Bias in Internet Measurement Infrastructure

Can we fix our broken glasses?

Pavlos Sermpezis
DATA&WEB SCIENCE LABORATORY

Lars Prehn
mpi max planck institut informatik

Emile Aben
RIPE NCC
RIPE NETWORK COORDINATION CENTRE

1st prize in RIPE Labs Article Competition

kentik fastly

Measuring RPKI ROV adoption with NetFlow

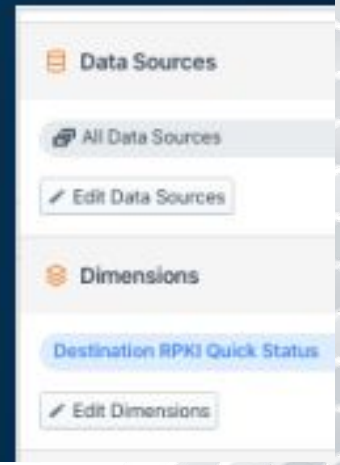
Doug Madory, dmadory@kentik.com, Kentik
Job Snijders, job@fastly.com, Fastly

talk by Doug Madory @ RIPE84 plenary Monday 17 May

Kentik's perspective can deepen understanding of RPKI

Kentik has over 300 customers and almost half have opted-in to the use of their data as part of aggregate analysis.

- Note: analysis is subject to biases of the customer set which includes (NSPs, CDNs and enterprises) and is skewed toward the US.



Internet measurement infrastructure & platforms

RIPE NCC
RIPE Atlas

https://atlas.ripe.net/

- data plane measurements
- > 11,000 probes & anchors
- in > 3000 ASNs

RIPE NCC RIS
RIPE NETWORK COORDINATION CENTRE

https://ris-live.ripe.net/

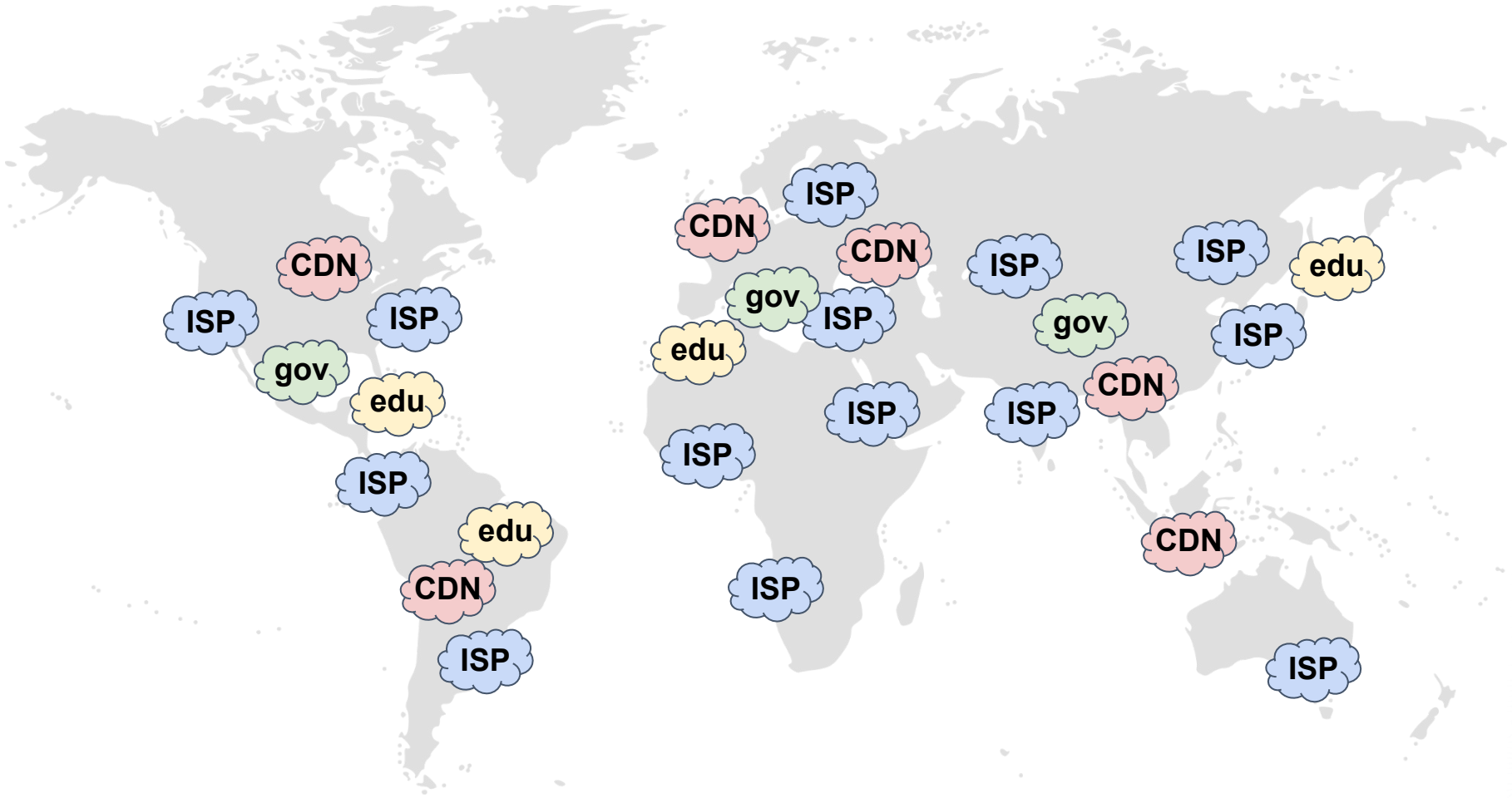
- BGP RIBs & updates
- 27 route collectors
- peering with > 500 ASNs

