

# Unbiasing Internet Measurements

## Tools and findings from the AI4NetMon project

**Pavlos Sermpezis**

Data and Web Science Lab

(<https://datalab.csd.auth.gr/>)

RIPE85 - RIPE Meeting - 2023 - 85th Meeting - 12-14 September 2023 - Paris, France



# The AI4NetMon project

- Funded by RIPE NCC [RACI project funding 2021](#) program



- Work conducted by the [Data & Web Science Lab](#)  
... special thanks to collaborating researchers:

Emile Abel (RIPE NCC)



Lars Prehn (MPII)



- Project's website: <https://ai4netmon.csd.auth.gr/>
- Github repo: <https://github.com/sermpezis/ai4netmon>



# The AI4NetMon project

- **Goal 1:** quantify bias in Internet measurement platforms (RIPE Atlas, RIPE RIS, etc.)
- **Goal 2:** unbias Internet measurement platforms
  - How? → (i) deploy extra vantage points, or (ii) select subset of vantage points

# The AI4NetMon project

- **Goal 1:** quantify bias in Internet measurement platforms (RIPE Atlas, RIPE RIS, etc.)
  - in this talk:
    - brief introduction to bias
    - results: bias in Internet measurement platforms
    - online exploration tools & API
  - extra material:
    - RIPE84 [presentation](#)
    - RIPE Labs [podcast](#)
    - RIPE Labs [article](#)
    - code (& [how to use it](#)) @ [github](#)

# The AI4NetMon project

- **Goal 2: unbias Internet measurement platforms**
  - How? → (i) deploy extra vantage points, or (ii) select subset of vantage points
    - in this talk:
      - results: bias vs. selection of Atlas probes
      - results: bias vs. RIPE RIS route collectors
      - online tools: recommendations for deploying extra vantage points
    - extra material:
      - MAT WG @ RIPE85, [presentation](#), by Thomas Holterbach
      - MAT WG @ RIPE85, [presentation](#), by Malte Tashiro

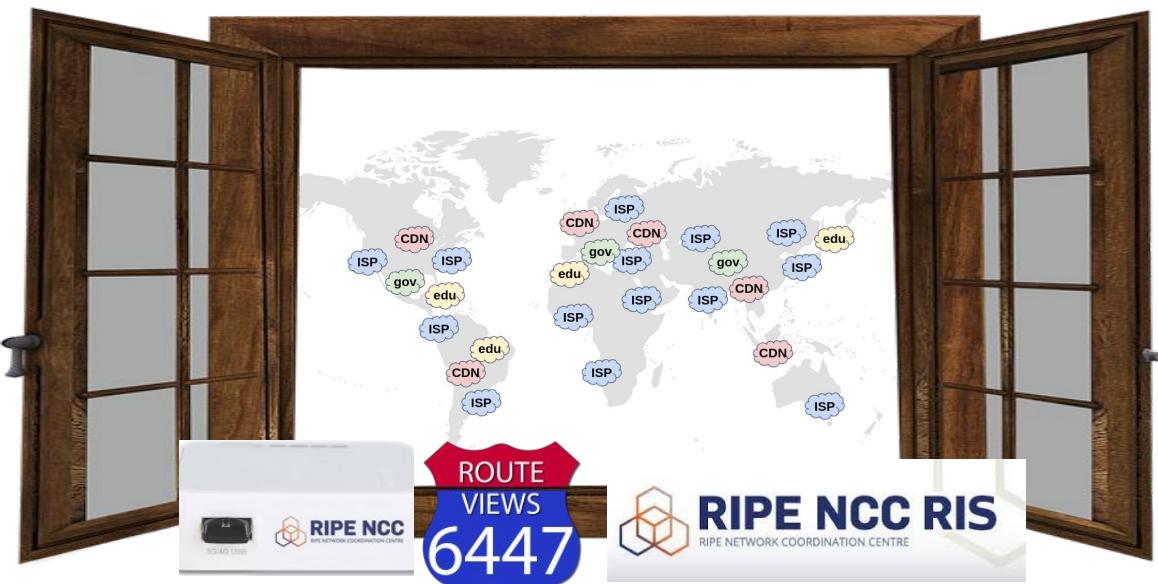


# Goal 1:

## Quantify bias in Internet measurement platforms

# Internet measurement platforms: a (stained glass) window to the Internet

Ideal (unbiased)

