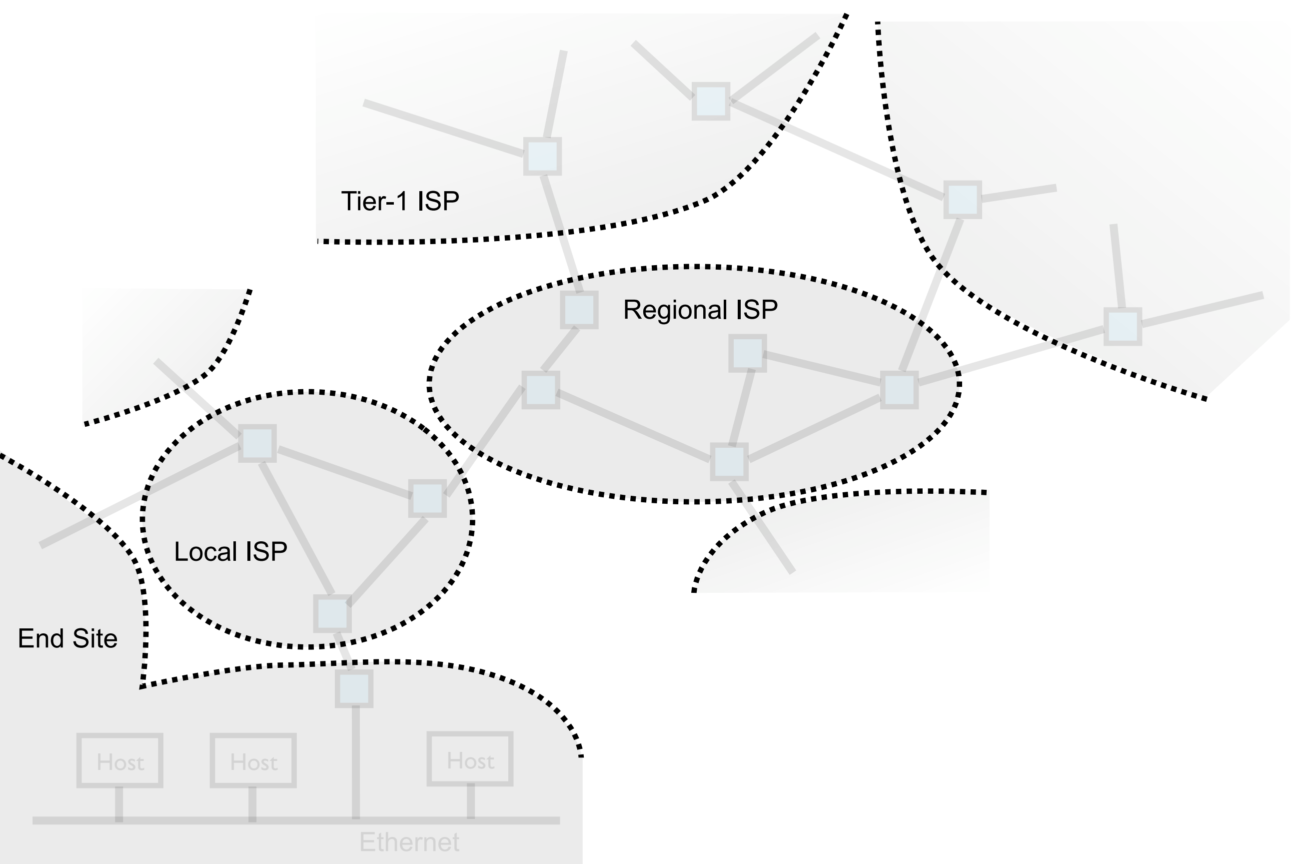
University of Glasgow

Compact Routing for the Internet

Colin Perkins
Stephen Strowes
Graham Mooney

Internet Routing

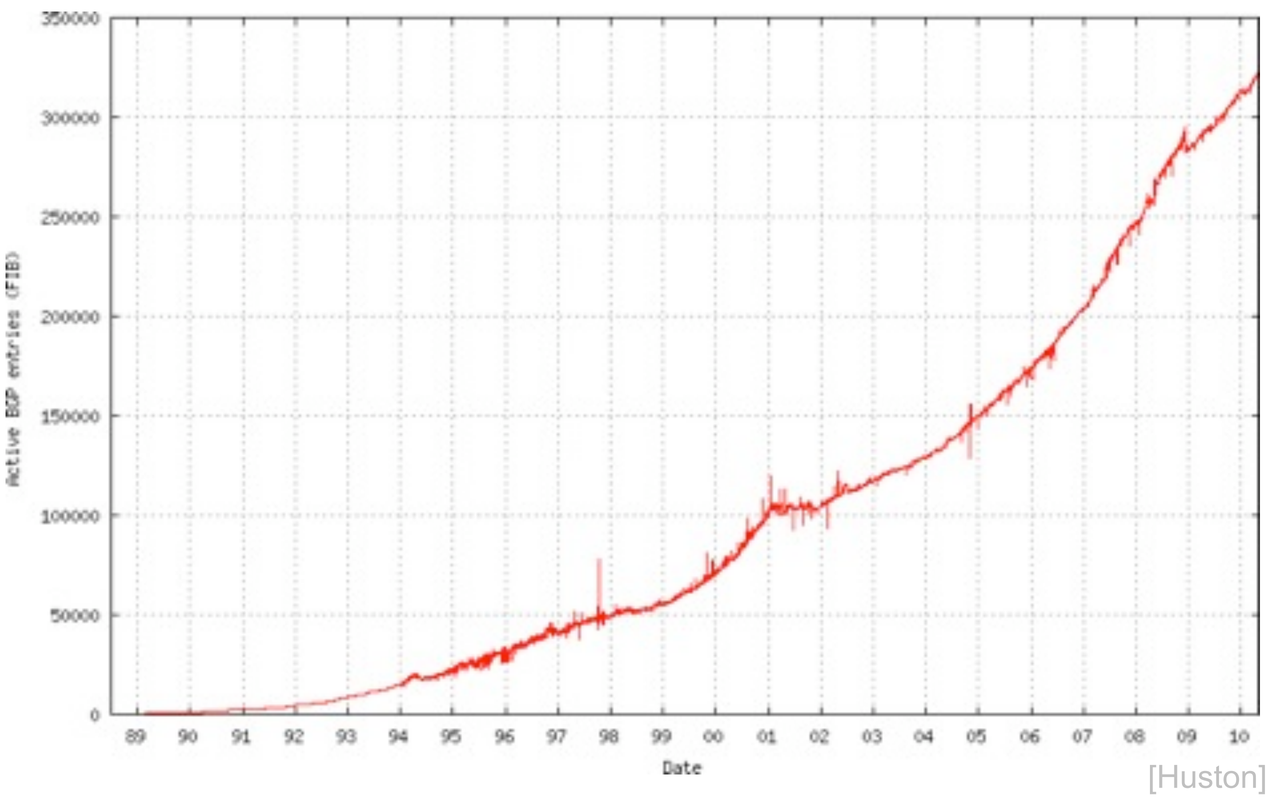
- Network of networks – the Internet AS graph
  - Each network is an Autonomous System (AS)
    - ~33,000 ASs in the Internet
  - Each network owns one or more address ranges, identified by prefix
- Inter-domain routing via BGP
  - Each AS advertises its own network prefixes, and those of it's customers



  - ~350,000 prefixes advertised
  - Path vector routing, longest prefix, shortest path
  - Policy via path inflation, selective advertisement, de-aggregation

Limitations of BGP

- Growth of routing table
  - Natural growth
  - Due to multi-homing
  - Due to traffic engineering
- Increased rate of change



[Huston]

- Concerns about long-term scalability
  - Exponential growth in routing updates [Huston]
  - Super-linear growth in routing tables sizes
    - Somewhat reduced in the past 18 months (recession?)
    - De-aggregation due to IPv4 exhaustion likely to cause increased rate of growth