ROUTE VIEWS 6447

http://www.routeviews.org

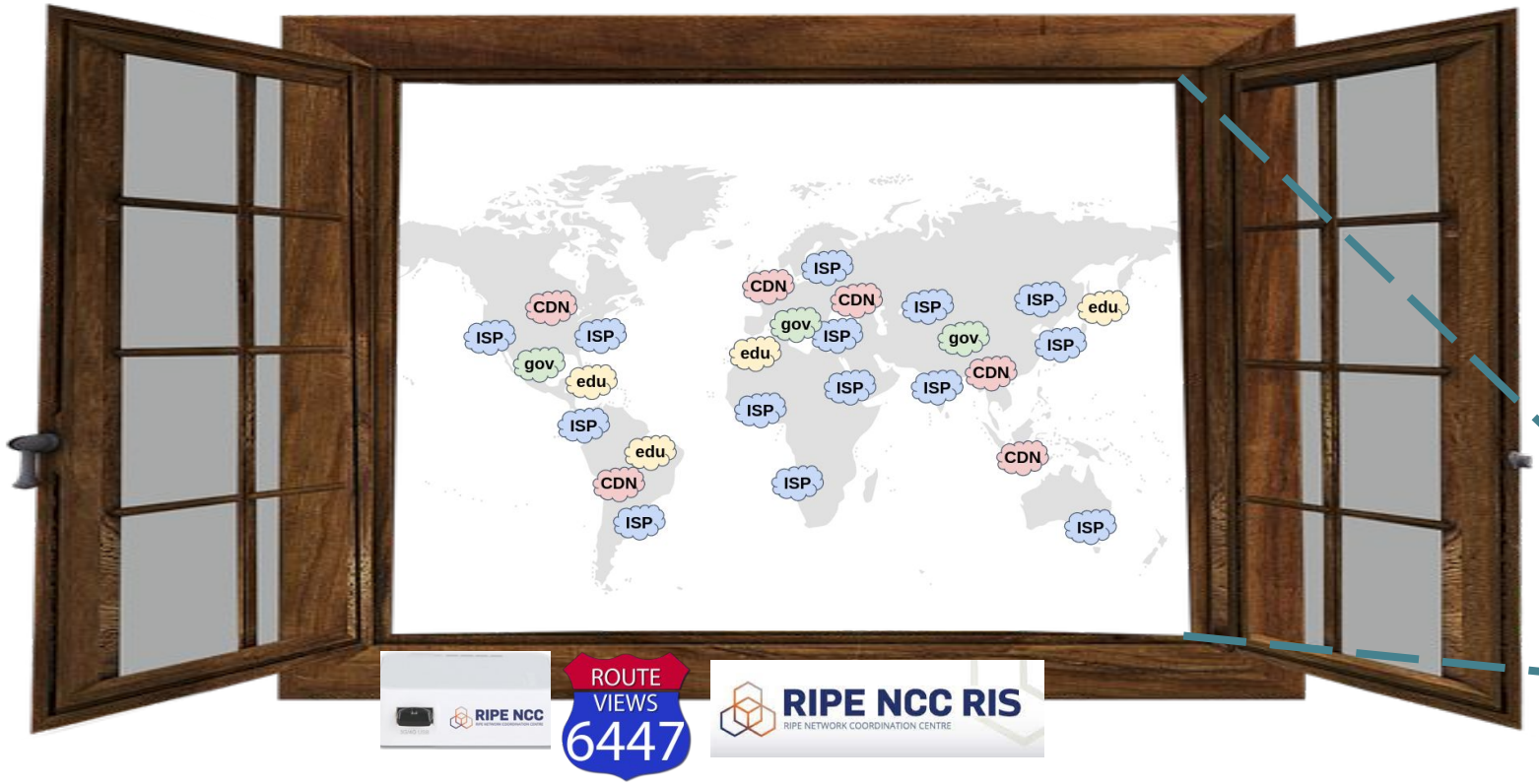
- BGP RIBs & updates
- 36 route collectors
- peering with > 300 ASNs

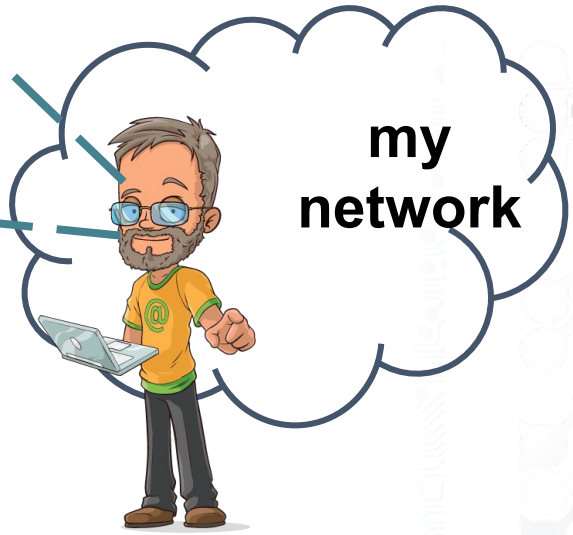
The Internet



Pavlos Sermpezis sermpezis@csd.auth.gr @PavlosSermpezis

Measurement platforms: a window to the Internet





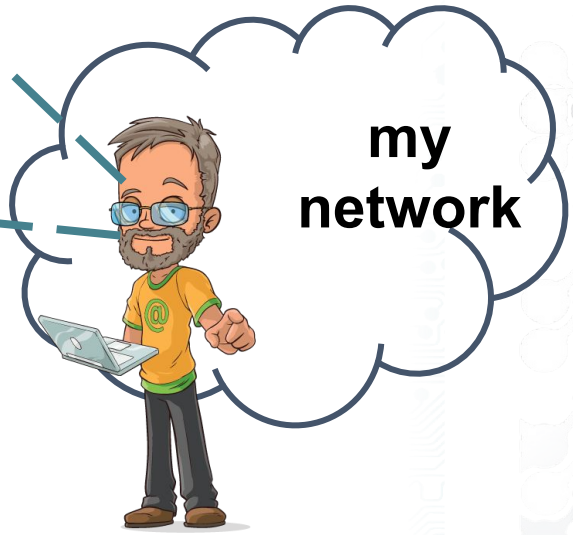
Pavlos Sermpezis sermpezis@csd.auth.gr @PavlosSermpezis

