Weighted Spectral Distribution: what a spectral topological metric can do for you?

Steve Uhlig

TU Berlin/Deutsche Telekom Laboratories

joint work with

Damien Fay (University of Cambridge), Hamed Haddadi (University of London), Liam Kilmartin (National University of Ireland), Andrew W. Moore (University of Cambridge), Jerome Kunegis (TU Berlin), Marios Iliofotou (UC Riverside)

Motivation

- Many graph measures
- Each captures specific structural property
  - local structure, e.g., clustering, communities,
  - global structure, e.g., degree distribution, assortativity
- How to compare arbitrary graphs?

Agenda

- Graph measures
- Internet topology evolution
- Topology generators
- Application discrimination
- Dk-random graphs

Topological measures
Examples

- Clustering coefficient: $\gamma(G) = 1/n \sum_i 2T_i/d_i(d_i-1)$, $d_i \geq 2$, with T_i the number of triangles of node i
- Assortativity coefficient: correlation coefficient between node degrees
  - Internet is disassortative: large nodes connect to small ones
  - Social networks are assortative
- K-core: remaining graph when removing nodes with degree less than k
- Centralities

Weighted Spectral Distribution
Definition

- A: adjacency matrix, D: diag(A)
- Let λ_i be the ith eigenvalue of the normalized Laplacian matrix $L = I - D^{-1/2} A D^{1/2}$
- WSD: $\omega(G,N) = \sum_k ((1-k)^N f(\lambda=k))$
- Depending on N, the WSD relates to specific topological properties
  - n=3: clustering
  - n=4: quasi-randomness

Weighted Spectral Distribution

Usage

- Graph distance metric [ToN'10]:
  $\delta(G_1,G_2) = \sum_k (1-k)^N (f_1(\lambda=k)-f_2(\lambda=k))^2$
- WSD + random projections [Fay'10]:
  $$Z = XT$$
  with X: WSDs of m graphs, T: random (Gaussian) matrix
- WSD + multi-dimensional scaling [Fay'10]:
  - Build dissimilarity matrix R based on δ
  - Find approximating vectors $Z_1, Z_2, \ldots, Z_k$ that minimize a cost function C

Agenda

- Graph measures
- Internet topology evolution
- Topology generators
- Application discrimination
- Dk-random graphs

Internet topology evolution

- Evolving Internet structure
- Past – strong hierarchy
  - Tier-1's
  - National ISPs
  - Stubs
- Present – big players rising [Arbor'10]
  - Google
  - Akamai
  - YouTube
- What do different data-sets tell us?

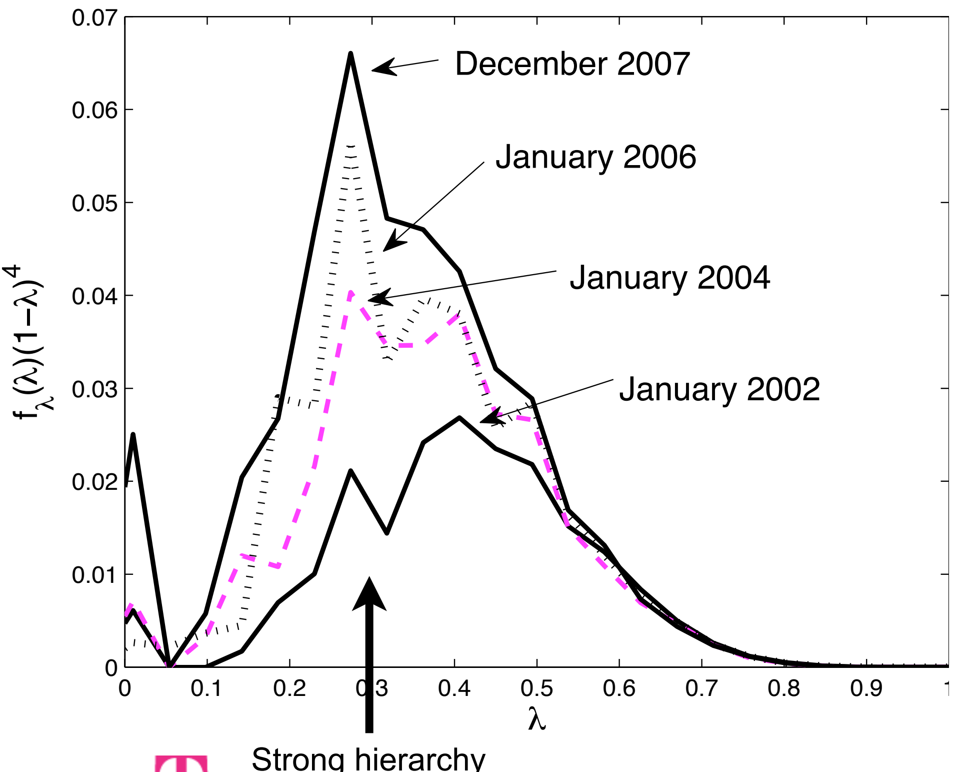
Internet topology evolution
Inference techniques

