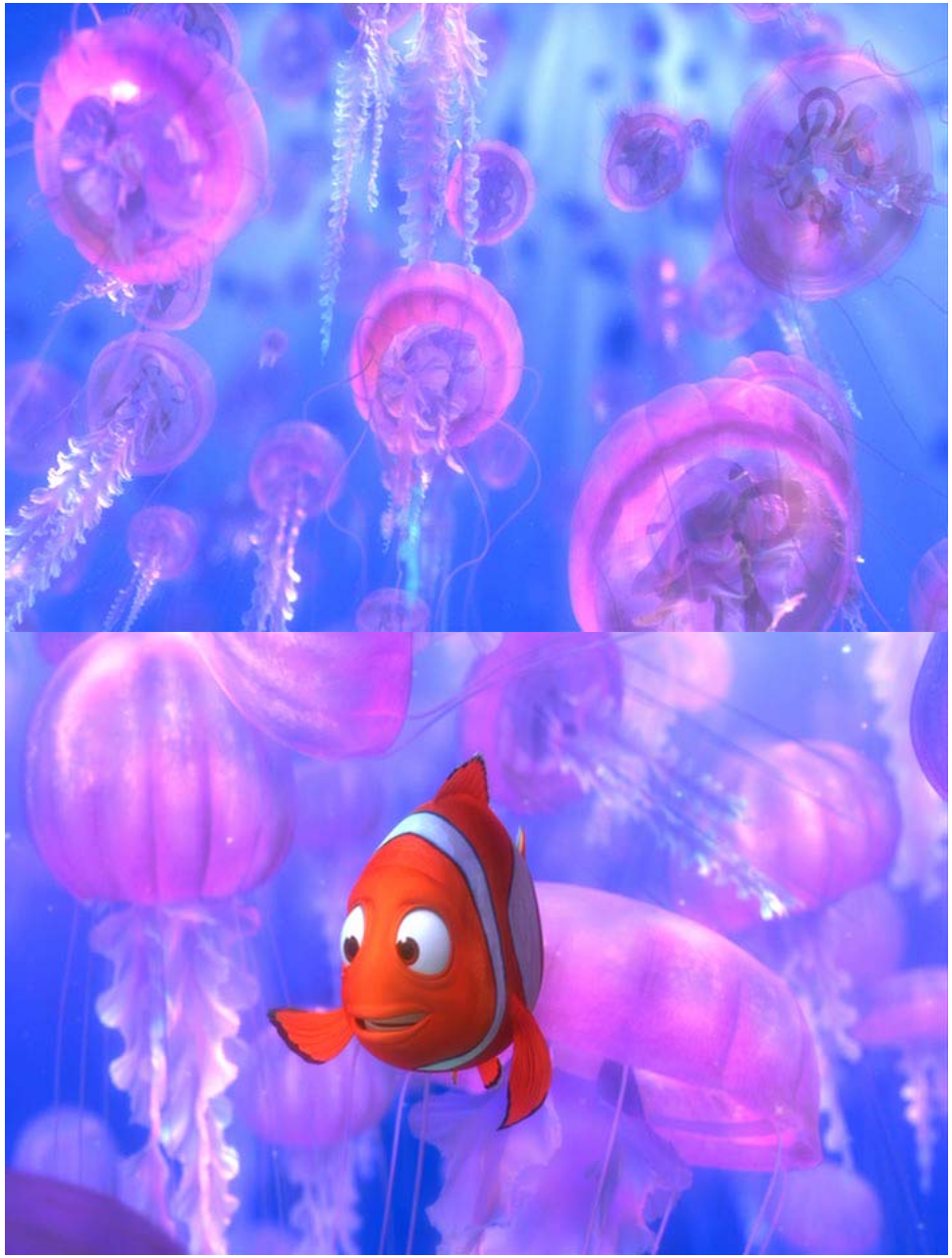


Copyright, Pixar, Inc. 2003

EVERGROW

Jellyfish, and other Interesting creatures Of the Internet

Scott Kirkpatrick, Hebrew University

with
Avishalom Shalit, Sorin Solomon,
Shai Carmi, Shlomo Havlin,
Eran Shir, and Yuval Shavitt

San Diego SDSC, May 11, 2006

Exploring the DIMES AS-graph

- We consider the Internet at the level of its autonomous systems (ASes)
- Previous studies have used degree as indicator to decompose networks
  - In particular, the Faloutsos' "jellyfish model"
    - Identify core of network as maximal clique (not a robust criterion)
    - Shells around network labeled by hop count from core (a small world)
    - Find that sites with few links often connect to those with high degree
- We consider longer-range connectivity, using k-pruning.
- K replaces degree as indicator of node's role
- K-core, K-shell, and K-crusts result
  - K-shell is "derivative" of K-core, K-crust is union of K-shells
  - Near power-law structure of a new "inflow" region is observed
  - K-shells are not connected, but K-crusts have a giant cluster
- For Erdos-Renyi graphs, K-core is w.h.p. K-connected. For scale free **random** graphs this should also be true.

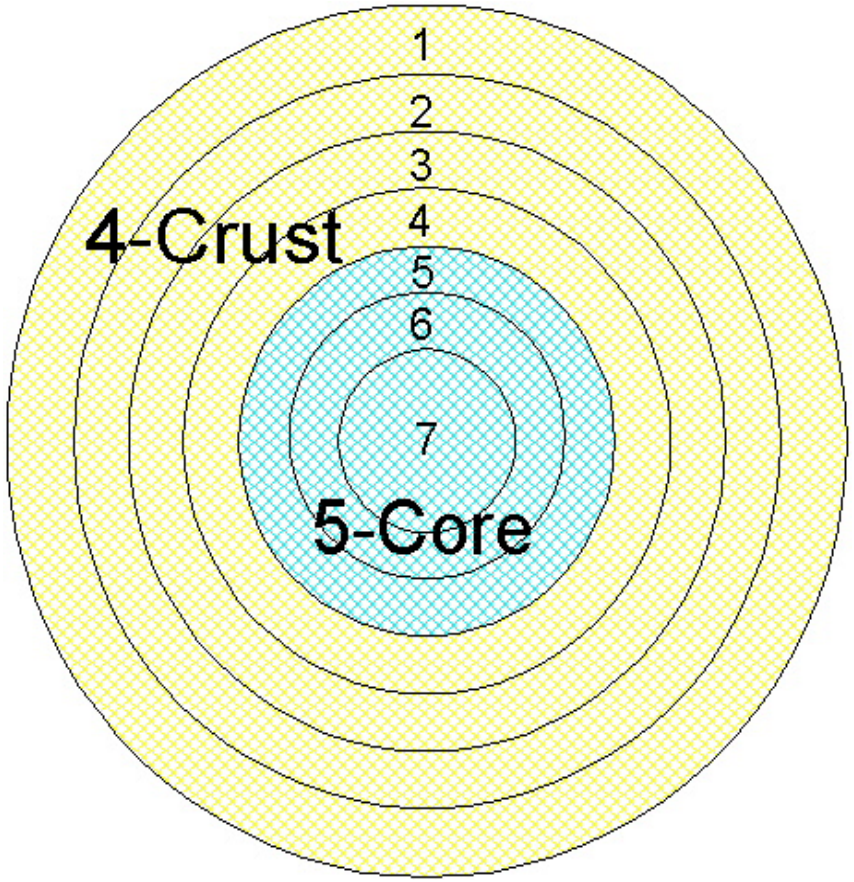
Using the k-core decomposition

- Study 3 components: Isolated sites, peer-connected sites, nucleus
  - since Jan 2005 in DIMES data, now using all of 2005
  - Isolated 25-30% of sites, peer-connected 70-75% of sites, nucleus is tiny
- Nucleus is unambiguously defined, very stable over time
  - 80-90 sites, kmax ~ 40, changes of 3-4 nodes 2Q to 4Q05.
    - Nucleus diameter is 2
  - Contrast with max-clique and extensions – 25% variation
  - World-wide set of international/national carriers, exchange pts
  - Betweeness metric shows congestion goes as N^2, BGP routing increases this