… but, in practice: a ***stained glass*** window

RIPE NCC

ROUTE VIEWS 6447

RIPE NCC RIS
RIPE NETWORK COORDINATION CENTRE



Pavlos Sermpezis sermpezis@csd.auth.gr @PavlosSermpezis

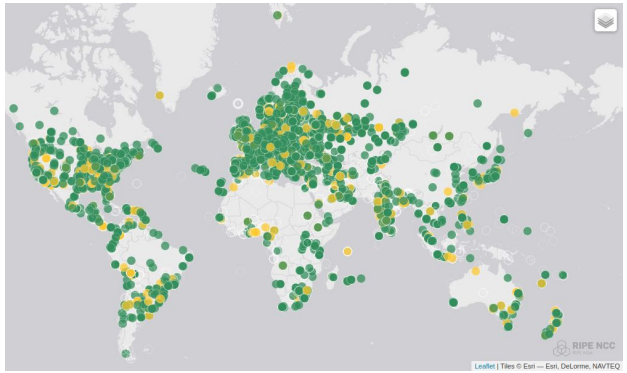
The “stained glass” view == Bias

not all network types can be equally seen by the platforms

→ our view of the Internet is **biased**

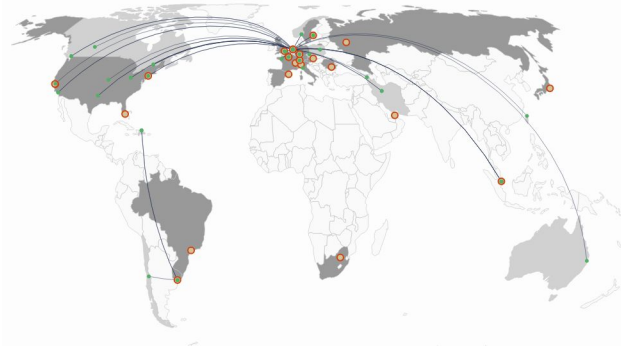
Example 1 (location bias)

- RIPE Atlas & RIPE RIS have more probes/peers in Europe



RIPE Atlas probes

https://atlas.ripe.net/results/maps/network-coverage/

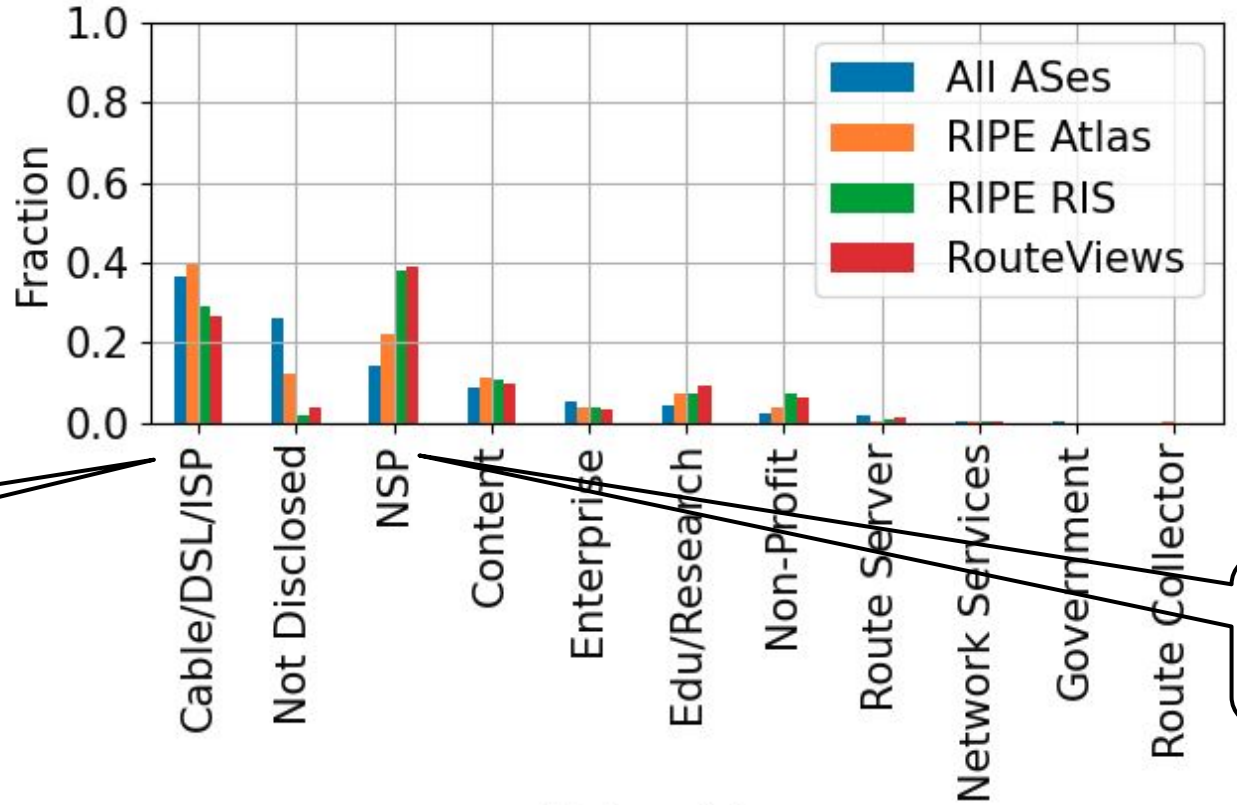


RIPE RIS route collectors

https://observablehq.com/@emileaben/ris-route-collectors-and-peer-locations

Example 2 (network-type bias)

- Peers of **RIPE RIS** and **RouteViews** do not equally represent all network types

