

## analysis of routing & topology data

sept/oct 2001

no aphorism is more frequently repeated...  
than that we must ask Nature few questions,  
or ideally, one question at a time.  
...this view is wholly mistaken.

Nature will best respond to a logically  
and artfully thought out questionnaire;  
indeed if we ask her a single question, she will often refuse  
to answer until some other topic has been discussed.  
-- perspectives in medicine and biology 1973 sir ronald a fisher

[broido@caida.org](mailto:broido@caida.org)  
[kc@caida.org](mailto:kc@caida.org)  
[ucsd/sdsc/caida](http://ucsd/sdsc/caida)

# outline

- motivation: macroscopic topology study
  - infrastructure-relevant research questions
  - led us to analyze BGP data on its own
  - realized that routing & topology inextricably related
- describe both kinds of data sets
- introduce techniques for analysis of Internet 'core'
  - ▶ graph theoretic model
  - ▶ combinatorial core vs giant component
  - ▶ dual graph
  - ▶ subprefix connected components
- introduce metrics for analysis
  - ▶ indegree/outdegree characteristics
  - ▶ transit versus origin role
- new contributions to the field of topology analysis
  - new granularities: BGP atoms, dual AS graph, ramified atoms
  - size distributions (weibull) with explanatory model (coalescence)
  - reachability functions
  - topological resilience

# motivating questions (a sample)

- number of network prefixes in table
  - reduction beneficial to [routing system] infrastructure
- trends in routability of IP space given by registries
- which ASes are most highly:
  - connected, vulnerable, controlling connectivity
- is there an Internet 'core'
- does the core grow due to multihoming?
  - are more-specifics what is driving up the table size?
- hop diameter of AS, AS diameter of IP paths
- optimizing active measurement architectures

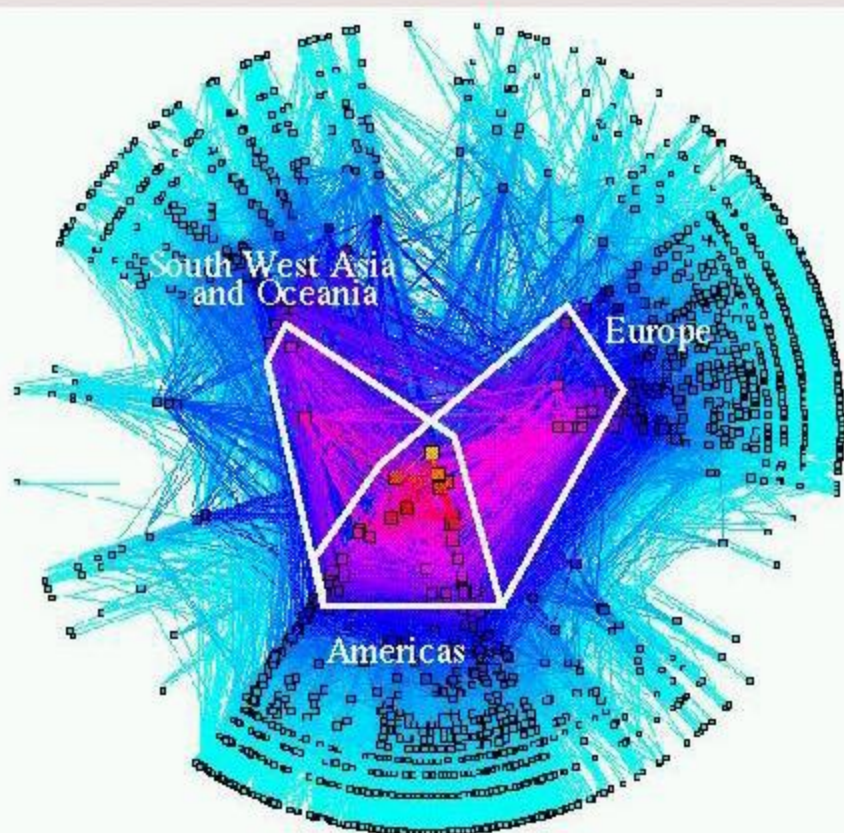
can topology be inferred from BGP tables  
for the purposes of modeling, simulation,  
and infrastructure analysis?

# motivating questions (more specific)

---

- how many prefixes does it take to cover the IPv4 address space?
- how many routing entries would be required if arbitrary intervals (as opposed to CIDR blocks) in IP space would be allowed as ranges of network addresses?
- what is the complexity of the system of AS paths associated with individual prefix?
- how many different types of networks are globally distinguishable with respect to routing policies?
- how many routing policies are applied by the Internet to addresses originated by one AS?

# global AS topology (from skitter data)



## topology data: skitter

---

- about 26 sources around world
- 500k destinations (diff by box)
- over 60% of prefixes covered (still building lists)
- parallel ICMP echo request reply
- lightweight, coarse temporal granularity
- used for variety of macroscopic studies
- most comprehensive in world (low bar)
- data available to researchers

[www.caida.org/tools/measurement/skitter/](http://www.caida.org/tools/measurement/skitter/)

## topology data: dynamic properties

---

- all graphs change with time
- new equipment (nodes, links, firewalls) added
- new IP blocks allocated
- renumbering out of old IP blocks
- "death rate" of IP addresses
- paths oscillate (load balancing)
- paths fluctuate (routing instability)
- paths flip (manual config)
- outages, cuts, blackouts

okay so data is messy...

## topology data: ambiguities

---

- non-responding hops (no IP returned)
- rate-limited response
- private addresses
- multicast addresses
- addresses in 0.-2./8 blocks
- no matching BGP prefix
- prefixes with multiple origin AS

okay so data is icky, too...

## private address use

---

- Internet is partitioned by  $> 2$  layers of NAT
- private addresses = 1.7% of skitter replies
  - (18K out of 1.07M, Jul.02-31,2001)
- assumption: private IPs are misconfigs
- we assume a generic misconfig rate of 1%
  - inadvertent transit ASes: 1.25%
  - multi-origin prefixes: 1.3%
- (assume 100x more private addrs than we observe)
- => private/unique IPs  $\sim 1.7$  (order of magnitude)

## correlation among data sets (perf., topology, routing)

---

### preliminary findings

- ~1% IP destinations disappearing monthly
  - (re-addressing, firewalls)
- route announced path not matching forward path
- indication of potential routing configuration errors
  - by no means automatic
- paths much less static than expected

## macroscopic question: DNS root server placement

---

- RSSAC, DNS technical advisory committee to ICANN
- goal: optimize root nameserver locations (13)
  - co-locate skitter hosts w root servers
  - demonstrate root server performance in serving target community
  - develop techniques for evaluating architectural optimality for root server placement
  - visualization to correlate data sources/types
- collaborative project to encourage proactive participation (network operators, researchers, others)

([www.caida.org/tools/measurement/skitter/](http://www.caida.org/tools/measurement/skitter/))

## skitter on-going daily summaries

---

<http://www.caida.org/tools/measurement/skitter/summary>

- path length (in IP hops) distribution
- RTT distribution
- RTT versus longitude,
- path dispersion
  - AS & country granularity
- AS core graph

## inter-domain routing (BGP) data

---

- not much instrumentation/diagnostics/data sources
  - UO's [www.route-views](http://www.route-views.org)
  - RIPE's [www.ripe.net/ris](http://www.ripe.net/ris)
  - Merit's IPMA /[www.merit.edu/ipma](http://www.merit.edu/ipma)
- really need real-time instrumentation
  - (w/o impacting forwarding performance)
  - ideal: route-lookups in real-time w/o kernel
- need data for realistic inter-domain routing models
  - identification/vis of flaps, outages, critical pts
  - correlation of perf. w/some measure of path 'length'
  - comparison of forward path with
    - ▶ BGP path
    - ▶ shortest path
    - ▶ reverse path

## other uses for inter-domain BGP data

- mapping topology data
  - aggregating IP to network prefixes
  - aggregating prefixes to origin AS
- inferring contractual relations
- "bird's eye view" of the net – (AScore)
- predicting AS path taken by a packet

important question: can you get essentially the same information from either dataset? (hint: no)

indeed, even though often covering fewer ASes than a full BGP table, skitter data shows bidirectional and transit connectivity for a significantly more ASes than BGP data of the best available quality and sampling. (totally non-obvious...)

## RouteViews BGP data format

---

192.172.226.0/24 134.24.127.30 64 1740 195 1909 i  
CAIDA network peer (cerf) med cerf sdsc caida

192.172.226.0/24 204.147.128.141 - 145 195 1909 i  
CAIDA network peer (vbns) med vbns sdsc caida

AS path: 1740 195 1909

Grows from tail to head – "prepending":

CAIDA advertizes prefix to SDSC

SDSC advertizes CAIDA's prefix to CERF

ergo,

CERF "knows" how to send traffic to CAIDA

## **new notion: policy-constrained dual AS graph**

- nodes: peering sessions
- links: pairs of sessions w/common AS and followed by traffic in skitter or BGP data
- derived from observed AS triples
- example:
  - 1909 195 1740 - AS path (skitter,BGP)
  - CAIDA SDSC CERF
  - nodes: 1909-195 and 195-1740
  - ordered AS pairs
  - directed link: from 1909-195 to 195-1740
- allows for capturing more policy

## dual AS graph features

---

- AS hop count = #sessions in a path
- easy to compare with AS hops
- If necessary, add: 0-AS & AS-0 for traffic origination/termination
- adds rigidity to the graph
- constrain path by continuity and direction
- like differential equation, but more choice at each step
- path continuation non-unique

general case:

nodes: AS  $n$ -tuples, links:  $n+1$ -tuples

## RouteViews data characteristics, 25 nov 2000

---

- 2.627M lines (prefix, peer, AS path)
- 34 contributing peers
- 28 peers with 81K–95K prefixes
- 2.534M lines in 28 tables

Of those:

- 2,243,524 (88.5%) have no repeated AS
- 290,392 (11.5%) lines with repeated AS
- 176,576 (7%) end with repeated AS
- 2/5 repeated had it only in the middle
- 0.45% suppressed, dampened or history

## taxonomy of RouteViews BGP tables (aug 2001)

■ table 'type'	prefix count	# RV tables
■ full size backbone	103K	37
■ filtered backbone	89-96K	6
■ non-backbone	<50K	8

filtered tables only partly useful

## analysis of BGP tables

---

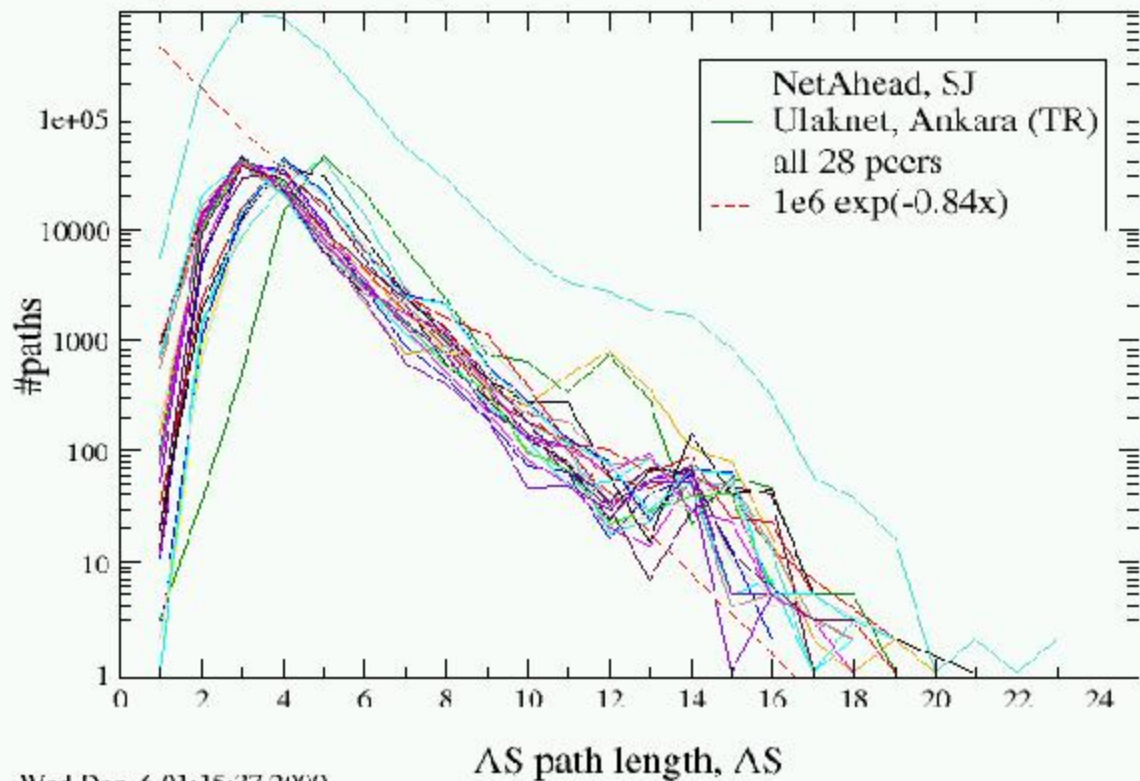
- data is clean, easy to get, parse
- appears a reasonable substitute for traceroute data
- fluctuations
  - NLANR's archives corrupted Dec.2000-Feb.2001
  - random daily fluctuations (~3%)
  - if a single /8 is not announced - 0.4% IP addresses
- misconfigs
  - routing loops (2 loops in 91 pref-peer pairs)
- discrepancies
  - e.g. multi-origin prefixes, 1.3%
- caveats: extremely sensitive to peer choice

# AS path length

- primary knob of BGP hammer (default metric)
- All AS versus unique AS count:
  - prepending – repeating of AS
  - makes AS path look longer
  - reduces traffic via this path
- bumps at AS path lengths 12,14
  - ▶ e.g., 202.183.247.0/24 3561 5400 5400 5727 4651 7568 7568 7568 7568 7568
  - ▶ 5 unique + 5 repeated ASes == AS path length 10
- routes are compared/chosen locally
- other metrics:
  - multiexit discriminator (MED)
  - communities

