Jan-23 Lecture

Today's Goals

- Say something about "research" papers
  - Since we'll be covering quite a few of them
- Review methods and conclusions of one of the *seminal papers* to study routing in the Internet
  - Ties to our look at Internet diagnostic tools

Research Papers

- Good source of material that represent the state-of-the-art on what researchers are working
- Even the "old" papers provide some useful insight, background, and context
- Types: workshops, conferences, and journals
  - Workshops are more work-in-progress
  - Journals are more for an archive of finished research
- All follow a similar process for reviewing papers
  - Papers are submitted
  - A committee (or editor(s)) evaluate the paper
  - A set of papers are accepted and published

Research Papers

- Workshops and conferences have a set submission date
- Papers are orally presented at the event

- Journals either don't have deadlines or have deadlines for specific, "special" issues focused on a specific topic

- Acceptance rates
  - at the top conferences are anywhere from 5%-20%
  - good conferences range up to about 30%

- The quality of a workshop/conference/journal is measured by its acceptance rate the organizing committee

- Flagship conferences for networking
  - http://www.sigcomm.org/events/sigcomm-conference (39 of 231)
  - http://www.ieee-infocom.org/ (292 out of 1395)

Routing Studies of the Internet

- Paxson, V., "End-to-End Routing Behavior in the Internet," IEEE/ACM Transactions on Network, vol. 5, no. 5, October 1997.
- Schwartz, Y., Shavitt, Y., Weinsberg, U., “On the Diversity, Stability and Symmetry of End-to-End Internet Routes,” IEEE Global Internet, San Diego, CA, March 2010.

What to Focus On

- What is the objective of the study?
- What is the methodology?
- What was learned?
- What are the weaknesses of the work?
- What ideas for improving the work are there?
  - Can target questions be better answered?
  - Are there other questions raised?

Paxson: E2E Routing in the Internet

- Basic idea: use traceroute and see what routes are, how they change, when they change, and is there symmetry
  - Turns out, using traceroutes was and is one of the best ways to discover the stability of routes
- What exactly can traceroutes tell us?
- What other information might someone want to know?

The Study

- 37 Internet sites
- 40,000 measurements
- Data collected over two periods
  - Nov 8 – Dec 24, 1994
  - Nov 3 – Dec 21, 1995

The Metrics

- *Pathologies*

- *Prevalence*: “Given that we observed route *r* at the present, how likely are we to observe *r* again in the future?

- *Persistence*: “Given that we observed route *r* at time *t*, how long before that route is likely to have changed?”

- [Q]: How can this information be determined by just taking periodic measurements?

Prevalence and Persistence

- [Q]: How can this information be determined by just taking periodic measurements?
  - The problem is, couldn't a route have been stable, changed, then changed back again by the next traceroute. How would you know it changed and changed back?

Prevalence and Persistence

- [Q]: How can this information be determined by just taking periodic measurements?
  – The problem is, couldn't a route have been stable, changed, then changed back again by the next traceroute. How would you know it changed and changed back?
- [A]: One of the more interesting ways in which they did the study was not to take samples at random intervals
  – For example: the route from A -> B could have been taken one day, one hour, or one second apart
  – By analyzing the changes when the interval was small, results could be extrapolated to larger and larger periods of time
  – Not perfect, but it provides good insight, even when the data set isn't perfect

Major Conclusions

- Likelihood of encountering a pathology increased from 1994 to 1995 (from 1.5% to 3.3%).
- Internet paths are heavily dominated by a single prevalent route, but that the time periods over which routes persist show wide variations.
- At the end of 1995, A->B and B->A paths differed by at least one AS approximately 30% of the time.
- Different sites (or pairs of sites) experienced very different routing characteristics

Accounting of Routing Pathologies

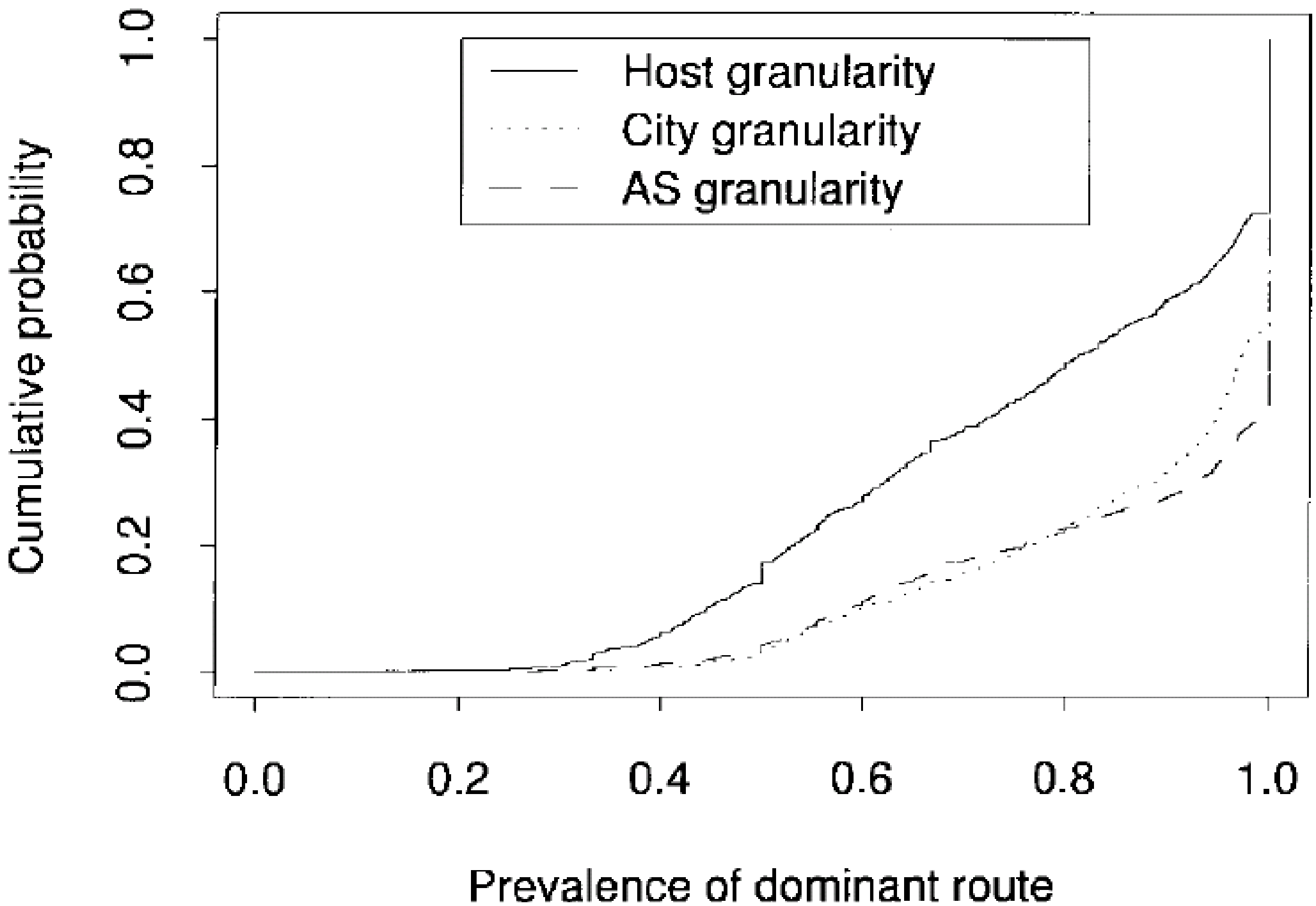
- Loops (forwarding, information, traceroute)
  - 1/10 of 1%
  - Generally on the order of hours (suprising)
  - Loops more common in high connectivity areas
- Erroneous routing (packets took the wrong path)
  - Infrequent
- Connectivity altered (clear non-contiguous path)
  - .16% and .44%
- Fluttering (now called flapping—oscillations in a path)
  - Minimal

