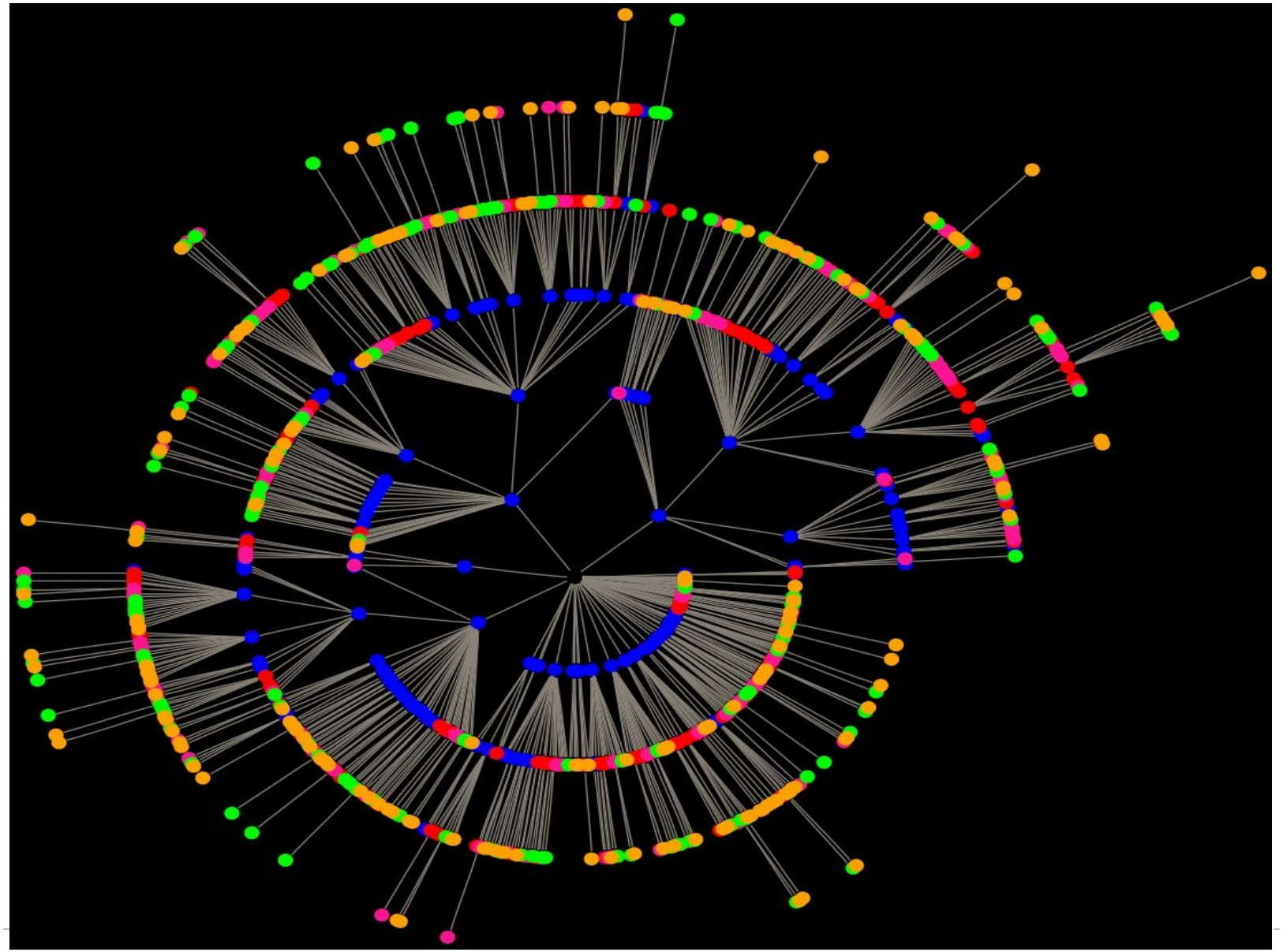
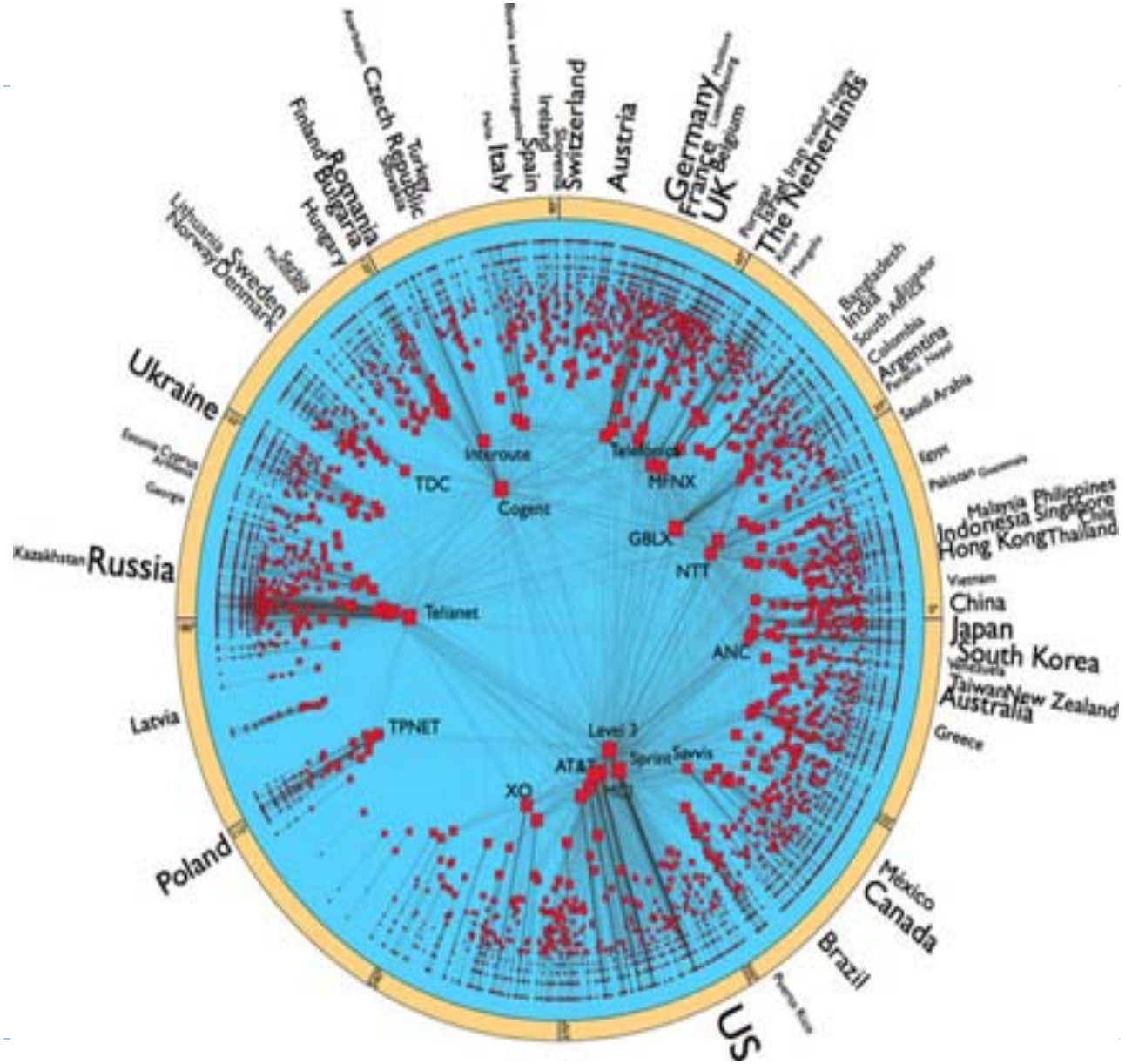
Internet Research with "Big (Internet) Data"

Part II (of III)

