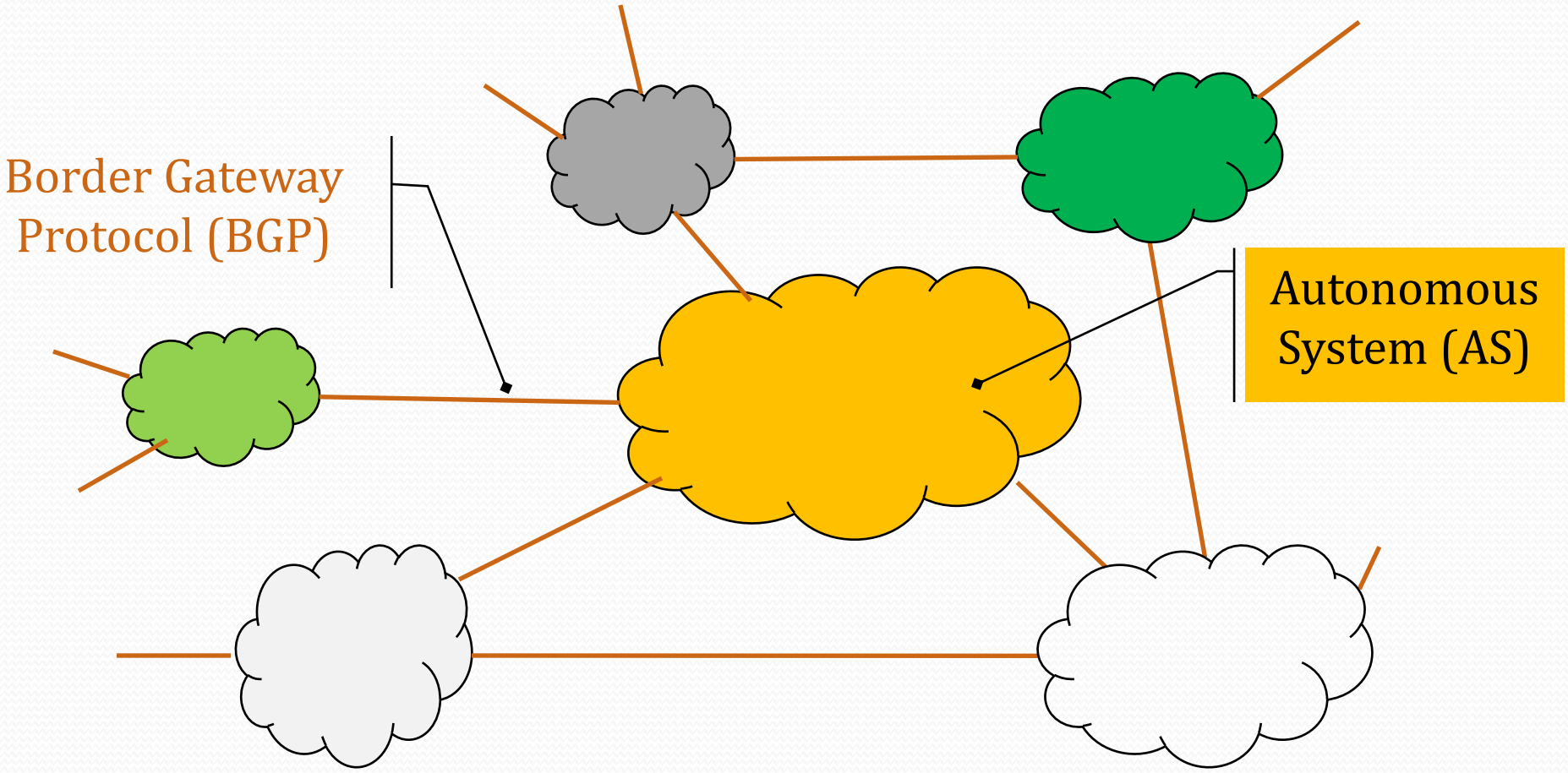
On the Quality of BGP Route Collectors for iBGP Policy Inference

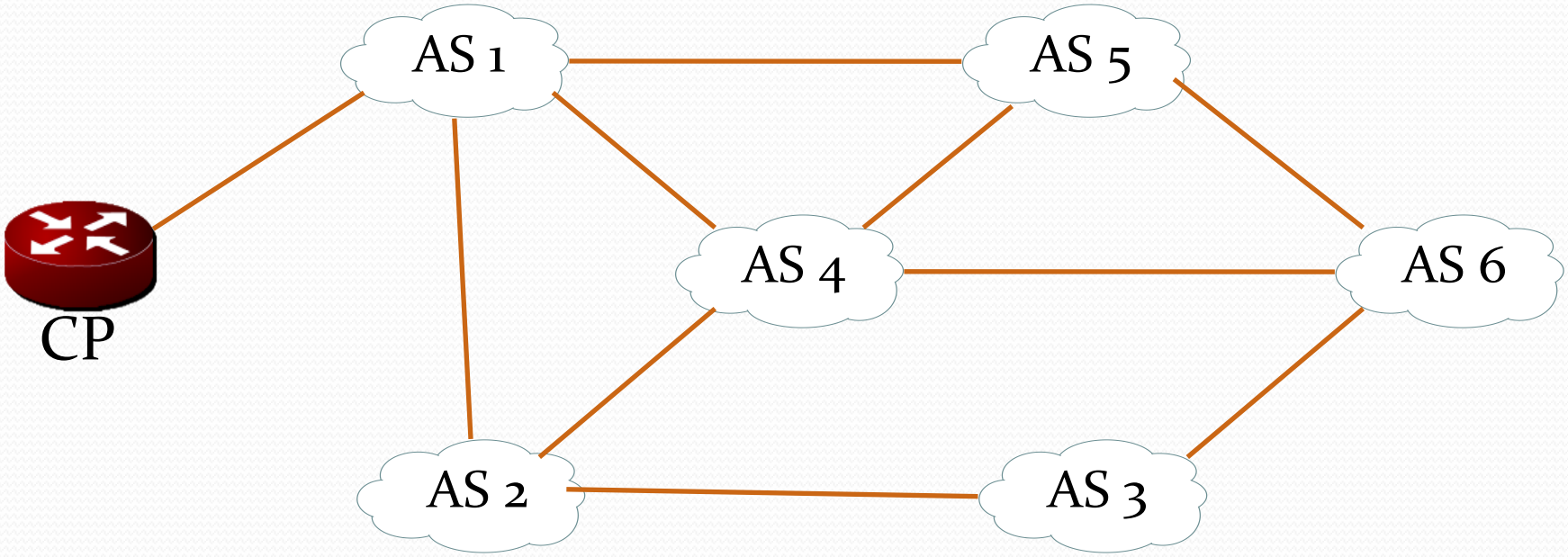
Luca Cittadini, Roma Tre University
Stefano Vissicchio, UCLouvain
Benoit Donnet, Université de Liege