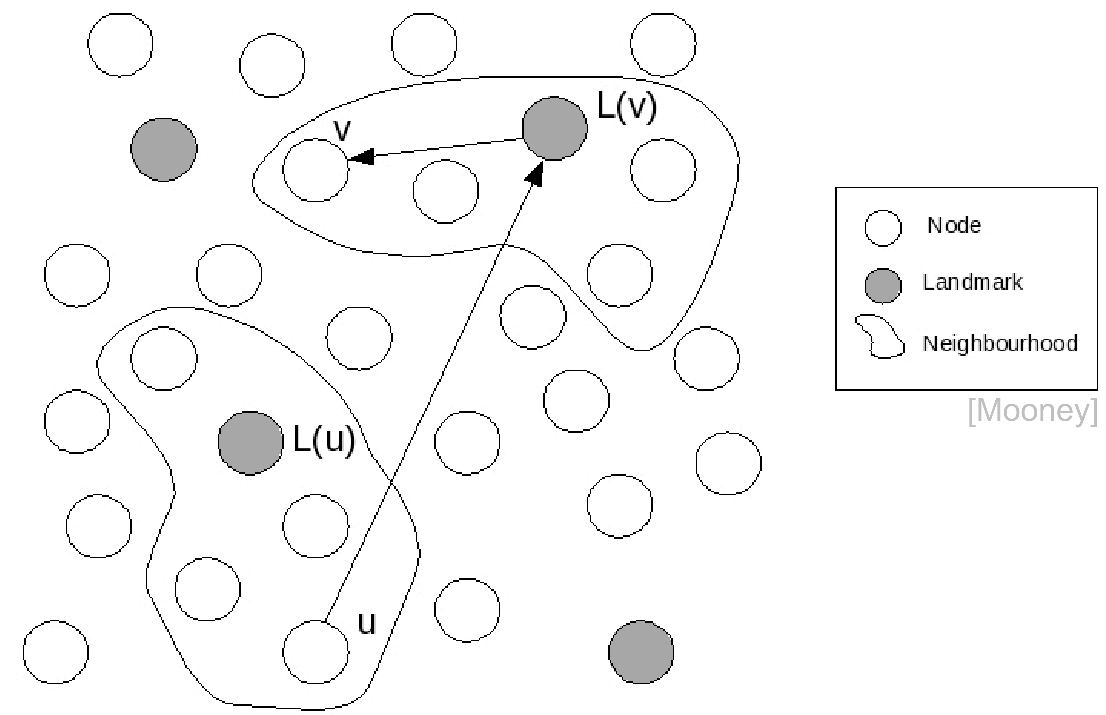
Compact Routing

- Key goal: routing tables with sub-linear scaling
  - Sub-linear growth in routing table size w.r.t. AS graph size
  - Cost: give up on shortest path routing – but stretch provably ≤ 3
- Algorithms only, not concrete routing protocols
  - Currently only defined for static graphs
  - Doesn't account for routing policy
  - Will require fundamental changes to the Internet architecture to deploy
- Promising initial results
  - Average stretch ~1.1 on Internet-like synthetic graphs [Krioukov, Infocom 2004]
  - Linear growth in routing updates [claffy, ACM CCR 37(3), 2007]

The Thorup-Zwick (TZ) Algorithm

[Thorup & Zwick, SPAA'01]

- Landmark-based
  - Random initial landmark set
    - Each has a neighbourhood of nodes closer to it, than it is to it's landmark
    - Iteratively balance neighbourhood sizes, creating new landmarks in large neighbourhoods
  - Route via landmarks
    - Packet headers contain destination and landmark addresses
    - Route towards landmark if outside destination neighbourhood, else route directly to destination
- Name-dependent
  - Small routing tables require a specific naming scheme
    - Cannot use AS-numbers or IP addresses for routing



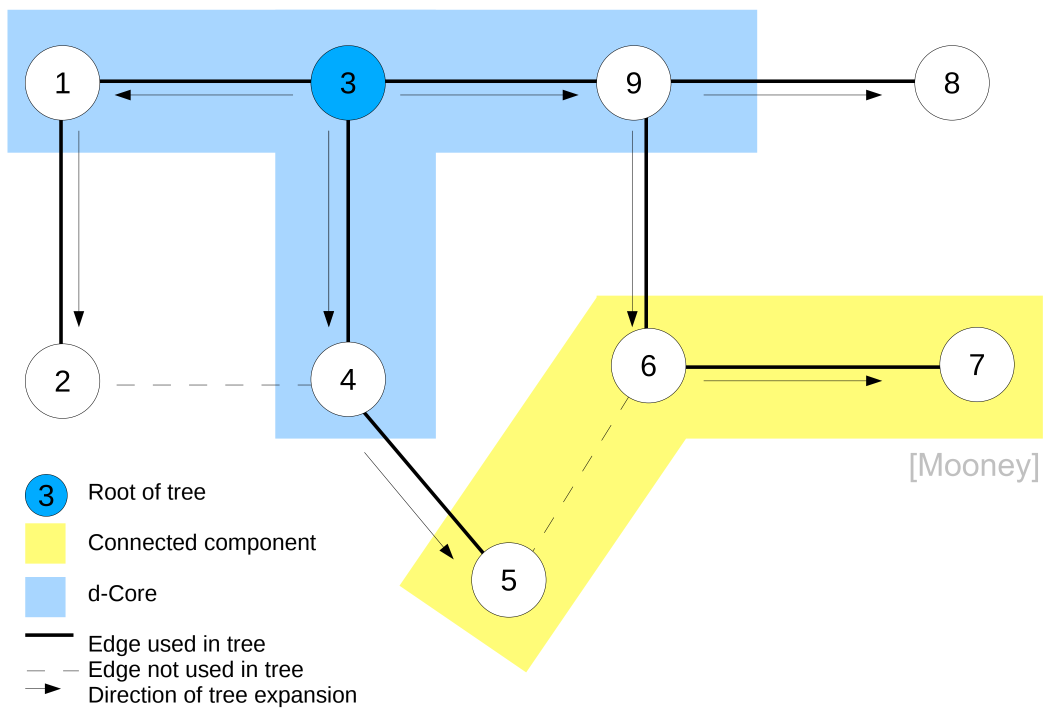
[Mooney]

