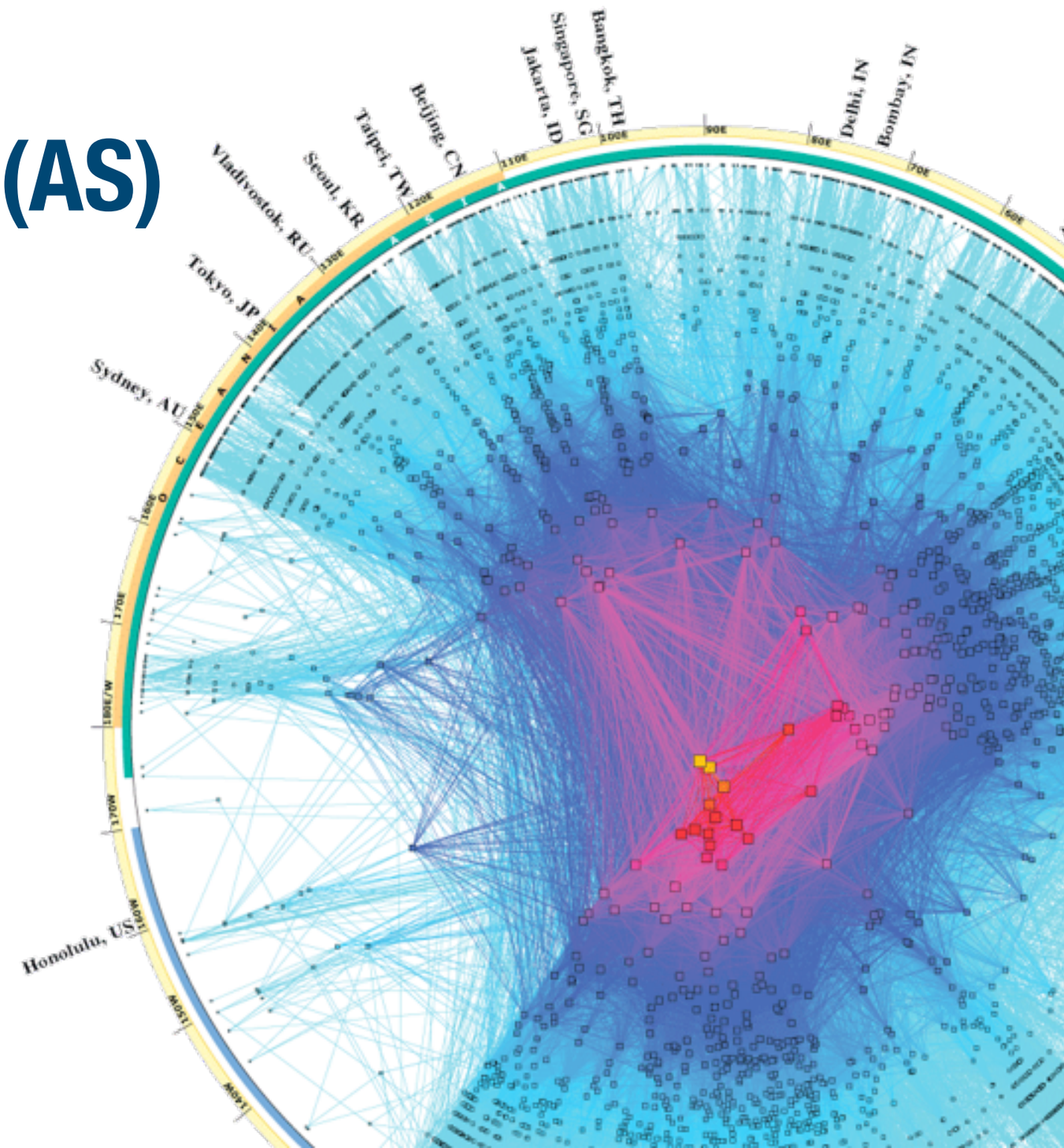
Autonomous Systems (AS) Introduction and Visualization

Bradley Huffaker
CAIDA
SDSC/UCSD

Data Visualization (CSE 199)
22 January 2016



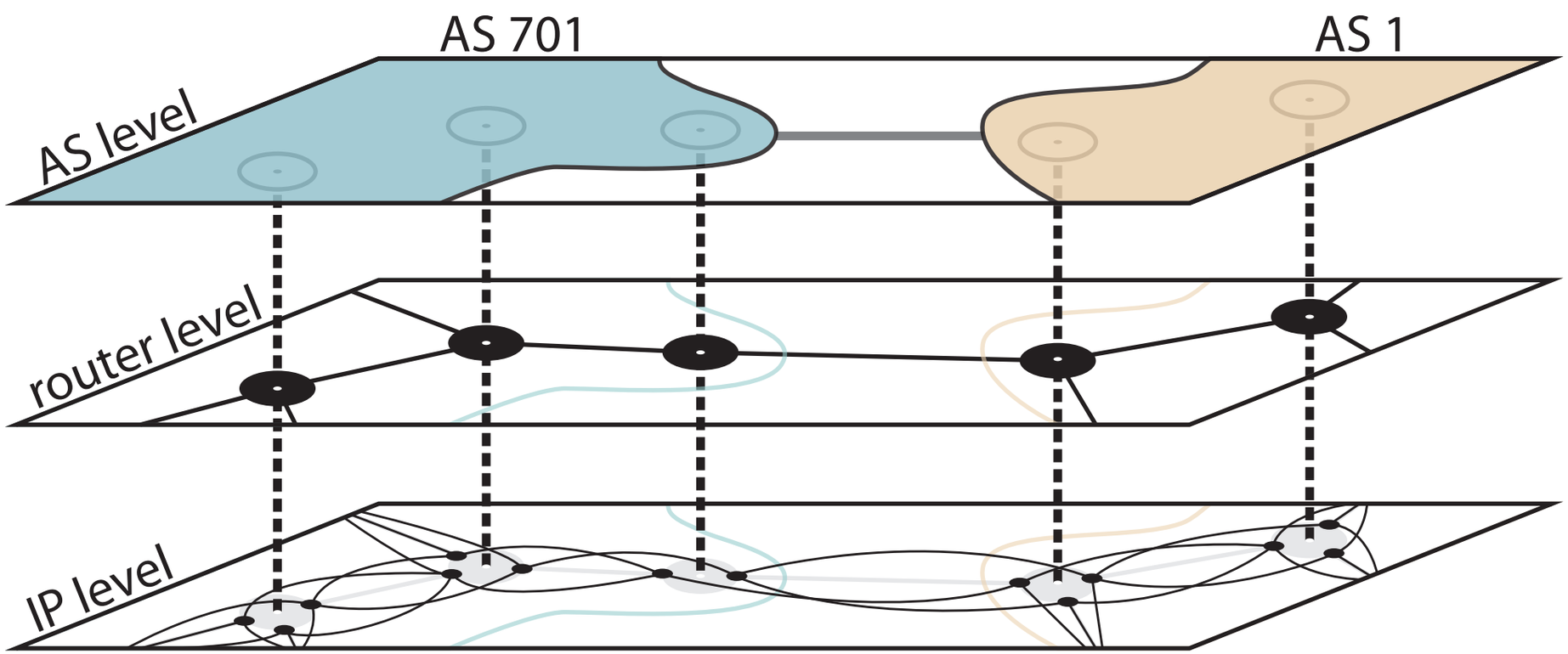
overview

- introduction
- visualizations
  - AS Core
  - AS Core single AS
  - AS Core single Organization
  - AS Relationships
- data files (URL)
- summary

Internet of layers

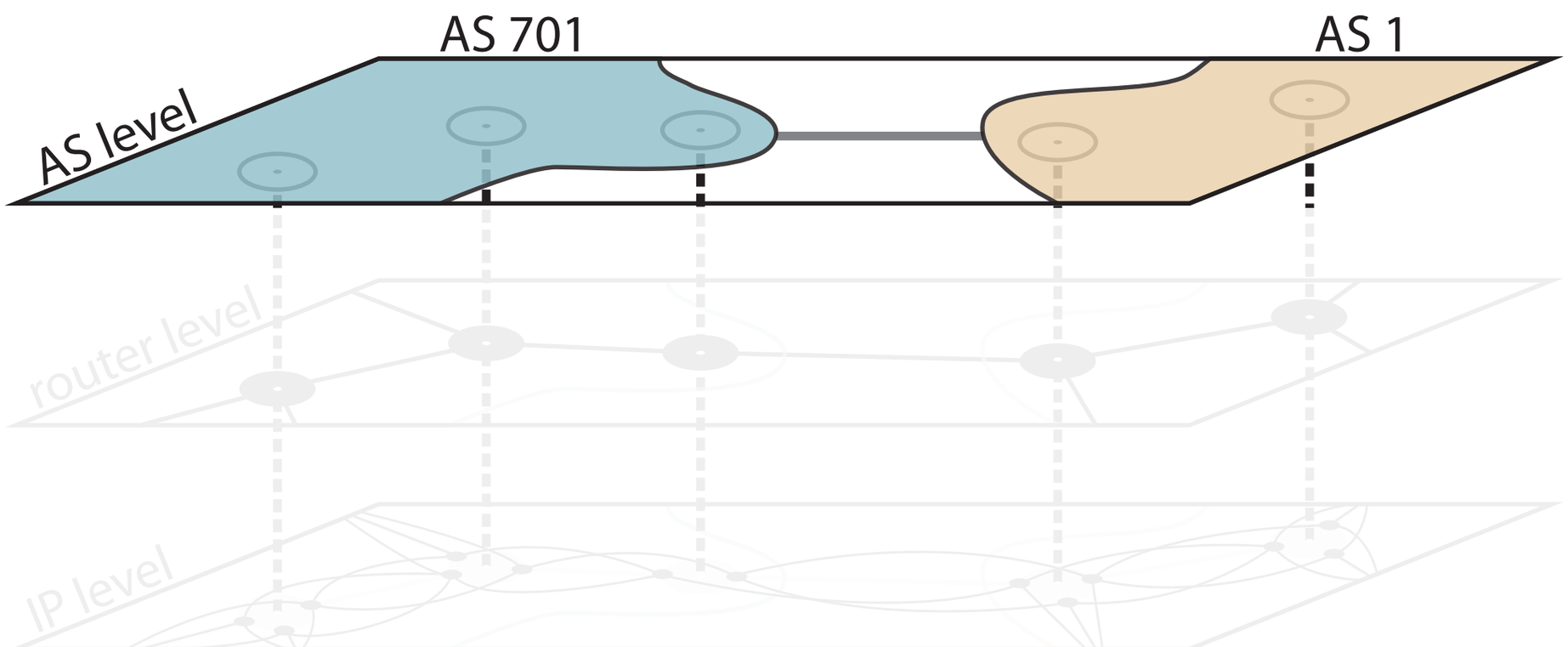
Internet maps can be grouped into three levels.

1. **IP addresses** that connect devices on to the Internet.
2. **Routers,** machines that route the traffic, interconnect via **IP addresses.**
3. **Autonomous Systems (AS)** are numbers used to route groups of **IP addresses**.



AS layer

We will focus on the AS layer.



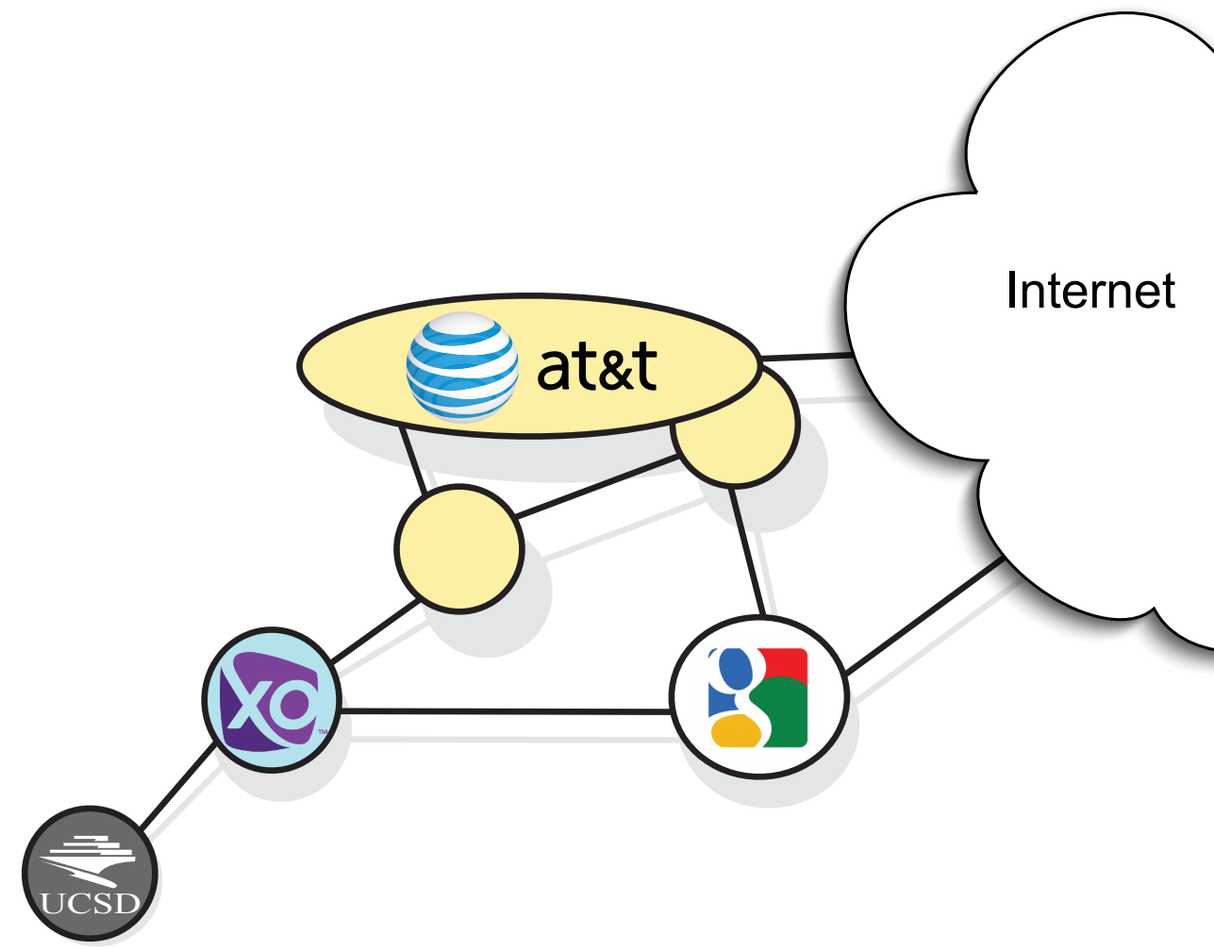
what is an AS?

An AS can roughly be thought of as a single organization.

Some companies use multiple ASes.

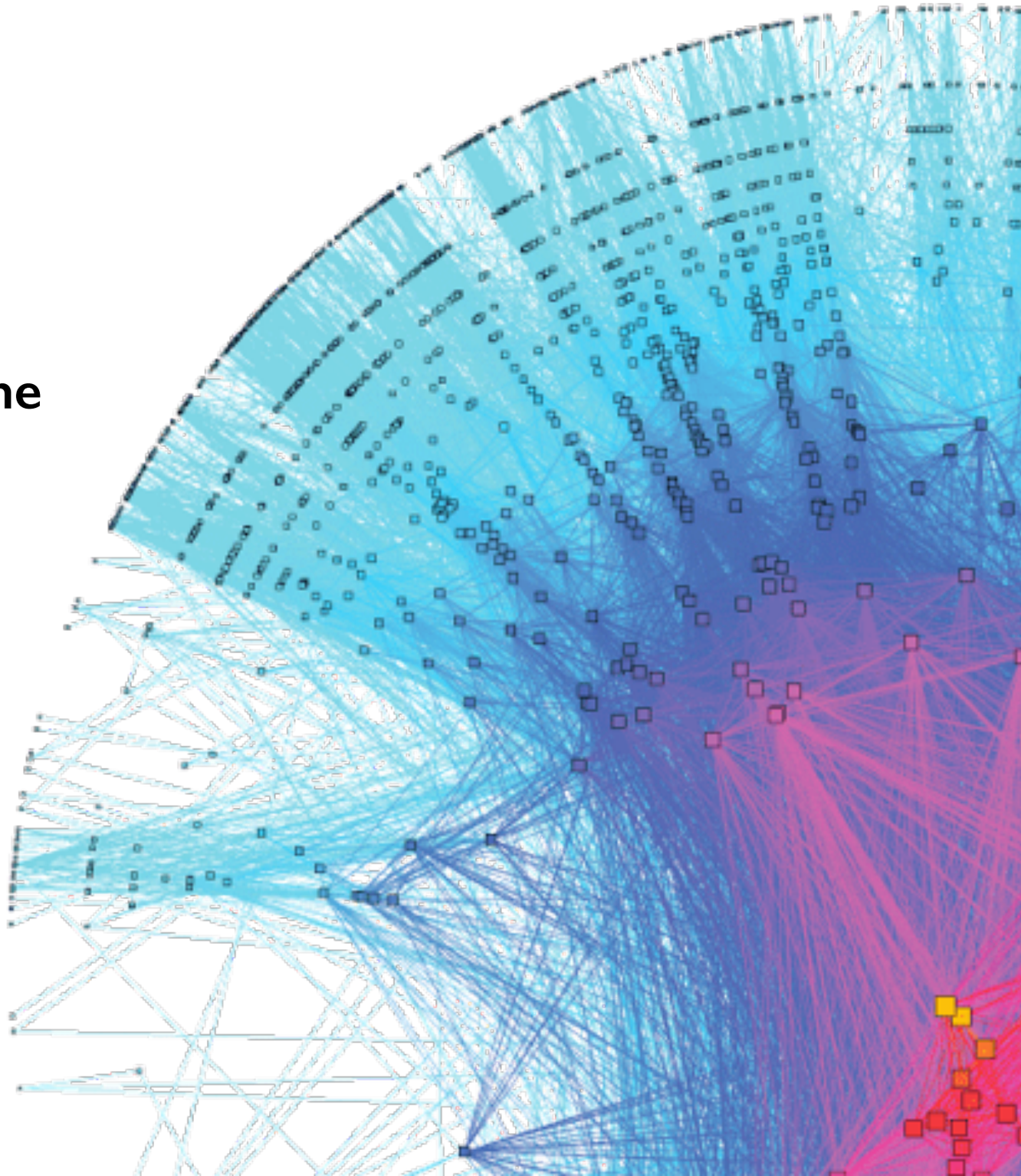
AS topology (observed January 2016)

- 51,000+ ASes
  - geographic location
  - country of registration
- 418,000+ AS links
  - business relationship



AS Core

AS Core focuses on visualizing the geographic scope and size of the ASes as a whole.

