

AI4NetMon tool: A Web app to Analyze bias in Internet Measurements

Pavlos Sermpezis & Sofia Kostoglou

joint work with: Lars Prehn, Marcel Flores, Athena Vakali, Emile Aben

Data & Web Science Laboratory (Datalab), <https://datalab.csd.auth.gr/>

Computer Science Dept., Aristotle University of Thessaloniki



RIPE NCC
RIPE NETWORK COORDINATION CENTRE

Edgio



Bias in data: a motivational example

- Assume an entire population of 100 people
 - 50 men, 50 women
 - 70 from country A, 30 from country B
- We do a survey with 10 participants
 - 8 men, 2 women
 - 8 from country A, 2 from country B

	Men	Women	Country A	Country B
Entire population	50%	50%	70%	30%
Survey sample	80%	20%	80%	20%

- Is there bias? → Yes! difference in the gender/country distributions between population & sample
- Is bias the same along gender/country? → No! sample is more biased wrt. the gender dimension
- Is bias a problem? → It depends!
 - Goal: estimate the average population height (gender bias **is** a problem, country bias **may be** a problem)
 - Goal: calculate % of native spoken languages (gender bias **is not** a problem, country bias **is** a problem)



Internet measurement platforms



RIPE NCC

RIPE Atlas



<https://atlas.ripe.net/>

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



<http://www.routeviews.org>

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

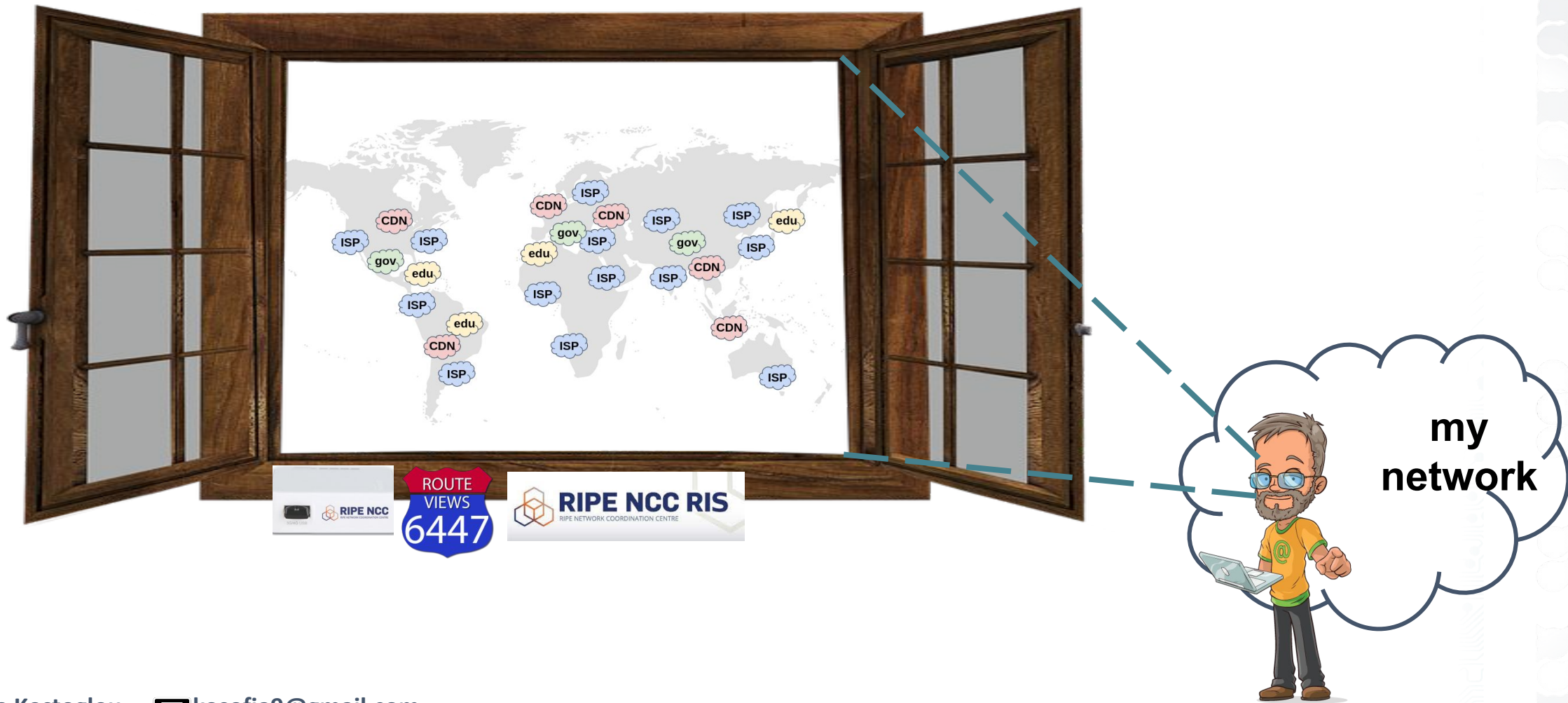


<https://ris-live.ripe.net/>

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



Measurement platforms: a window to the Internet





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





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/>



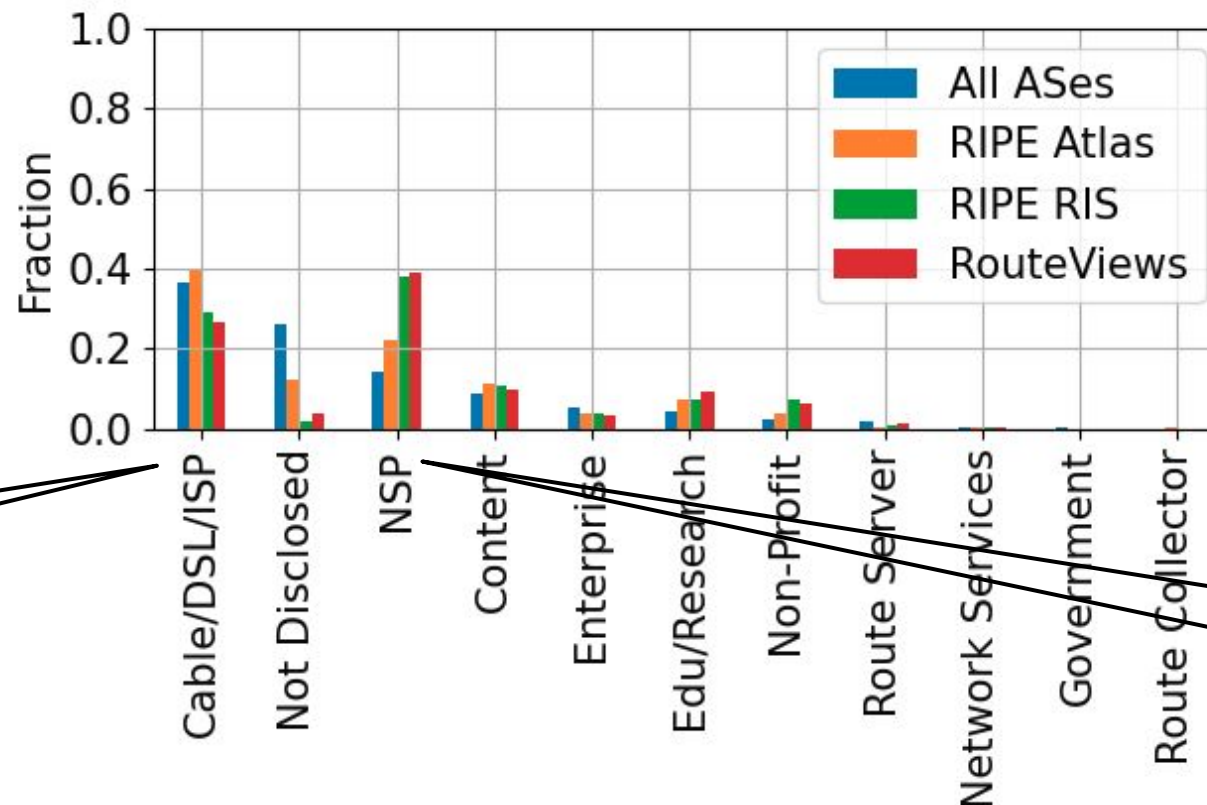
RIPE RIS route collectors

<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

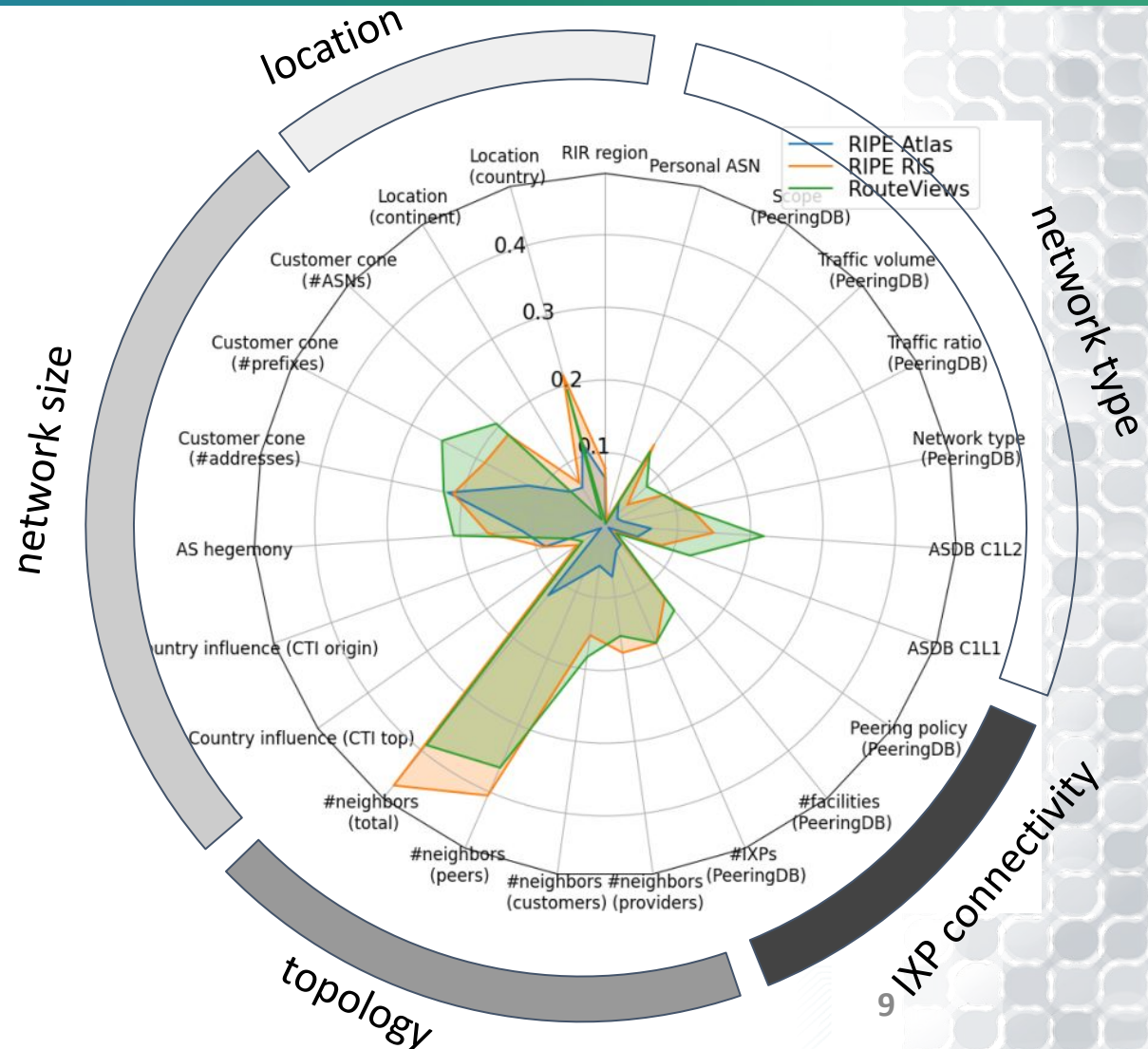


Cable/DSL/ISP are *under-represented*

NSPs are *over-represented*

Quantifying bias

- Many dimensions of bias
 - *location, network size, topology, IXP connectivity, network type, etc.*
- Bias score per dimension
 - Bias == Difference between two distributions (**all networks** vs. **networks with vantage points**)
 - **Bias score**: Kullback-Leibler divergence metric
 - i.e, 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