IFIP Networking 2014 Conference
June 2-4, 2014 – Trondheim, Norway

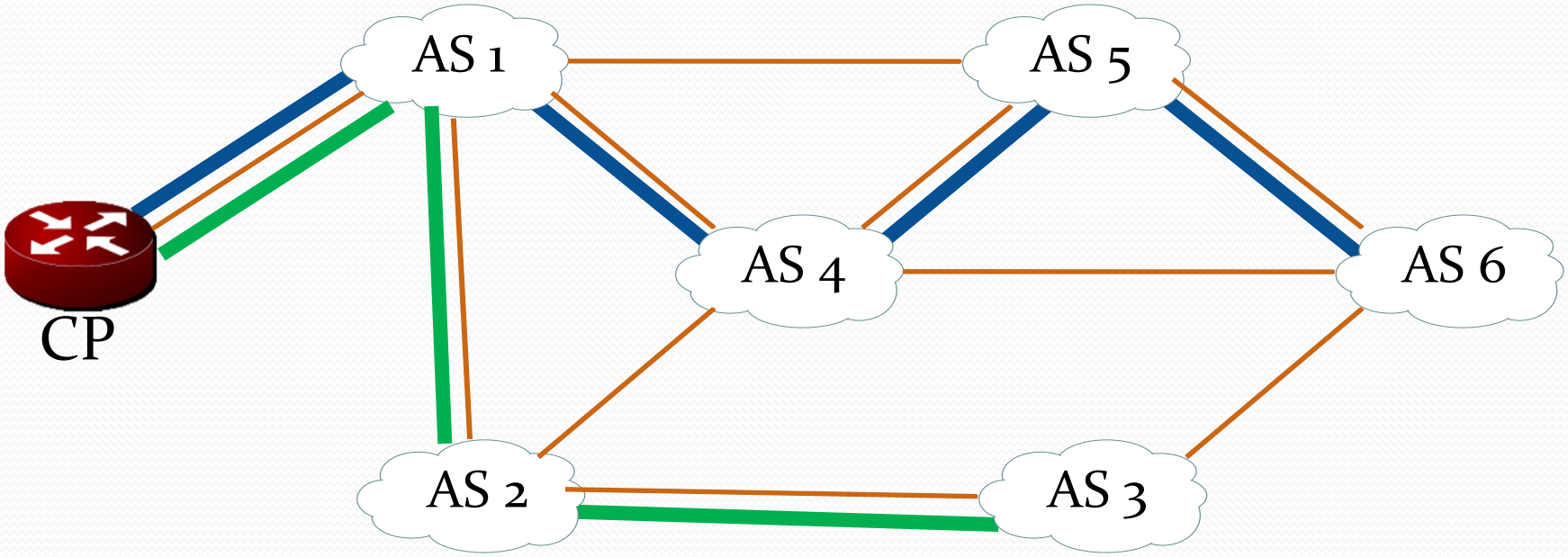
BGP glues the Internet



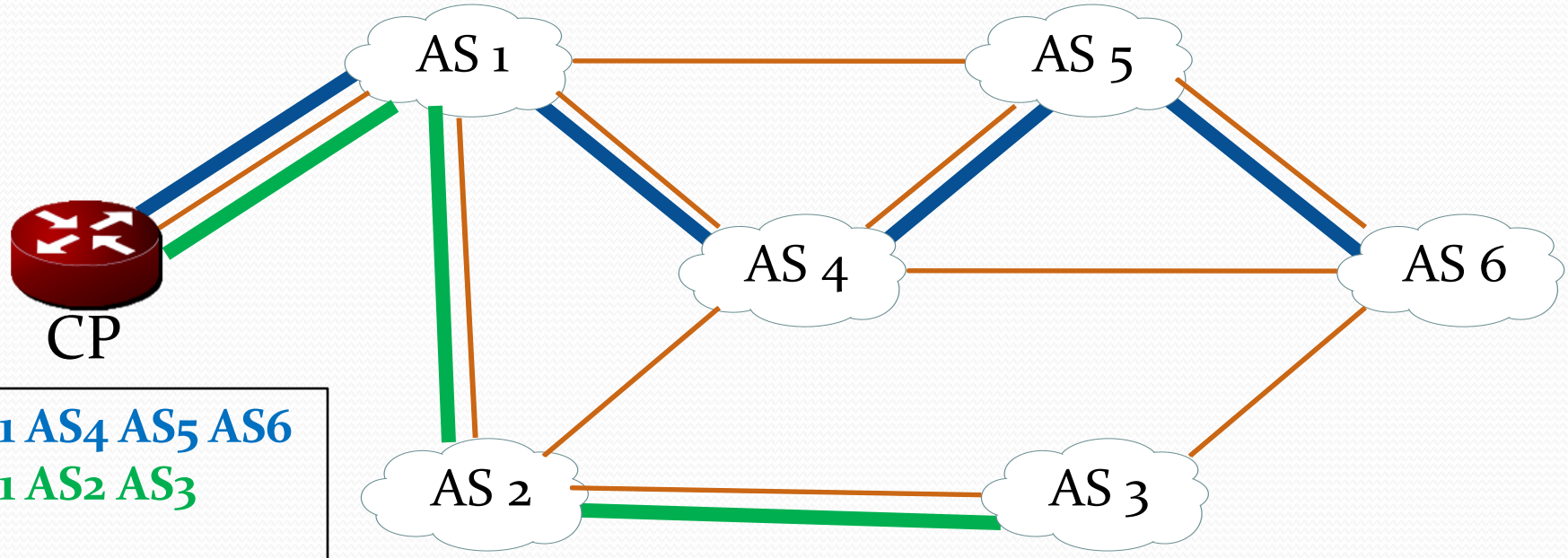
Some BGP data are public



Some BGP data are public

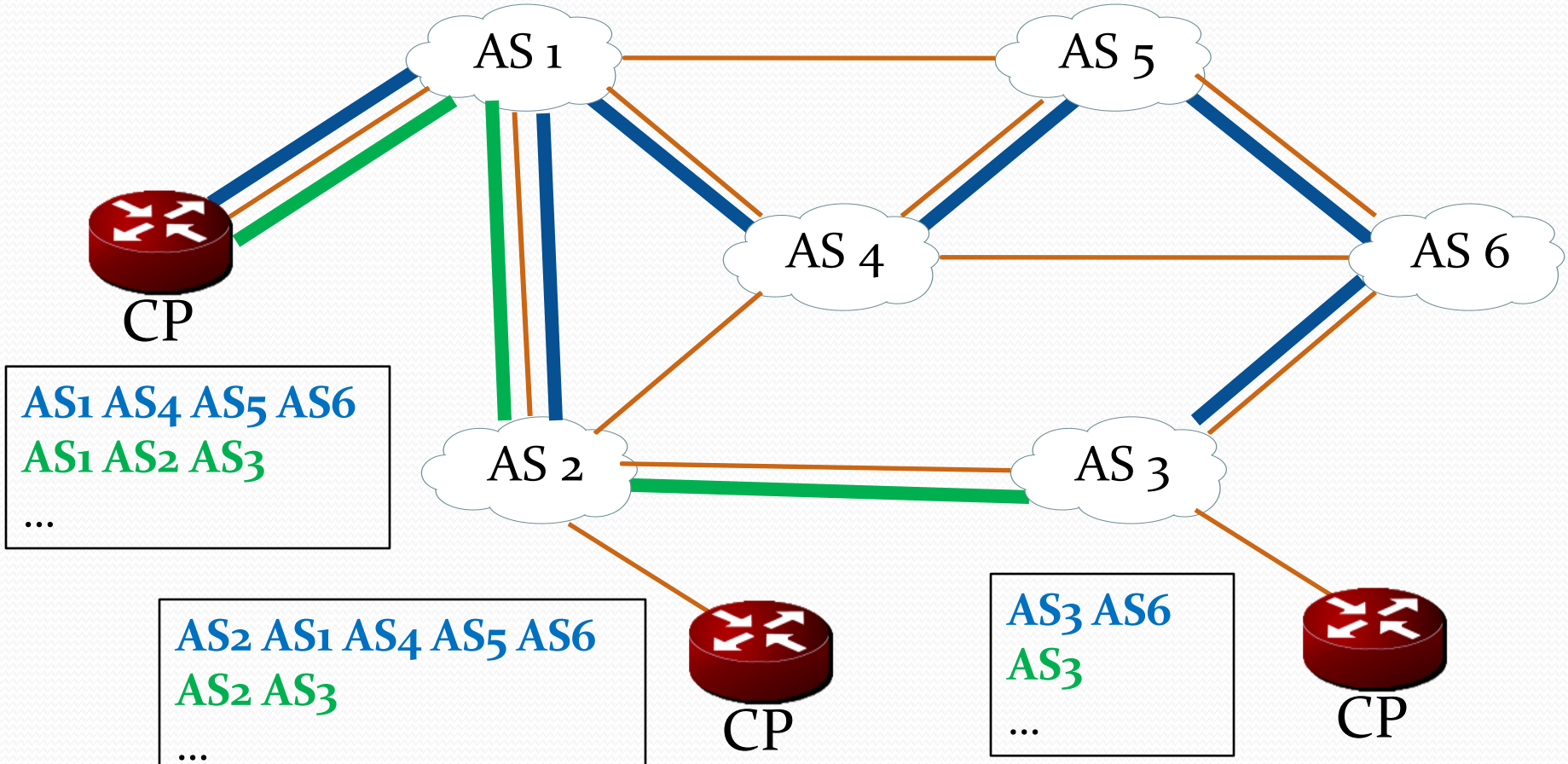


Some BGP data are public



AS1 AS4 AS5 AS6
AS1 AS2 AS3
…

Some BGP data are public



BGP data are used in research

- Prefix reachability (e.g., [Bush09])
- AS-level topology discovery (e.g., [Gregori12])
- Commercial relationships (e.g., [Gao01])
- Route diversity (e.g., [Muhlbauer06])
- …

Public BGP data are biased

- Because of many factors
  - CPs are not everywhere
  - the information reported is incomplete
  - BGP is based on information hiding
- Biases are partially assessed in prior works
  - prefix reachability (e.g., [Bush09])
  - AS-level graph (e.g., [Oliveira10,Roughan11])

Public BGP data are biased

- Because of many factors
  - CPs are not everywhere
  - the information reported is incomplete
  - BGP is based on information hiding
- Biases are partially assessed in prior works
  - prefix reachability (e.g., [Bush09])
  - AS-level graph (e.g., [Oliveira10,Roughan11])
- **Research question:**

  How general are known biases?