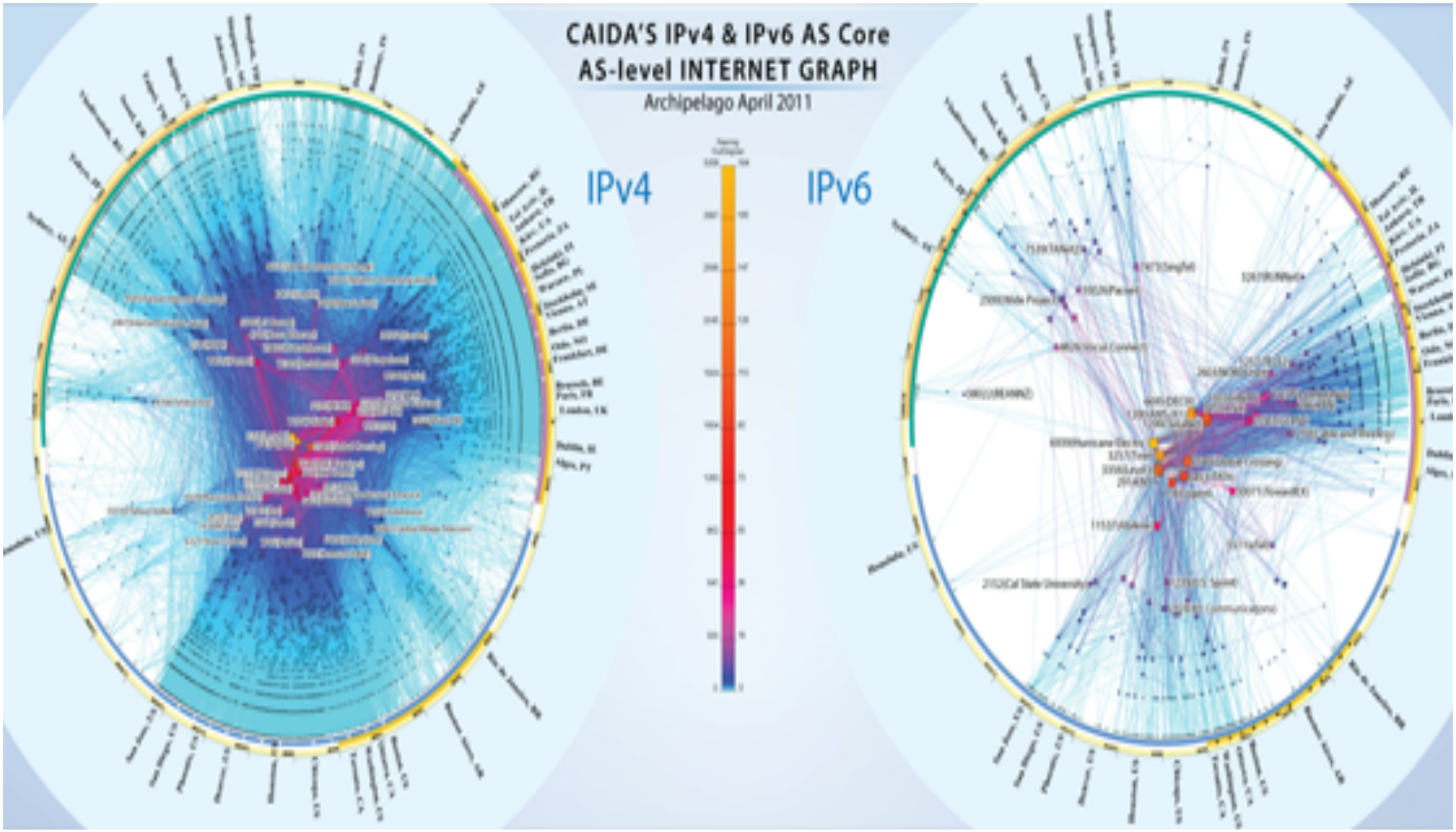
Walter Willinger
NIKSUN, Inc.

wwillinger@niksun.com




Italy
Switzerland
Austria
Germany
France
UK
The Netherlands
Czech Republic
Romania
Bulgaria
Hungary
Finland
Sweden
Denmark
Norway
Lithuania
Ukraine
Russia
Kazakhstan
Latvia
Poland
Belgium
Argentina
Colombia
India
South Africa
Saudi Arabia
Egypt
Malaysia
Philippines
Singapore
Indonesia
Hong Kong
Thailand
China
Japan
South Korea
Taiwan
New Zealand
Australia
Greece
México
Canada
Brazil
US
Interoute
TDC
Cogent
Telefonica
MFNX
GBLX
NTT
ANC
Telianet
TPNET
Level 3
AT&T
Sprint
Savvis
XO


The Internet
peer1 hosting
S8 naukanet
S11 neo telecoms group
O8 neterra ltd
G17 netgate
L18 netia commercial auton
H15 new world network
P16 new edge networks
G14 newline datagroup inter
M7 nlayer communications
K15 nordunet
P12 ntl group limited
F12 orange pcs limited
L5 ovh
N10 peer 1 network inc
H7 pointshare corporation
H13 polish telecom
F7 pt telekomunikasi indon
A9 qwest
I4 rcn corporation
L11 reach network


CAIDA'S IPv4 & IPv6 AS Core
AS-level INTERNET GRAPH
Archipelago April 2011
IPv4
IPv6

The **AS-level** Internet

The **AS-level** Internet

- The Internet as a logical/virtual construct
  - Nodes: Autonomous systems/domains (ASes)
  - Links: Logical (i.e., protocol-defined) connections between ASes
  - AS-link: the two ASes exchange reachability information
  - Reachability: "active" BGP session(s) between border routers
  - AS-link is defined via a protocol: Border Gateway Protocol (BGP)
- AS-level topology of the Internet
  - The Internet as a "network of networks"

The **AS-level** Internet (since ~1995)

- Challenges (due to decommissioning of NSFNET)
  - No one entity has a complete view of the network
  - Networks come in many shapes and forms …
  - What geography for networks?

- Popular "recipe" for studying the AS-level Internet
  - Step 1: Use BGP measurements (routing tables, updates)
  - Step 2: Obtain the data from multiple route monitors
  - Step 3: Combine BGP-derived AS-level paths to obtain the Internet's AS-level topology

Step 1: BGP RIBs

- BGP RIBs (routing information base)
  - RIBs contain routing information maintained by BGP-speaking routers
  - Typical Routing table size: ~200-300K entries
  - Augment with constantly exchanged announcement/withdrawal messages

Typical BGP RIB table entry

```
PREFIX:        4.21.252.0/23
FROM:          194.85.4.55  AS3277
TIME:          2004-12-31 20:07:56
TYPE:          MSG_TABLE_DUMP/AFI_IP
VIEW:          0  SEQUENCE: 440
STATUS:        1
ORIGINATED:    Fri Dec 31 06:26:51 2004
AS_PATH:       3277 13062 20764 701
               6389 8063 19198
NEXT_HOP:      194.85.4.55
COMMUNITIES:   3277:13062 3277:65301
               3277:65307 20764:3000
               20764:3011 20764:3020
               20764:3022
```