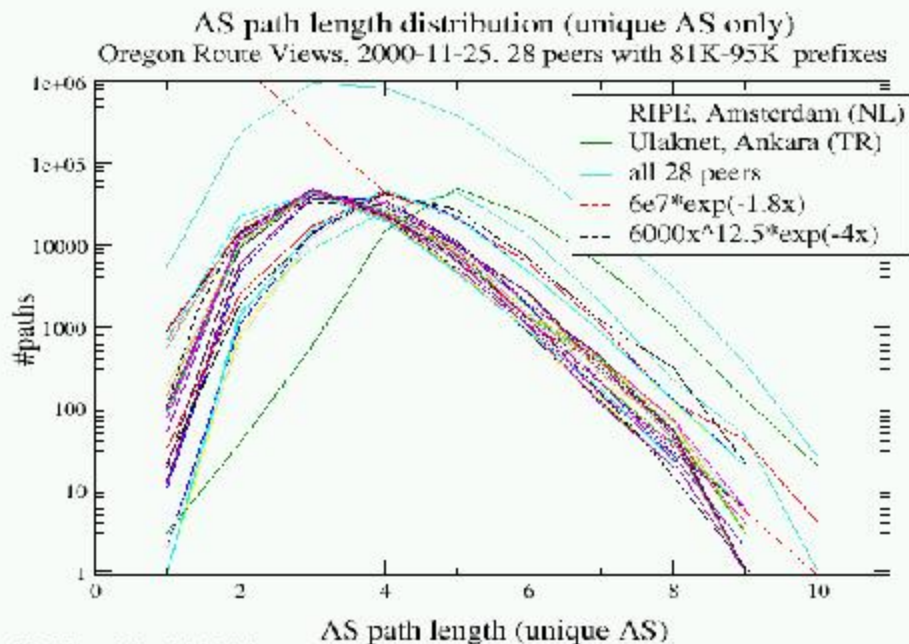
# AS path length distribution

Oregon Route Views, 2000-11-25. 28 peers with 81K-95K prefixes



# AS path length measured in unique AS

- Much smaller tail:  $\exp(-1.8x)$  vs.  $\exp(-0.84x)$
- Much smoother, close to Gamma distrib.
- (reachability function has same shape, we'll see later)
- invariance across levels of aggregation (surprising)



## evolution of average AS Path length

---

- for 2 years 2000–2001
- common Route View peers
- some peer AS increase average length slightly
- some other slightly decreased
- overall close to being invariant
- nov 2000: 3.60 hops
- may 2001: 3.64 hops
- same as noise level (1%)
- no statistically significant change
- reachability function invariant as a whole, 1999–2001
- (exception: RIPE)

## Internet evolution – BGP table view

---

### Internet trends:

- expansion (numeric growth)
- sophistication (structural complexity)
- refinement/fragmentation
- churn (objects grow, split, recombine, die)
  
- multiple causes of changes compose, so net change does not reflect trends

### Internet is a "two nines" system

- 1% noise a rule, not exception
- => trends w/in few % are likely noise/uncertainty

## Building an AS topology graph from BGP data

- directed from backbone downstream to router
- more hierarchical, captures less 'lateral' peering
- Gao e.a. find 90% provider-customer links
- In reality, # peers  $\sim$  #customers (order of magnitude)
  
- Zocalo: medium-sized ISP in Bay Area
  - 8 transit providers
  - 100 downstream networks
  - 12 have their own AS
  - 225 peers
  - => total degree  $\sim$  250
  - in sum of 43 backbone tables: indeg.11, outdeg.63, total 74
  - => 3 times less

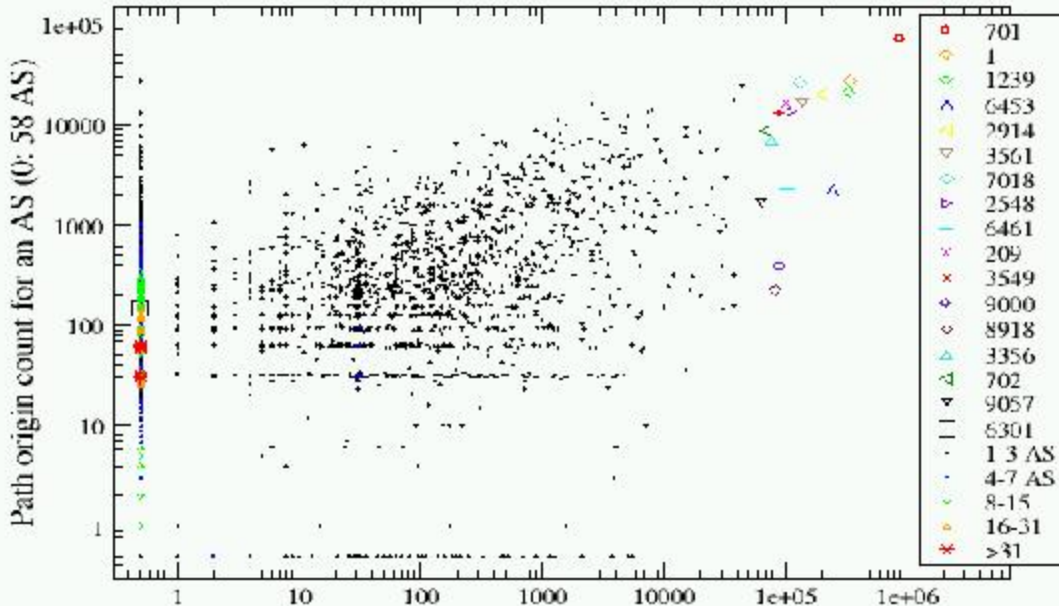
# building a 'core' AS graph from two data types

- deriving topological combinatorial core
  - recursively strip outdeg.0 and 2-loops
  - bidirectional connectivity remains
  - should be 100% if sufficient coverage
  - compare using (1) BGP and (2) forward IP topology data
- (1) BGP routing data AS graph (RouteViews tables)
  - 2.77% in combinatorial core (327 out of 11823)
  - accumulation does not change much
  - adding links from less to more specifics:
    - ▶ adds 4400 links to 25K
    - ▶ increases combin.core from 300 to 900 AS
- (2) probed forward topology AS graph, 1 Aug 2001:
  - 9858 nodes out of 11.5K
  - 2556 (26%) in comb.core
  - 2102 (21%) in g.c. which reaches 9835
  - caveat: traceroute links with third party addresses
    - ▶ multipath noise
    - ▶ (may be 1% cases, exact number unknown)

# transit versus origin AS

AS 701 is in 31% of all paths (and is origin for about 2.5% of prefixes)

AS origin count vs. transit path count  
Oregon BGP Route Views, 2000-11-29. 30 peers over 80K pref.

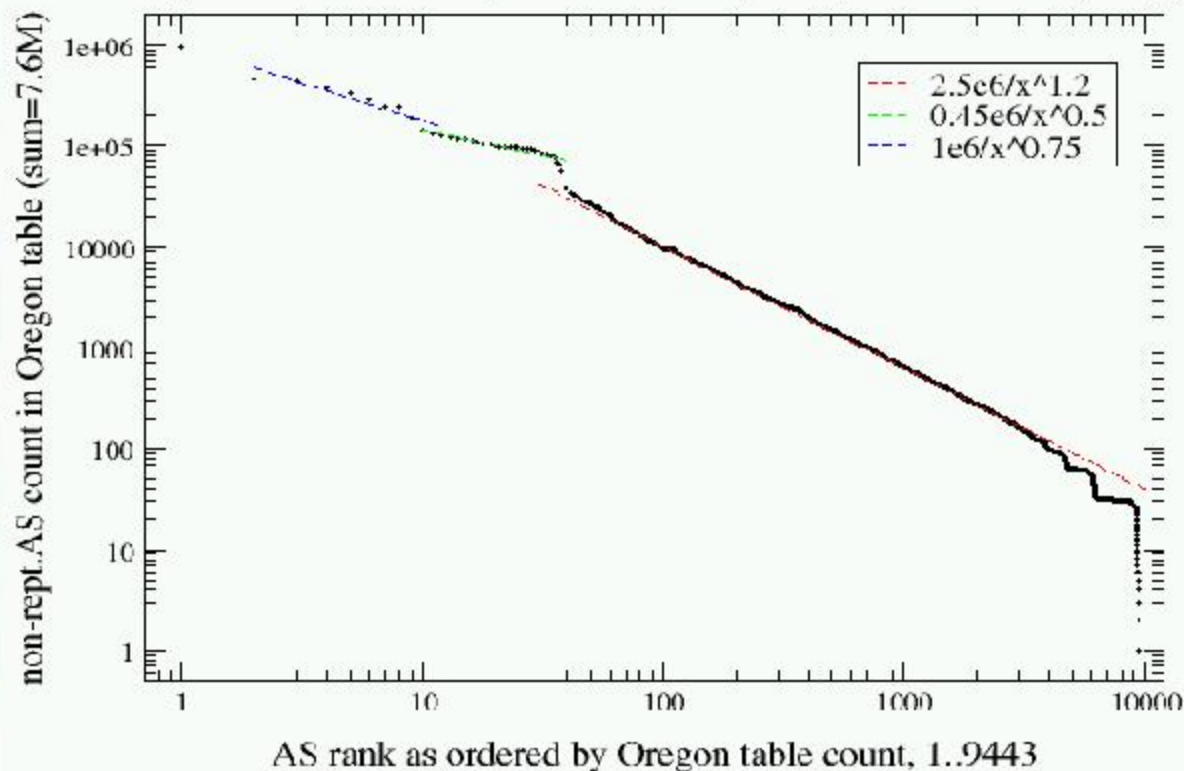


Transit (middle of a path) count for an AS. 0: 7842 AS; >0: 1608. 0's are at 0.5

Tue Jan 9 23:33:56 2001

# AS counts in non-repeating-AS paths

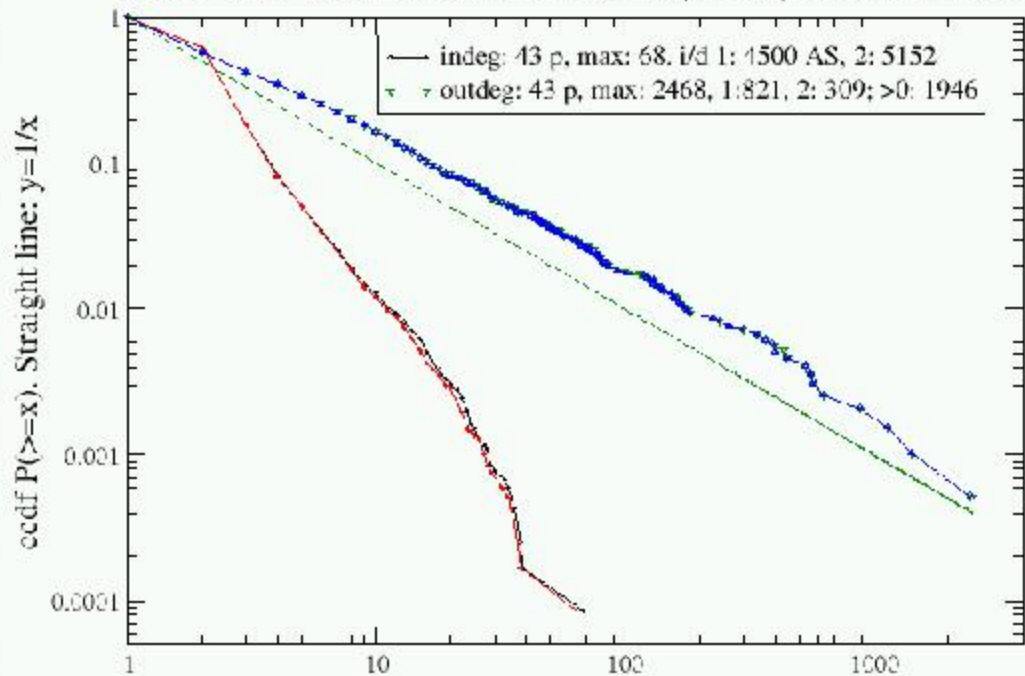
Oregon Route View data, 2000-11-29. 37 peers (25 with over 90,000 pI)



# AS in- and outdegrees in Route Views data

## In- and outdegrees for BGP AS graph

Route View BGP data 2001-09-01, 43 (37 full) backbone tables



## navigating complexity of BGP AS graph

---

- measured by downstream peering richness
- entropy of outbound links distribution (for ~2K transit AS)
- $\sum p \log p$  where  $p \sim$  #prefixes routed via an AS link
- takes into account how often each link is taken
- nov 2000: 4.23 bits
- may 2000: 4.24 bits
- lots of individual ASes changed though
- opposite changes w.net effect 0
- another invariant?

# taxonomy of prefixes

- standalone (no more specifics, no less specifics)
- root (have more specifics, are not more specific)
- more specifics – subsets
  - subsets of other prefixes
  - express fine-grained policies
  - litter routing space?
  - getting free ride?
- taxonomy by prevalence in tables:
  - globally visible: all BGP tables
  - highly visible: in most BGP tables
  - semi-global: in 1/2 of all BGP tables
  - local (a few tables)

highly visible  $\approx$  semi-global

"bathtub curve" phenomenon (fuzzy dichotomy)

## BGP table prefix growth

route-views data does not confirm conjecture that more specifics are currently driving up the BGP table sizes

- indeed, for 1.3 years since aug 2001, standalone prefixes grow faster than more specifics (standalone 38%, more spec 32%)
- currently at parity
- more specifics do churn (appear/disappear) faster
- also convert to standalones
- but overall change in more specifics not statistically significant

increase in # of top prefixes is caused by:

- coverage of new address space 55%
- deaggregation/contraction of existing prefixes 37.5%
- expansion of existing prefixes 5%
- aggregation 2.5%

but growth was in top prefixes, not more specifics

## BGP origination policies

---

- AS indeg  $\geq 3$ , pf indeg  $\geq 3$  26.2%
- AS indeg 1, pf indeg 1 14.8%
- AS indeg 2, pf indeg 2 11.6%
- AS indeg 2, pf indeg 1 6.0%
- AS indeg 1, od.2, pf indeg 1 4.6%

## new unit of connectivity analysis: BGP atoms

---

BGP atoms are:

- equivalence class of prefixes that share same set of AS paths as seen e.g. by peers in Oregon Route Views BGP tables
- unit of granularity in between prefixes and ASes
  - grow proportionally
- new framework to analyze routing system
  - reduce granularity
  - strong potential

## atom sizes

- about 50% of atoms contain just one prefix
- largest atom contains over 2200 prefixes
- size distribution is close to Weibull curve or power fn
- address space size distribution more Kantor-like
- ramified atom
  - several paths to same prefix branch or loop at an AS
  - maps to graph with positive outdegree/cycl.#
  
- reducing the number of atoms
  - **focal point**
    - AS present in all paths of atom, at same relative position
  - **crown point**
    - AS with largest number of following ASes among focal points
  - **crown atom**
    - atom whose system of AS paths has been truncated at crown point

# Crown and ramified atoms in RouteViews

(01 september 2001 tables)

## ■ crown atoms

- take advantage of leaf and AS/leaf prefixes
- chop away non-branching "stalk" from the atom
- reduce to 72% atoms (Sept.01: from 24270 to 17566)