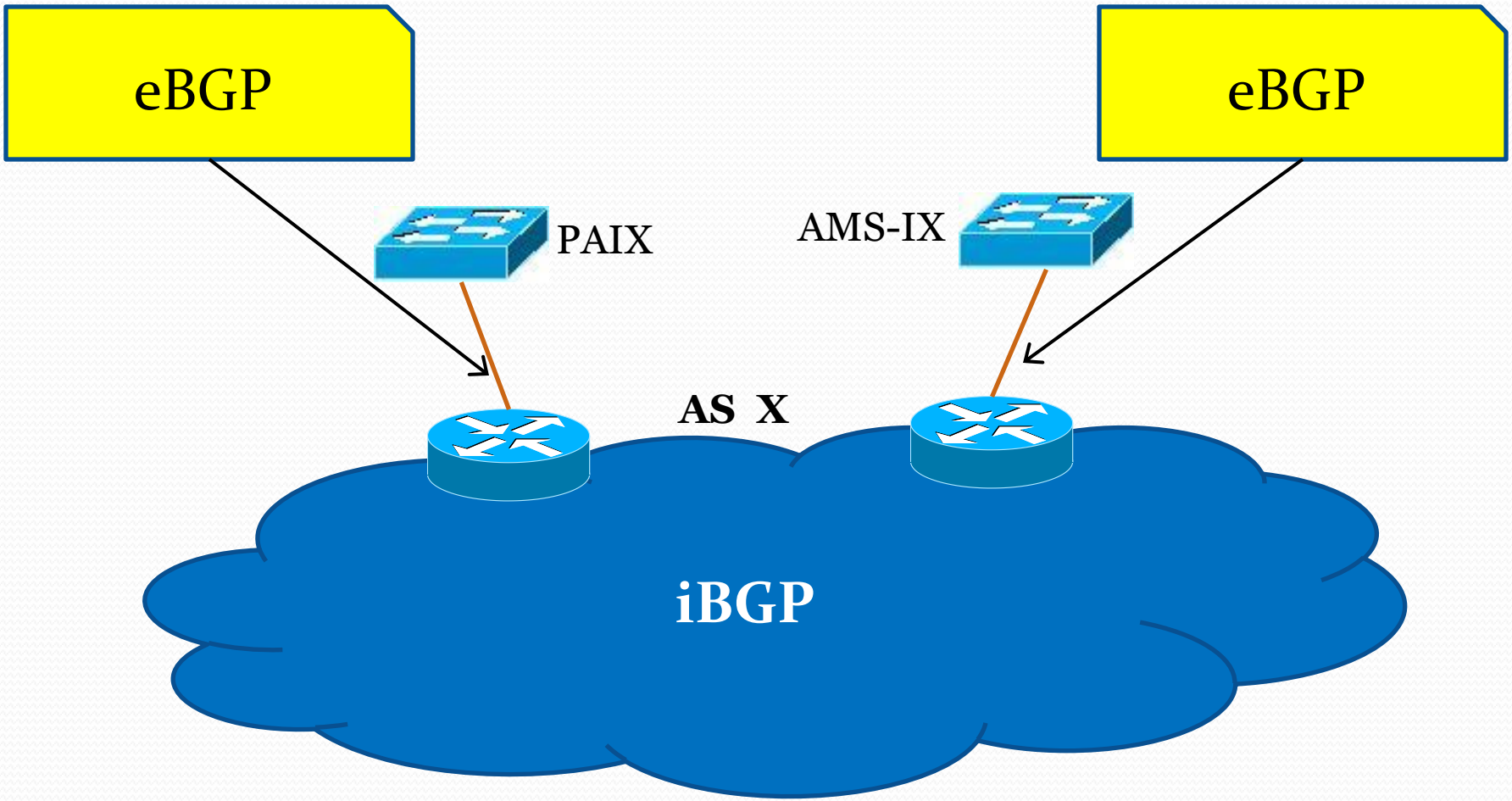
Our contributions

- We consider iBGP policy inference
  - as a different metric wrt AS-level graph discovery
- We propose a bias comparison methodology
  - not relying on any ground truth
- We perform multiple sensitivity analyses
  - showing how diverse biases apply to different metrics
- We conduct a marginal utility study
  - evaluating location strategies for new CPs

