# Network Syntax – *Inter-Domain Traffic Estimation for the Outsider*

Mario A. Sánchez, **Fabián E. Bustamante**, Bala Krishnamurthy, Walter Willinger, Georgios Smaragdakis, Jeffrey Erman

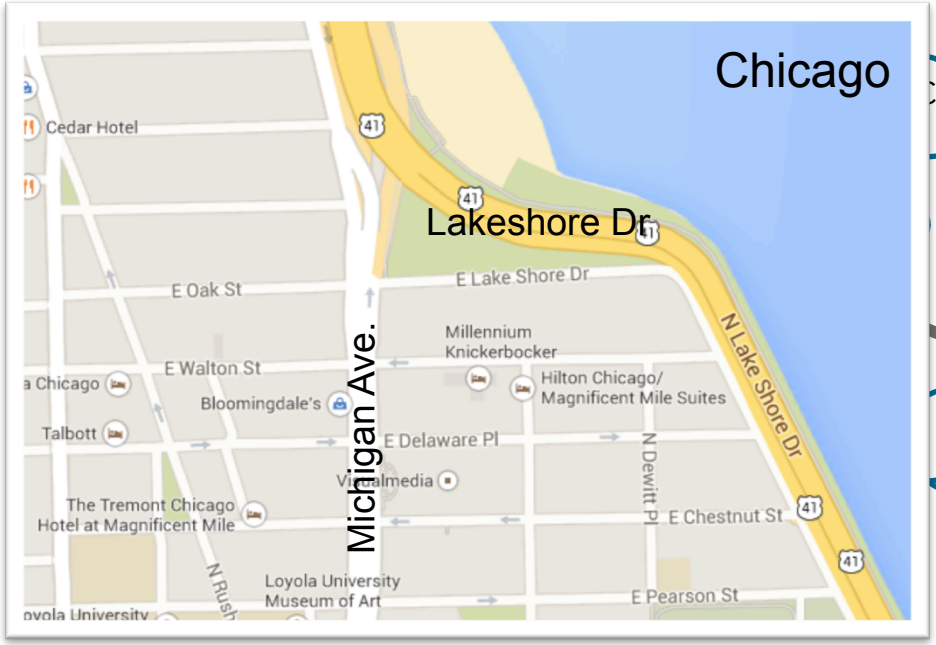
*Northwestern U., AT&T Labs-Research, Niksun, MIT/TU Berlin*

# Mario's babysitting his newborn…

*Daniella Sánchez García*

# Internet inter-domain system

- Most Internet inter-domain studies
  - Focused on network connectivity and dynamics …



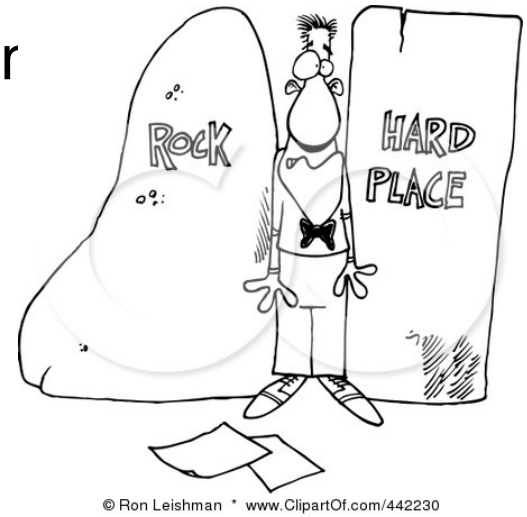


- Shifting focus to traffic …

# Inter-domain traffic studies – challenge

- Traffic is all that matters …
  - Network engineering, anomaly detection, economics …
- Challenge – few available traffic datasets

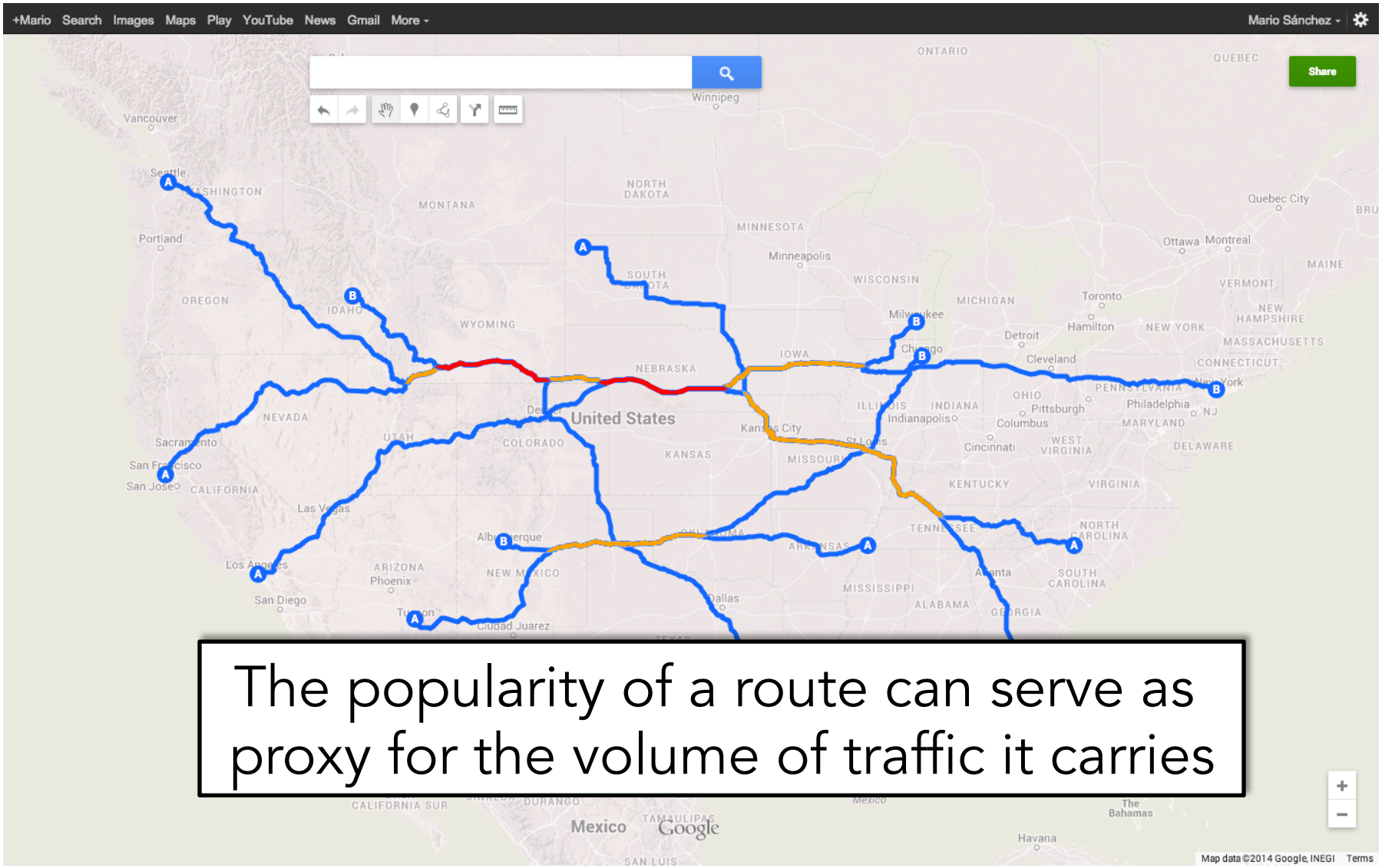
*Fine-grained* data or a small slice
- For insiders
- Can't reproduce
- Can't scale




Publicly available, *coarse* datasets
- Few
- Limited scope

- *Inter-domain traffic estimation for the outsider?*

# Pushing the analogy – Paths and traffic

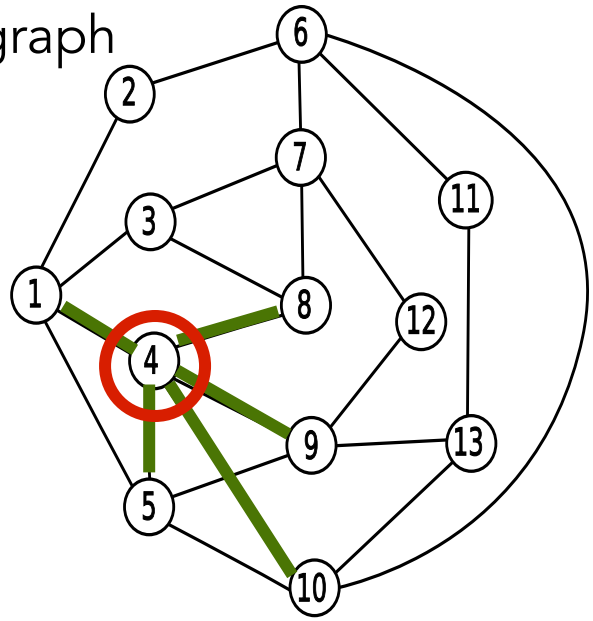


# Pushing the analogy – Paths and traffic

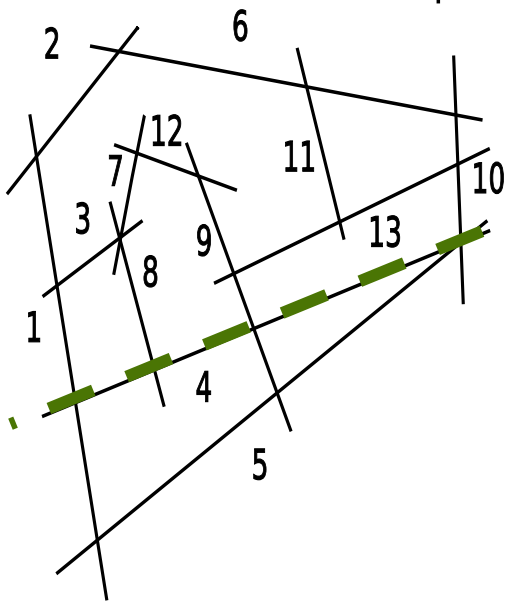
- Space syntax* – from urban planning
  - Some streets are more *central* than others
  - Predicting path traffic based on road structure

Connectivity (dual) graph

*Central node – Node with higher connectivity*




Axial map




*Hillier and Hanson, 1984

# Basic idea

- Use AS-level connectivity graphs as carved by large *traceroute* datasets
- Apply structural analysis to identify popular routes
  - Traceroutes tells us the route of packets
  - Many traceroutes let us infer popularity
- Use popularity as a proxy of traffic volume

*Network Syntax*

# Network Syntax *for the outsider*

- Leverage existing traceroute datasets
  - Publicly available data, no need for special buddies
  - Enabling a large-scale perspective
- Not just *any* AS-level connectivity graph
  - Basic AS connectivity says little about traffic (BGP)
- Several centrality metrics to capture popularity
  - Robust to idiosyncrasies of traceroute datasets

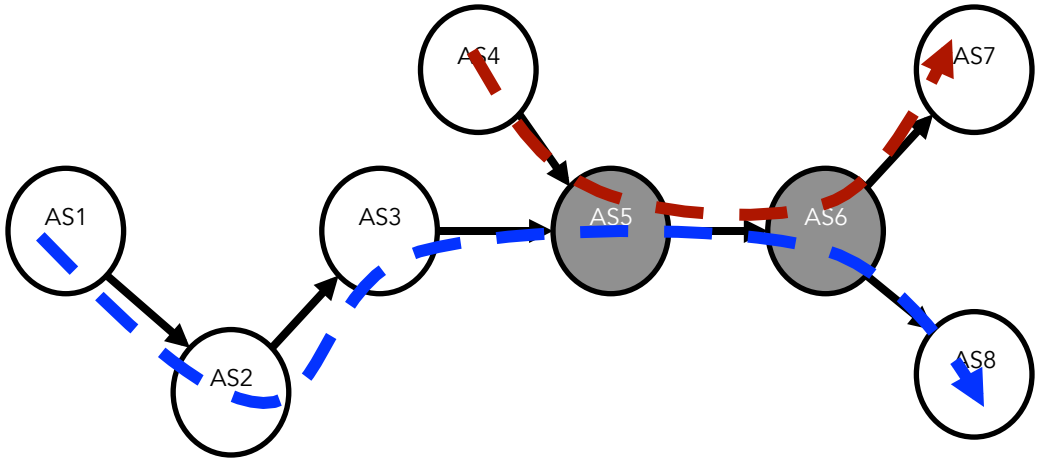
# Evaluating Network Syntax

- Compute metrics from traceroute datasets
  - Ono campaigns April 2011 & 2013
  - ~13M & 14M probes
  - 1-2K sources & 12K destination Ases

- Use ground-truth to check correlation to traffic
  - Global Tier-1 ISP – traffic exchanged with all customers
  - Large IXP – traffic exchanged over public peering fabric

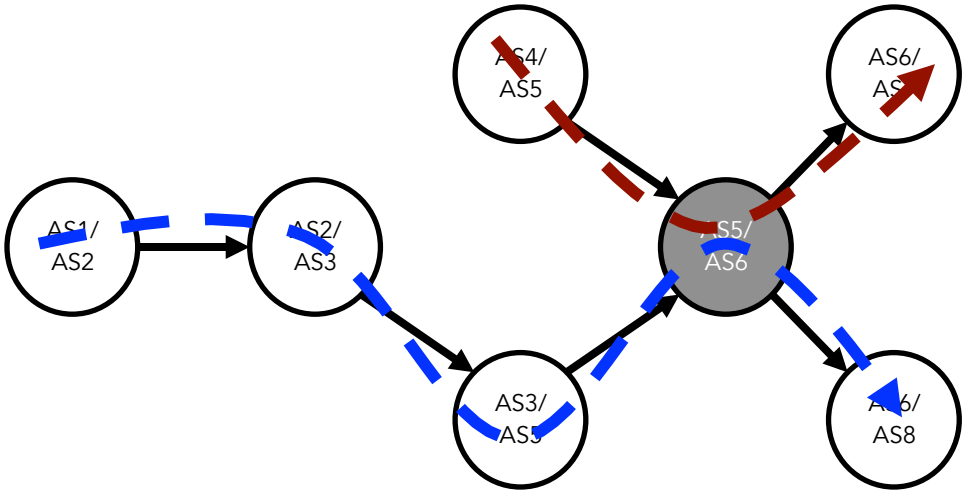
| # links | 2011 | 2013 |
|---|---|---|
| **IXP** | 5.6K | 4.5K |
| **ISP** | 257 | 147 |