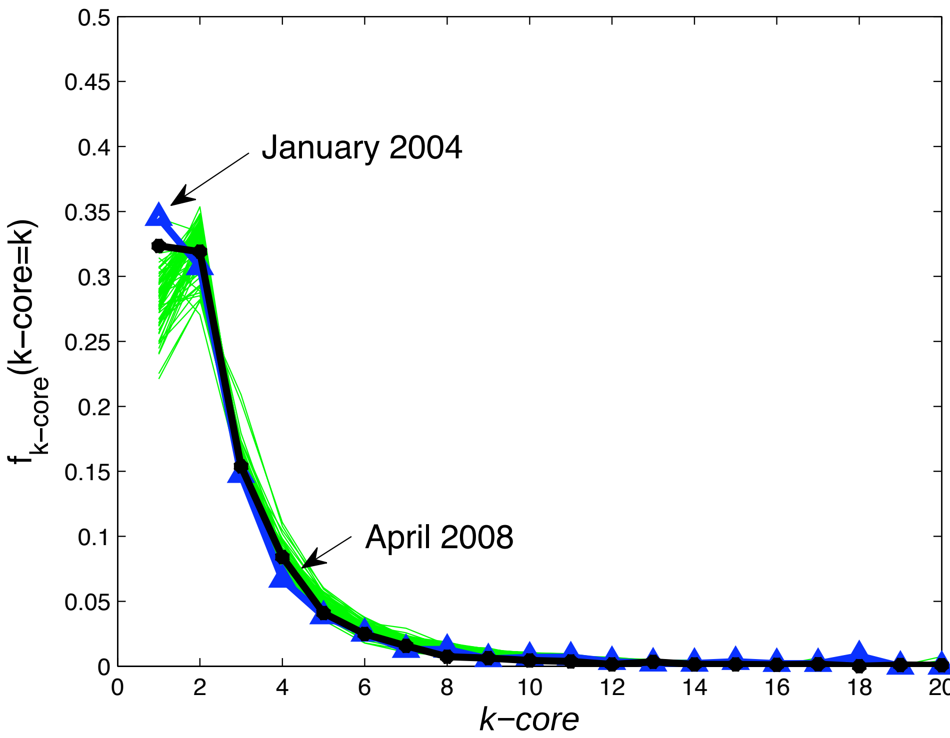
- Control plane (e.g., BGP): routing paths towards all destinations seen by a set of observation points
- Data plane (e.g., traceroute): probe packets sent from a few boxes towards a large set of IP addresses
- Limitations:
  - BGP does not sampling all path diversity, biased by routing policies and observation points
  - Traceroute: finding a wide set of destinations to probe is hard, NAT and firewalls block probes at the edge