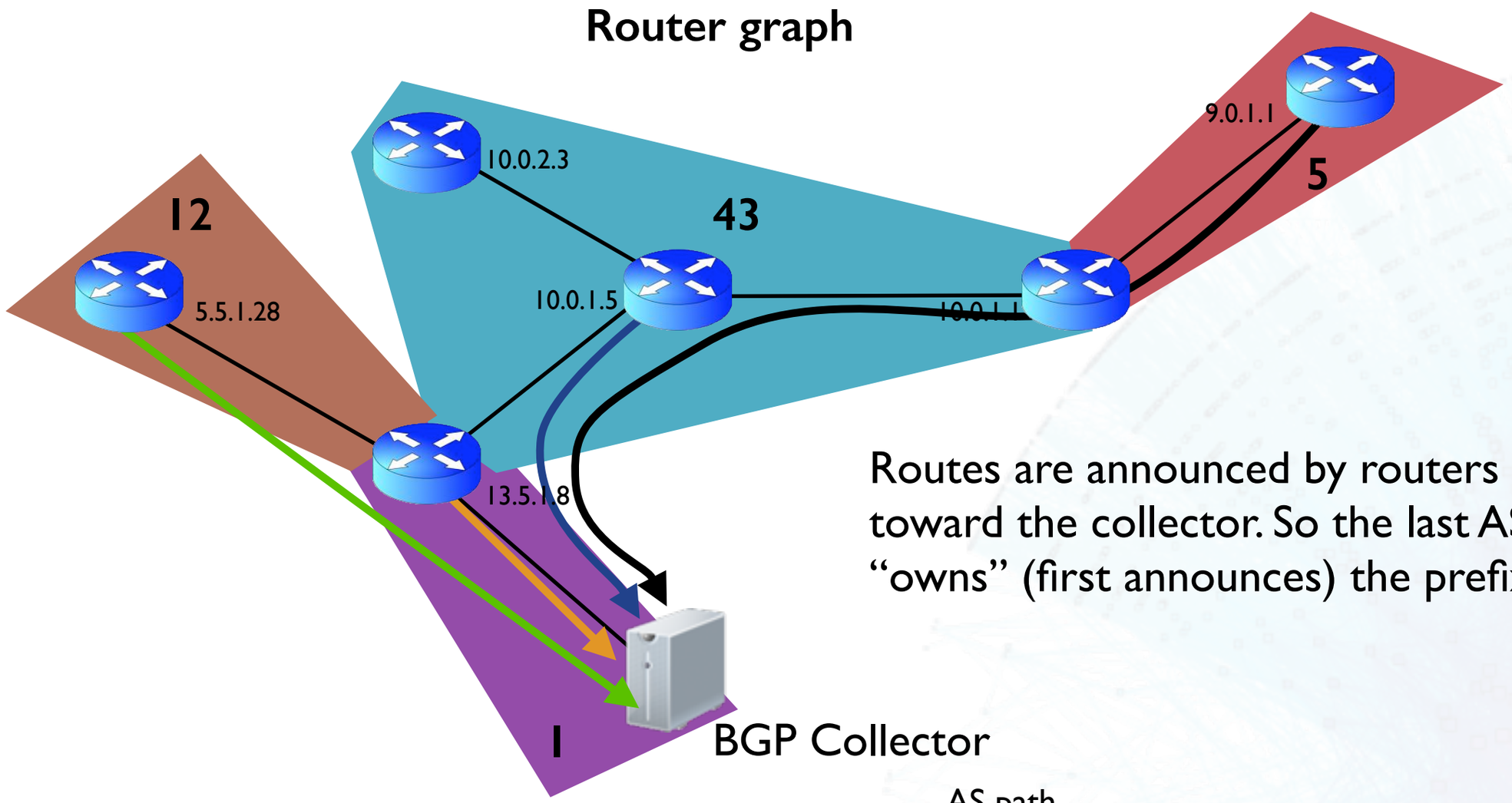


AS Core data

- AS paths
  - 20150101.paths.bz
- AS longitude
  - 20150101.as2loc.txt.bz

BGP Routes

as core



Routes are announced by routers and forwarded toward the collector. So the last AS, the "origin" AS, "owns" (first announces) the prefix.

	prefix	AS path		origin AS
route 1	5.5.1.0/24		1	12
route 2	13.5.1.0/24			1
route 3	10.0.0.0/16		1	43
route 5	9.0.1.0/24	1	43	5

BGP Collectors

as core

Collecting and sharing global routing [Border Gateway Protocol (BGP)] data:

- University of Oregon, Route Views Project
  - http://www.routeviews.org
- RIPE NCC (Regional Internet Registry for Europe/Middle East)
  - http://www.ripe.net/data-tools/stats/ris/ris-raw-data

BGP dump

origin AS

```
TABLE_DUMP2 | 127  1 649600 | B | 157.130.10.233 |   701     | 4.21.103.0/24  |  70 1 3549 46
TABLE_DUMP2 | 127  1 649600 | B | 203.62.252.186 |   122     | 4.21.103.0/24  | 122 1  4637 3549 46
TABLE_DUMP2 | 127  1 649600 | B | 12.0.1.63      |  7018     | 4.21.103.0/24  | 7018 3549 46133
```

source IP | source AS | prefix | AS path

AS path (file)

20150101.paths.bz2

```
# source:topology|BGP|20150104|routeviews|wide
# source:topology|BGP|20150105|routeviews|wide
513|3320|1299|24961|13301
28917|3356|22773|16653
28571|1251|20080|6939|4766|38420
25220|1299|2914|4648|4610|17746
25091|2914|286|8529
202109|1299|12778|49725
9002|9304|17408|131149
14840|6453|4755|45820|22853
```

The top of the file lists the sources used to generate the file.
source:<type>|<source>|<date>|<organization>|<server>

The actual AS path. <source>|<hop0>|<hop1>….|<destination>
The source AS is where the path was observed.

Your view of the topology is limited by the number of sources and origin ASes. Depending on what you are doing with the data, it often helps to filter paths to only sources that have all or almost all origin AS.

degree types

as core

Based on the set of AS paths, different types of degrees may be used.

paths: A → B
B → C
E → B → D

relative to B

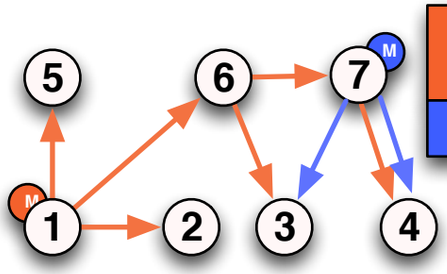
- global degree: A, C, E, D
  - all ASes that connect to the given AS
- out degree: C, D
  - ASes that followed after the given AS in at least one path
- transit degree: E, D
  - ASes that followed after the given AS in at least one path

AS degree types

as core

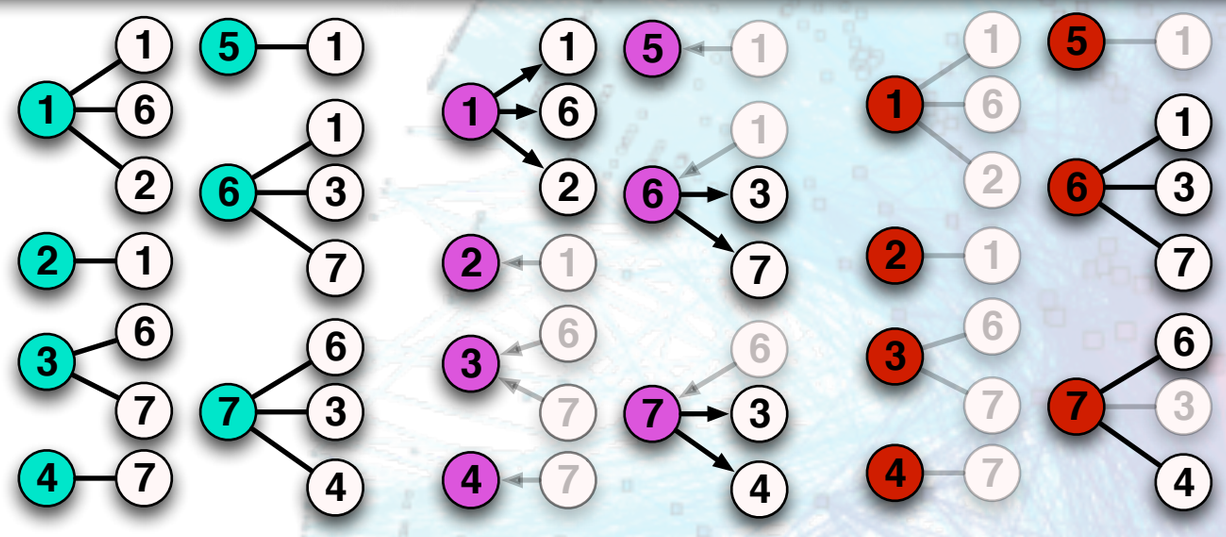
Based on the set of monitors, you will observe different types of paths and so a different set of degrees.

monitors	undirected paths		directed paths		triplets
M1	1-2 1-6-3	1-5 1-6-7-4	1→2 1→6→3	1→5 1→6→7→4	1-6-3 1-6-7 6-7-4
M7	7-3	7-4	7→3	7→4	



degrees

AS	global	out	transit
1	3	3	0
2	1	0	0
3	2	0	0
4	1	0	0
5	2	0	0
6	3	2	3
7	3	2	2

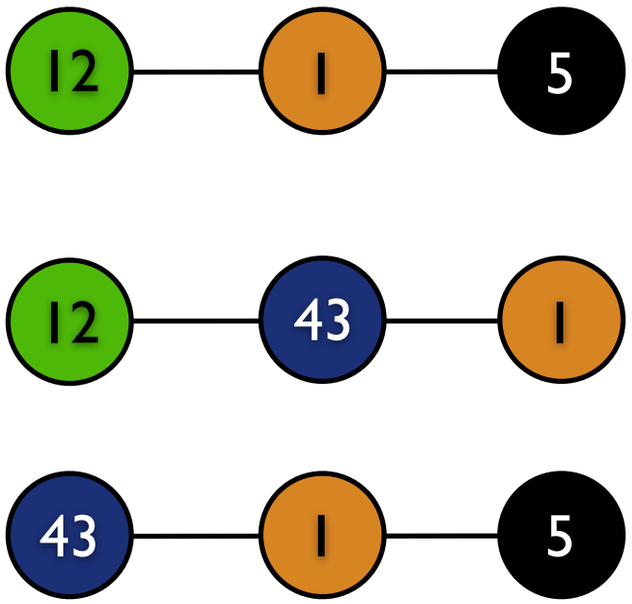


global degree out degree transit degree

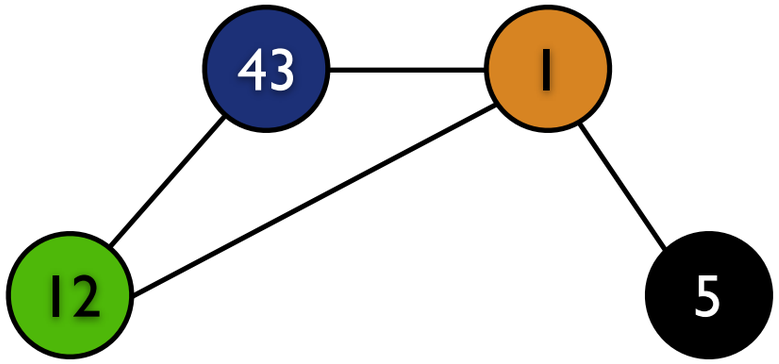
AS paths

data process

paths



Graph



AS	name	longitude	neighbors	transit degree
1			5, 43	3
5			1	0
12			43	0
43			1, 12	2

AS geolocation

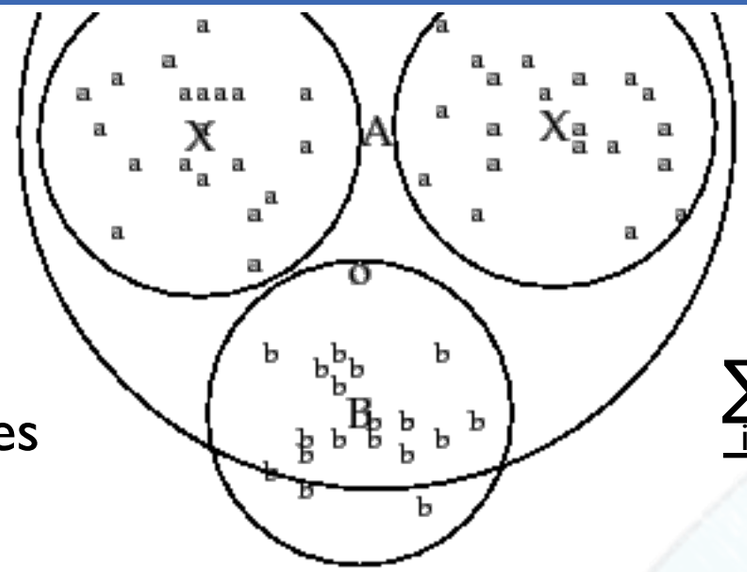
- Geolocation
  identification of real-world geographic location of Internet identifiers
- Digital Envoy's Netacuity commercial geolocation server
- MaxMind GeoLite is a free service
  http://www.maxmind.com/app/geoip_country

Netacuity geographic dump

IP first	IP last	country state city	latitude longtiude
192.172.226.0	192.127.226.255	usa ca la jolla	32.855 -117.249
137.164.23.0	137.164.23.255	usa ca tustin	33.736 -117.823
137.164.46.0	137.164.46.255	usa ca los angeles	33.973 -118.248
74.125.49.0	74.125.49.255	usa il chicago	41.886 -87.623

geolocation to longitude

as core



An AS's longitude equals the weighted average of the longitudes assigned by Netacuity to the addresses it announces.

$$\frac{\sum_i block_i.longitude * block_i.size}{\sum_i block_i.size}$$

origin AS	prefix	geolocation blocks: IP block	geolocation blocks: longitude	weighted average longitude
12	5.5.1.0/24	5.5.1.0 - 5.5.1.255	-103	-103
1	10.0.0.0/16	10.0.0.0-10.0.127.255	25	37.5
		10.0.128.0-10.0.255.255	50	
43	13.5.1.0/24	13.5.1.0-13.5.1.255	-23	-23
5	9.0.1.0/24	9.0.1.0-9.0.1.255	45	45

AS centroid (file)

20150101.as2loc.txt

```
#format:as|continent|country|region|city|lat|long|total-prefixes
1|||||36.023026628203|-88.3649892886005|55
2|||||39.9882562979862|-75.4179414754833|2
3|||||42.3643803602359|-71.1006790537166|33
```

number of prefixes announced by AS

longitude

latitude

continent, country, state, city
empty in this file

AS number

AS organization (file)

organization name

20150101.as-org2info.txt

```
# format:org_id|changed|org_name|country|source
5LINX-ARIN|20140811|Globalinx|US|ARIN
……
# format:aut|changed|aut_name|org_id|source
46191|20141120|5LINX-ASN|5LINX-ARIN|ARIN
……
```

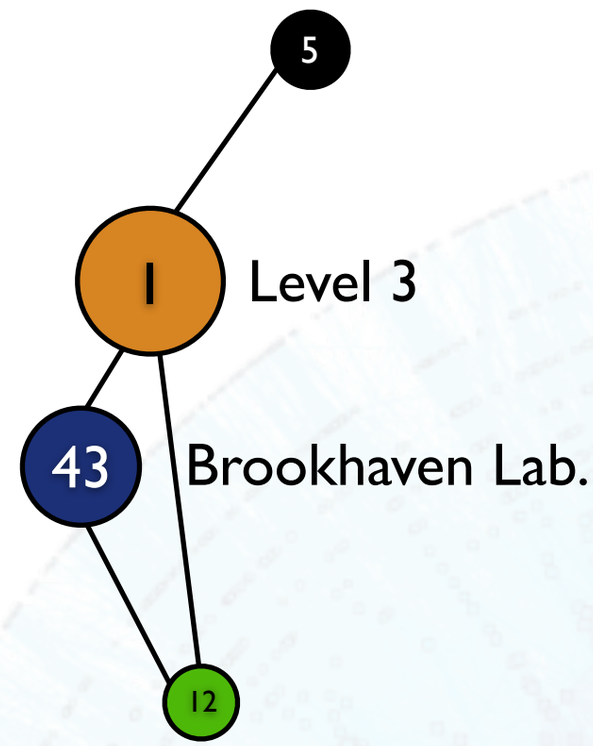
org-id

AS number

This file contains both organization and ASes. The type changes at every line that starts with "# format:<type>". Connect between entries with org-id. In this example the AS 46191 belongs to organization Globalinx.

bring it all together

as core



We now have everything we need to build the graph.

AS	name	longitude	neighbors	transit degree
1	Level 3	37.5	5, 43	3
5	Symbolics, Inc.	45	1	0
12	New York University	-103	43	0
43	Brookhaven Laboratory	-23	1, 12	2

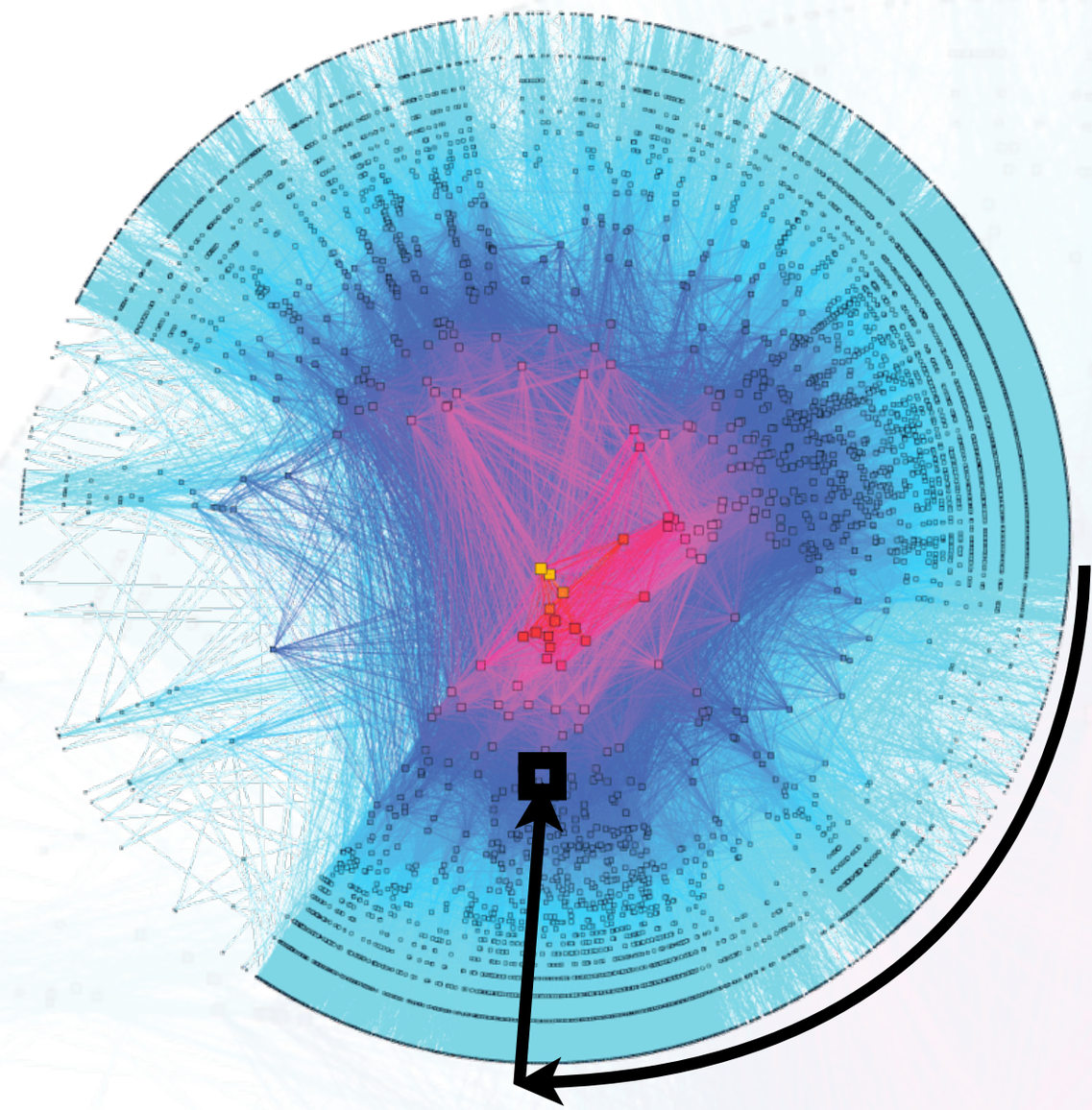
node location

as core

Each node represents a single AS.