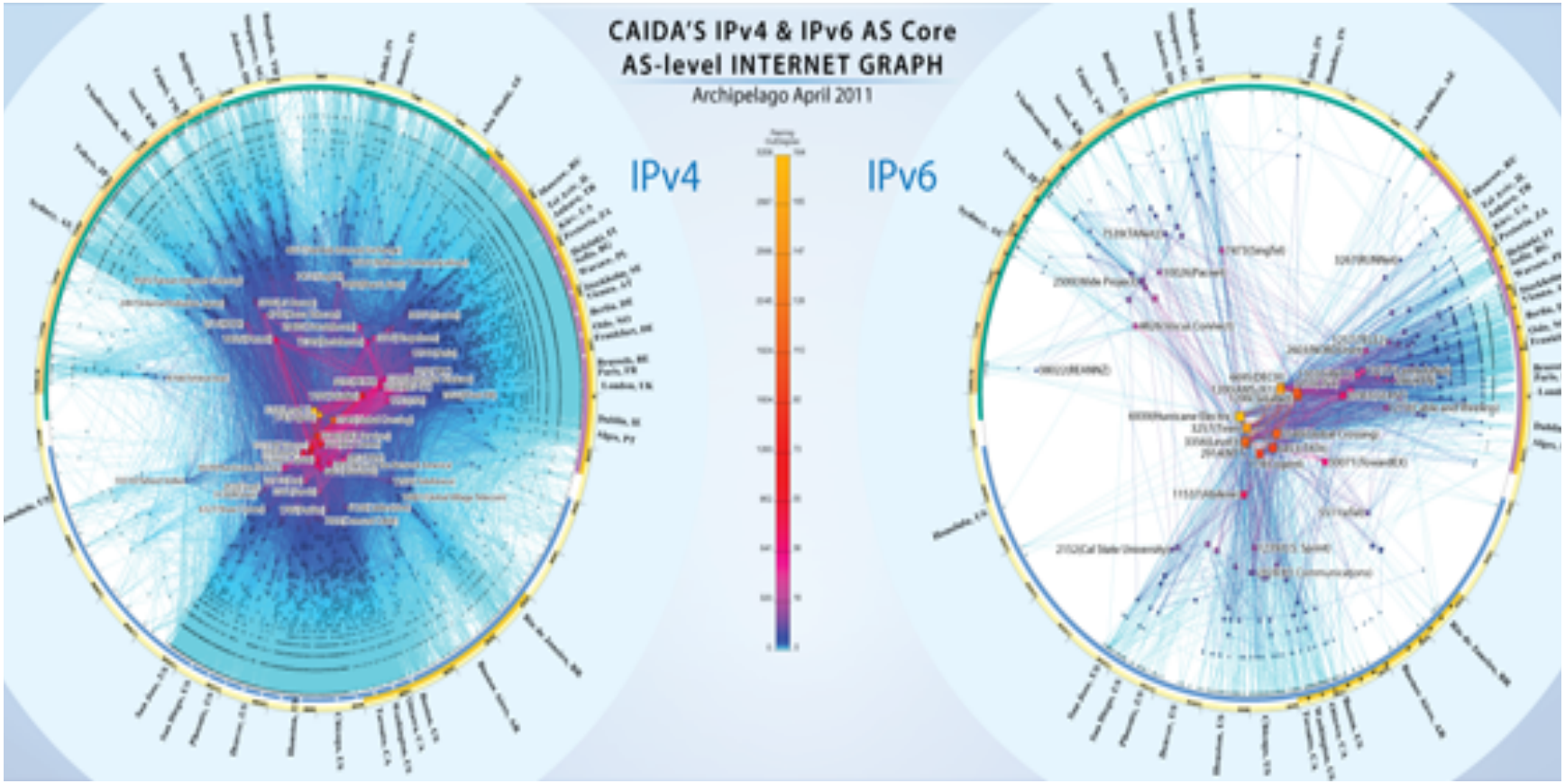
Step 2: BGP data collections

- Commonly-used publicly available large BGP datasets
  - RouteViews project (Univ. of Oregon, since ~1997)
    - www.routeviews.org/
  - RIPE RIS project (RIPE NCC, Netherlands, since ~2000)
    - www.ripe.net/data-tools/stats/ris/routing-information-service

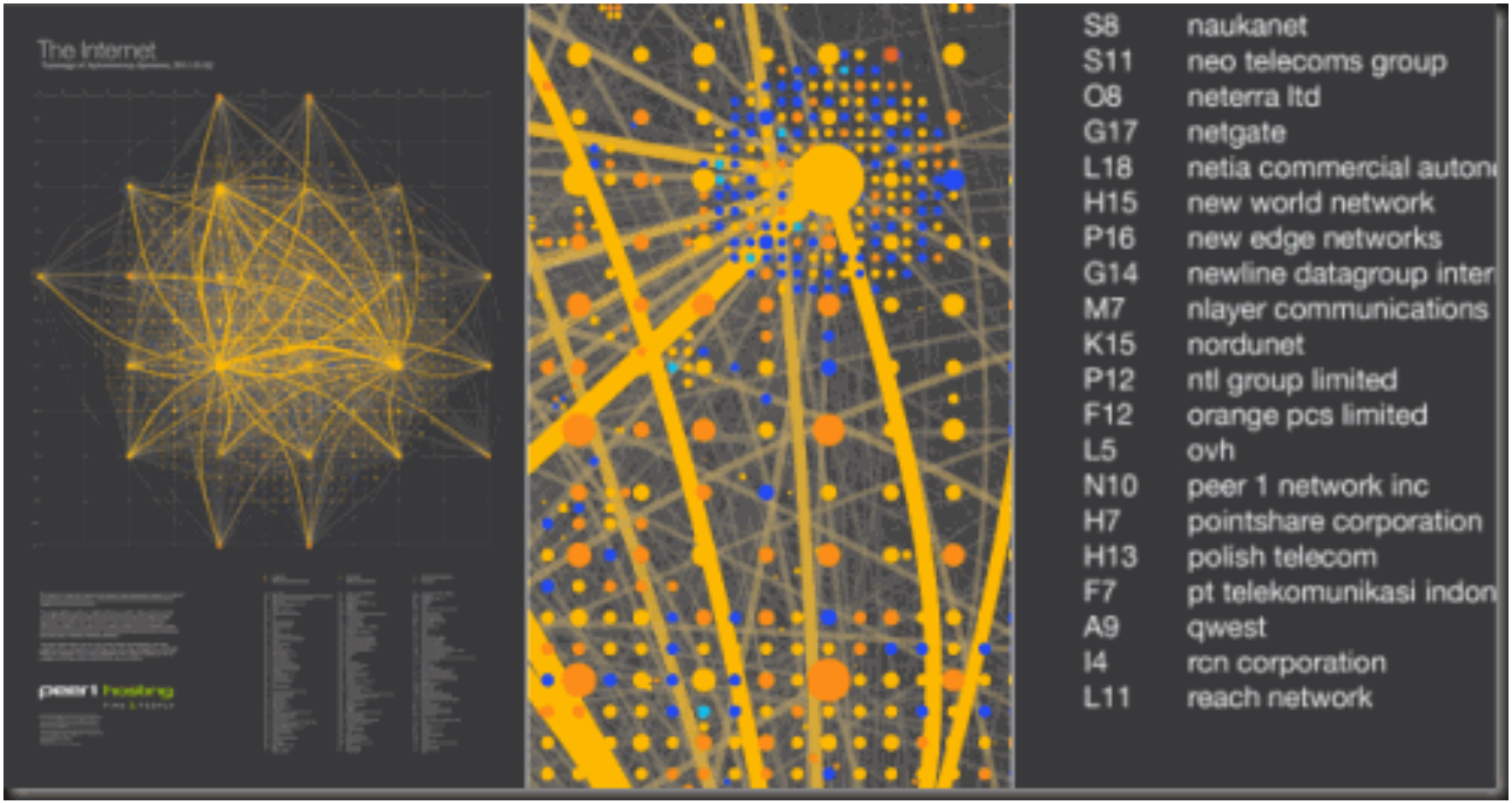
Step 3: Combine AS-level paths

- Another example of "big (Internet) data"
  - Currently, there are some 14 RIS route collectors, and each one of them collects an entire BGP routing table every eight hours
  - 1 table (~ 200-400K RIB entries) is about 500MB (uncompressed)
  - Some 4 billion BGP-derived AS-level paths (~ 7 PB of data) per year
- Working assumption
  - With billions of BGP-derived AS paths, it is possible to recover the Internet's AS-level topology
  - The produced visualizations provide "insight" into the Internet's router-level topology

The "**AS-level**" Internet (caida.org)