# From traceroutes to connectivity graphs



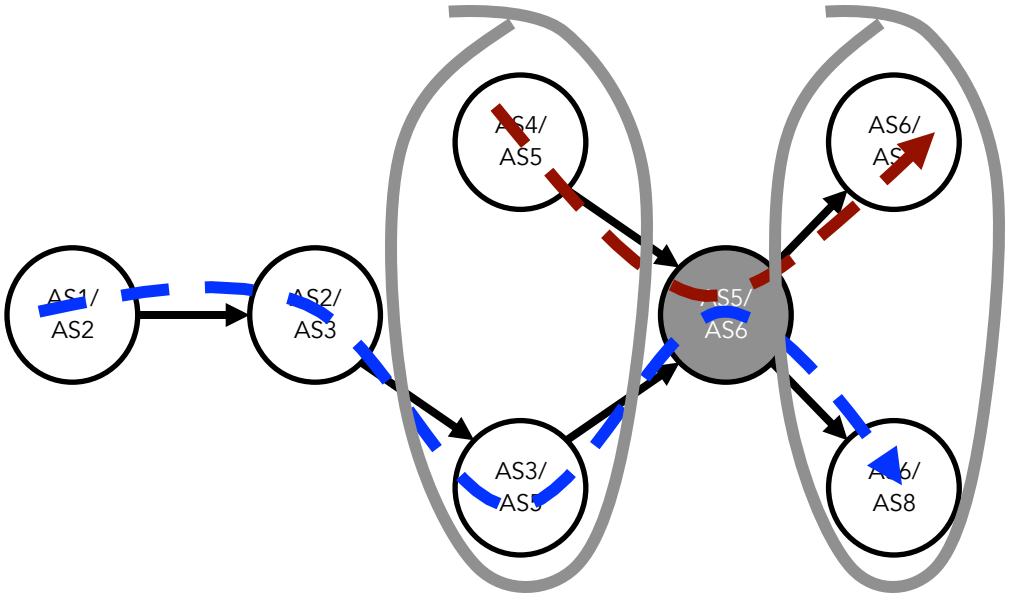

AS-link Traversing Path (ALTP) between AS5 and AS6

Partial connectivity graph



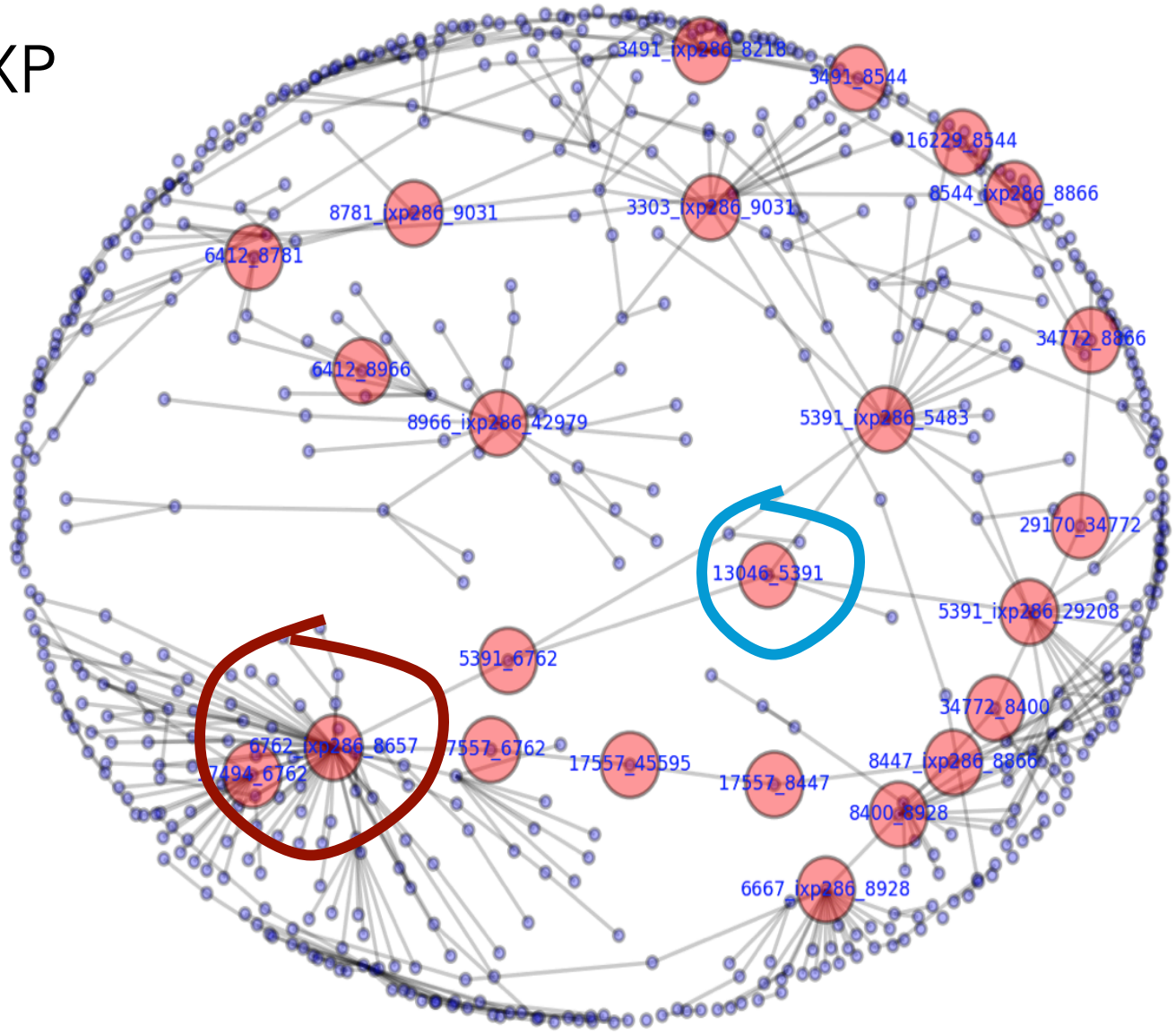
# Network Syntax' metrics of centrality

- New and adapted metrics from Space Syntax
- Two examples
  - **Connectivity** – #AS links that precede/succeed a node
  - **ALTP-frequency** – Relative cardinality of ALTP-set



# From traceroutes to connectivity graphs

Part of the IXP connectivity graph …

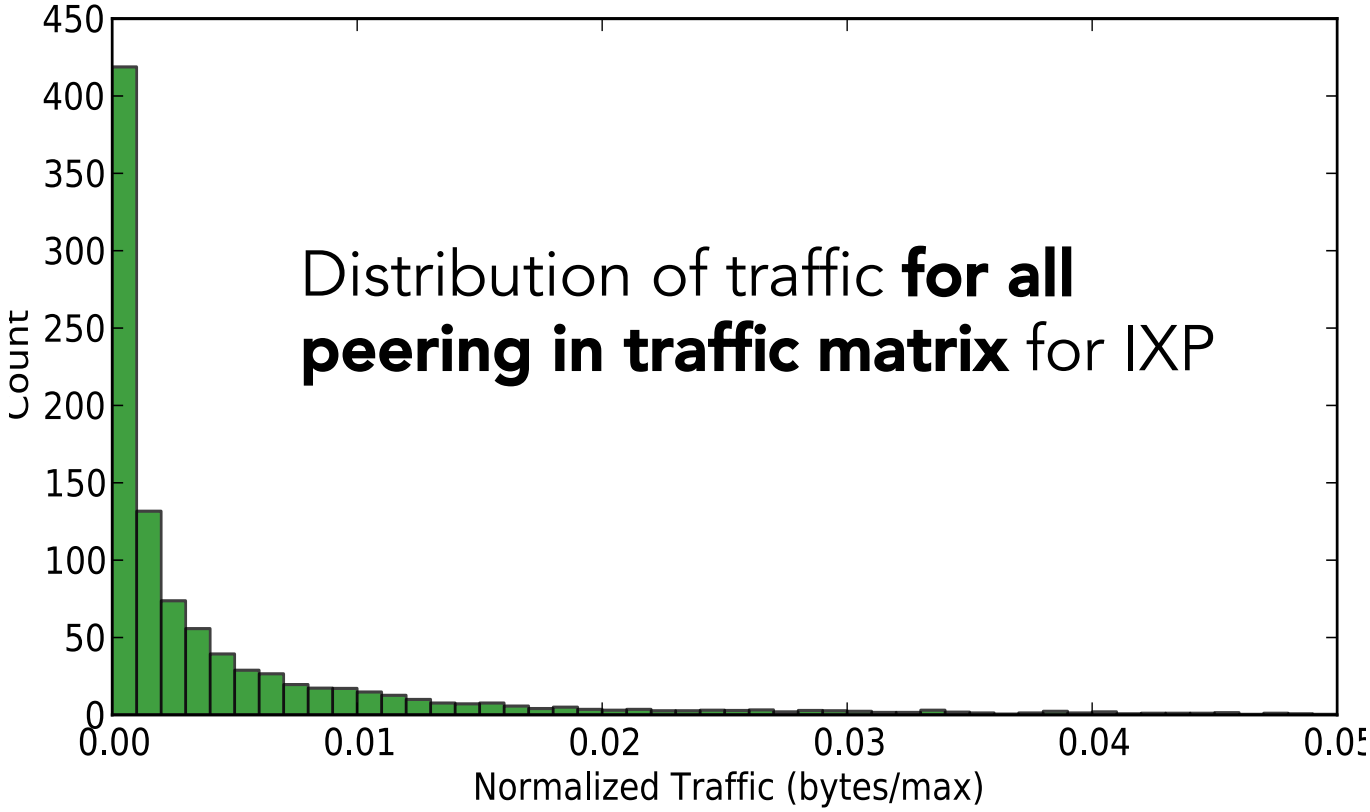


# Methodology…

- Generate AS-link connectivity graphs
- Compute Network Syntax metrics for each AS-link
- Order AS-links using each metric
- Cluster links in equal-size sets to reduce noise (e.g., from sampling)
- *Evaluate the correlation with traffic (validation)*
- Apply it to; e.g. predicting missing traffic link volumes

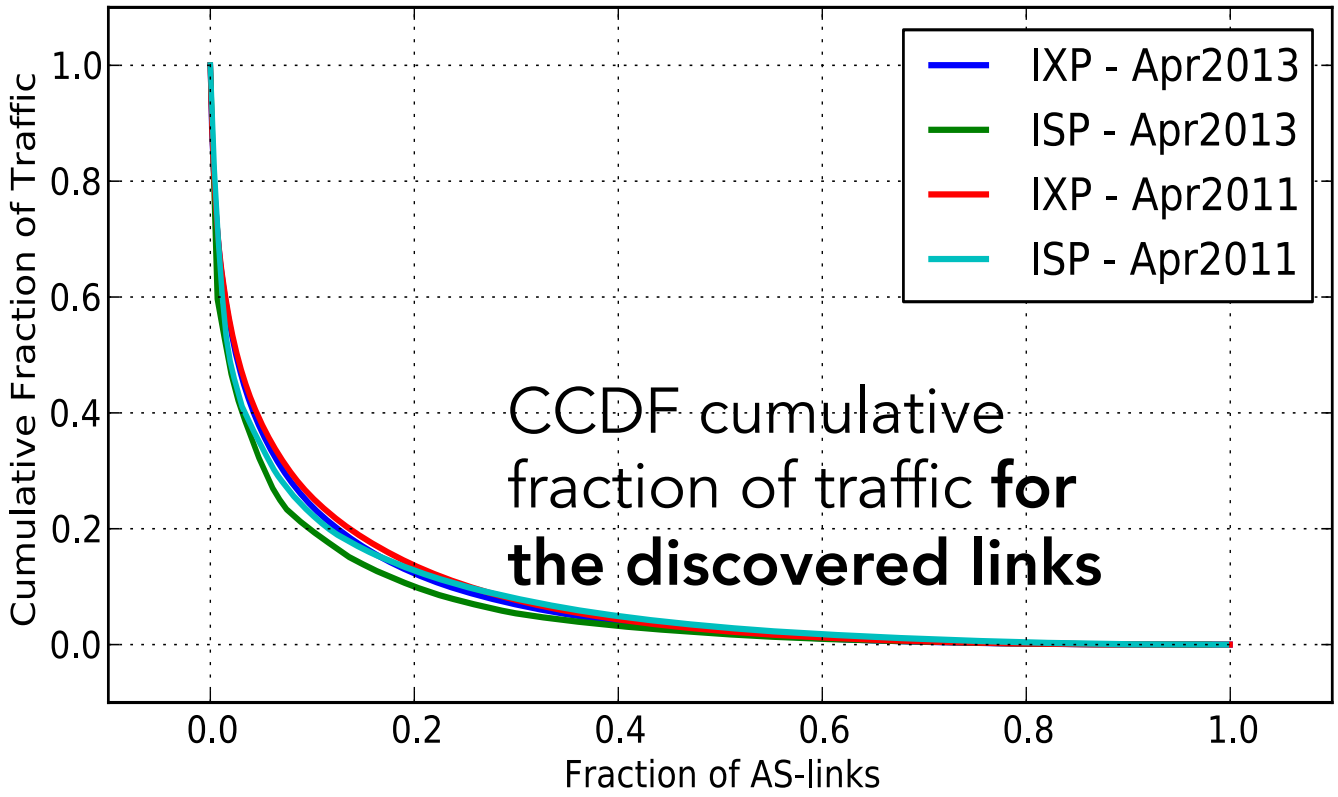
# Traffic and popularity – highly skewed

- As most popularity measurements, long-tail distributions



# Traffic and popularity – highly skewed

- As most popularity measurements, long-tail distributions
- Transform to remedy failure of expected normality