AS topology evolution
Traceroutes

- Reduced complexity
- More dominant core

- No significant change

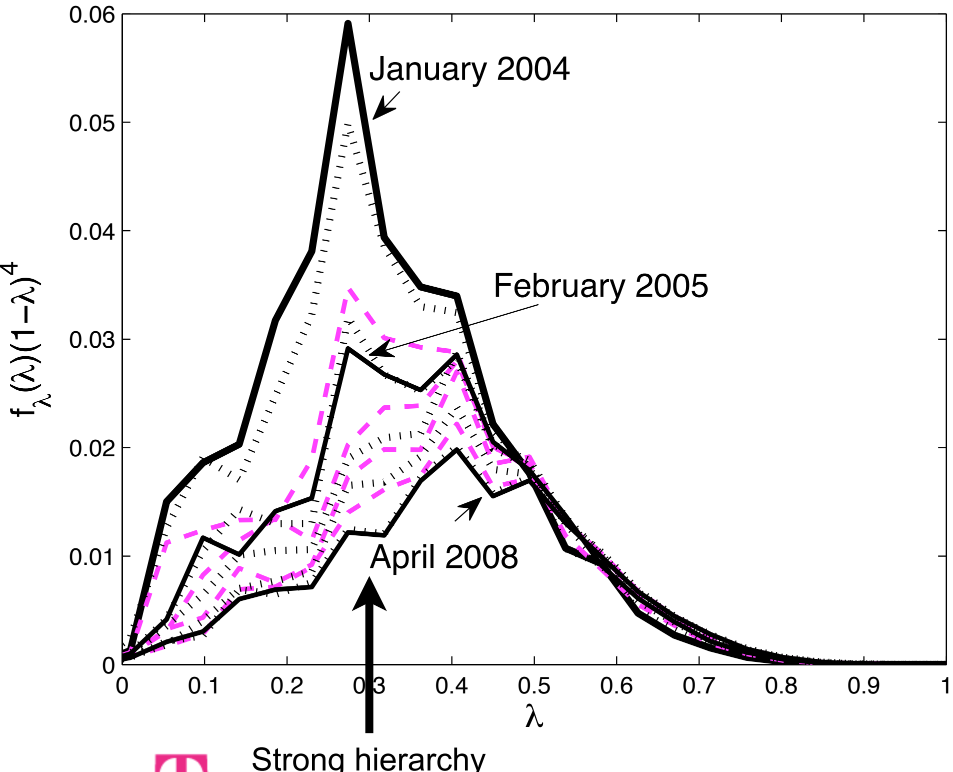




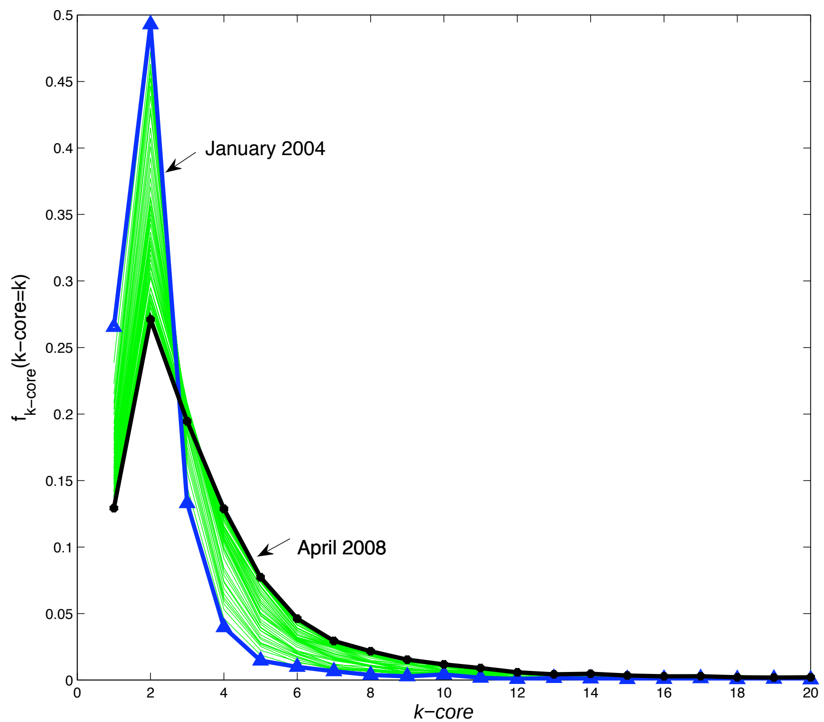
AS topology evolution

BGP

- Increased complexity
- Less dominant core



- Small degree nodes get better connected



Agenda

- Graph measures
- Internet topology evolution
- Topology generators
- Application discrimination
- Dk-random graphs