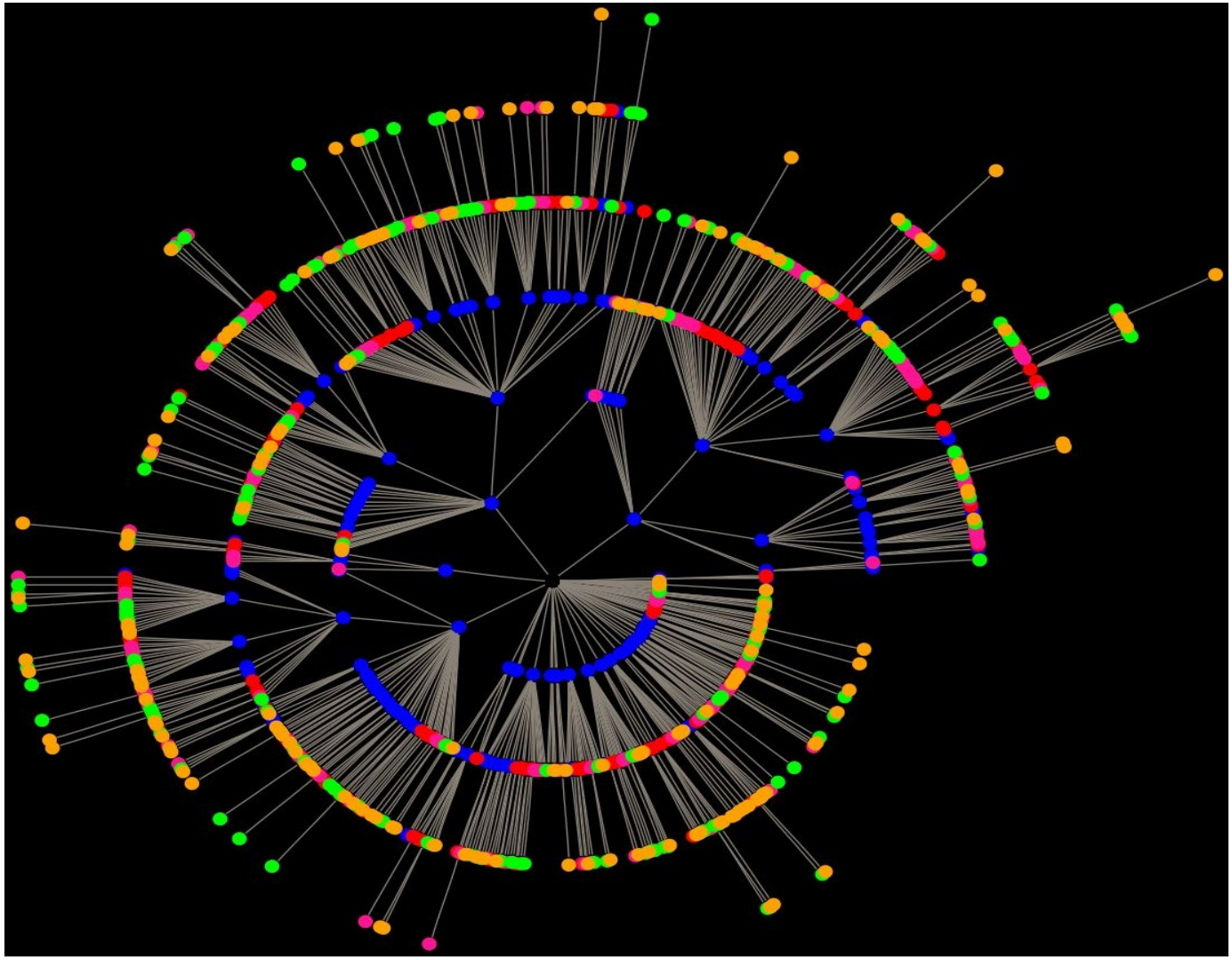


The "**AS-level**" Internet (Peer1.com)



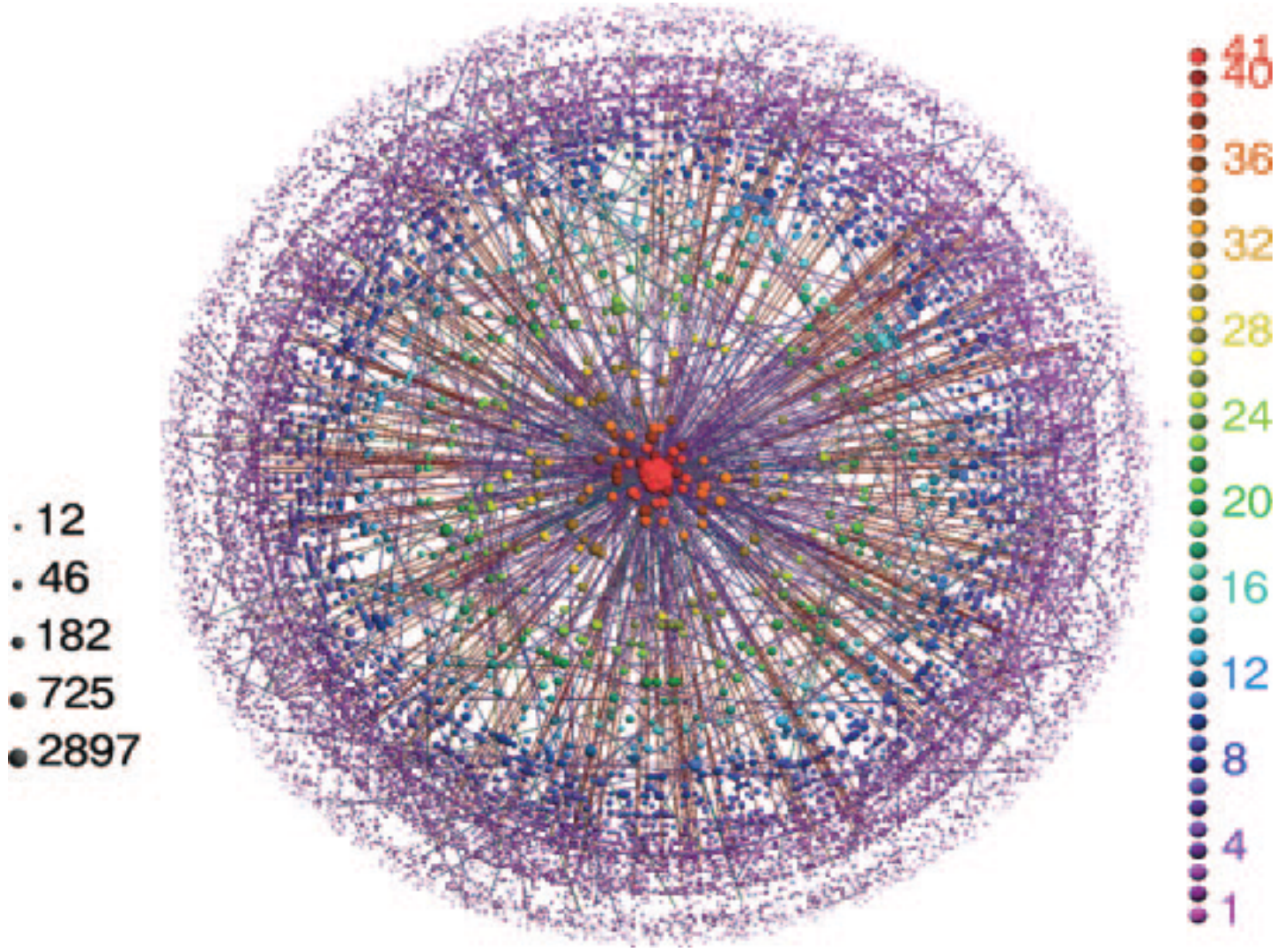
http://www.peer1.com/blog/2011/03/map-of-the-internet-2011

The "**AS-level**" Internet (2007)