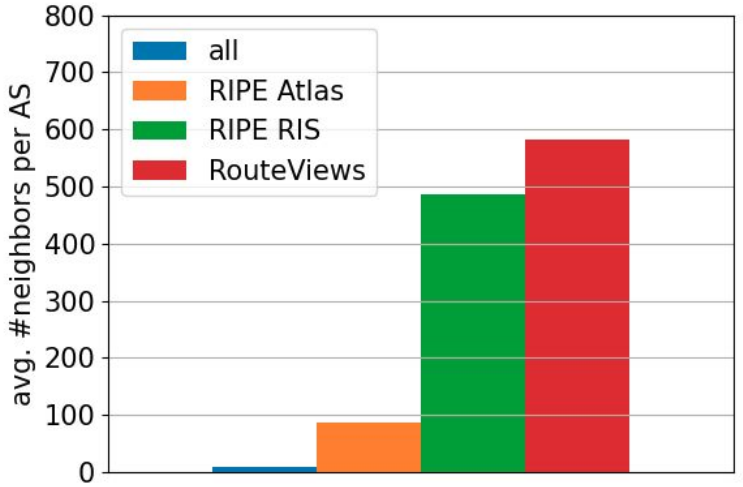


Cable/DSL/ISP are ***under-represented***

NSPs are ***over-represented***

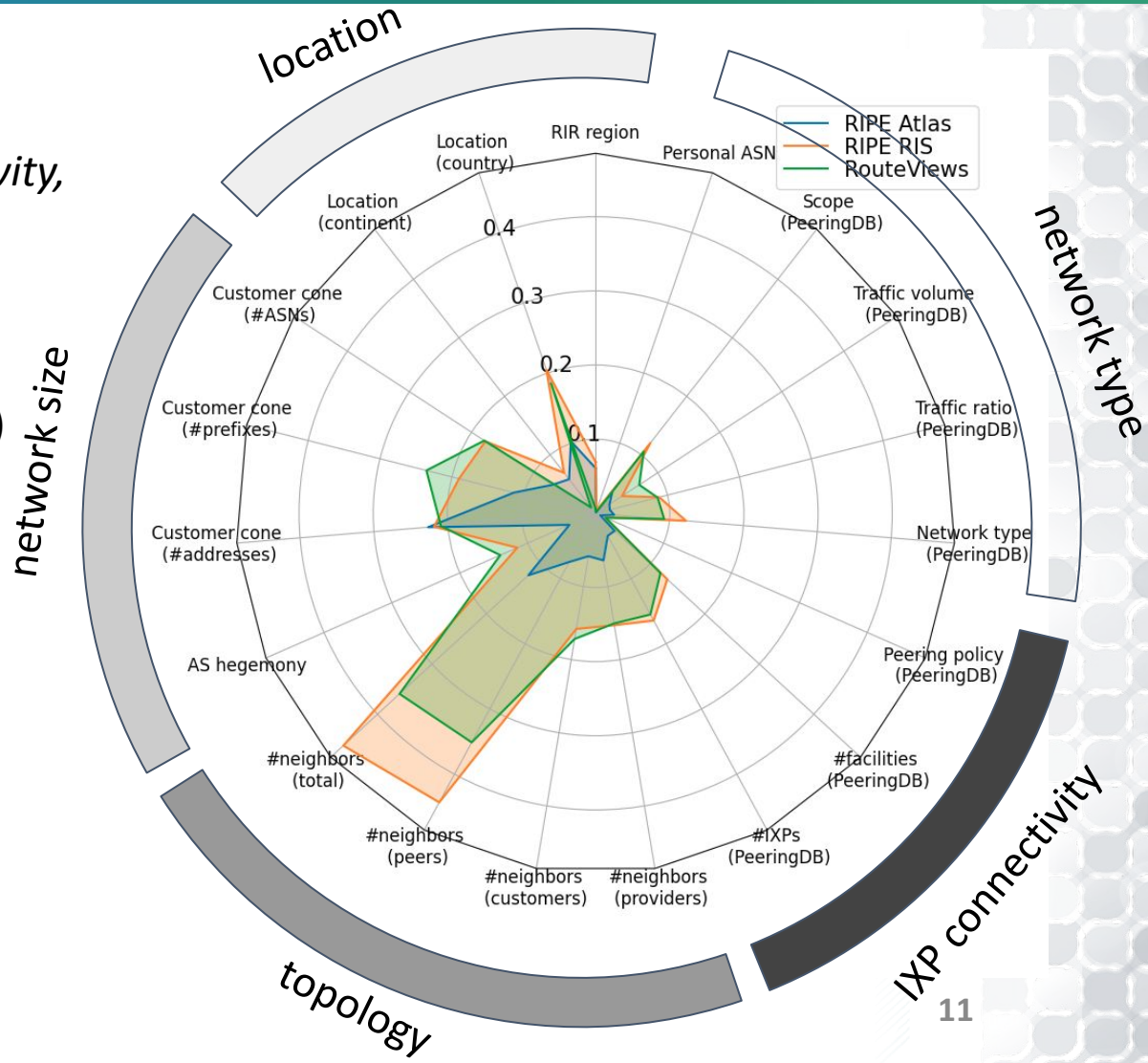
Example 3 (topological bias)

- ASes that feed to **RIPE RIS**/**RouteViews** or host **RIPE Atlas** probes, are networks that typically peer with many other networks