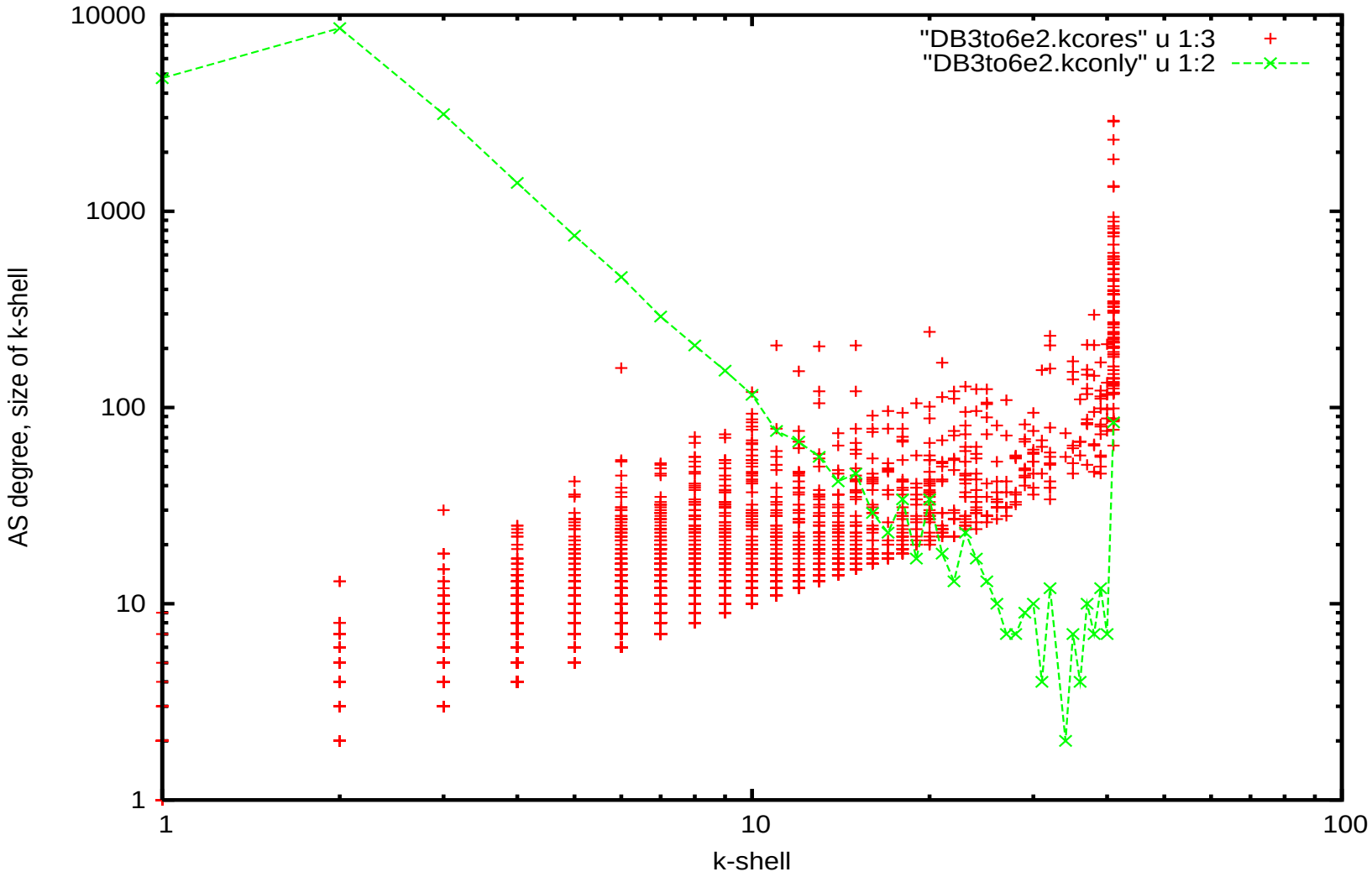
- New ASes predominate in the low-k shells
- The peer-connected crusts have structure which shifts from fractal to regular with increasing k

k-Core Method

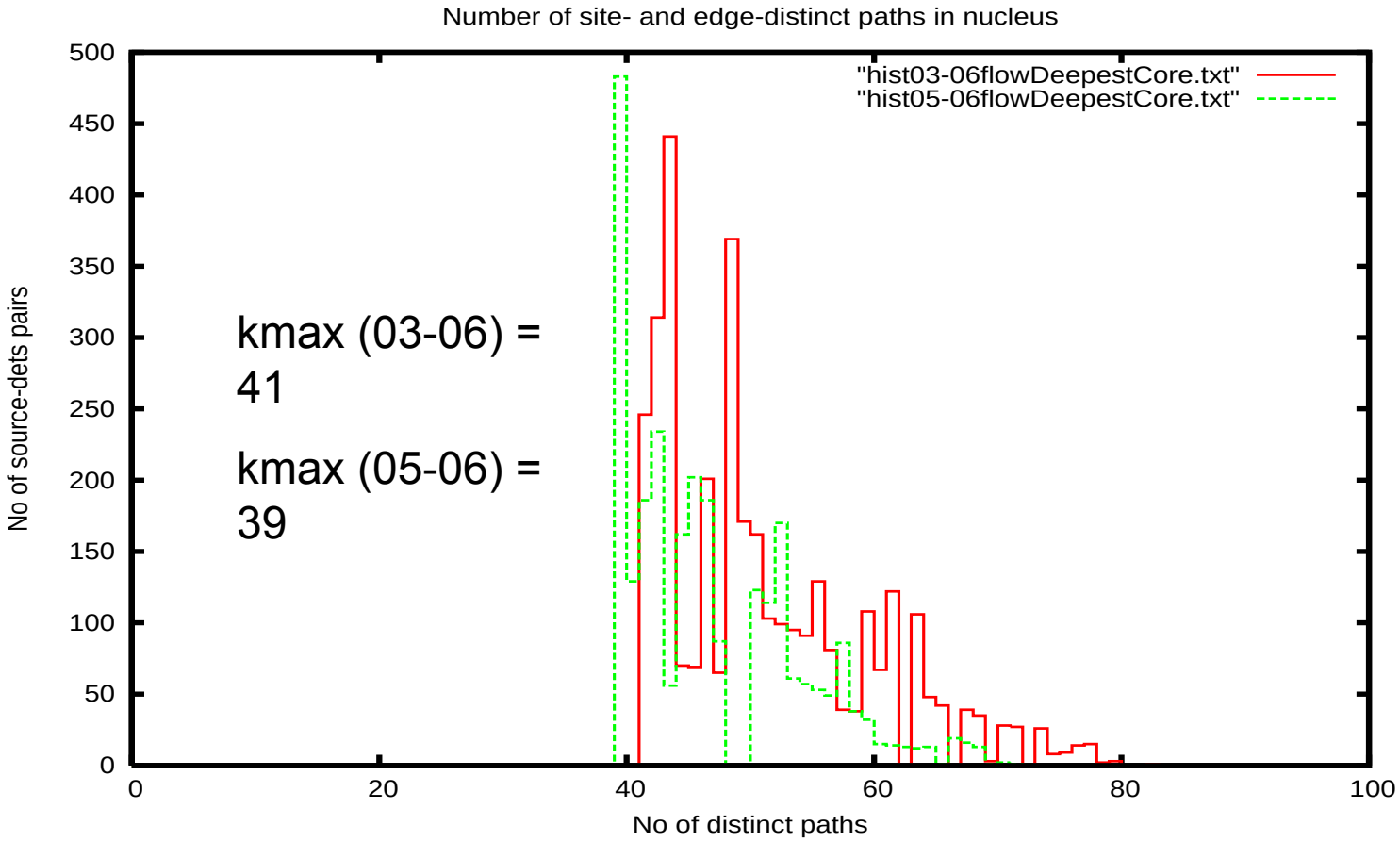
- Some definitions :
- k-Core – union of all shells with indices >= k.
- k-Crust – union of all shells with indices <= k.



How does original degree map into k-shell?



Numbers of site-distinct paths in the nucleus



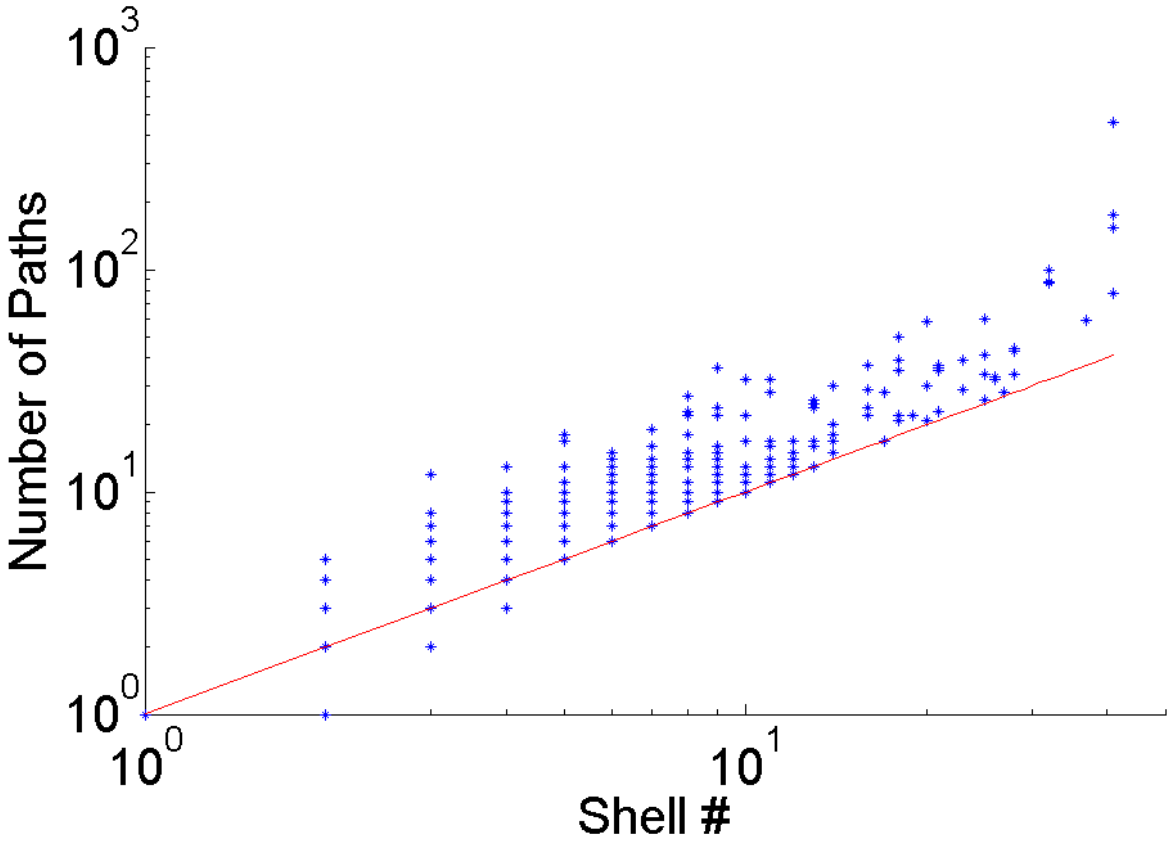
Conclusion: innermost k-cores are k-connected. But outer k-cores (2,3,4) show exceptions (sites with 1,2,3 paths).