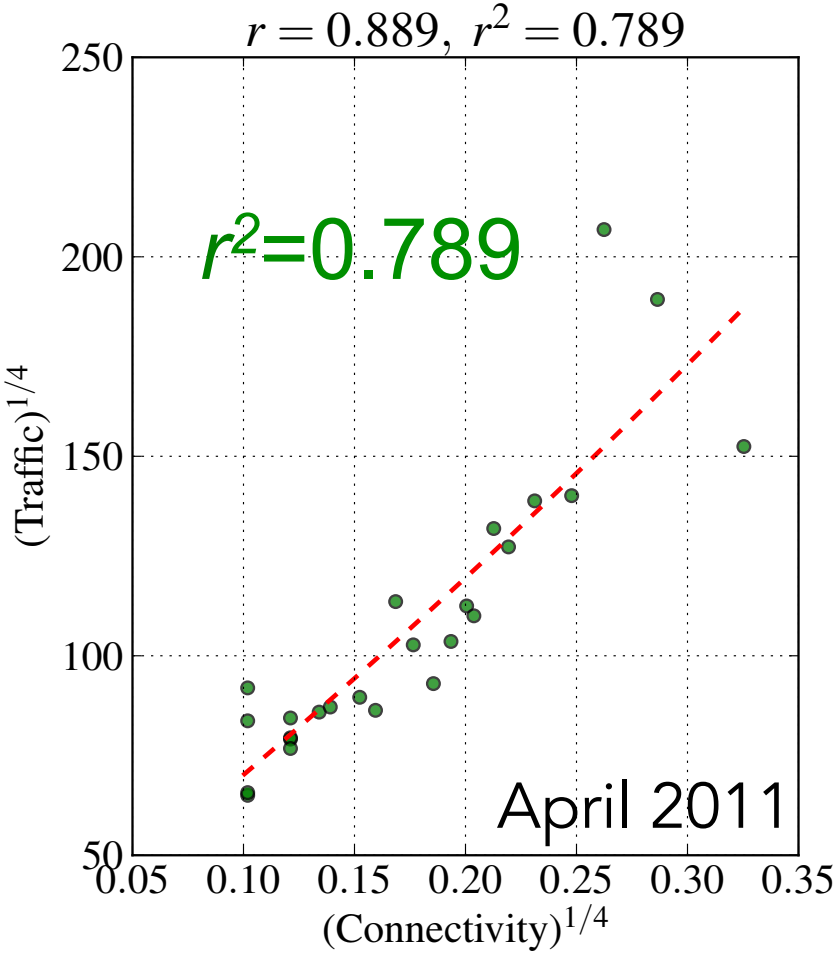
# Let's now look at those correlations

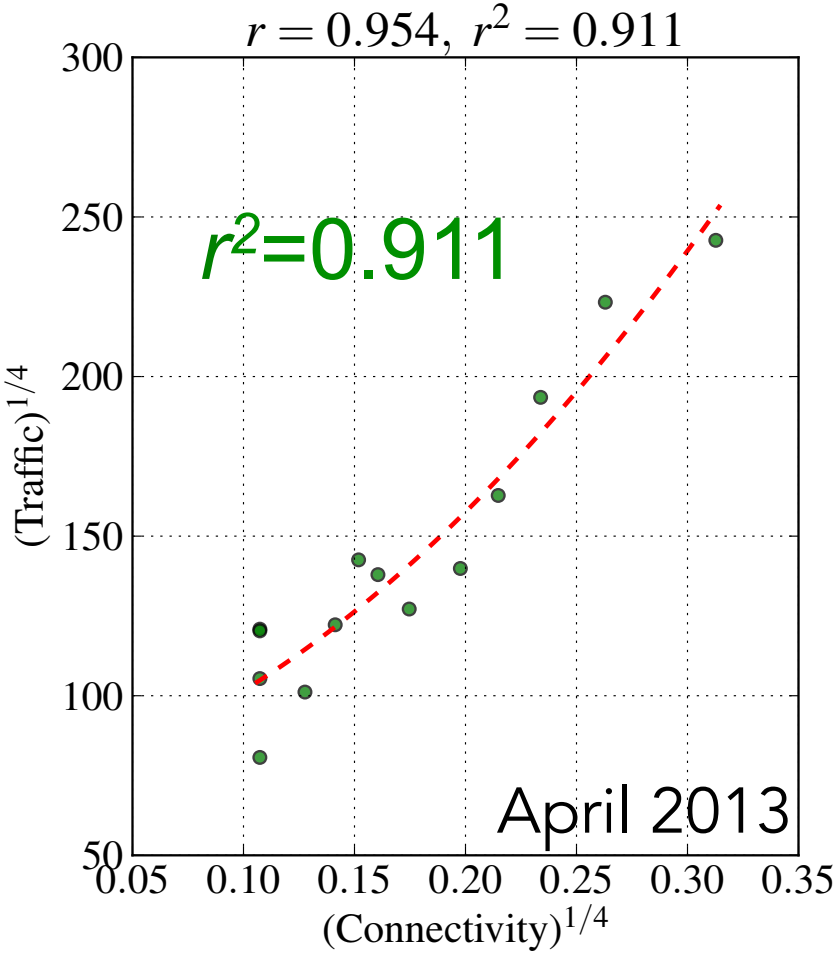
- Correlation between centrality and traffic in both ISP and IXP
- For both datasets, 2011 and 2013
- Different centrality metrics
- and different traceroute datasets

- … this is just a sample …

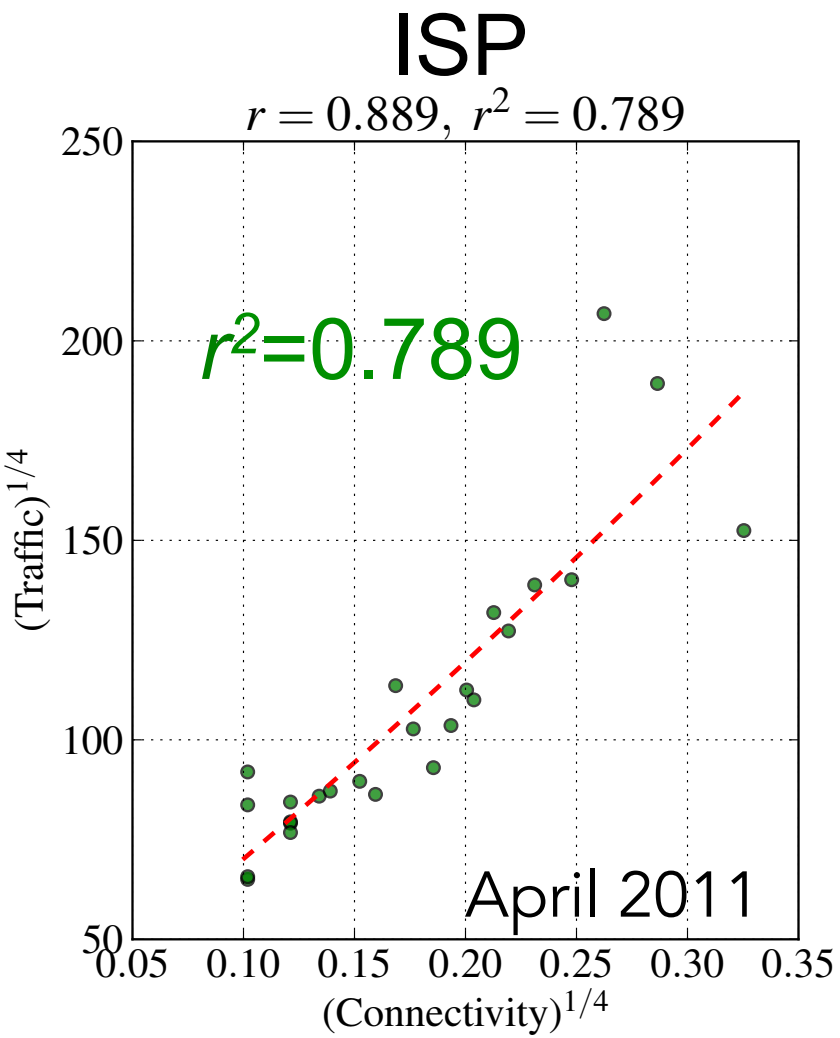
# Network Syntax – Popularity and Traffic