Dataset, code, API, Web app

- AI4NetMon project <https://ai4netmon.csd.auth.gr/>
 - You can find all the information about the project!
- Code & Data @ GitHub <https://github.com/sermpezis/ai4netmon/>
- API <https://ai4netmon.csd.auth.gr/api/>
 - Documentation @ GitHub
- Web app <https://app-ai4netmon.csd.auth.gr/>





Find bias at a glance...

- Give a set of vantage points (ASNs)
- See if the set is representative through visualizations



Web app: "Show me the bias"

- Available at <https://app-ai4netmon.csd.auth.gr/>



Select a custom set of vantage points. In the boxes below, add a list of ASNs/Probe IDs (only numbers, separated with commas, no spaces; e.g., 174,1299,3333)

Select the type of list of numbers

- ASNs probe IDs

Custom Set #1 (ASNs)

Custom Set #2 (ASNs)

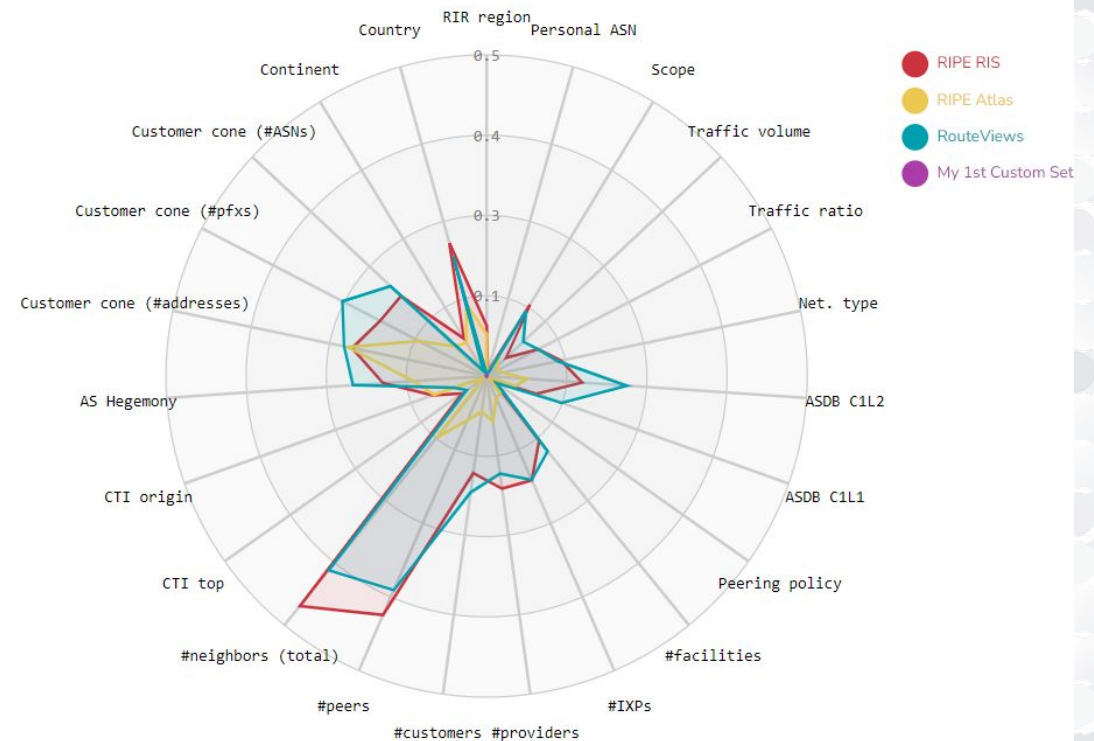
Custom Set #3 (ASNs)

Monitor sets

- RIPE RIS RIPE Atlas RouteViews My 1st Custom Set
 My 2nd Custom Set My 3rd Custom Set

Bias dimensions

- RIR region Country Continent Customer cone (#ASNs)
 Customer cone (#pfxs) Customer cone (#addresses)
 AS Hegemony CTI origin CTI top #neighbors (total)
 #peers #customers #providers #IXPs #facilities
 Peering policy ASDB C1L1 ASDB C1L2 Net. type
 Traffic ratio Traffic volume Scope Personal ASN



Web app: "Show me the bias"

- Available at <https://app-ai4netmon.csd.auth.gr/>



Select a custom set of vantage points. In the boxes below, add a list of ASNs/Probe IDs (only numbers, separated with commas, no spaces; e.g., 174,1299,3333)