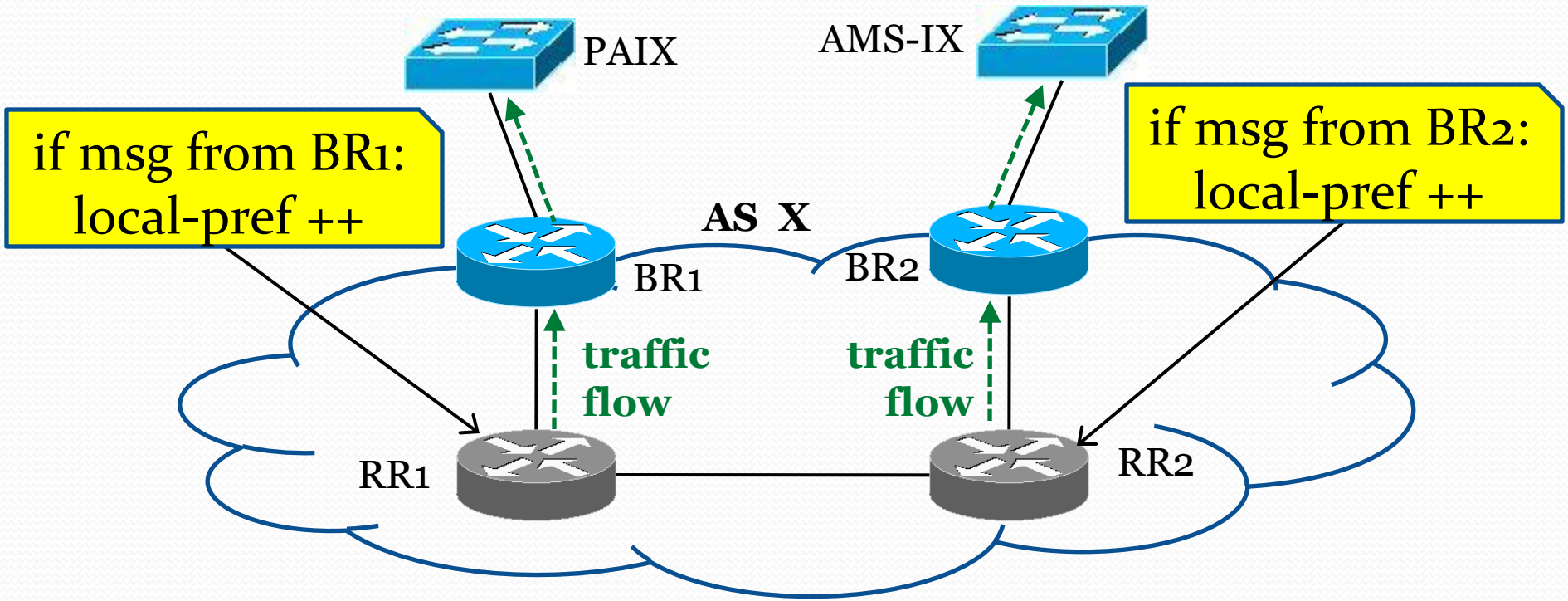
Our contributions

- We consider iBGP policy inference
  - as a different metric wrt AS-level graph discovery
- We propose a bias comparison methodology
  - not relying on any ground truth
- We perform multiple sensitivity analyses
  - showing how diverse biases apply to different metrics
- We conduct a marginal utility study
  - evaluating location strategies for new CPs

iBGP is used within each AS



iBGP can feature policies

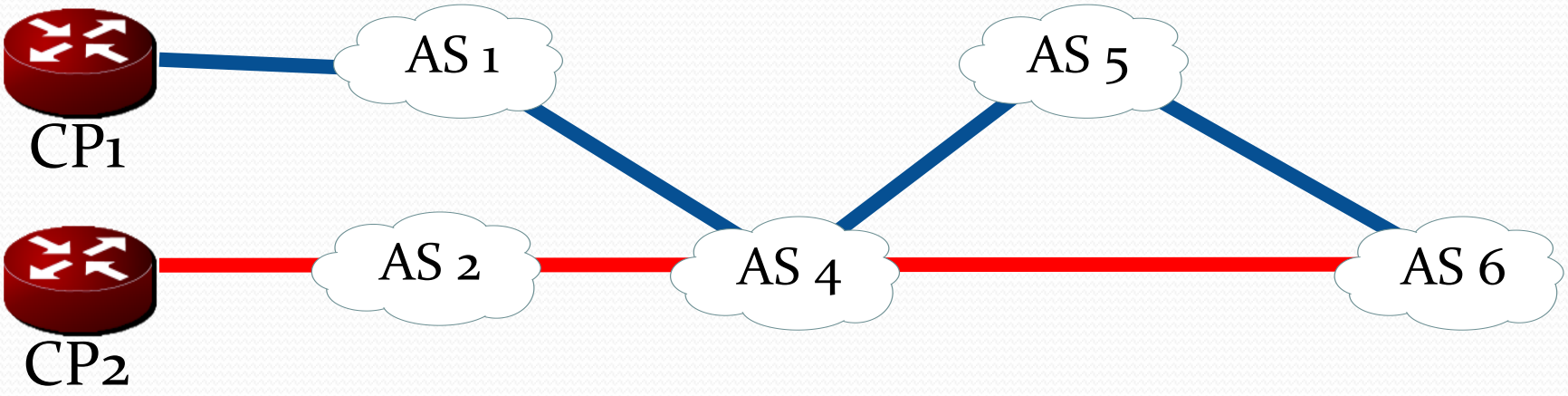


iBGP policies are relevant

- Common in transit providers [Vissicchio14]
  - e.g., to implement TE objectives
- Affect previous research results
  - on iBGP correctness (e.g., [Griffin02])
  - on route prediction (e.g., [Flavel10])
- Provide information on AS internals
  - partially disclose AS structure and configuration

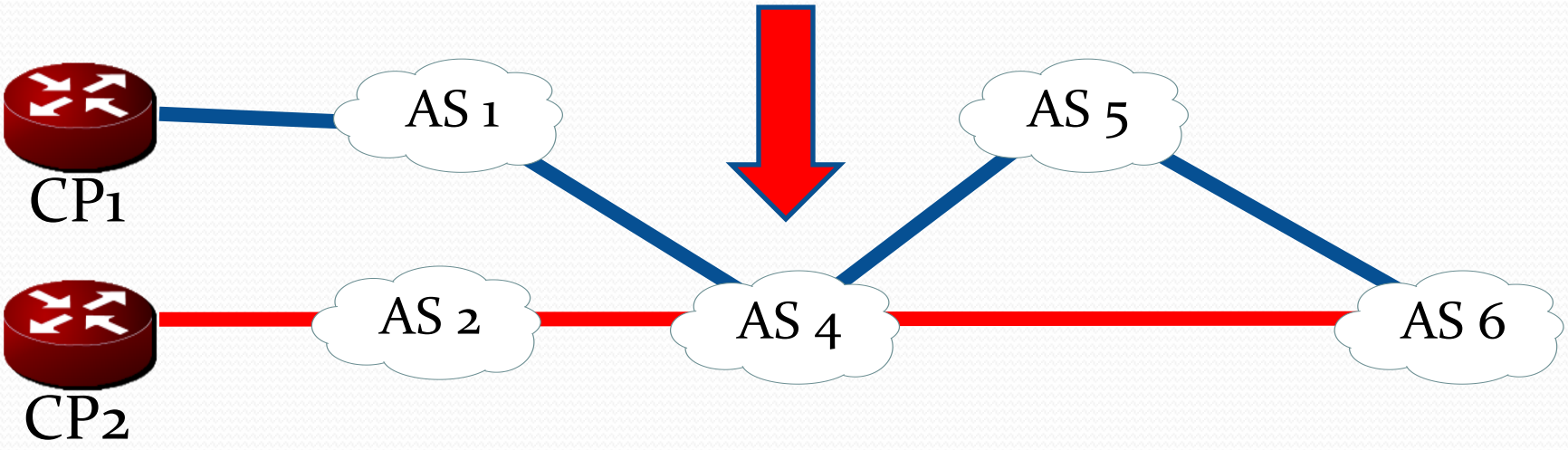
iBGP policies can be inferred

- By analyzing BGP routes from public datasets
  - simultaneously active on different CPs
  - to the same destination, e.g., prefix 1.0.0.0/8 at AS 6



iBGP policies can be inferred

- By analyzing BGP routes from public datasets
  - simultaneously active on different CPs
  - to the same destination, e.g., prefix 1.0.0.0/8 at AS 6
- Different AS-path lengths == iBGP policy