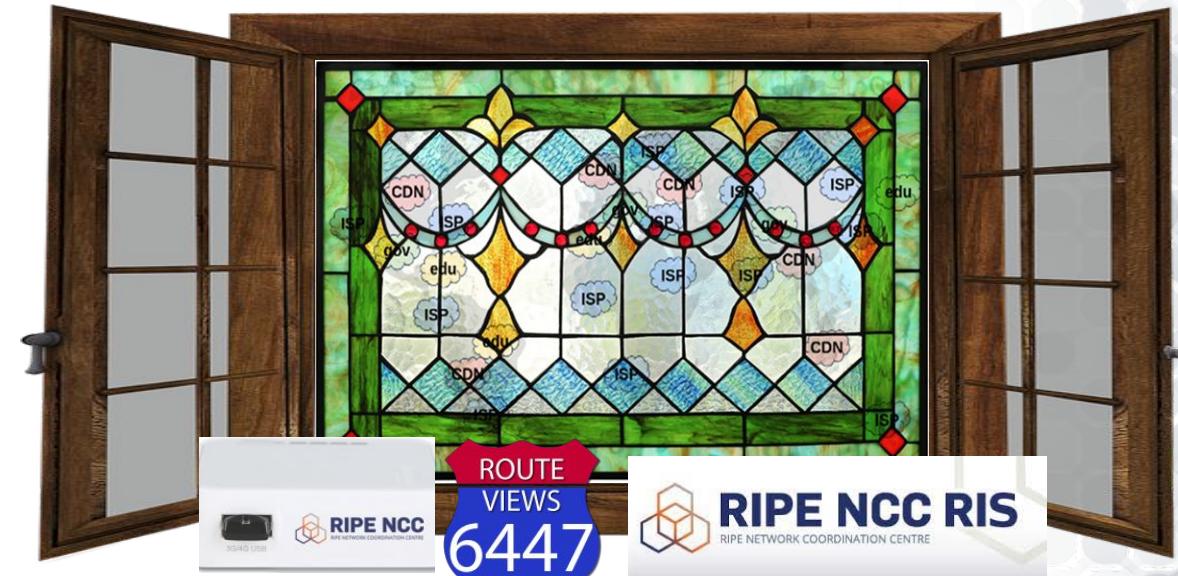


# Internet measurement platforms: a (stained glass) window to the Internet

Ideal (unbiased)



In reality (biased)