Generating realistic graphs

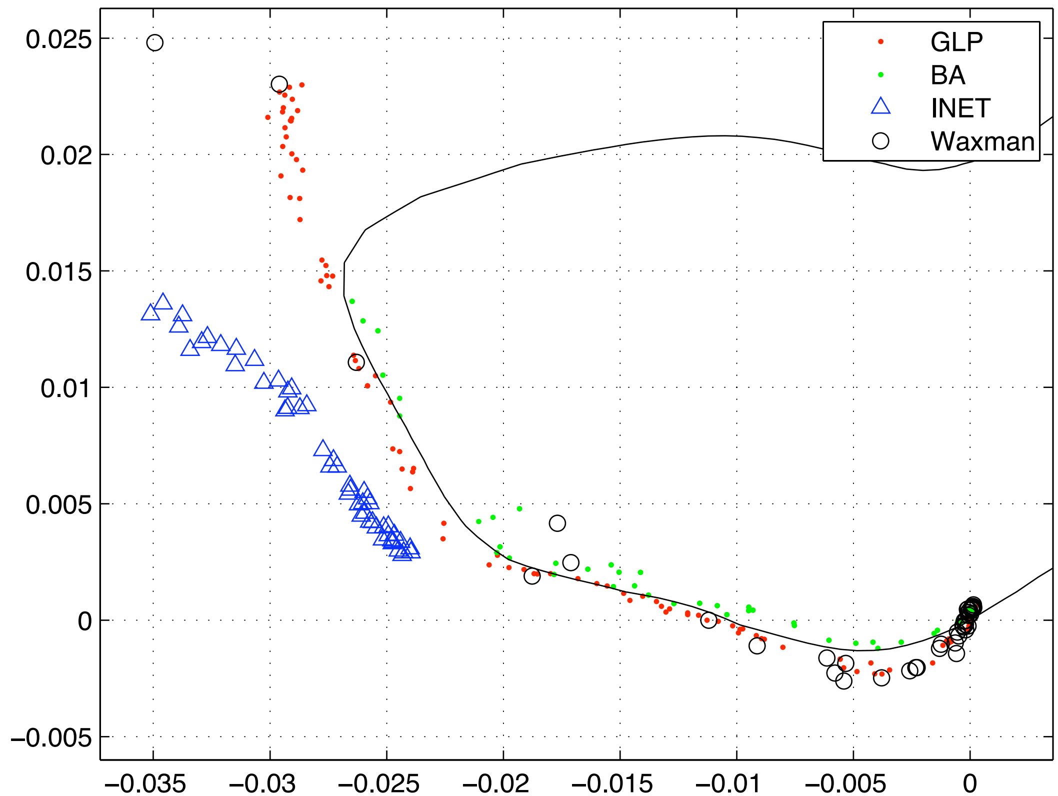
- Existing graph models
  - Random graphs
  - Power-law
  - Small world
- Each model has specific structural properties
- How randomly do existing generators sample the graph space?

Topology generators

- **Waxman**: random graphs $p(i,j) = \alpha\, e^{-d/(\beta L)}$ where d is euclidian distance, $0 < \alpha,\beta < 1$, and L is network diameter
- **BA**: linear preferential attachment (LPA) $p(i,j) = d_j / \sum_{k \in V} d_k$ where d_j is degree of node j and V is the set of nodes that have joined the network so far
- **GLP**: preferential attachment + path length and clustering match
- **Inet**: LPA + full-meshed Internet core + 30% of single-homed ASes

Topology generators
How random is your graph?