Our contributions

- We consider iBGP policy inference
  - as an additional metric wrt AS-level graph discovery
- We propose a bias comparison methodology
  - not relying on any ground truth
- We perform multiple sensitivity analyses
  - showing how diverse biases apply to different metrics
- We conduct a marginal utility study
  - evaluating location strategies for new CPs

Bias comparison methodology

- Works in the absence of ground truth
- We rely on sub-datasets
  - slices of the initial dataset
  - deliberately biased in a controlled way
  - exposing the impact of a specific factor
- We perform side-by-side comparison
  - iBGP Policy Inference (***pol***) vs interdomain link discovery (***links***) on sub-datasets

We applied our methodology

- Main dataset
  - BGP RIB dumps from RIPE RIS CPs on Sept. 16th, 2012
- Validation datasets
  - RIB dumps from RIS CPs on Sept. 16th, 2009-2011
  - RIB dumps from RIS CPs on random days in Sept. and Oct.

Our contributions

- We consider iBGP policy inference
  - as an additional metric wrt AS-level graph discovery
- We propose a bias comparison methodology
  - not relying on any ground truth
- We perform multiple sensitivity analyses
  - showing how diverse biases apply to different metrics
- We conduct a marginal utility study
  - evaluating location strategies for new CPs

We evaluate different factors

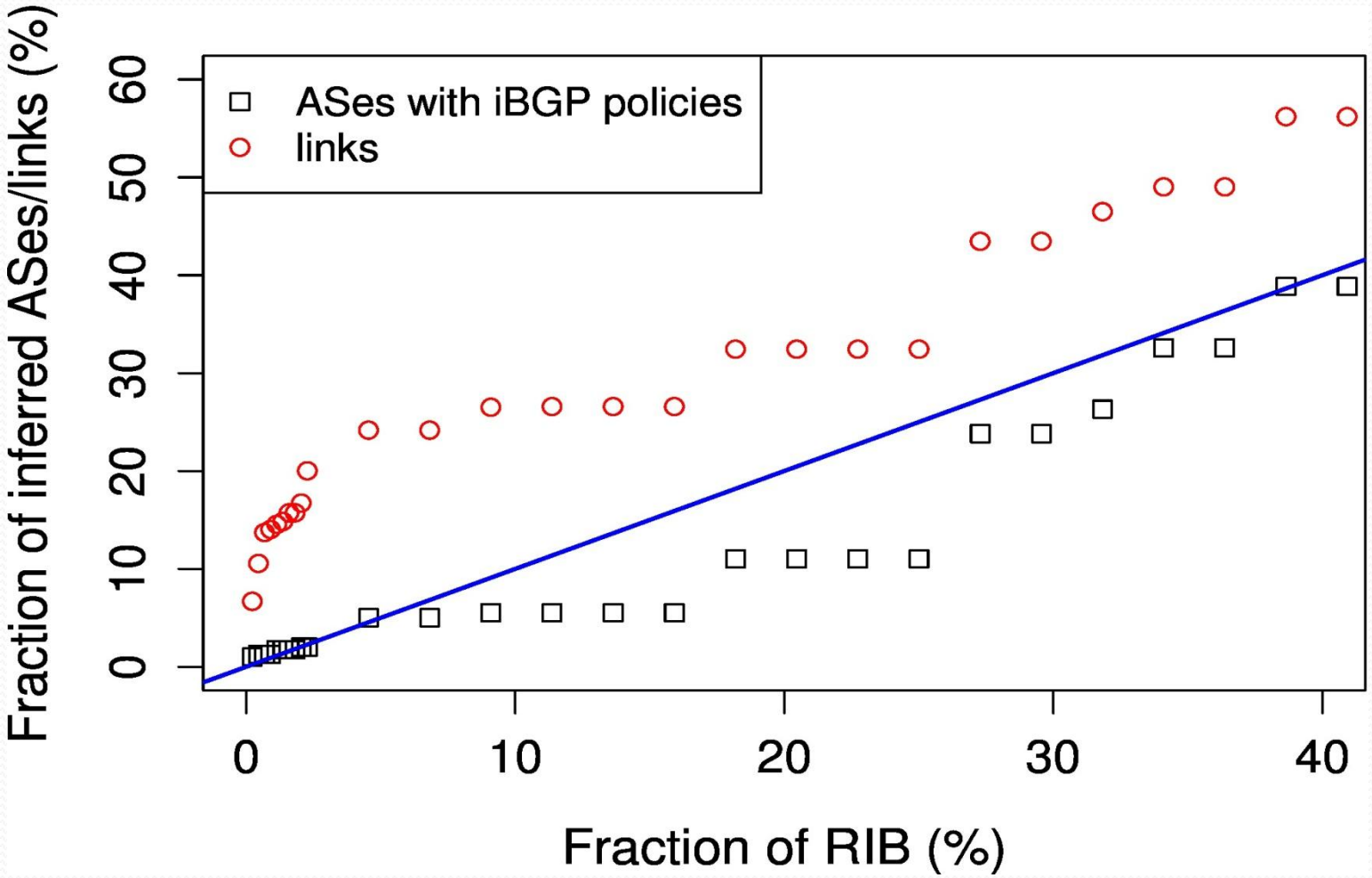
- Prefix visibility from CPs
- Number of CPs
- Position of CPs

Those factors impact the amount of useful BGP information in the dataset

Prefix visibility: experiments

- *Sub-datasets*: CPs with ≤ K% of the full RIB
  - how much can we infer for any given K?
- *Goal*: impact of info from single prefixes
  - redundancy
  - utility of partial vs full CPs
  - hints on optimal CP position