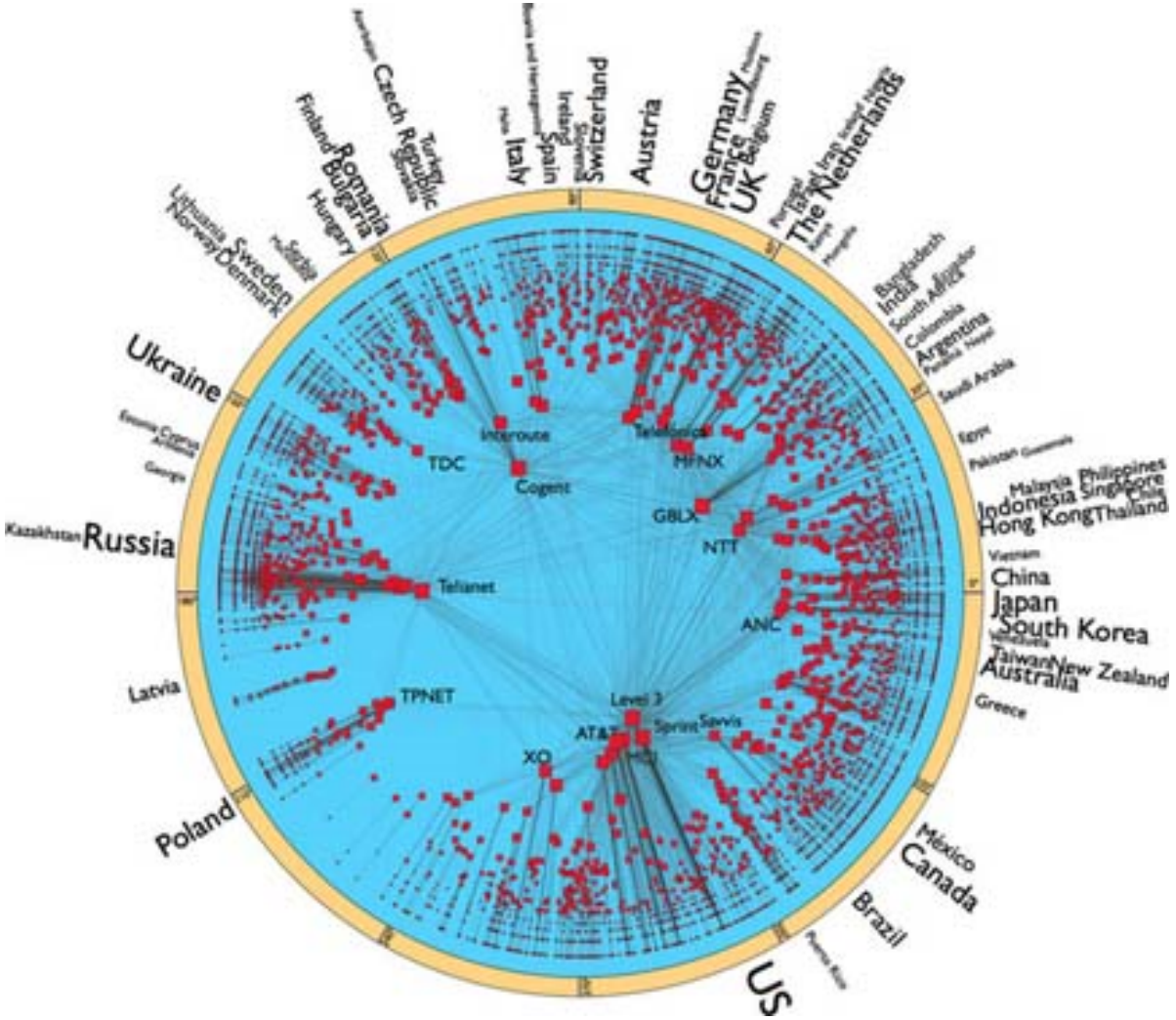
R. D'Souza, C. Borgs, J. Chayes, N. Berger, and R. Kleinberg (PNAS 2007)

The "**AS-level**" Internet (2007)



S. Carmi, S. Havlin, S. Kirkpatrick, Y.Shavitt, and E. Shir (PNAS 2007)

The "**AS-level**" Internet (2010)



- M. Boguna, F. Papadopoulos, and D. Krioukov (Nature Communications, 2010)

The "AS-level" Internet (since ~2000)

- "Discoveries" and "insights"
  - Random (e.g., scale-free) graphs appear to be suitable models
  - There are ASes with high-degree nodes in the Internet
  - Removal of these high-degree ASes destroys the Internet
  - Discovery of an "Achilles' heel" for the AS-level Internet

… problem solved!?

- Solid? – Based on real data!
- Rigorous? – Scale-free network models!
- Relevant? – Discovery of another serious vulnerability!

Back to basics ("engineer's view") …

- How did we get here ("physicist's view")?
  - Available data is taken at face value ("don't ask …")
  - No or only little domain knowledge is required
  - The outcome often leaves little room for further efforts
- The "Engineer's" perspective
  - Available data tends to be scrutinized (not enough, though)
  - Domain knowledge is "king" – details matter!
  - The results often give rise to new questions/problems

The Engineer's view

- The inter-domain routing system
- The inter-domain routing protocol BGP

- BGP-based measurements
- BGP data collection projects for the public good

Engineer's view – some details

- Basic problem re measuring AS-level connectivity
  - Individual ASes know their (local) AS-level connections
  - AS-specific connectivity data is considered proprietary
  - AS-level connectivity cannot be measured directly
  - No central agency exists that collects this data
- Practical solution
  - Rely on BGP, the de facto inter-domain routing protocol
  - Use BGP RIBs (routing information base)
  - RIBs contain routing information maintained by the router

Engineer's view – some details (cont.)

- Inter-domain routing system
  - Foundation for Internet wide-area communication
- Inter-Domain Routing Protocol BGP4
  - De facto standard inter-domain routing protocol RFC 1771/4271
  - Enables ASes to implement/realize their routing policies
- Scalable, expressive, flexible information-hiding protocol
  - Exchange of routing information w/o revealing AS-internals
  - Support for complex and evolving AS-specific business policies

Engineer's view - some "details"

- Key observation
  - BGP is not a mechanism by which ASes distribute connectivity information
  - BGP is a protocol by which ASes distribute the reachability of their networks via a set of routing paths that have been chosen by other ASes in accordance with their policies.

- BGP-based measurements to map the AS-level Internet
  - Engineering hack – BGP is an information-hiding and not an information-revealing protocol
  - An example of "What we can measure is typically not what we want to measure!"

Engineer's view – some details (cont.)

- Use BGP RIBs (routing information base) and updates
  - RIBs contain routing information maintained by the router
  - Typical Routing table size: ~200-300K entries
  - Focus has been on AS_PATH attribute
- Typical BGP RIB table entry

```
PREFIX:          4.21.252.0/23
FROM:            194.85.4.55  AS3277
TIME:            2004-12-31  20:07:56
TYPE:            MSG_TABLE_DUMP/AFI_IP
VIEW:            0   SEQUENCE:  440
STATUS:          1
ORIGINATED:      Fri Dec 31  06:26:51  2004
AS_PATH:         3277  13062  20764  701
                 6389  8063  19198
NEXT_HOP:        194.85.4.55
COMMUNITIES:     3277:13062  3277:65301
                 3277:65307  20764:3000
                 20764:3011  20764:3020
                 20764:3022
```

Engineer's view – some details (cont.)

- Daily BGP table dumps and updates are collected from multiple monitors that are connected to numerous routers across the Internet
- RouteViews project (University of Oregon)
  - Started ~1997
  - Initially connected to large providers, recently also to IXPs
  - http://www.routeviews.org/
- RIPE RIS project (RIPE NCC, Netherlands)
  - Started data collection around 2000
  - Similar approach as RouteViews
  - http://www.ripe.net/data-tools/stats/ris/routing-information-service

Engineer's view – some details (cont.)

- RouteViews/RIPE RIS data were never meant to be used to infer the Internet's AS-level connectivty
- BUT: value/benefit of the data for operators is huge!

From RouteViews/RIPE RIS websites

- *"The RouteViews project was originally conceived as a tool for Internet operators to (i) obtain real-time information about the global routing system from the perspectives of several different backbones and locations around the Internet, and (ii) determine how the global routing system viewed their prefixes and/or AS space."*

- *"The goal of the Routing Information Service (RIS) is to collect routing information between ASes and their development over time from a number of vantage points in the Internet. One important application for this data will be debugging. For example, if a user complains that a certain site could not be reached earlier, the RIS will provide the necessary information to discover what caused the problem."*

Engineer's view – some details (cont.)