# Why there is bias?

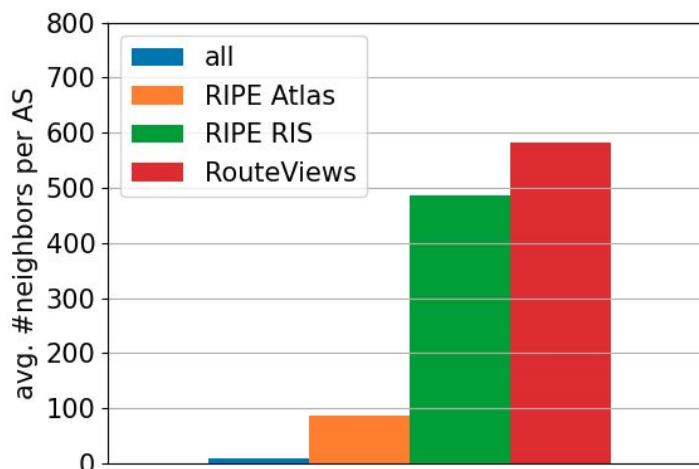
## Example 1: location bias



**RIPE Atlas probes**

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

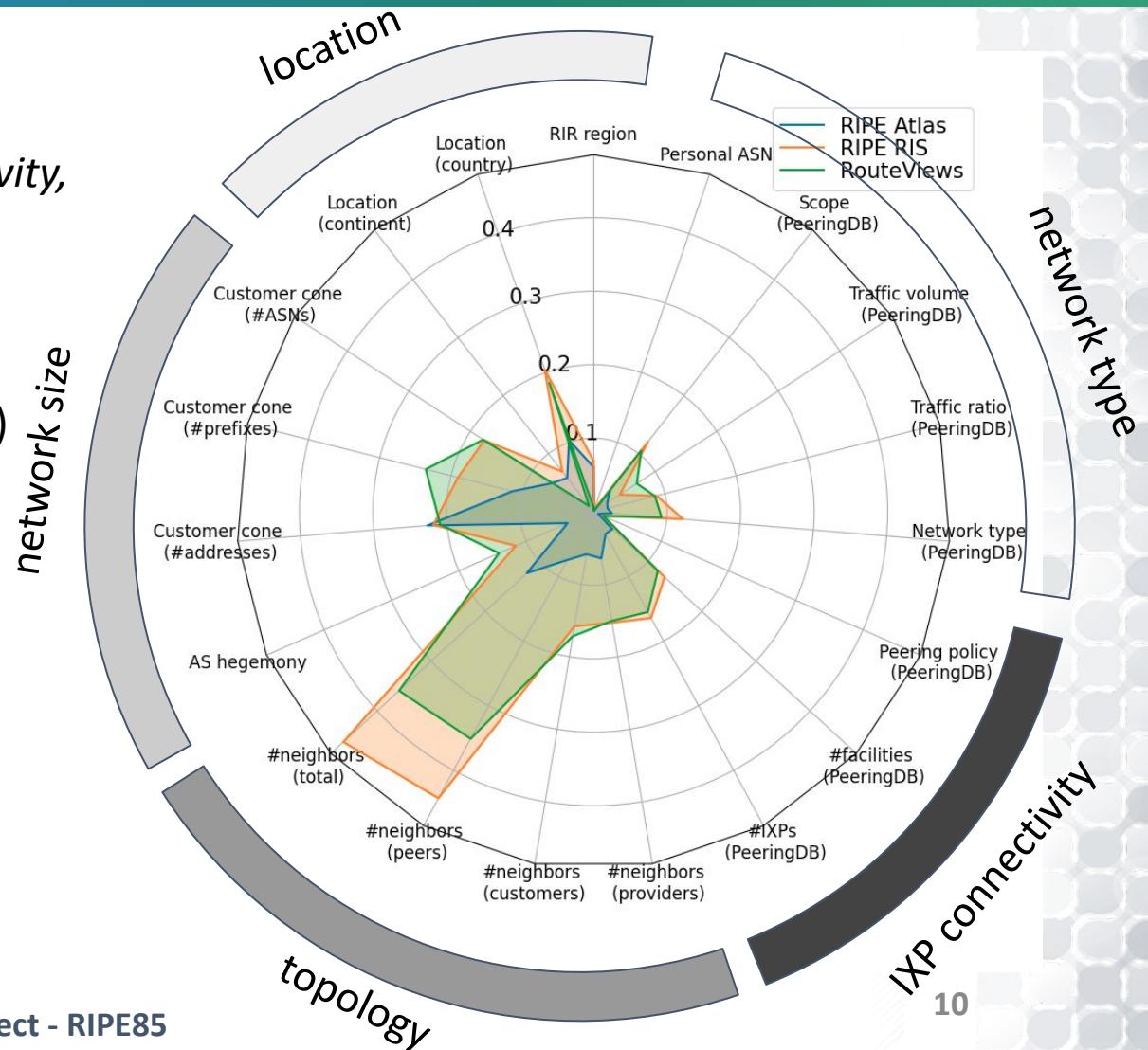
## Example 2: topology bias



**RIPE RIS / RouteViews peers have much more peering links (“neighbors”) compared to an average AS**

# Results: Bias in