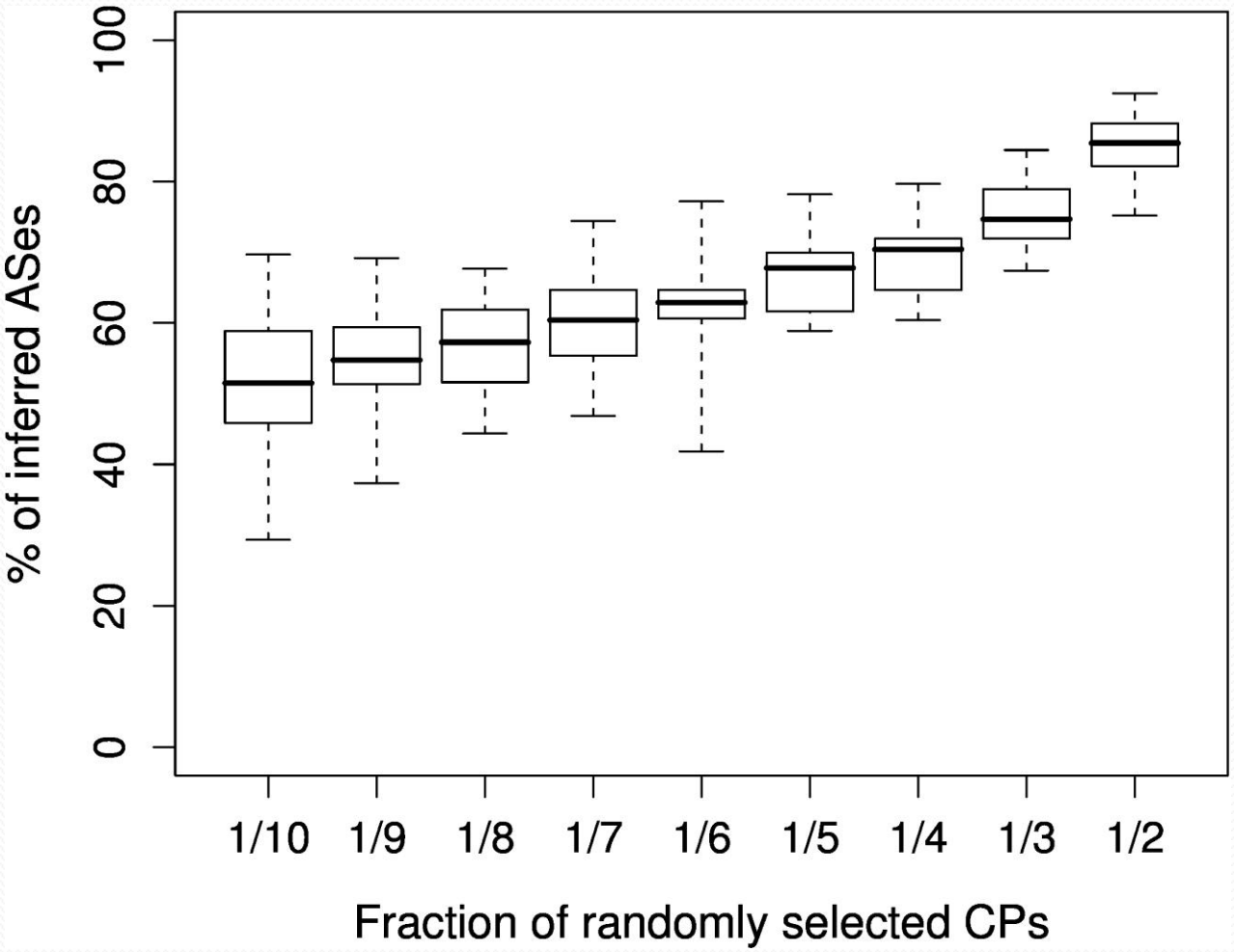
Prefix visibility: results



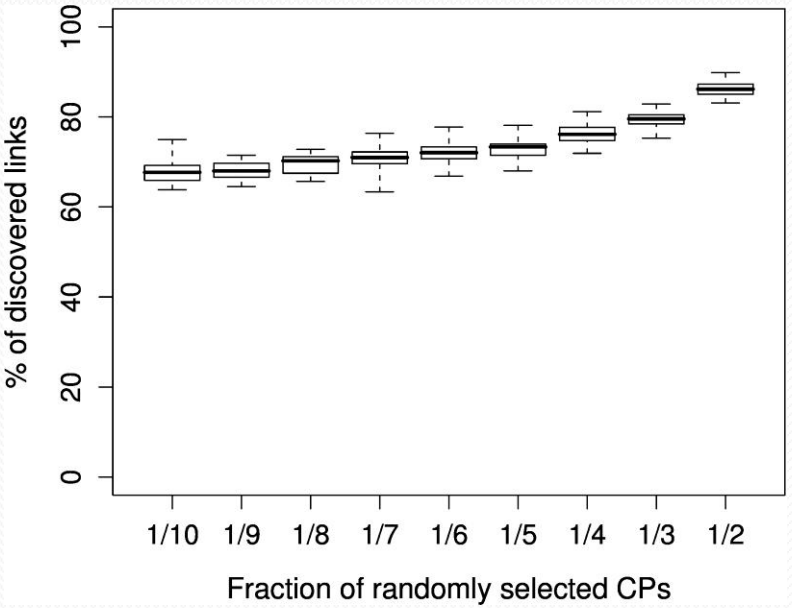
Number of CPs: Experiments

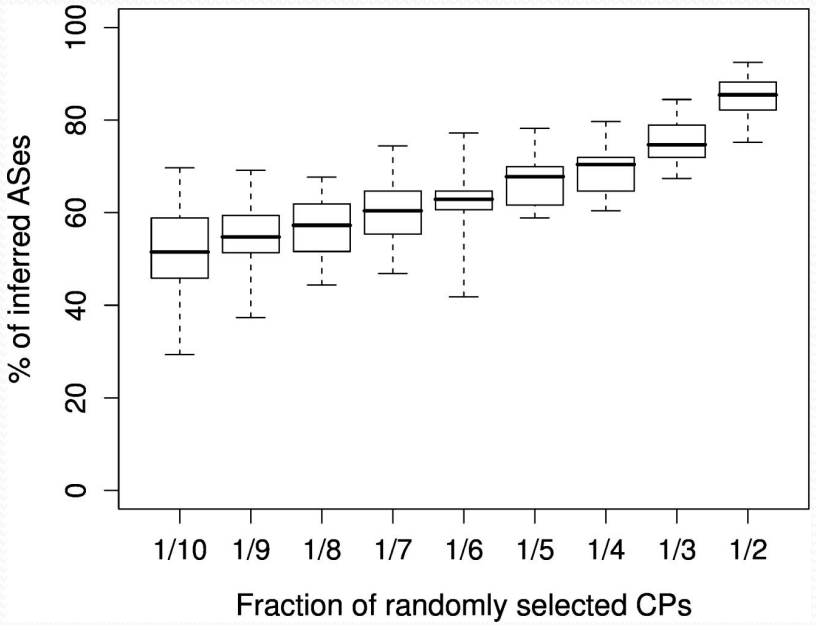
- *Sub-datasets*: random sets of CPs of fixed size K
  - what is the impact of K on the inference power (for *links* and *pol*)?
- *Goal*: sensitivity to the number of CPs
  - hints on utility of randomly adding new CPs

Number of CPs: Results for *pol*



Number of CPs: Analysis





- The number of CPs is more critical for *pol*
- Variability → importance of *specific* CPs for *pol*
  - i.e., less info redundancy in policy inference

Position of CPs: Recap

- *Sub-datasets*: random sets of 15 CPs in AS class X
  - how the position of CPs in the Internet hierarchy influences the results of our inferences?
- *Analysis*: results expose differences
  - big contributors (ECs, and LTPs) are the same
  - CPs in a single class are more useful for *links*
  - CPs in multiple classes are more critical for *pol*
    - variability stresses the importance of specific groups of CPs

Our contributions

- We consider iBGP policy inference
  - as an additional metric wrt AS-level graph discovery
- We propose a bias comparison methodology
  - not relying on any ground truth
- We perform multiple sensitivity analyses
  - showing how diverse biases apply to different metrics
- We conduct a marginal utility study
  - evaluating location strategies for new CPs

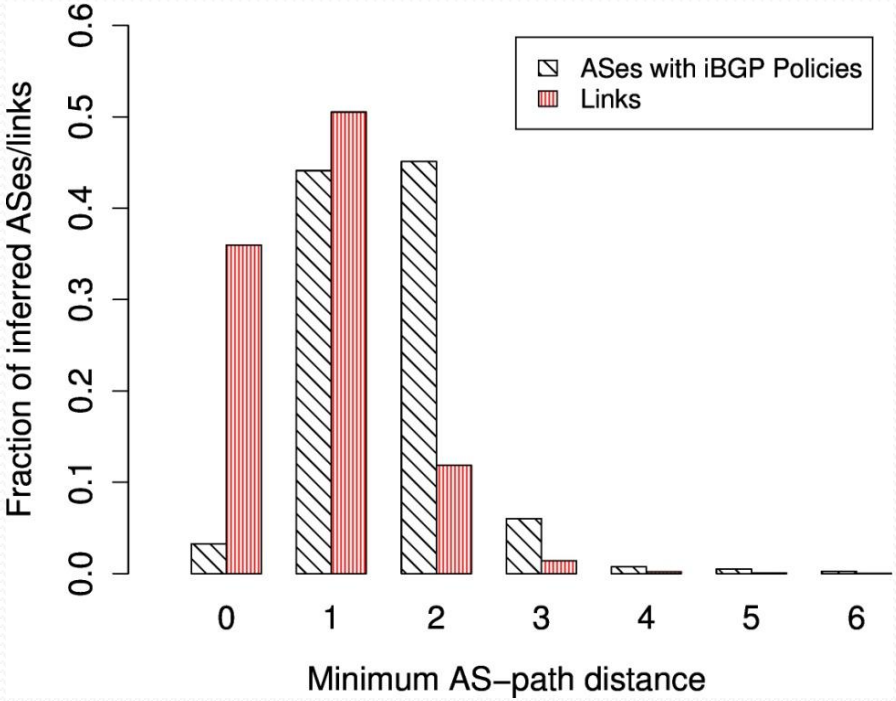
We compare CP location metrics

- Marginal utility of CPs is not practically useful
