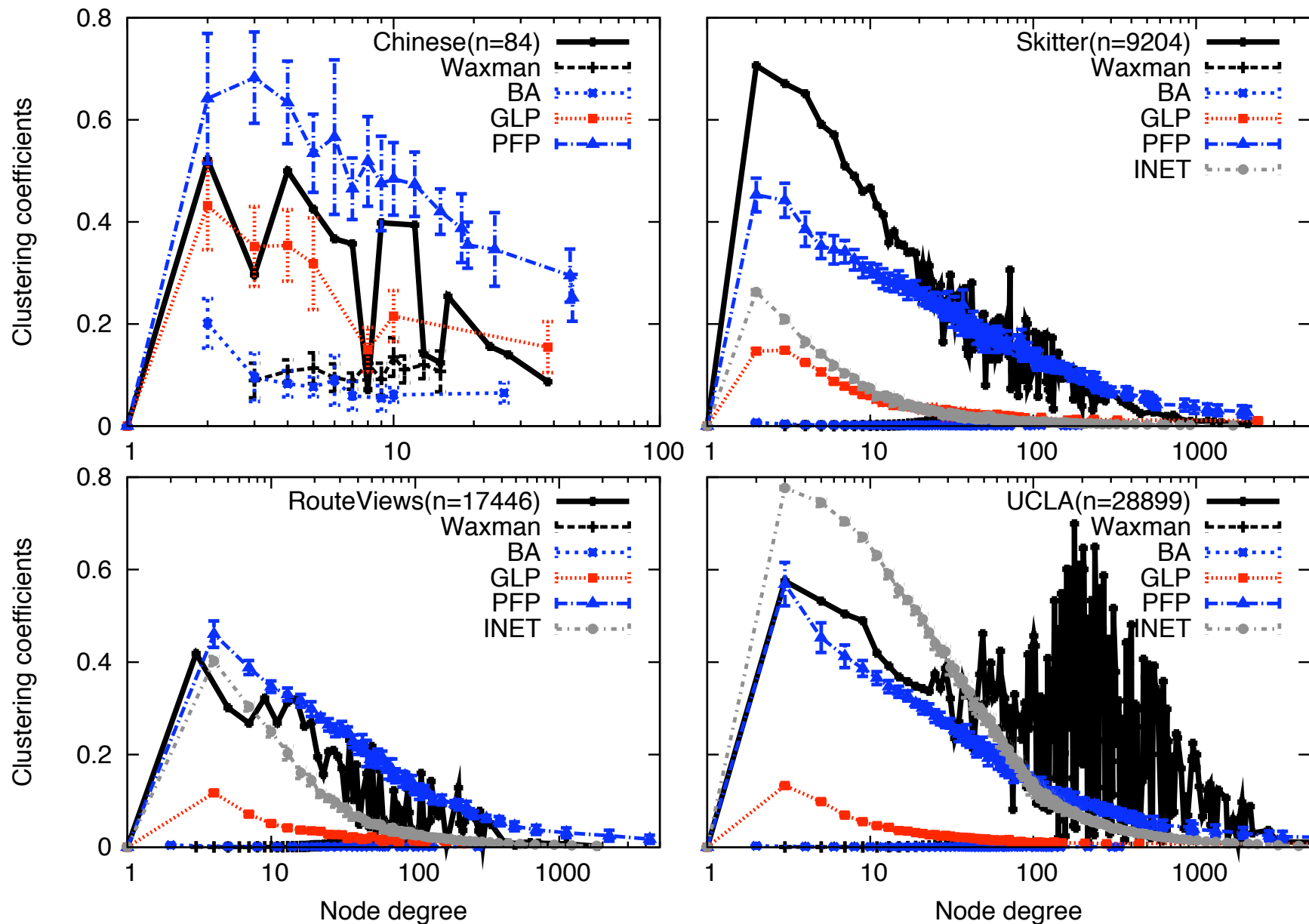
Degree distribution



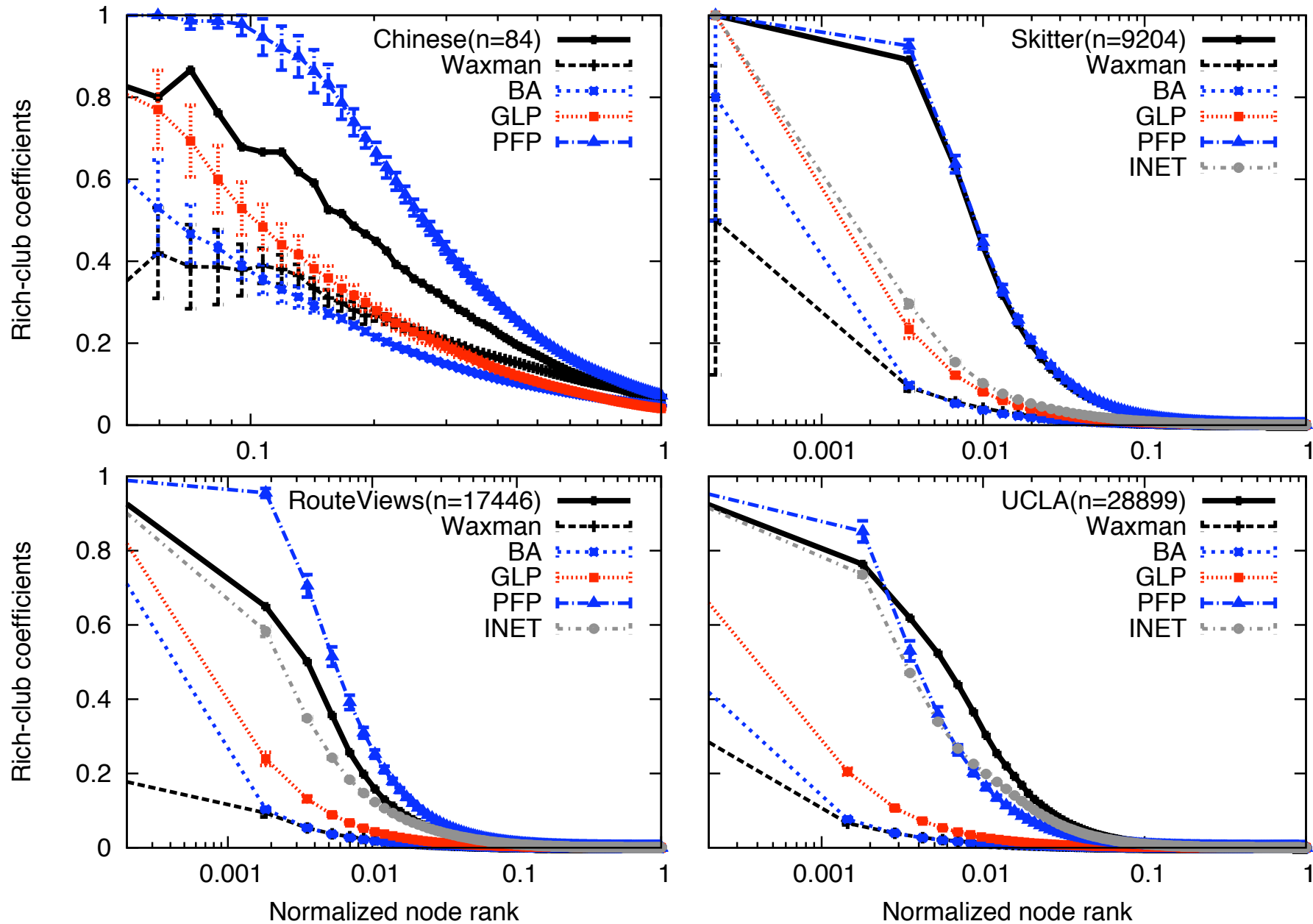
Average neighbor degree



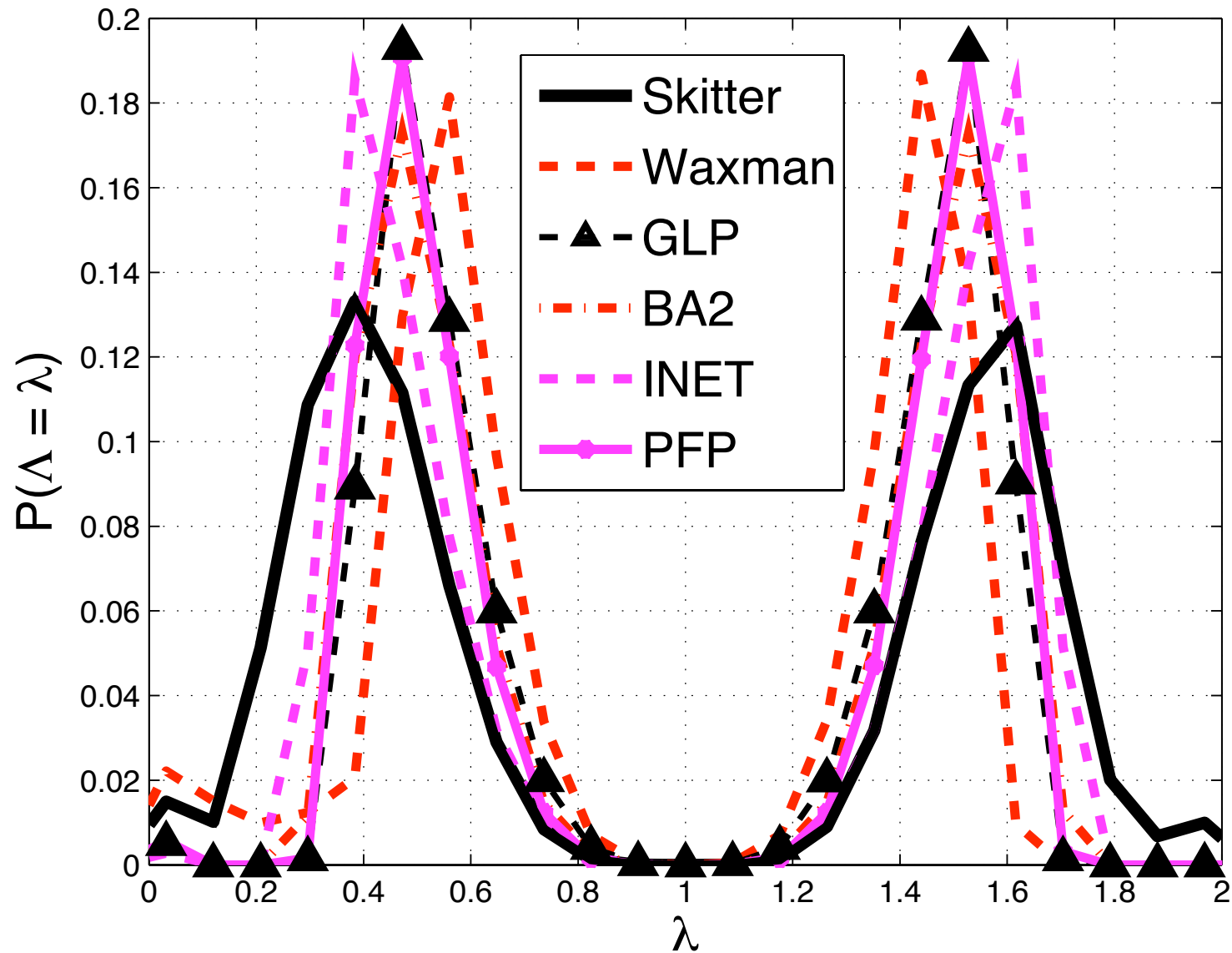
Clustering Coefficients