Select the type of list of numbers: ASNs probe IDs

Custom Set #1 (ASNs): 21232, 8222, 6939, 1090, 2000, 1110, 4444

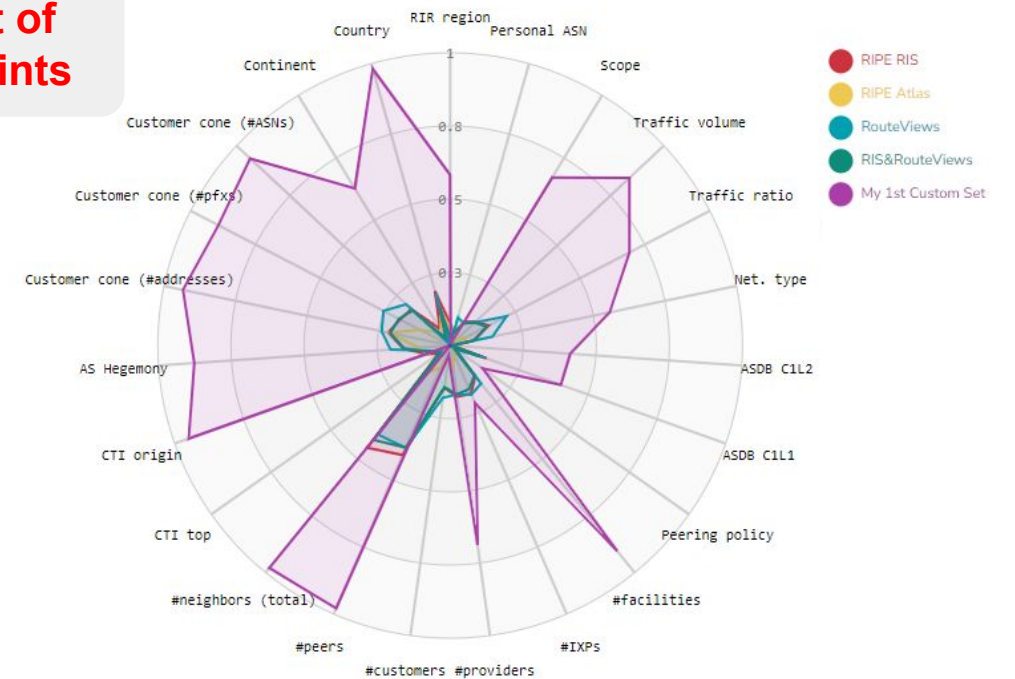
Custom Set #2 (ASNs):

Custom Set #3 (ASNs):

Monitor sets: RIPE RIS RIPE Atlas RouteViews RIS&RouteViews My 1st Custom Set My 2nd Custom Set My 3rd Custom Set

Bias dimensions: RIR region Country Continent Customer cone (#ASNs) Customer cone (#pfxs) Customer cone (#addresses) AS Hegemony CTI origin CTI top #neighbors (total) #peers #customers #providers #IXPs #facilities Peering policy ASDB C1L1 ASDB C1L2 Net. type Traffic ratio Traffic volume Scope Personal ASN

Custom set of vantage points





Web app: "Show me the bias"

- Available at <https://app-ai4netmon.csd.auth.gr/>



Select a custom set of vantage points. In the boxes below, add a list of ASNs/Probe IDs (only numbers, separated with commas, no spaces; e.g., 174,1299,3333)

Select the type of list of numbers ASNs probe IDs

Custom Set #1 (ASNs)

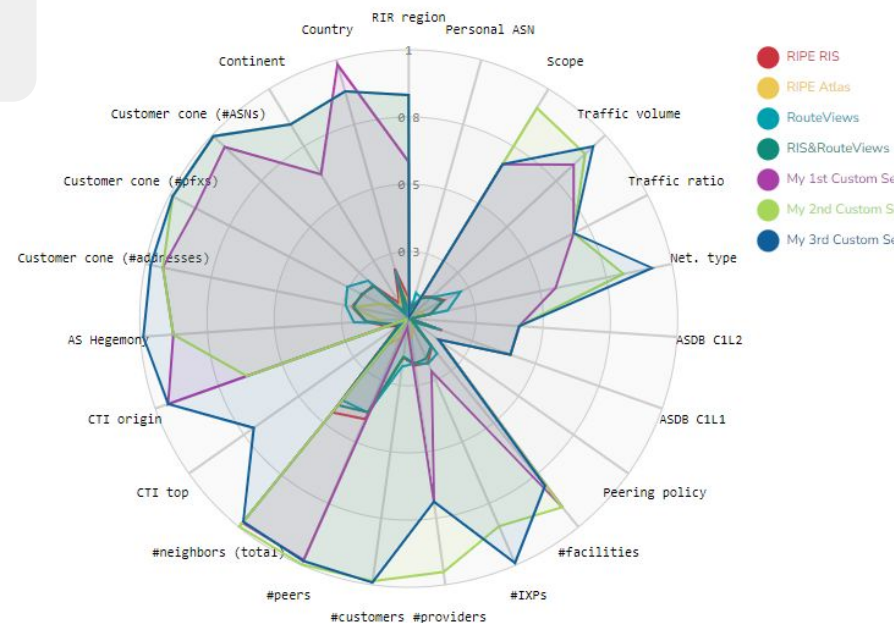
Custom Set #2 (ASNs)

Custom Set #3 (ASNs)

Monitor sets RIPE RIS RIPE Atlas RouteViews RIS&RouteViews My 1st Custom Set My 2nd Custom Set My 3rd Custom Set

Bias dimensions RIR region Country Continent Customer cone (#ASNs) Customer cone (#pfxs) Customer cone (#addresses) AS Hegemony CTI origin CTI top #neighbors (total) #peers #customers #providers #IXPs #facilities Peering policy ASDB C1L1 ASDB C1L2 Net. type Traffic ratio Traffic volume Scope Personal ASN

...up to 3 custom sets of vantage points





Web app: "Show me the bias"

- Available at <https://app-ai4netmon.csd.auth.gr/>



Select a custom set of vantage points. In the boxes below, add a list of ASNs/Probe IDs (only numbers, separated with commas, no spaces; e.g., 174,1299,3333)

Select the type of list of numbers ASNs probe IDs

Custom Set #1 (ASNs)