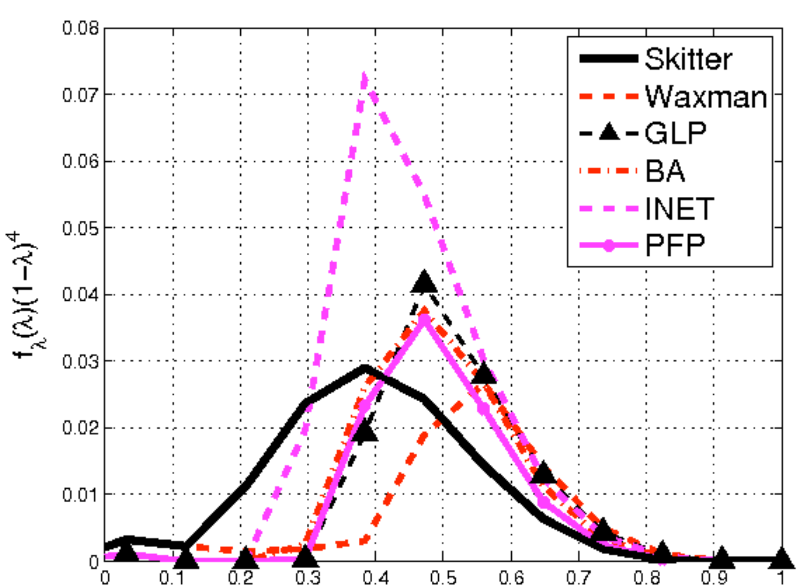
- Generate graphs with well-known topology generators
- Project their WSD onto low-dimensional space
- Each generator samples graph space in a very specific way

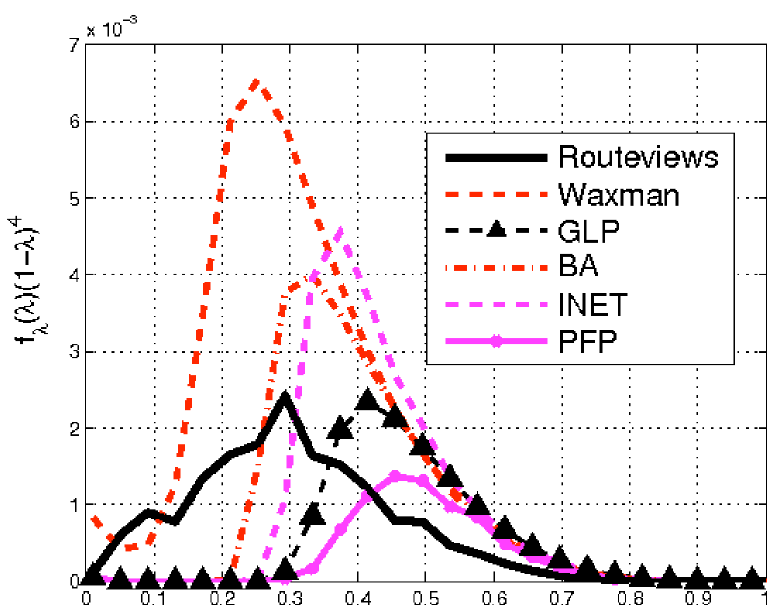


Topology generators

Optimization

- Optimize generator parameters to minimize distance with a given topology
- No generator manages to closely match any given graph