Quantifying & visualizing bias

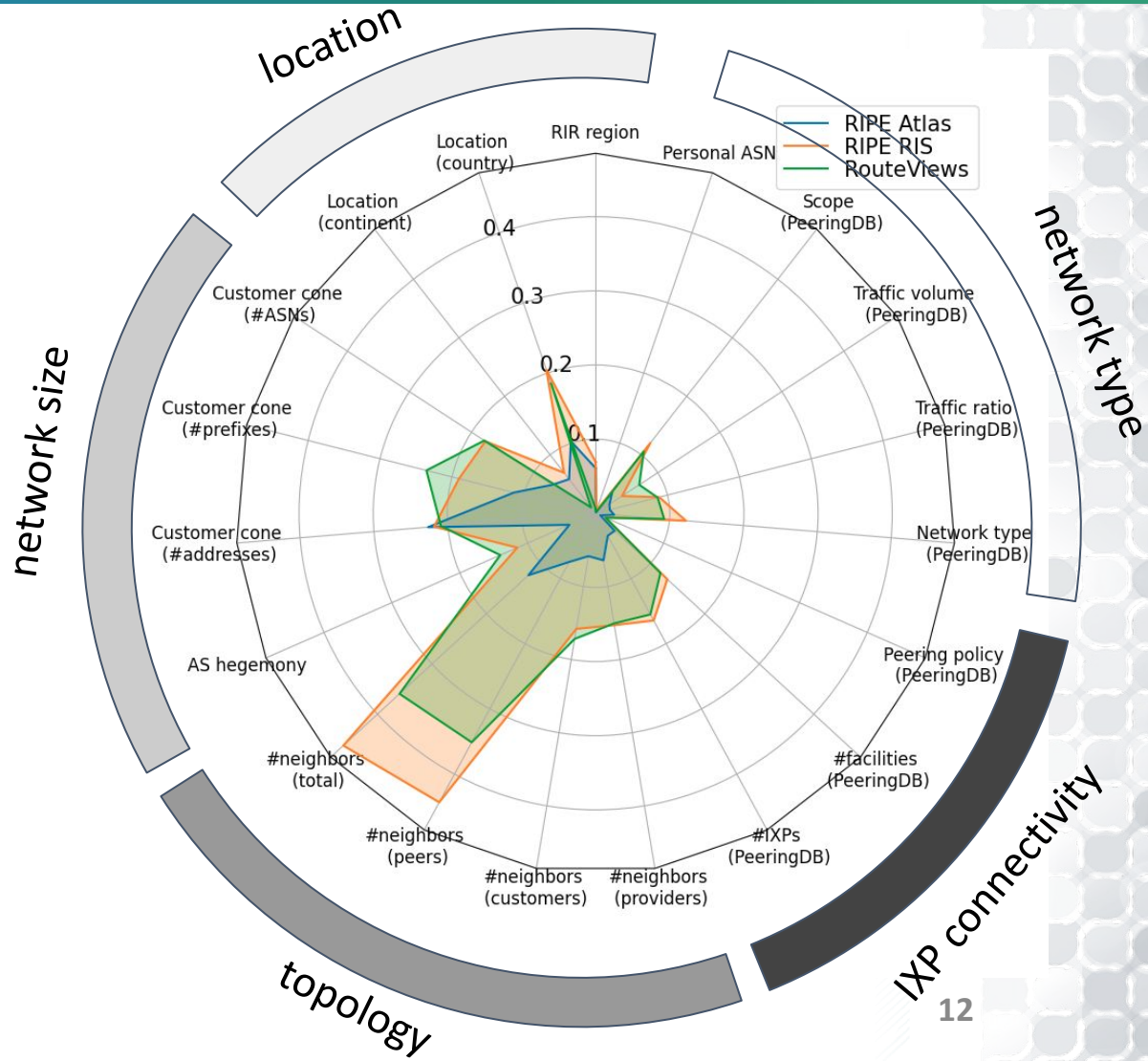
- Many dimensions of bias
  - *location, network size, topology, IXP connectivity, network type, etc.*
- Bias score per dimension
  - a value between 0 (low bias) and 1 (high bias)
- Radar plot of bias
  - each radius ⟶ a bias dimension
  - colored lines/areas ⟶ bias score
  - high bias ⟶ far from center



Pavlos Sermpezis sermpezis@csd.auth.gr @PavlosSermpezis

Bias in Internet Measurement Infrastructure

RIPE Atlas is *significantly less biased* than **RIPE RIS** and **RouteViews** in almost all dimensions