- Basic problem #1: Incompleteness
  - Many peering links/relationships are not visible from the current set of BGP monitors
  - A well-known problem of vantage point locations
- Basic problem #2: Ambiguity
  - Need heuristics to infer "meaning" of AS links: customer-provider, peer-to-peer, sibling, and a few others
  - Existing heuristics are known to be inaccurate
  - Renewed recent efforts to develop better heuristics …

Engineer's view – some details (cont.)

- The dilemma with current BGP measurements
  - Parts of the available data seem accurate and solid (i.e., customer-provider links, nodes)
  - Parts of the available data are highly problematic and incomplete (i.e., peer-to-peer links)
- Bottom line
  - (Current) BGP-based measurements are of questionable quality for accurately inferring AS-level connectivity
  - It is expected that future BGP-based measurements will be more useful for the purpose of AS-level inference
  - Very difficult to get to the "ground truth"

But the issues with BGP data is not new!!

- R. Govindan and A. Reddy, 1997. An analysis of Internet inter-domain topology and route stability. IEEE INFOCOM.

- The Govindan & Reddy 1997 paper is an early textbook example for what information a measurement paper should provide.
- The Govindan & Reddy 1997 paper is now hardly cited and largely forgotten!
- An example of the influence that secondary citations can and do have …

The missing link problem in BGP data

- Estimated size of the AS-level Internet (around 2010)
  - # nodes: about 30,000
  - # customer-provider links: about 60,000
  - # peering links: about 20,000
- Estimated size of the AS-level Internet (around 2012)
  - #nodes: about 40,000
  - # customer-provider links: about 80,000
  - # peering links: more than 200,000

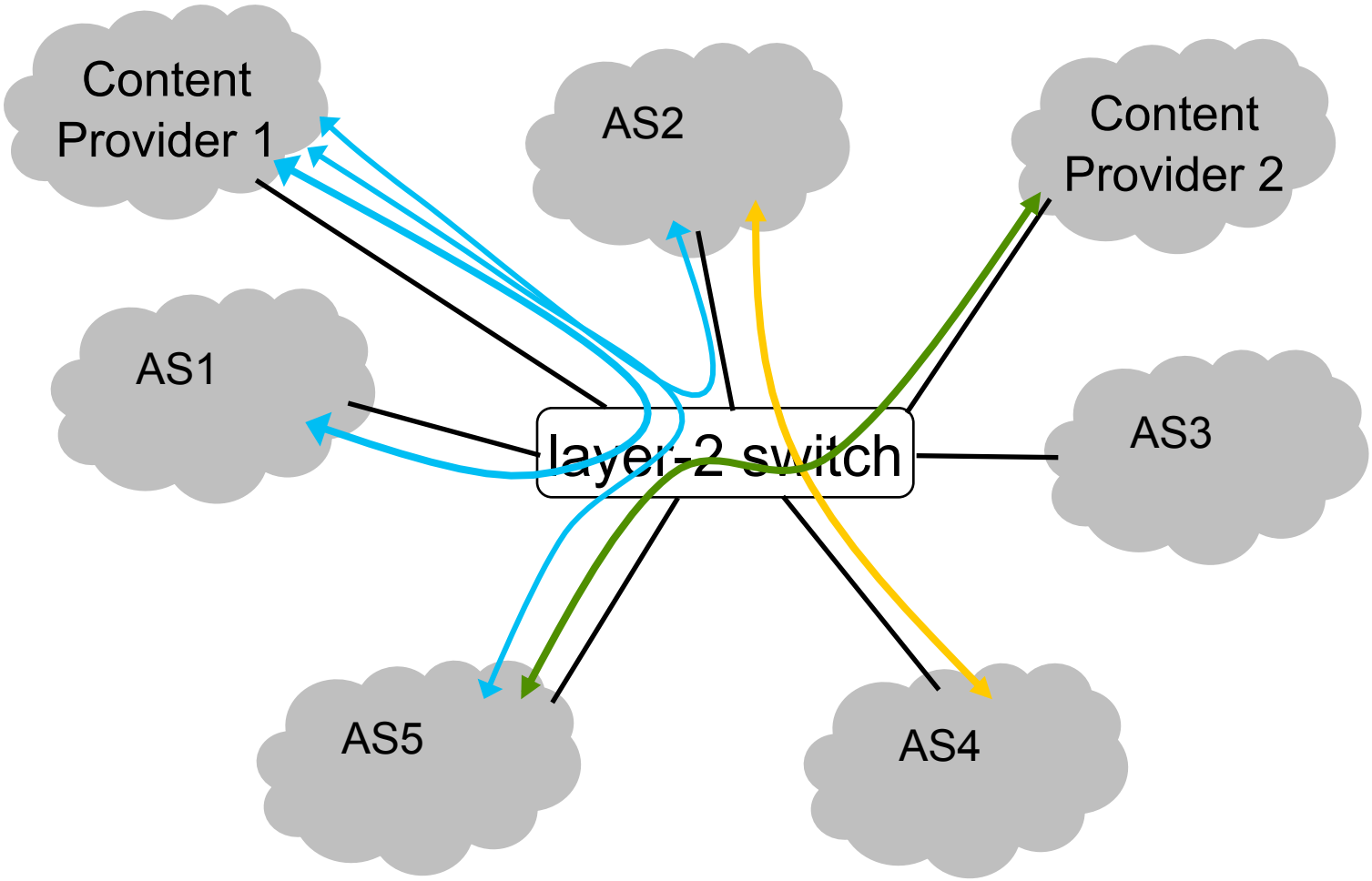
What happened between 2010-2012?

- We got lucky …! Anja Feldmann's group (T-Labs/TU Berlin) obtained high-quality traffic data from one of the largest IXPs in Europe

- *B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, W. Willinger Anatomy of a Large European IXP, ACM Sigcomm 2012*

- Internet eXchange Point (IXP): An IXP is a physical facility with a switching infrastructure for the primary purpose to enable networks to interconnect and exchange traffic directly (and essentially for free) rather than through one or more 3rd parties (and at a cost).

Internet eXchange Points (IXPs)



Key findings from mining the IXP data

- At this one IXP, there are more than twice as many peering links in use as compared to the most current estimate of the total number of peerings in the entire Internet
- The total number of peering links at the 300+ active IXPs worldwide is at least 200,000
- The vast majority of these 200,000 peerings are not visible to the commonly-used BGP measurements.

Discussion

- **Details do matter!**
- **BGP data are not meant for mapping the AS-level Internet!**
- **Every networking student is assumed to know BGP, so why is this domain knowledge not used?**
- **Why has the 1997 Govindan & Reddy paper never been referenced in subsequent Internet topology papers?**