- Routing table scales as $O(n^{1/2})$
  - Nodes store landmark set and addresses of neighbourhood nodes

The Brady-Cowen (BC) Algorithm

[Brady & Cowen, ALENEX'06]

- Build forest of spanning trees
  - First routed at highest-degree node
  - *d*-core is all nodes within $^{d}/_{2}$ hops of highest degree node; *d*-fringe is the remainder
  - Build additional spanning trees to cover connected components in the *d*-fringe
- Re-label nodes in trees
  - Algorithms due to Thorup-Zwick & Peleg
  - Efficient routing in trees with small labels
- Routing
  - Choose appropriate spanning tree
  - Routing in the tree based on node labels



Choice of *d* critical for performance

Routing table scales as $O(\log^2 n)$

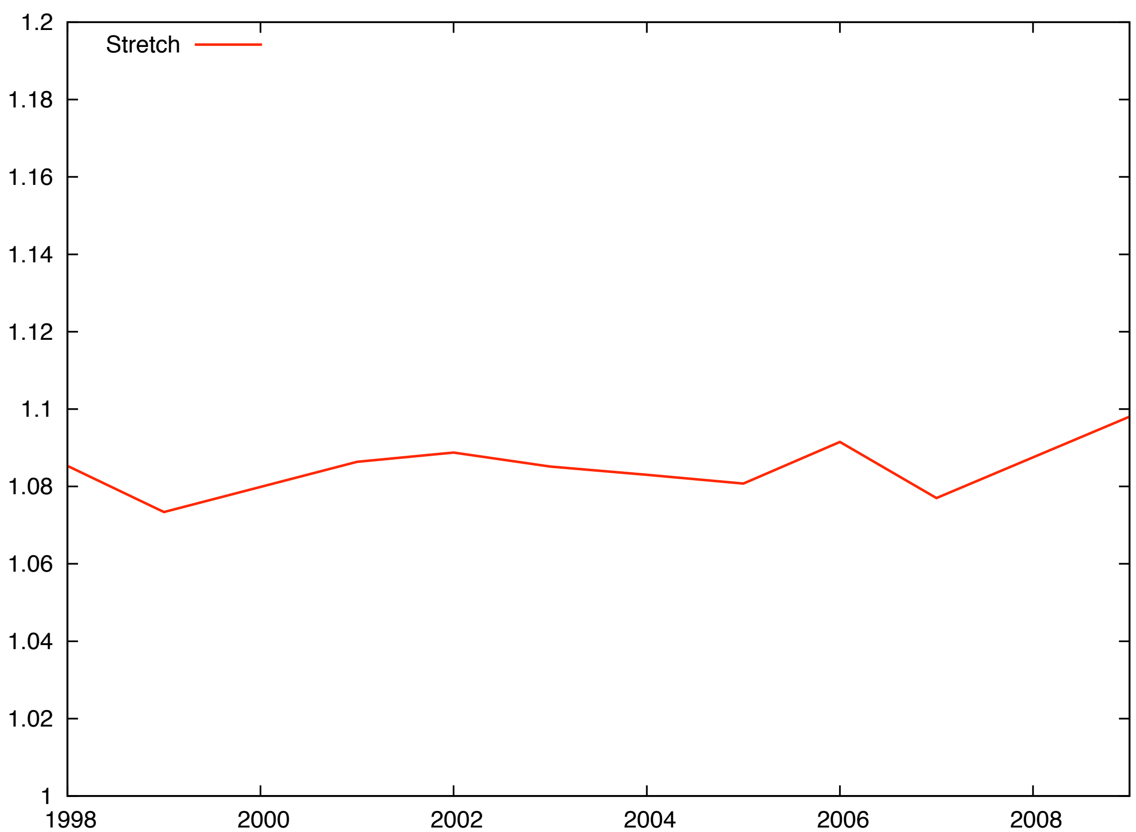
Performance Evaluation

- Evaluate stretch and routing table size of TZ and BC algorithms on snapshots of Internet AS graph
  - (Path stretch simulations for BC algorithm for future work)
  - Use BGP routing table data from CAIDA and RouteViews
  - Annual snapshots from March 1998 to March 2009

- Determine whether results from synthetic graphs are repeated on the real-world Internet topology