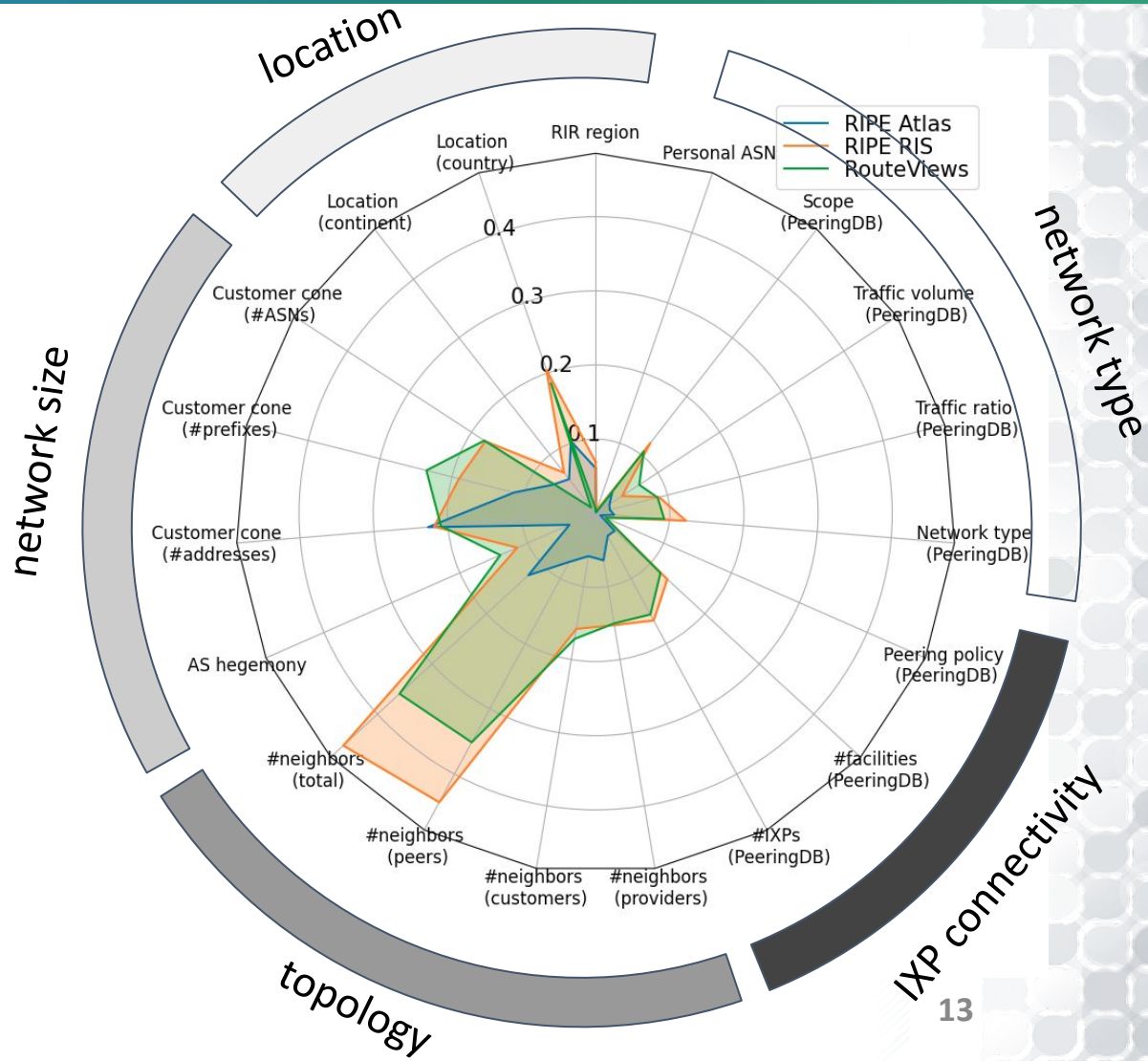


Pavlos Sermpezis sermpezis@csd.auth.gr @PavlosSermpezis

Bias in Internet Measurement Infrastructure

RIPE Atlas is *significantly less biased* than **RIPE RIS** and **RouteViews** in almost all dimensions

RIPE RIS has *high topology bias* (due to route collectors at IXPs) and *high network size bias* (peers are large networks)



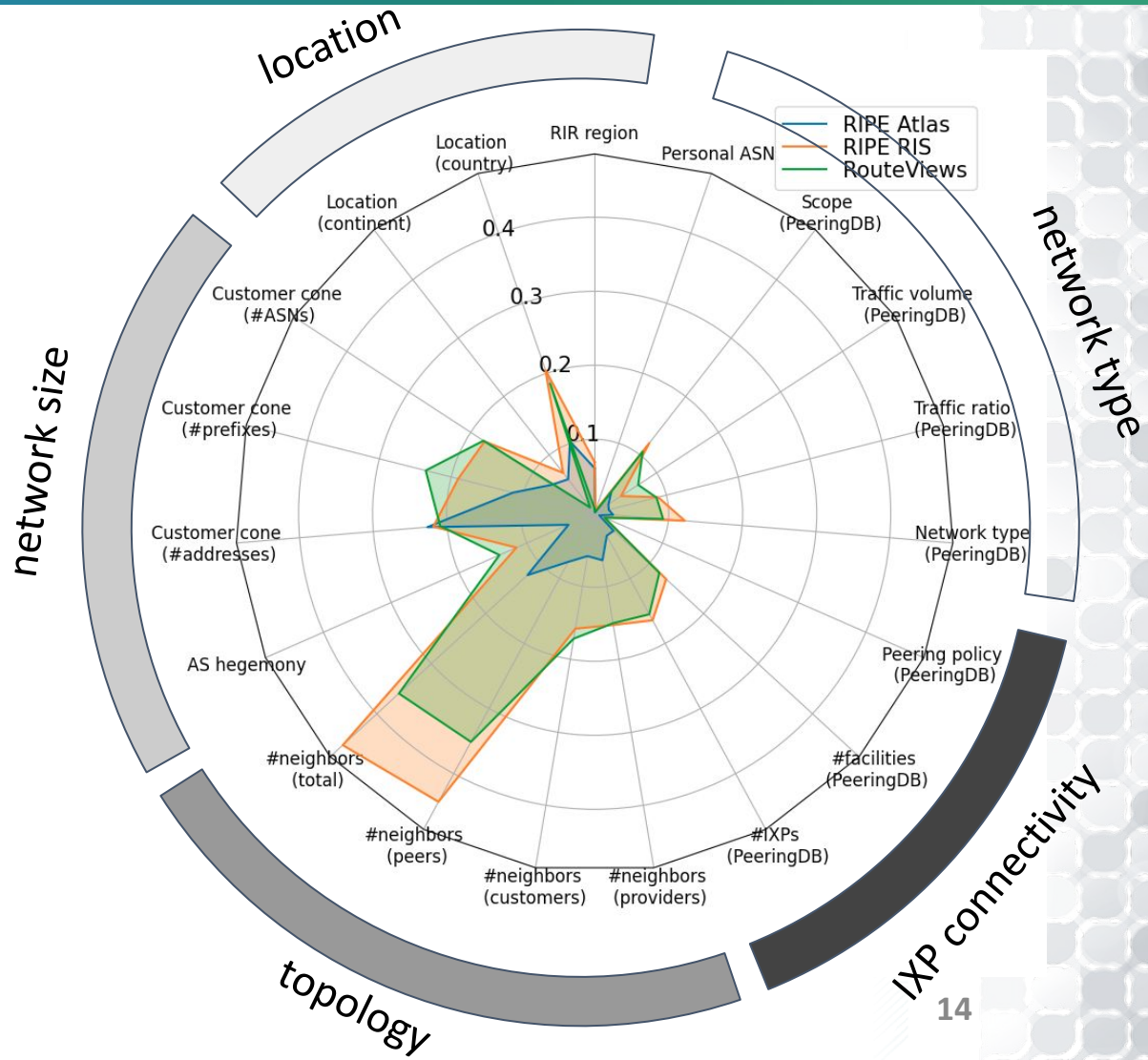
Pavlos Sermpezis sermpezis@csd.auth.gr @PavlosSermpezis

Bias in Internet Measurement Infrastructure

RIPE Atlas is *significantly less biased* than **RIPE RIS** and **RouteViews** in almost all dimensions

RIPE RIS has *high topology bias* (due to route collectors at IXPs) and *high network size bias* (peers are large networks)