Path counting for outer k-shells

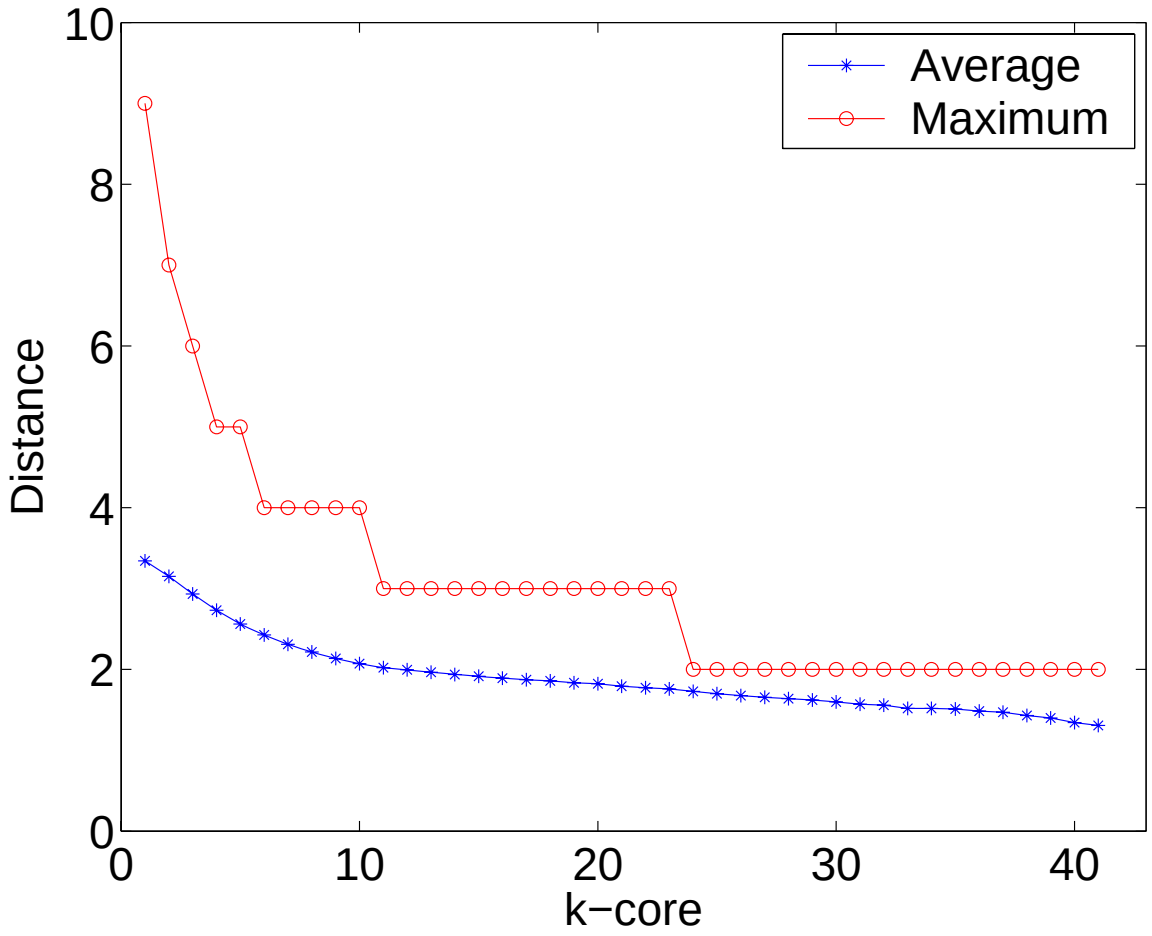
Sampling many pairs.

The shell # is the smaller shell index of each pair.

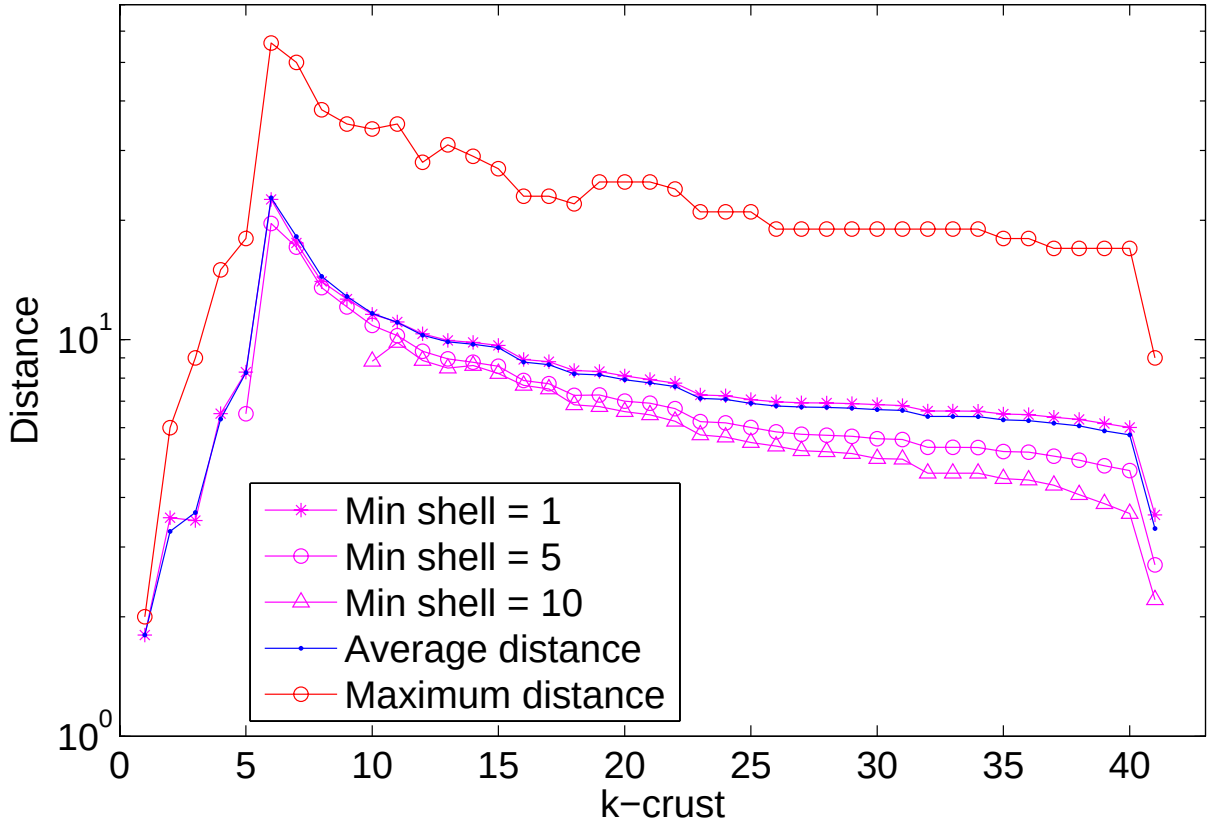
With rare exceptions: #paths >= shell index