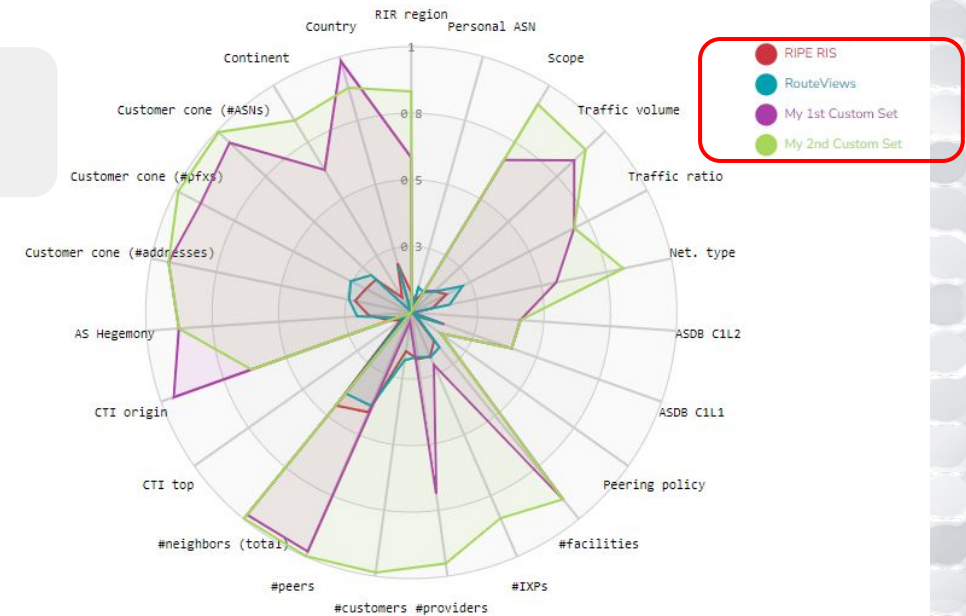
Custom Set #2 (ASNs)

Custom Set #3 (ASNs)

Monitor sets RIPE RIS RIPE Atlas RouteViews RIS&RouteViews My 1st Custom Set My 2nd Custom Set My 3rd Custom Set

Bias dimensions RIR region Country Continent Customer cone (#ASNs) Customer cone (#pfxs) Customer cone (#addresses) AS Hegemony CTI origin CTI top #neighbors (total) #peers #customers #providers #IXPs #facilities Peering policy ASDB C1L1 ASDB C1L2 Net. type Traffic ratio Traffic volume Scope Personal ASN

select sets to show in the radar-plot



Web app: "Show me the bias"

- Available at <https://app-ai4netmon.csd.auth.gr/>



Select a custom set of vantage points. In the boxes below, add a list of ASNs/Probe IDs (only numbers, separated with commas, no spaces; e.g., 174,1299,3333)

Select the type of list of numbers ASNs probe IDs

Custom Set #1 (ASNs)

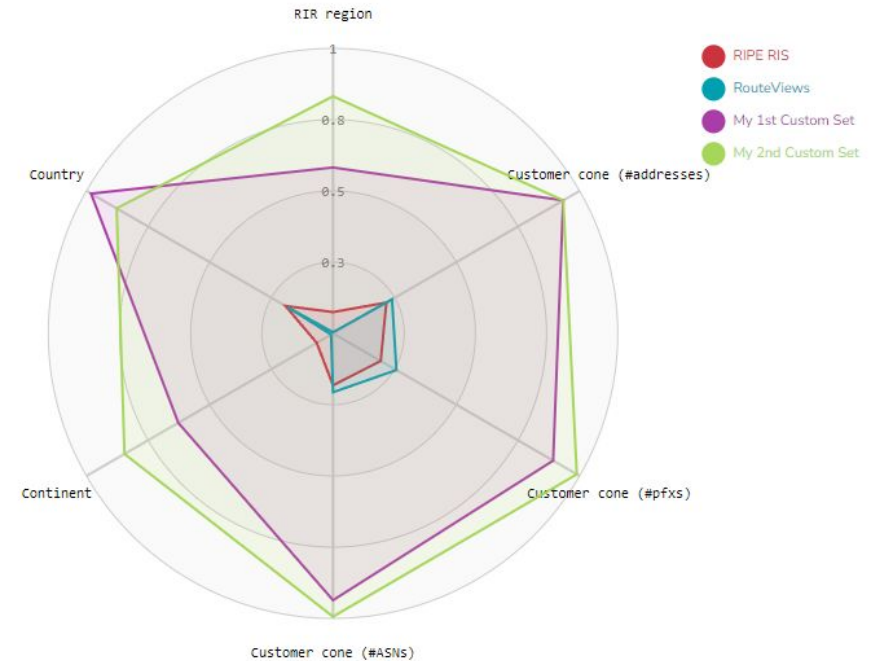
Custom Set #2 (ASNs)

Custom Set #3 (ASNs)

Monitor sets RIPE RIS RIPE Atlas RouteViews RIS&RouteViews My 1st Custom Set My 2nd Custom Set My 3rd Custom Set

Bias dimensions RIR region Country Continent Customer cone (#ASNs) Customer cone (#pfxs) Customer cone (#addresses) AS Hegemony CTI origin CTI top #neighbors (total) #peers #customers #providers #IXPs #facilities Peering policy ASDB C1L1 ASDB C1L2 Net. type Traffic ratio Traffic volume Scope Personal ASN

keep only subset of bias dimensions





But...why does this bias exist??

- See in more detail through distribution plots
- Understand why sets are not representative in each one of the different bias dimensions
- Check the bias causes through relative distribution differences

Web app: "Show me the bias"