## ■ ramified atoms

- different AS paths from the same-AS peers

0	4578
1	8636
2	2028
3	70

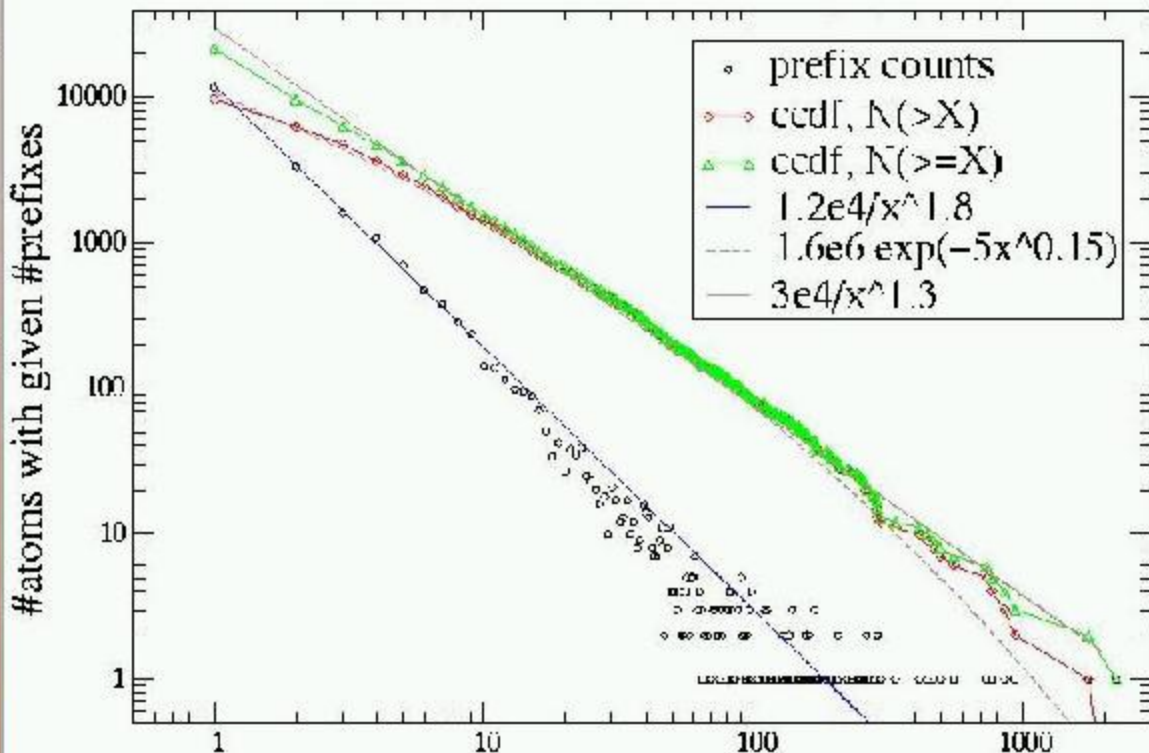
- different path lengths in ramified atom

1	22142
2	5887
3	14
1	1

# atom size in prefixes

## Distribution of prefix counts for BGP atoms

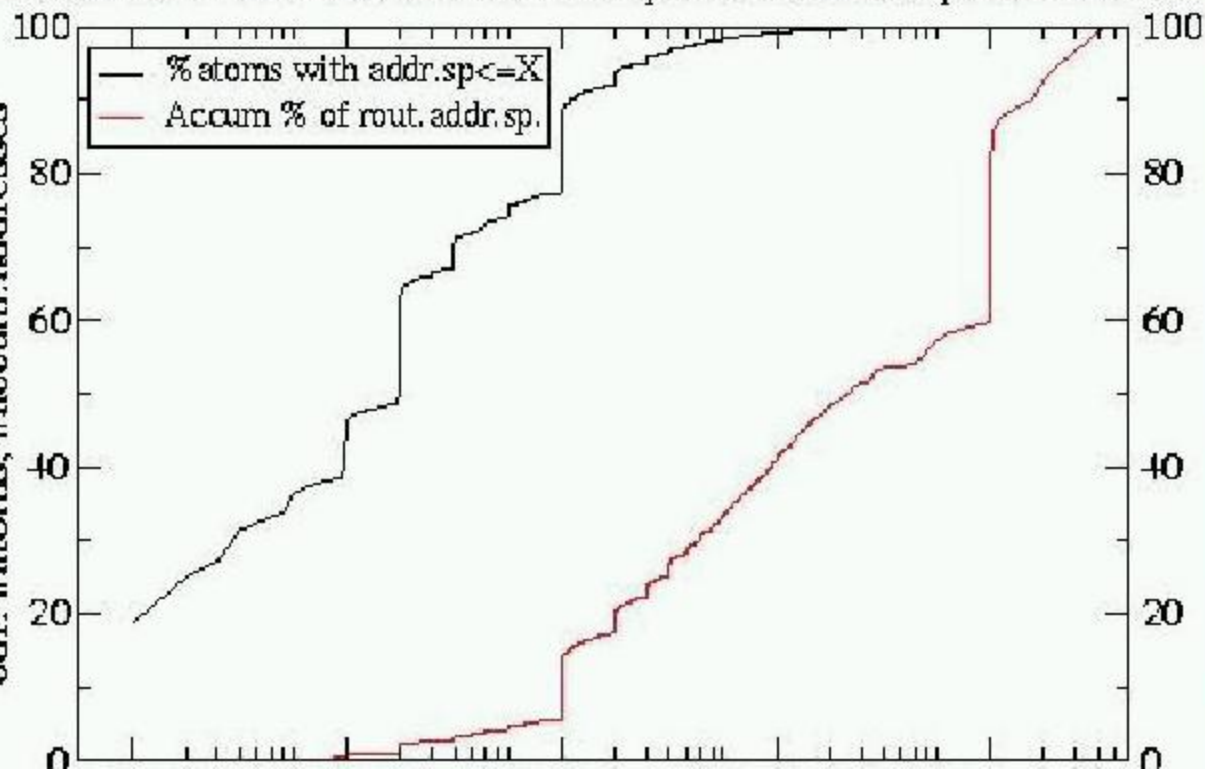
Oregon Route Views BGP data, 2001-05-01, 27 peers, 98K common prefixes



# address space size of atoms

## Address space size of routing policy atoms

RouteView 2000-05-03, 12:00, 33 peers with >85K pref. 19870 at.



atoms	#AS	percent
-------	-----	---------

1	4943	59.84
---	------	-------

2	1647	19.94
---	------	-------

3	750	9.08
---	-----	------

4	347	4.20
---	-----	------

5	191	2.31
---	-----	------

6	111	1.34
---	-----	------

7	67	0.81
---	----	------

8	49	0.59
---	----	------

9	34	0.41
---	----	------

10	18	0.22
----	----	------

11	23	0.28
----	----	------

12	9	0.11
----	---	------

13	10	0.12
----	----	------

14	11	0.13
----	----	------

15	13	0.16
----	----	------

16	9	0.11
----	---	------

17	3	0.04
----	---	------

18	4	0.05
----	---	------

19	5	0.06
----	---	------

21	2	0.02
----	---	------

22	1	0.01
----	---	------

23	2	0.02
----	---	------

## new unit of connectivity analysis: cones

---

- set of nodes that
  - start at a given node
  - go through only non-core nodes reachable from this node.
- nodes that wholly or in part depend on a given node (tip of cone) for connectivity

## new connectivity unit: connected subcomponents

---

- intra-prefix

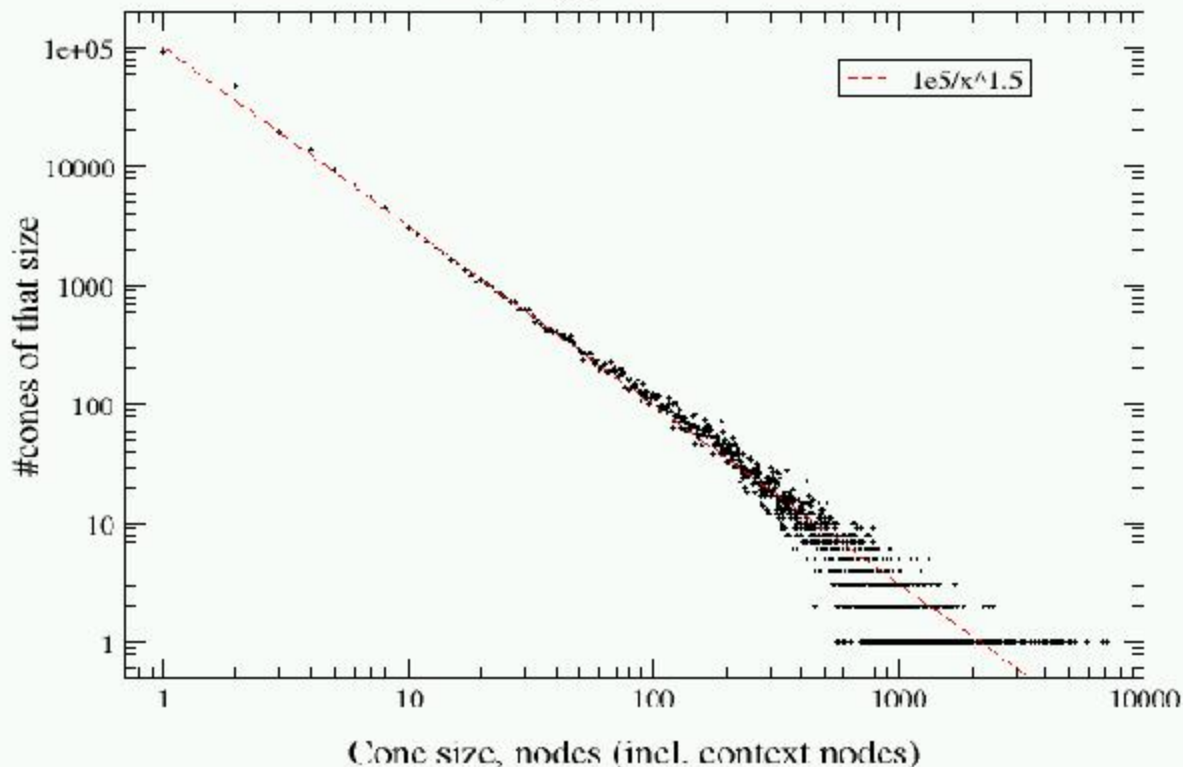
- motivation

- aggregating IPs into prefixes can create artificial connectivity
- 152.130 (customer gateways for large backbone, not intraprefix-connected)
- prefix outdegree thus misleading/specious

...so: deaggregate prefixes to find (really) connected subcomponents

# Cone sizes for IP nodes on full skitter graph (all levels)

Data: 2000-11-23..29 (7 days). nnode=1336707 nlink=2115274



## BGP observations

---

estimates for number of distinct policies

- somewhat independent of peer selection
- internal details of global routing varies
- between 2% and 70% of atoms ramified, depending on peer choice
- suggests vastly different transit policies among/within networks
- BGP tangle
  - atom with directed loop (apparent routing loop)
  - should not happen...

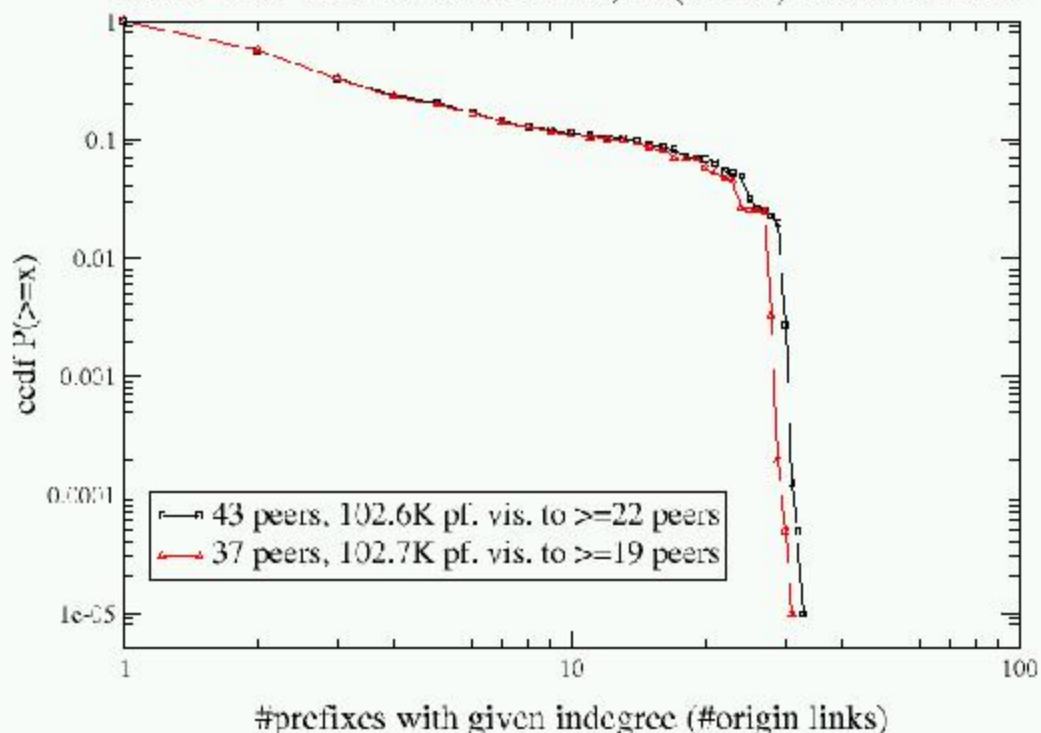
use of atoms can reduce table size in half while still preserving AS path and specificity relationships

- theoretical result...

# prefix indegree counts

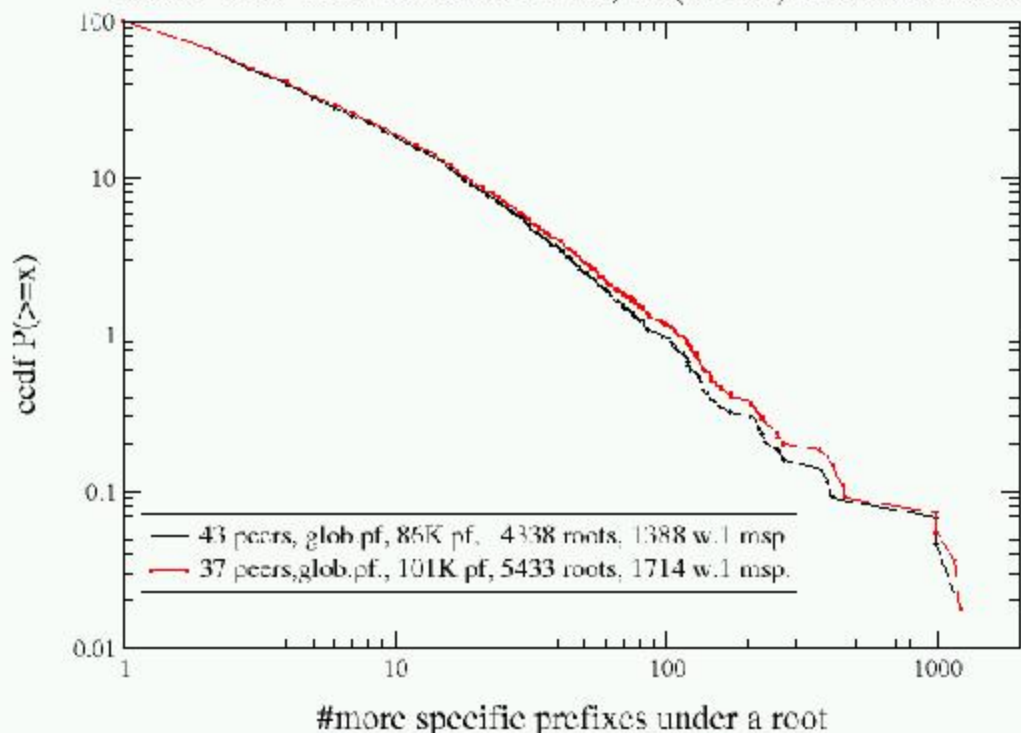
## Prefix indegrees (origin link set counts)