A fresh look at the **AS-level** Internet …

Where **AS-** meets **physical-level** Internet

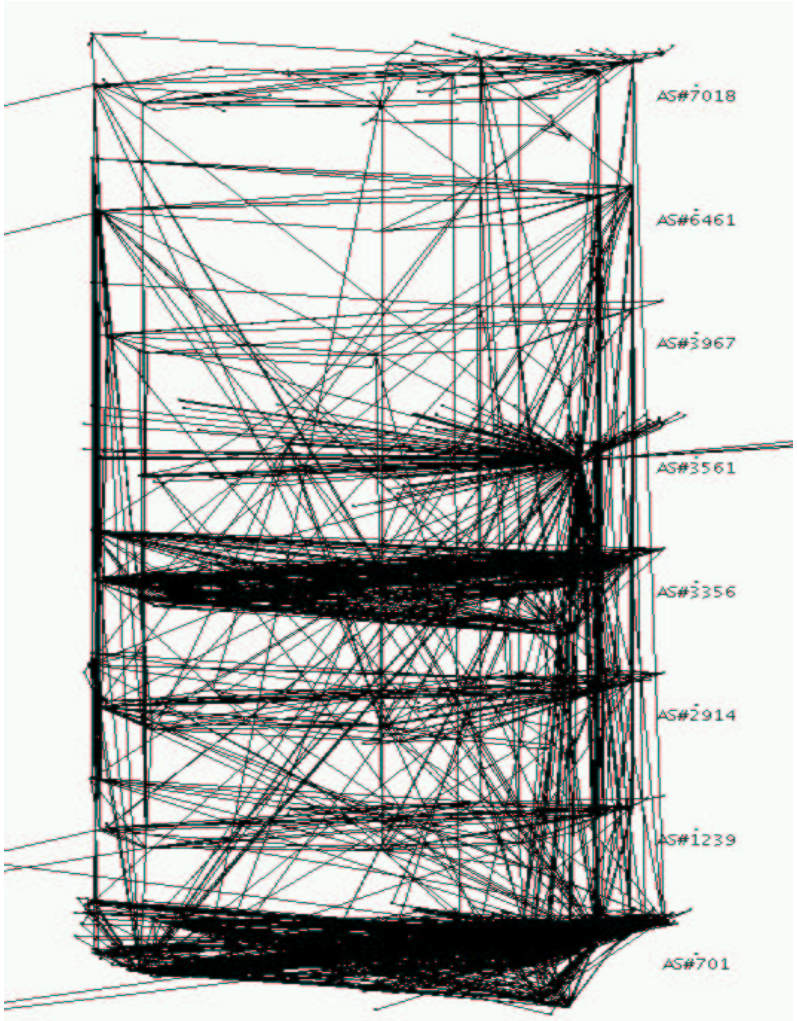
- The critical role of colocation facilities
  - IXPs are housed in one or more colo facilities
  - Routers of the different ASes are housed in colos/router hotels
  - (Some US Tier-1s have separate facilities/buildings)

- Internet connectivity
  - Who is connecting with whom in which colocation facility?
  - Inter-AS links manifest themselves in different physical connections (between distinct border routers)
  - Intra-AS connectivity is the router-level view of the AS

DE-CIX: Colocation facilities in FRA

- Equinix, FR4, Lärchenstr. 110
- Equinix, FR5, Kleyerstr. 90
- Equinix, FR2, Kruppstr. 121-127
- e-shelter, Eschborner Landstr. 100
- I.T.E.N.O.S., Rebstöckerstr. 25-31
- Interxion, FRA1, Hanauer Landstr. 302
- Interxion, FRA2, Hanauer Landstr. 304A
- Interxion, FRA3, Weissmüller Str. 21
- Interxion, FRA4, Weissmüller Str. 19
- Interxion, FRA5, Hanauer Landstr. 308a
- Interxion, FRA6, Hanauer Landstr. 300a
- Interxion, FRA7, Hanauer Landstr. 296a
- KPN, Kleyerstr. 90
- Level3, Kleyerstr. 82 (Building A)
- Level3, Kleyerstr. 90
- NewTelco, Rebstöckerstr. 25-31 (Building B, Room B.1.10)
- TelecityGroup, Gutleutstr. 310
- Telehouse, Kleyerstr. 79 (Building K)
- Telehouse, Kleyerstr. 79 (Building I)

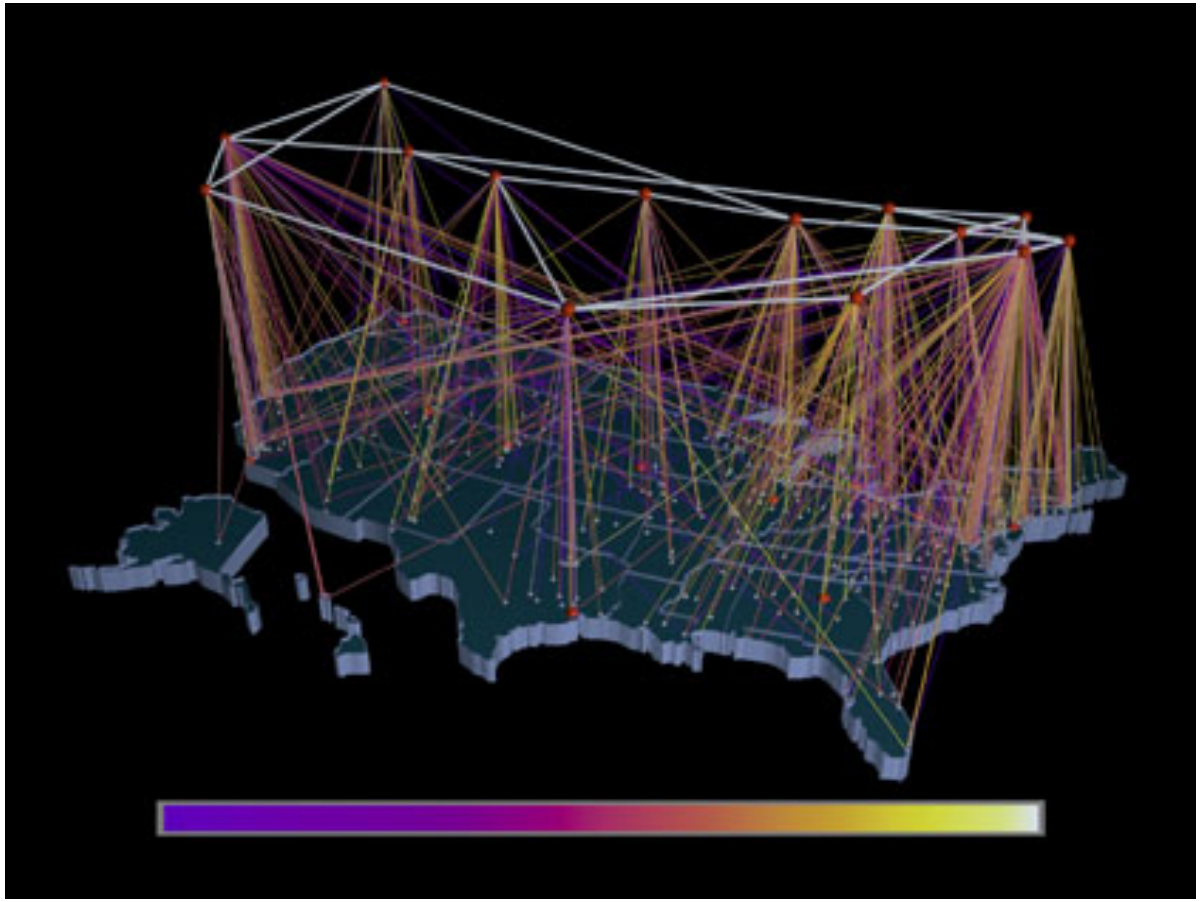
An early attempt (D. Nicol et al. 2003)



http://users.cis.fiu.edu/~liux/research/papers/topo-wsc03.pdf

Grand Challenge – What we have …

- Visualization of the Internet in 1994 (topology & traffic)