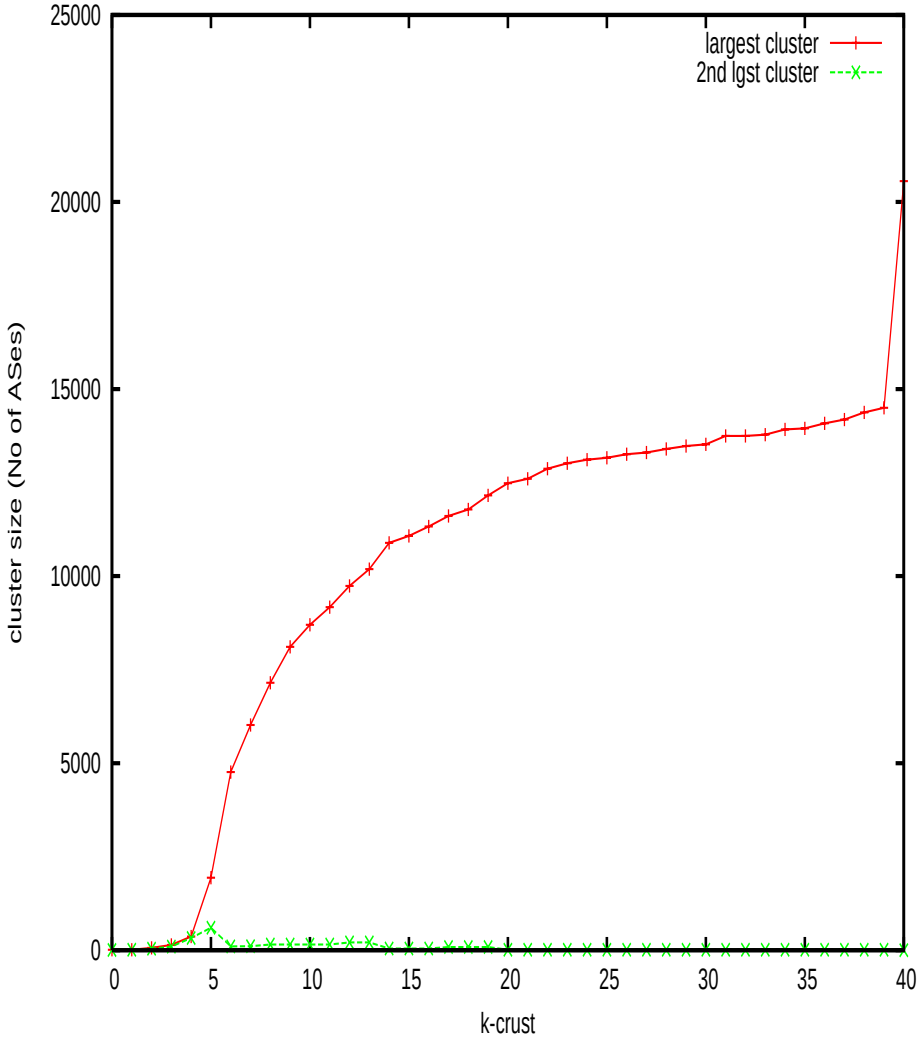
Distances and Diameters in cores



Distances and Diameters



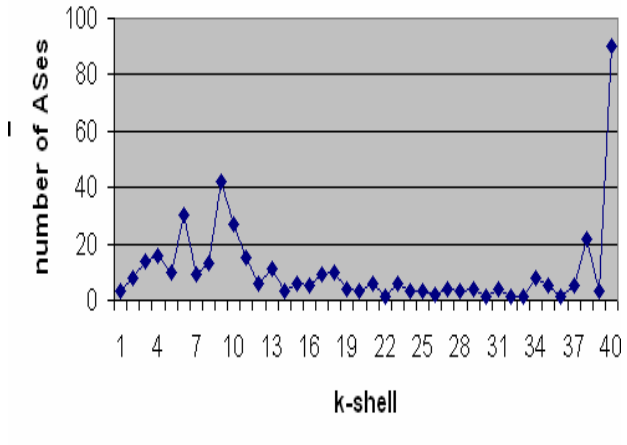
K-crusts show percolation threshold



← These are the hanging tentacles of our (Red Sea) Jellyfish

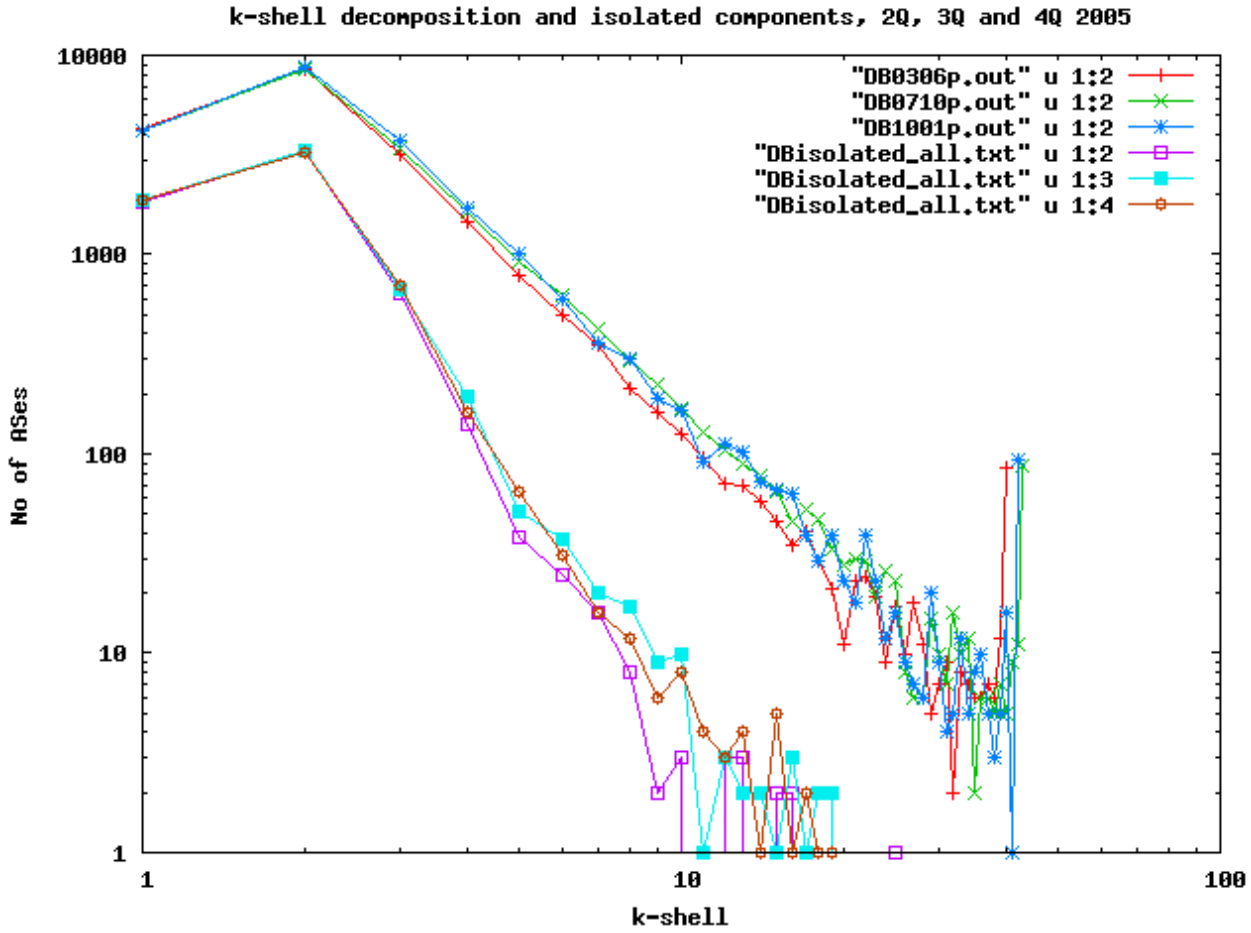
For subsequent analysis, we distinguish three components:
Core, Connected, Isolated