Agenda

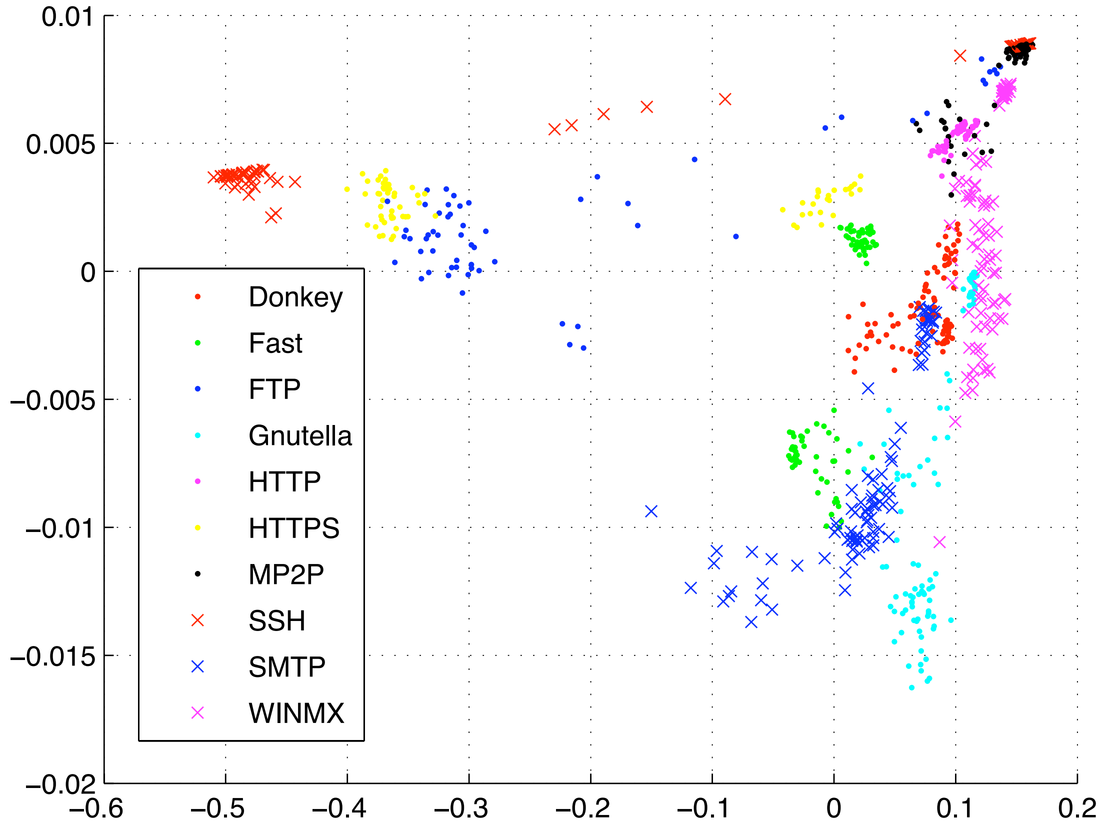
- Graph measures
- Internet topology evolution
- Topology generators
- Application discrimination
- Dk-random graphs

Application identification

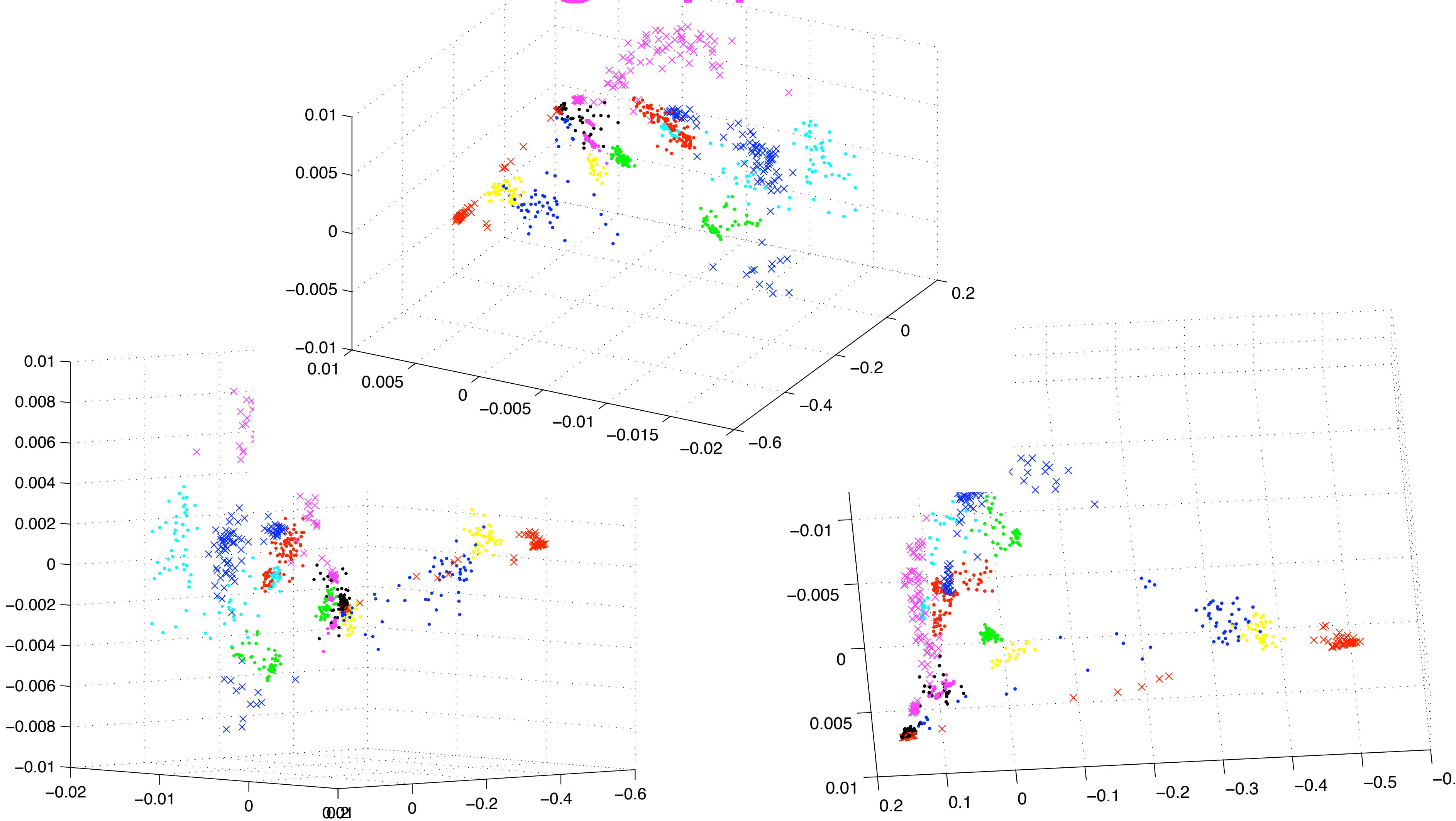
- Different applications have different communication patterns
- Can we discriminate between them only based on their structure?
- [CoNEXT'09] requires about 10 metrics to classify such graphs
- How many dimensions are really necessary?

