Characterizing AS Relationships by Recursive Analysis of AS Adjacency Matrix

Hirochika Asai <panda_at_hongo.wide.ad.jp>

Hiroshi Esaki <hiroshi_at_wide.ad.jp>

The University of Tokyo

CAIDA-WIDE-CASFI workshop @Osaka,

April 24-25, 2010

Terminology (you probably know)

- Autonomous System (AS)
  - administrative network domain operated by ISP, company and university
- AS Relationships
  - transit
    - provider-customer relationship
      - provider to customer link : p2c
      - customer to provider link : c2p
  - peering
    - peer-to-peer relationship
      - peer-to-peer link : p2p

Summary

- AS magnitude quantification method
  - quantify AS' network scale by using a simple traffic transition model
    - To calculate the magnitude, we use eigenvalue analysis.
    - from AS adjacency matrix (not AS paths)
- Characterize AS relationships
  - analyze differences in magnitude by AS relationships
    - show the proposed method appropriately characterize the relationships

INTRODUCTION

Background

- AS relationships inference has been used in many research fields.
  - Traffic optimization
    - e.g., application layer inter-domain traffic optimization [Asai 2008]
      - high-cost transit traffic reduction
  - Routing
    - e.g., resilient overlay network [Andersen et al. 2001]
  - Security
    - e.g., prefix hijack detection [Zhang et al. 2008]

Related work

- AS relationships inference based on "*valley-free path model*"
  - heuristics [Gao 2001]
    - annotate links, eliminating contradictions to valley-free path model by analyzing AS paths in routing tables
  - (weighted) MAX2SAT [Battista et al. 2003, 2007, Dimitropoulos et al. 2005, 2007]
    - maximize the (weighted) number of valley-free paths in routing tables

Problem of related works and our solution (1/2)

- Requiring enough (a number of) AS paths
  - lower availability for AS paths

    → use AS adjacency matrix; some adjacencies are available from Internet routing registries etc. as well.
- Annotating links
  - difficulty annotating invisible links; AS paths in BGP routing tables constitute a (quasi) spanning subgraph of the Internet.

    → quantify ASes before characterize the relationships; since almost all ASes are visible, this makes it easy to characterize newly visible links.

Problem of related works and our solution (2/2)

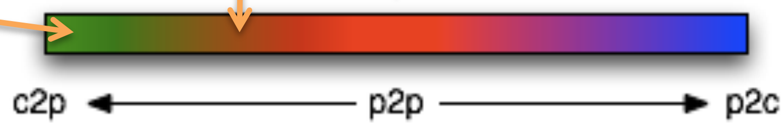
- Classifying links into two (transit and peering) or three (+sibling)
  - do not represent the relationships numerically
    - – Req. adding precision of inference
    - – Req. inferring complex relationships such as paid peer

→ characterize inter-AS links quantitatively

e.g., inferred as transit but may peering

notion

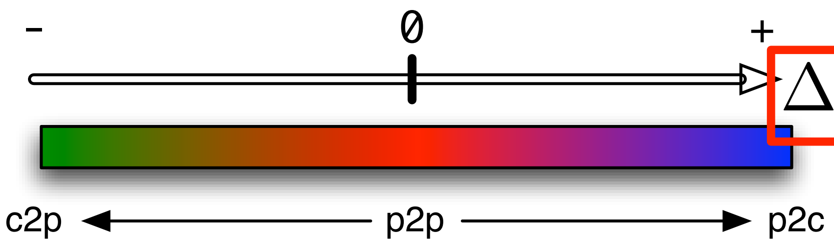
e.g., inferred as transit and should be



p2c: provider to customer
c2p: customer to provider
p2p: peer to peer

Well-known way to represent the relationships quantitatively

- Degree; i.e., #number of neighbors
  - related works also use this to determine the orientation of transit links
    - high degree = large AS
      - Larger ASes tend to be providers.
    - low degree = small AS
      - Smaller ASes tend to be customers.