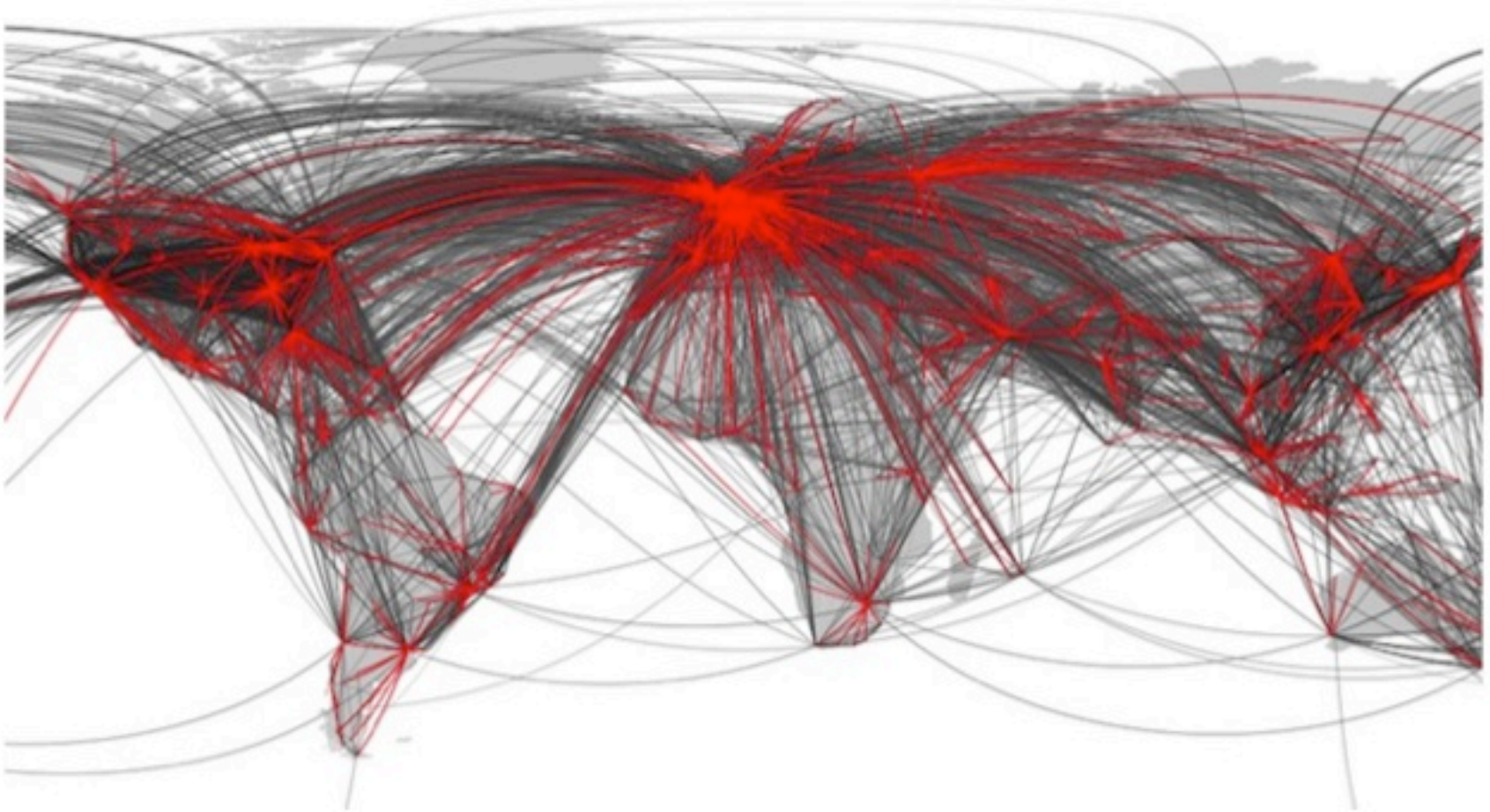
Grand Challenge – What we want …

- Visualization of the Internet in 2015 (topology & traffic)??

Why is this (very) hard?

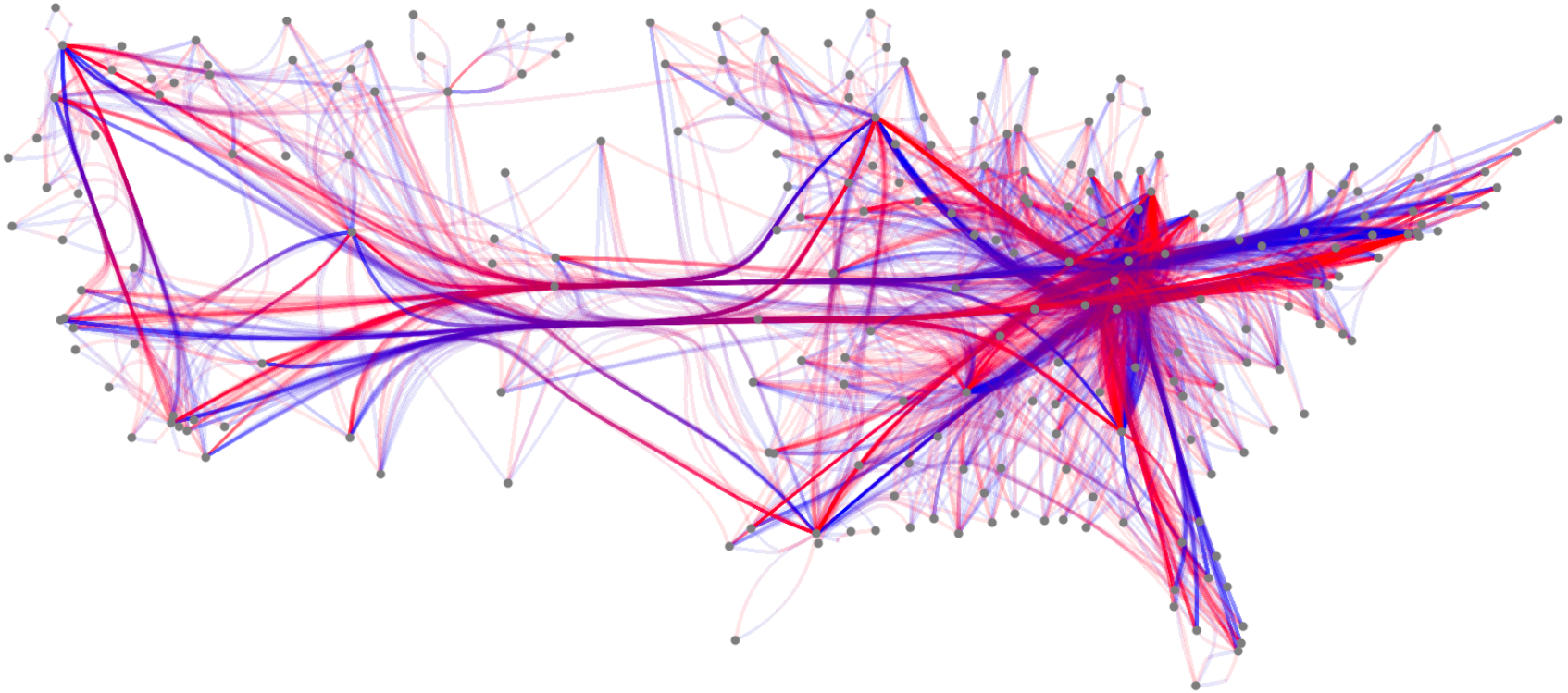
- What topology?
  - AS-level Internet topology
  - Physical Internet topology
  - Main focus of some 15-20 years of Internet topology research
  - We know much less about this than we thought we did …
- What traffic?
  - Inter-domain traffic (AS traffic matrix)
  - How much traffic is exchanged between any pair of ASes?
  - We know next to nothing about this …
- What visualization?
  - ????

An analog: Worldwide airline system …



http://www.northwestern.edu/newscenter/stories/2012/06/complex-networks-skeletons.html

… or US airline traffic



http://vis.stanford.edu/files/2011-DividedEdgeBundling-InfoVis.pdf

Conclusion

- Past 15-20 years of research on the Internet's AS-level (and router-level) topology
  - Example of Grossman's (mis)quote of H.L Mencken:

    "Complex problems have simple, easy-to-understand wrong answers."

- The future of Internet research with "big (Internet) data"
  - Learn form past mistakes!
  - "What we want to measure is typically not what we can measure!"
  - Details matter!
  - Domain knowledge is critical!
  - There is no "free lunch"!

Collaborators

- TU Berlin
  - Anja Feldmann, George Smaragdakis, Philipp Richter
- Akamai/Duke University
  - Nikos Chatzis, Jan Boettger
  - Bruce Maggs, Bala Chandrasekaran
- Northwestern University
  - Fabian Bustamante, Mario Sanchez
- University of Oregon (joint NSF grant, 2013/15)
  - Reza Rejaie, Reza Motamedi
- AT&T Labs-Research
  - Balachander Krishnamurthy, Jeff Erman
- University of Adelaide (joint ARC grant 2011/14)
  - Matt Roughan
- University of Wisconsin
  - Paul Barford, Ramakrishnan Durairajan

Thanks!

Questions?