Path Stretch Distribution – TZ

[Mooney]



Mean path stretch over time

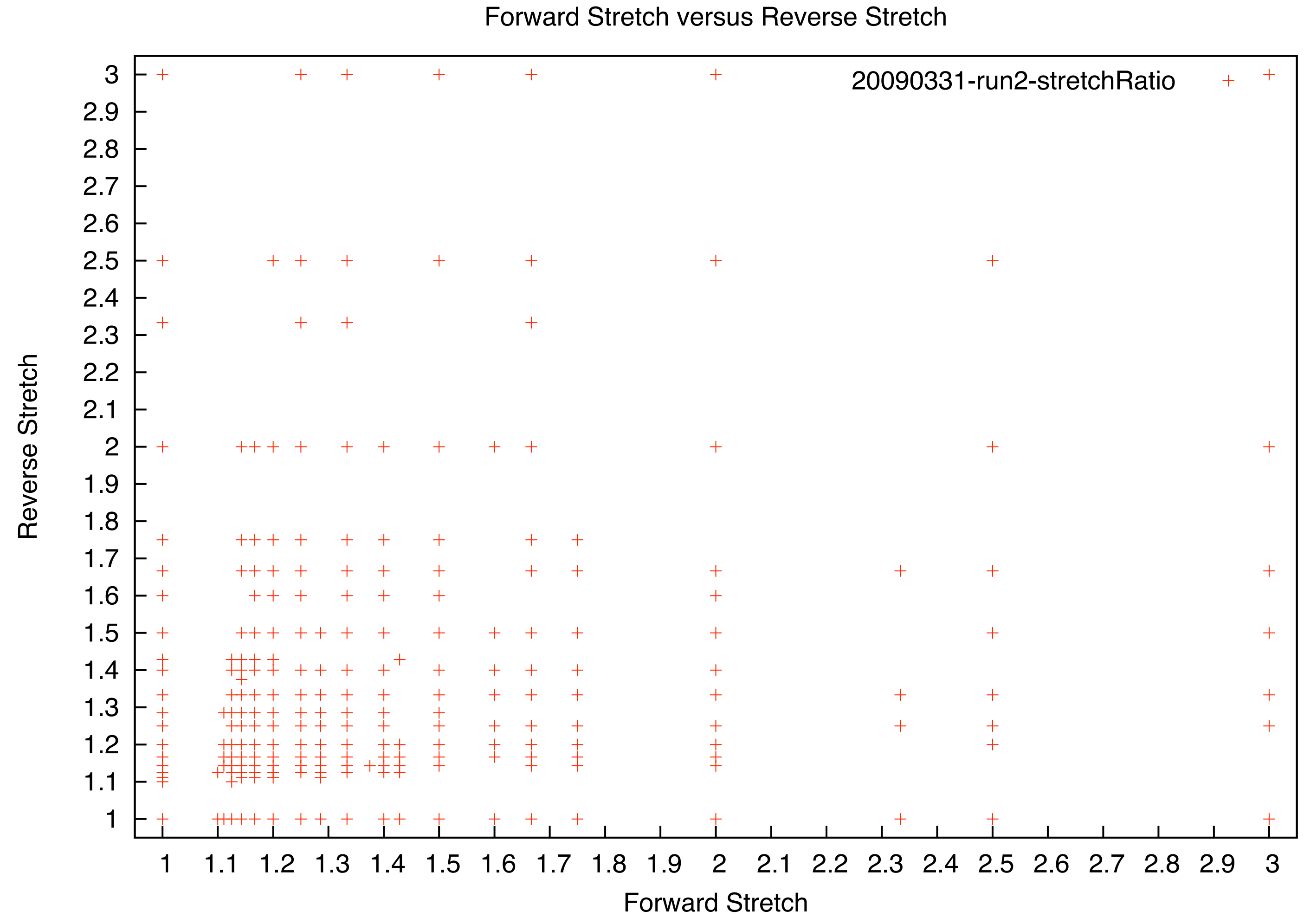
Sample path stretch distribution (log y-axis)

- Measure stretch from each node to 1% of the other nodes, randomly chosen (measure both forward and reverse path stretch)
- Average stretch slightly better than Krioukov's results on Internet-like graphs; Remarkably stable average stretch and stretch distribution over time – the Internet appears to be a near-ideal network for TZ compact routing