Internet measurement platforms

- 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





# AI4NetMon tool 1: “Show me the bias”

# “Show me the bias”: online tool

- Online tool: <https://observablehq.com/@pavlos/ai4netmon-bias>

[Optional] Select a different **custom set** of vantage points

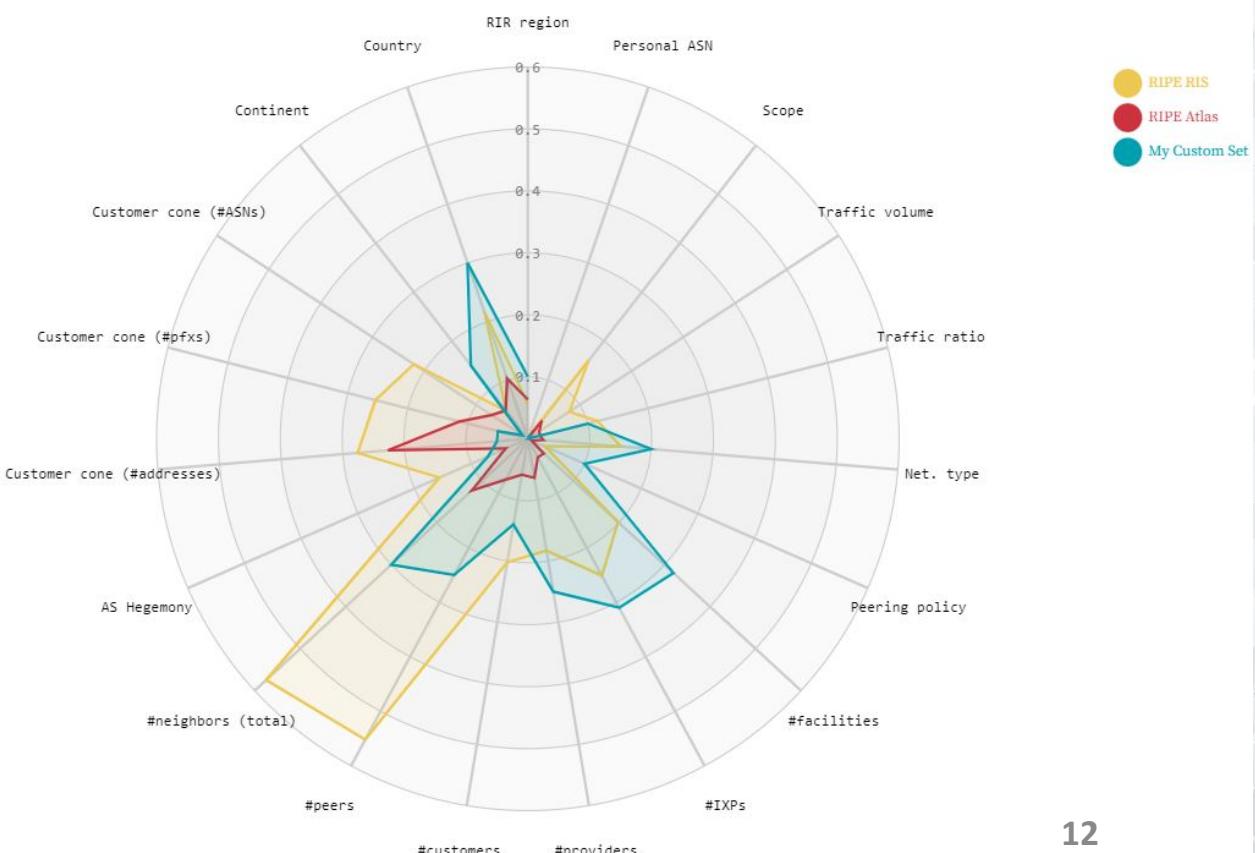
Set of vantage points (ASNs)