Largest cluster in each shell



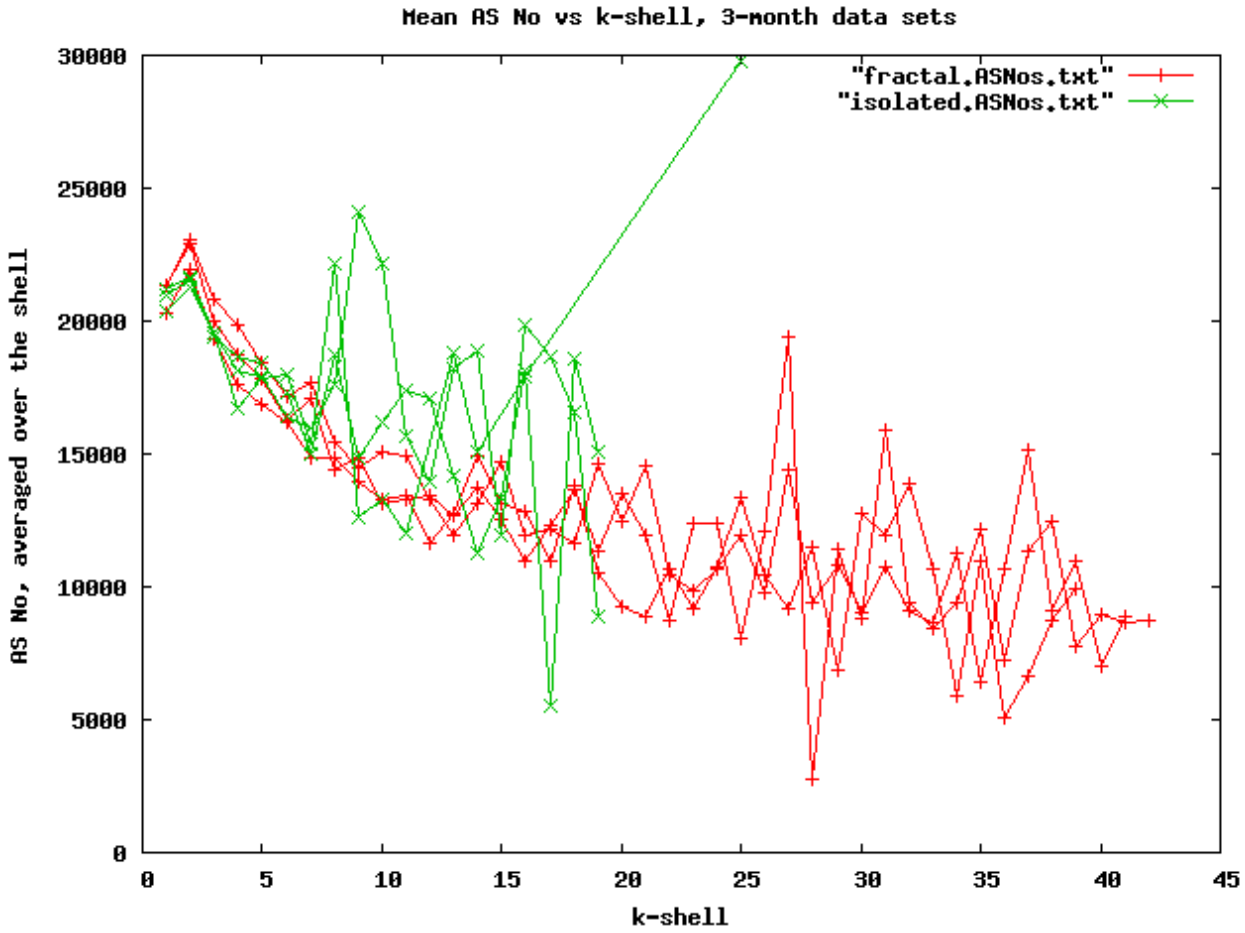
Data from 01.04.2005

Structure persists over 2Q, 3Q, 4Q 2005

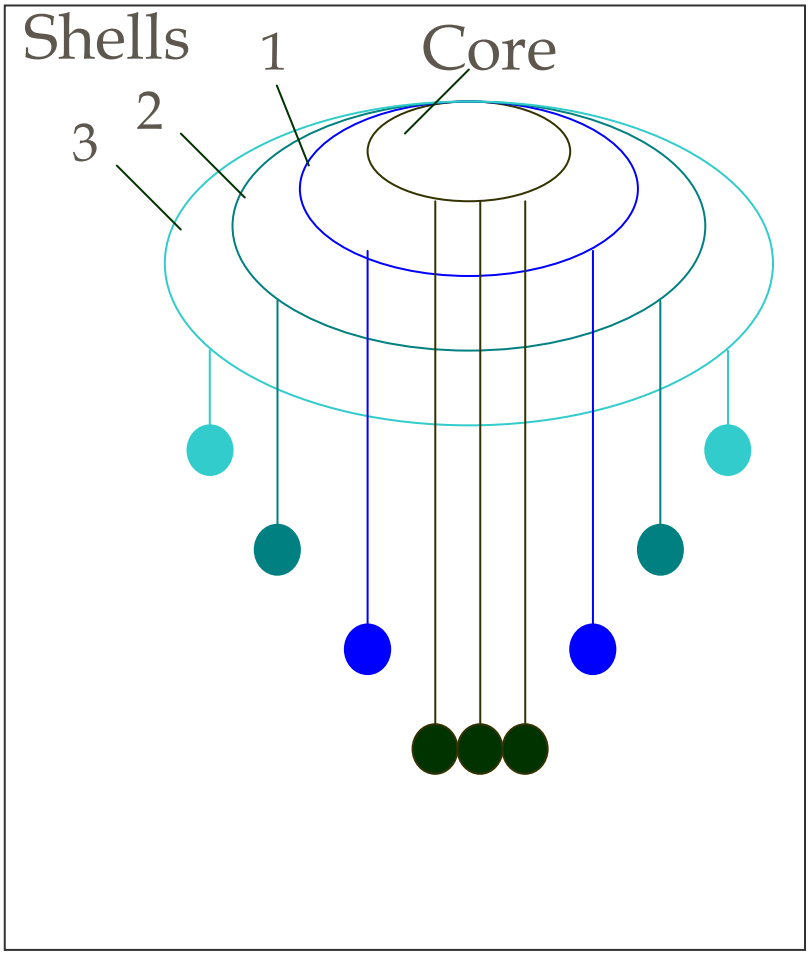


Source: DIMES cleaned data (no AS loops)

Newer ASes (higher AS No) found at low k

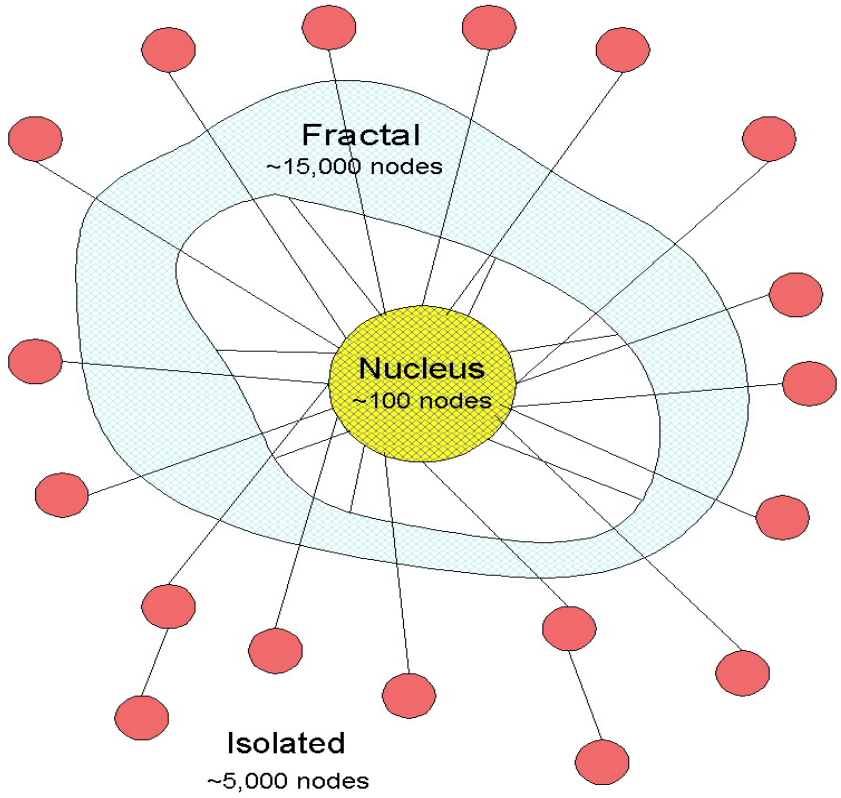


Michalis Faloutsos' Jellyfish



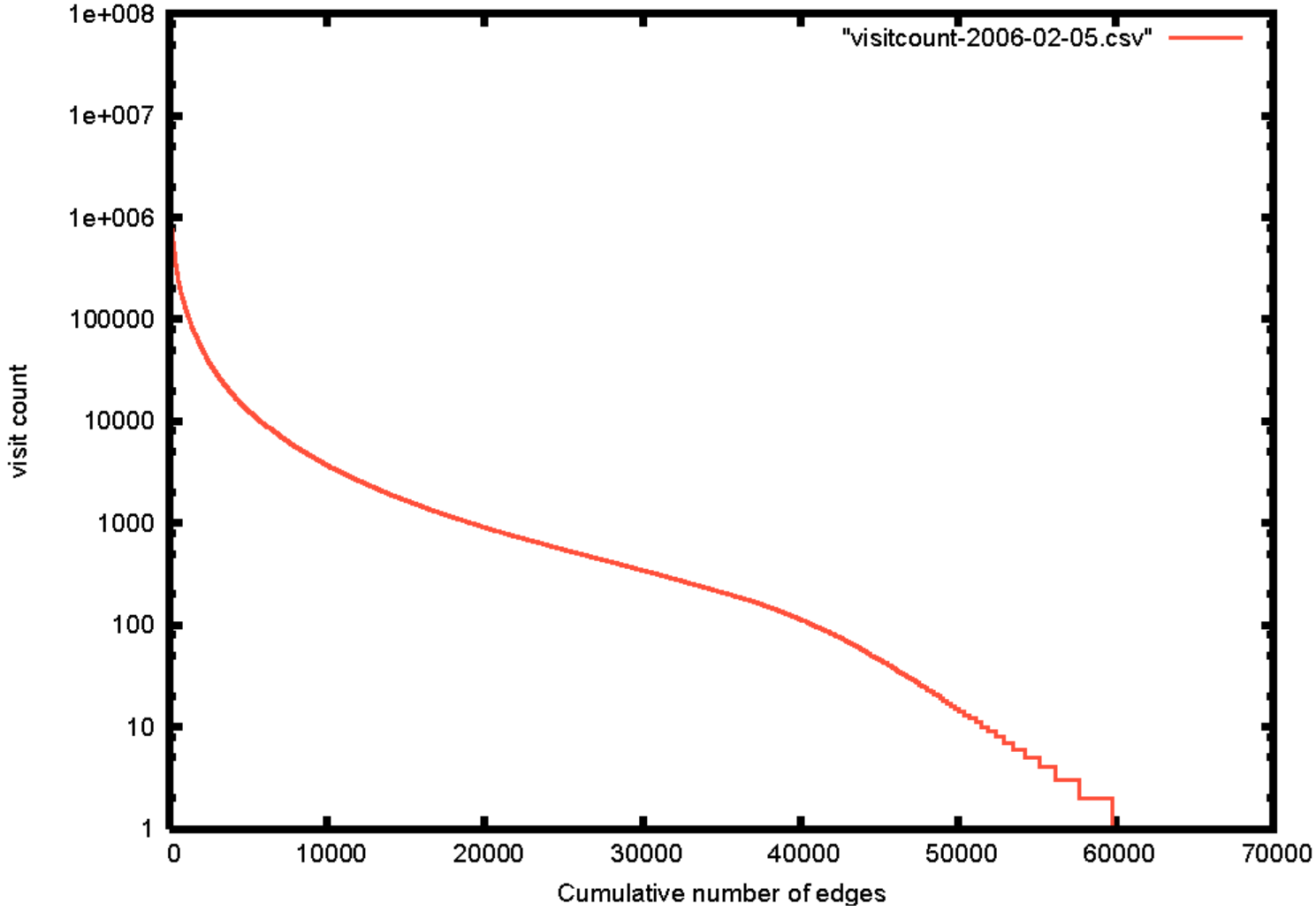
- Highly connected nodes form the core
- Each Shell: adjacent nodes of previous shell, except 1-degree nodes
- **Importance** decreases as we move away from core
- 1-degree nodes hanging
- The denser the 1-degree node population the longer the stem

Meduza (מדוזה) model



This picture has been stable from January (kmax = 30) to present day, with little change in the nucleus composition. The precise definition of the tendrils: those sites and clusters isolated from the largest cluster in all the crusts – they connect only through the core.

What sort of coverage is obtained?



Who's "tier-1" in Medusa?