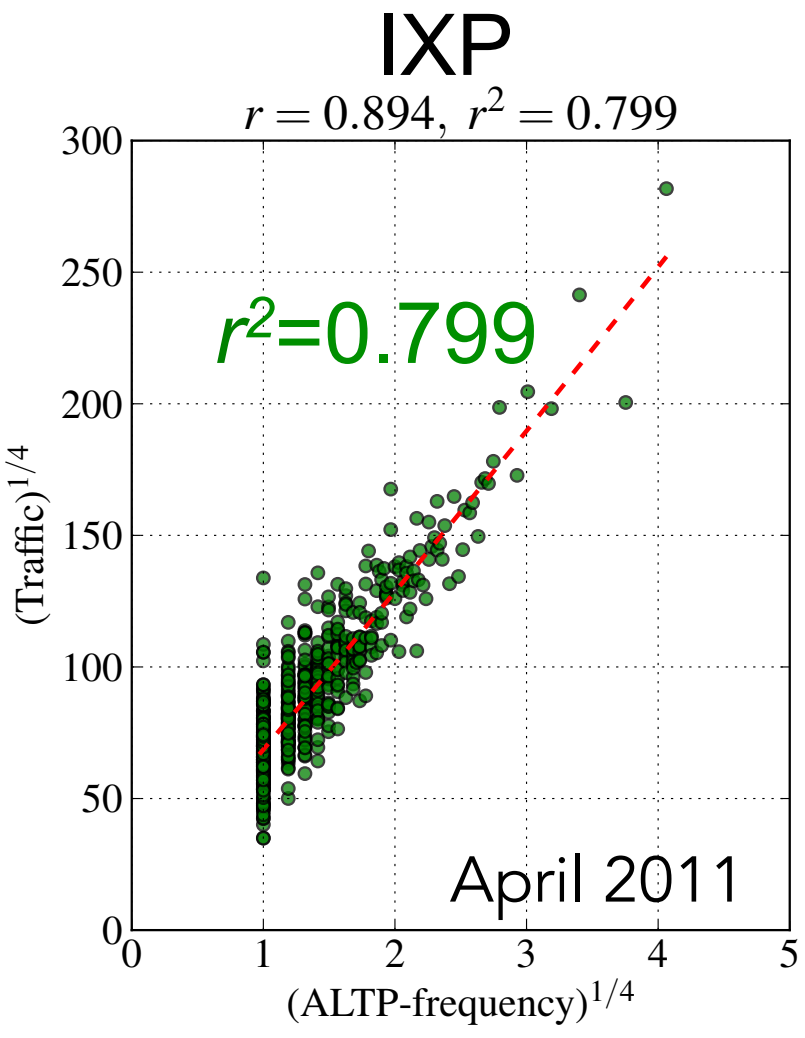
ISP



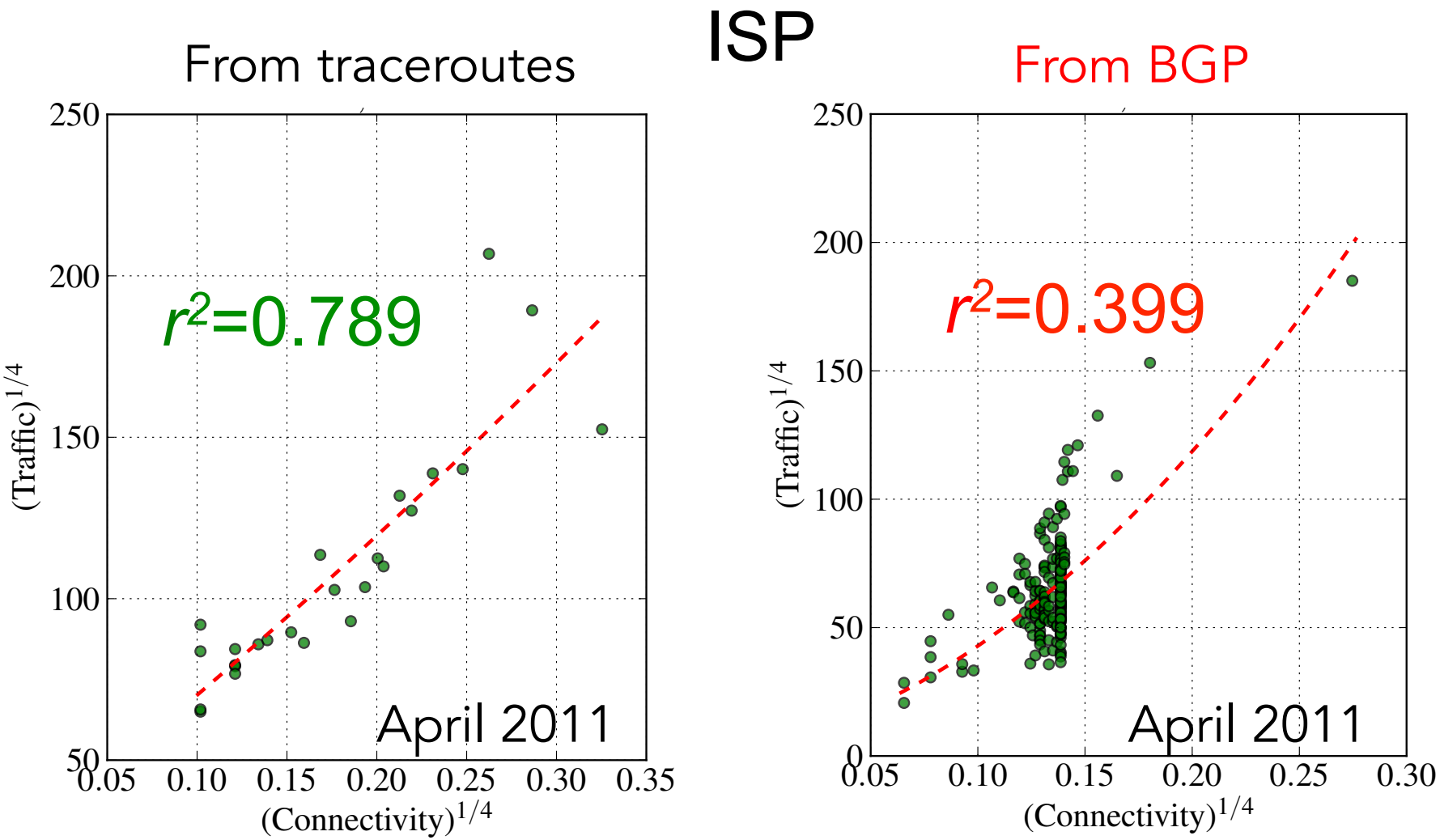



# Network Syntax – Popularity and Traffic



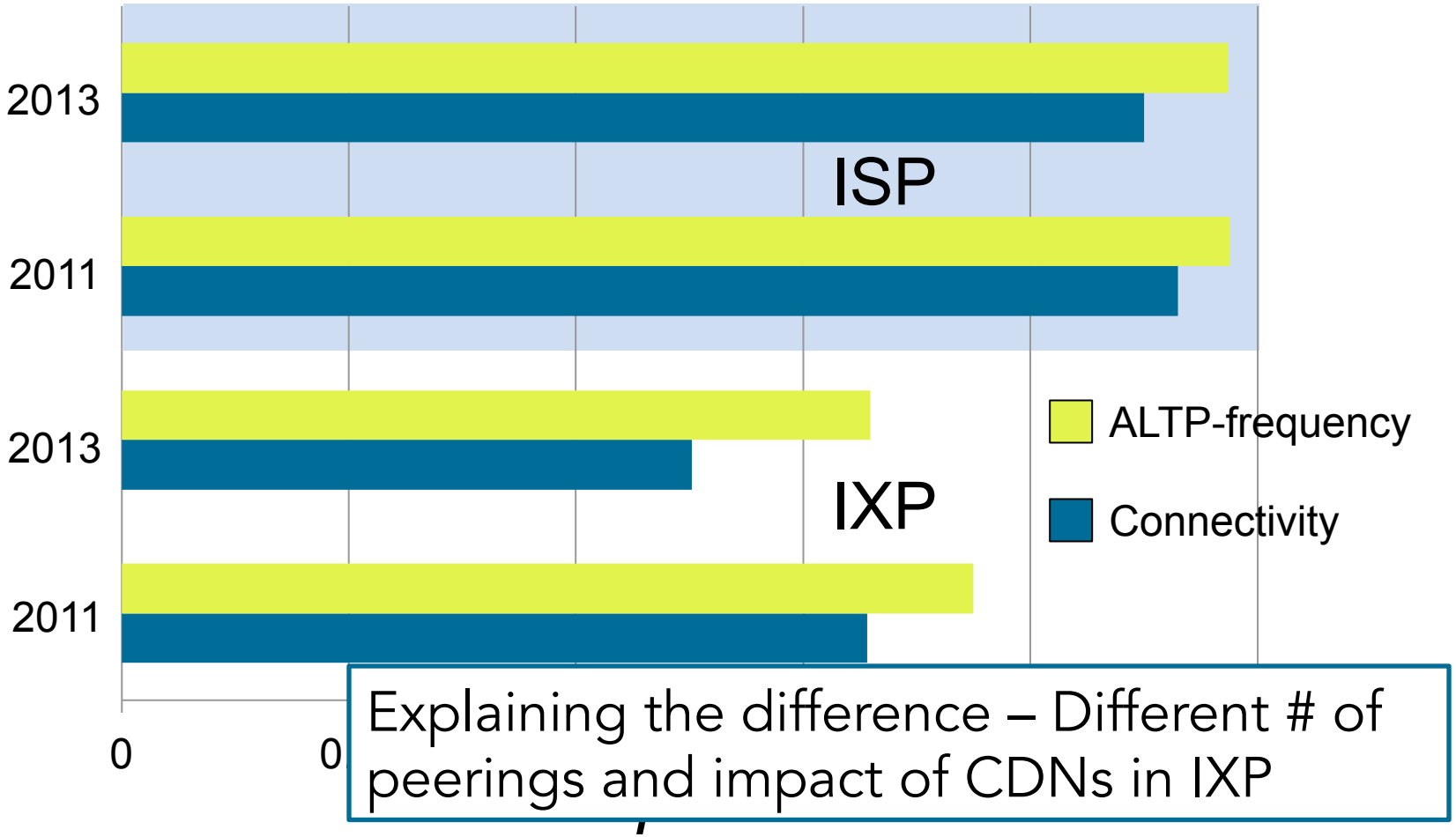
# Not just *any* AS-level connectivity graph!



# So, what can you do with that?

- Rank AS-link based on traffic volume
  - To check – rank AS-link based on traffic using ground-truth data, look at the correlation
- Predict link traffic
  - Use (different) subsets to compute correlation, use regression line to estimate traffic

# Estimating AS-links traffic-based ranking



# Summary

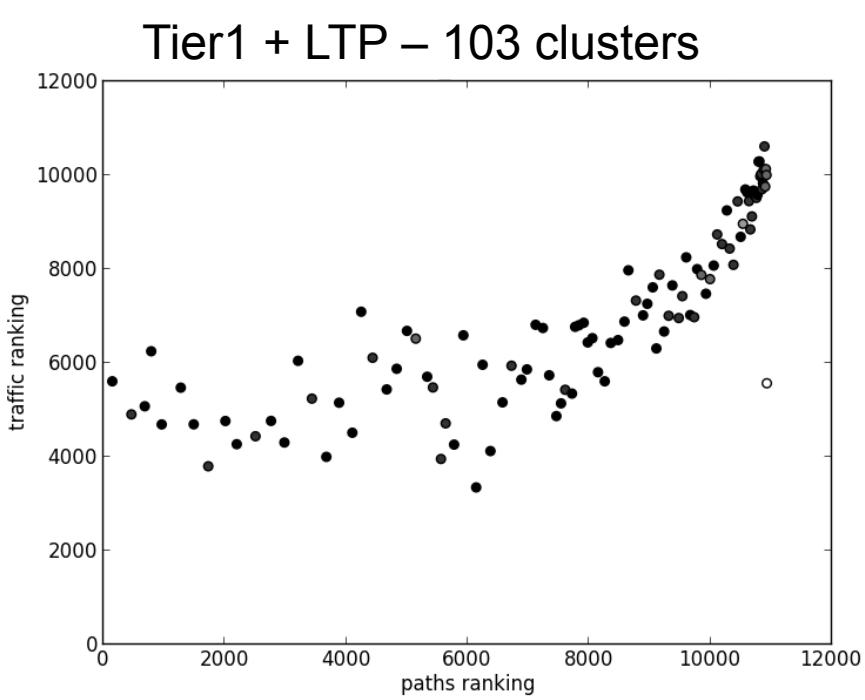
- Network Syntax – Inter-domain traffic for the outsider
  - Leveraging existing datasets
  - With easy-to-compute metrics of popularity
  - As proxies for traffic volumes

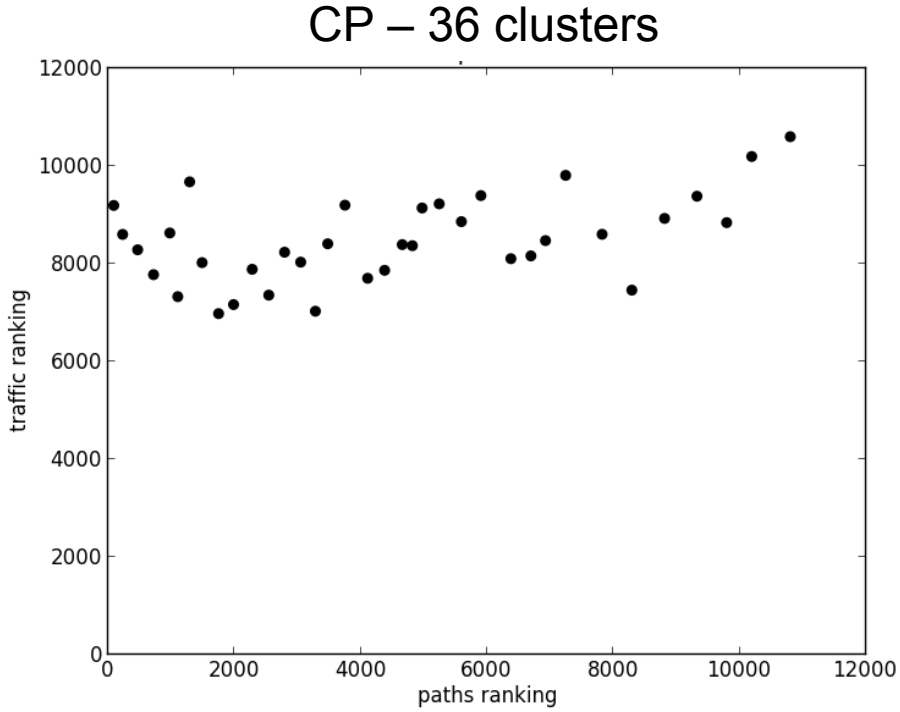
- A rich open field
  - *What characterizes a good traceroute dataset?*
  - *What happens in the reverse path?*
  - *What other applications are possible?*
  - …

# Network Syntax – *Inter-Domain Traffic Estimation for the Outsider*

Mario A. Sánchez,
**Fabián E. Bustamante**,
Balachander Krishnamurthy,
Walter Willinger,
Georgios Smaragdakis,
Jeffrey Erman

# Cluster Breakdown by AS type

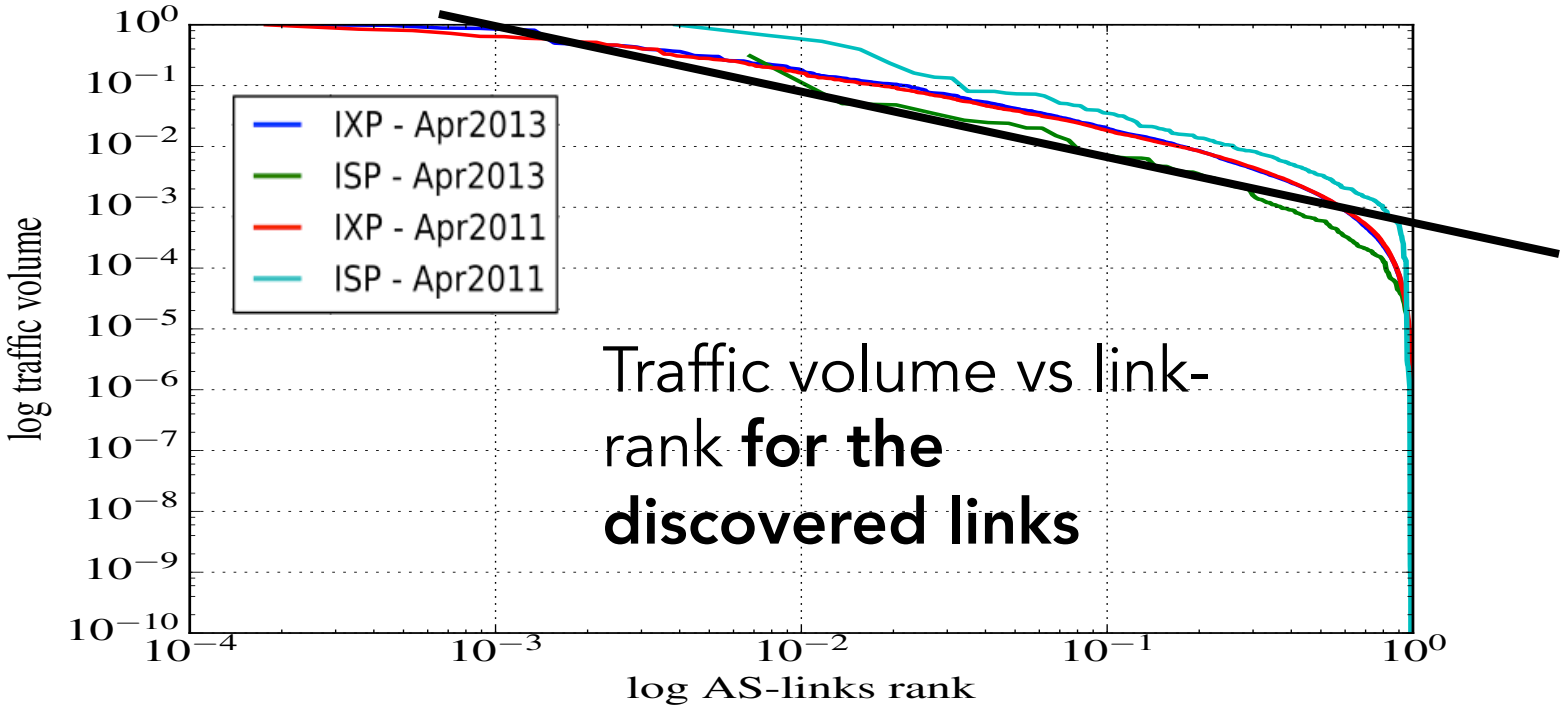



Regardless of ranking based on ALTP-frequency… traffic is high -> we are unable to capture it (only see one side of the peering?)