  - close to zero for all CPs in both metrics
  - … consistently with discovered redundancy
- We focus on two indirect indicators
  - AS-path distance
  - normalized Routing State Distance (RSD*)

AS-path distance: Experiments

- AS-path distance = minimum distance in the AS path
- *Sub-datasets*: CPs at distance <= K
  - what is the minimum distance K needed for a given inference?
- *Goal*: hints on optimal "topological" position of CPs
  - in the Internet AS-level graph

AS-path distance: Analysis



- For both metrics, marginal utility of CPs is *localized*
  - distant CPs are rarely useful
- Different inference target may require different CP placements

RSD*: Recap

- RSD*=adaptation+normalization of RSD [Gursun12]
  - quantify difference in BGP view between CPs
- *Sub-dataset*: CPs at a given RSD*
  - how the BGP view difference relates to the marginal utility of CPs?
- *Analysis*: RSD* is not a good indicator for both metrics
  - higher RSD* is better for *links*
  - a more complex balance is needed for *pol*

Putting all together

- We proposed a bias comparison methodology
  - applicable to BGP datasets with no ground truth
- BGP dataset biases likely depend on the metric
  - different sensitivity to the same factors
- No one-size-fit-all for BGP monitoring infrastructures
  - optimal monitor location depends on the metric
  - known placement algorithms (e.g., [Gregori12]) likely not good for other metrics than topology discovery

Thanks for your Attention!

- Questions?

stefano.vissicchio@uclouvain.be