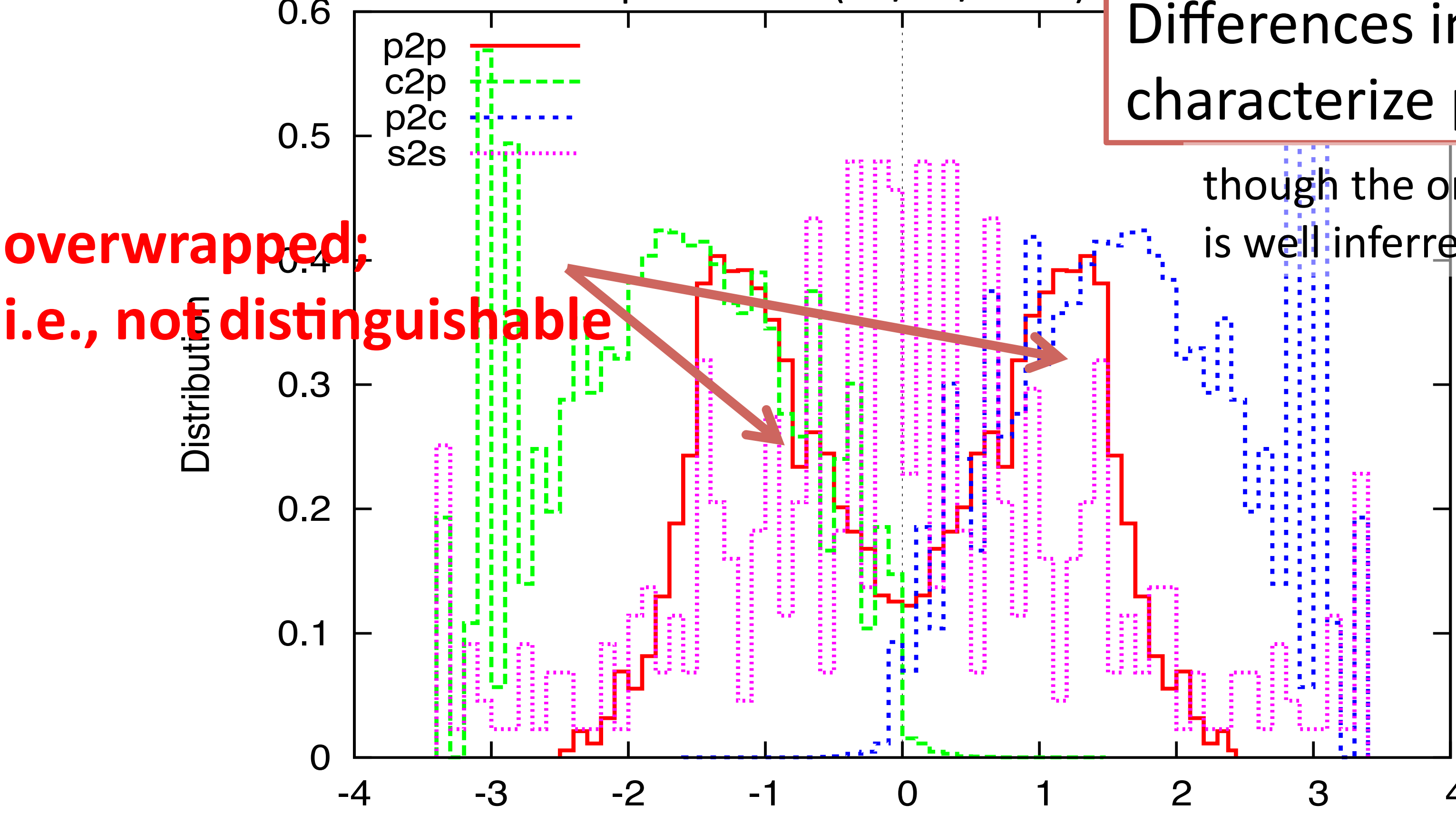


$\Delta\rho$: difference of degree

p2c: provider to customer
c2p: customer to provider
p2p: peer to peer

In reality...
PDF of difference in degree

Dataset: CAIDA AS Relationships Dataset (10/08/2009)



Δρ : difference of degree in logarithmic

Differences in degree do not characterize peering well.

though the orientation of transit is well inferred with degree.

Note; the distribution is normalized by area for each type of relationships.

PROPOSED METHOD

Concept

- Input (available information)
  - AS-level (quasi) spanning subgraph
    - contains almost all ASes
    - contains visible and invisible inter-AS links
      - We use “CAIDA AS Relationships Dataset (10/08/2009)” in this presentation.
- Method
  1. quantify AS size, which we call “*magnitude*”
  2. analyze differences in magnitude