Select the sets of vantage points to visualize

Monitor sets  RIPE RIS  RIPE Atlas  My Custom Set

Select the bias dimensions to visualize

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



# “Show me the bias”: online tool

- Online tool: <https://observablehq.com/@pavlos/ai4netmon-bias>

[Optional] Select a different **custom set** of vantage points

Set of vantage points (ASNs) e.g., AS123, AS234, AS456

Select the sets of vantage points to visualize

Monitor sets  RIPE RIS  RIPE Atlas  My Custom Set

Select the bias dimensions to visualize

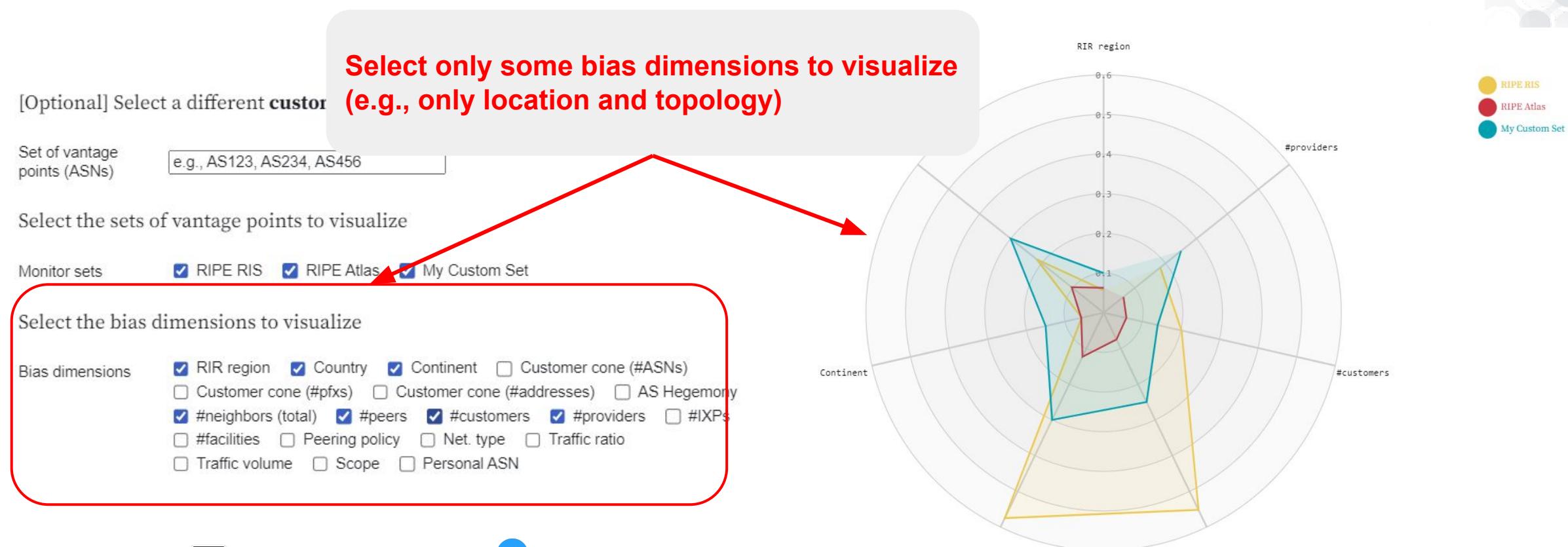
Bias dimensions  RIR region  Country  Continent  Customer cone (#ASNs)  
 Customer cone (#pfxs)  Customer cone (#addresses)  AS Hegemony  
 #neighbors (total)  #peers  #customers  #providers  #IXPs  
 #facilities  Peering policy  Net. type  Traffic ratio  
 Traffic volume  Scope  Personal ASN

- Select custom set of vantage points
- Automatic visualization of bias
- Easy comparison between platforms / sets

The radar chart displays the relative values of various network bias dimensions for three sets: RIPE RIS (yellow), RIPE Atlas (red), and My Custom Set (cyan). The dimensions are: Customer cone (#ASNs), Customer cone (#pfxs), Customer cone (#addresses), AS Hegemony, Peering policy, #neighbors (total), #peers, #customers, #providers, #facilities, #IXPs, and Traffic volume. The chart shows that RIPE RIS has high values in Customer cone (#ASNs) and #peers, while RIPE Atlas has high values in Customer cone (#pfxs) and #customers. My Custom Set shows a more balanced distribution across most dimensions.

# “Show me the bias”: online tool

- Online tool: <https://observablehq.com/@pavlos/ai4netmon-bias>





# AI4NetMon tool 2: API

# The AI4NetMon API

- The API is available at <https://ai4netmon.csd.auth.gr/api/>
- Endpoint for bias calculation: **/bias/{params}**
- Options for {params} - Internet measurement platforms
  - Atlas
  - RIS
  - RouteViews
  - RIS&RouteViews
  - rrc00
  - ...
  - rrc26
- Options for {params} - any custom list of ASNs / vantage points
  - ?asn=123&asn=456&...&asn=789

# The AI4NetMon API

- The API is available at <https://ai4netmon.csd.auth.gr/api/>
- Endpoint for bias calculation: **/bias/{params}**
- Options for {params} - Internet measurement platforms
  - Atlas
  - RIS
  - RouteViews
  - RIS&RouteViews
  - rrc00
  - ...
  - rrc26
- Options for {params} - any custom list of ASNs
  - ?asn=123&asn=456&...&asn=789

Example 1: bias of RIPE Atlas platform  
<https://ai4netmon.csd.auth.gr/api/bias/Atlas>

Example 2: bias of a custom set of ASNs  
<https://ai4netmon.csd.auth.gr/api/bias/?asn=174&asn=1299&asn=3333>

# The AI4NetMon API

Example 1: bias of RIPE Atlas platform  
<https://ai4netmon.csd.auth.gr/api/bias/Atlas>



```
{  
    "AS_rank_source": "0.069",  
    "AS_rank_iso": "0.1205",  
    "AS_rank_continent": "0.0659",  
    "AS_rank_numberAsns": "0.0708",  
    "AS_rank_numberPrefixes": "0.1223",  
    "AS_rank_numberAddresses": "0.2563",  
    "AS_hegemony": "0.1232",  
    "AS_rank_total": "0.1289",  
    "AS_rank_peer": "0.0762",  
    "AS_rank_customer": "0.0584",  
    "AS_rank_provider": "0.0749",  
    "peeringDB_ix_count": "0.0367",  
    "peeringDB_fac_count": "0.0357",  
    "peeringDB_policy_general": "0.0056",  
    "peeringDB_info_type": "0.0246",  
    "peeringDB_info_ratio": "0.0206",  
    "peeringDB_info_traffic": "0.023",  
    "peeringDB_info_scope": "0.0352",  
    "is_personal_AS": "0.0016",  
    "list_name": "Atlas"  
}
```