ASes with nearby/same degree and longitude will overlap.

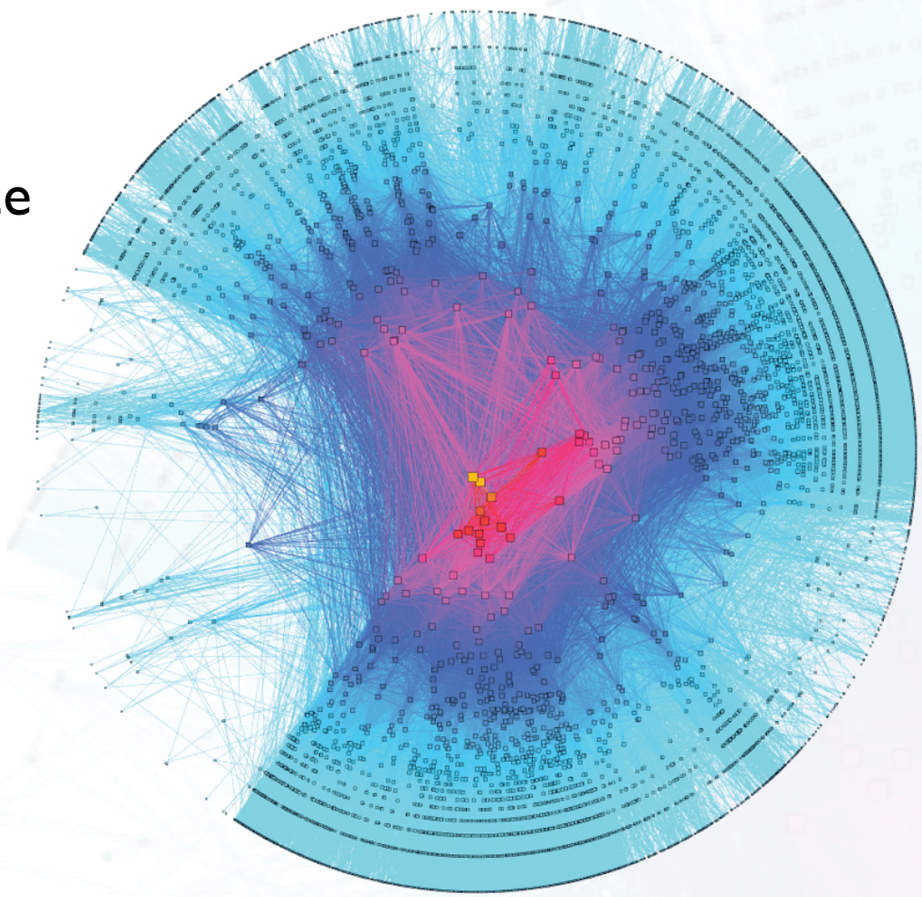
node: color/radius/size	$1 - \log\left(\frac{\text{degree (AS)} + 1}{\text{maximum.degree} + 1}\right)$
node: angle	longitude of the AS's BGP prefixes
link: color	node color with smallest degree

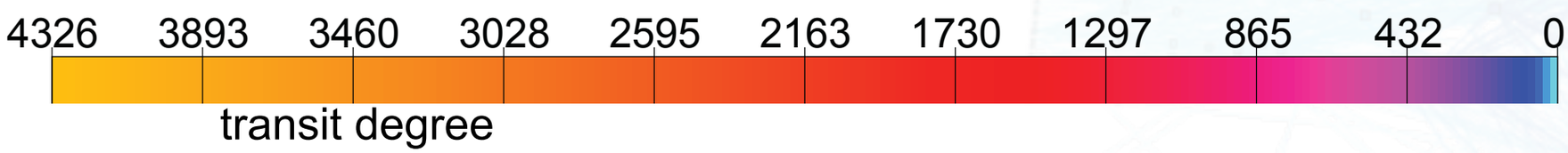


node color

as core

ASes are colored based on their transit degree. Since transit degree is a long tailed distribution, it is best to use a logarithmic color scheme.

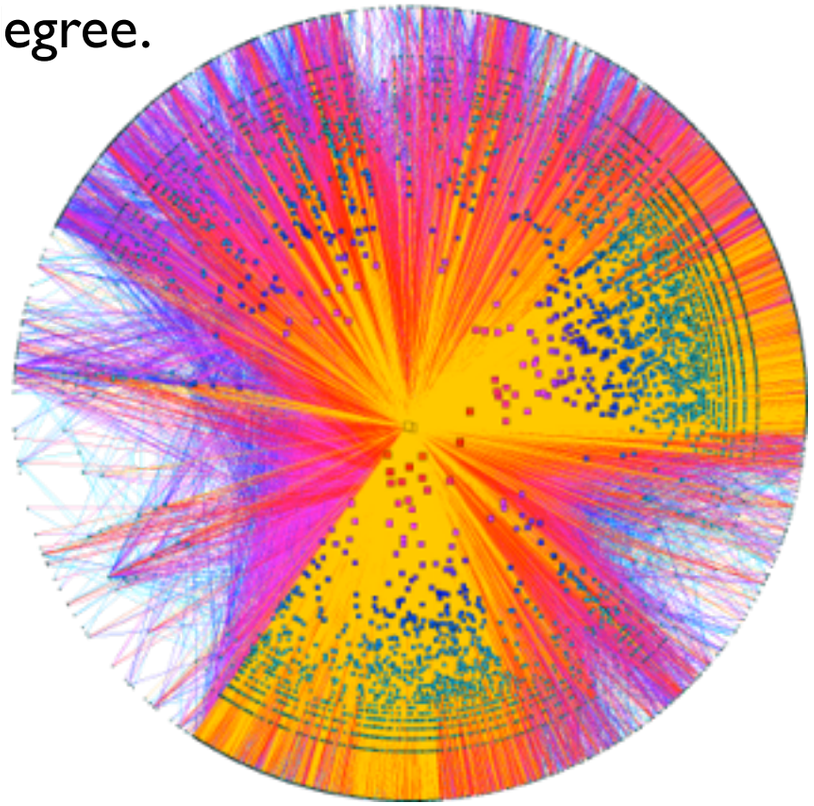




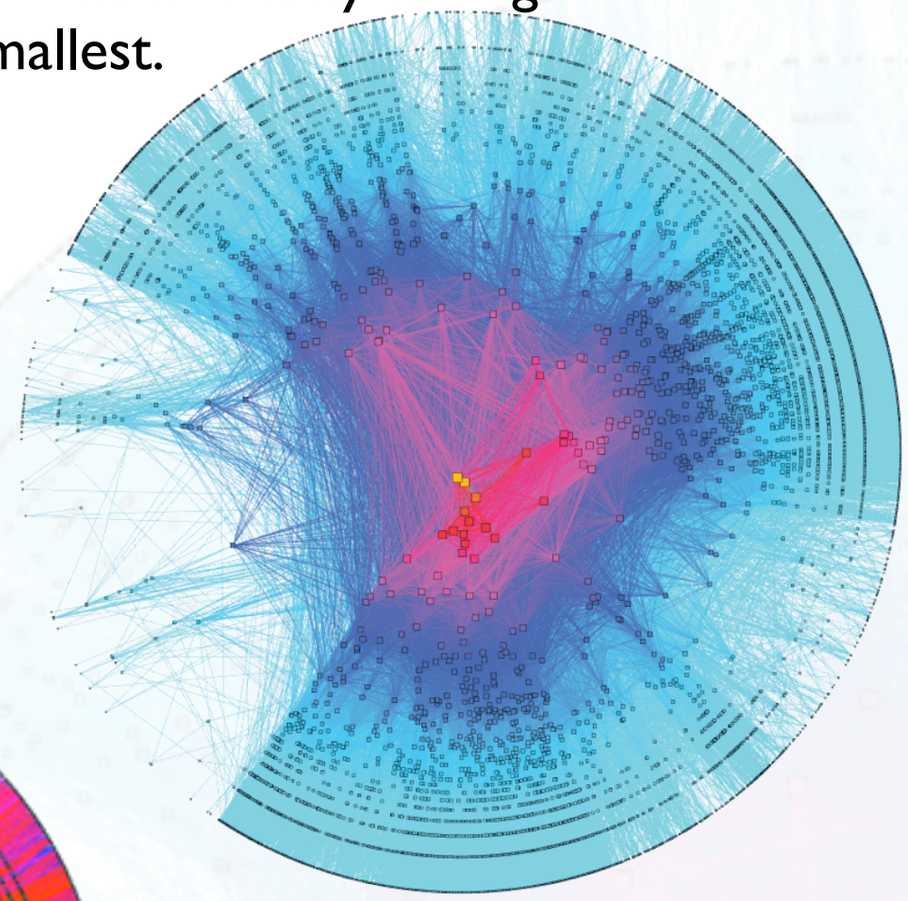
link color

as core

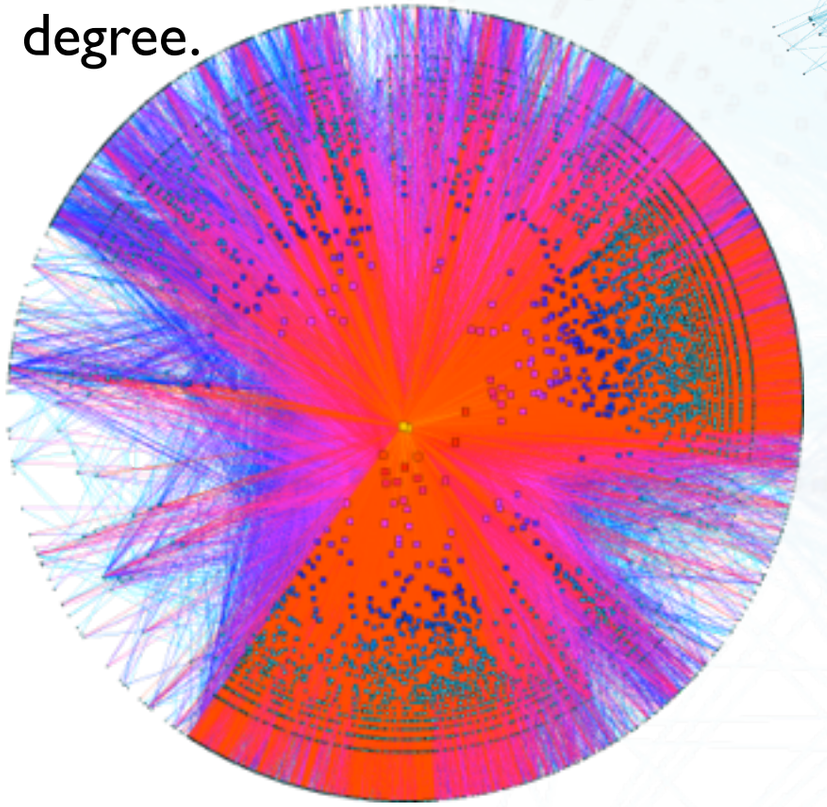
Link color set by largest degree.



Link color set by average smallest.



Link color set by average degree.