# Traffic and popularity – highly skewed

- As most popularity measurements, long-tail distributions
- Transform to remedy failure of expected normality

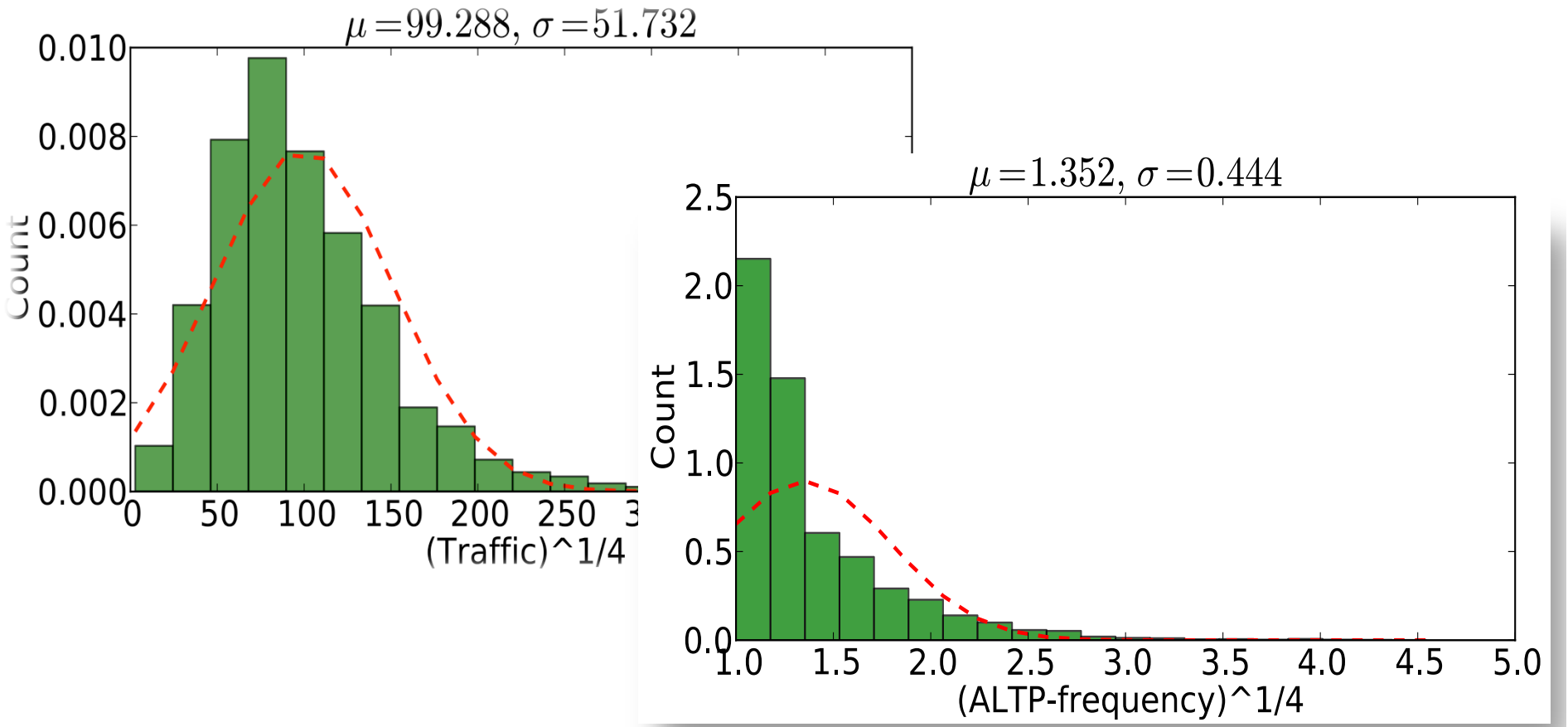


# Network Syntax' metrics

- **Connectivity** (degree centrality) – # of directly linked or neighboring nodes
- **Control value** (linked to clustering coefficient) # of alternative connections each of these neighbors has; measures the degree to which space controls access to its neighbors
- **Global choice** (betweenness centrality) – captures how often each line is used on topologically shortest paths from all lines to all other lines in the system
- **Integration** (a type of normalized closeness centrality) – measures the mean distance between every segment and all other segments in the system
- **ALTP-frequency** - …

# Transformations for normality

- Transform to remedy failure of expected normality



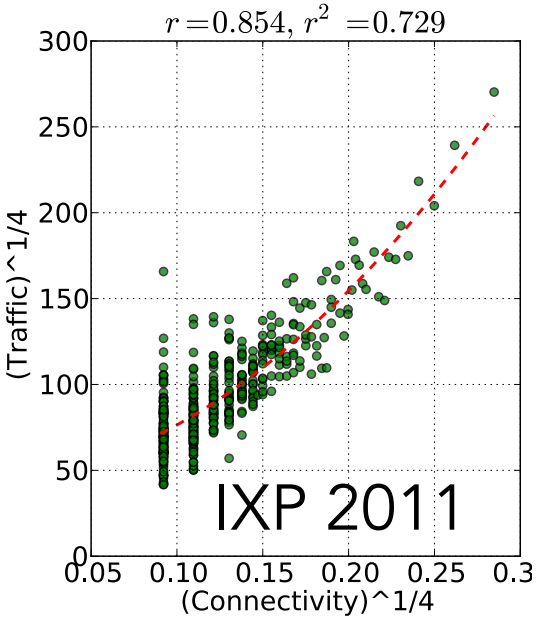
# Strong correlation – proxies for traffic

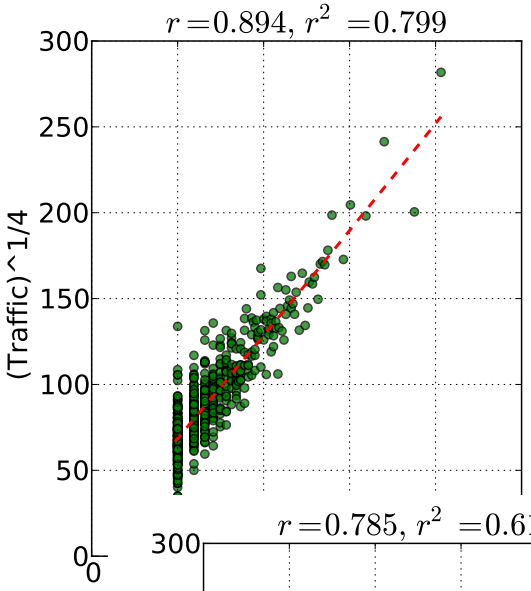
- $r^2$ values for all metrics with ISP and IXP

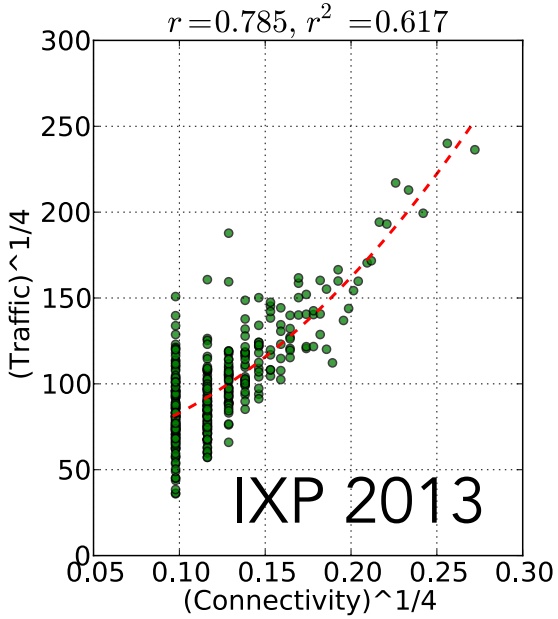
| | IXP | | ISP | |
|---|---|---|---|---|
| | Apr 2011 | Apr 2013 | Apr 2011 | Apr 2013 |
| Connectivity | 0.720 | 0.617 | 0.789 | 0.954 |
| Control value | 0.685 | 0.521 | 0.759 | 0.750 |
| Global choice | 0.661 | 0.580 | 0.653 | 0.903 |
| Integration | 0.575 | 0.356 | 0.826 | 0.629 |
| ALTP-frequency | 0.799 | 0.713 | 0.965 | 0.958 |

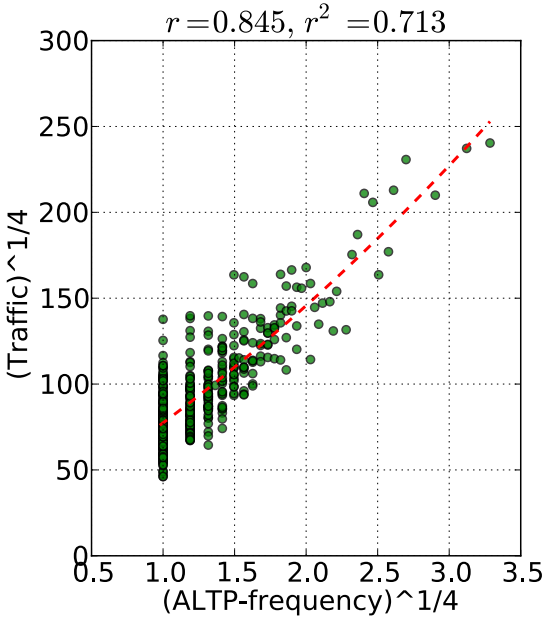
*Network Syntax centrality metrics are great proxies for traffic*

# Centrality metrics as proxies for traffic



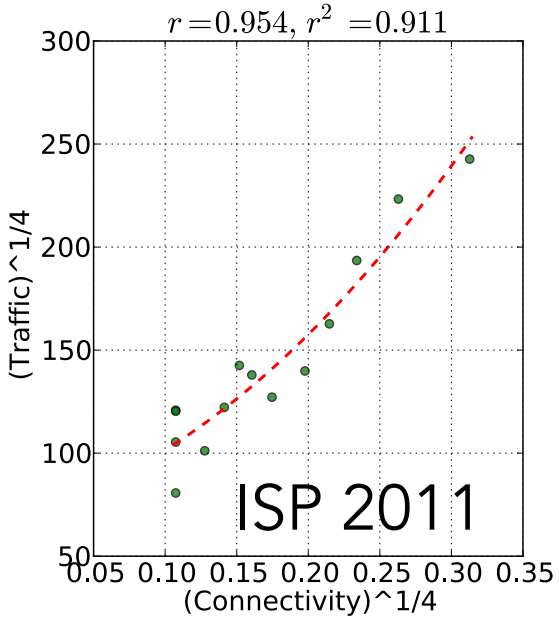



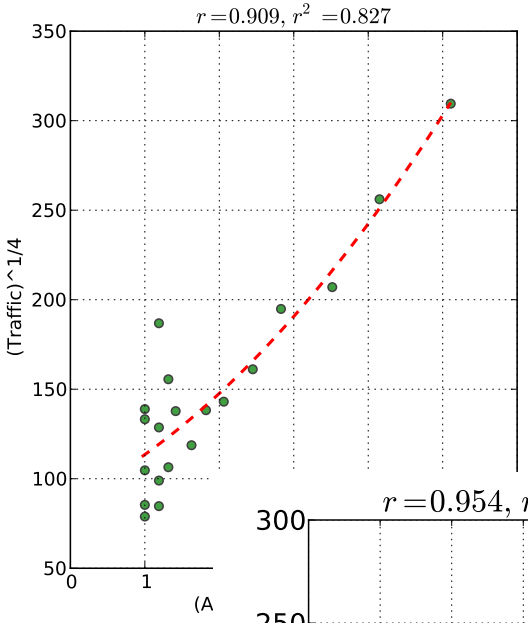



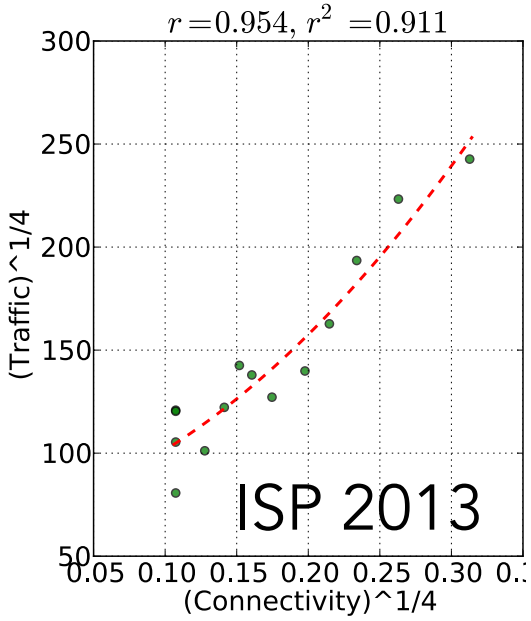

Correlation between connectivity and ALTP-frequency, and IXP traffic

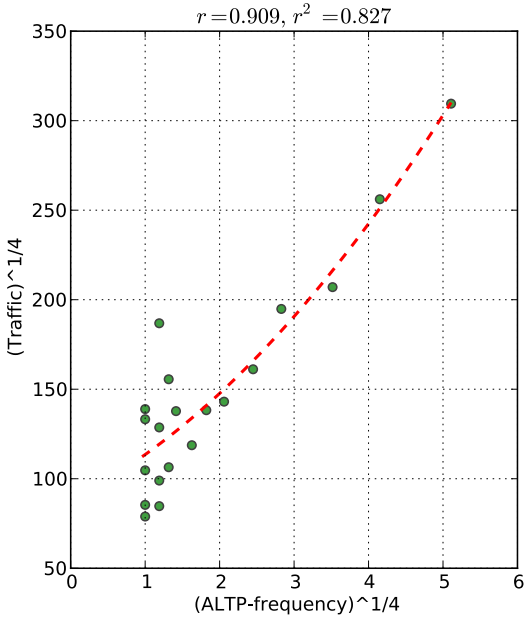
# Centrality metrics as proxies for traffic