701	2992
7018	2766
3356	2665
1239	2619
174	1967
209	1387
12956	1261
1299	1251
3549	1219
3561	1215
2914	998
7132	951
702	923
6730	923
6461	907
4323	772
1273	728
3491	687
6453	644
3303	612
3320	590
6939	584
2828	577
4513	570
4637	544
7911	542
8220	531
5400	522
1221	508

1668	496
16150	460
6395	453
3257	450
286	391
3246	389
8342	387
5511	384
4766	367
25462	365
8928	360
7473	359
3292	347
3786	343
2516	330
3209	329
12989	327
6539	317
6320	283
10026	283
6695	277
3352	263
8001	259
1257	258
22773	250
6327	247
5650	245
19151	239
13237	237

8075	226
2497	225
15412	213
6762	208
19029	206
4589	203
5459	202
5089	197
852	180
5462	176
15290	174
577	156
2856	153
8546	153
9318	145
6079	137
13768	136
4725	133
22822	128
293	122
4134	122
3300	117
4355	113
6830	110
12322	108

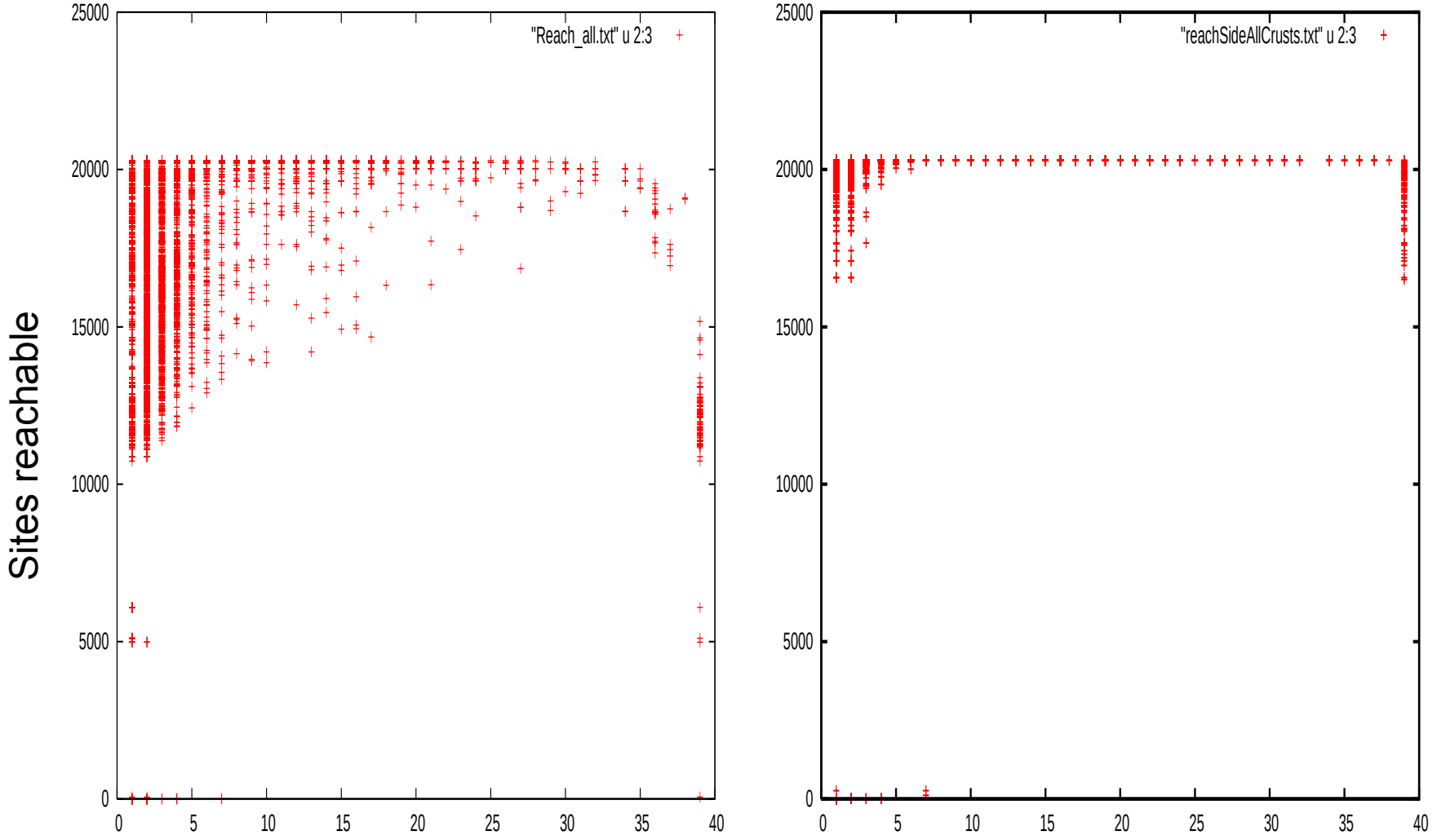
4436	98
6389	96
8210	95
4788	93
23352	89
19548	87
23342	80
10310	75
812	64
15169	50

Data from months 10-12,
2005 kmax = 42, 93 nodes

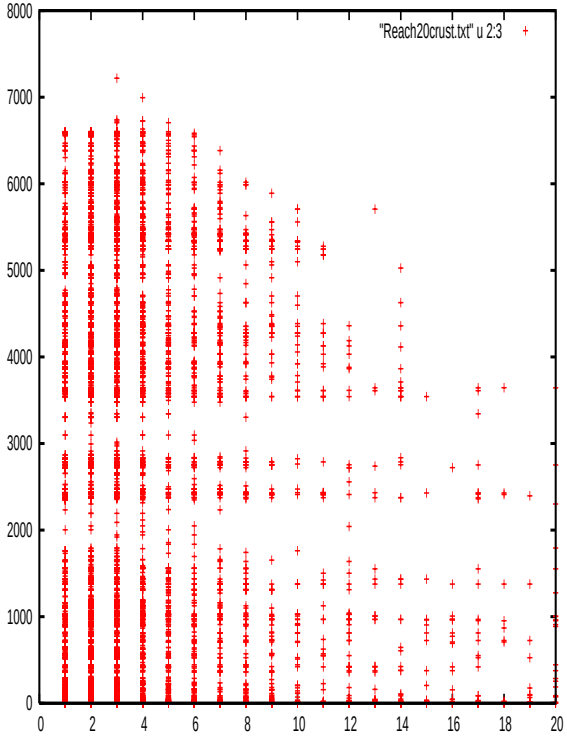
Summary and next steps

- New data permits closer study of AS graph structure
  - Much data not seen in previous BGP-based studies
  - Analyze as a function of k-shell, instead of simply degree
  - Major deviations from simple random models
    - Must be understood to develop good Internet generator
- Reachability is not percolation, but can be evaluated
  - Decision to transmit a message depends on sender and destination, not simply on the existence of a link
  - Cost of evaluating uphill-downhill reachability is comparable to shortest path

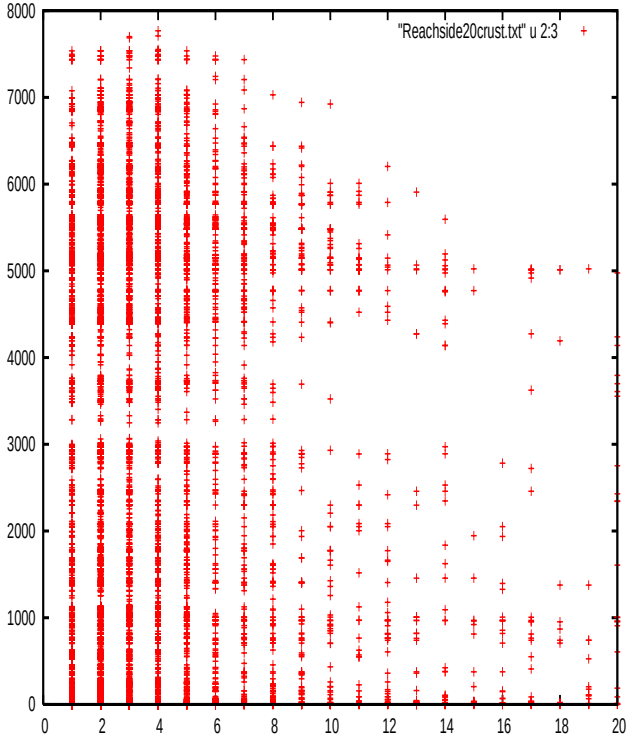
Preliminary reachability data (using whole graph)