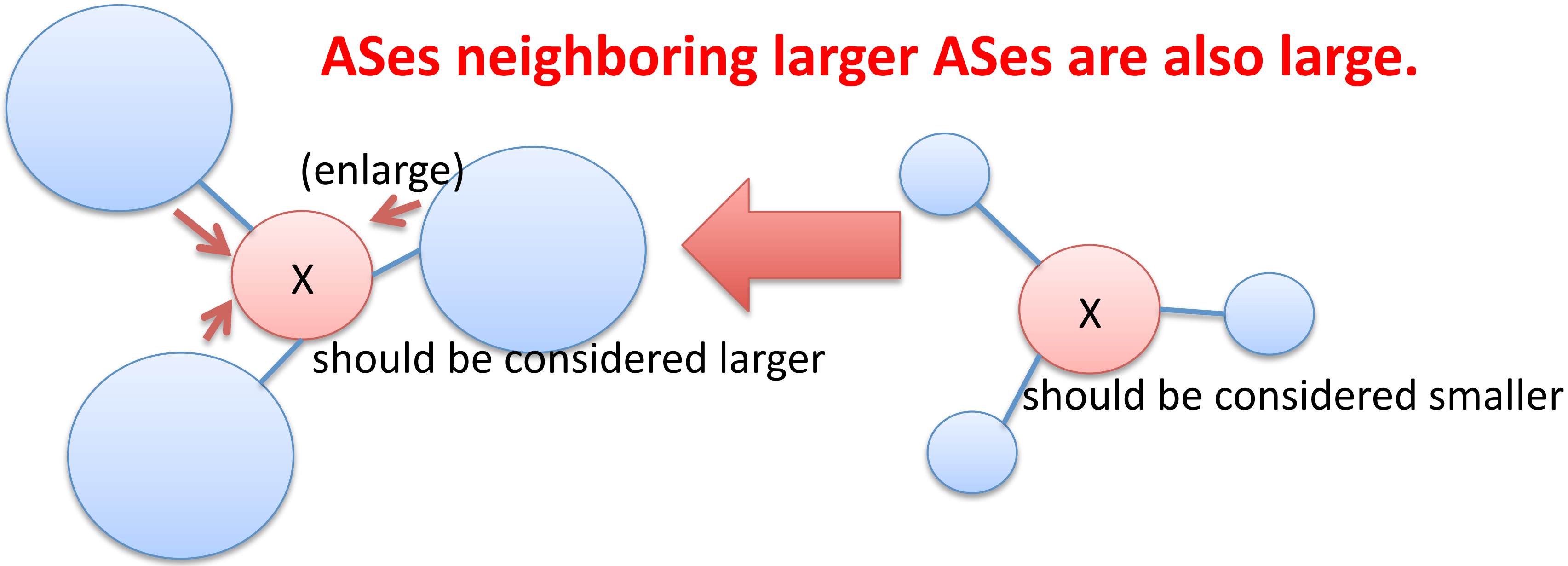
AS Magnitude Quantification

- AS magnitude
  - represents network scale of the AS
    - e.g., degree [Tangmunarunkit et al. 2001]
      - Note: Differences in degree do not represent peering well.
- For more accurate quantification
  - take into account the scale of neighbor ASes
    - e.g., An AS connecting to larger ASes is also larger, even though the AS has low degree.

How do we calculate AS magnitude? Main idea

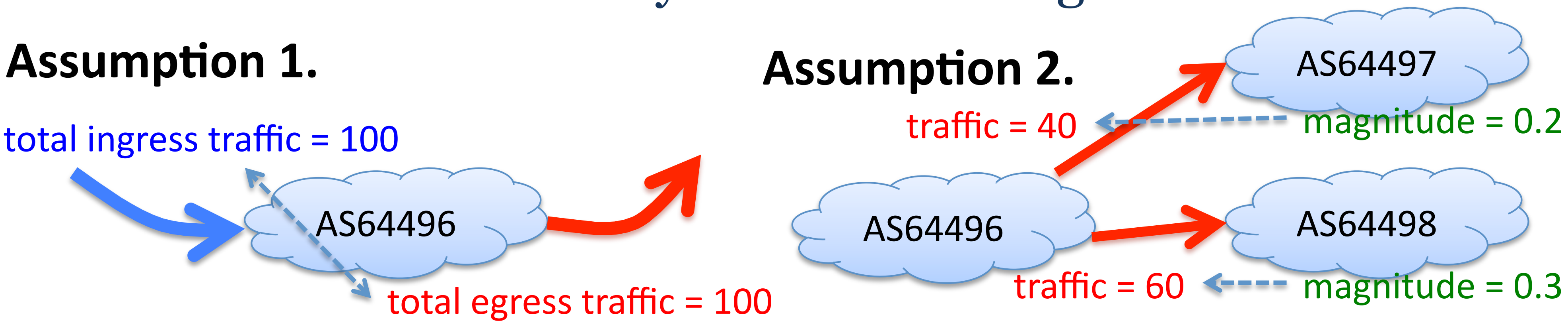
- take into account the magniutde of neighbor AS
  - note this results in recursive definition

ASes neighboring larger ASes are also large.



How do we calculate AS magnitude? Mapping into traffic transition model

- Simple model with three assumptions
  1. total ingress traffic = total egress traffic
  2. egress traffic: proportional to the neighbor AS's magnitude
  3. magnitude: proportional to the total ingress traffic in steady state of exchanged traffic



AS magnitude quantification – calculation procedure

Idea: calculate the traffic distribution and map it to the magnitude

(1) Define a weighted AS adjacency matrix

$$ {}^nA := \begin{pmatrix} {}^na_{11} & \cdots & {}^na_{1j} & \cdots & {}^na_{1m} \\ \vdots & \ddots & \vdots & & \vdots \\ {}^na_{i1} & \cdots & {}^na_{ij} & \cdots & {}^na_{im} \\ \vdots & & \vdots & \ddots & \vdots \\ {}^na_{m1} & \cdots & {}^na_{mj} & \cdots & {}^na_{mm} \end{pmatrix} $$

(i) $n = 0$ **random walk model for initial case**

$$ {}^na_{ij} = \begin{cases} 1 & : \text{if AS } i \text{ and AS } j \text{ are adjacent} \\ 0 & : \text{otherwise} \end{cases} $$

(ii) $n \geq 1, n \in \mathbb{Z}$

$$ {}^na_{ij} = \begin{cases} {}^{(n-1)}\rho_j & : \text{if AS } i \text{ and AS } j \text{ are adjacent} \\ 0 & : \text{otherwise} \end{cases} $$