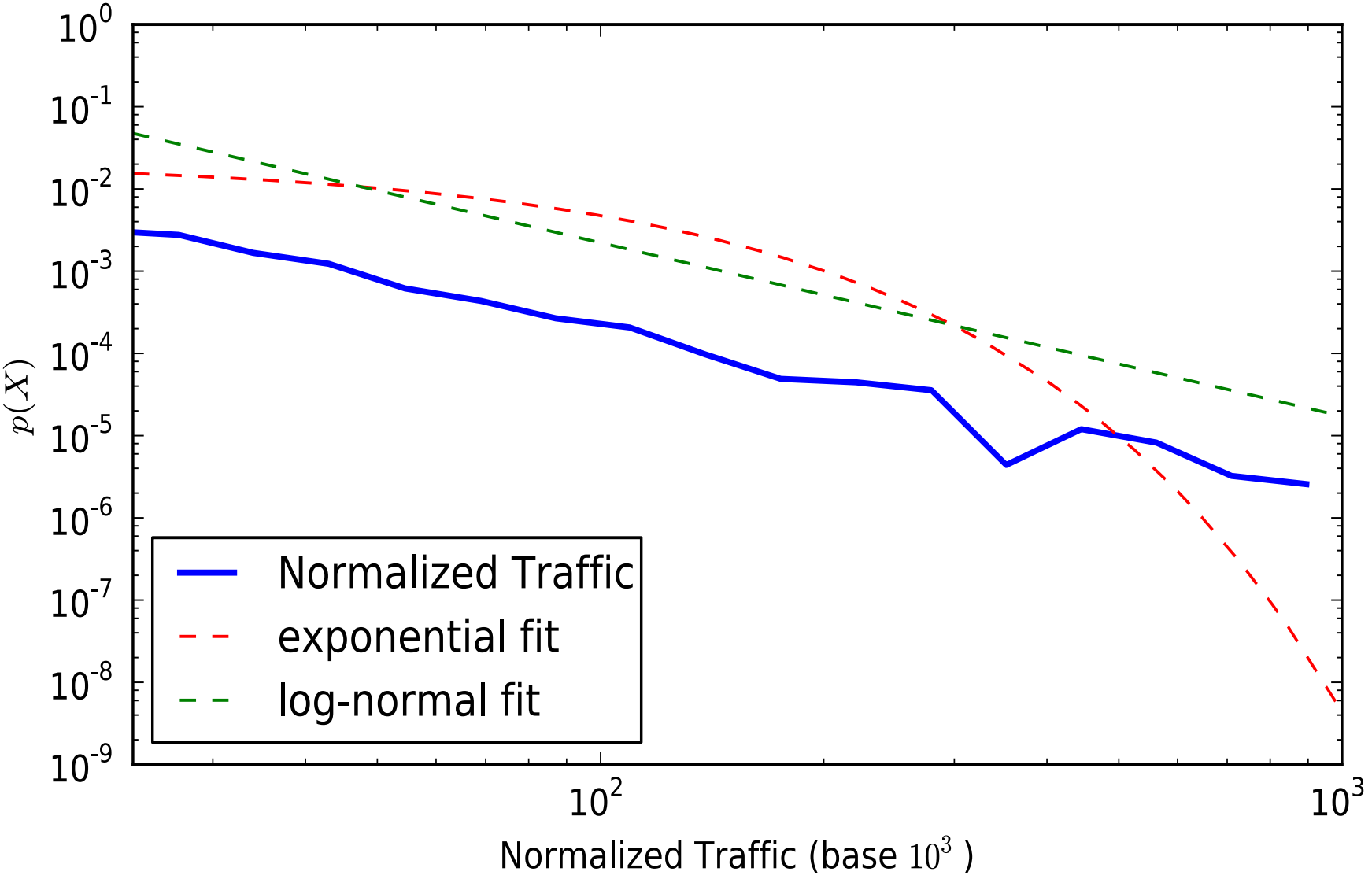




Correlation between connectivity and ALTP-frequency, and ISP traffic

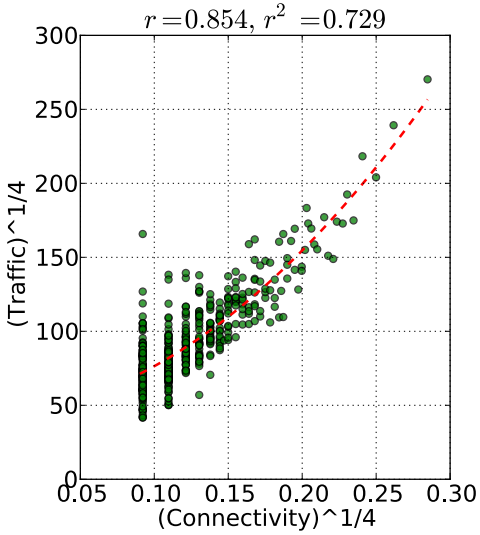


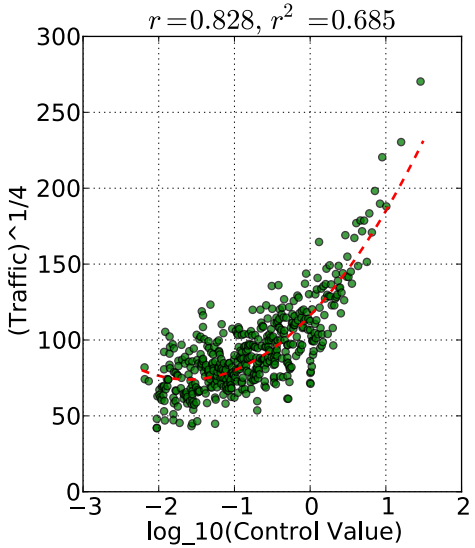



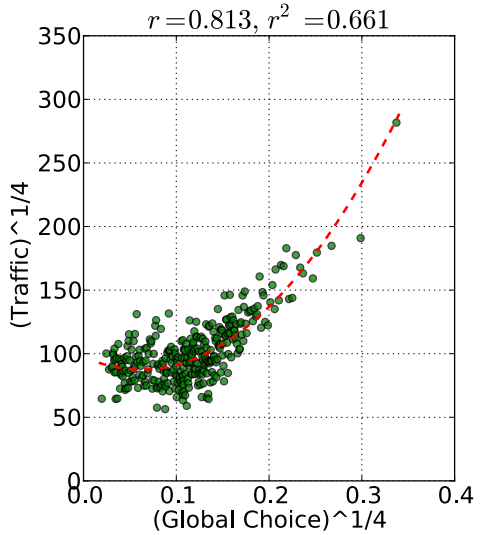
# Goodness of fit normalize traffic

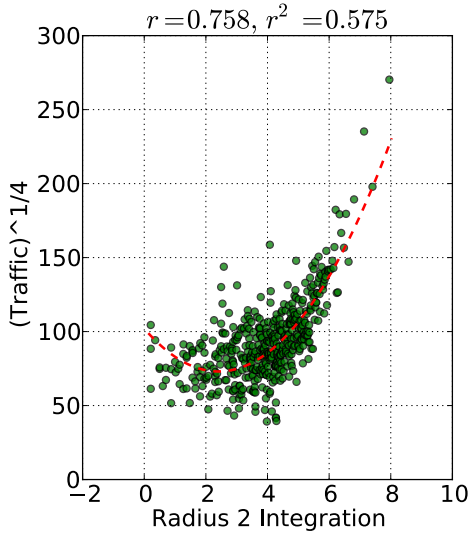


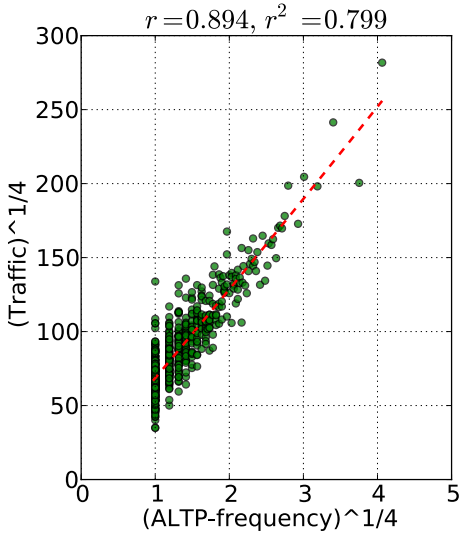
# IXP 2011


$r = 0.854$, $r^2 = 0.729$
(Traffic)^1/4
(Connectivity)^1/4

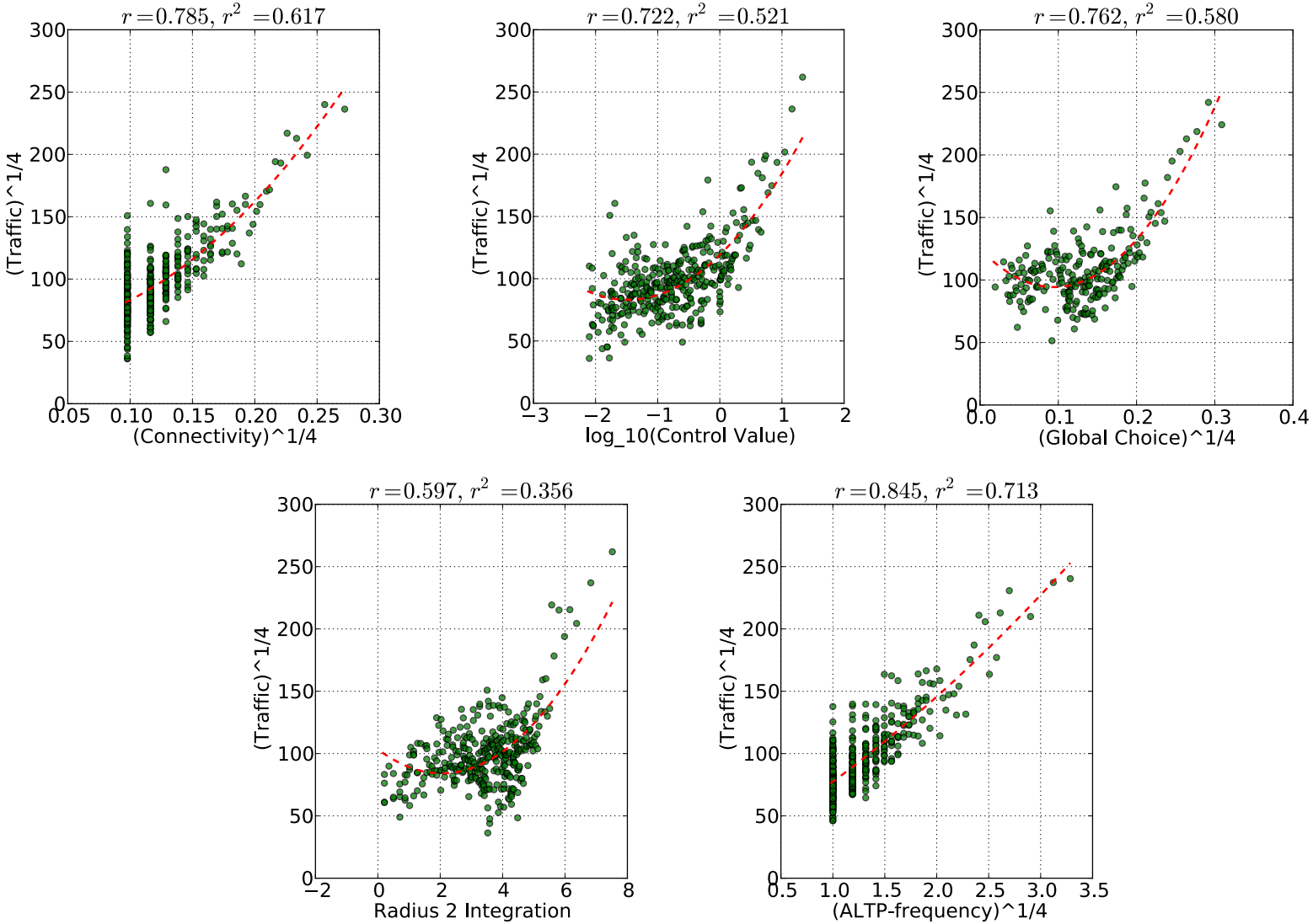

$r = 0.828$, $r^2 = 0.685$
(Traffic)^1/4
log_10(Control Value)


$r = 0.813$, $r^2 = 0.661$
(Traffic)^1/4
(Global Choice)^1/4

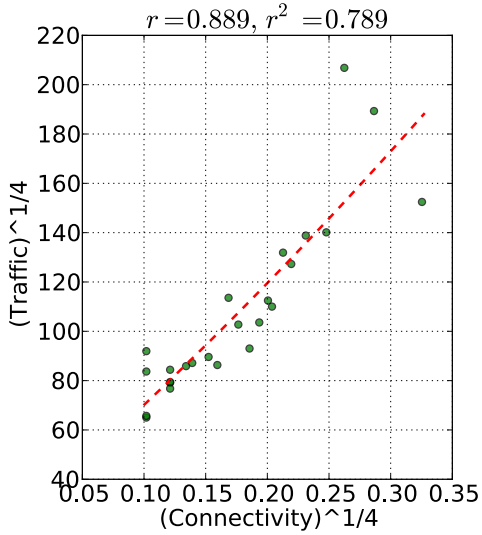

$r = 0.758$, $r^2 = 0.575$
(Traffic)^1/4
Radius 2 Integration

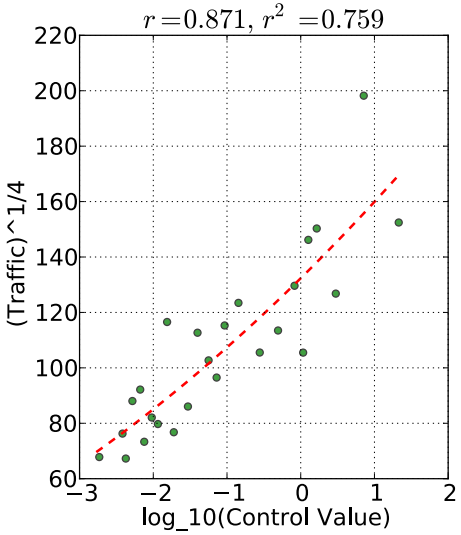

$r = 0.894$, $r^2 = 0.799$
(Traffic)^1/4
(ALTP-frequency)^1/4

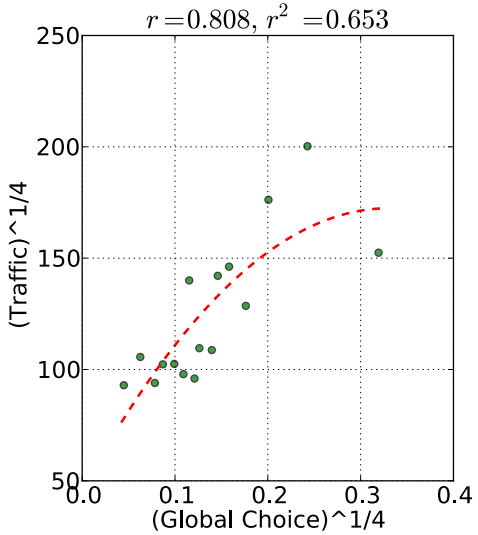
# IXP 2013

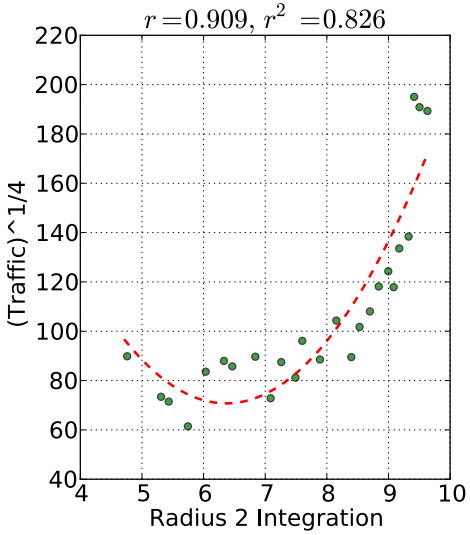

$r = 0.785$, $r^2 = 0.617$
(Traffic)^1/4
(Connectivity)^1/4
$r = 0.722$, $r^2 = 0.521$
log_10(Control Value)
$r = 0.762$, $r^2 = 0.580$
(Global Choice)^1/4
$r = 0.597$, $r^2 = 0.356$
Radius 2 Integration
$r = 0.845$, $r^2 = 0.713$
(ALTP-frequency)^1/4