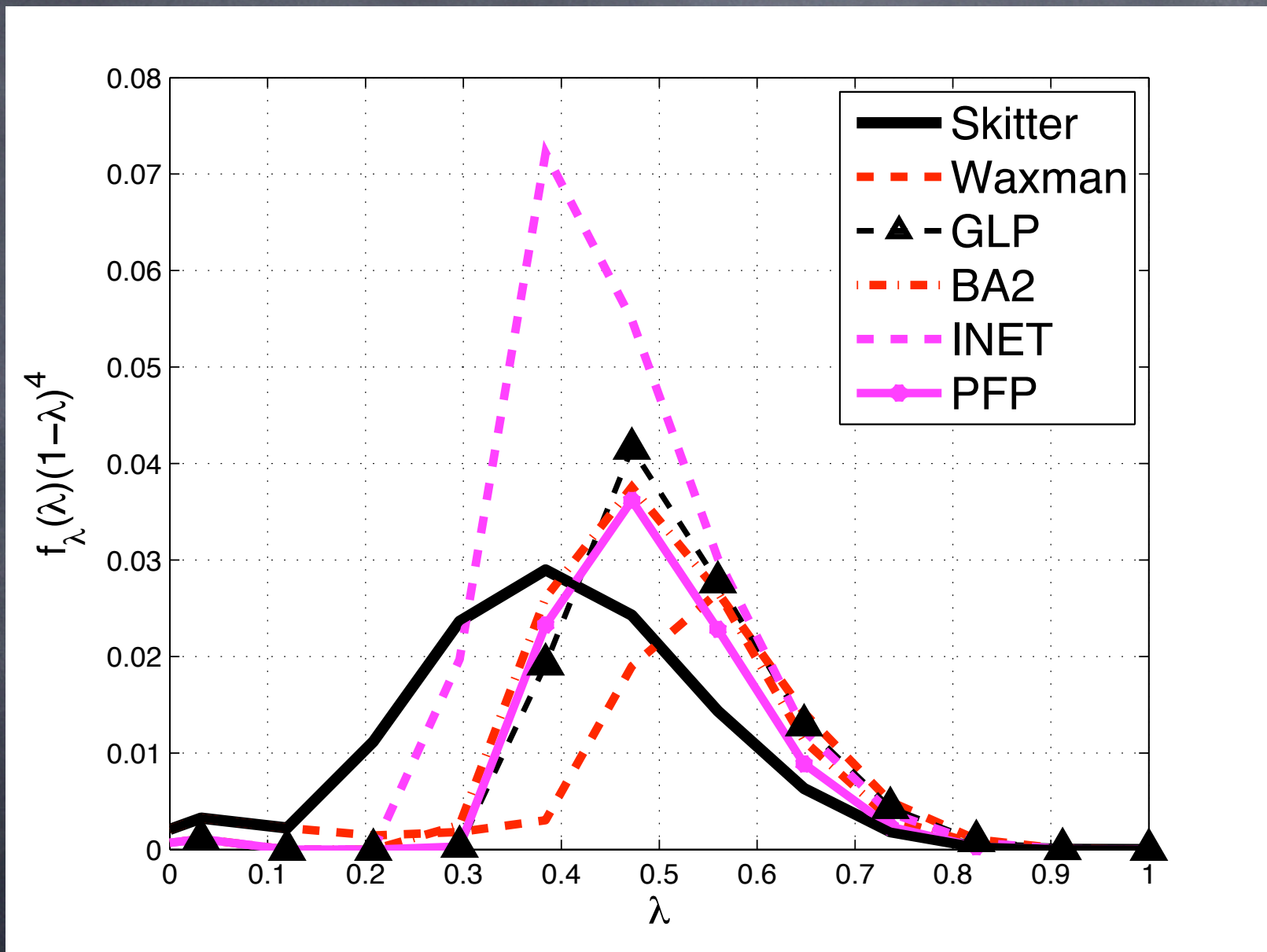
Rich Club connectivity



Application 2: topology



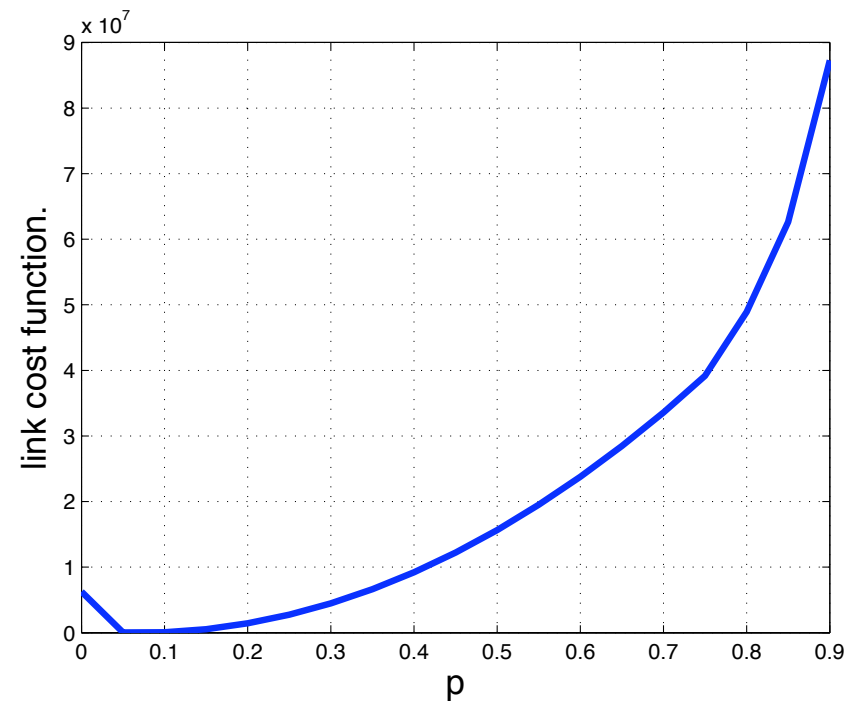
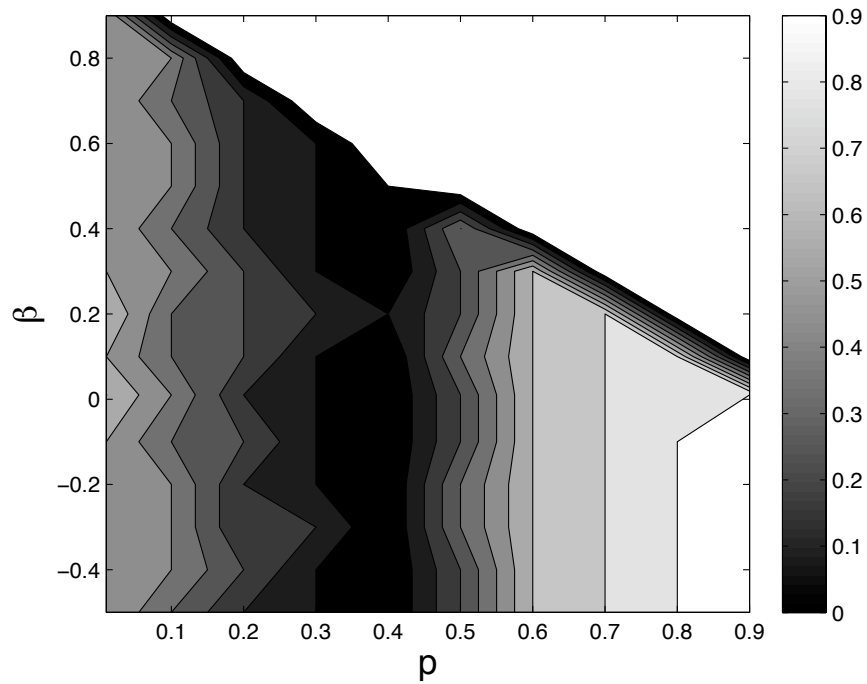