Path Stretch Distribution – TZ

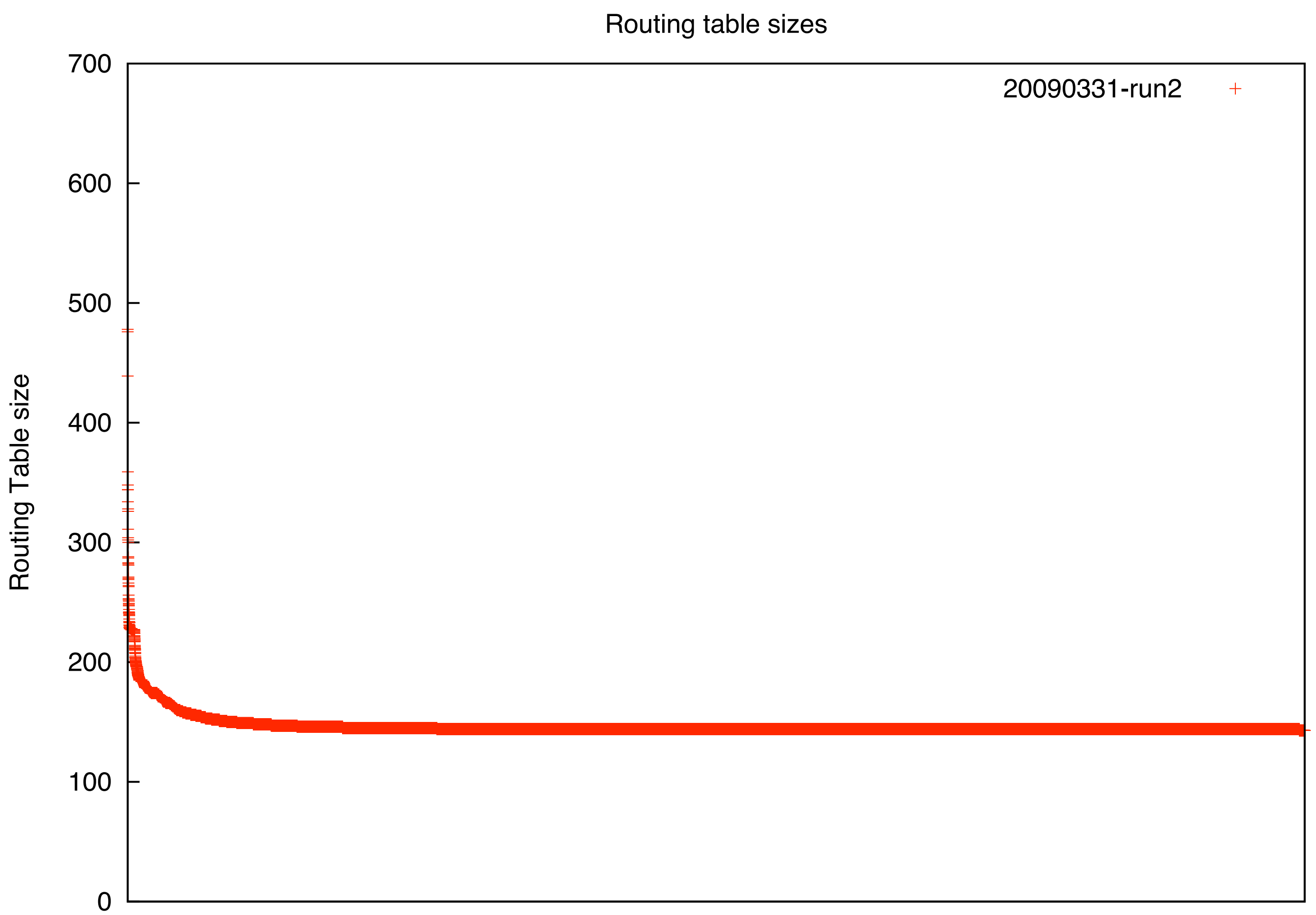
[Mooney]

- Majority of paths are low-stretch in forward and reverse direction
- High degrees of path asymmetry exist, but are uncommon



Routing Table Size – TZ

[Mooney]