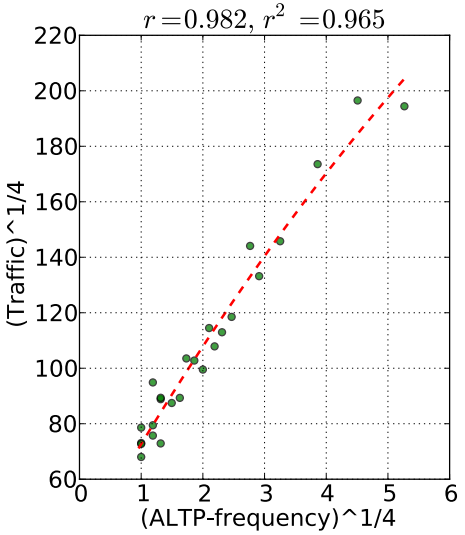
# ISP 2011


$r = 0.889$, $r^2 = 0.789$
(Traffic)^1/4
(Connectivity)^1/4

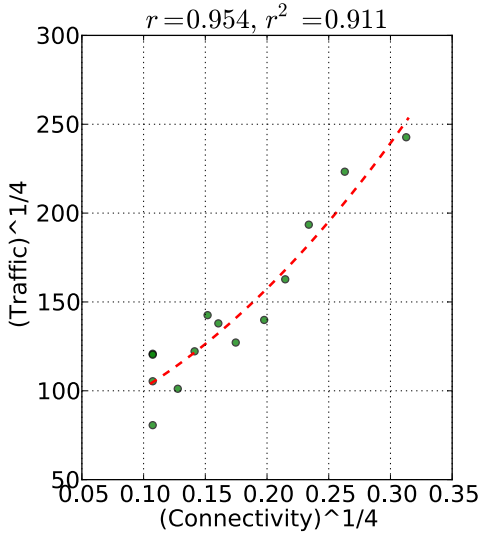

$r = 0.871$, $r^2 = 0.759$
(Traffic)^1/4
log_10(Control Value)

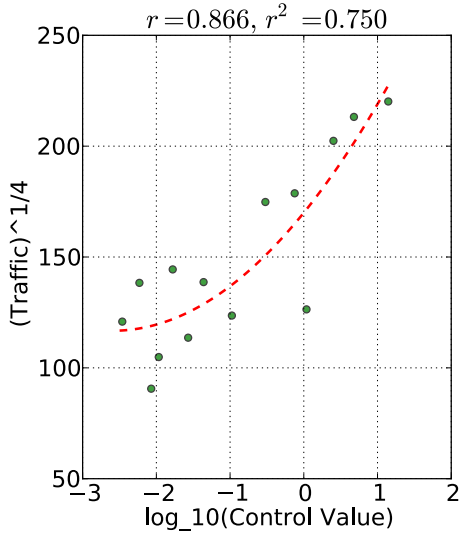

$r = 0.808$, $r^2 = 0.653$
(Traffic)^1/4
(Global Choice)^1/4

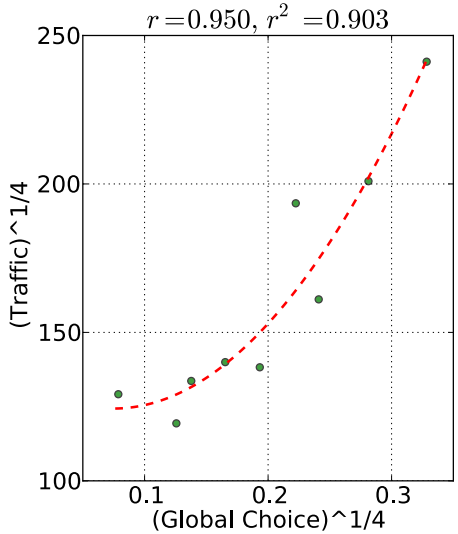

$r = 0.909$, $r^2 = 0.826$
(Traffic)^1/4
Radius 2 Integration

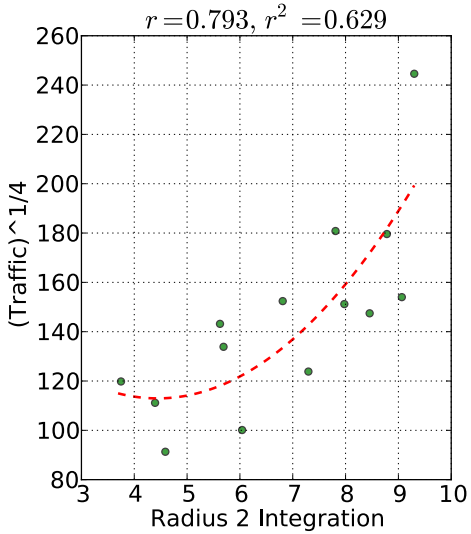

$r = 0.982$, $r^2 = 0.965$
(Traffic)^1/4
(ALTP-frequency)^1/4

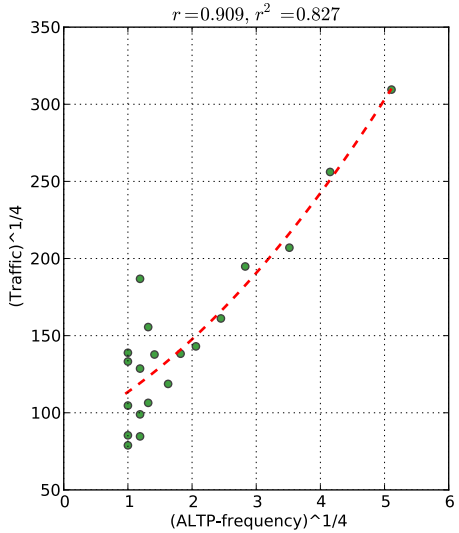
# ISP 2013


$r = 0.954$, $r^2 = 0.911$
(Traffic)^1/4
(Connectivity)^1/4

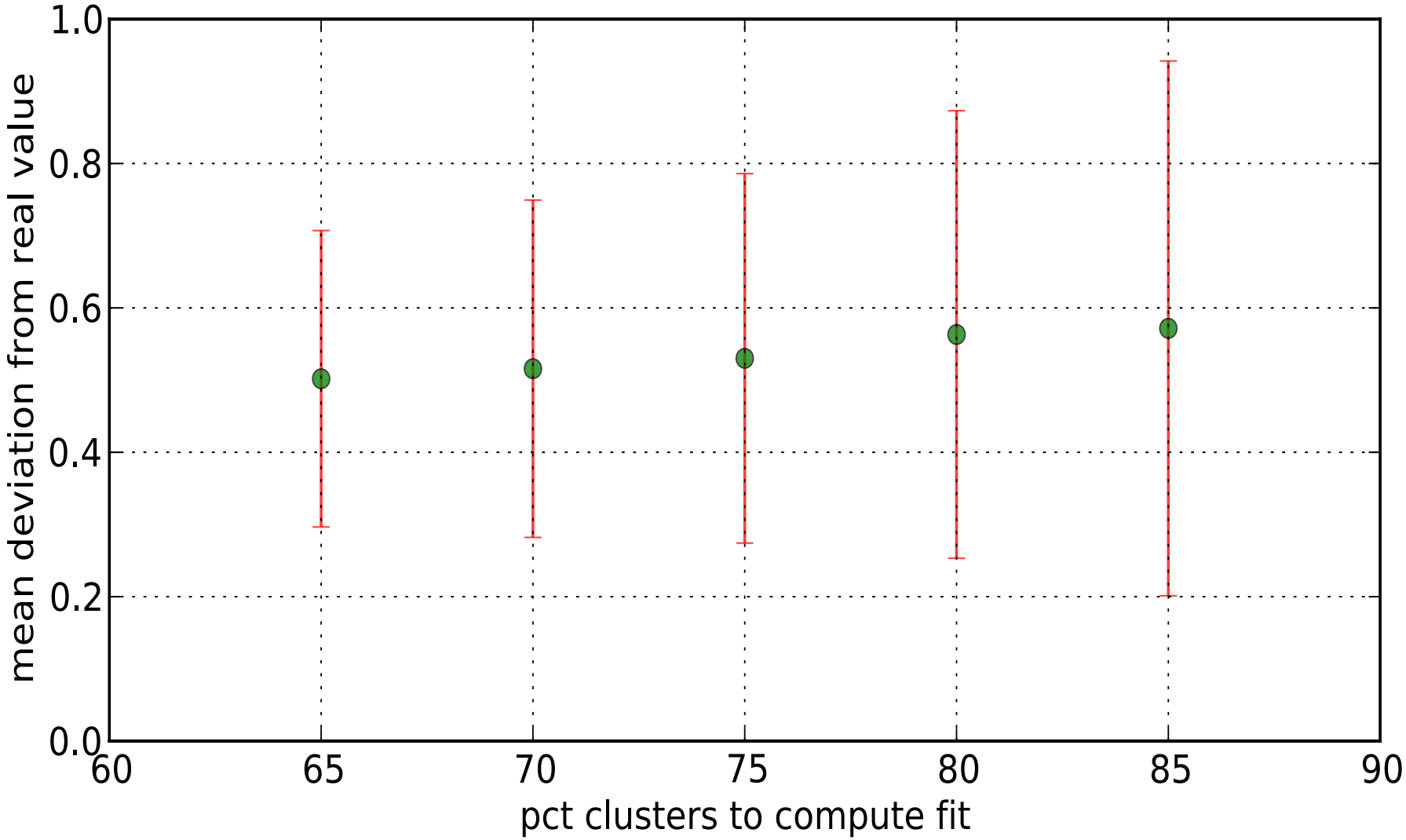

$r = 0.866$, $r^2 = 0.750$
(Traffic)^1/4
log_10(Control Value)