Accounting of Routing Pathologies

- Infrastructure failures (unreachables)
  - .21% in D1, slightly higher in D2
- Too Many Hops (more than 30 hops)
  - Minimal (none in D1 and 6 in D2)
  - Average hop count was 15.6 and 16.2 respectively
  - Noted exceptions to hop count being correlated with physical dist
- Temporary Outages (series of lost traceroute packets)
  - 0 losses: about half
  - 1-5 losses (likely due to congestion): about half
  - >5 losses: about 1%
- Time of Day Patterns
  - Routing affected by network use (load leading to congestion)
  - Some effects maybe from router maintenance

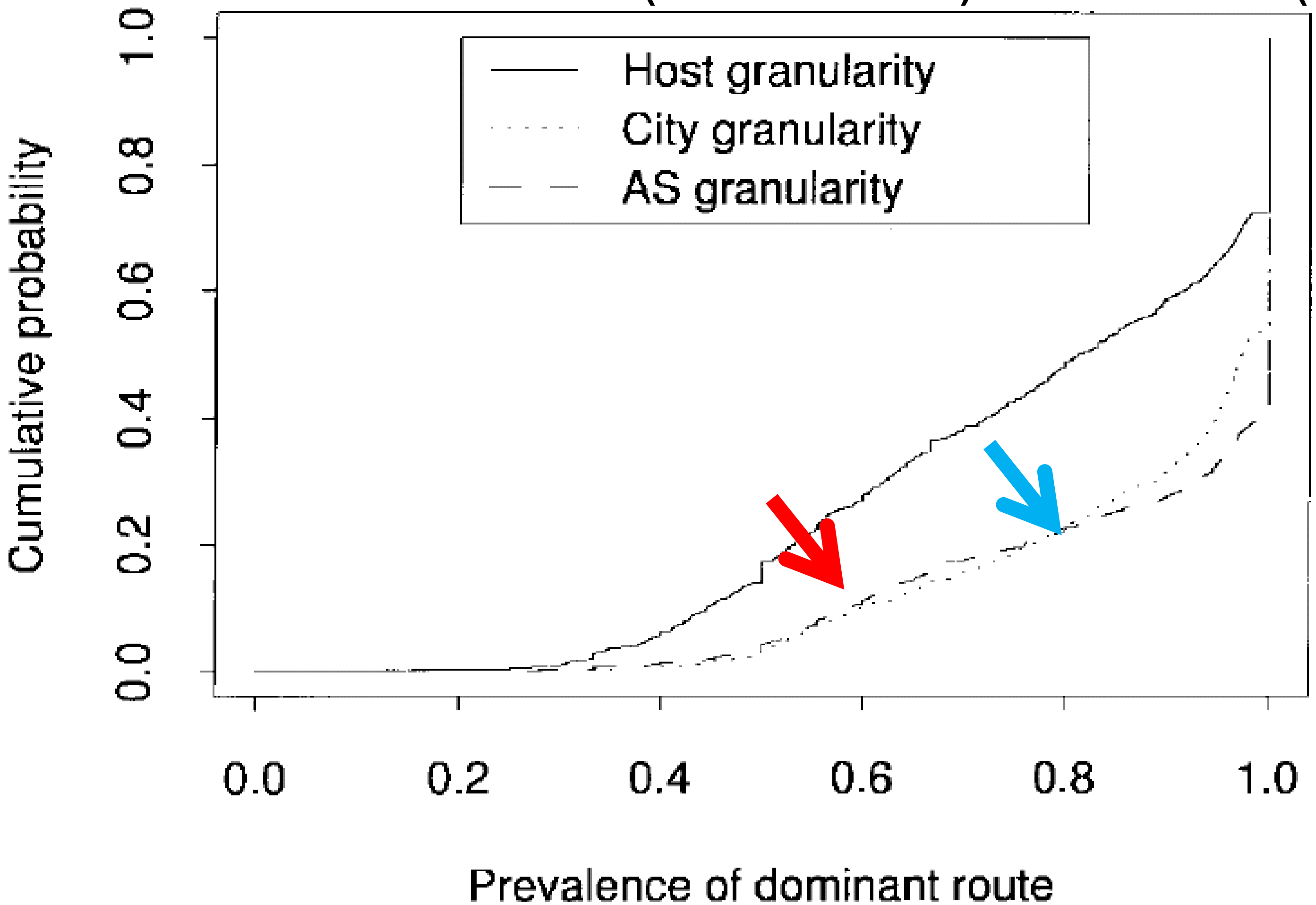
Prevalence Graph