(2) Equalize ingress and egress traffic; i.e., converting to <u>traffic transition matrix</u>

$$ {}^nT = \left(\frac{{}^na_{ij}}{\sum_k {}^na_{ik}} \right) $$

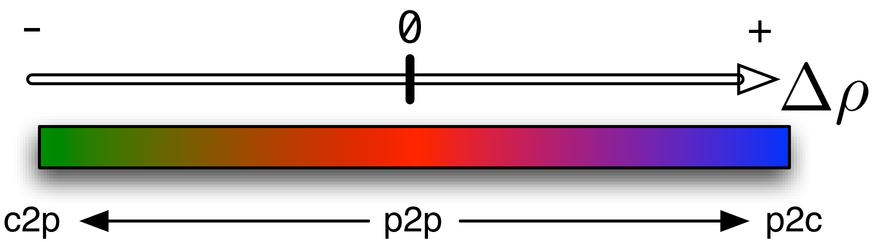
recursive definition

(3) Calculate the left eigenvector of *T* corresponding to the maximum eigenvalue

$\boxed{{}^n\boldsymbol{\rho}}$: the left eigenvector; the *i*-th element denotes the magnitude of AS *i*.

AS relationships estimation: the difference in magnitude

Idea: estimate the relationships from differences in magnitude



p2c: provider to customer
c2p: customer to provider
p2p: peer to peer

$$\Delta^{n}\rho_{i,j} := \log_{10}\left(\frac{{}^{n}\rho_i}{{}^{n}\rho_j}\right) = \log_{10}\left({}^{n}\rho_i\right) - \log_{10}\left({}^{n}\rho_j\right)$$

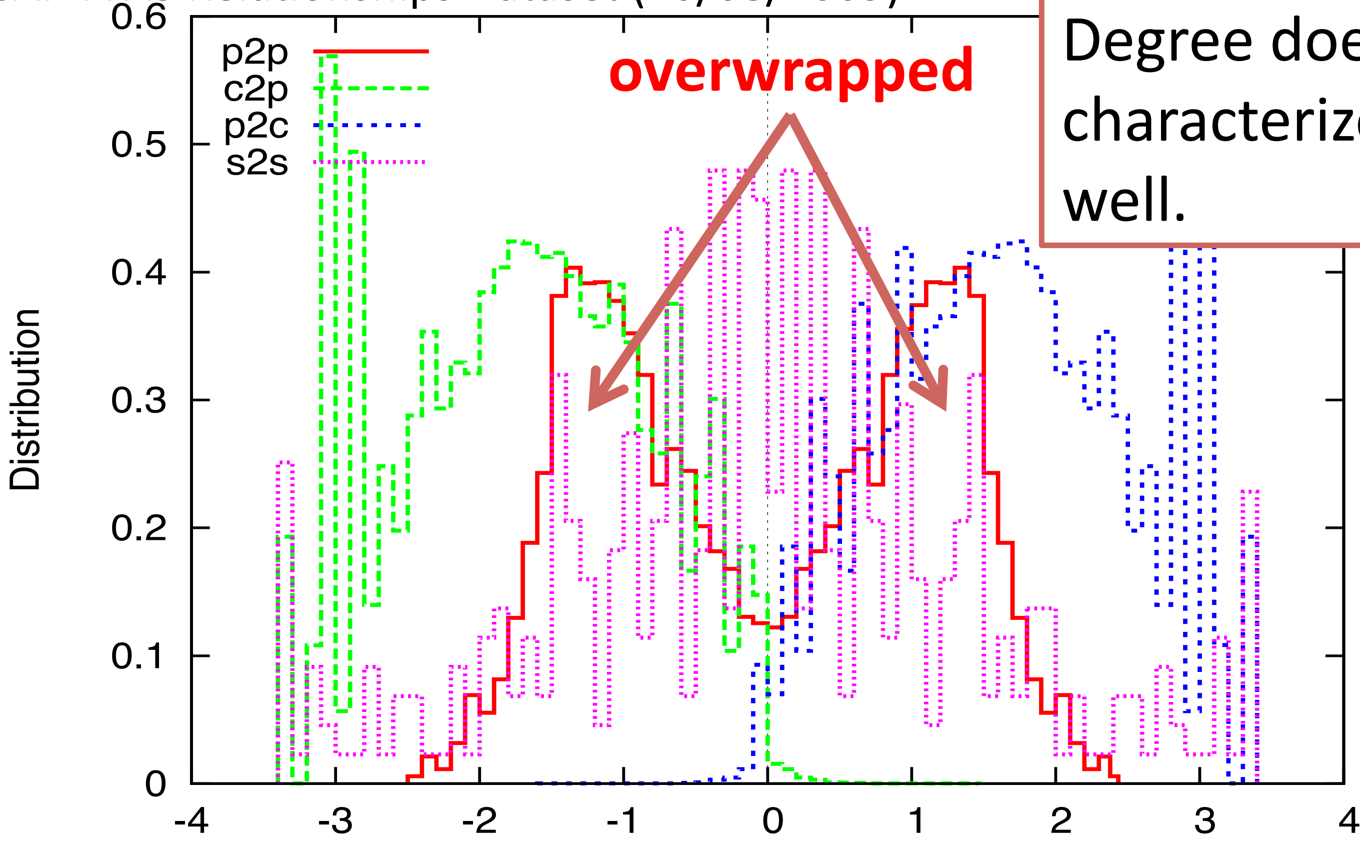
Evaluation and the Result

Evaluation 1

- Datasets
  - for quantification
    - CAIDA AS Relationships Dataset (10/08/2009)
      - as a spanning subgraph
  - for verification
    - CAIDA AS Relationships Dataset (10/08/2009)
      - as a "correct dataset"
- Evaluation method
  - draw distribution of differences in magnitude by each type of relationships
  - ROC analysis