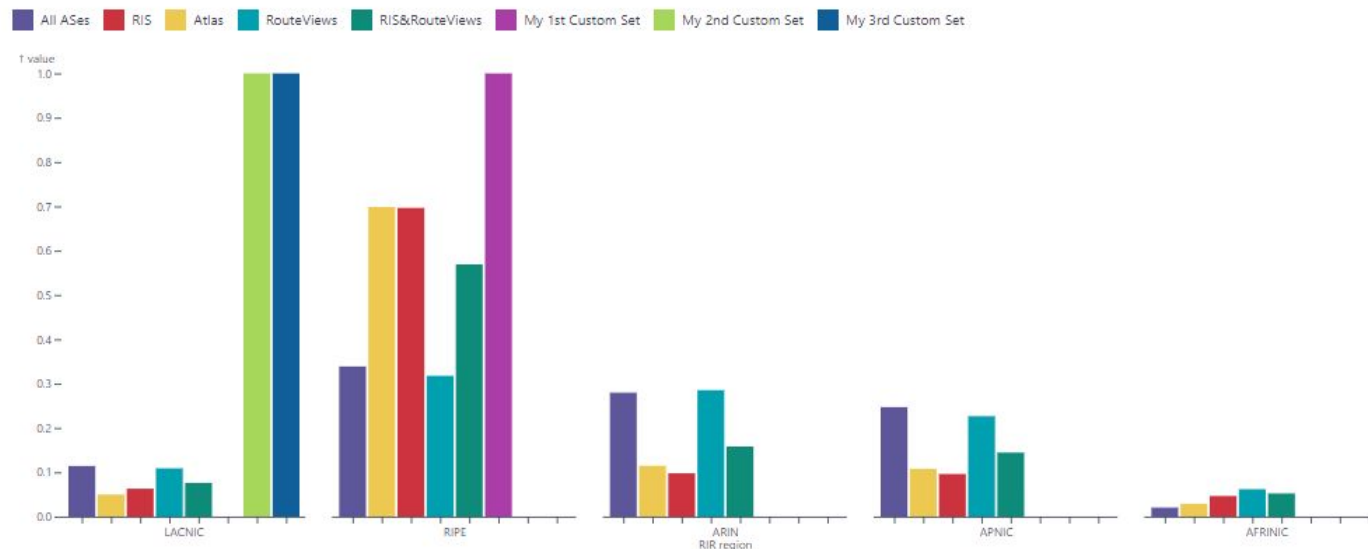
- Available at <https://app-ai4netmon.csd.auth.gr/>



!! Bias = difference in "population" (All ASes) and "sample" distributions (monitors sets).

Bias dimensions

check distribution per bias dimension





Web app: "Show me the bias"

- Available at <https://app-ai4netmon.csd.auth.gr/>



Dive deeper into bias "causes": for all different values of each dimension, shed light into the relative change of distributions

Vantage point set: My 1st Custom Set

Filter by: Bias Dimension, Value or Difference

Search: 199 results

Bias Dimension	Value	Difference %
Customer cone (#ASNs)	1.0-3.0	-93.836
Customer cone (#ASNs)	3.0-9.0	96.274
Location (continent)	North America	-30.224
Location (continent)	Europe	70.523

different value-bins for num values (pointing to 3.0-9.0)

different values for categ values (pointing to Europe)



Web app: “Show me the bias”

- Available at <https://app-ai4netmon.csd.auth.gr/>



Check the largest and smallest percentage difference in the bias causes

Top Recommendation - Largest Difference %

The characteristic **Traffic ratio (PeeringDB)** with value **Not Disclosed** is the most **overrepresented** in the selected set, by a percentage of **30.31 %**

Top Recommendation - Smallest Difference %

The characteristic **#neighbors (peers)** with value **1.0-3.0** is the most **underrepresented** in the selected set, by a percentage of **-81.2564 %**



Detailed page: Route Collector Measurements

- Check all the bias analysis and tools, especially for route collectors



Web app: Route collectors measurements

- Available at <https://app-ai4netmon.csd.auth.gr/>



Select a custom set of vantage points. In the boxes below, add a list of ASNs/Probe IDs (only numbers, separated with commas, no spaces; e.g., 174,1299,3333)

Select the type of list of numbers
 ASNs probe IDs

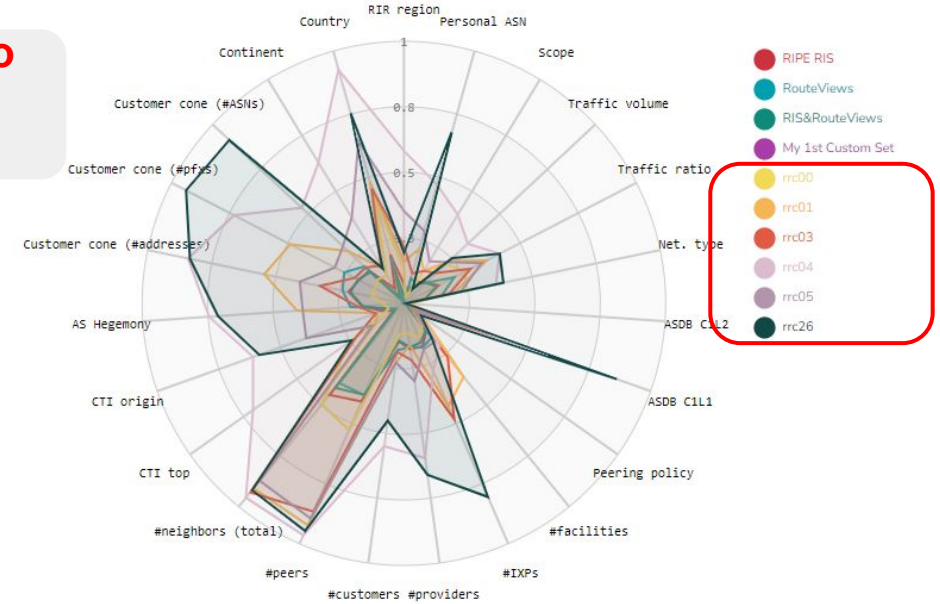
Custom Set #1 (ASNs)

Custom Set #2 (ASNs)

Custom Set #3 (ASNs)

select RRCs to show and compare

- Monitor sets
- RIPE RIS RouteViews RIS&RouteViews My 1st Custom Set My 2nd Custom Set My 3rd Custom Set rrc00 rrc01 rrc03
 - rrc04 rrc05 rrc06 rrc07 rrc10 rrc11 rrc12 rrc13 rrc15 rrc16 rrc18 rrc20 rrc21 rrc22
 - rrc23 rrc24 rrc25 rrc26
- Bias dimensions
- RIR region Country Continent Customer cone (#ASNs) Customer cone (#pfxs) Customer cone (#addresses) AS Hegemony
 - CTI origin CTI top #neighbors (total) #peers #customers #providers #IXPs #facilities Peering policy ASDB C1L1
 - ASDB C1L2 Net. type Traffic ratio Traffic volume Scope Personal ASN