- Graph is a bit challenging to understand



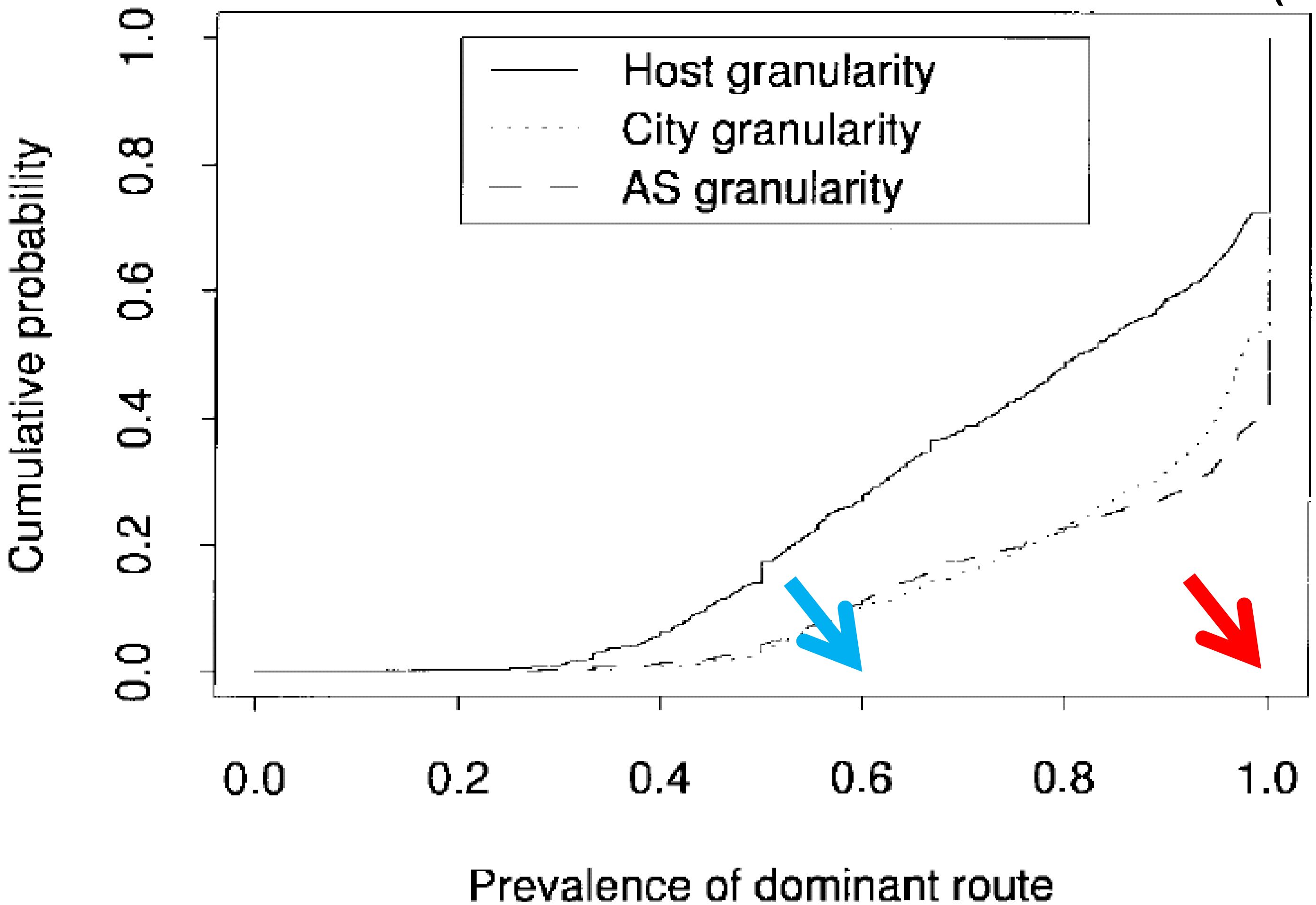
Examples from the Prevalence Graph

- For 90% of the Internet paths, the most common series of cities occurred at least 60% (60%-100%) of the time (red)
- For 82% of the Internet paths,, the most common series of cites occurred at least 80% (80%-100%) of the time (blue)



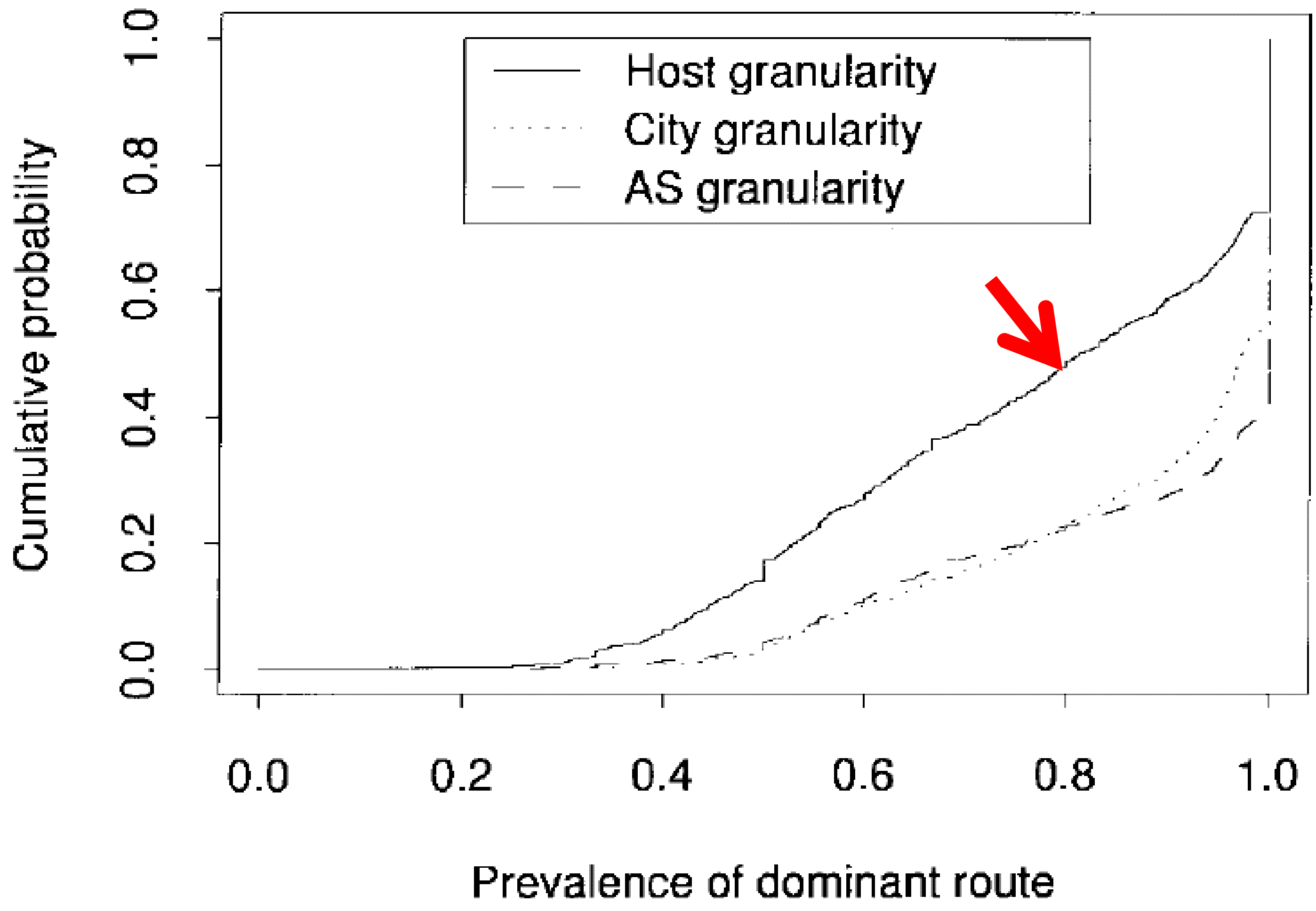
Examples from the Prevalence Graph

- Extreme example: for 100% of the Internet paths, the most common series of cities occurred 100% of the time (red)
- Extreme example: for 60% of the Internet paths, the most common series of cities occurred 100% of the time (red)