Route View BGP data 2001-09-01, 43 (37 full) backbone tables



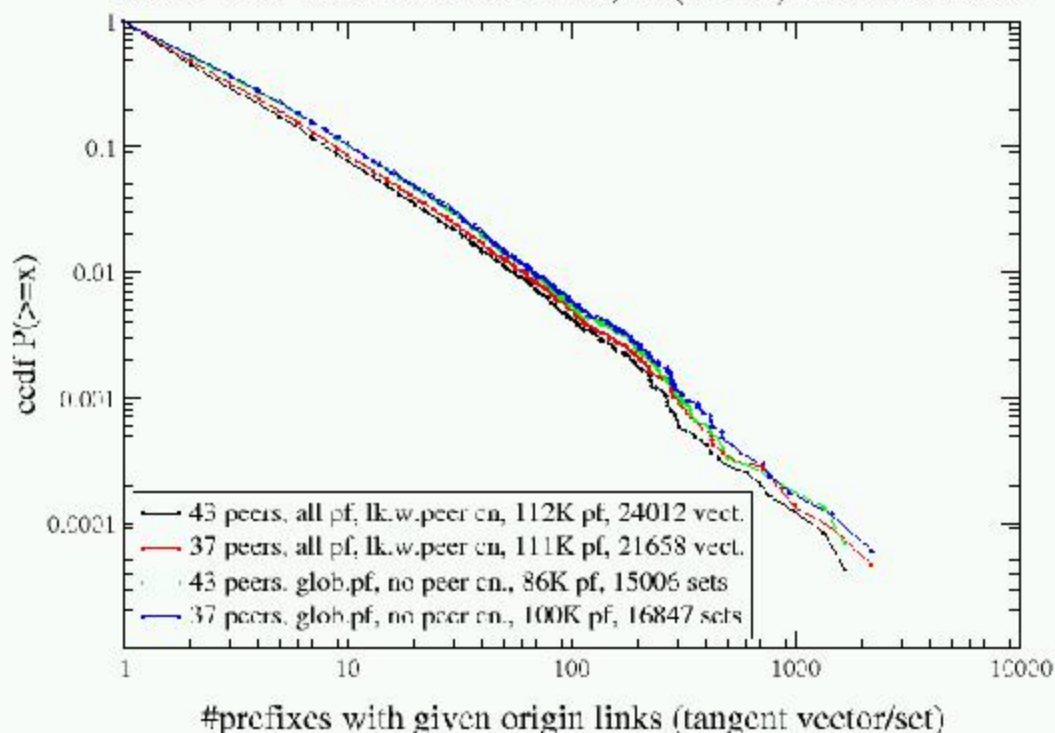
# trees of more specifics

Counts of more specifics under the same root prefix  
Route View BGP data 2001-09-01, 43 (37 full) backbone tables



# tangent vectors (prefix indegree multisets)

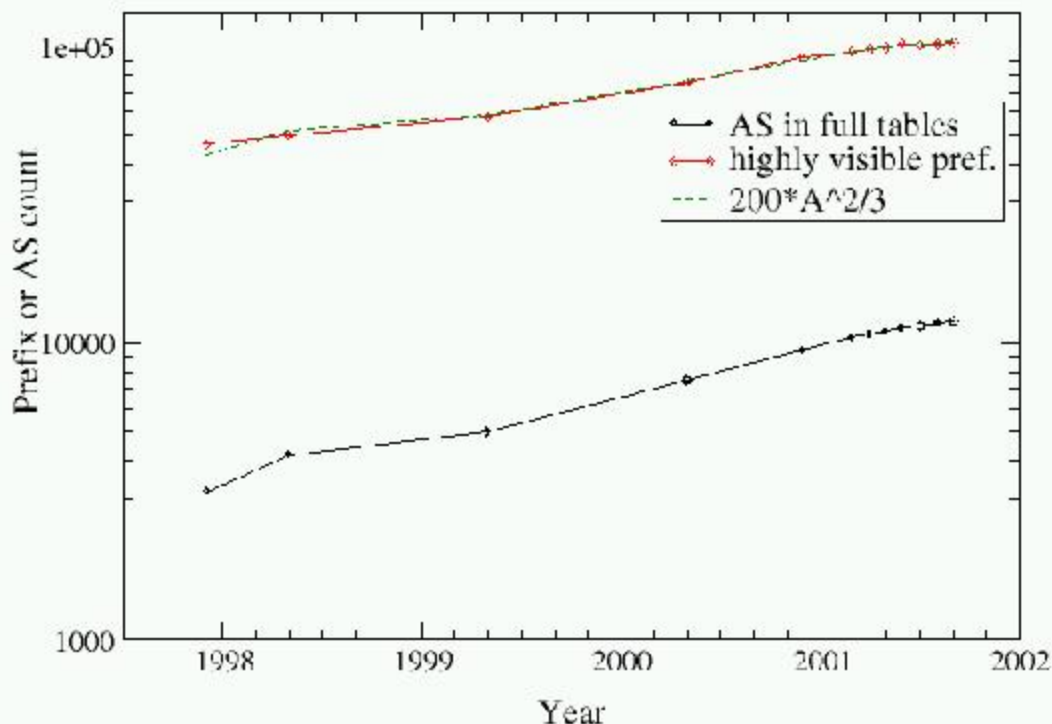
Prefix counts for origin link sets with and w/o peer counts  
Route View BGP data 2001-09-01, 43 (37 full) backbone tables



# global invariant: Broido's '200 formula'

Prefix vs. AS growth:  $P = 200 * A^{2/3}$

Route Views 1998-2001. Prefixes present in  $> 1/2$  of all full tables



---

end part 1

# global Internet topology analysis on IP level

---

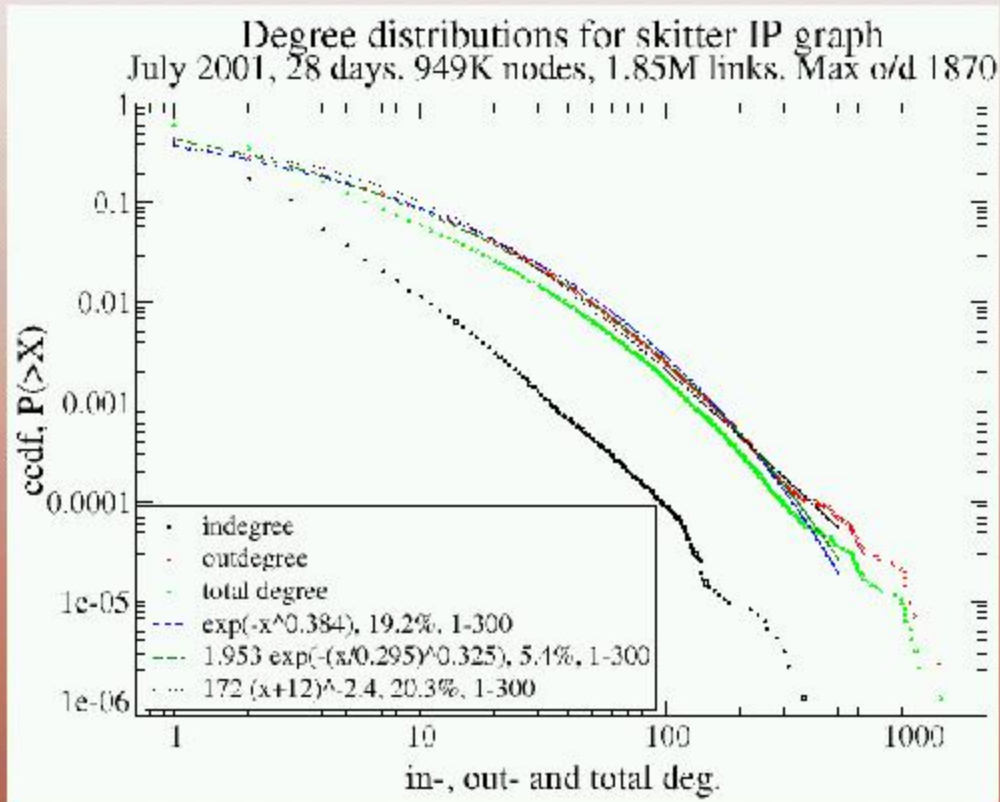
caida probed topology data of Jul.15-18 2001

- 826265 nodes
- 60329 (7%) in combinatorial core
  - stripped in 24 iterations
- 52330 nodes in giant connected component
  - bidirectionally connected components of the core
  - 200 times larger than any other connected component
  - almost 90% of the core
  - largest of small components are /24s, /25s

Most Internet object size distributions  
are closer to Weibull  
than to power functions.

- $P(X > x) = a \exp(-(x/b)^c)$
- decreases faster than power function, slower than exponential
- exception: AS outdegree in RouteViews BGP AS graph is on some days closer to power function
- another possible family more commonly used by physicists  
 $a x^c \exp(-x/b)$

# full IP graph degree distribution, Jul 2001



## full IP graph degree distribution, jul 01

---

Approximation properties:  
outdegree interval 1–300

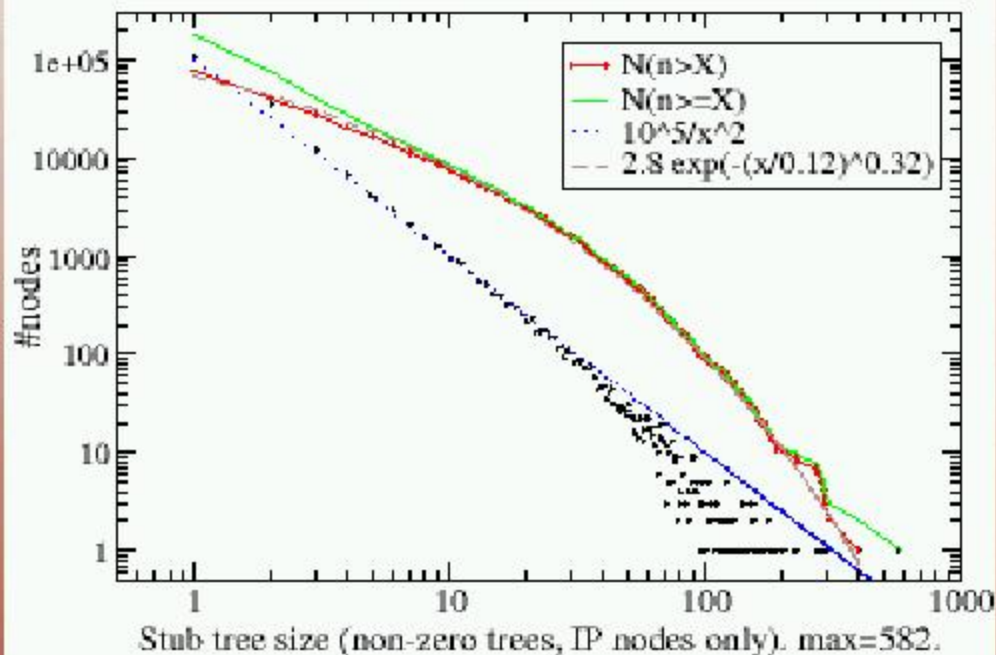
Relative error:  $\max(y(x)/a, a/y(x) \mid 1 \leq x \leq 300) - 1$

Weibull with three parameters:	5.4%
Weibull with 1 parameter:	19.2%
Shifted power function:	20.3%

# tree size distribution

55% IP nodes are in stub trees

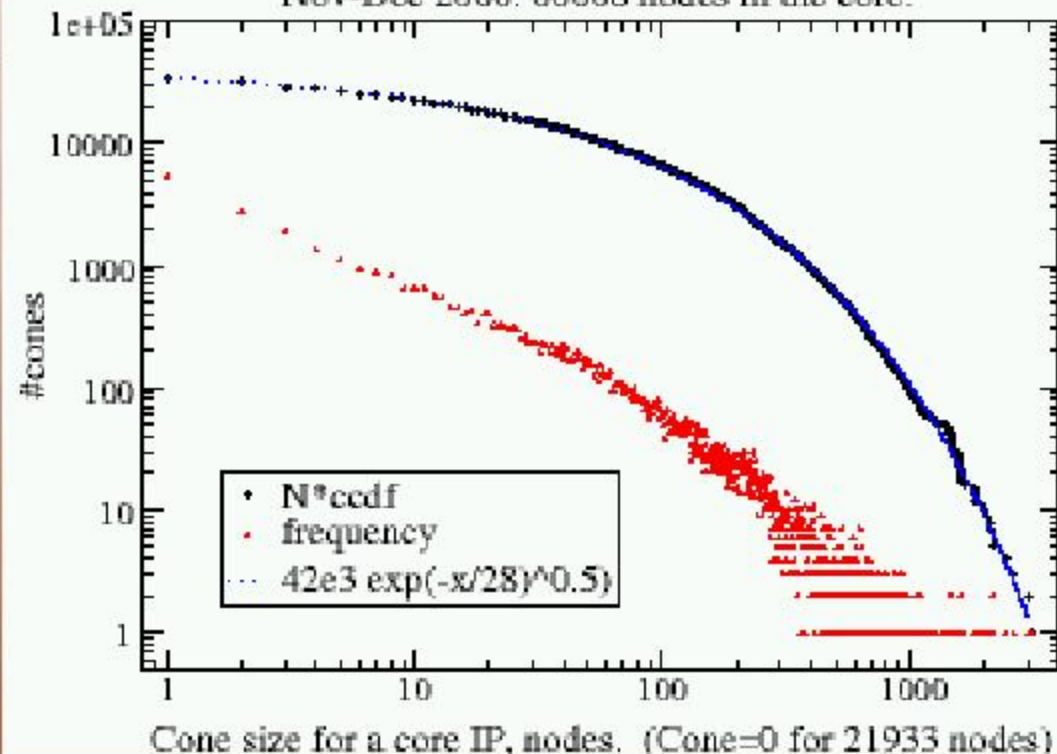
Stub tree size distributions for IP graph. 348K tree nodes  
28 days, Nov-Dec.2000, 17 mon. 182K trees; size 1: 107 K



(Trees node counts do not include roots)

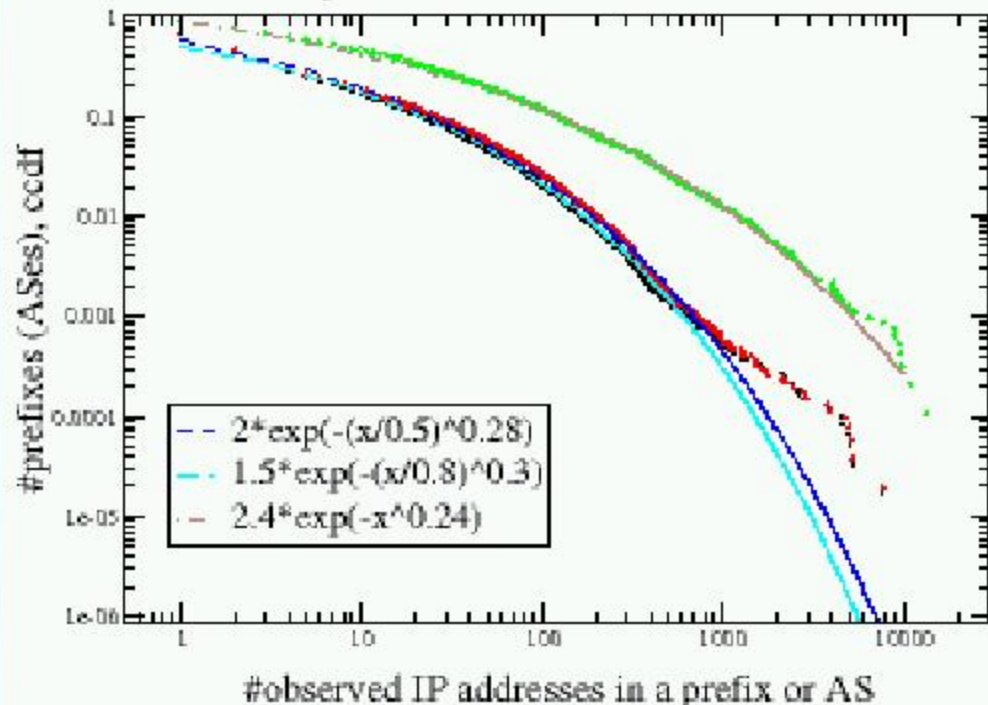
# cone size distribution

Cone sizes for IP core nodes  
Nov-Dec 2000. 60008 nodes in the core.