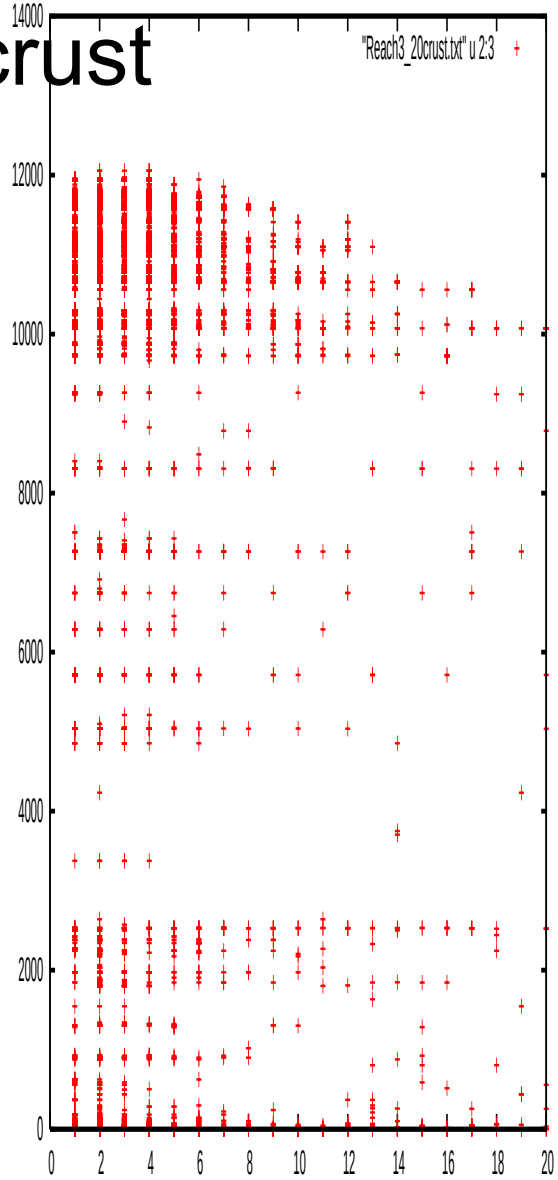
Now restrict to the 20-crust



Up then down



Side step at top



Three sidesteps

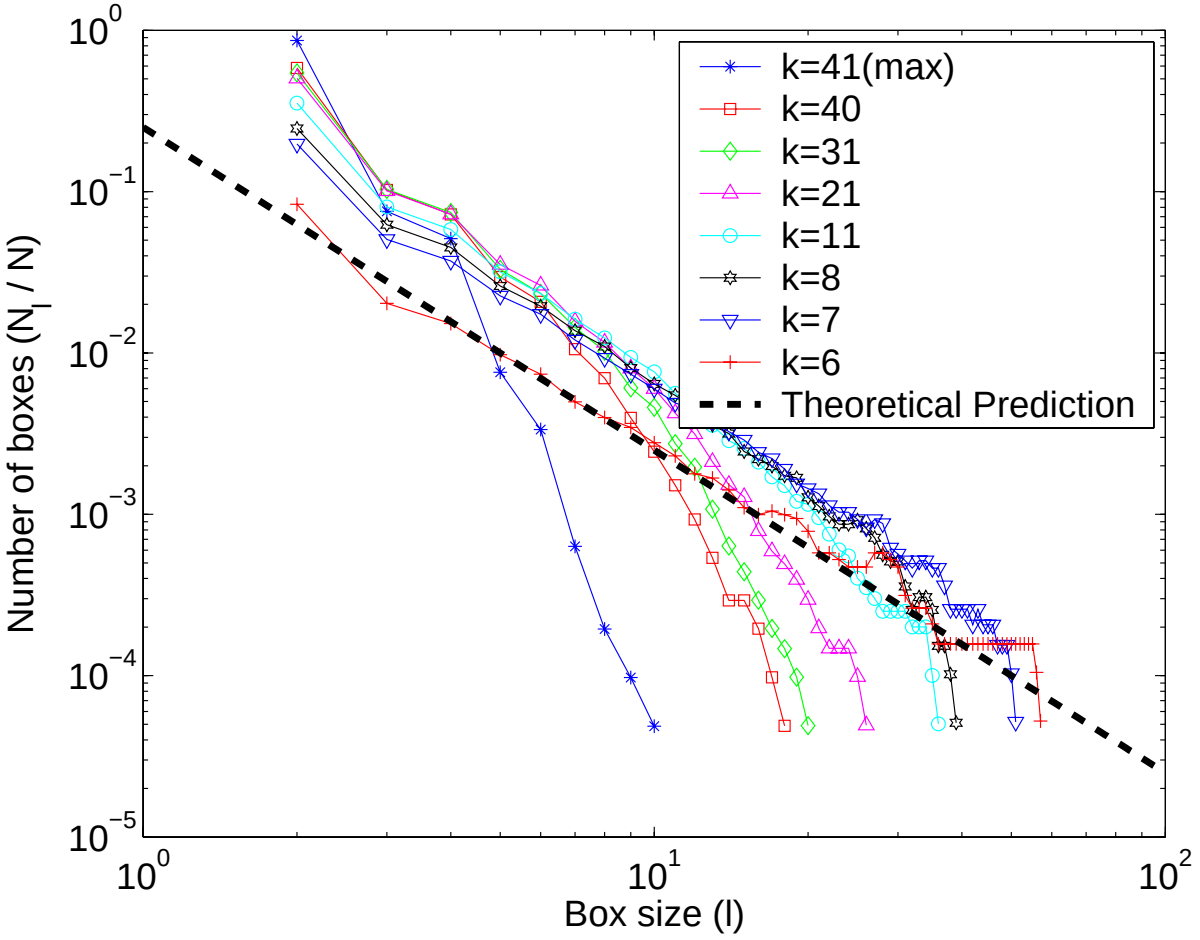
Some further conclusions

- Up-down routing is much more restricted than possible connectedness, a deficit due to BGP routing, which reflects present business arrangements. (Casts them into concrete, actually.)

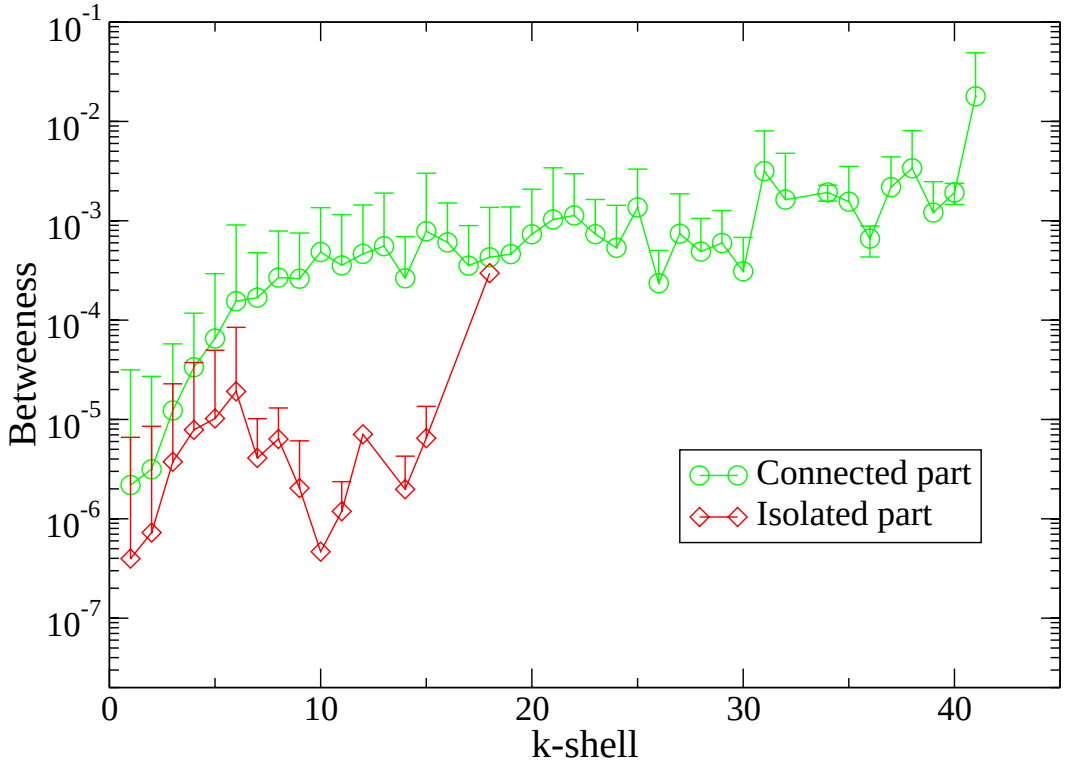
- This difference will inhibit growth of P2P distributed solutions.

Backup slides

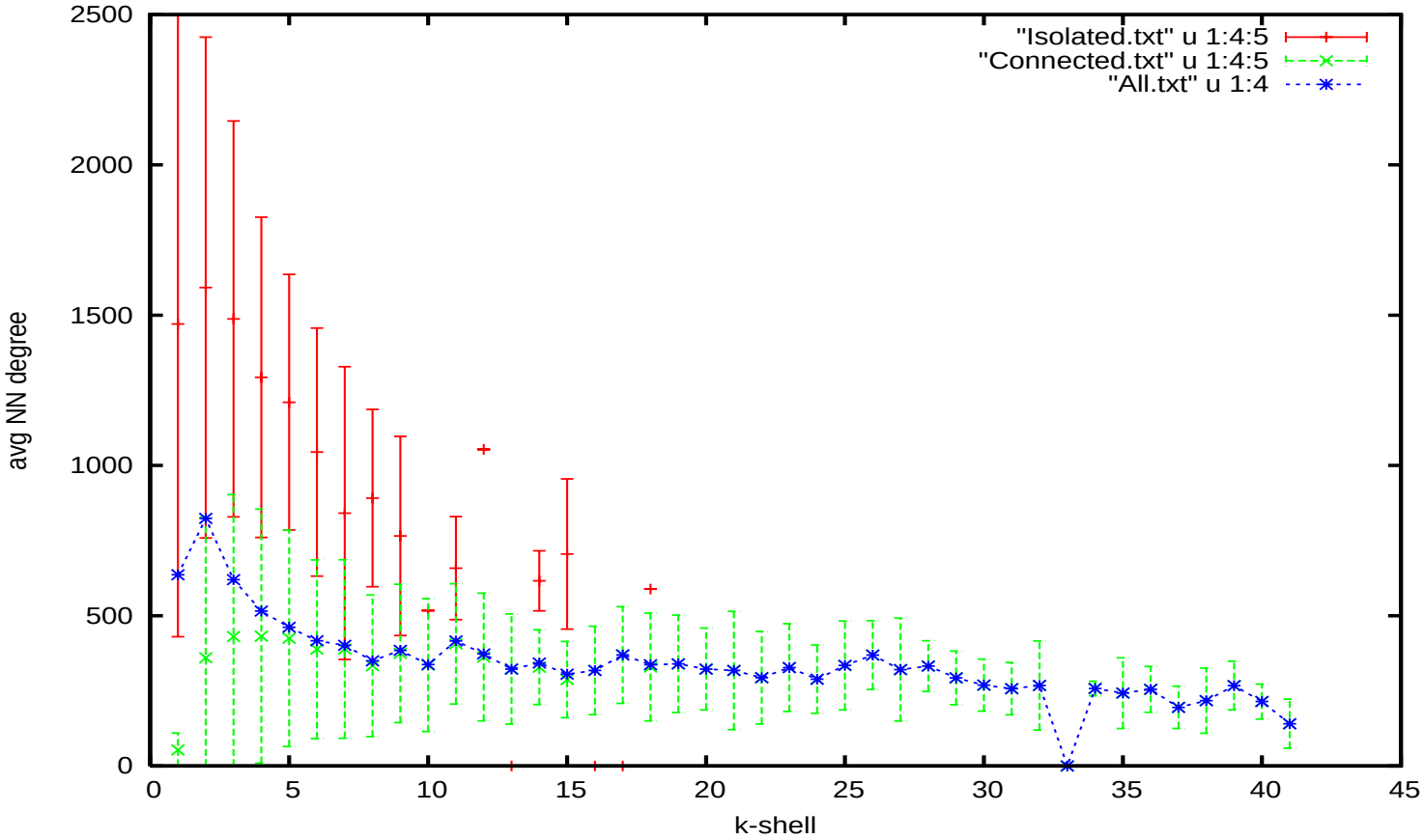
OK, we've got power laws, but is it a fractal topology?