Discriminating applications

- Using WSD and projecting on 2 dimensions already discriminates
- 86% of the flows are correctly classified
- Misclassification occurs due to low dimensionality of projection and similarity of specific applications



Discriminating applications



Agenda

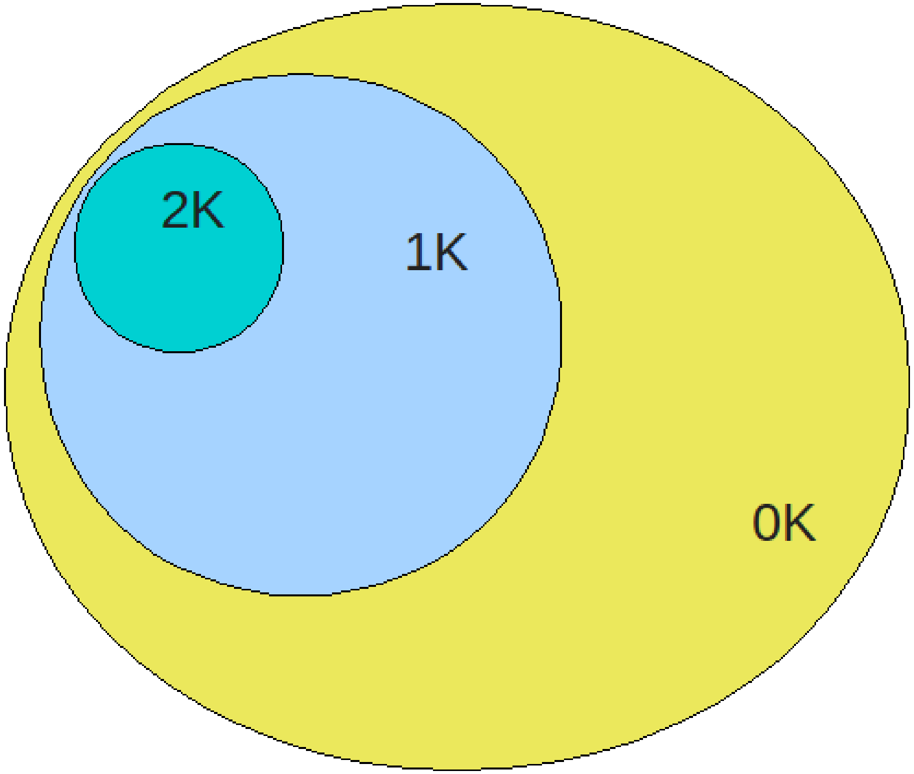
- Graph measures
- Internet topology evolution
- Topology generators
- Application discrimination
- Dk-random graphs