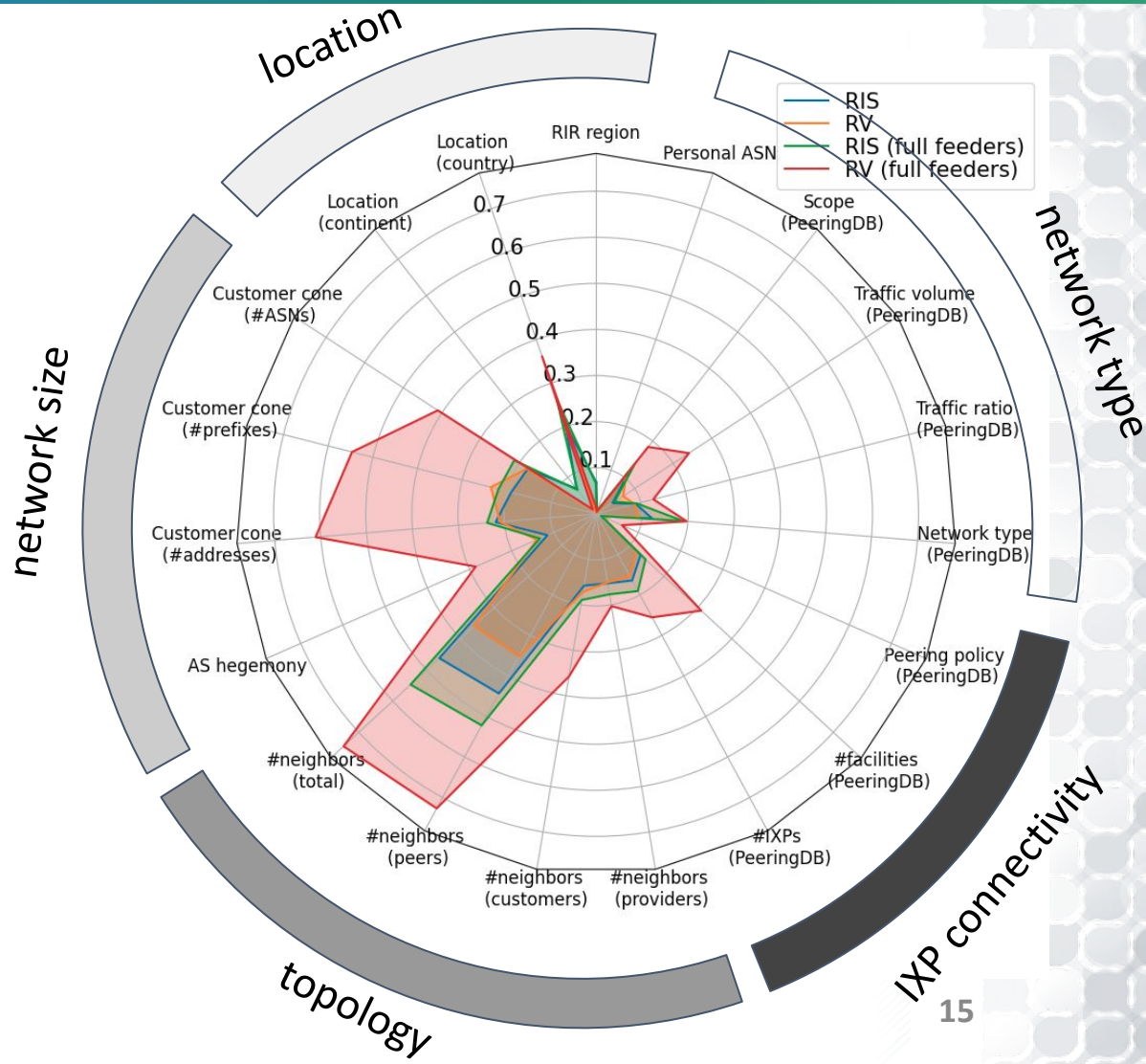
RIPE Atlas, **RIPE RIS** and **RouteViews** have relatively *low network-type bias*