# Size of [least specific and all] prefixes, and of ASes

Distributions for #observed IP addresses in a prefix/AS  
28 days: 2000-11-21..12-18. 17 monitors



## why do Weibull/power functions arise?

---

- simple model: full binary tree
- $2^k$  nodes  $k$  hops from the root
- subtree for each node has size  $2^{(n-k)}$
- #subtrees =  $2^n /$  subtree size

# Why do all size distributions look Weibull?

Where do Weibull or shifted/cut-off power functions come from?

## Suggestion: Use Coalescence Theory

- used to model size distributions of populations of physical objects that merge in nature
  - applied to: formation of cloud droplets, bubbles, polymers, stars, Universe...
- 
- Idea: Objects merge together, size = sum of both sizes
    - Rate of aggregation depends upon both objects' sizes
    - Special case: multiplicative coalescence, where probability of merging  $\sim$  product of sizes

# Why do all size distributions look Weibull?

---

Math model: Smoluchowski coagulation equation

- M. v. Smoluchowski, Drei Vortraege ueber Diffusion, Brownische Bewegung und Koagulation von Kolloidteilchen. Physik. Z., 17:557-585, 1916
- D.J. Aldous. Deterministic and stochastic models for coalescence (aggregation, coagulation): a review of mean-field theory for probabilists. July 10, 1997  
[www.stat.berkeley.edu/users/aldous](http://www.stat.berkeley.edu/users/aldous)

(Thanks to Reka Albert for the reference)

# Coalescence probability is size-dependent

(rationally moderated chaos)

- merging 100,000 objects of size 1
- randomly select two objects to merge
  - fractional multiplicative coalescence
  - $P(x_1, x_2) = P(x_1) P(x_2|x_1)$
  - $P(x_1) \sim x_1^a$
  - $P(x_2|x_1) \sim x_2^a$ , with  $x_1$  is excluded ( $\Rightarrow$  asymmetric)
  - $a$  has the meaning of inverse temperature
  - higher  $a \Rightarrow$  lower temperature  $\Rightarrow$  larger clusters
  - $\Rightarrow$  freezing (gelation) sooner
  - $a=0$  (size-independent):
    - infinite temperature
    - very small clusters
- take snapshots every 10K size distributions
  - (after 10K, 20K, ... 99.99K mergers)

# Coalescence simulation

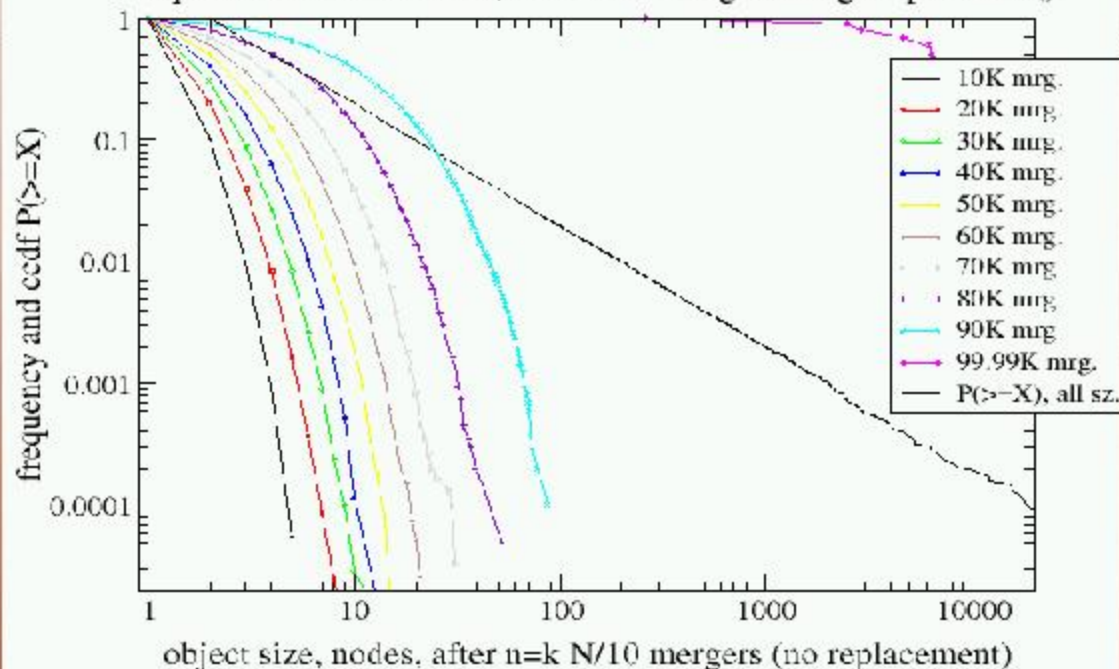
---

- next few graphs will show for coalescence simulation
  - object size distributions
  - maximum object size distribution
  - for size-independent or multiplicative or submultiplicative (sqrt(multipl)) coalescence
- resulting size distribution: exponential
  - as predicted by analytic solution to Smoluchowski equation

# Coalescence object size (additive)

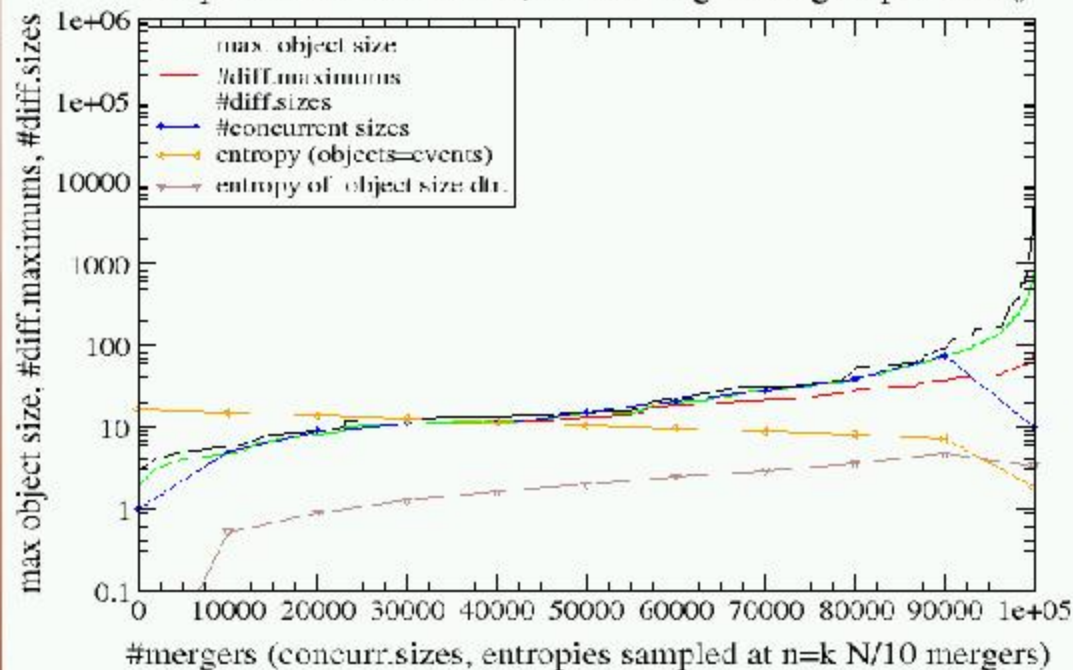
Object size distribution after  $n$  mergers, no replacement

Sel.prob.  $\sim$  size<sup>-0.1</sup>. 100K nd, 10..99.99K mrg. Rand.gen. perl rand()



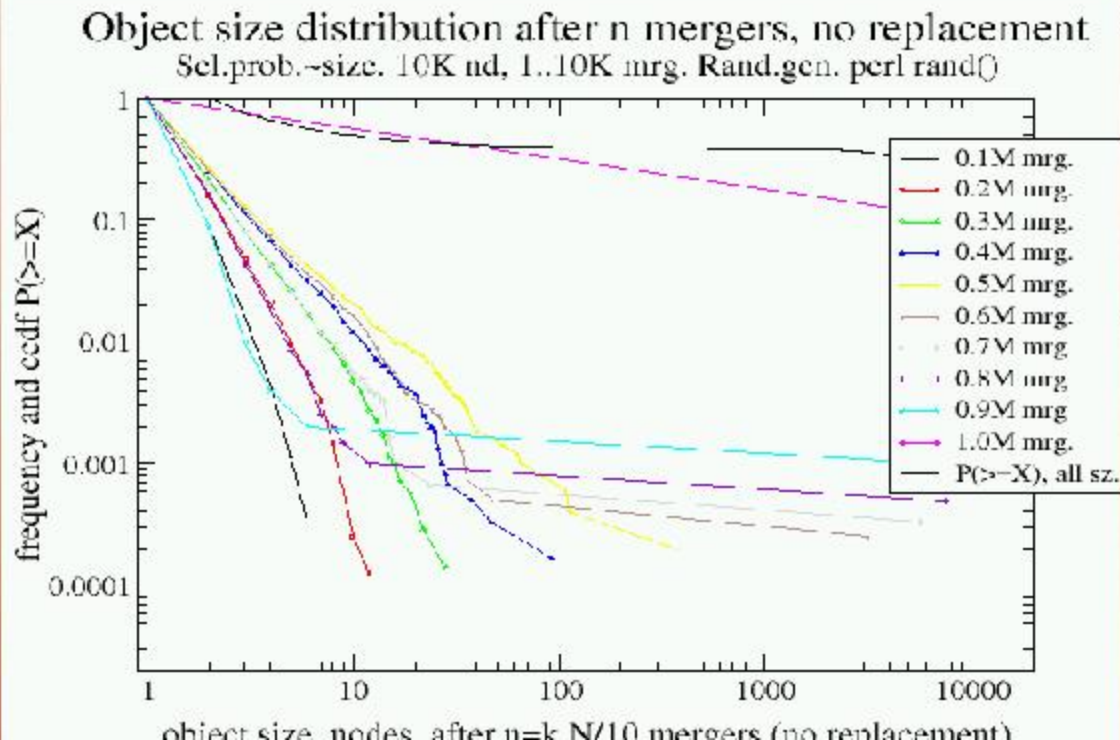
# Coalescence: evolution of maximum obj size

Maximum object size after  $n$  mergers, no replacement  
Sel.prob.  $\sim$  size<sup>0.100K</sup>nd, 99.99K mrg. Rand.gen. perl rand()



# If coalescence probability $\sim$ product of sizes

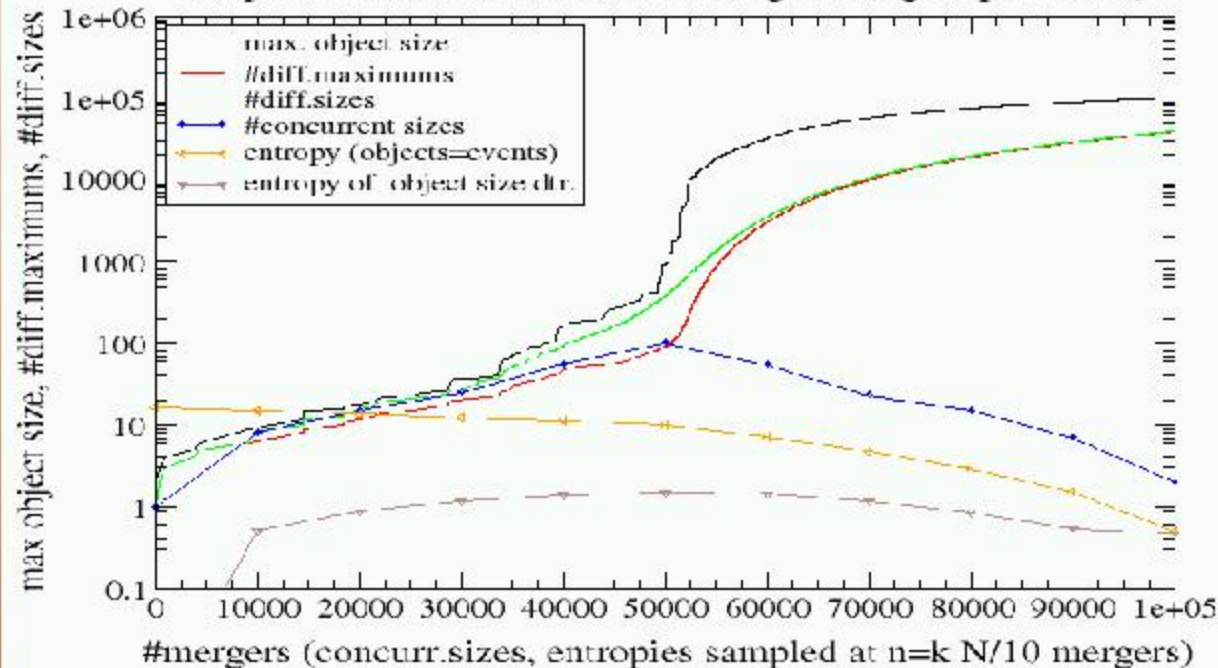
- size distribution is close to power function
- giant component emerges after 50% are aggregated