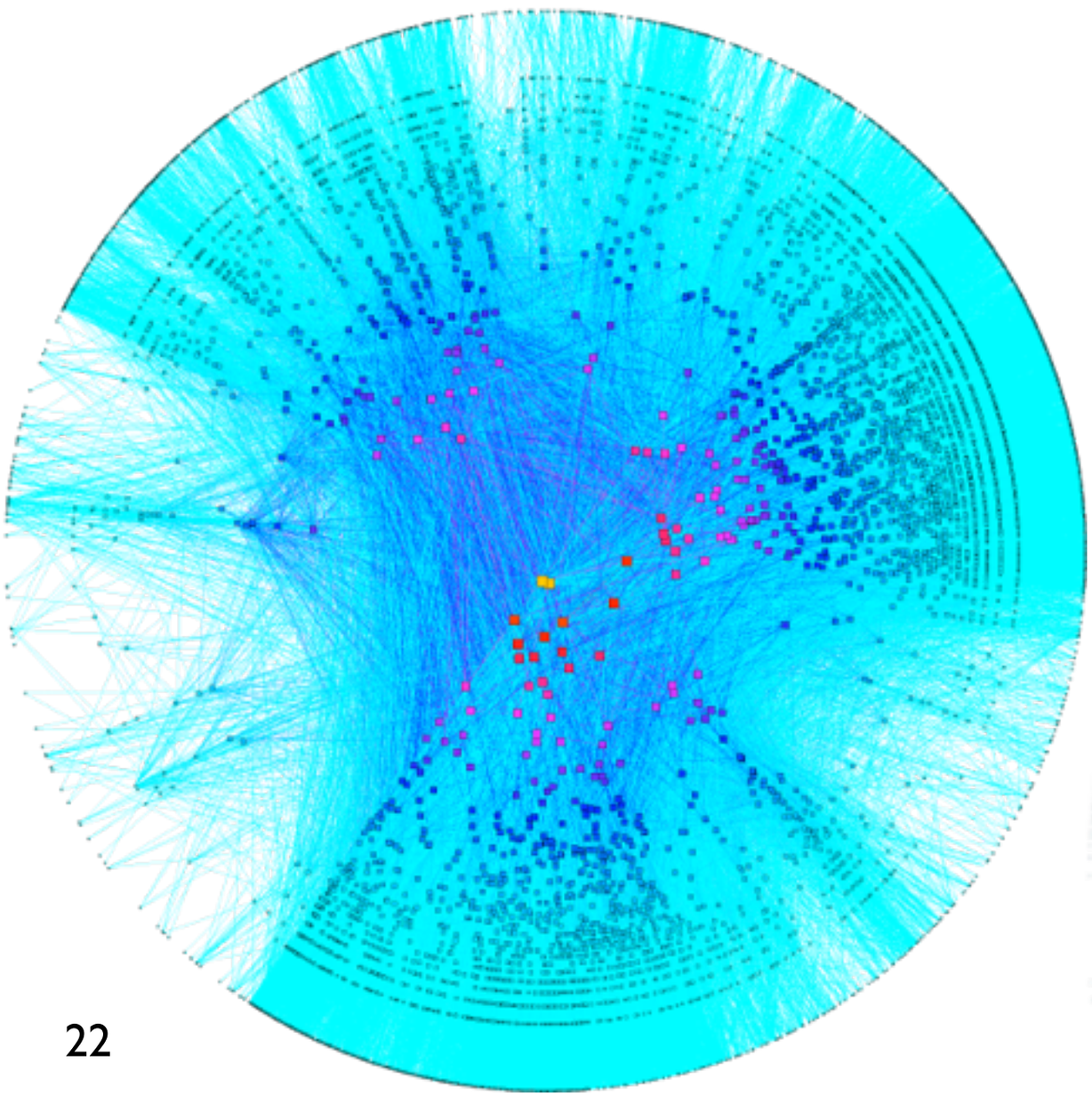


draw order

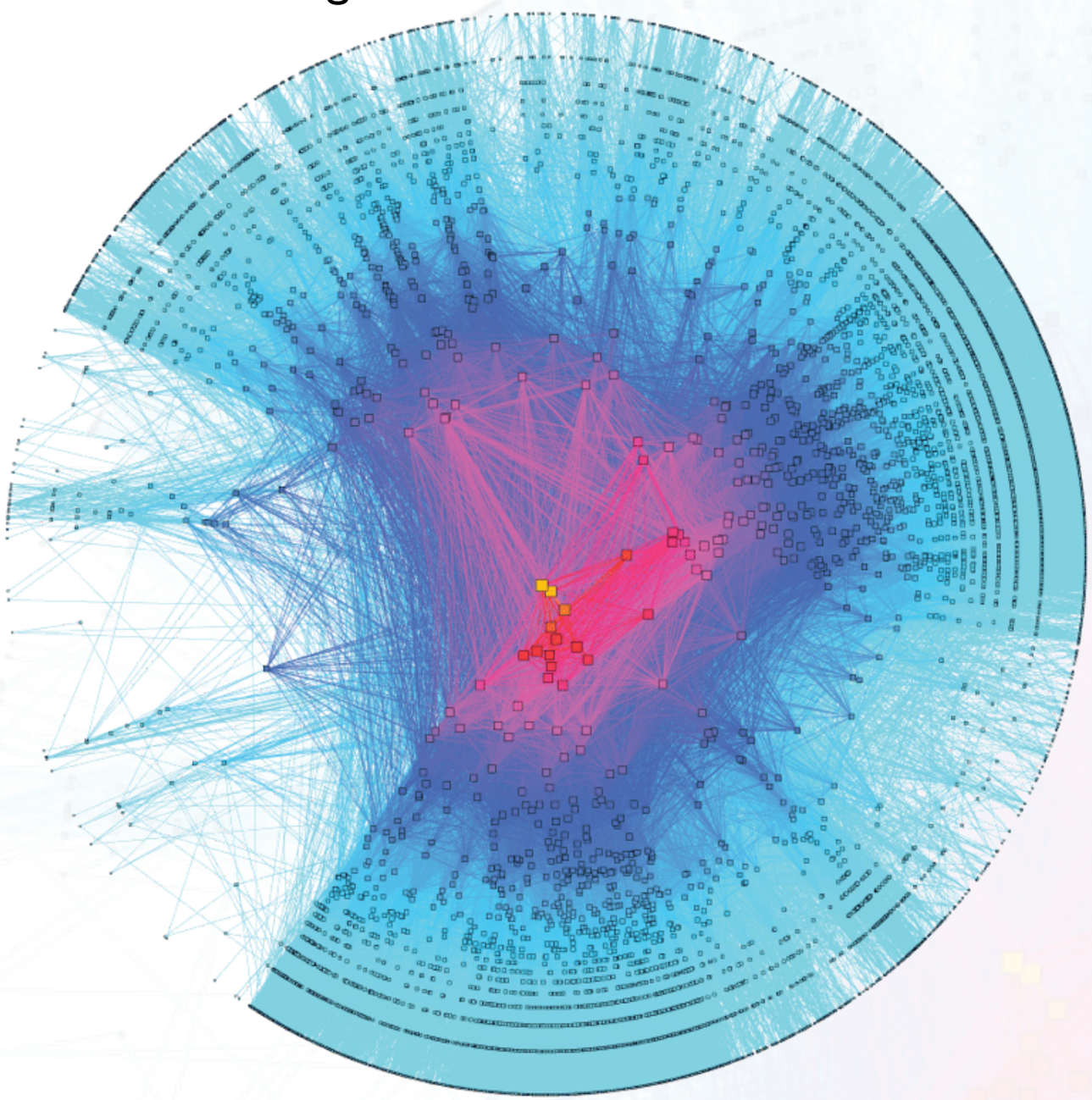
as core

nodes draw over links

unsorted link draw order

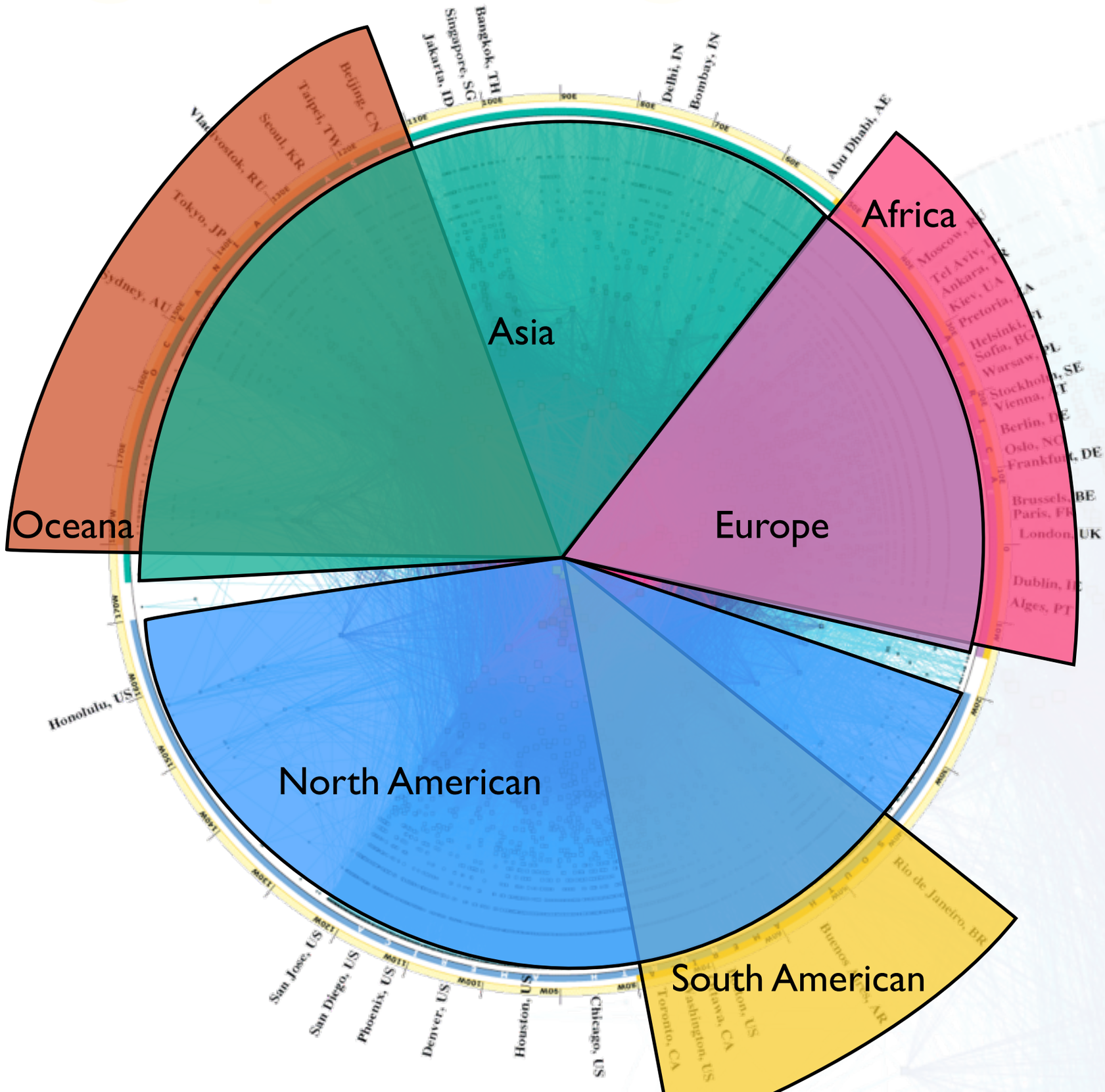


draw largest link last



geographic regions

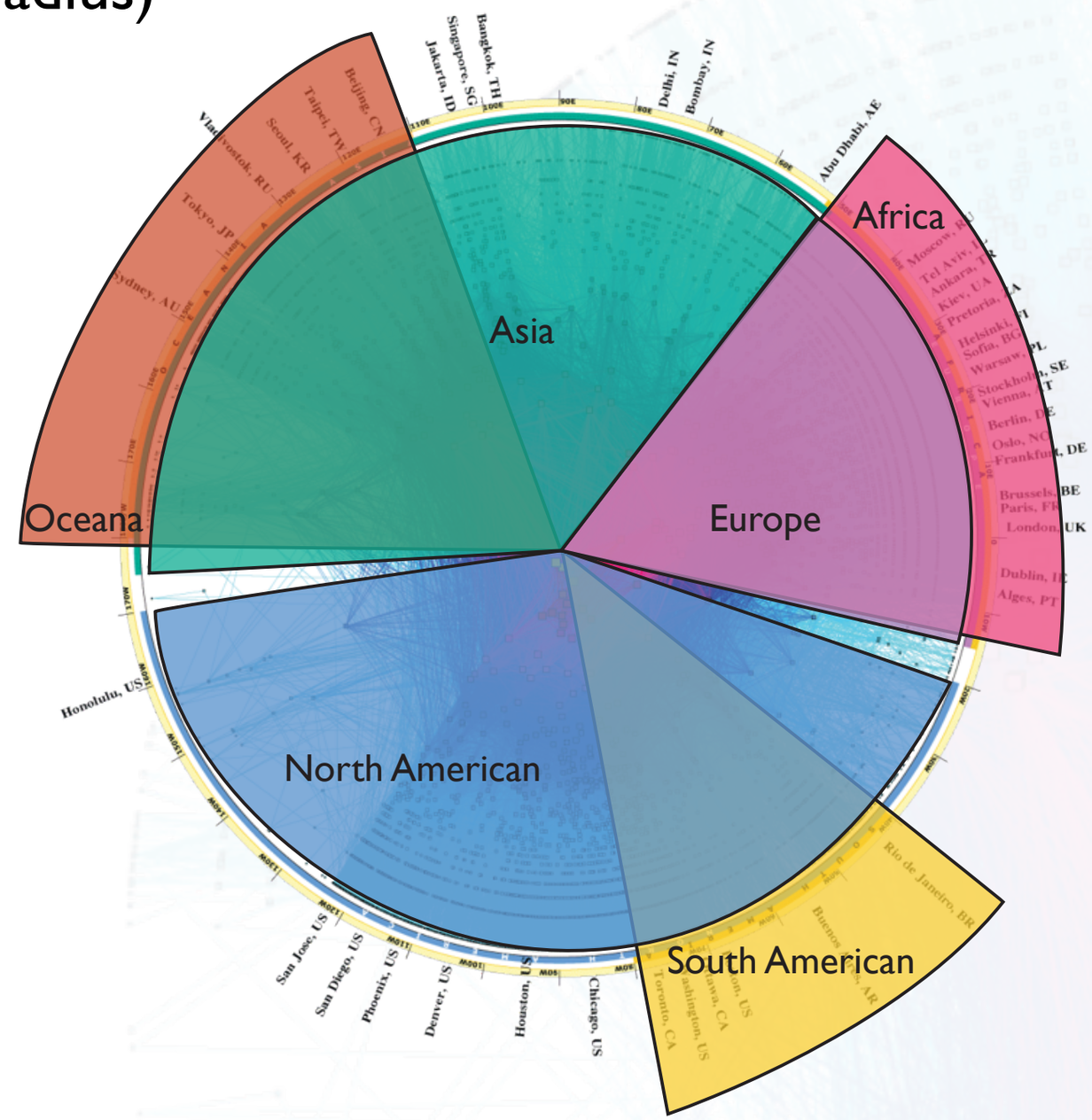
as core



limitations

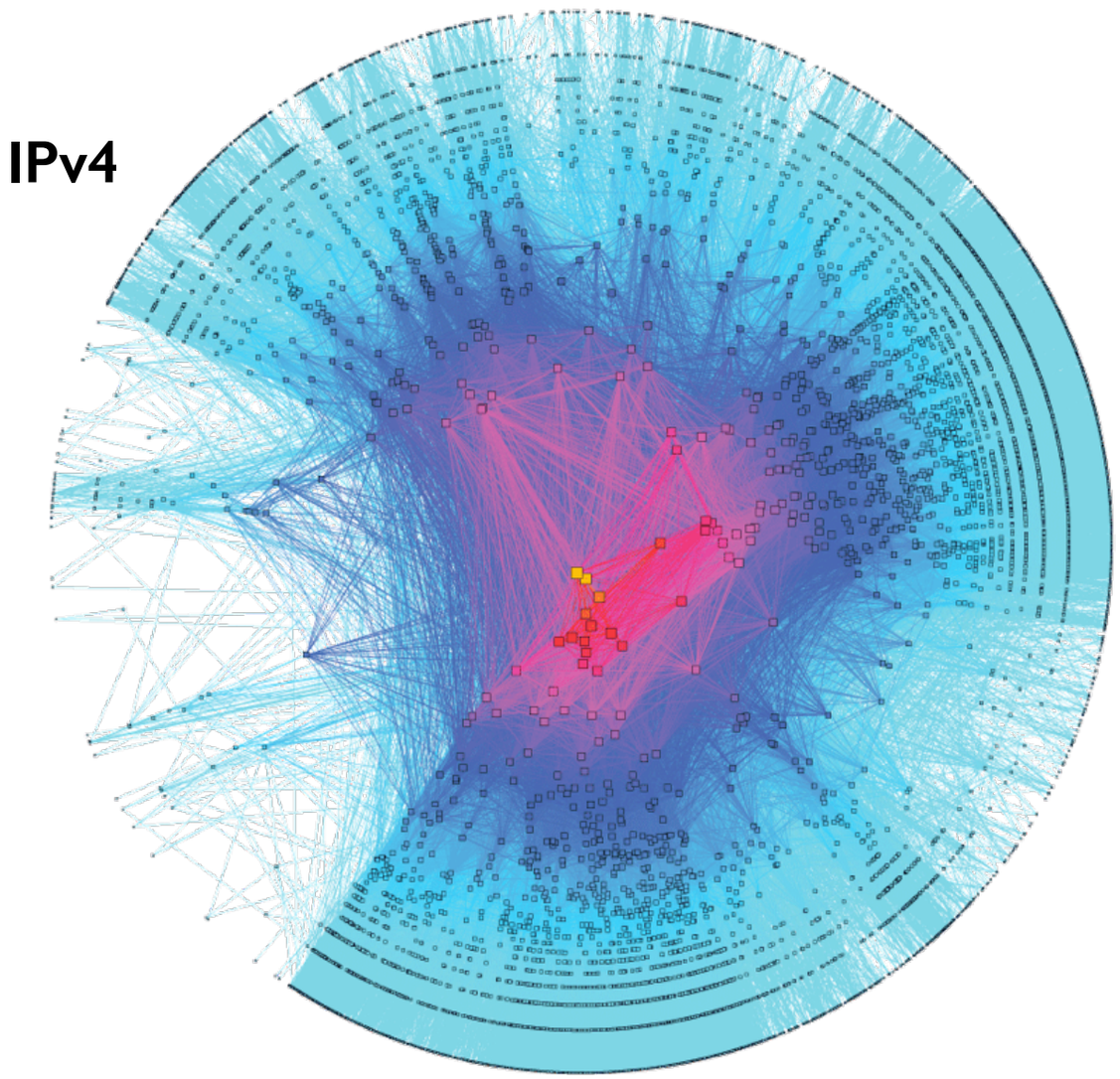
as core

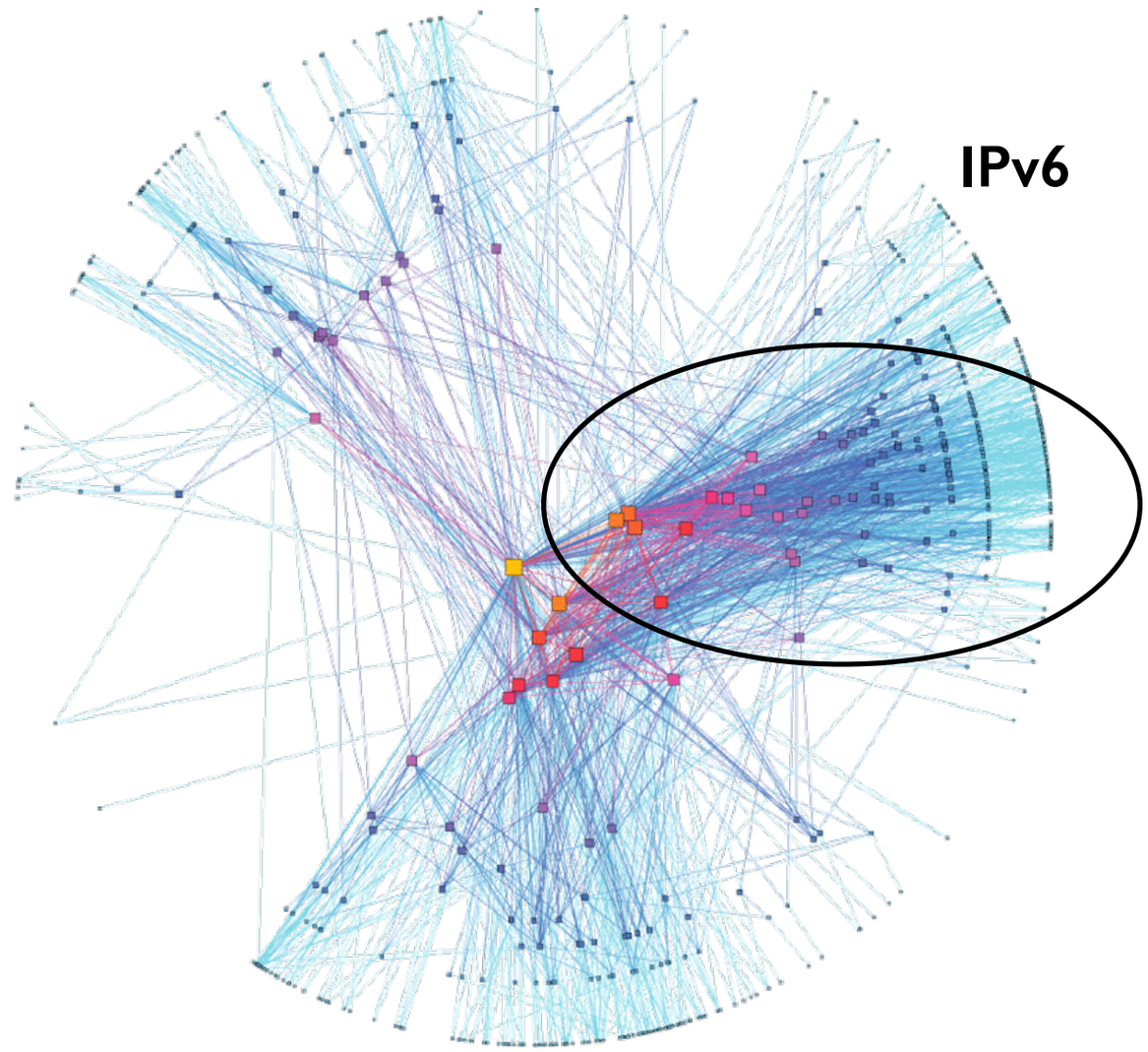
- degree encoded twice (size/radius)
  - could encode additional metrics
- latitude node encode
  - overlap between continents