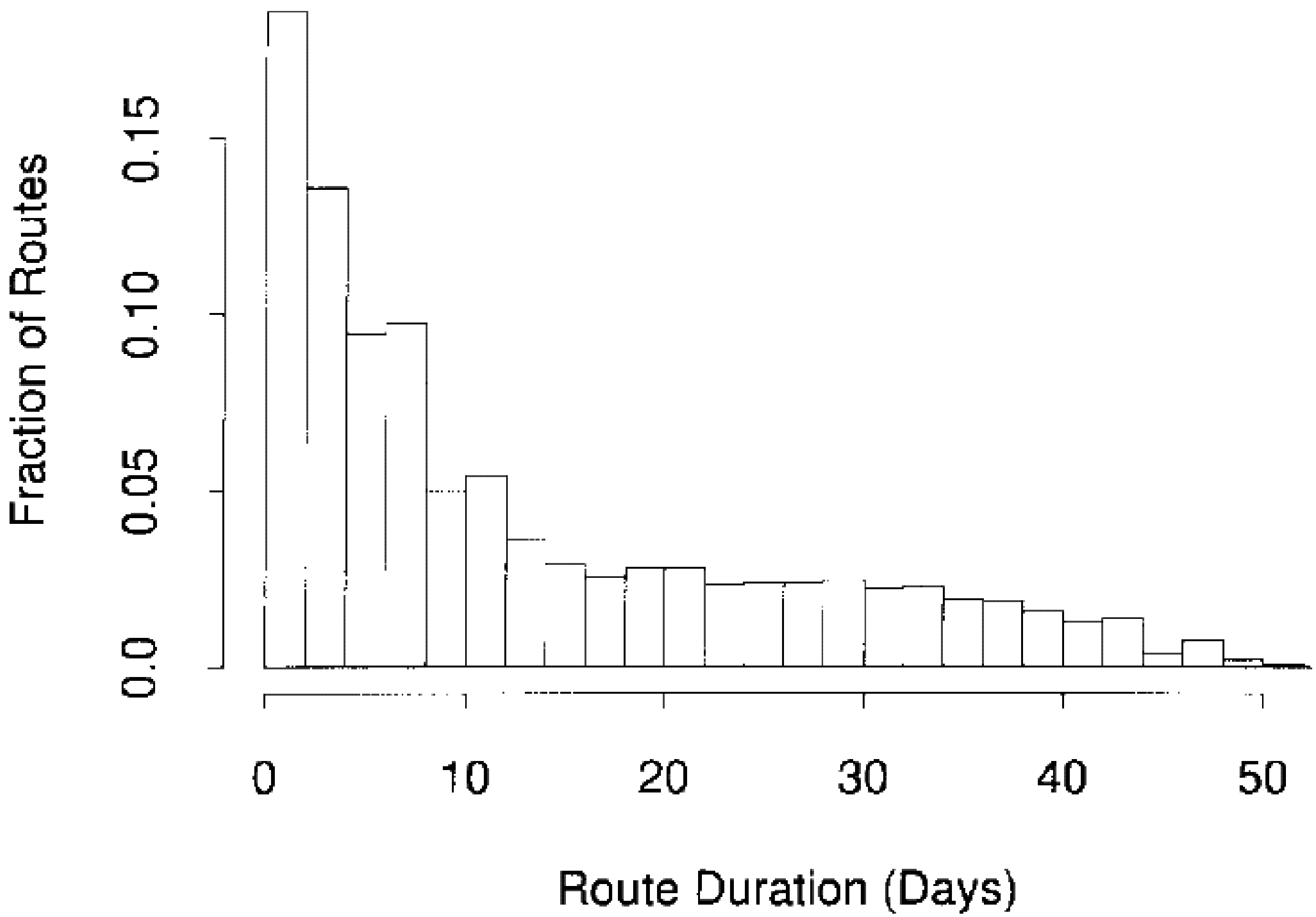
Examples from the Prevalence Graph

- For 50% of the Internet paths, the most common series of hosts occurred at least 80% (80%-100%) of the time



Persistence Graph

- Easier to understand



Persistence Table

SUMMARY OF PERSISTENCE AT DIFFERENT TIME SCALES

Time scale	%	Notes
seconds	N/A	"Flutter" for purposes of load balancing. Treated separately, as a pathology, and not included in the analysis of persistence.
minutes	N/A	"Tightly-coupled routers." We identified five instances, which we merged into single routers for the remainder of the analysis.
10's of minutes	9%	Frequent route changes inside the network. In some cases involved routing through different cities or AS's.
hours	4%	Usually intra-network changes.
6+ hours	19%	Also intra-network changes.
days	68%	Two regions. 50% of routes persist for under 7 days. The remaining 50% account for 90% of the total route lifetimes.

Routing Asymmetries

- 49% of the measurements observed an asymmetric path that visited at least one *different city*

- 30% of the measurements observed an asymmetric path that visited at least one *different AS*

- Majority of asymmetries are confined to a single city or AS difference

- 1/3 of city asymmetries differed by 2 or more hops
  - One example: UCL in London to Univ of Mannheim in Germany
    - One direction was through Amsterdam and Heidelberg and the other was through Princeton and College Park

Next Paper: Schwartz, Shavitt, Weinsberg

- Looked at many of the same basic questions
- Used a similar technique, but used 100 different locations (called vantage points)
- Two 96-hour measurement periods: late ’06 and early ‘09
  - Much shorter time period
  - 100K traceroutes per each measurement

Major Conclusions

- Routes are diverse but relatively stable with high variance between sites
- Stability properties are consistent for both directions between two sites
- At the country level (higher level than previous study), routes are very stable
- Diversity and stability are consistent over time (between 2006 and 2009)

(See also their characterization of other studies)