PDF of difference in magnitude (n=0; i.e., degree)

Dataset: CAIDA AS Relationships Dataset (10/08/2009)



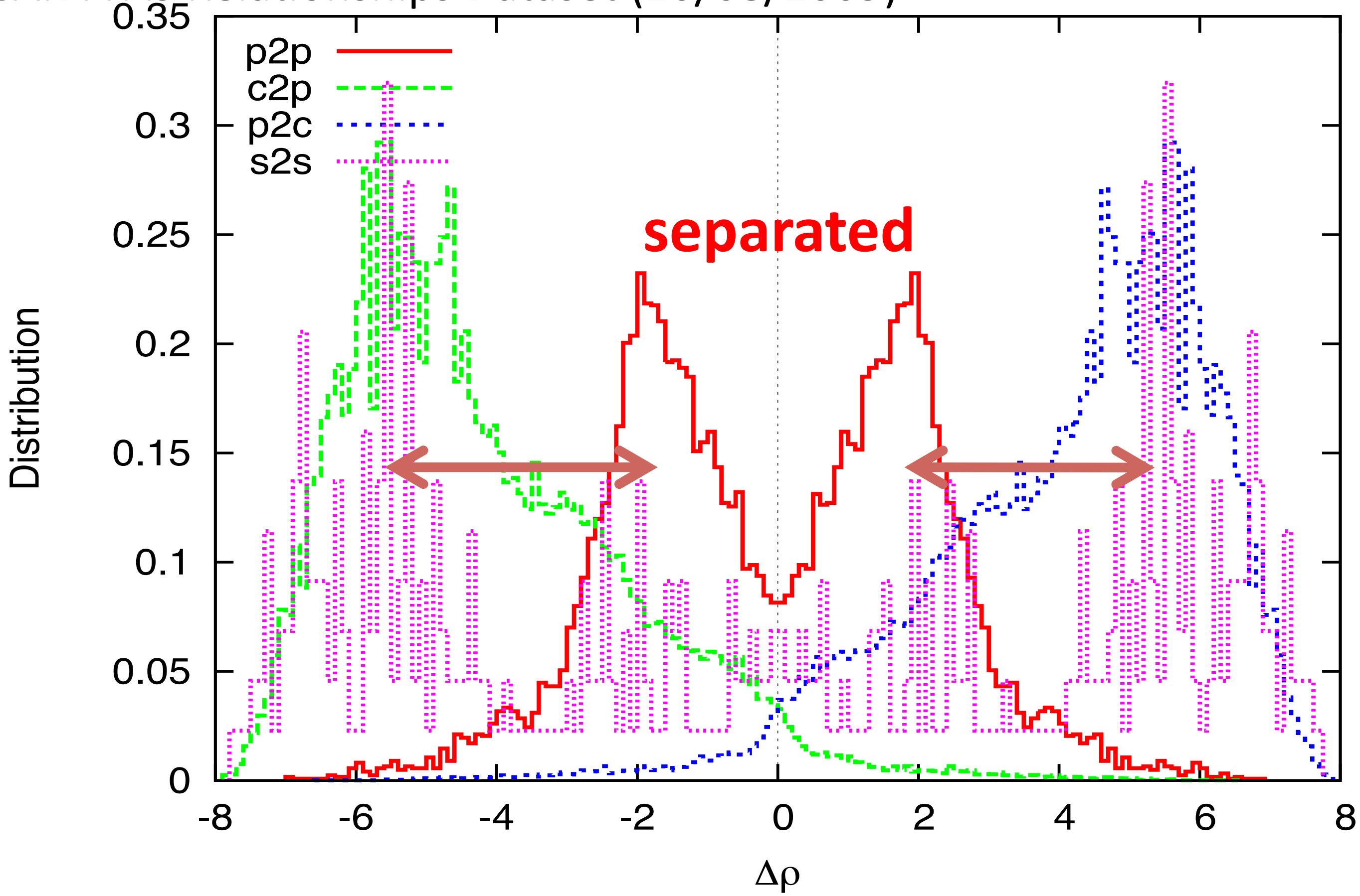
Degree does not characterize peering well.

$\Delta\rho$: difference of degree in logarithmic

Note; the distribution is normalized by area for each type of relationships.