IPv4 vs IPv6 graphs

IPv4 vs IPv6



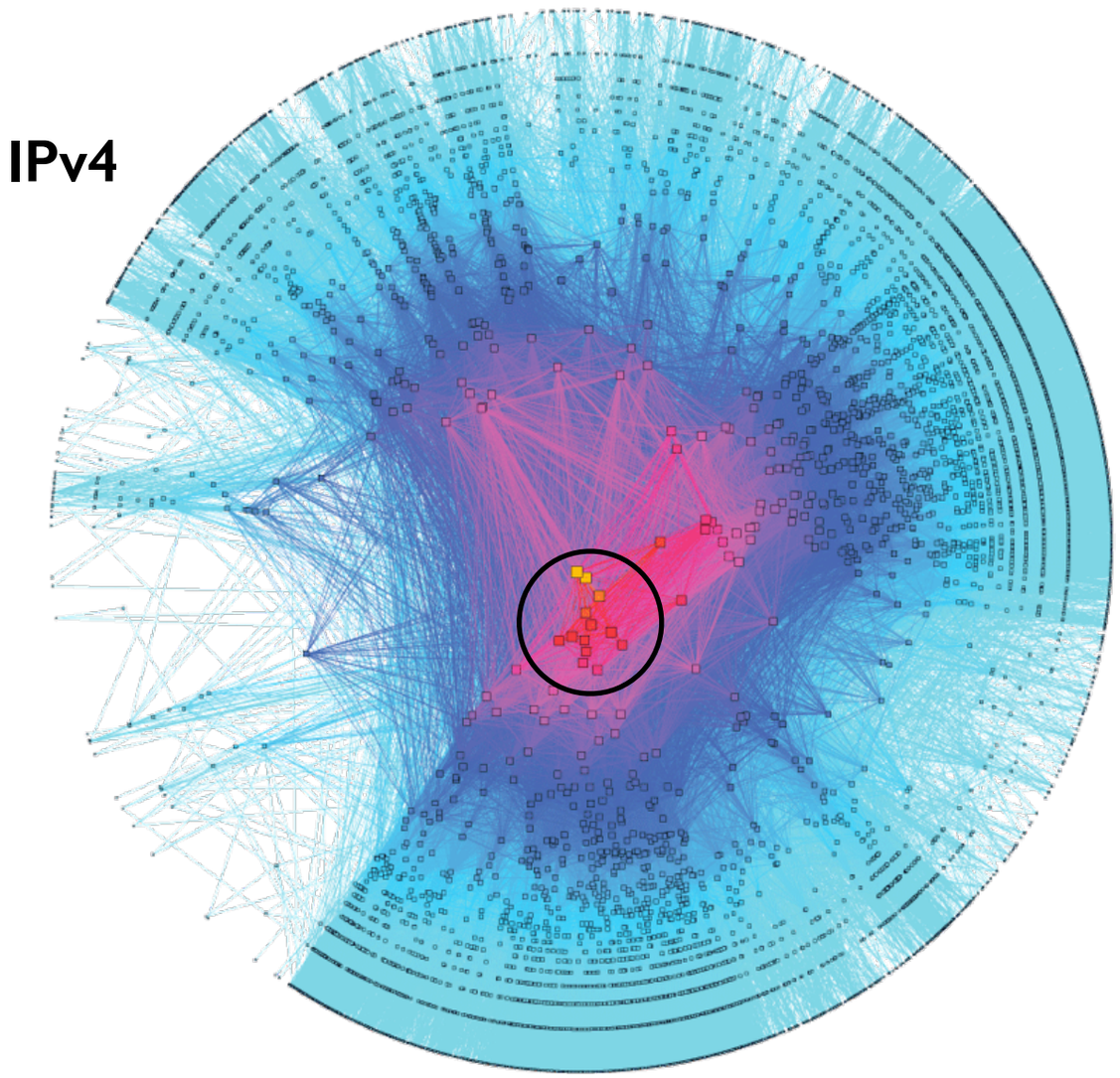


IPv4 high density in Asia, America, and Europe

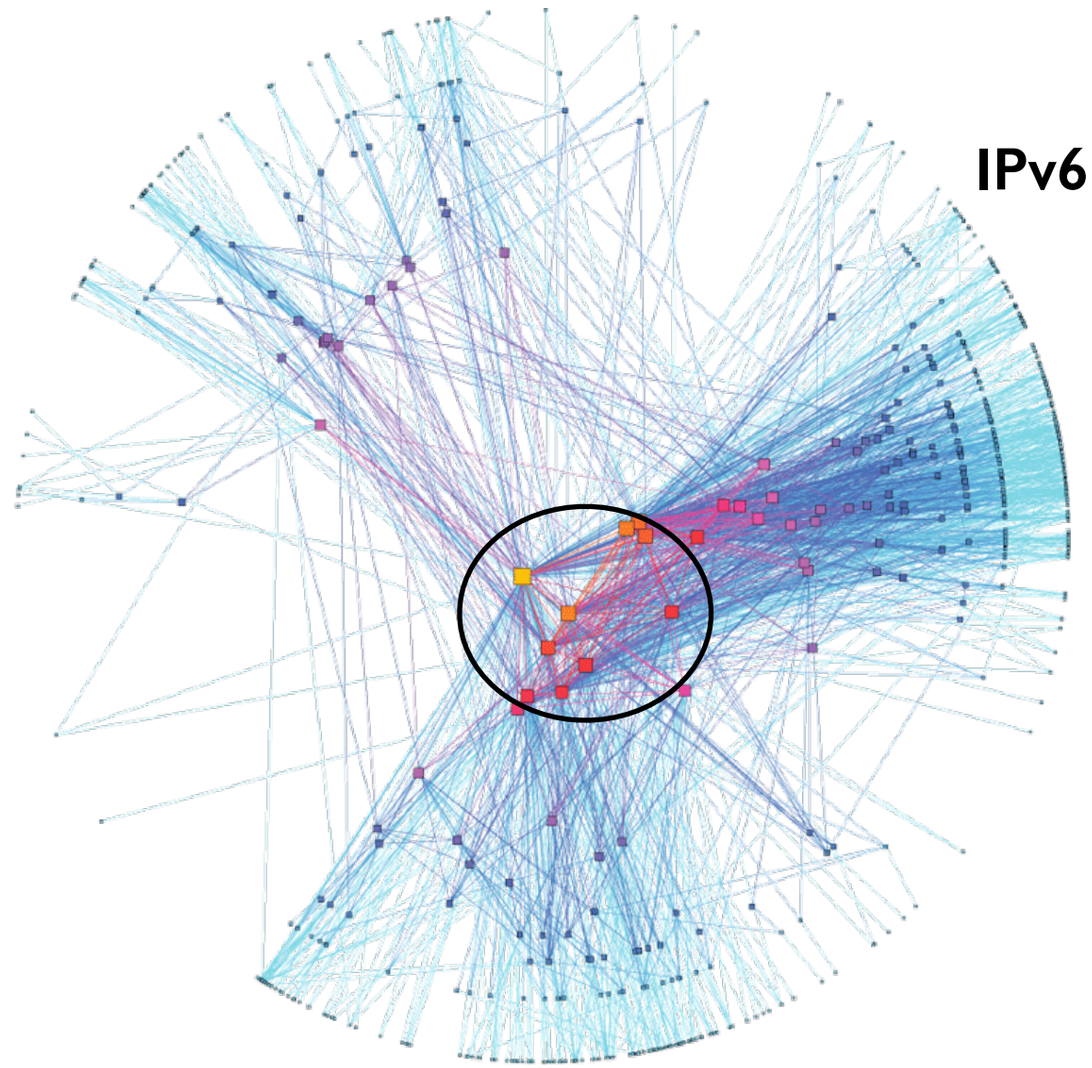
IPv6 highest area of density in Europe

IPv4 vs IPv6 cores

IPv4 vs IPv6



IPv4 core primarily in North America

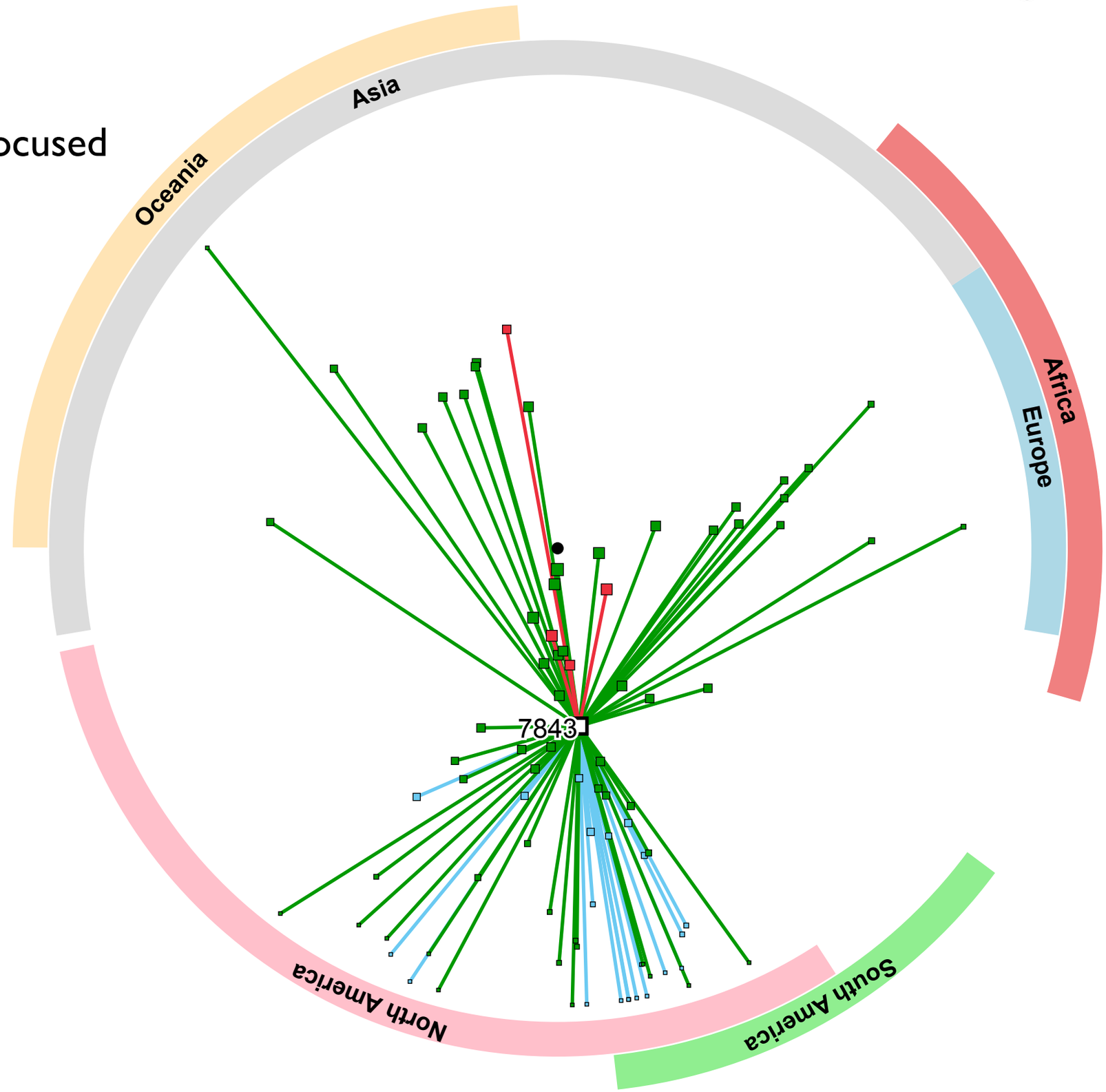


IPv6 core spread between America and Europe

AS Core single AS

as core single

Showing the same placement, but focused on a single AS.



AS core single AS data

as core single

- AS customer cone
  - 20150101.ppdc-ases.bz
- AS longitude
  - 20150101.as2loc.txt.bz
- AS relationships
  - 20150101.as-rel.txt.bz

AS Relationships

as core single

provider

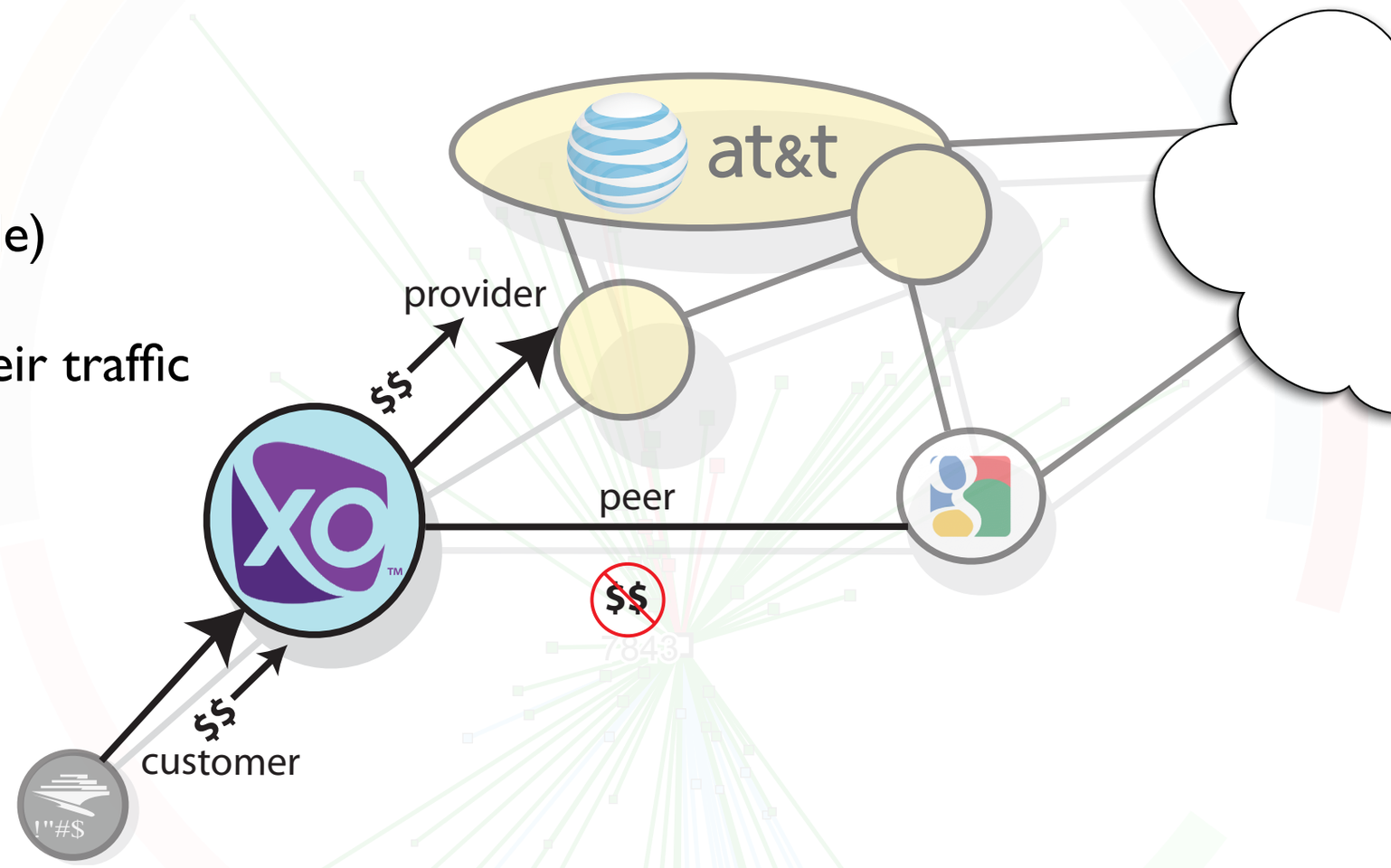
- you pay them to transit your traffic (XO pays AT&T)

peer

- unpaid exchange of traffic (between XO and Google)

customer

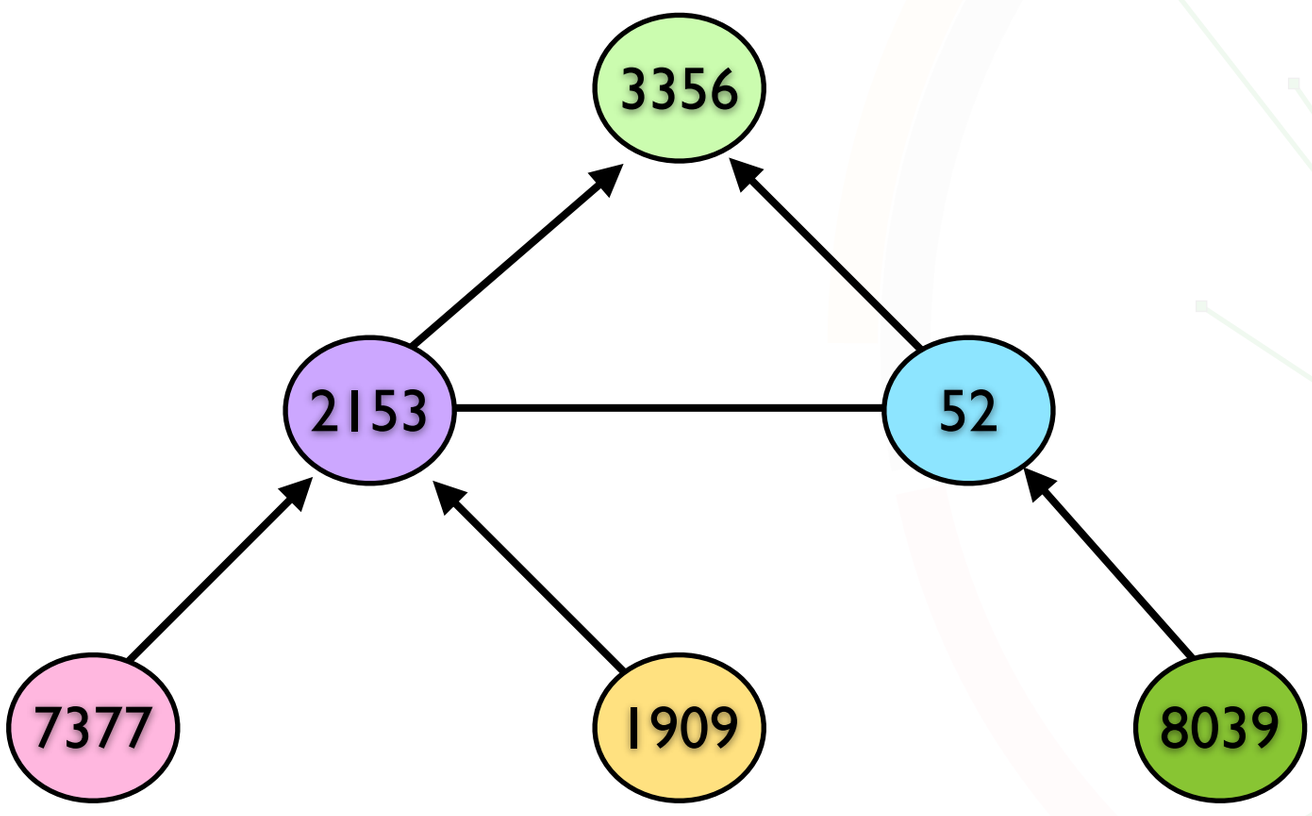
- they pay you to transit their traffic (UCSD pays XO)

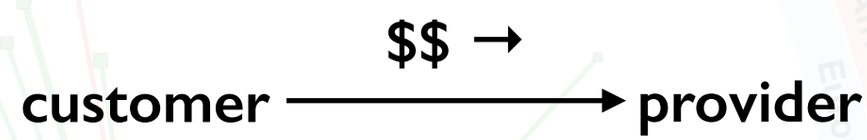


AS hierarchy

as core single

ASes can be organized into a hierarchical structure based on the type of business relationships they form between themselves.





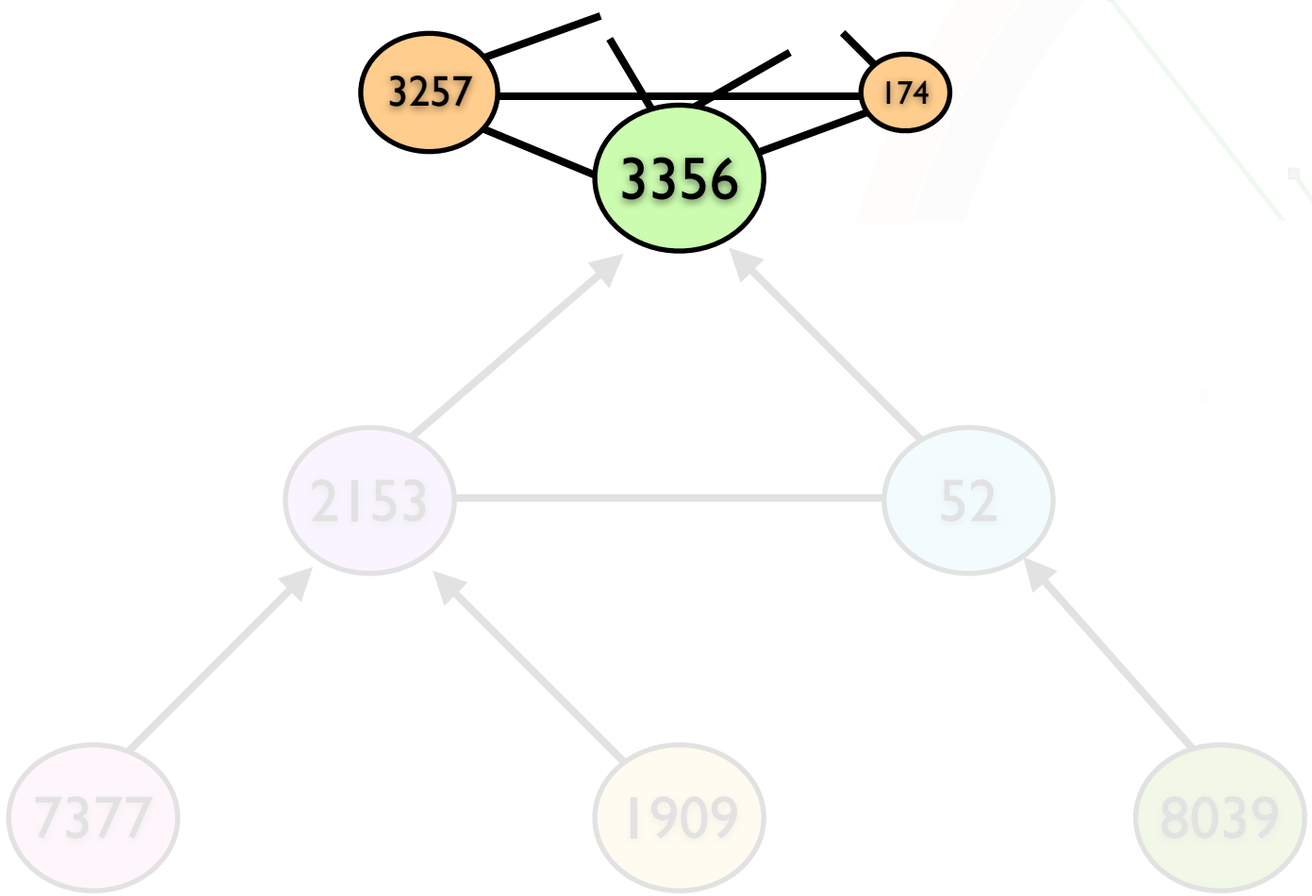
customer pays provider for transit

peers do not pay to accept each other's traffic

AS hierarchy

as core single

ASes at the top of the hierarchy are call Tier-1.
Together they form the backbone of the Internet.