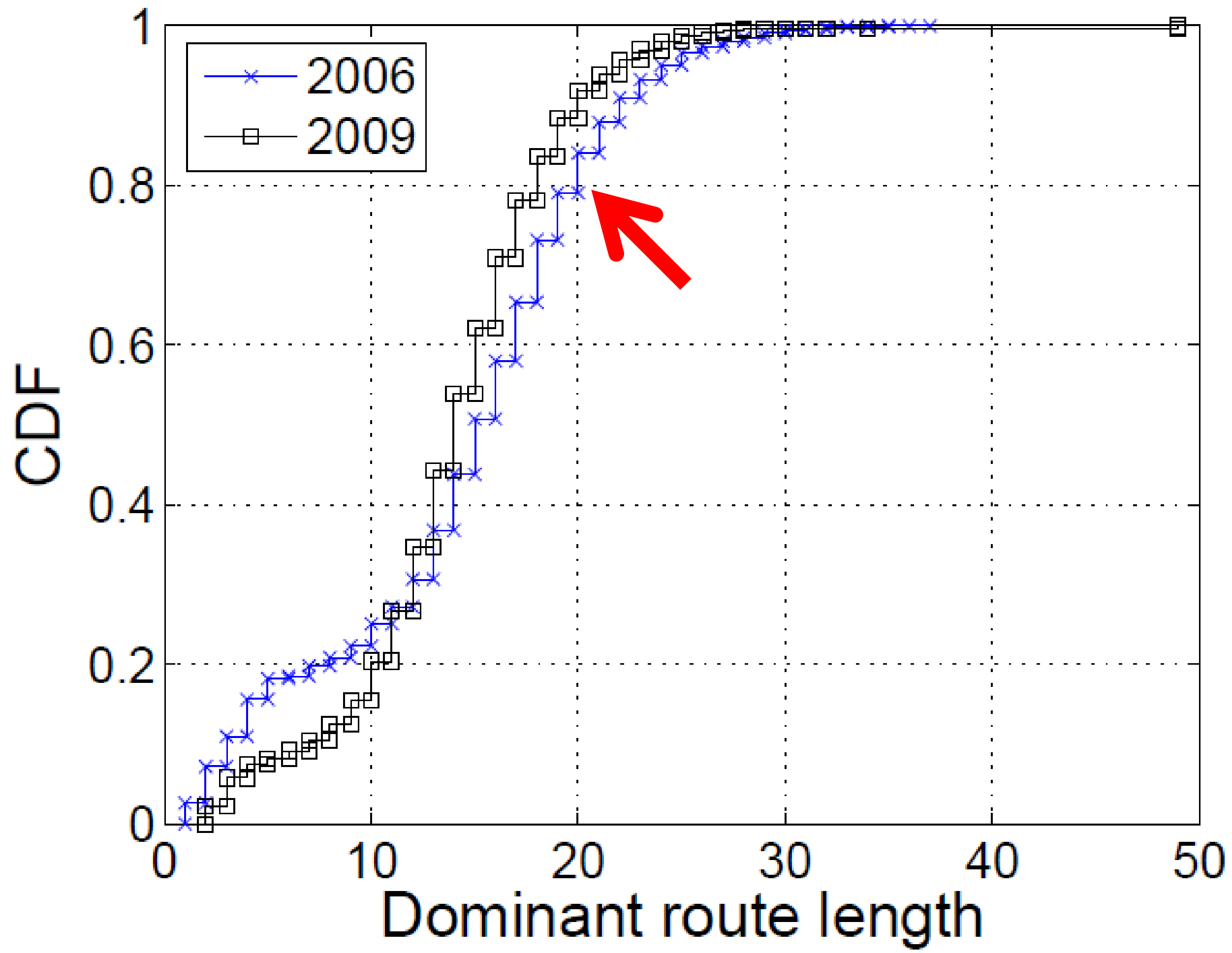
Some Differences in Methodology

- Newer study (SSW) used fixed intervals between measurements where Paxson used exponentially distributed times

- Newer study (SSW) used ICMP and UDP traceroutes
  - The key part is still the timeout of the TTL
  - But different protocols and ports mean load balancing might kick in
  - Routers avoid flapping by considering other information when deciding to route out redundant interfaces (protocol, addresses, ports, etc.)

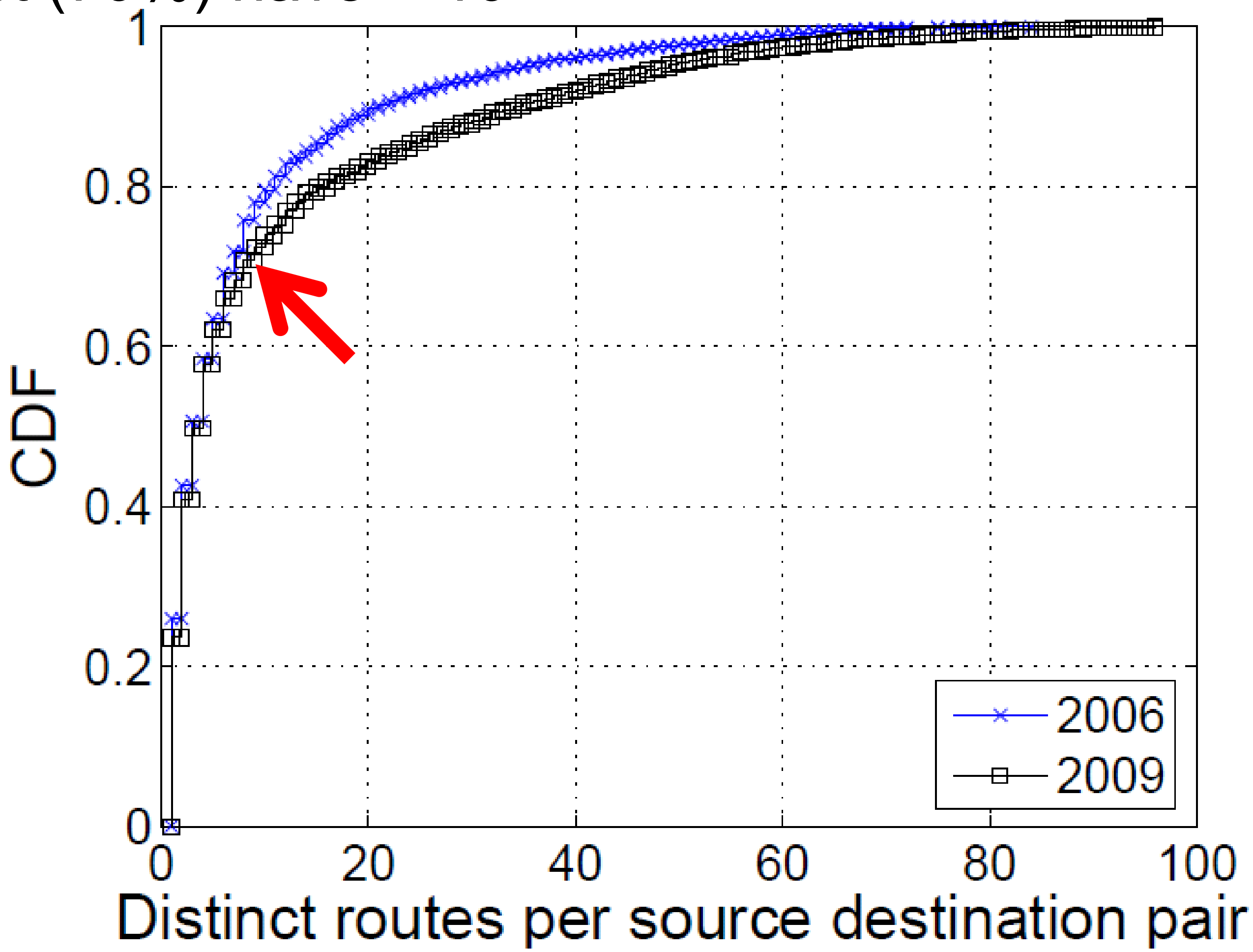
Dominant Route Length

- Another CDF: 80% of routes are 20 hops or less; the other 20% have more than 20 hops or more