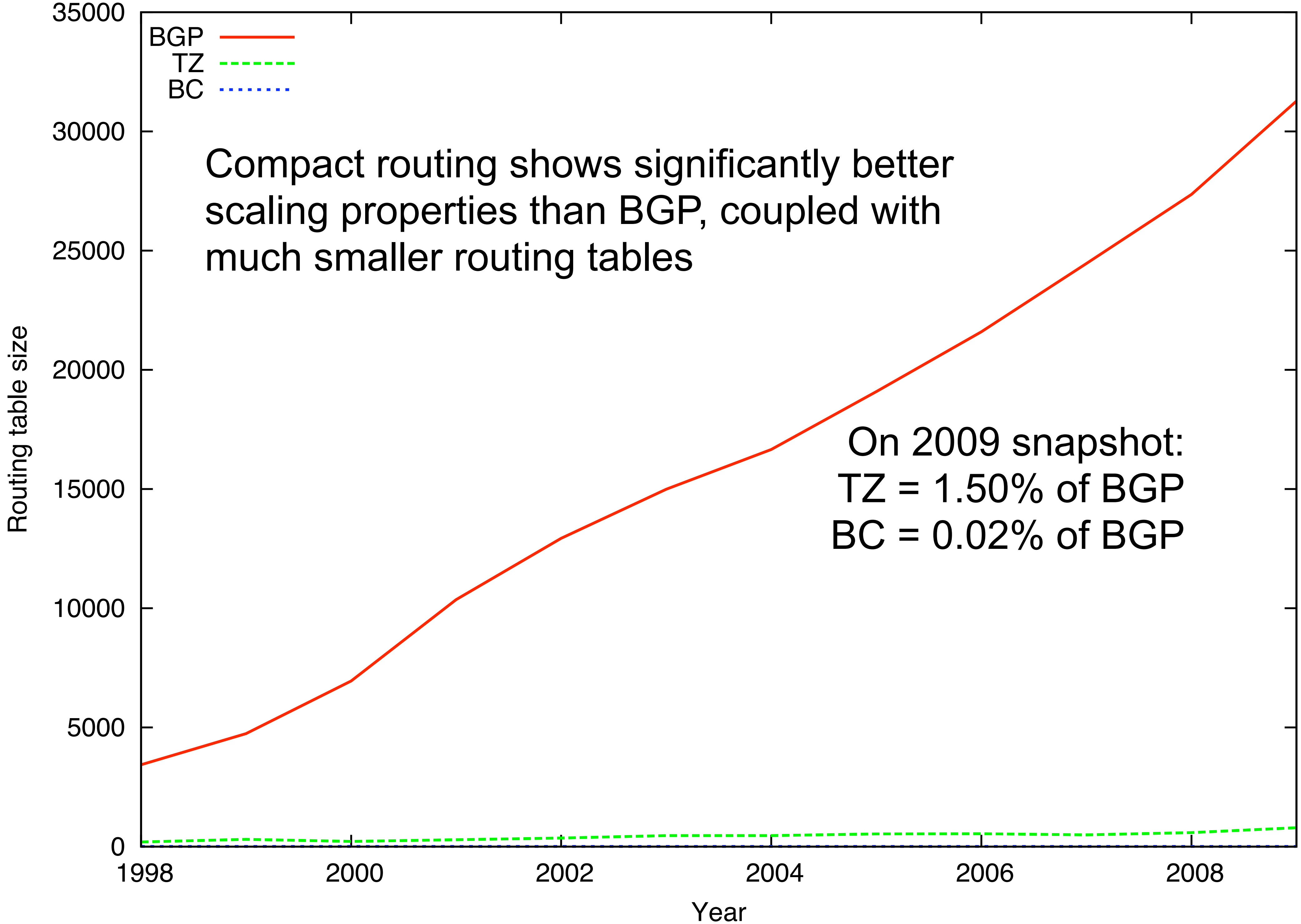


Sample routing table size distribution

- Per-node routing table size depends on node location
- Average size of routing table: 140 entries
- Worst case: 476 entries