# The AI4NetMon API

- The API is available at <https://ai4netmon.csd.auth.gr/api/>
- Endpoint for bias calculation: **/bias/{params}**
- Options for {params} - Internet measurement platforms
  - Atlas
  - RIS
  - RouteViews
  - RIS&RouteViews
  - rrc00
  - ...
  - rrc26
- Options for {params} - any custom list of ASNs
  - ?asn=123&asn=456&...&asn=789

Note: API is still in v0.1

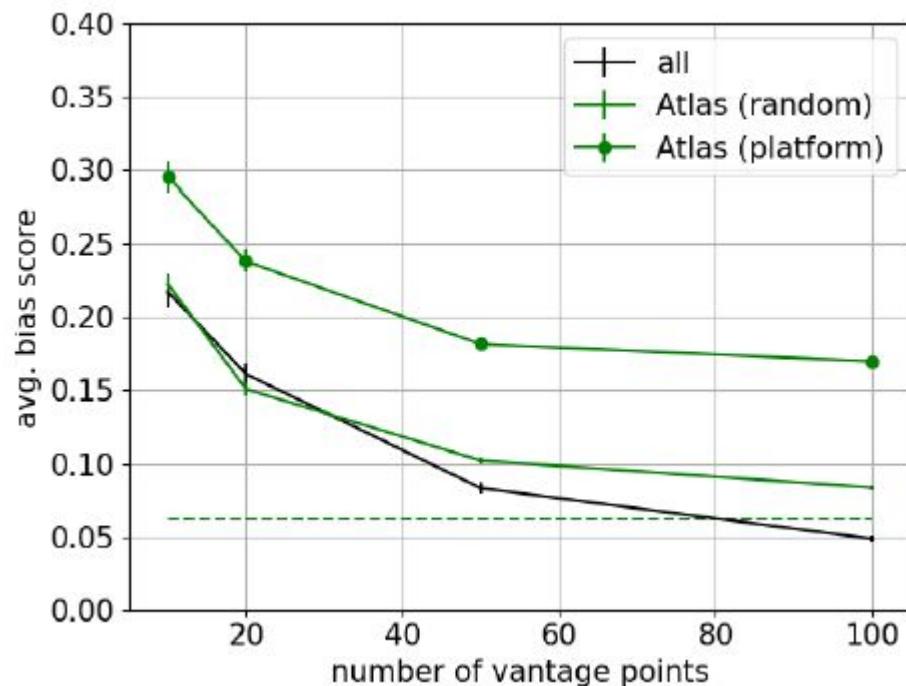
- Be gentle with it!
- Let us know about any problems or suggestions
- Stay tuned for more functionality (updates @ [our docs](#))

## Goal 2:

# Unbias Internet measurement platforms

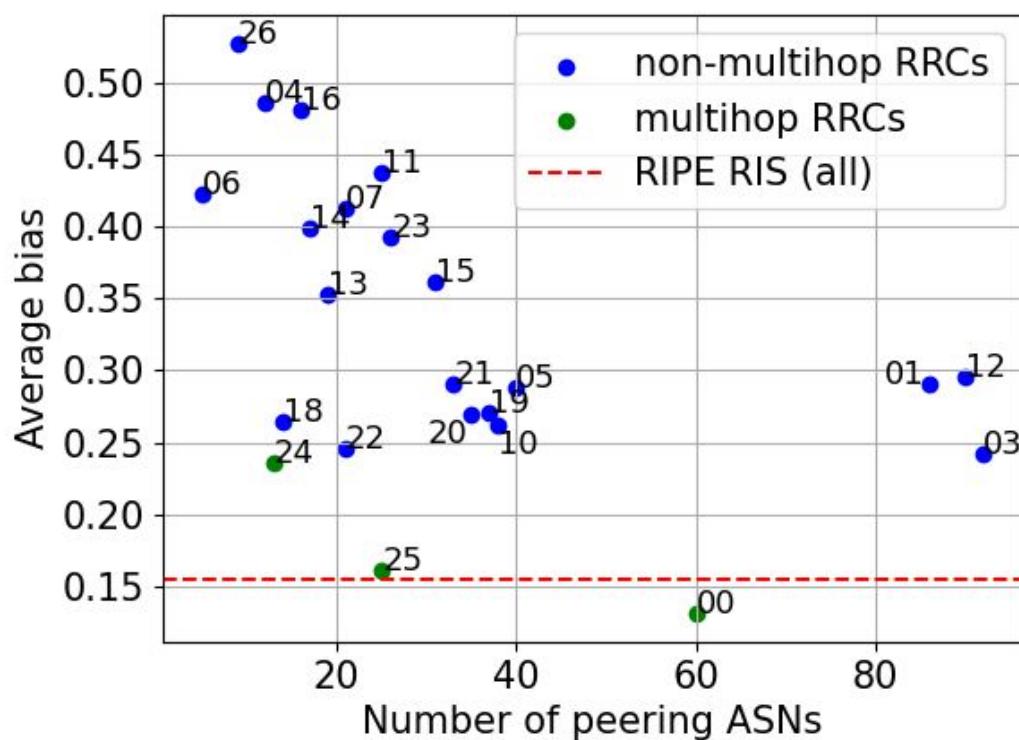
# Bias in RIPE Atlas measurements with few probes