PDF of difference in magnitude (n=2)

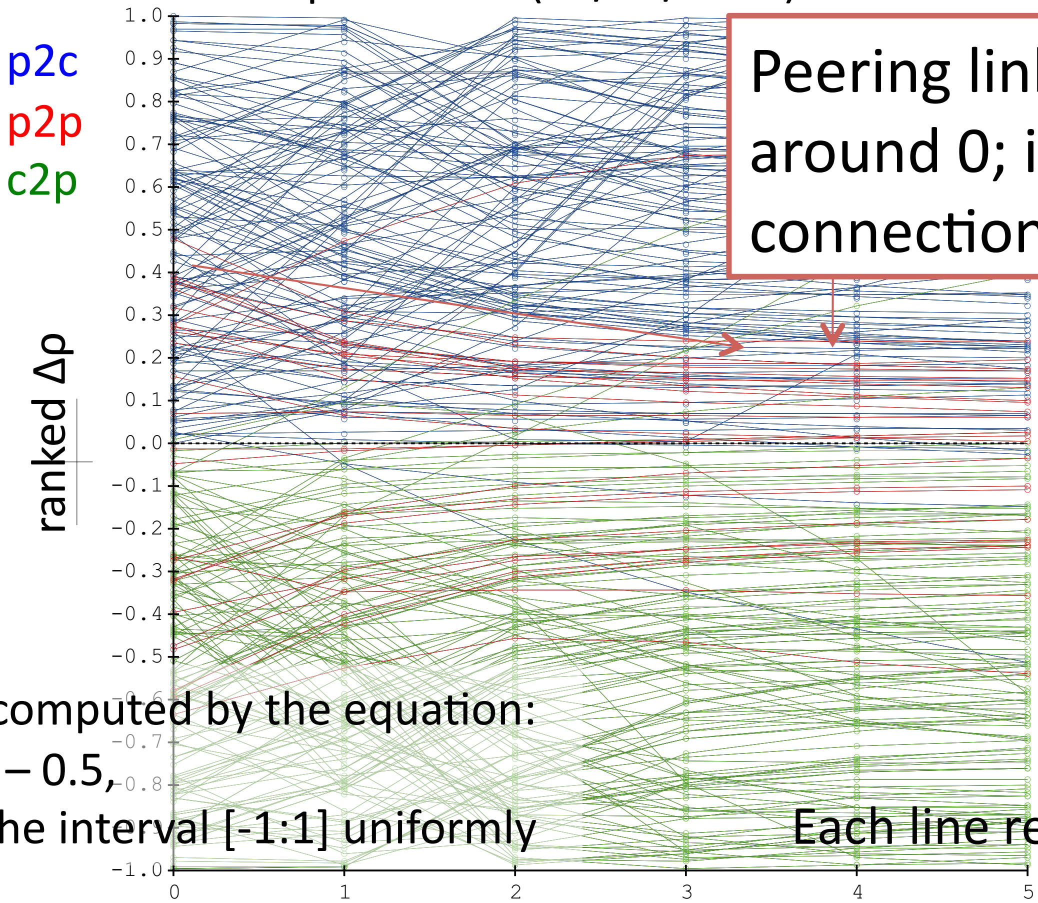
Dataset: CAIDA AS Relationships Dataset (10/08/2009)



Note; the distribution is normalized by area for each type of relationships.

How do the differences in magnitude change?

Dataset: CAIDA AS Relationships Dataset (10/08/2009)



Peering links' Δρ comes to around 0; i.e., equal-scale connection.

Note; ranked Δρ is computed by the equation:
rank-of-link/#links – 0.5,
i.e., distributed to the interval [-1:1] uniformly

Each line represents inter-AS link.

n (#recursion)