Pavlos Sermpezis sermpezis@csd.auth.gr @PavlosSermpezis

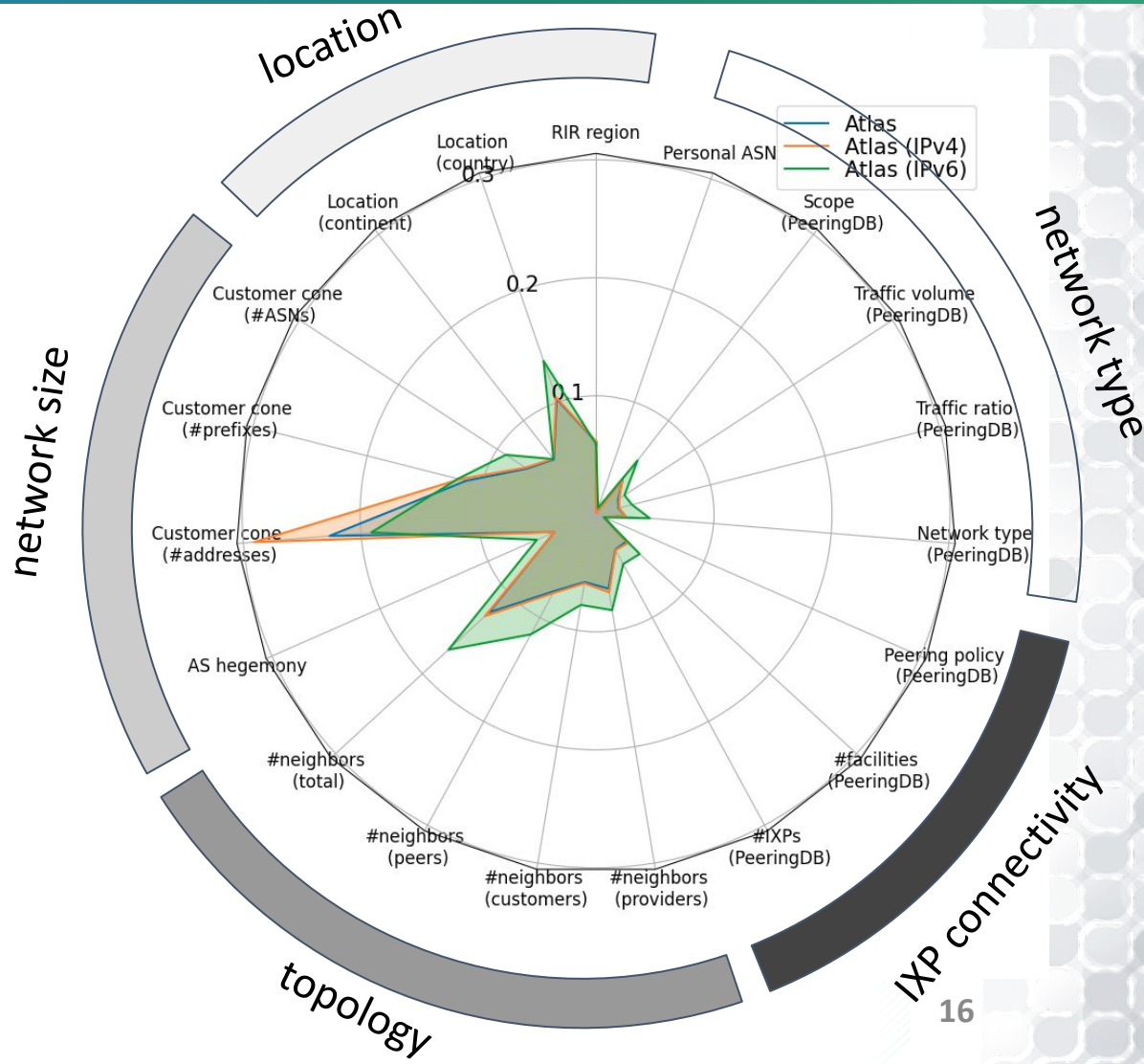
Bias in Internet Measurement Infrastructure

Full feeds are *more biased*



Bias in Internet Measurement Infrastructure

IPv6 Atlas probes are *more biased* than **IPv4**



Pavlos Sermpezis sermpezis@csd.auth.gr @PavlosSermpezis

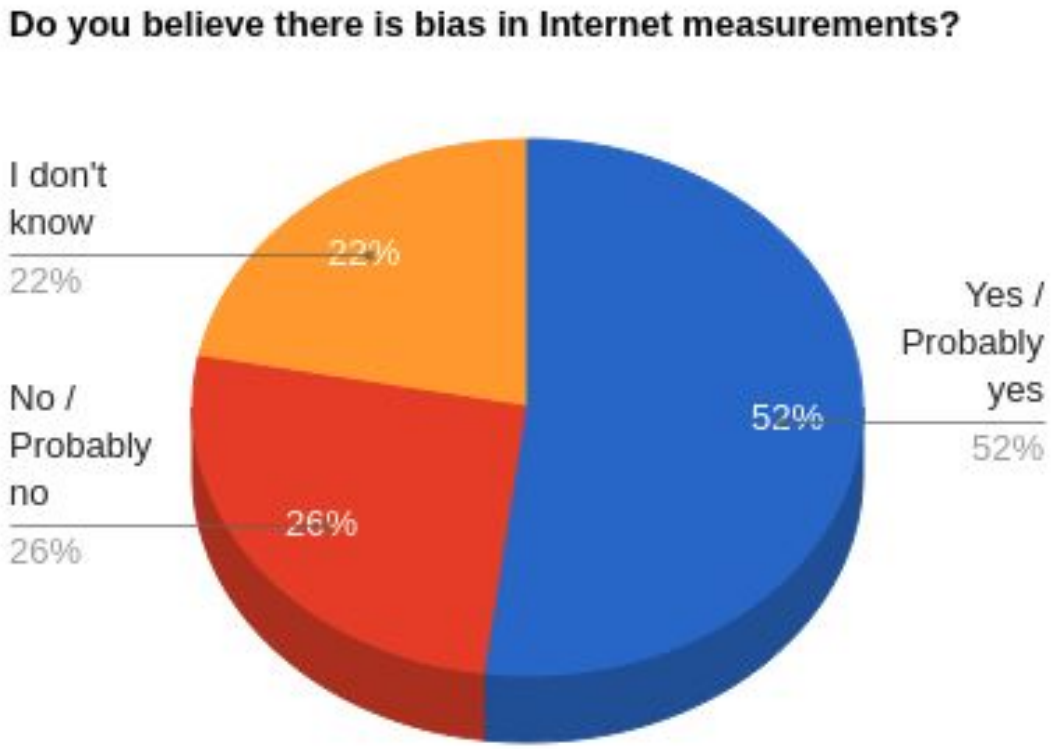
So… what?

- Should I care? → Yes! Bias may affect the insights you get from your measurements
  - *e.g., "Estimating the Impact of BGP Prefix Hijacking", IFIP Networking, 2021 [link]*

- Be aware of bias! Carefully interpret your results
  - "Which dimensions affect my measurements?"
    [tell us about your use cases @ our survey]
  - "Is there bias in my dimensions?"
    [check the radar plot & find out more details @ our project's website]

Do people know?

- Not all people know! → our main goal: raise awareness & deepen our understanding



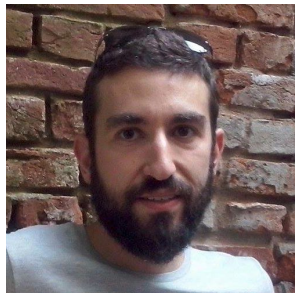
How can we fix it?

- Option 1: design bias-free methodologies to select Atlas probes or RIS/RouteViews feeds
- Option 2: extend the infrastructure in a bias-free way

@ the AI4NetMon project

Q&A

- Do you find our work useful?
- Tell us what you **don't like** about our work and why!
- We'll be around at RIPE84, let's chat!

Pavlos 

Emile