Betweeness of the components

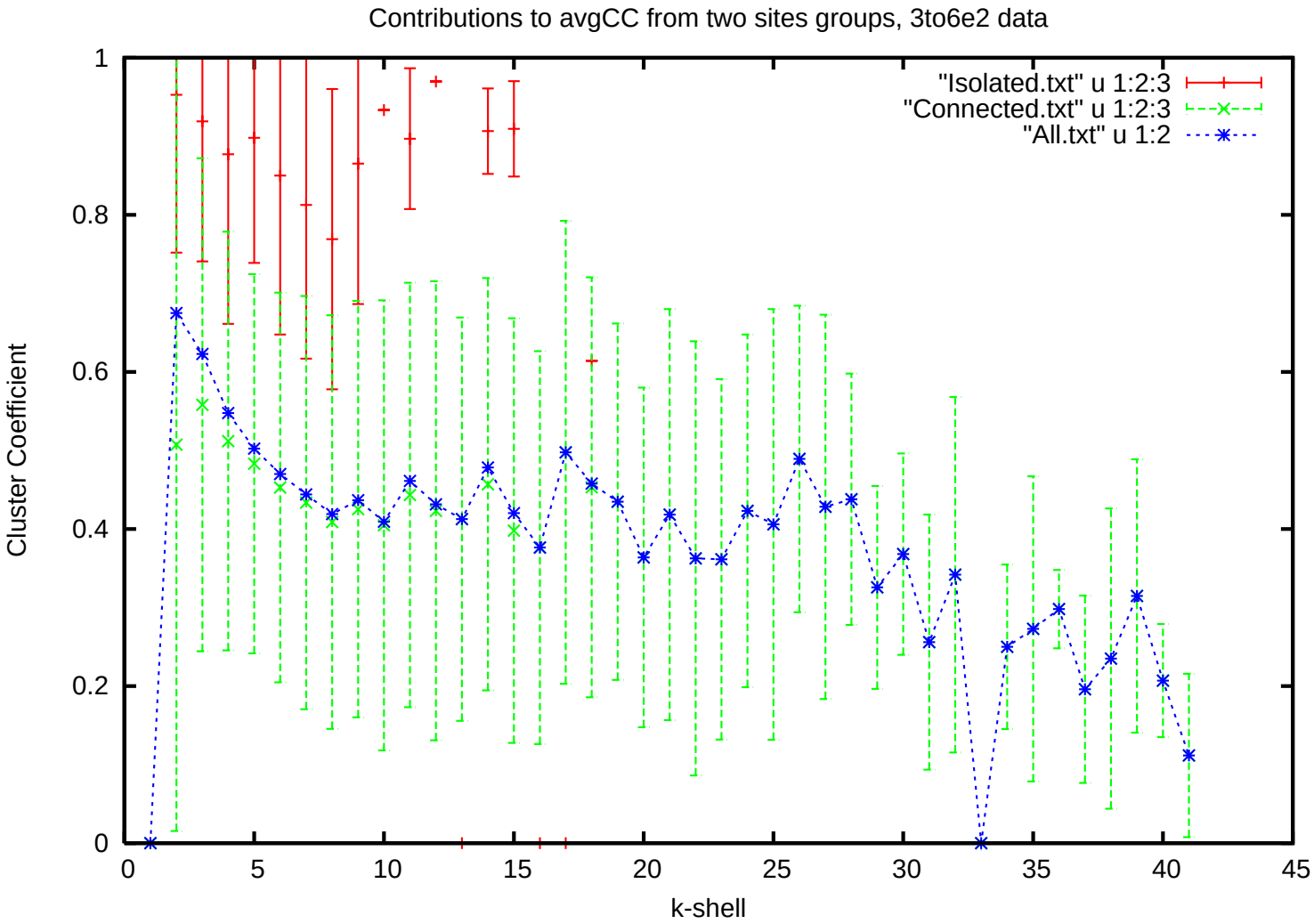


Avg NN degree also separates

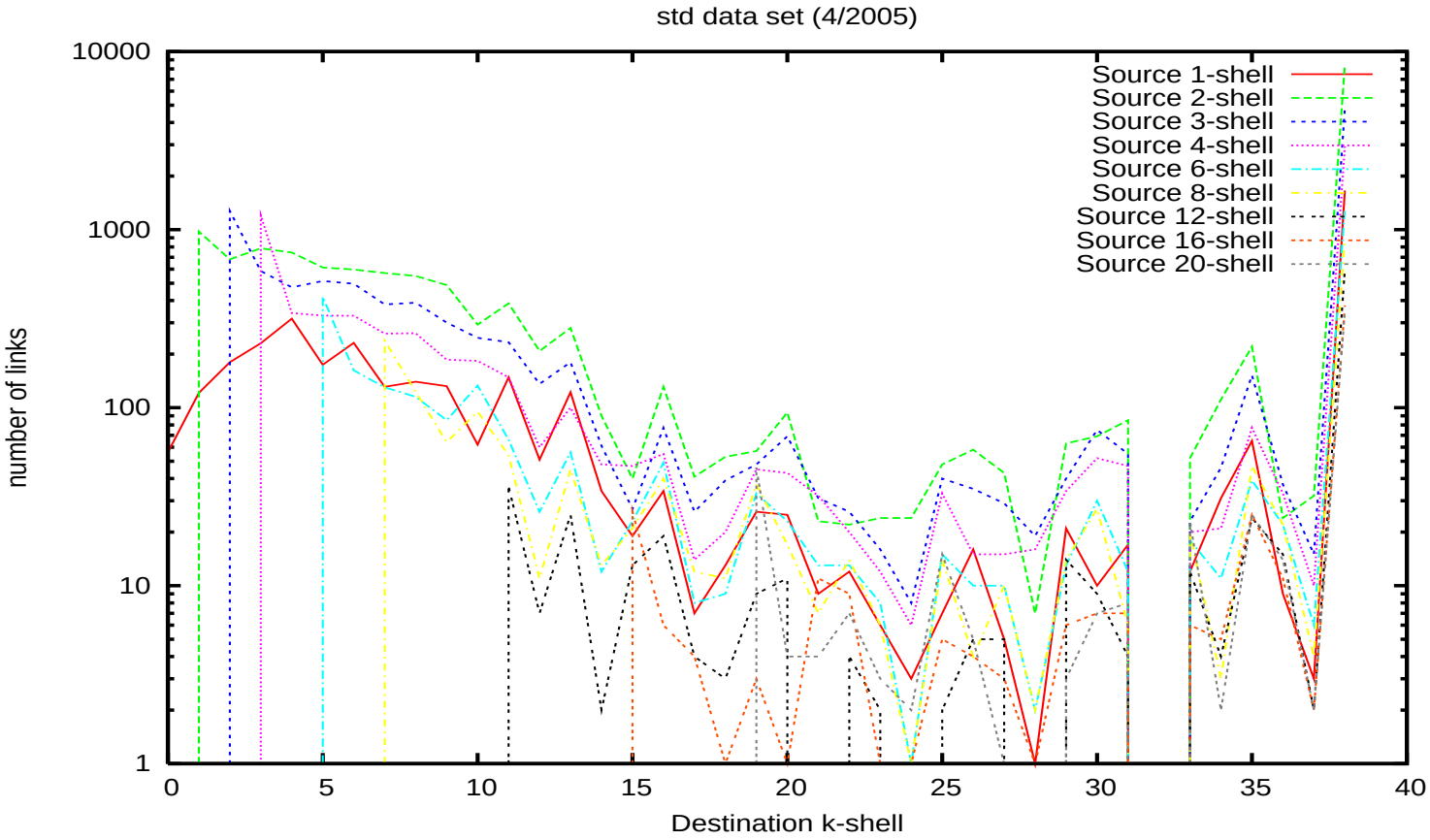


Data from March to June, 2005
(seen at least twice)

Similarly for cluster coefficient



Where do the links go in Medusa?



Early shells (1-10) link to intermediate shells as well as to the core.

Links per site of k-shells to k-core (above) and to k-crust (below)