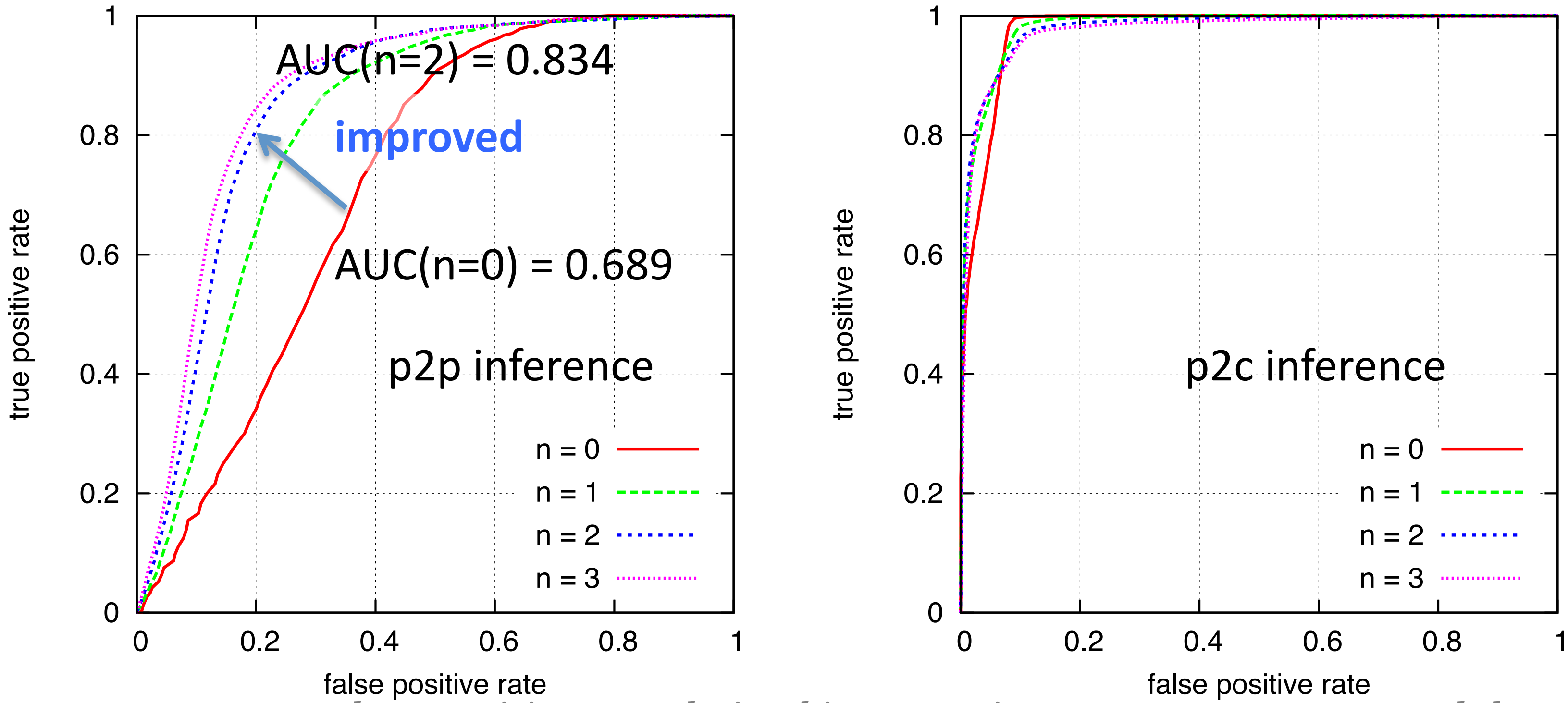
(*) randomly sampled at rate 1/500

ROC Analysis by giving a threshold

setting up a threshold

$$\begin{cases} \Delta^n\rho_{i,j} > {}^n\tau & \rightarrow \text{p2c (AS } i\text{: provider, AS } j\text{: customer)} \\ \Delta^n\rho_{i,j} < -{}^n\tau & \rightarrow \text{c2p (AS } i\text{: customer, AS } j\text{: provider)} \\ -{}^n\tau \leq \Delta^n\rho_{i,j} \leq {}^n\tau & \rightarrow \text{p2p} \end{cases}$$

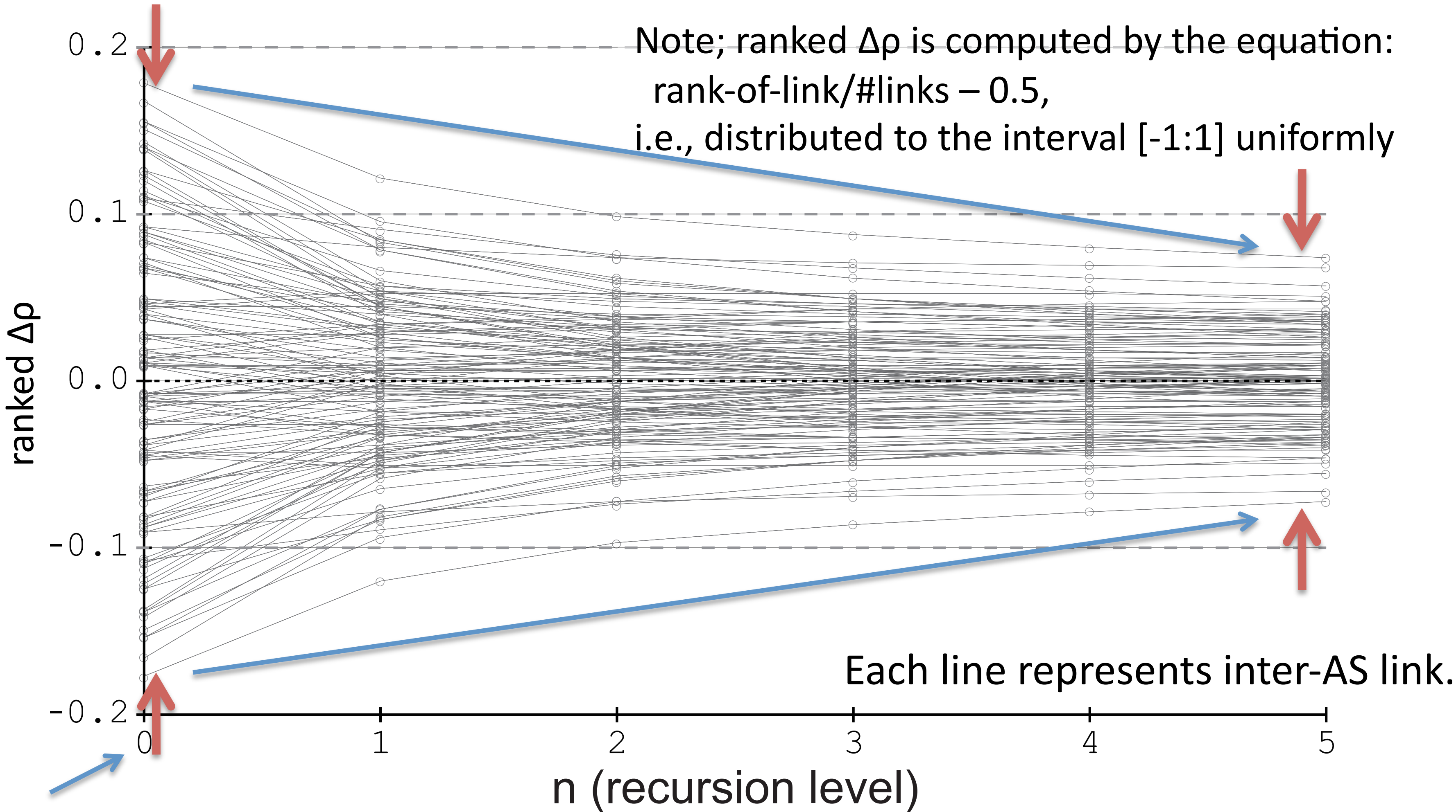
$$({}^n\tau \geq 0, {}^n\tau \ : \ \text{threshold})$$