Dk-random graphs

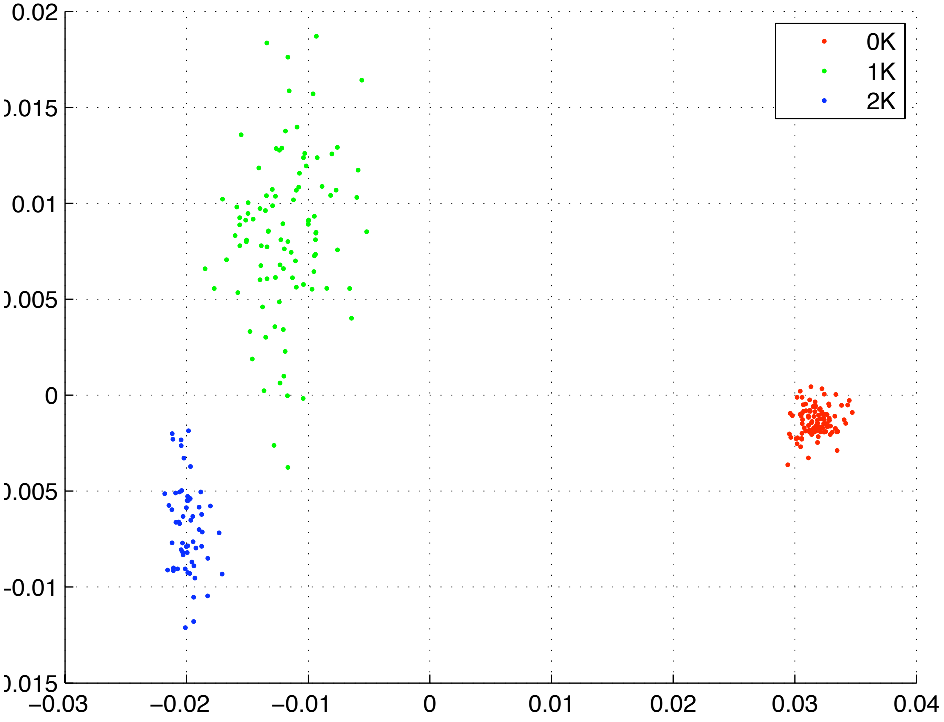
- Configuration model [Bollobas'80]
  - Builds graph with given degree distribution
- Orbis topology generator relies on dK series:
  - 0K: set of graph with average degree k
  - 1K: set of graphs with degree distribution p(k)
  - 2K: set of graphs with joint degree p(k,j)
  - ...
- Are graphs generated by Orbis random?

Dk-random graphs

- Theoretical space of Dk-random graphs



- Generate 100 graphs from 0k, 1k, and 2k
- WSD + MDS



Conclusions

- Sets of graphs are not random
  - Do not assume that topology generators sample randomly a given graph space
  - Real-world graphs even less so
- Graphs metrics like WSD can do a lot for you
  - Compare structural evolution through specific property, e.g. mixing
  - Find most appropriate topology generator
  - Discriminate specific graphs, e.g. Internet applications
  - Assess randomness in graph space, e.g. Dk-series

Selected references

- [Arbor'10] C. Labovitz et al. Internet Inter-domain traffic. Proc. of ACM SIGCOMM, 2010.
- [Bollobas'80] B. Bollobas. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. Eur. J. Combinatorics, vol. 1, 1980.
- [ToN'10] D. Fay et al. Weighted Spectral Distribution for Internet Topology Analysis: Theory and Applications. IEEE/ACM Transactions on Networking, 19(1), 2010.
- [CoNEXT'09] M. Iliofotou et al. Exploiting Dynamicity in Graph-based Traffic Analysis: Techniques and Applications. Proc. of ACM CoNEXT, 2009.
- [Fay'10] D. Fay et al. Discriminating graphs through spectral projections. Under submission, 2010.