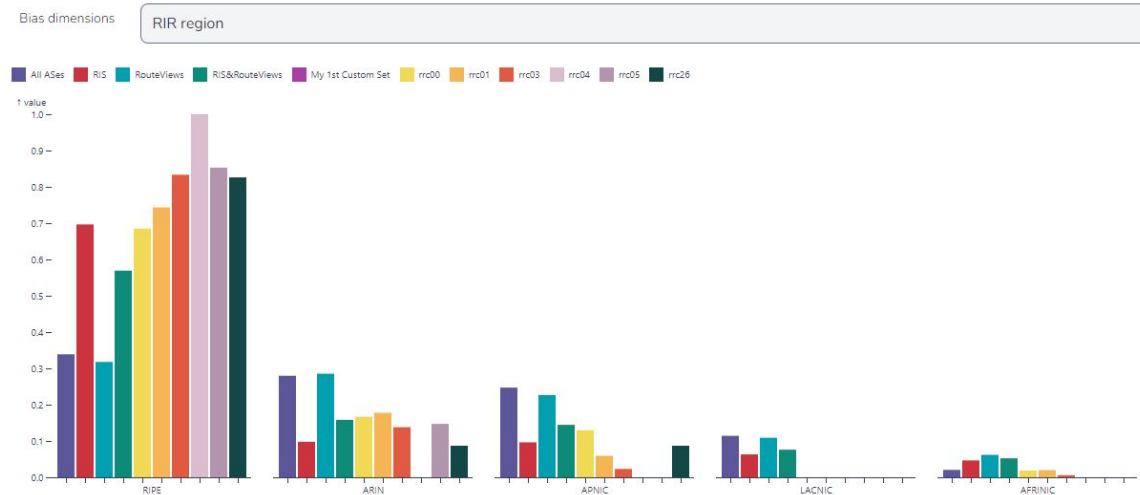
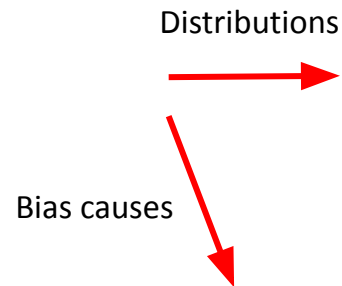


Web app: Route collectors measurements

- Available at <https://app-ai4netmon.csd.auth.gr/>



Same options
as bias page



Monitor Set:

Filter by: Bias Dimension, Value or Difference: 8 results

Bias Dimension	Value	Difference %
Network type (PeeringDB)	Not Disclosed	25.52
Network type (PeeringDB)	Educational/Research	-4.82



Detailed page: RIPE Atlas Measurements

- Check all the bias analysis and tools, especially for RIPE Atlas probes

Web app: RIPE Atlas

- Available at <https://app-ai4netmon.csd.auth.gr/>



Select a custom set of vantage points. In the boxes below, add a list of ASNs/Probe IDs (only numbers, separated with commas, no spaces; e.g., 174,1299,3333)

Select the type of list of numbers ASNs probe IDs

Custom Set #1 (ASNs)

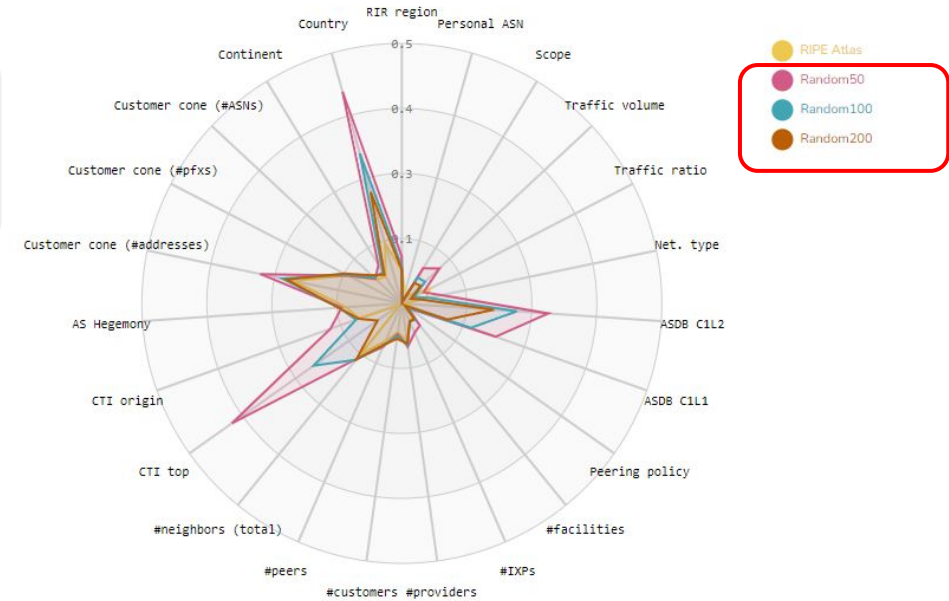
Custom Set #2 (ASNs)

Custom Set #3 (ASNs)

Monitor sets RIPE Atlas Random50 Random100 Random200 My 1st Custom Set My 2nd Custom Set My 3rd Custom Set