# Coalescence: evolution of maximum obj size

probability  $\sim$  product of sizes

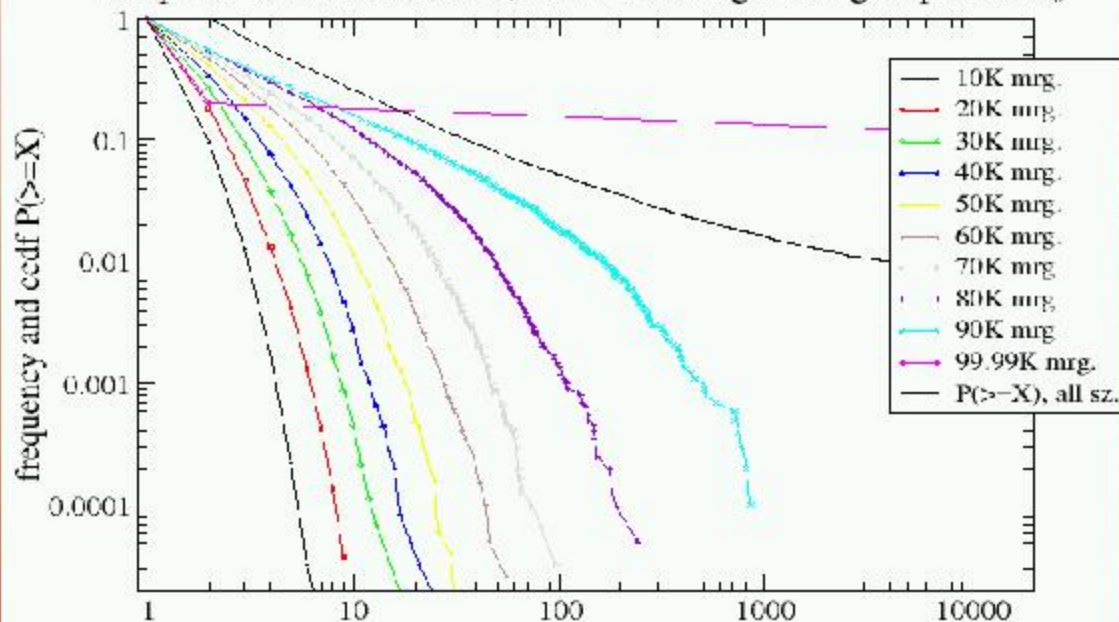
Maximum object size after n mergers, no replacement  
Sel.prob.  $\sim$  size. 100K nd, 99.99K mrg. Rand.gen. perl rand()



## Coalescence prob. $\sim \text{sqrt}(\text{product of sizes})$

- size distribution is close to power function
- giant component emerges after 90% are aggregated
- size distributions close to Weibull (rel. err.  $< 20\%$  for generic sizes)

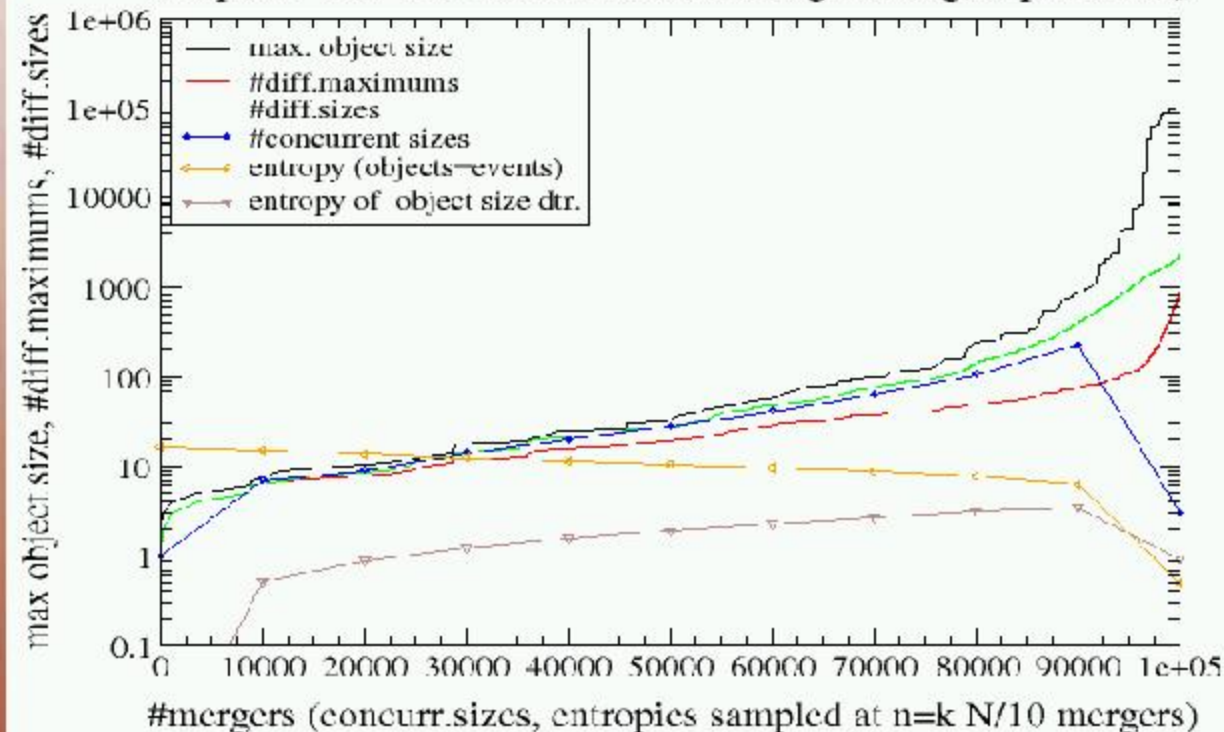
Object size distribution after  $n$  mergers, no replacement  
Sel.prob.  $\sim \text{size}^{0.5}$ , 100K nd, 10..99.99K mrg. Rand.gen. perl rand()



# Coalescence: evolution of maximum obj size

Maximum object size after n mergers, no replacement

Sel.prob.  $\sim \text{size}^{0.5}$ . 100K nd, 99.99K mrg. Rand.gen. per l rand()



# topology: combinatorial core vs gcc

---

## filtering of the graph

### combinatorial core

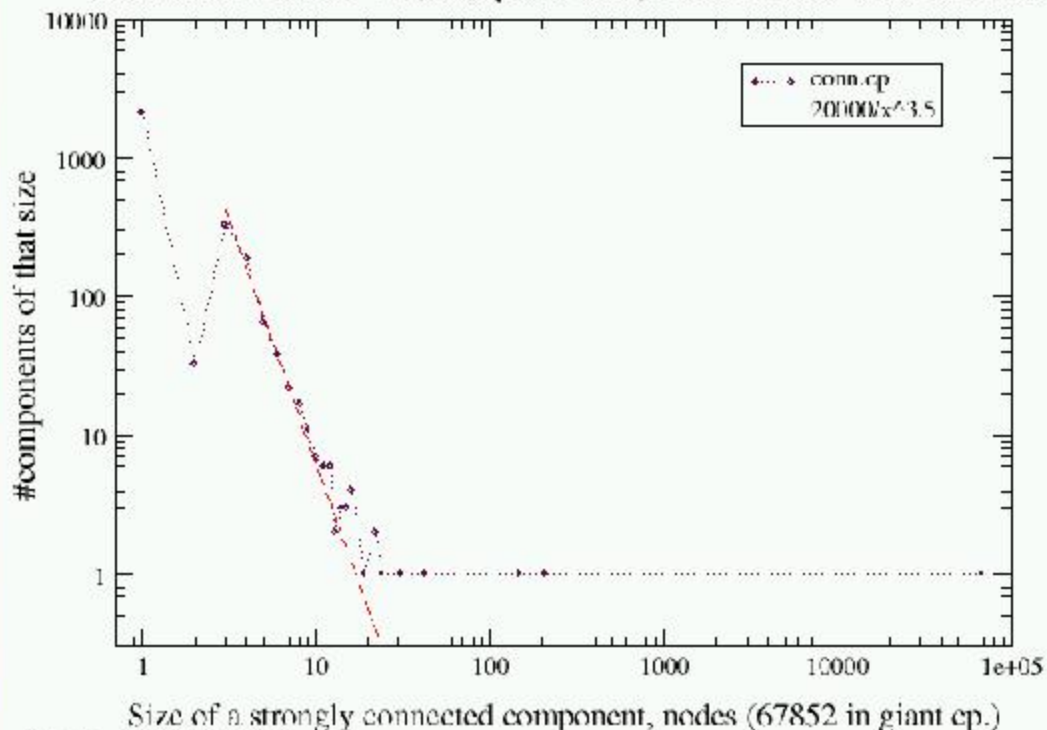
- remove nodes of outdegree 0 (no outgoing links)
  - destinations have outdegree 0
- strip nodes of outdegree 0
- stub network's routers now have outdegree 0
- repeat (recursive) stripping of outdegree 0

### giant connected component

- bidirectionally connected components of the graph

# topology data: connected components

Connected component size distribution for skitter IP arc core graph  
Data 2000-11-21..12-18 (28 days, 17 mon). nnode=73687 nlink=710770



# hop metric of 'centrality'

- shortest path distance on IP graph
- find maximum distance from any given node to all core
- center of the core: those with min maximum distance
- e.g., take top 50 by this metric

## top 3 prefixes

Prefix	Outdegree
157.130.0.0/16	1172
144.232.0.0/16	757
4.0.0.0/8	655

# reachability functions

---

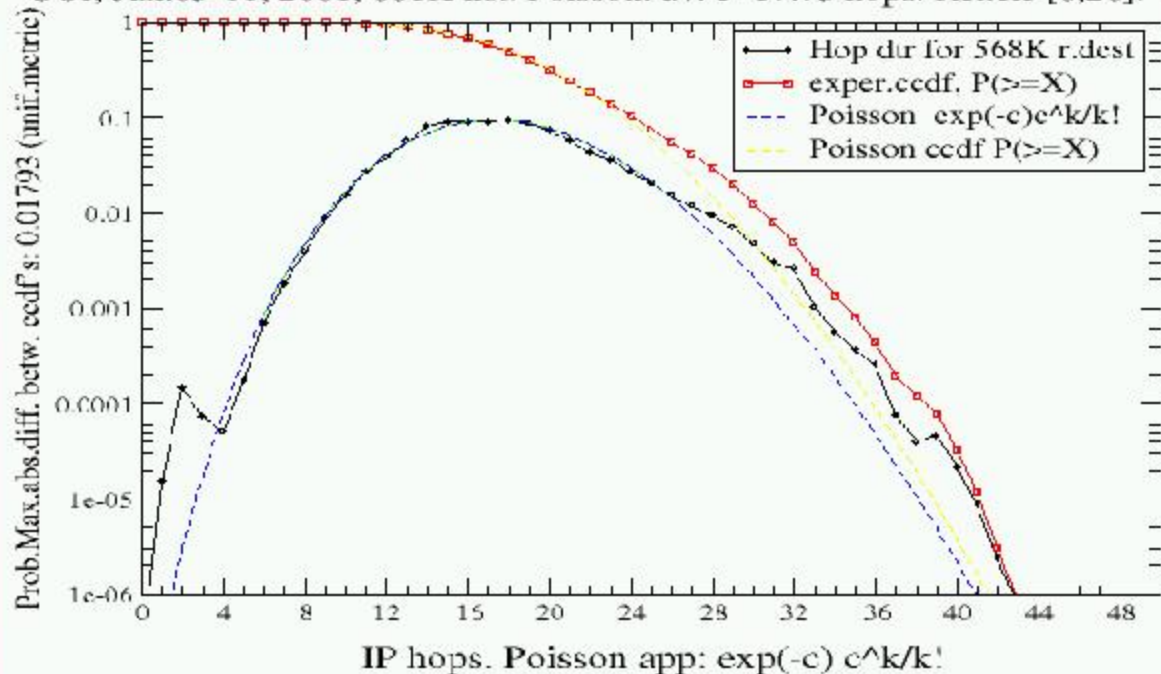
previous work: counts of nodes at shortest path distance  $x$   
(H.Tangmunarunkit, R.Govindan)

here: reachability functions are for individual nodes.

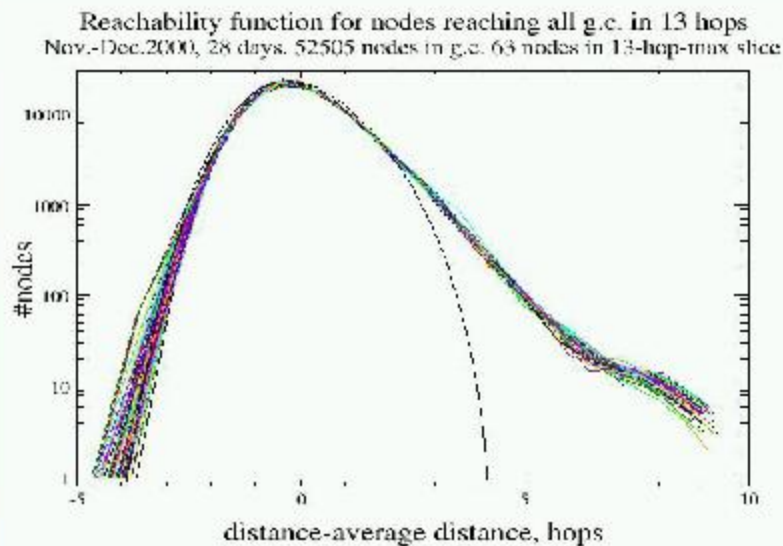
# reachability functions

Poisson approximation to monitor's hop count distribution

UOr, Jun.05-11, 2001, 661K list. Poisson: av.  $c=17.73$  hops. Rel.err [6,26]: 20.18%



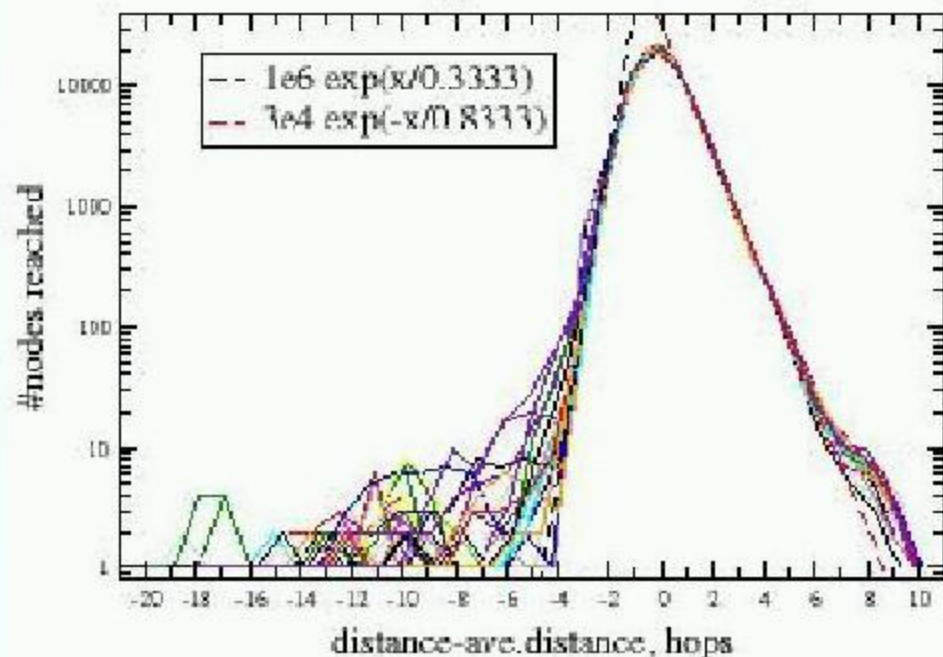
# neighbourhood size vs. hop distance



- 'neighbourhood': nodes at radius  $R$
- hypothesis: there exists a mother function that fits all these curves
- such a function would provide a parameteric method for finding topological distance from internet core