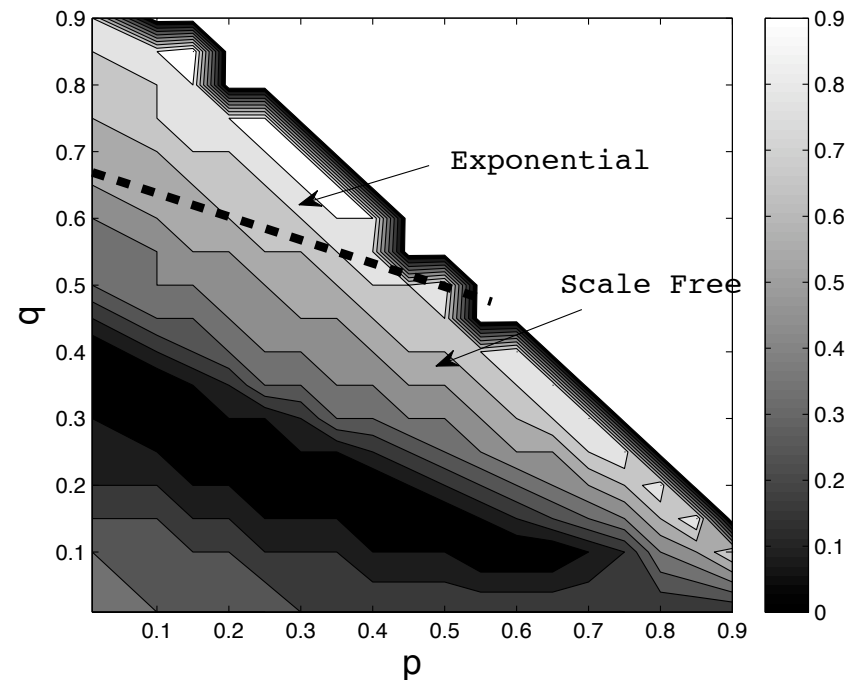
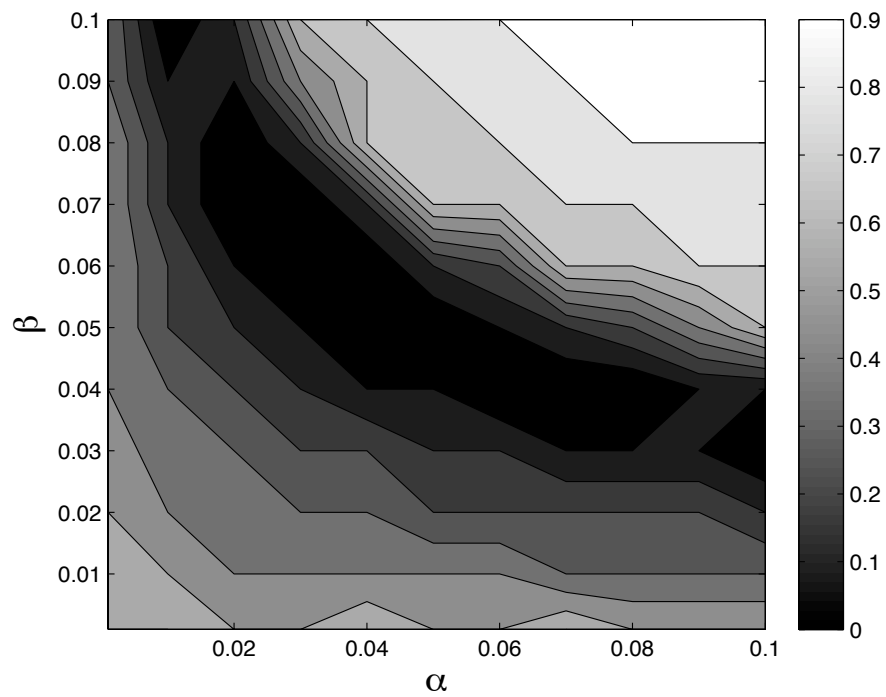
Application 2: topology

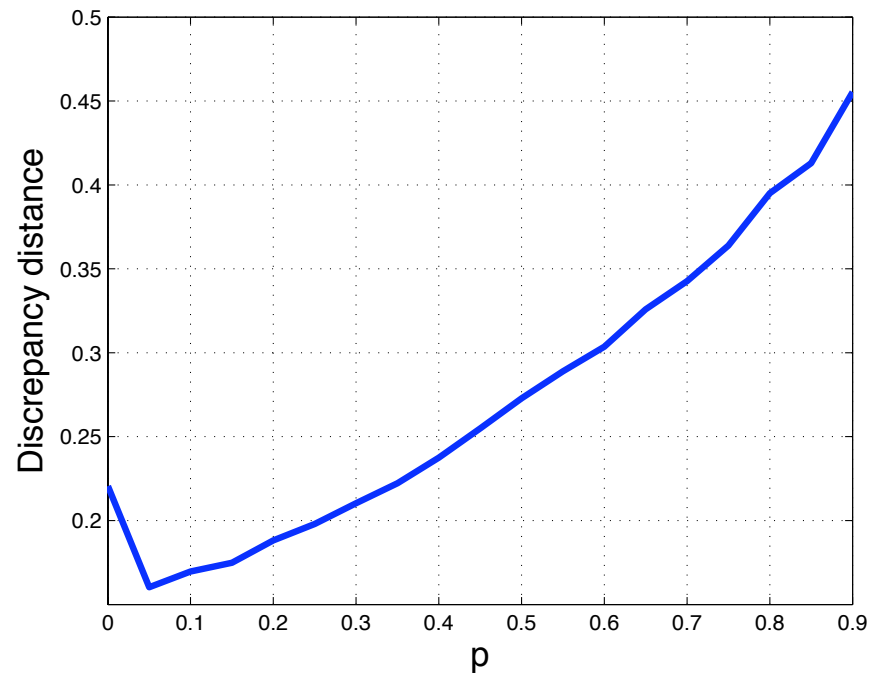
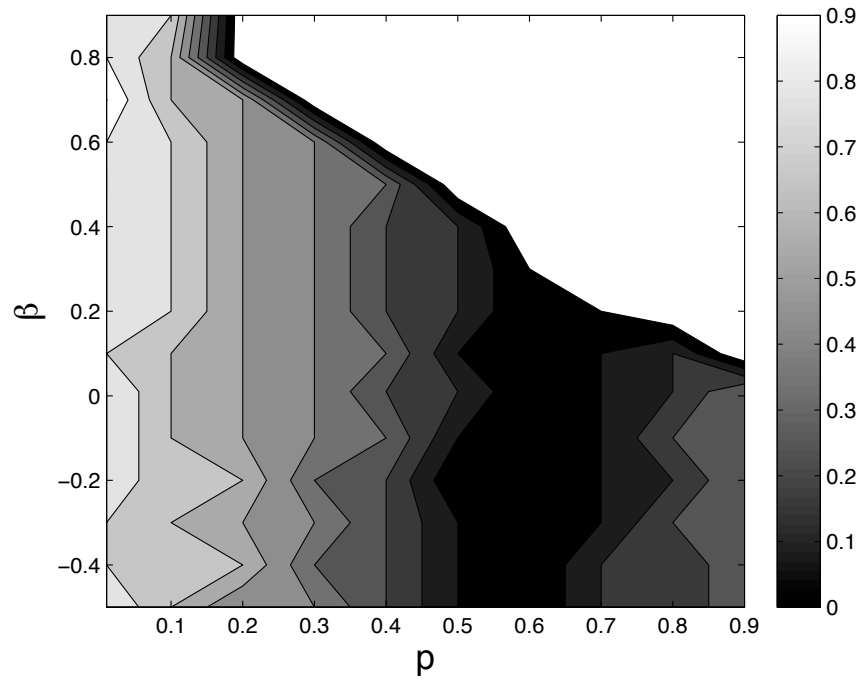