Bias dimensions RIR region Country Continent Customer cone (#ASNs) Customer cone (#pfxs) Customer cone (#addresses) AS Hegemony CTI origin CTI top #neighbors (total) #peers #customers #providers #IXPs #facilities Peering policy ASDB C1L1 ASDB C1L2 Net. type Traffic ratio Traffic volume Scope Personal ASN

select sets of random sampled Atlas probes

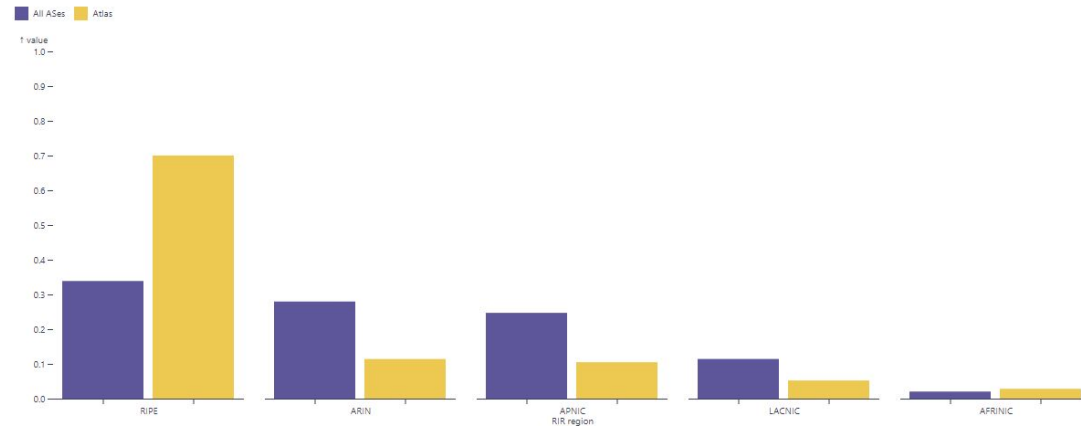
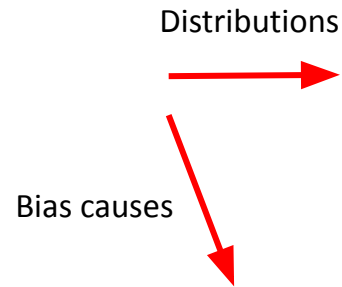


Web app: RIPE Atlas

- Available at <https://app-ai4netmon.csd.auth.gr/>



Same options as bias page



Monitor Set: RIPE Atlas

Filter by: Bias Dimension, Value or Difference

Search: 199 results

Bias Dimension	Value	Difference %
Customer cone (#ASNs)	1.0-3.0	-34.258
Customer cone (#ASNs)	3.0-9.0	13.518



Network Information per ASN or Atlas probe id

- Find a big variety of network information
- Select an individual network by ASN or probe id



Web app: Network Information

- Available at <https://app-ai4netmon.csd.auth.gr/>



check in which
percentile each
feature belong

Type ASN

1299

Property	Value	Percentile
AS_rank_continent	Europe	29.37%
is_personal_AS		98.96%
peeringDB_info_ratio	Balanced	24.68%
peeringDB_info_traffic	100000000.0	0.1%
peeringDB_info_scope	Global	9.01%
peeringDB_info_type	NSP	12.62%
peeringDB_info_prefixes4	600000.0	
peeringDB_info_prefixes6	130000.0	
peeringDB_policy_general	Restrictive	2.02%
peeringDB_ix_count	0.0	0.0%



A use case: Usage patterns in RIPE Atlas measurements

- How do users use RIPE Atlas in practice?
- Do they select biased sets of vantage points?
- Analyze common patterns using the AI4NetMon web app

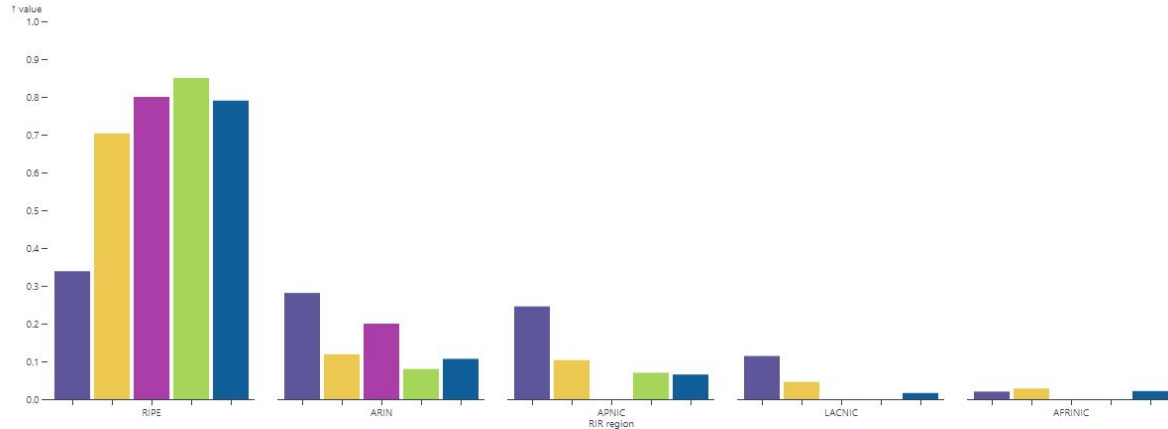


Analysis of most frequent ASNs (results)

Bias dimensions

RIR region

All ASes Atlas My 1st Custom Set My 2nd Custom Set My 3rd Custom Set



Top 10 Most Frequent ASes in our sample, have:

- 1) **Location:** Bias towards the RIPE region (same for the rest of the sample).
- 2) **Network size:** Significantly larger reach (in terms of the customer cone).
- 3) **Topology:** Significantly larger number of neighbors (peers and customers).
- 4) **Interconnection:** Larger number of connections to IXPs. They also have mostly “restrictive” and “selective” peering policies.
- 5) **Network Type:** The highest levels of traffic.

**Users tend to use in their measurements
ASes corresponding to large ISPs!**



Summarizing...

- Our contributions
 - A framework (data, definitions, etc.) to quantify bias
 - Analysis of bias in Internet measurement platforms
 - Code & tools
 - Website <https://ai4netmon.csd.auth.gr/>
 - Web app <https://app-ai4netmon.csd.auth.gr/>
- Next steps
 - **Unbias** Internet measurements [ongoing work] :
 - (a) extend platforms (add extra vantage points)
 - (b) carefully select vantage points (subsampling)
 - **Use cases**: When the bias really hurts our findings?
 - **Bias in ML models** based on data from measurements