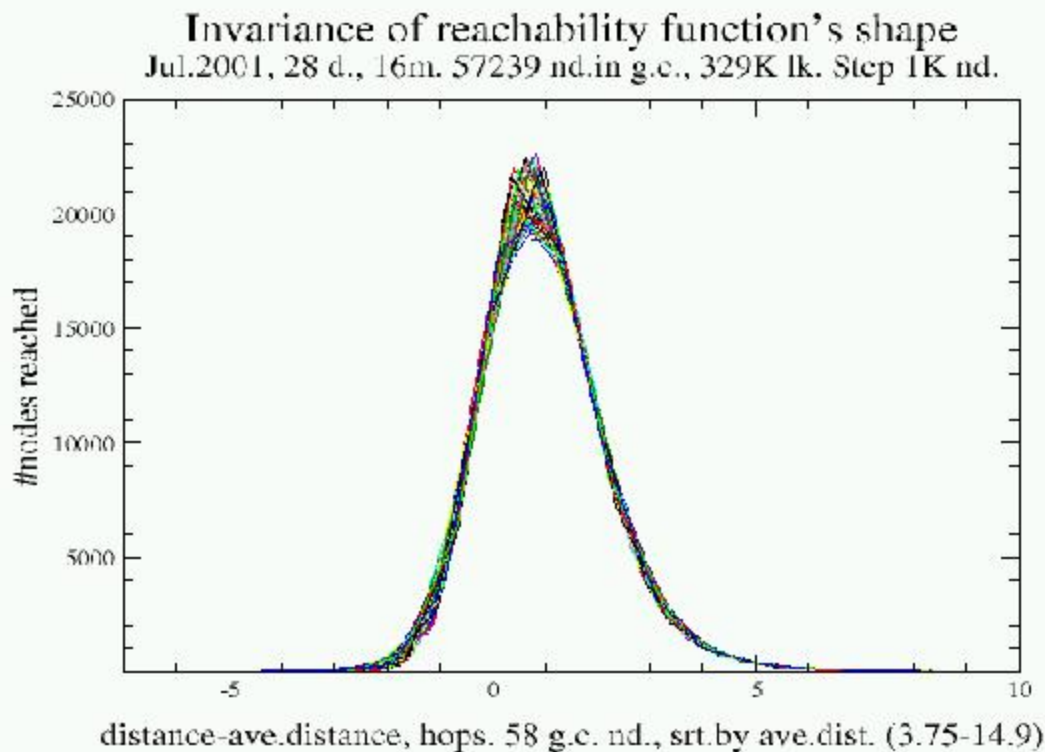
# Neighbourhood size

Reachability function for peripheral nodes  
Nodes which reach all g.c. in 25 or more hops (75 curves)



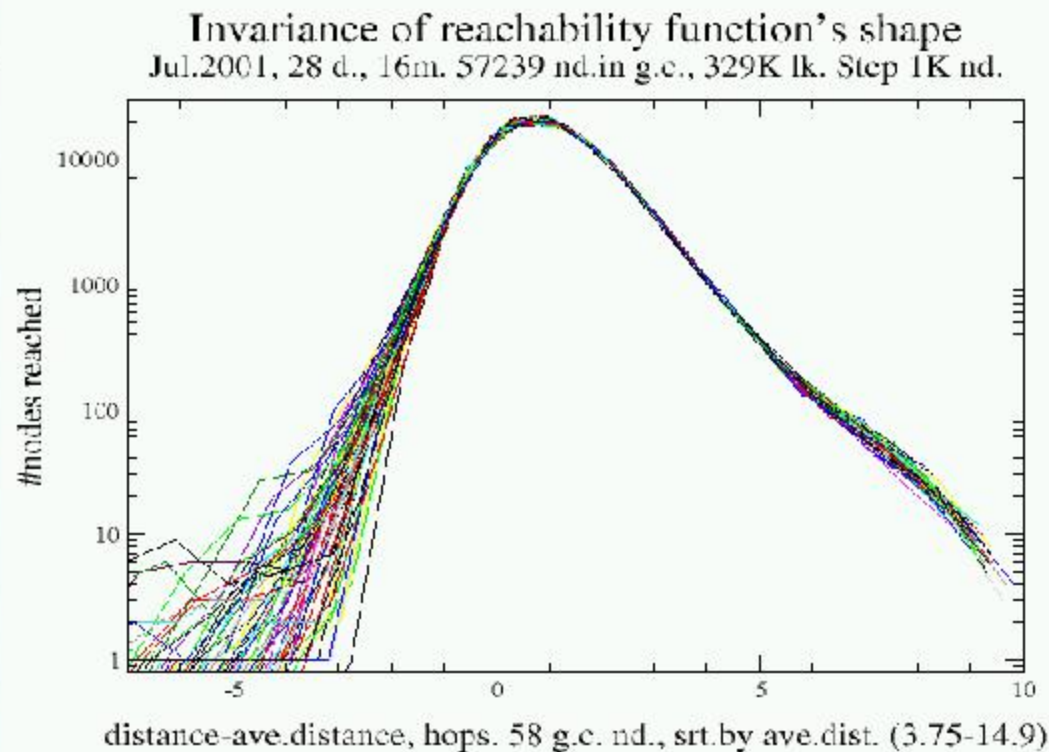
Neighbourhood size: (step 1000 thru giant comp.)

Skitter data of Jul.02-29, 2001



Neighbourhood size: (step 1000 thru giant comp.)

Skitter data of Jul.02-29, 2001



## So... neighbourhood size suggests:

---

- mother function is invariant of graph
  - shifted and stretched for each node
  - shift, stretch = average, std dev of hop distance from a node

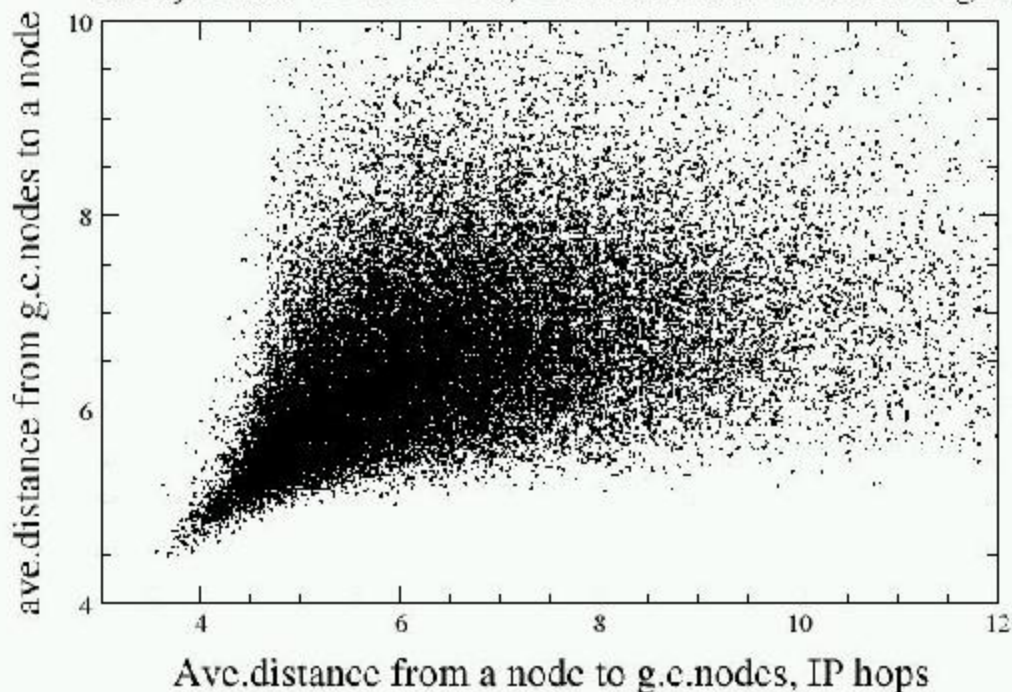
### likely implications:

- there IS a core of the graph
- neighbourhood is small before we reach core
- starts growing as soon as we reach core
- "Gaussian" because minimums are a limiting case of addition

# shortest path distances in GCC

## Shortest path distances in giant component

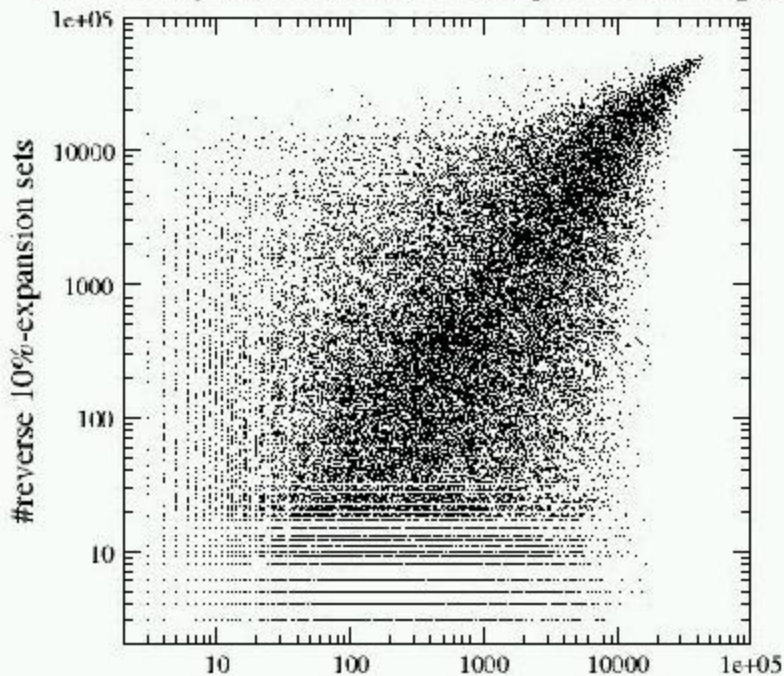
28 days: 2000-11-29..12-18, 17 mons. 52505 IP nodes in g.c.



# topology: forward & reverse expansion sets

## Forward and reverse expansion sets

How many times a node belongs to 10% expn. set



#sets which reach forward to 10% of g.c., containing a node

## topological resilience to outbound link removal

---

algorithm:

- forwarding deactivated (outbound links removed) on all nodes with outdegree over  $x$

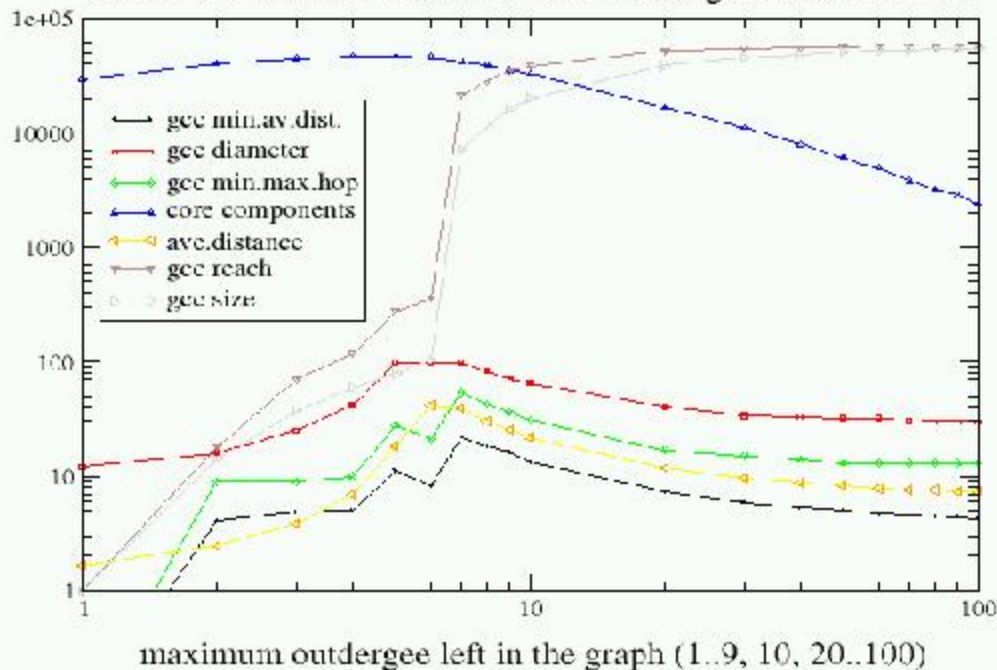
rationale:

- will still have these nodes reachable (do not remove inbound links and/or nodes)
- removing all nodes with outdegree  $> x$  is unambiguous

# topological resilience to outbound link removal

IP giant component (skitter, Jul 2001)

IP giant comp. metrics with nodes over outdegree  $\times$  deactivated  
Skitter 20010702-29. 28 d. 16 mon. 57239 g.e. nd, 329074 lk.