AS Relationship (file)

as core single

inferred Tier 1 ASes

20150101.as-rel.txt.bz2

Internet eXchange Points

```
# inferred clique: 174 209 701 1239 1299 2828 2914 3257 3320 3356 3549 5511
# IXP ASes: 1200 4635 5507 6695 7606 8714 9355 9439 9560 9722 9989 11670 17819
1|11537|0
```

AS0 AS1 type

-1	AS0 is a provider of AS1	AS0 ← AS1
0	AS0 and AS1 are peers	AS0 — AS1
1	AS0 is a customer of AS1	AS0 → AS1

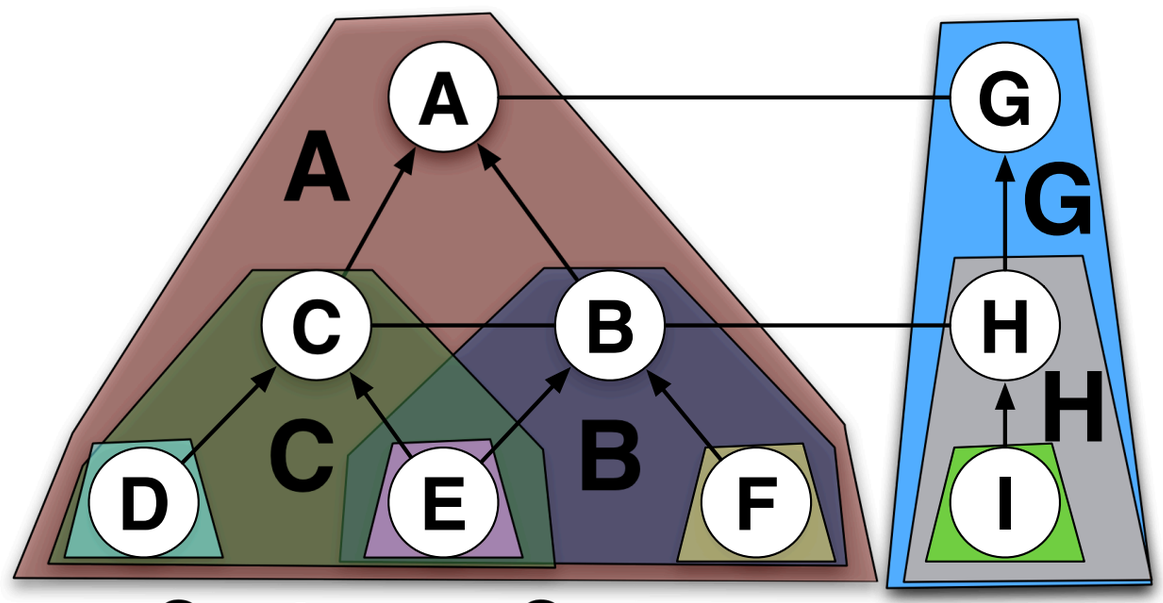
http://www.caida.org/data/as-relationships/

AS Customer Cone

as core single

An AS's customer cone contains the set of ASes we observe the AS announce to its peers or providers. In practice, this is the set of ASes it can reach through its customers.

AS	size
A	6
B	3
D	1



Customer Cones

http://www.caida.org/data/as-relationships/

AS Customer Cone (file)

as core single

20150101.ppdc-ases.txt.bz2

```
2345   2345
14     14 209 701 1239 1299 2828 2914 3257 3320 3356 3549 5511
1200   435 1200 5507 6695 7606 8714 9355 9439 9560 9722
```

AS

customer cone members

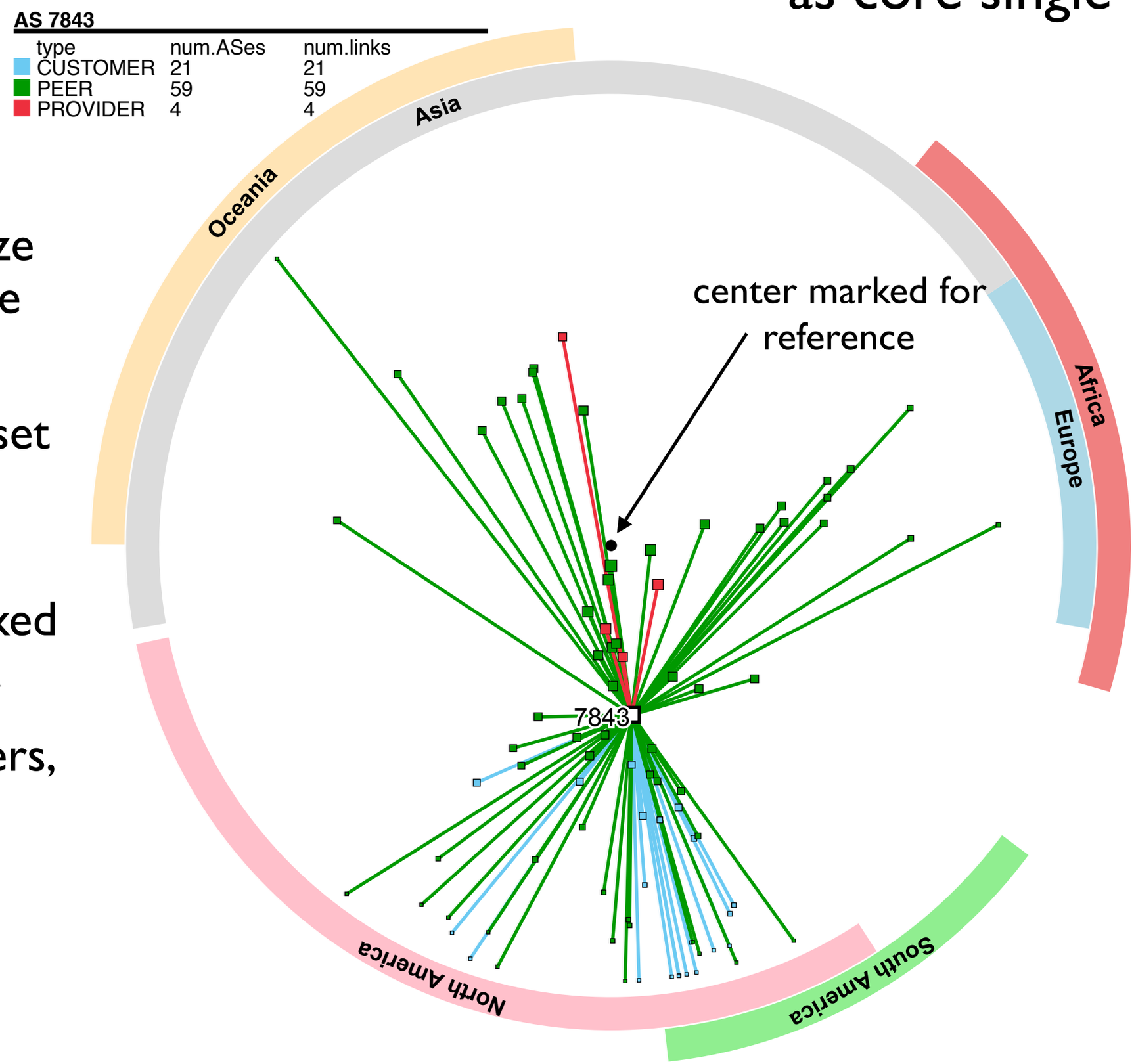
AS	size	members
2345	1	2345
14	12	14 209 701 1239 1299 2828 2914 3257 3320 3356 3549 5511
1	10	435 1200 5507 6695 7606 8714 9355 9439 9560 9722

http://www.caida.org/data/as-relationships/

AS core single AS

as core single

- node location and size set by customer cone size
- node and link color set by relationship to selected AS
- selected AS size is fixed regardless of degree.
- draw order: customers, peers, and providers



AS's Organization (file)

20150101.as-org2info.txt.bz2

org-id

```
# format:org_id|changed|org_name|country|source
RRMA-ARIN|20130425|Time Warner Cable Internet LLC|US|ARIN
..............
# format:aut|changed|aut_name|org_id|source
7017|20110607|SCRR-7015        |RRMA-ARIN|ARIN
7756|20110607|SCRR-7756        |RRMA-ARIN|ARIN
7843|20110607|TWCABLE-BACKBONE|RRMA-ARIN|ARIN
```

AS number

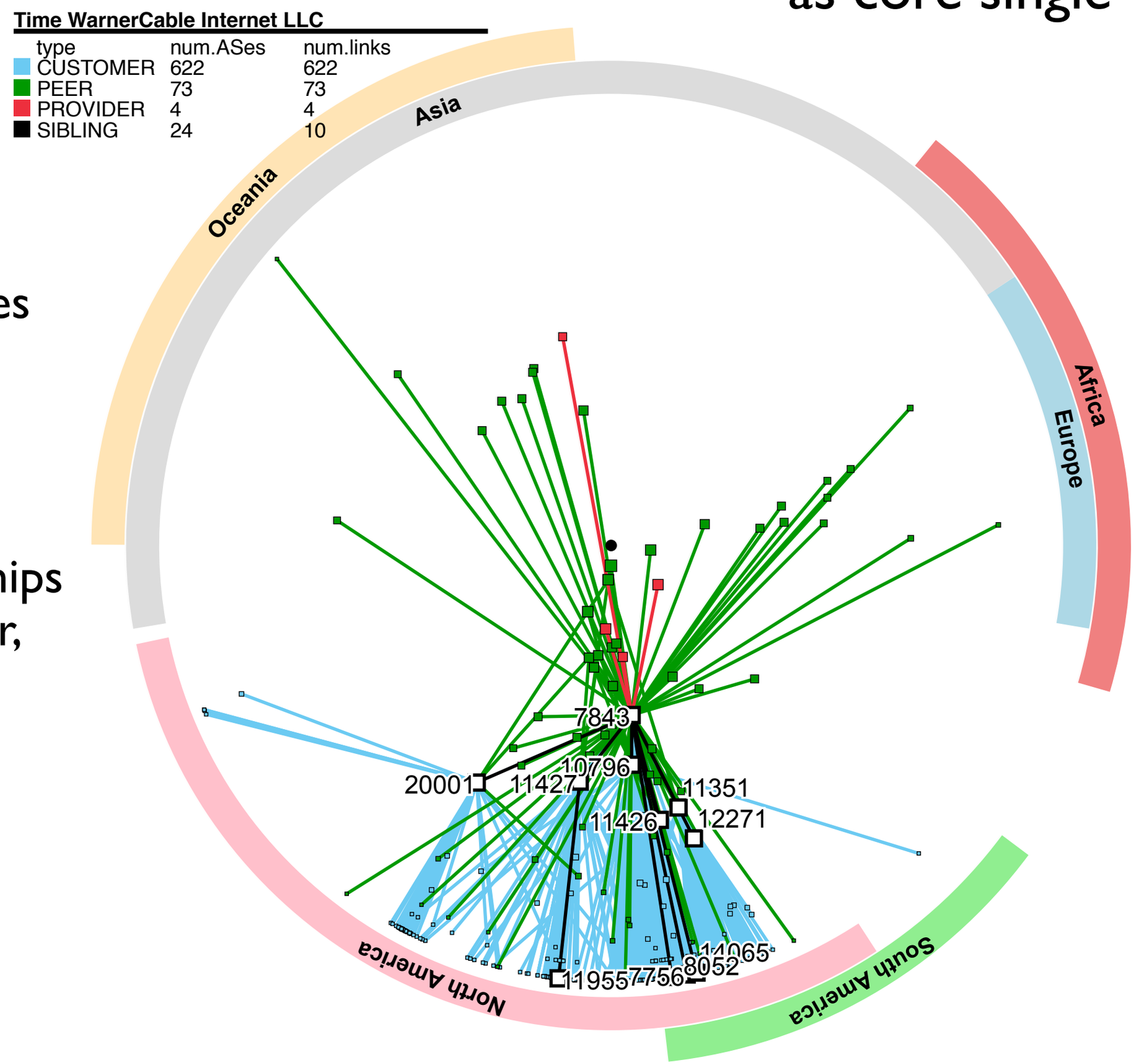
org-id

We can build a list of an organization's ASes by filtering on its org-id.

http://www.caida.org/data/as-organizations/

AS core Organization

as core single



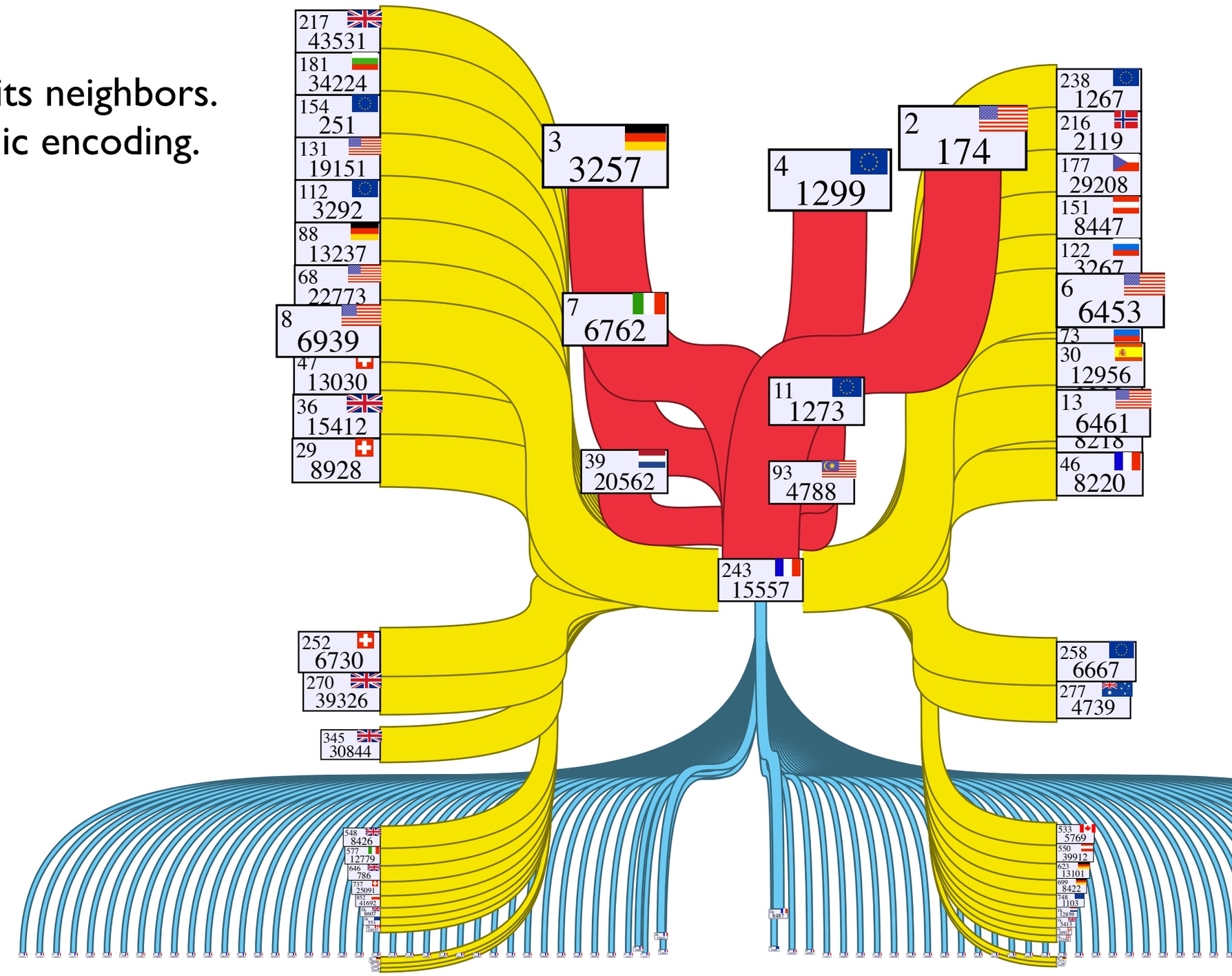
An organization's ASes and their neighbors.

- **black** links connect "**sibling**" ASes
- Conflicting relationships are resolved: provider, peer, customer.

AS relationships

as relationships

Information about an AS and its neighbors.
Shown without the geographic encoding.



data

as relationships

- AS customer cone
  - 20150101.ppdc-ases.bz
- AS country
  - 20150101.as-org2info.txt.bz
- AS relationships
  - 20150101.as-rel.txt.bz

AS ranking

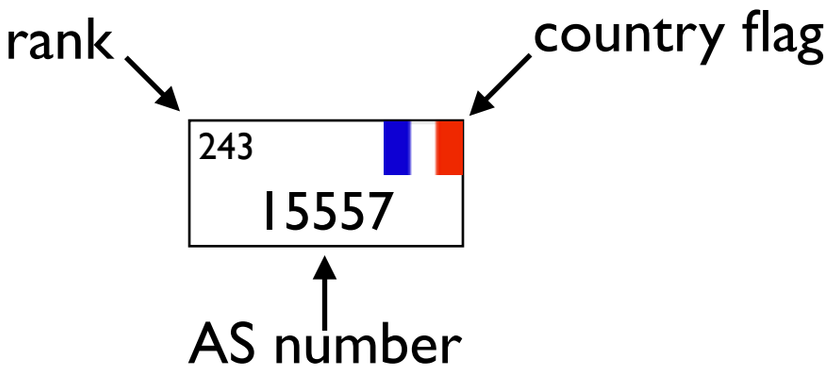
as relationships

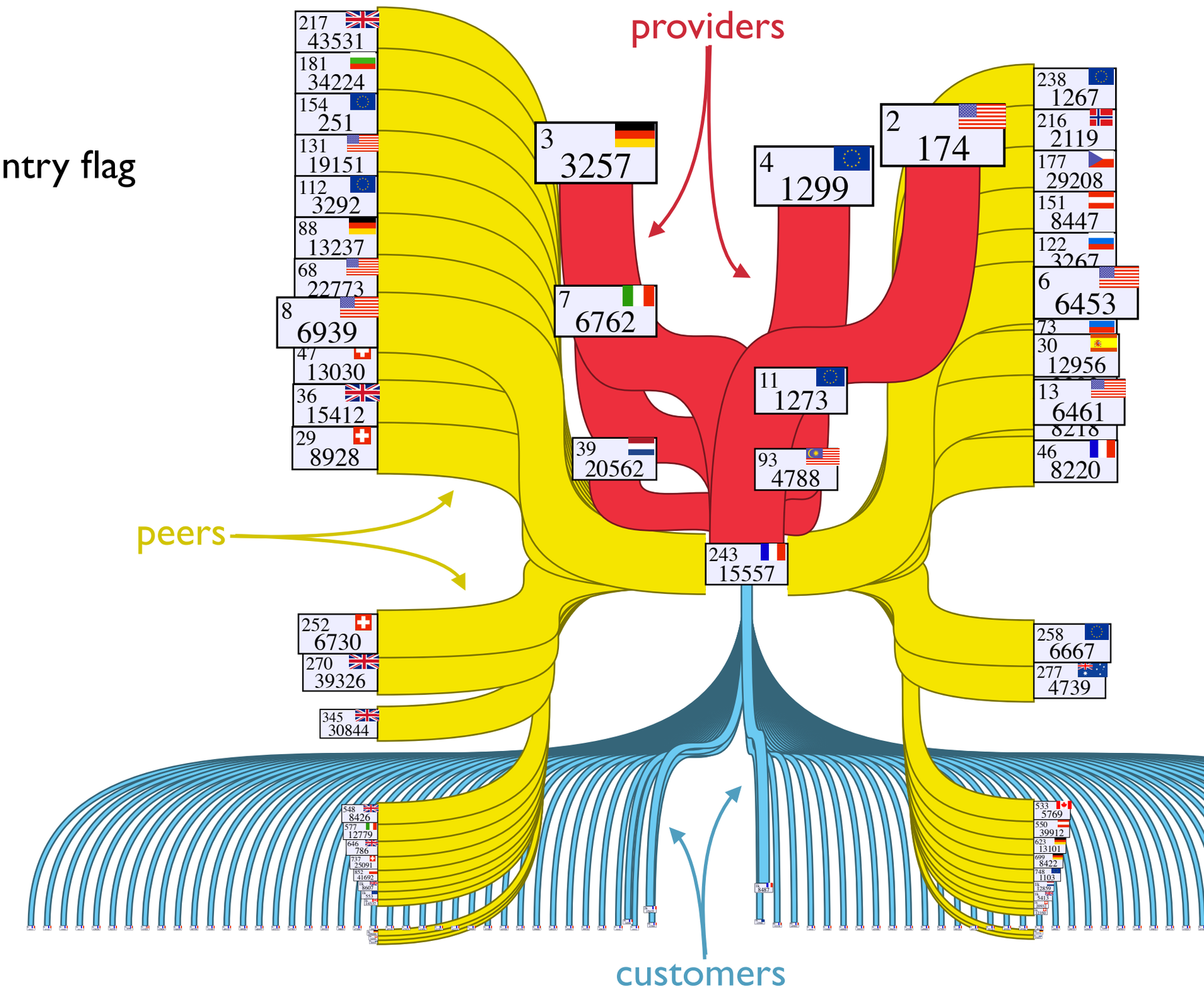
We define an AS's ranking to be equal to the number of ASes with a customer cone size greater then the AS plus 1.
(+1, allows rank to start at 1)