- Bias vs. number of probes
- Less probes → higher bias
- Automatic selection by Atlas (“Atlas platform”) is ***more biased*** than randomly selecting probes (“Atlas random”)!



# Bias in RIPE RIS: per route collector analysis

- Each route collector (RC) has different set of peering ASes
- Bias is different for each RC
- Bias vs. number of peers



# Unbias measurements? Select subset of vantage points

- We designed a subsampling algorithm → selects unbiased sets of vantage points
  - Details at our [documentation](#) @ github
  - Online tool: <https://observablehq.com/@pavlos/ai4netmon-bias-per-route-collector>
- Main findings: ***our subset of 50 RIPE RIS peers reduces 4 times the bias of RIPE RIS***
- Main findings: ***our subset of 300 RIPE Atlas probes can lead to almost zero bias***



# Unbias measurements? Deploy extra vantage points



# AI4NetMon tool 3: Recommendations for vantage point deployment

# Unbiasing measurement platforms: online tool

- Online tools (current version)
  - RIPE RIS: <https://observablehq.com/@pavlos/ai4netmon-ripe-ris-extension>
  - RIPE Atlas: <https://observablehq.com/@pavlos/ai4netmon-ripe-atlas-extension>

Quick view: top-5 recommendations (see below the complete data & options)

ASN	RIR	Country	TOTAL bias
AS138222	APNIC	IN	-3.08%
AS134324	APNIC	IN	-3.02%
AS134873	APNIC	IN	-3.02%
AS138248	APNIC	IN	-3.02%
AS138287	APNIC	IN	-3.02%

RIPE RIS

- Calculate how bias decreases if a new vantage point is added
- Provide recommendations for best locations for deployment

Quick view: top-5 recommendations (see below)

ASN	RIR	Country	TOTAL bias
AS329031	AFRINIC	NG	-1.40%
AS28345	LACNIC	BR	-1.37%
AS26162	LACNIC	BR	-1.35%
AS61524	LACNIC	CL	-1.27%
AS263290	LACNIC	BR	-1.23%

RIPE Atlas

# Unbiasing measurement platforms: online tool

Location bias	0.7	
Network size bias	1	
Topology bias	0.1	
IXP-related bias	0.7	
Network type bias	1	