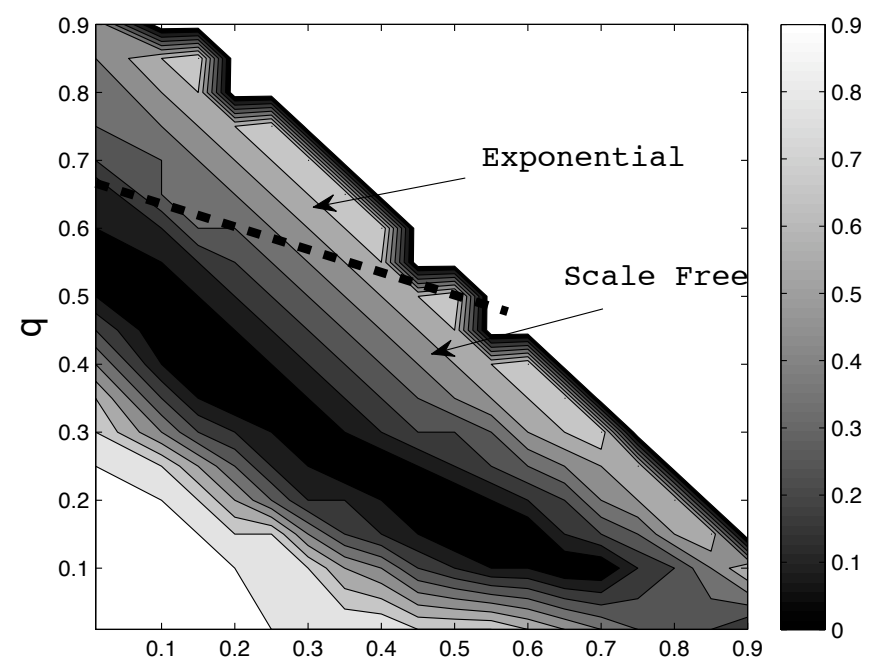
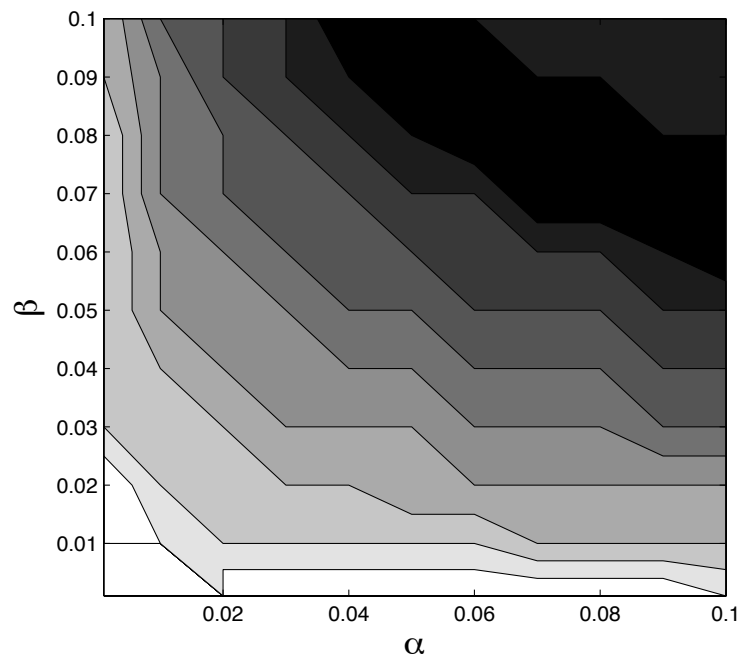


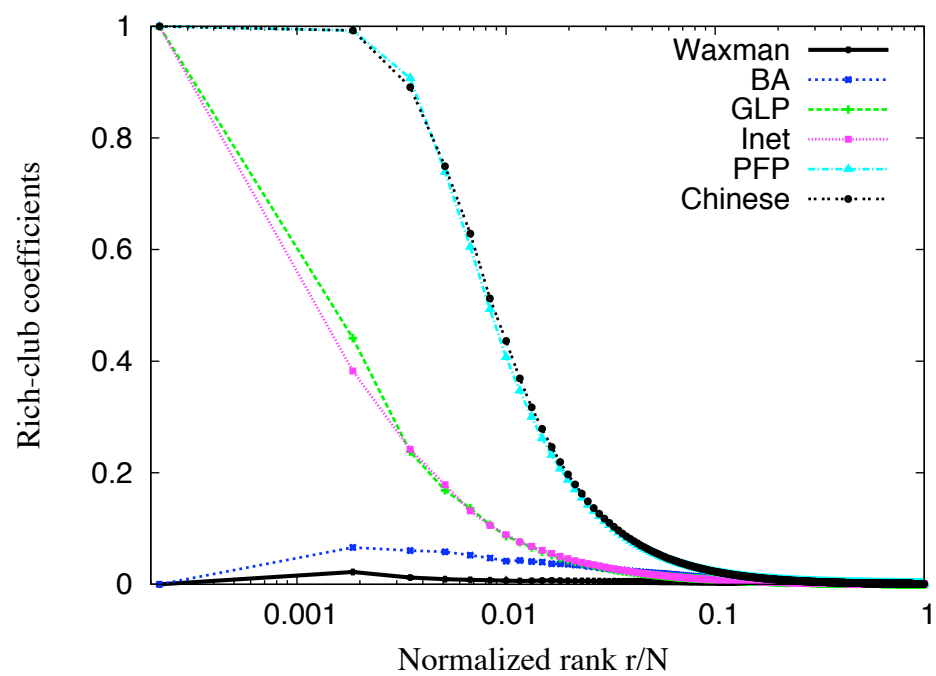
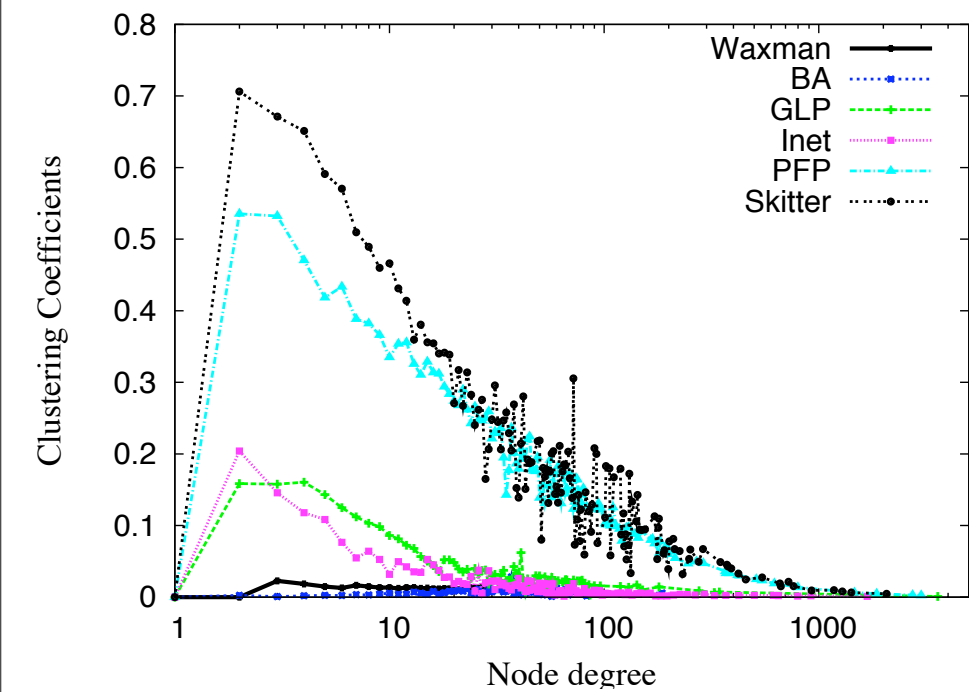
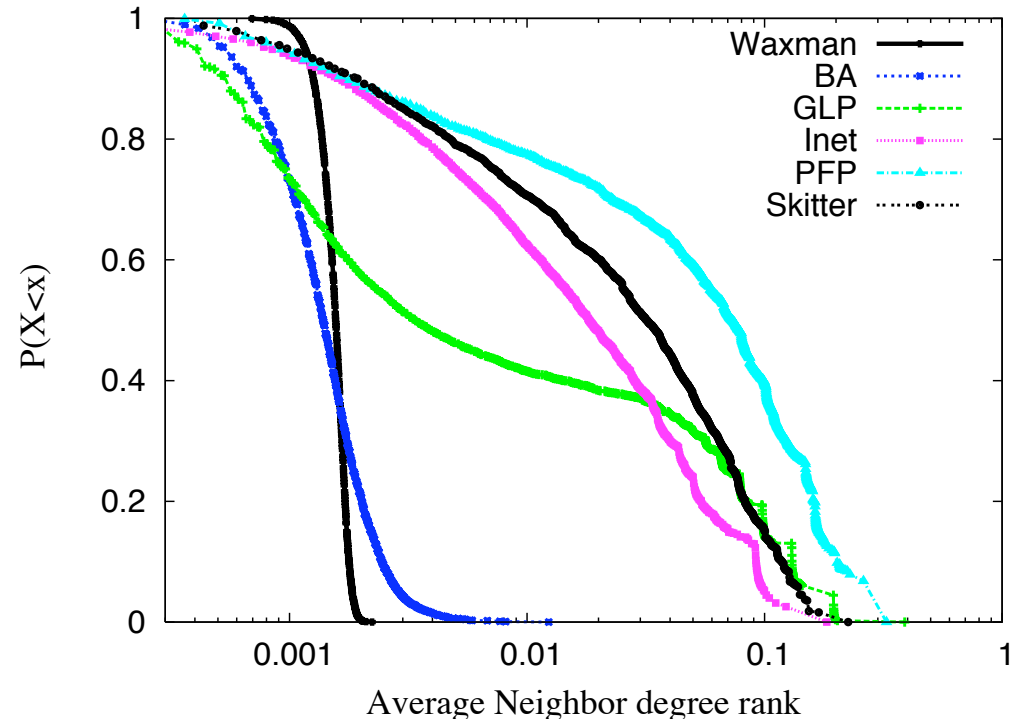