Evaluation 2

- Datasets
  - for quantification
    - CAIDA AS Relationships Dataset (10/08/2009)
      - as a spanning subgraph
  - for verification
    - inter-AS links between well-known tier-1 ISPs
      - The links between tier-1 ISPs are considered "peering".
- Evaluation method
  - draw ranked difference in magnitude

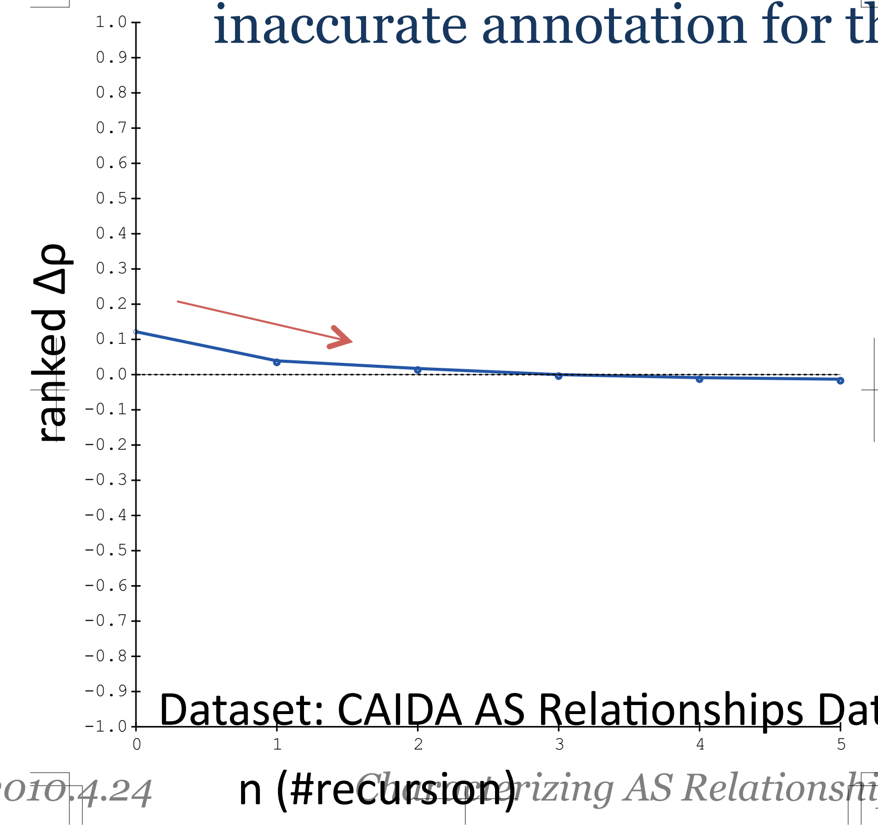
Peering characteristics (magnitude distance between Tier-1 ISPs)



recursion level = 0 → rank of difference in degree in logarithmic scale

Potential of finding inaccurate annotations: Is Verison-Verio transit?

- According to CAIDA dataset, Verison (AS701) is provider of Verio (AS2914).
  - Both are considered “Tier-1” ISP. CAIDA’s algorithm made inaccurate annotation for this link?



Dataset: CAIDA AS Relationships Dataset (10/08/2009)

The difference in degree is larger, but the difference in magnitude (n≥1) becomes smaller.

CAIDA’s annotation may be inaccurate. i.e., this link may be peering.

Conclusion

- We presented followings
  - quantify AS magnitude
    - by eigenvalue analysis
  - characterize AS relationships
    - by comparing the differences in magnitude
- contribution
  - proposed path-less (i.e., not paths but adjacencies) characterization method for AS relationships
  - showed the proposed method characterized the relationships appropriately
    - consider whether the proposed method is applicable to find "paid peer" in future

Thank you for your attention