rank	AS	customer cone size
1	336	10
2	42	8
3	876	7
3	91	7
5	23	4

breakdown

as relationships

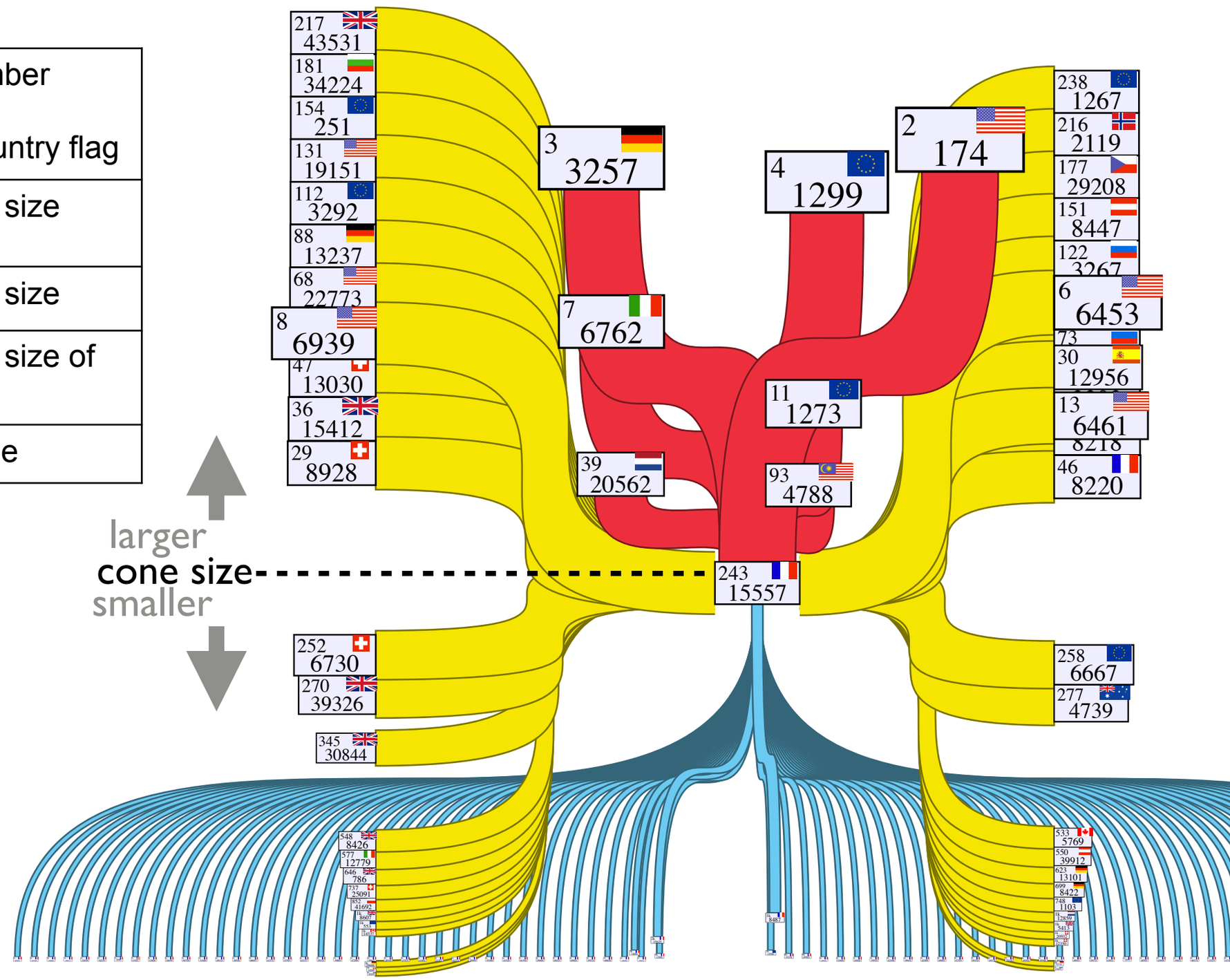




AS placement

as relationships

node label	center: AS number upper left: rank upper right: country flag
node vertical position	customer cone size
node size	customer cone size
link width	customer cone size of neighbor
link color	relationship type

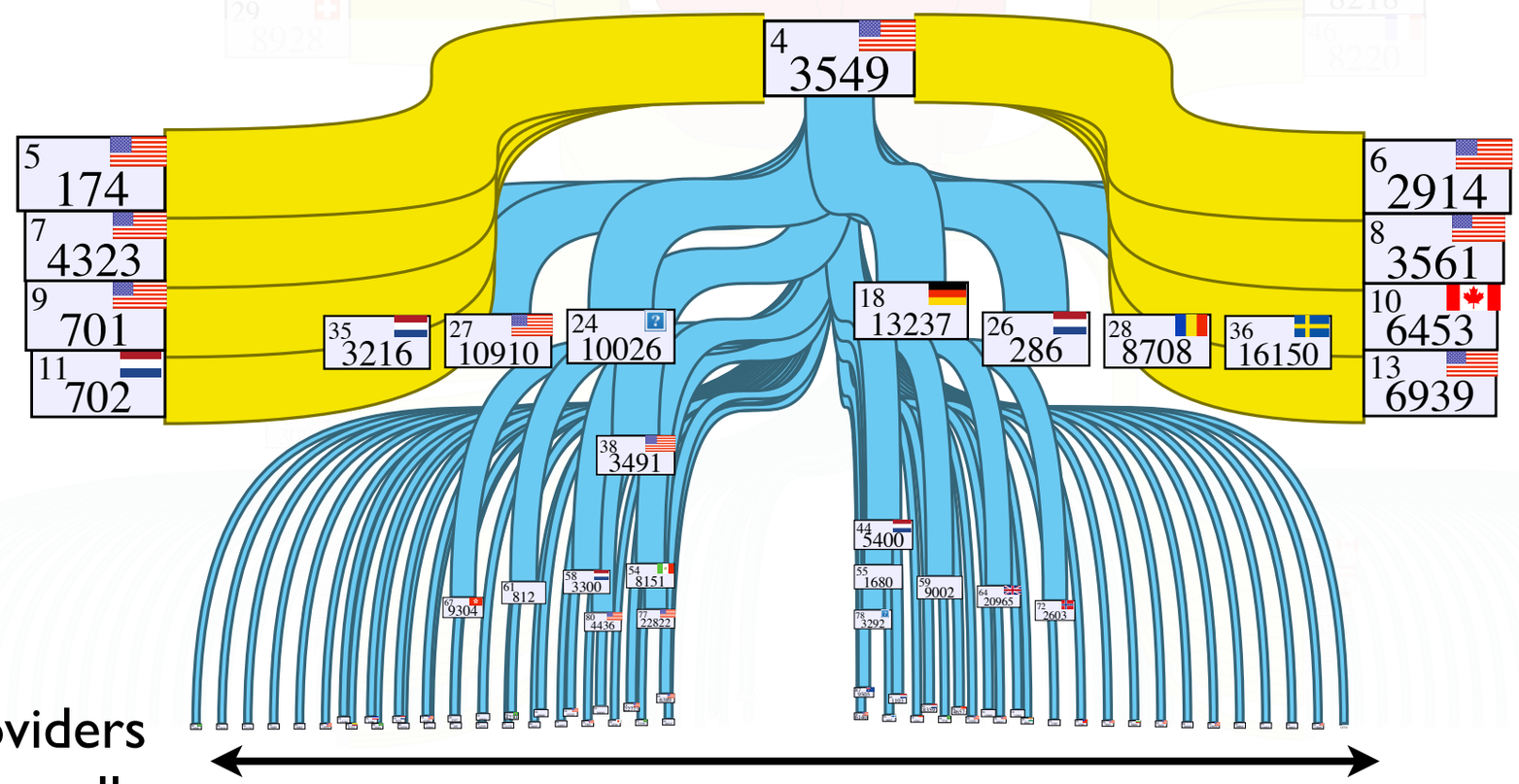


AS placement

as relationships

- **AS node placement**
  - overlapping provider / customers move along the X axis
  - overlapping peers move along the Y axis (distorts cone size mapping)
- **large links drawn over small links**
- **AS always drawn over links**
  - visually disconnects AS and its link
  - but nodes can been seen and read

overlapping peers move vertically

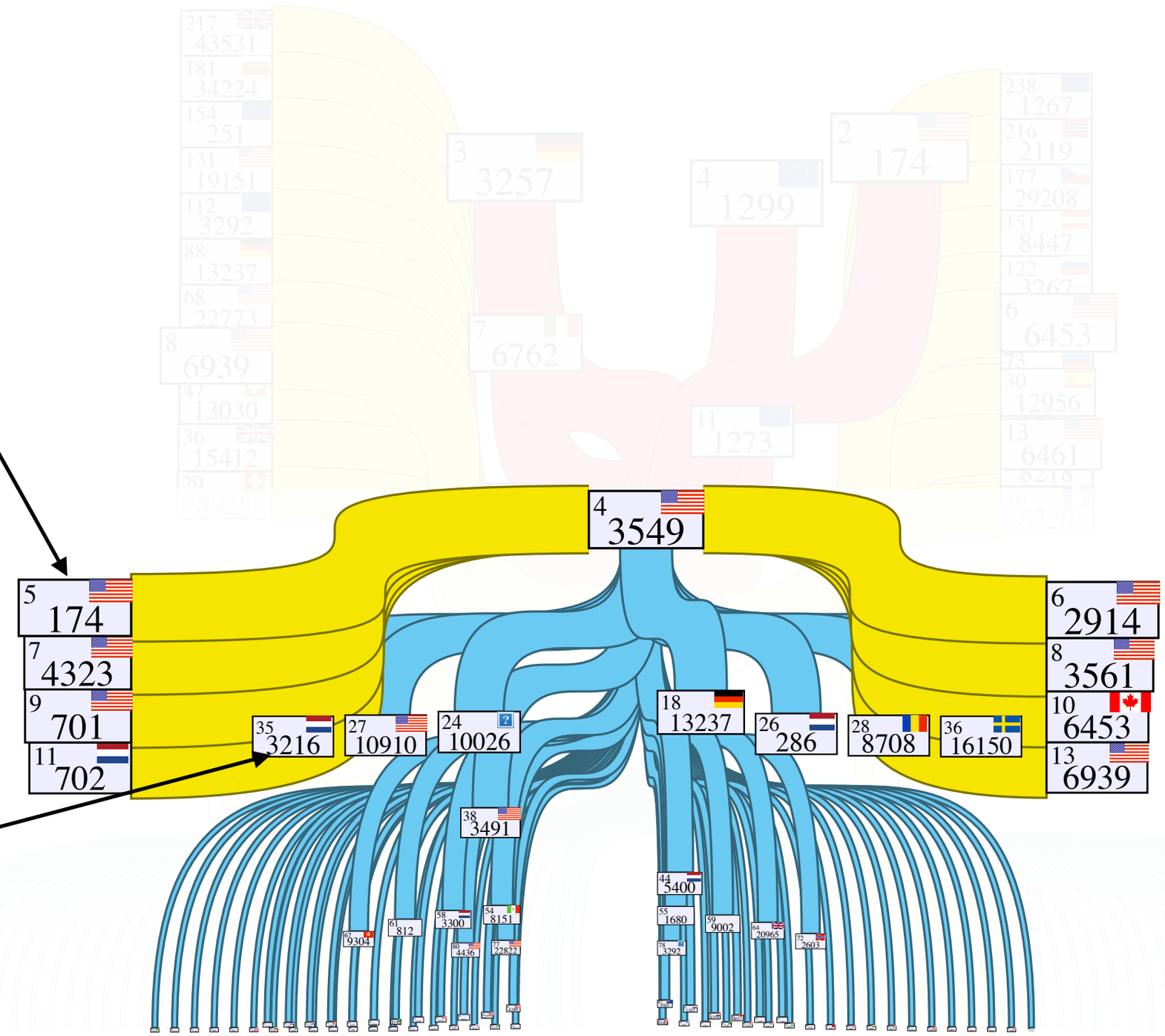


overlapping customers/providers move horizontally

AS placement

as relationships

More important, i.e. larger cone sizes, ASes have larger and more readable information.

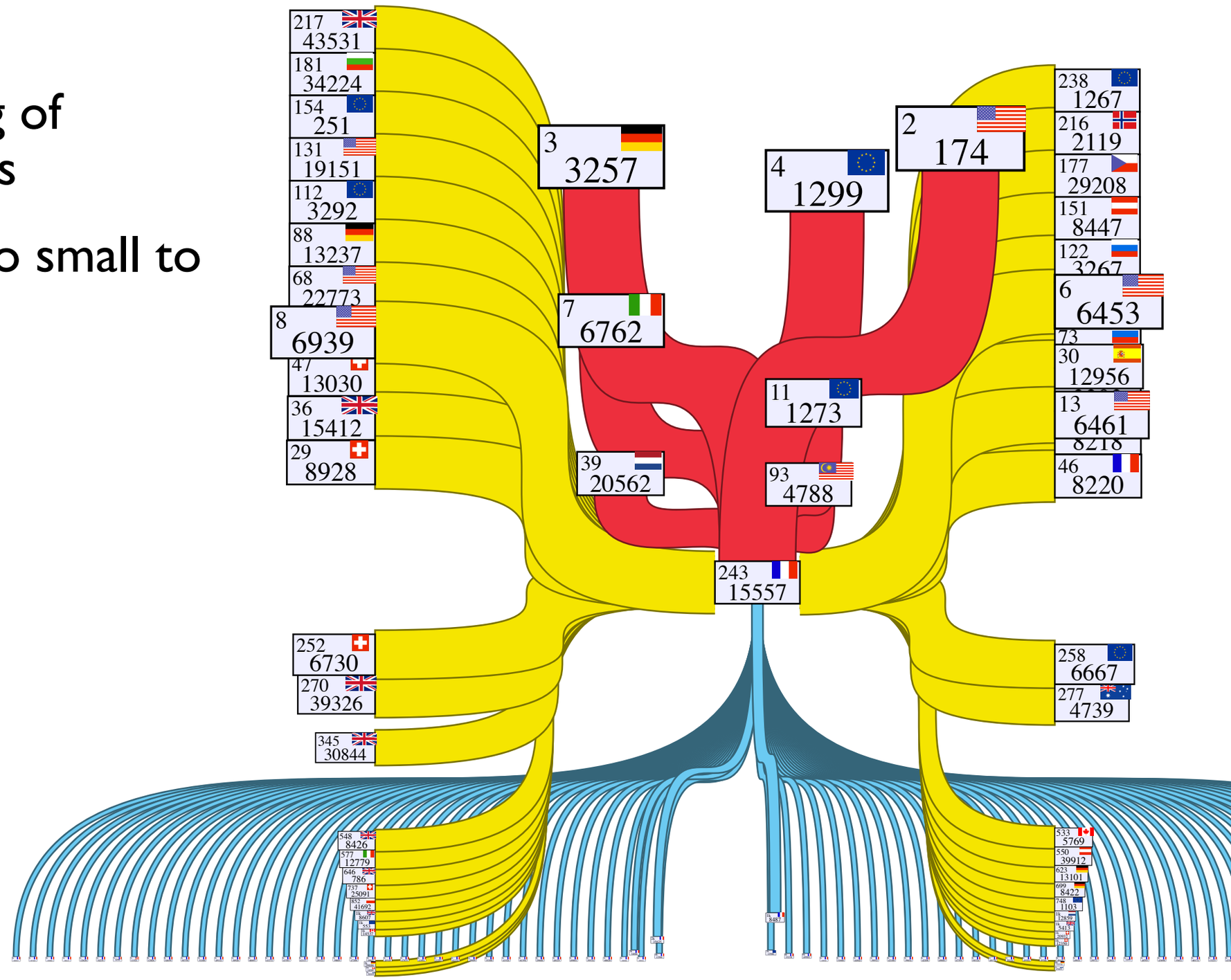


Node placement means a node is visible, even if its link is not.

limitations

as relationships

- redundant encoding of customer cone sizes
- smaller ASes are too small to read

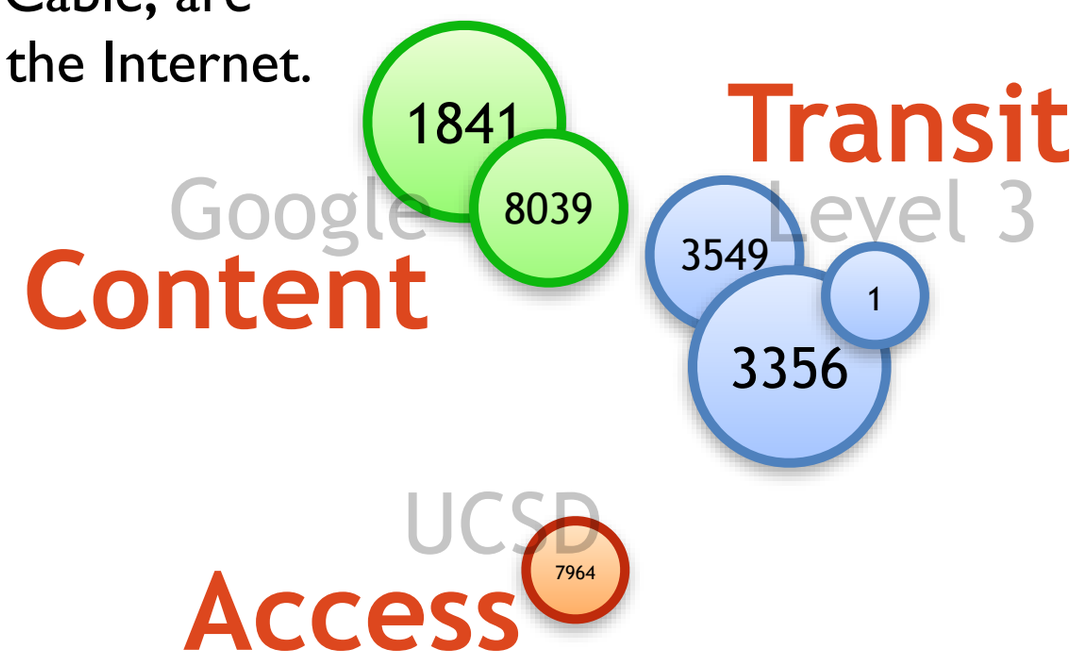


AS types

bonus

AS are often divided into Transit, Content, or Access depending on the type of companies that operate them. Larger organizations can combine all three types.

- **Transit ASes**, such as AT&T, form the backbone of the Internet by carrying traffic between ASes.
- **Content ASes**, such as Netflix, provide the videos, webpages, and other content.
- **Access ASes**, such as UCSD or Time Warner Cable, are used by smaller organizations to get access to the Internet.