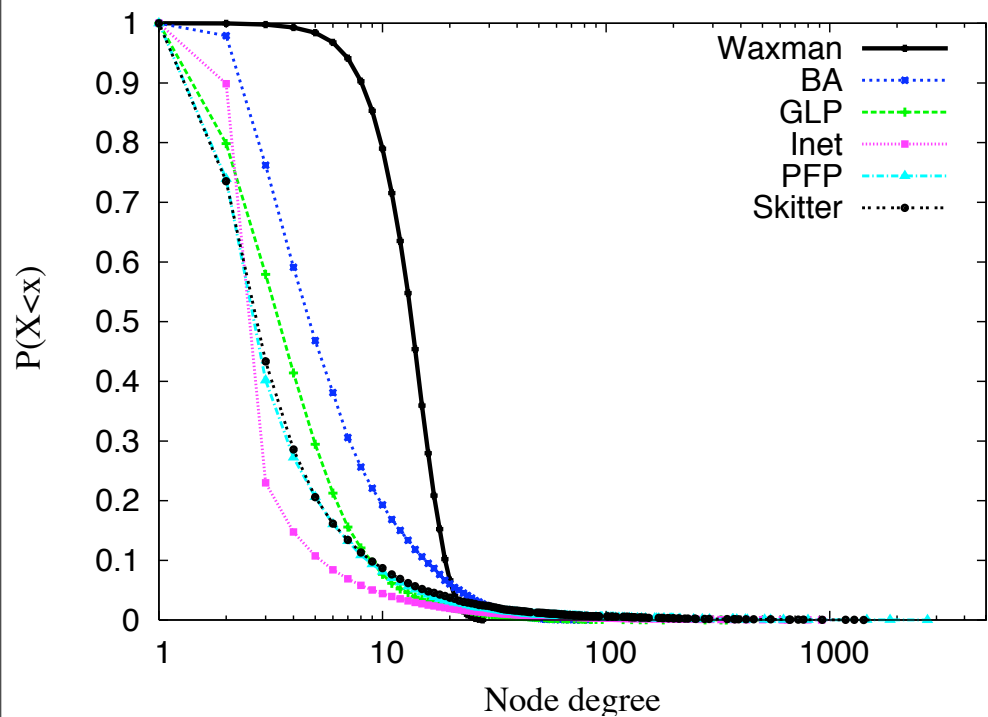
Parameter estimates

Waxman	$\alpha = 0.08$ (default= 0.15)	$\beta = 0.08$ (default= -0.2)
BA	$p = 0.2865$ (default= 0.6)	$q = 0.3145$ (default= 0.3)
GLP	$p = 0.5972$ (default= 0.45)	$\beta = 0.1004$ (default= 0.64)
Inet	$\alpha = 0.1013$ (default= 0.3)	—
PFP	—	—

- Waxman model is the worst.
- INET has too much structure but at the right point.
- None of the models examined is capable of mimicking the spectrum of the skitter data set.







Conclusions

- The metric allows identification of where in the structure two graphs differ.
- The metric behaves as expected and allows optimized parameters to be estimated.
- The raw spectrum is not sufficient to explain the structure in a topology.
- Future work?
- Questions?