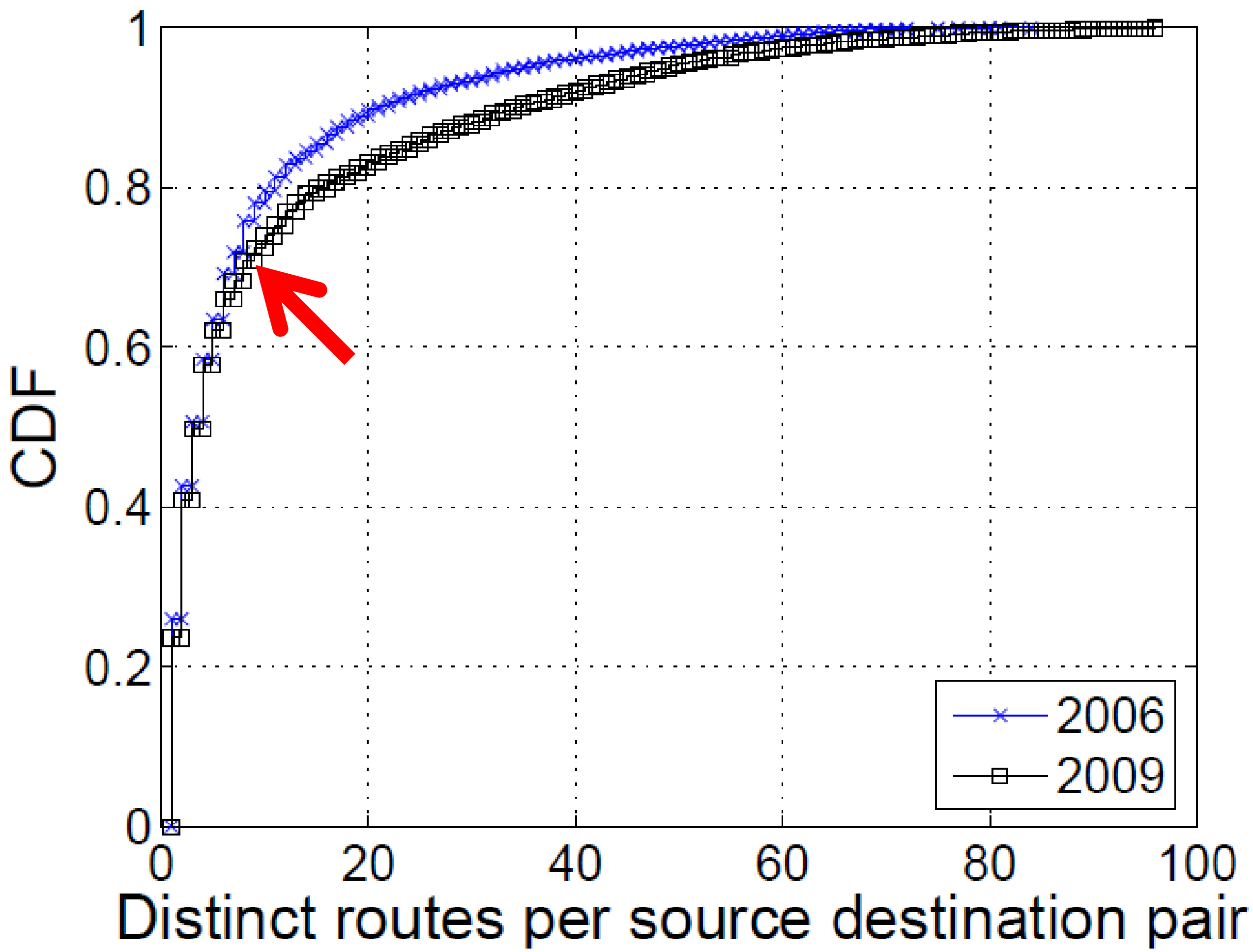
Route Diversity

- Another CDF: approximately 30% have *more than 10 routes* between the source-destination pair
- The rest (70%) have < 10