**Customization options for:**  
**- Calculating bias**  
**- Filtering specific candidate locations for deployment**

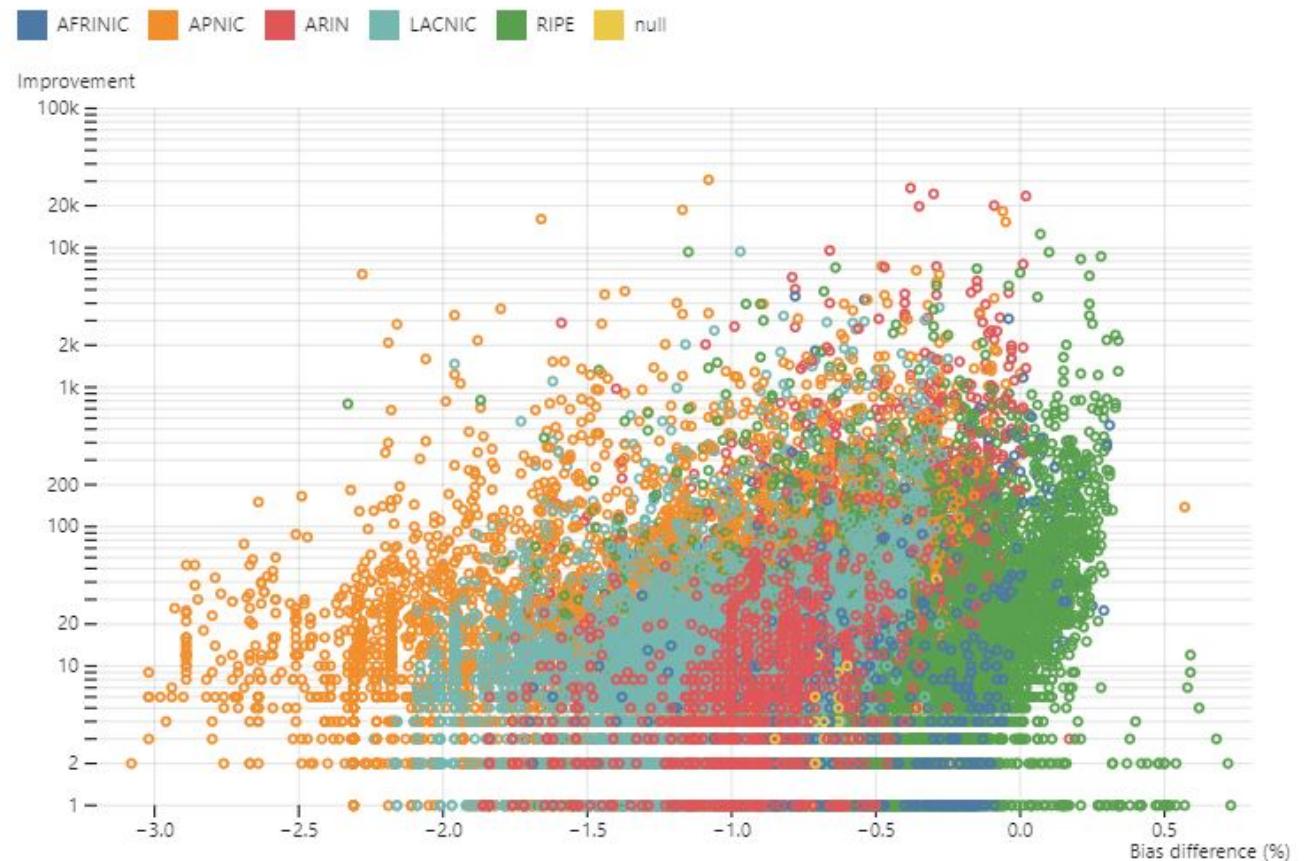
Filter: #neighbors (total)	<input checked="" type="radio"/> None <input type="radio"/> larger than <input type="radio"/> smaller than
Value	3062
Filter: Customer cone (#ASNs)	<input checked="" type="radio"/> None <input type="radio"/> larger than <input type="radio"/> smaller than
Value	23630
Filter: Customer cone (#prefixes)	<input checked="" type="radio"/> None <input type="radio"/> larger than <input type="radio"/> smaller than
Value	401598
Filter: min_dist	<input checked="" type="radio"/> None <input type="radio"/> larger than <input type="radio"/> smaller than
Value	5
Filter: Improvement proximity	<input checked="" type="radio"/> None <input type="radio"/> larger than <input type="radio"/> smaller than
Value	15357
Filter by condition:	<input type="checkbox"/> Is in PeeringDB <input type="checkbox"/> Has Atlas probe <input type="checkbox"/> Is RouteViews peer <input type="checkbox"/> Is NOT RouteViews peer <input type="checkbox"/> Peers at IXP(s)
Filter by peering policy:	<input type="checkbox"/> Restrictive <input type="checkbox"/> Open <input type="checkbox"/> Selective <input type="checkbox"/> No <input type="checkbox"/> N/A

# Unbiasing measurement platforms: online tool

Select locations taking into account other non-bias characteristics →

**Visualization:**

- dots correspond to candidate vantage points
- bias difference (x-axis; lower is better) vs. monitoring improvement (y-axis; higher is better)



# Summarizing...

- Goal 1 → Raise awareness about bias in Internet measurement platforms
  - There is bias & you should be informed about it!
  - hopefully our tools can help
- Goal 2 → Unbiasing platforms
  - our analysis shows that there is room for improvement
  - hopefully our recommendations are a first step
  - more people have started looking into it (MAT@RIPE85)