# topological resilience to outbound link removal

---

IP giant component (skitter, Jul.02-29 2001)

giant component breaks only when outdegree 1-6 remain

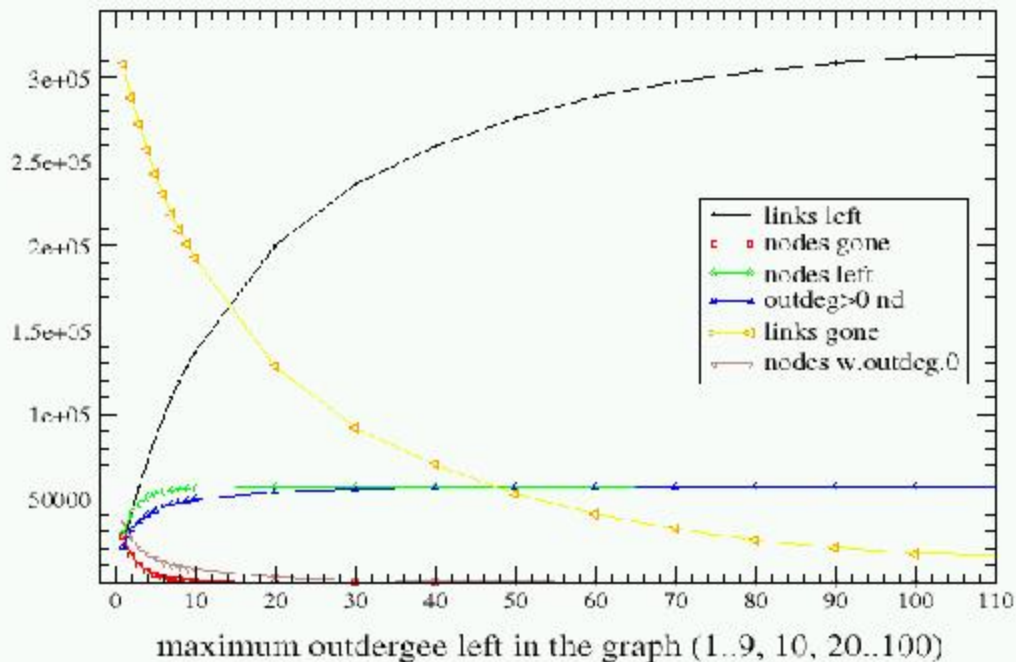
- maximum outdegree was 329
- forwarding deactivated on 21% nodes
- 70% links removed

astonishingly robust graph

# topological resilience to outbound link removal

IP giant component (skitter, Jul 2001)

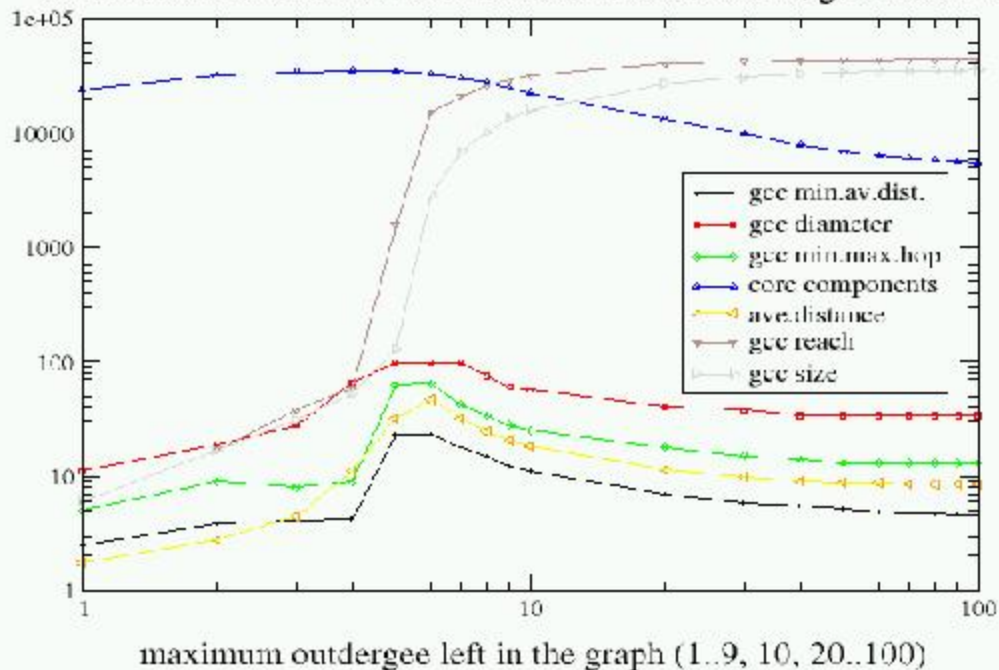
IP giant ep. metrics when nodes over outdegree  $x$  are disabled  
Skitter 20010702-29. 28 d. 16 mon. 57239 nd., 329K lk. in g.c.



# topological resilience to outbound link removal

combinatorial IP core (skitter, 1 Aug 2001, 10 days)

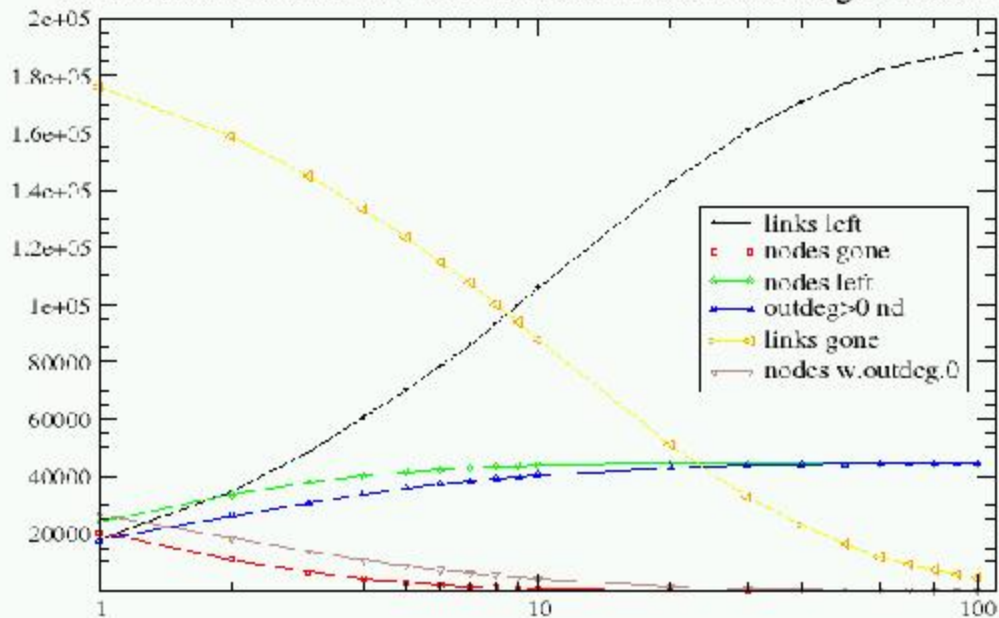
IP core metrics when nodes over outdegree  $x$  are disabled  
Skitter 20010723..0801. 10 d. 16 mon. 44356 core nd, g.c. 36090 nd.



# topological resilience to outbound link removal

combinatorial IP core (skitter, 1 Aug 2001)

IP core metrics when nodes over outdegree  $x$  are disabled  
Skitter 20010723..0801. 10 d. 16 mon. 44356 core nd, g.c. 36090 nd.

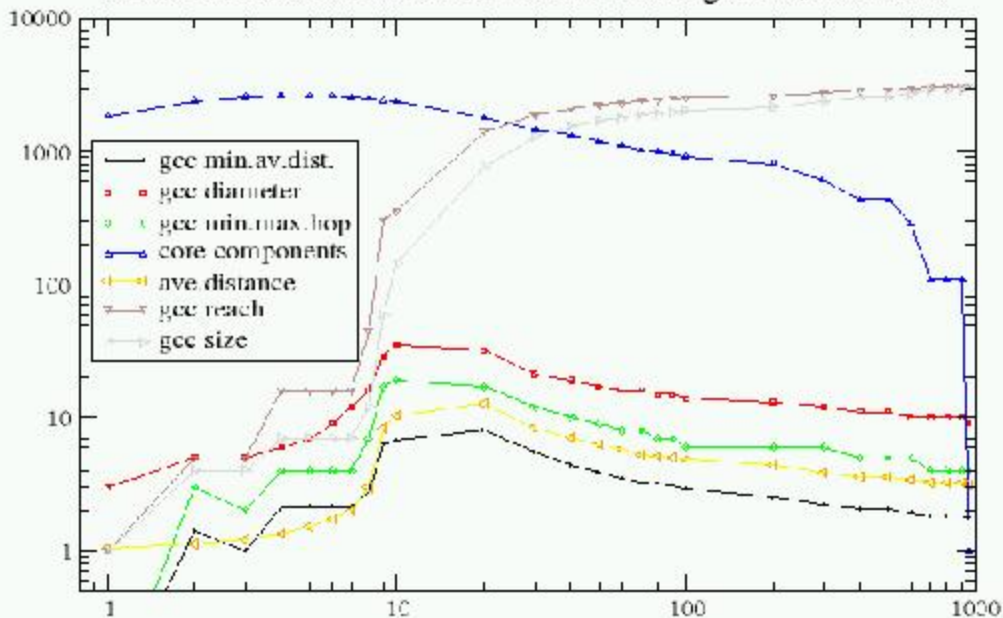


maximum outdegree left in the graph (1..9, 10, 20..100)

# topological resilience to outbound link removal

Giant component of skitter AS core (skitter, Jul 2001)

AS giant comp. metrics with nodes over outdegree  $\times$  deactivated  
Skitter 20010702-29. 28 d. 16 mon. 3045 g.e. nd, 20619 lk.



maximum outdegree left in the graph (1..9, 10, 20..100,200..900)

## topological resilience to outbound link removal

---

Giant component of skitter AS core (skitter, Jul 2001)

- Contains 30% of AS nodes
- (BGP AS graph's core: 3.3%, Route Views Aug.01)

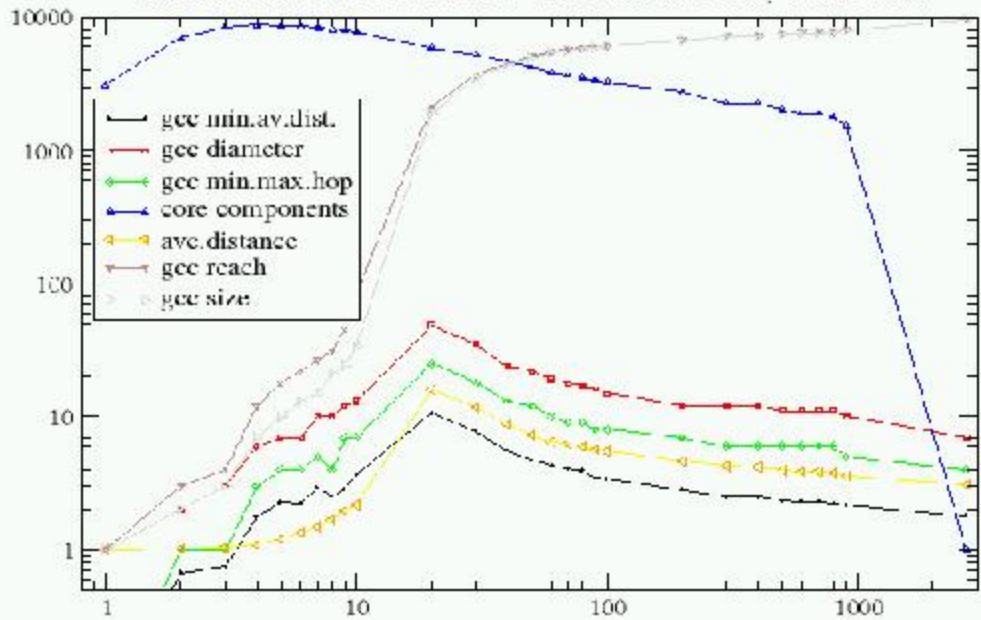
Giant component breaks only when outdeg.1-10 remain

- maximum outdegree was 949
- forwarding deactivated on 10% nodes
- 68% links removed

# topological resilience to outbound link removal

symmetrized skitter AS graph (Jul 2001)

Symm. AS graph metrics w.nodes over outdegree x deactivated  
Skitter 20010702-29. 28 d. 16 mon. 9602 nd, 54614 lk.



maximum outdegree left in the graph (1..9, 10, 20..100,200..1000)

## topological resilience to outbound link removal

---

symmetrized skitter AS graph (Jul 2001)

Giant component breaks only when outdeg.1-10 remain

- maximum outdegree was 949
- forwarding deactivated on 6.4% nodes
- 66% links removed

---

(end part 2 (end talk...))

## note on forward IP topology vs BGP RouteViews data

---

for july 2001

- BGP routeviews: 34 peers, 9.4k ASes, ~100K prefixes
- skitter: 17 monitors, 300k dsts, 50K prefixes

to equalize, have to reduce skitter to AS graph:

- 0) pick one trace per prefix per day  
    ignore nodes next to non-responding hops
- 1) strip IP leaves
- 2) convert IP -> prefixes
- 3) strip prefix leaves
- 4) convert prefix -> AS
- 5) strip AS leaves

## forward IP topology vs BGP RouteViews data

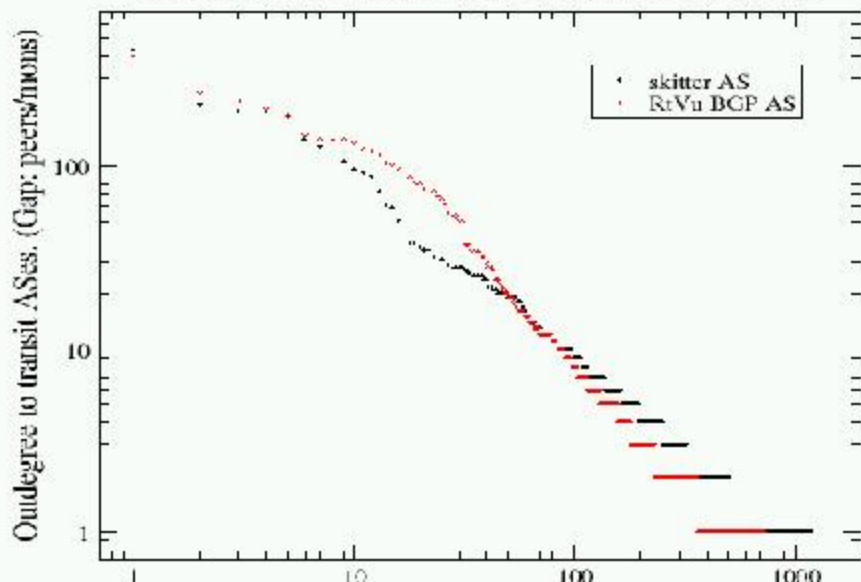
- basically a ton of decimation on a day of skitter data
- skitter AS graph: 6949 AS nodes, 16145 AS links (peering sessions)
- remove any potential advantage of skitter probing
  - frequency
  - finer (IP) granularity
- nonetheless, skitter captures much larger share of the Internet's bidirectional connectivity.
- skitter combinatorial core (full-transit portion) contains 988 AS nodes, as opposed to BGP's 299 nodes (3.3X more)

# ranked list of skitter vs RV outdegrees

level 2 transit AS (those with at least one outbound connection to another transit AS, as opposed to stub AS (customers)).

Of those: 732 in Oregon Nov.29 BGP data and 1189 in skitter data.

Comparing skitter and Oregon BGP AS outdegree  
2000-11-29, 37 peers, 17 mon, 1 tr/pref, no mult.or.



AS rank: 1-1189 skitter, 1-732 BGP. Top 5: 701, 1239, 3561, 6461, 1

## upshot

---

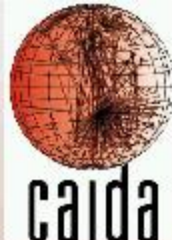
even the best available inter-domain routing (BGP) data serves very weak substitute for IP probed topology data

and yet this BGP data is an essential tool for sensible comprehensive macroscopic Internet topology analysis

## ongoing/future

---

- continue building infrastructure
  - sources
  - destination lists
- dynamically discover topology, b/w
- correlation
  - among sources
  - with perf, topology
- identify critical pieces of infrastructure
  - continue w metrics and analysis
    - ▶ 'center', 'core', 'close'
  - (optimize dist. server architectures)
  - filter new root locations?
  - determination of connectivity metrics
    - ▶ persistence of paths
    - ▶ metrics for 'change'
  - better GLS support
  - model of the Internet (no, really)
  - graphing/layout techniques....



[www.caida.org/Presentations/](http://www.caida.org/Presentations/)

andre broido  
kc claffy  
UCSD/SDSC/CAIDA  
[broido@caida.org](mailto:broido@caida.org)  
[kc@caida.org](mailto:kc@caida.org)  
[www.caida.org](http://www.caida.org)