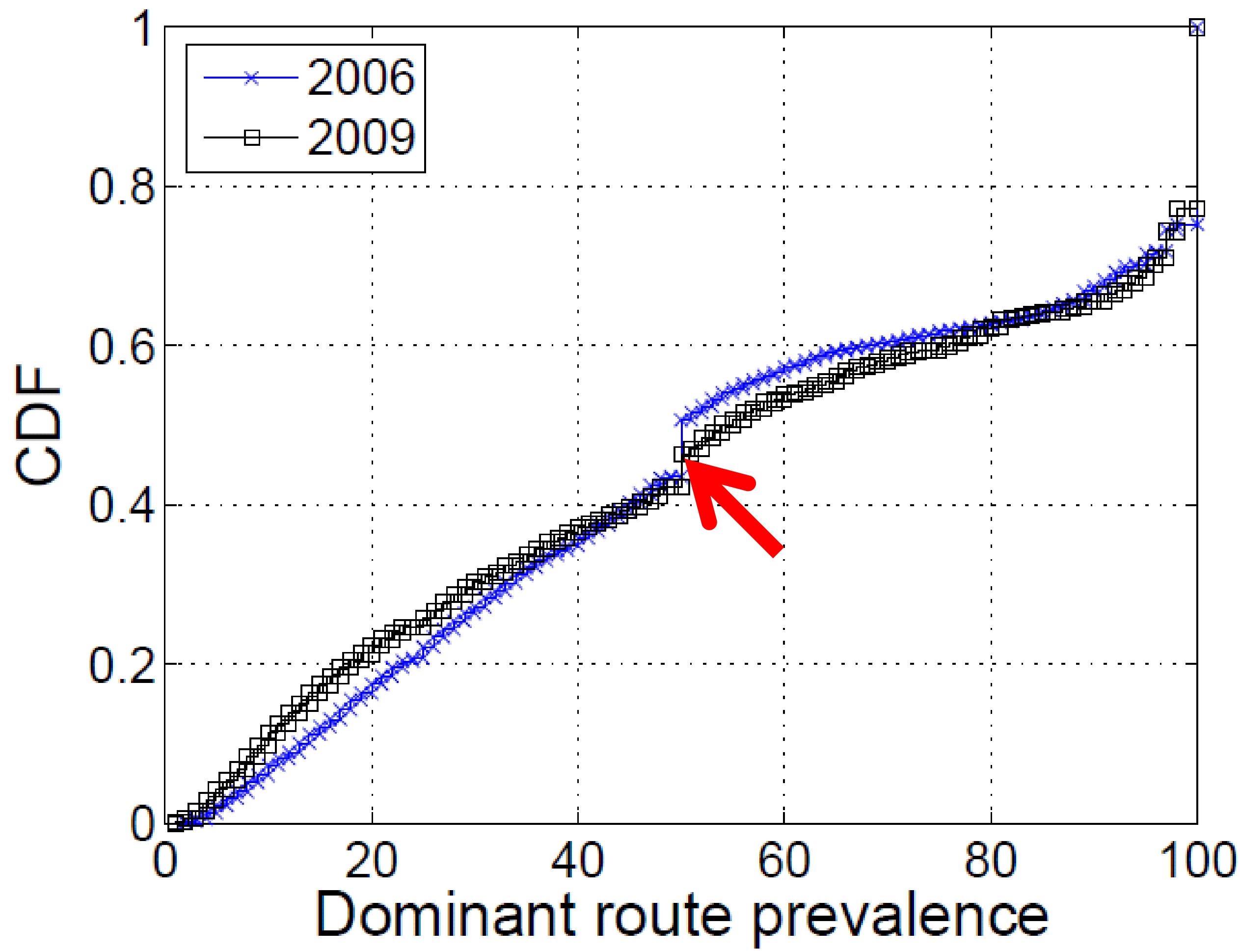
Route Diversity

- More route diversity in 2009 than in 2006



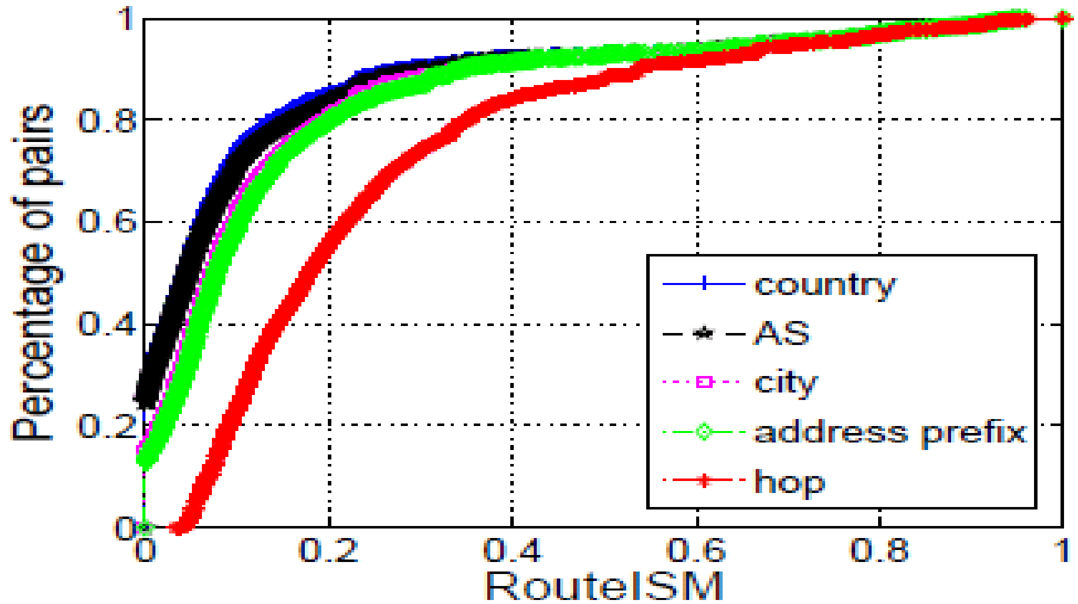
Route Prevalence

- Another CDF: for 50% of the Internet paths, the most common series of routers occurred 50% of the time

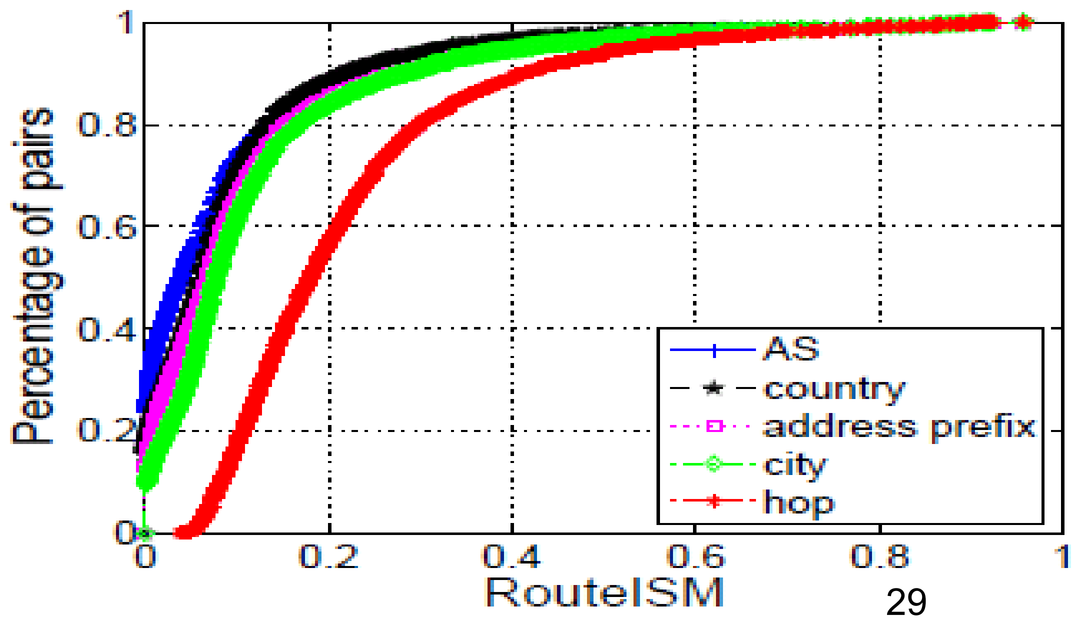


Higher Level Granularity

- As the granularity goes up (from host level to city, AS, or country, stability increases