AS types (file)

bonus

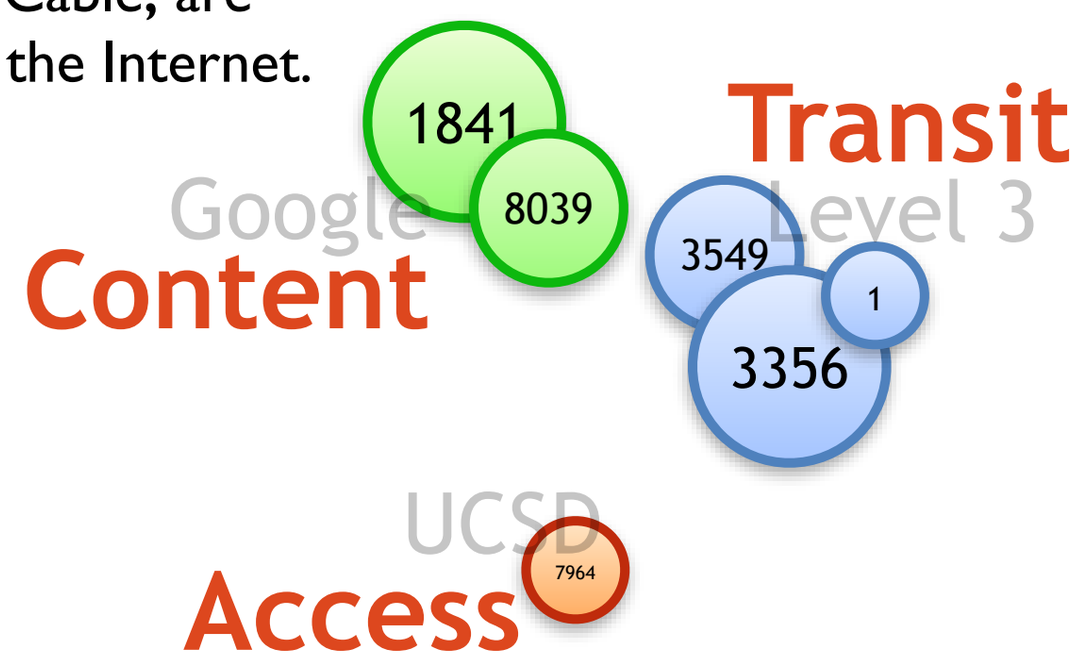
Cases are often divided into Transit, Content, or Access depending on the type of companies that operate them. Larger organizations can combine all three types.

- **Transit ASes**, such as AT&T, form the backbone of the Internet carrying traffic between ASes.
- **Content ASes**, such as Netflix, provide the vides, webpages, and other content.
- **Access ASes**, such as UCSD or Time Warner Cable, are used by smaller organizations to get access to the Internet.

20150101.as2type.txt.bz

```
# name: as2type
# ../as-graph-201409/bin/type-convert-amogh.pl
# types:Access|Content|Transit/Access
1|Transit/Access|CAIDA_class
2|Transit/Access|CAIDA_class
```

AS number — type — inference method



key points

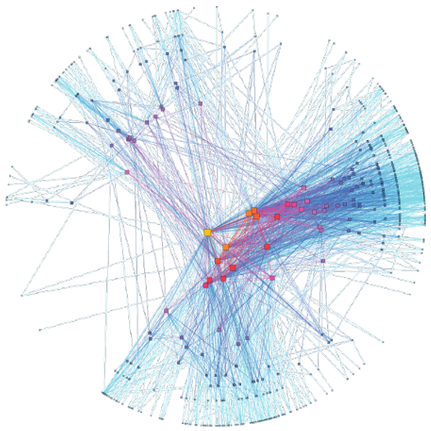
- **important objects drawn last** (i.e. over less important objects)
  - rare objects over common (typically more links then nodes)

- **highlight a few features in a complex graph**

- **not all objects must be readable / distinguishable**
  - it is often enough to be able to get an idea of the frequency

data

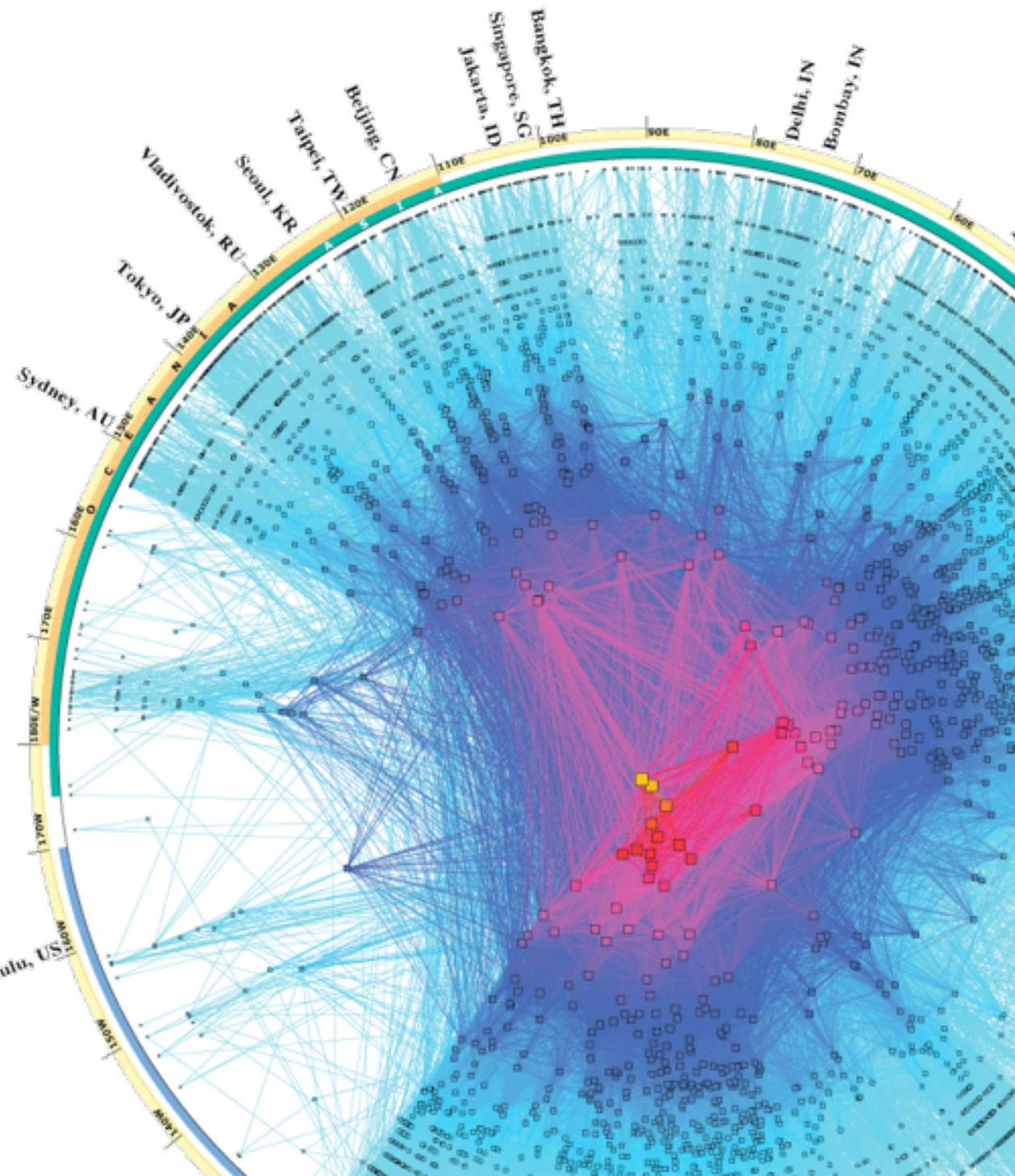
summary



http://www.caida.org/publications/presentations/2016/as_intro_visualization_ucsd/

20150101.paths.txt.bz	all observed AS paths	slide 10
20150101.as2loc.bz	centroid of AS's address space	slide 16
20150101.as-org2info.txt.bz	provides country and organization information	slide 17,36
20150101.as-rel.txt.bz	types of business relationships between pairs of ASes	slide 32
20150101.ppdc-ases.bz	the set of ASes that are customers, or customer's customers of an AS	slide 34
20150101.as2type.txt.bz	an AS's type of business	slide 47

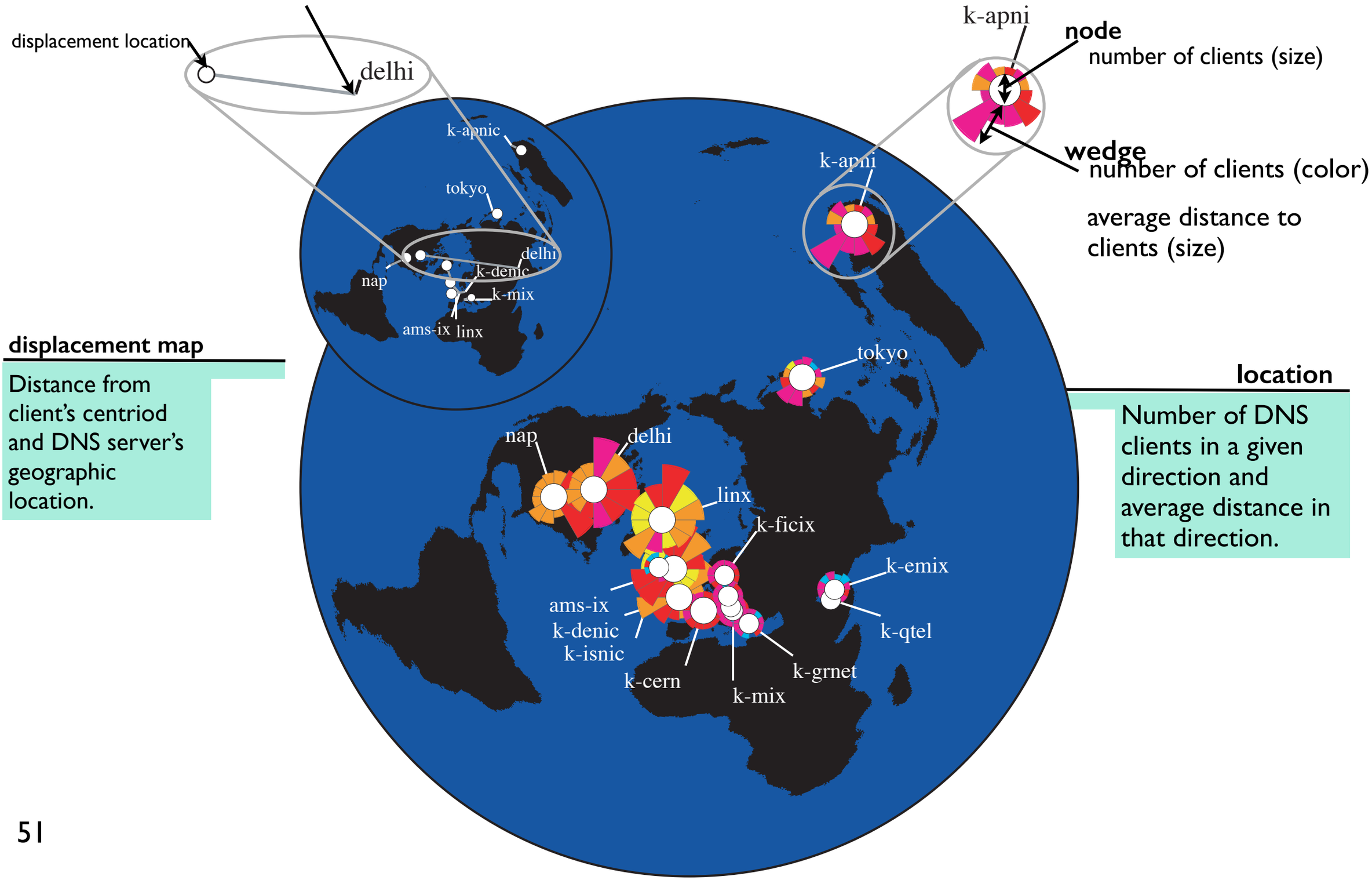
Questions?



Internships:
http://www.caida.org/home/jobs/

anycast

bonus



displacement map

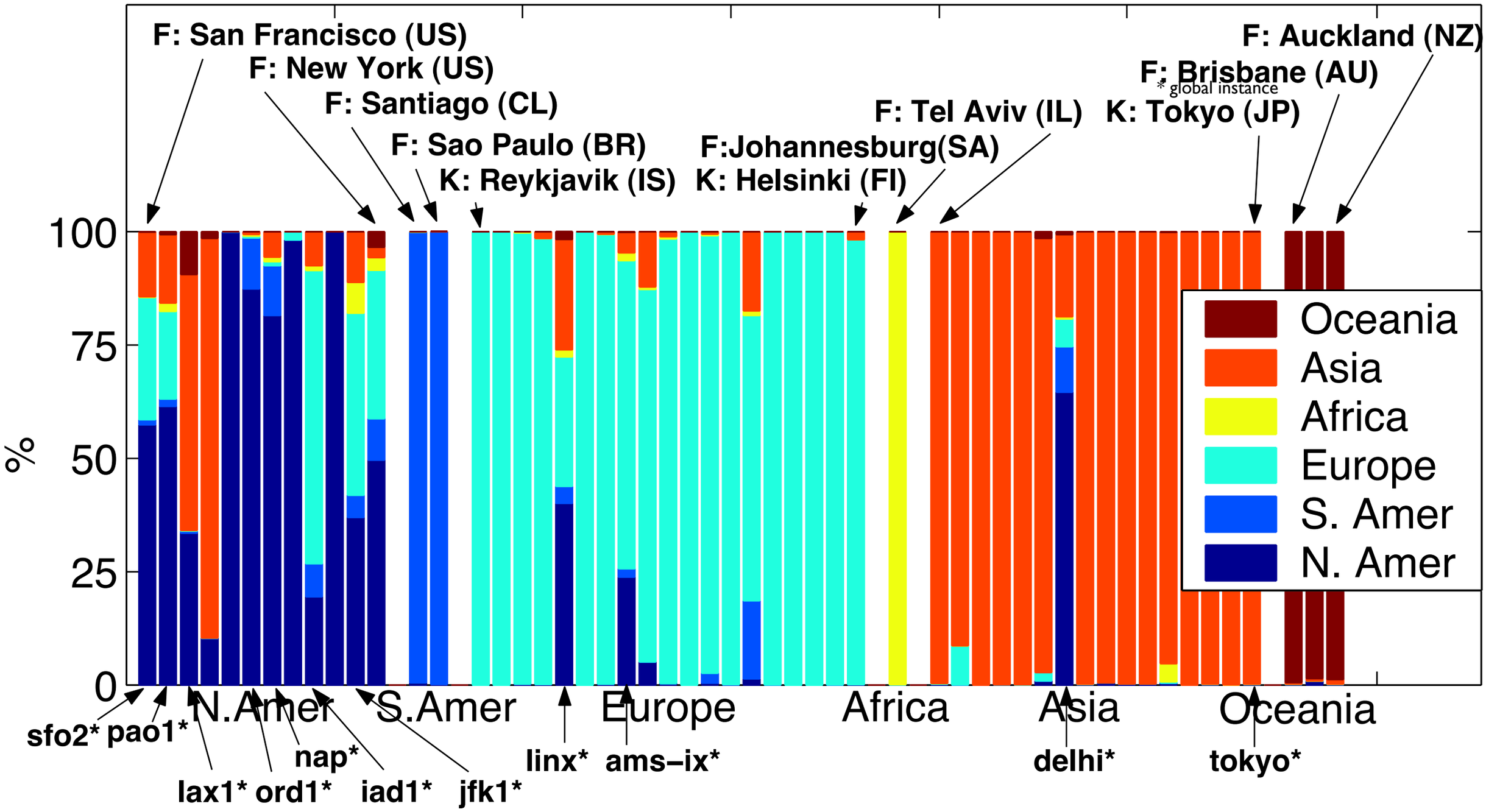
Distance from client's centriod and DNS server's geographic location.

location

Number of DNS clients in a given direction and average distance in that direction.

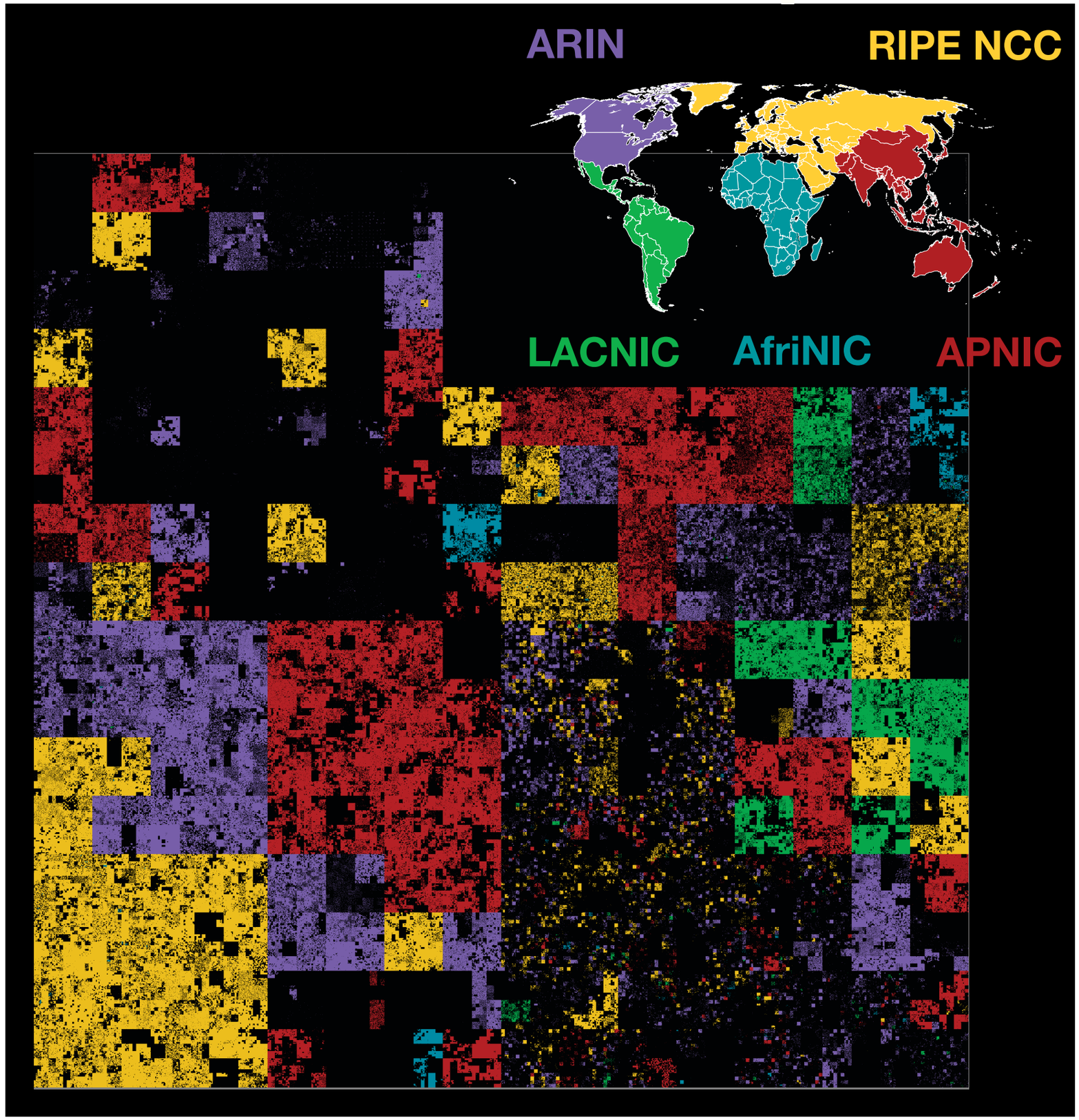
Geographic sorted bars

bonus



used address space

bonus



Code red

bonus