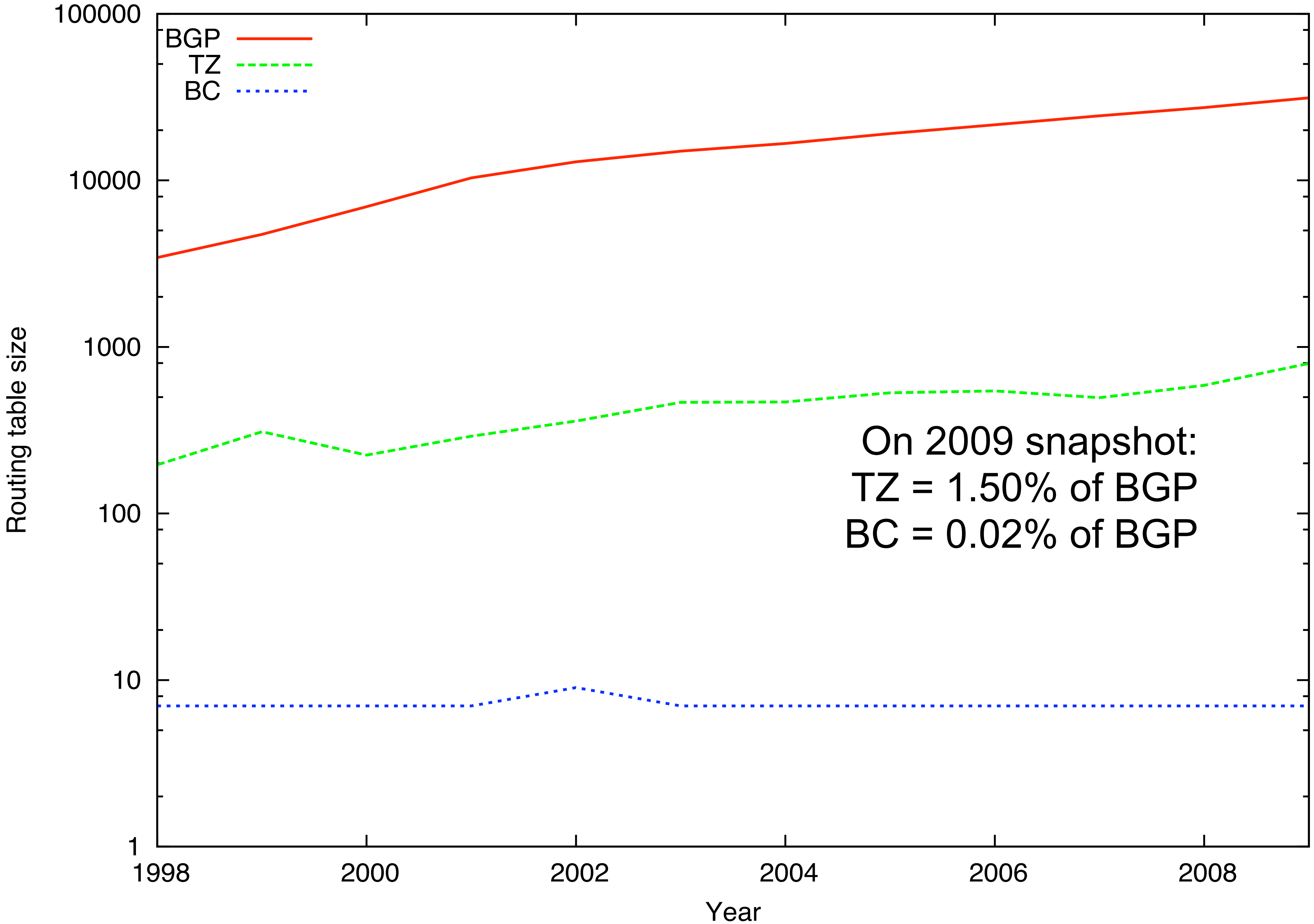
Evolution of Routing Table Size

[Mooney]



Evolution of Routing Table Size

[Mooney]



Discussion

- Evaluation of compact routing on snapshots of the Internet AS graph shows excellent scaling
    - Results for average path stretch and path stretch distribution for TZ on synthetic power-law graphs are confirmed for the Internet AS graph
    - Routing tables sizes are extremely compact, and grow slowly
- But, neither algorithm is developed into a realistic protocol

Future Directions

- Complexity of BC algorithm seems unjustified
  - Spanning tree and labelling algorithms computationally expensive
  - Compared to TZ algorithm, reduction in routing table size not significant
- Can the TZ algorithm be developed into a robust protocol?
  - Topology awareness in choice of landmarks
  - Support for dynamic networks
  - Support for policy routing

Improving TZ: Topology Aware Landmarks

[Strowes]

- TZ landmark selection algorithm is naïve
  - Random initial landmark set, iterated to balance neighbourhood sizes, can lead to poorly placed, ill-connected, nodes becoming landmarks
- New landmark selection: *k*-shell decomposition
  - Decompose using the *k*-shells algorithm

    Recursively remove degree 1 nodes → 1-shell; then degree 2 nodes → 2-shell; until all nodes assigned; highest degree shell is the nucleus

    [Carmi, PNAS, 104(27)]
  - The nucleus comprises well-connected core networks
    - A few dozen nodes, relatively stable over time
    - Reasonable correlation with “tier-1” ASes and other core networks
  - Initial results indicate that nodes in *k*-shells nucleus are compatible with TZ landmark selection constraints
  - Experiments ongoing:
    - Don’t expect significant change in stretch distribution
    - Do expect landmarks to be *more robust* and *better connected*

Conclusions

- Compact routing algorithms show promise for a clean-slate Internet routing architecture
    - First comprehensive evaluation of these algorithms on snapshots of the Internet AS-graph topology
- Much work remains to be done to develop the algorithms into robust protocols
    - *k*-shells decomposition promising for topologically meaningful landmarks, leading to more robust routing
    - Longer term challenges to handle dynamic networks and routing policy