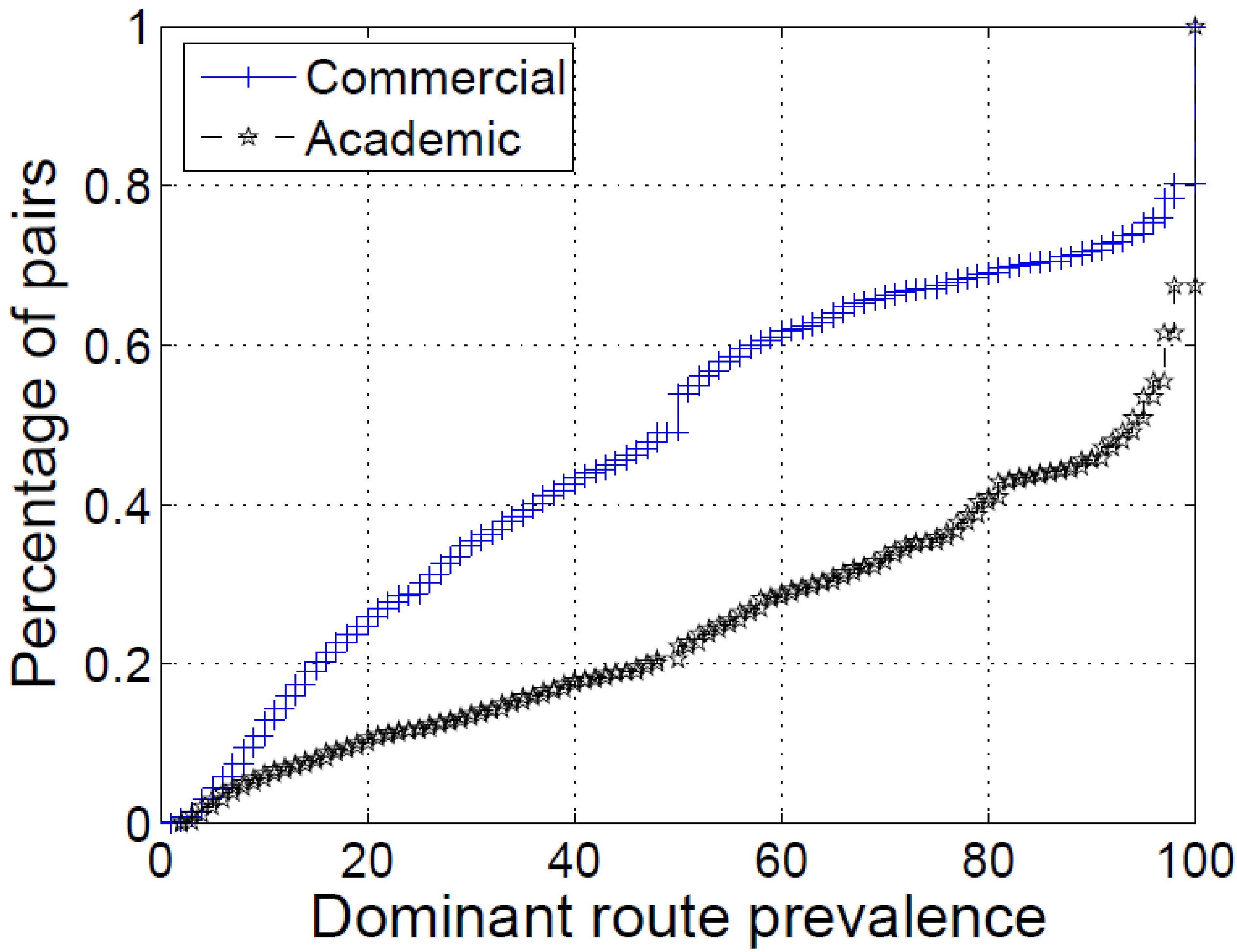
(a) 2006



(b) 2009

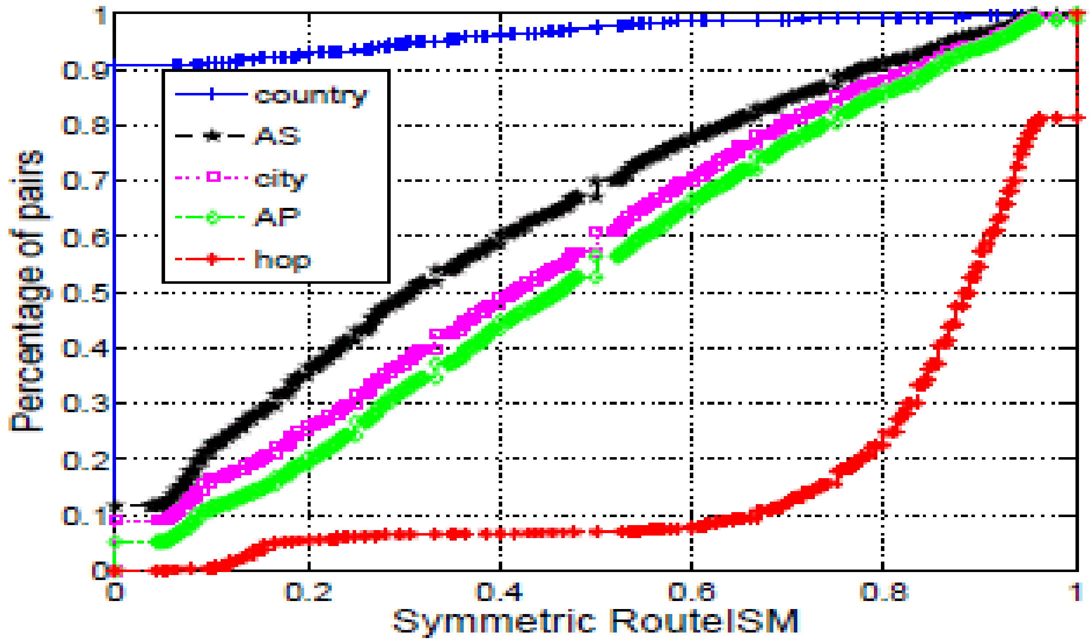
Academic v. Commercial Networks

- Based on this graph, which has the more prevalent routes?



Symmetry

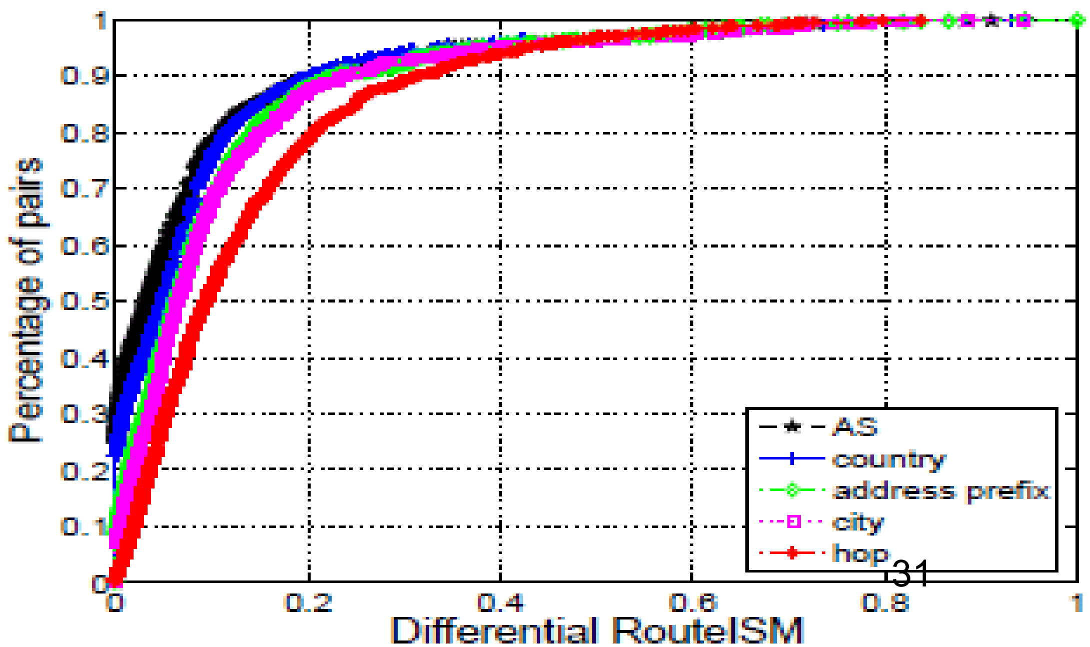
- Host level is not symmetric because routers have different addresses in different directions



(a) RouteISM symmetry

Differential Symmetry

- Routes in the two directions are either both stable or both unstable



What to Focus On

- What is the objective of the study?
- What is the methodology?
- What was learned?
- What are the weaknesses of the work?
- What ideas for improving the work are there?
  - Can target questions be better answered?
  - Are there other questions raised?