$r = 0.950$, $r^2 = 0.903$
(Traffic)^1/4
(Global Choice)^1/4

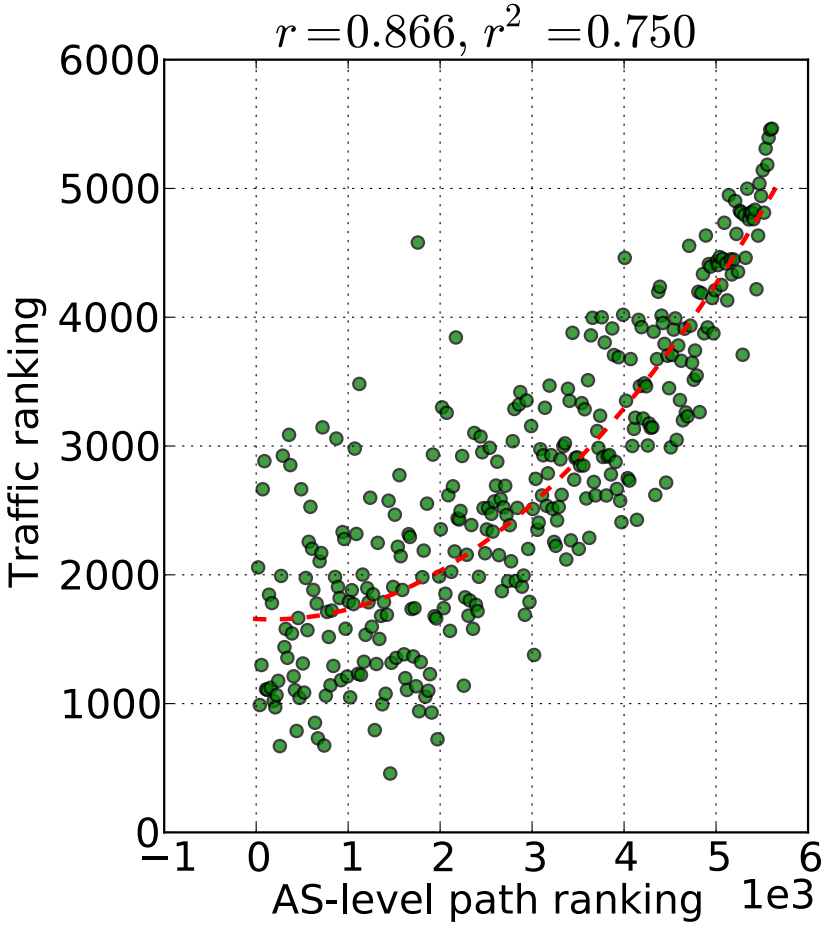

$r = 0.793$, $r^2 = 0.629$
(Traffic)^1/4
Radius 2 Integration

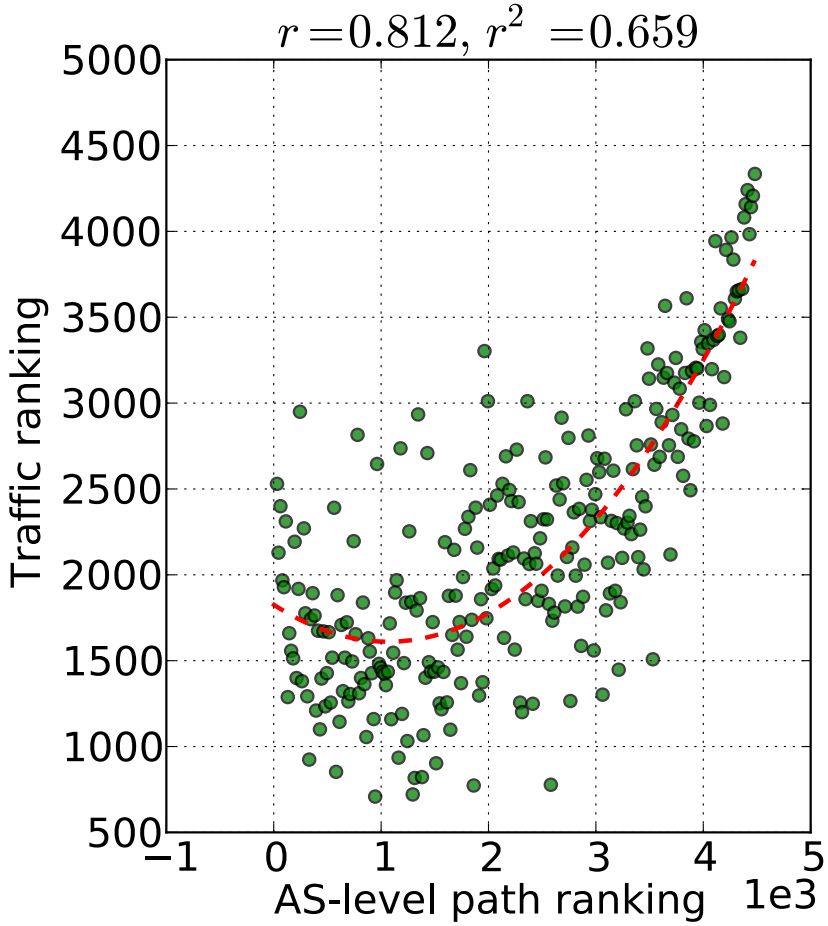

$r = 0.909$, $r^2 = 0.827$
(Traffic)^1/4
(ALTP-frequency)^1/4

# Traffic Prediction Error

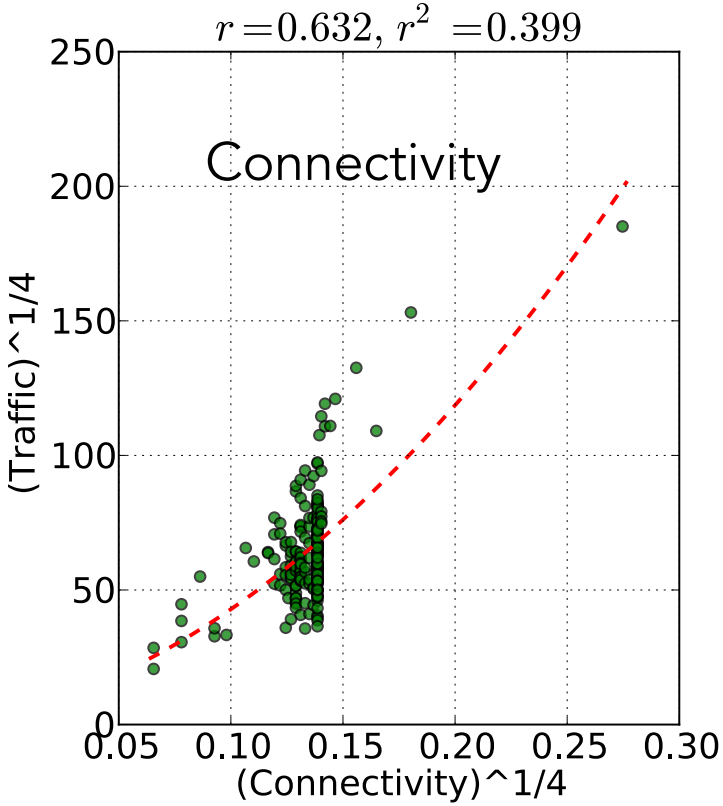


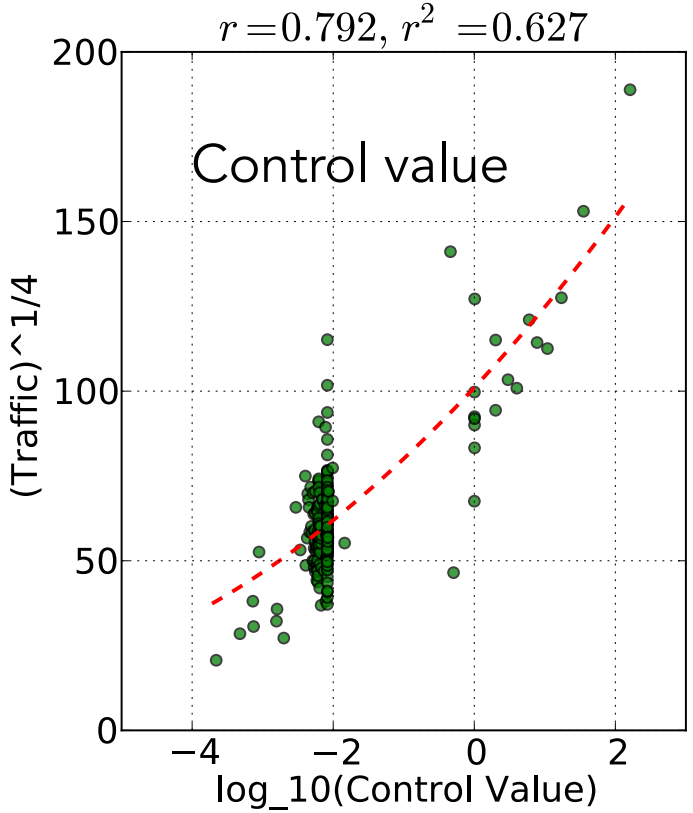
# Path Ranking IXP2011/2013



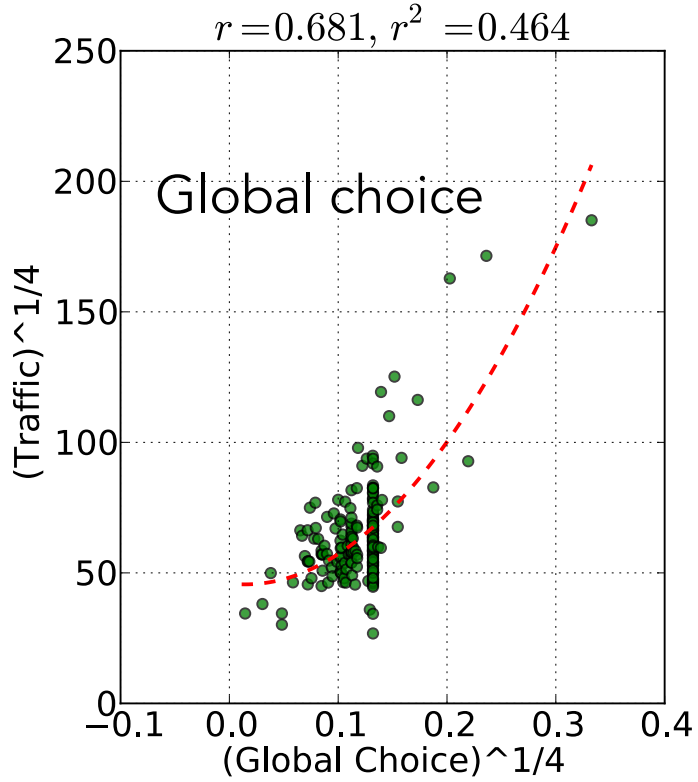



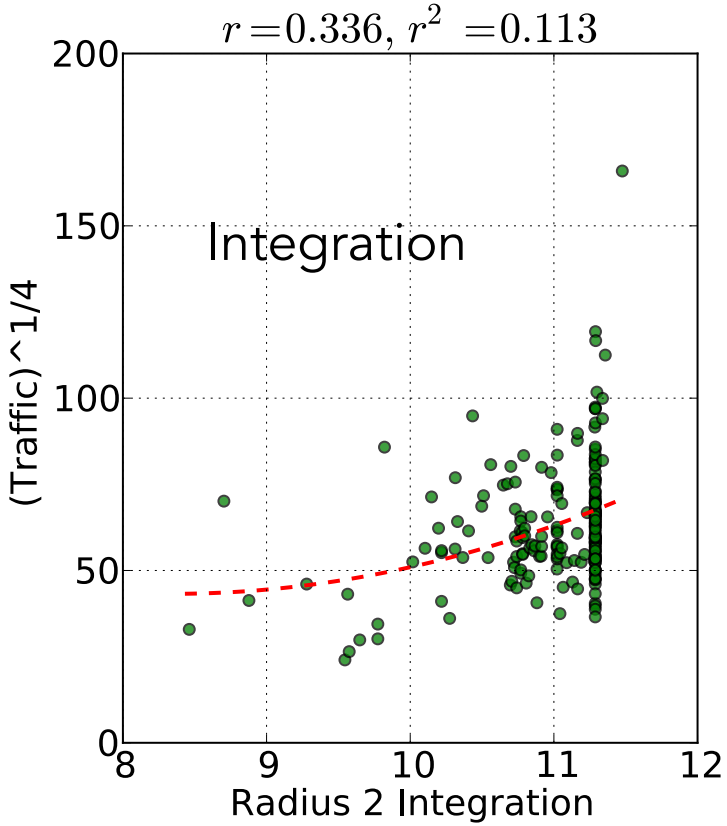
# BGP – ISP 2011


$r = 0.632$, $r^2 = 0.399$
Connectivity
(Traffic)^1/4
(Connectivity)^1/4
0
50
100
150
200
250
0.05
0.10
0.15
0.20
0.25
0.30

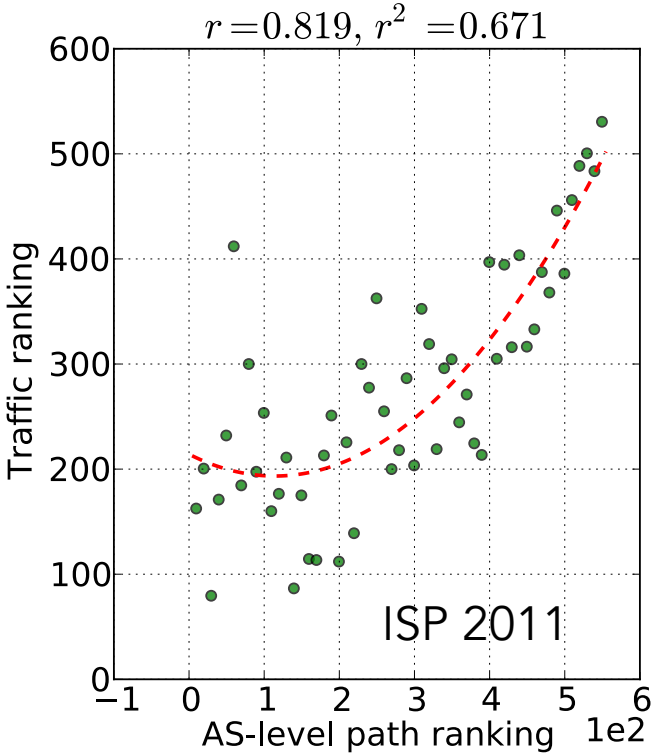

$r = 0.792$, $r^2 = 0.627$
Control value
(Traffic)^1/4
log_10(Control Value)
0
50
100
150
200
−4
−2
0
2

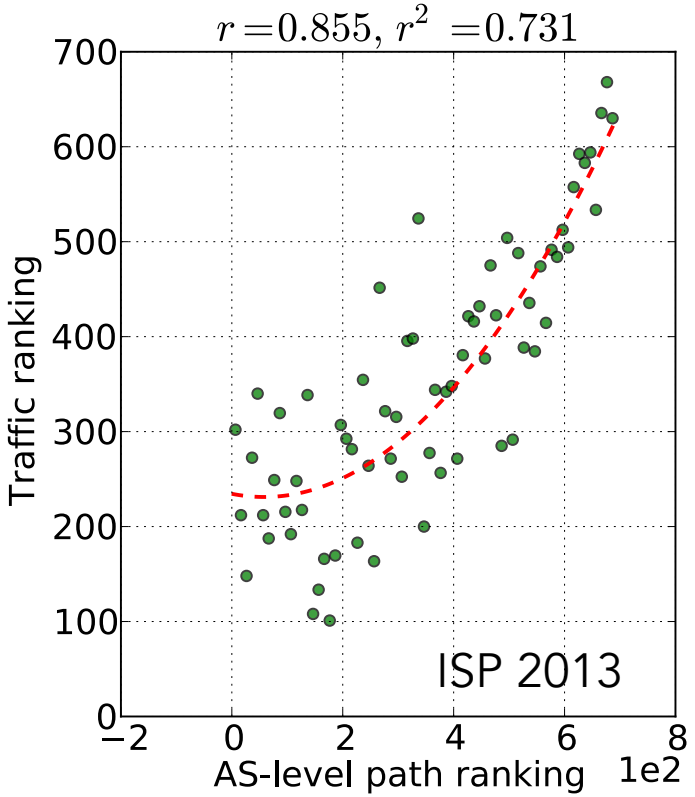
# BGP – ISP 2011


$r = 0.681$, $r^2 = 0.464$
Global choice
(Traffic)^1/4
(Global Choice)^1/4
250
200
150
100
50
0
−0.1
0.0
0.1
0.2
0.3
0.4

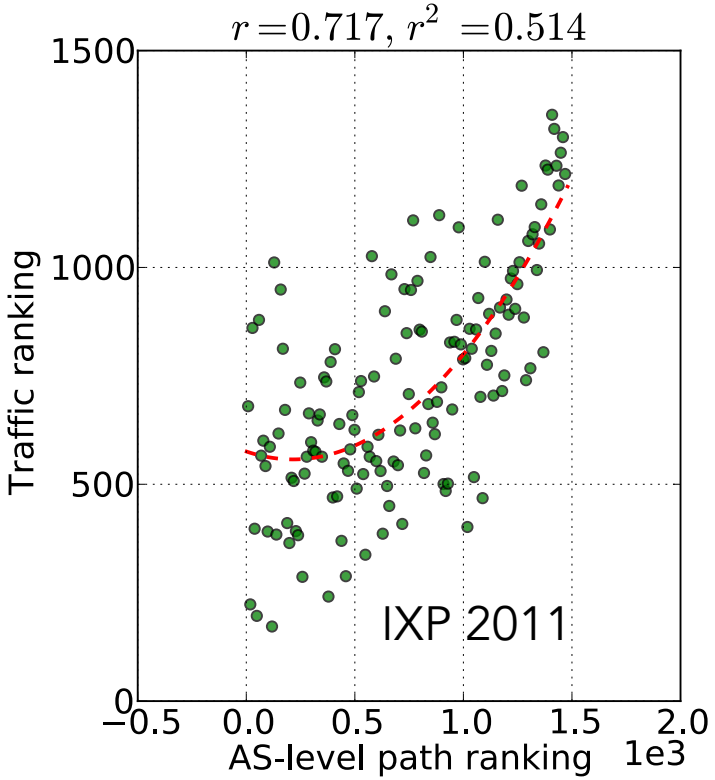

$r = 0.336$, $r^2 = 0.113$
Integration
(Traffic)^1/4
Radius 2 Integration
200
150
100
50
0
8
9
10
11
12

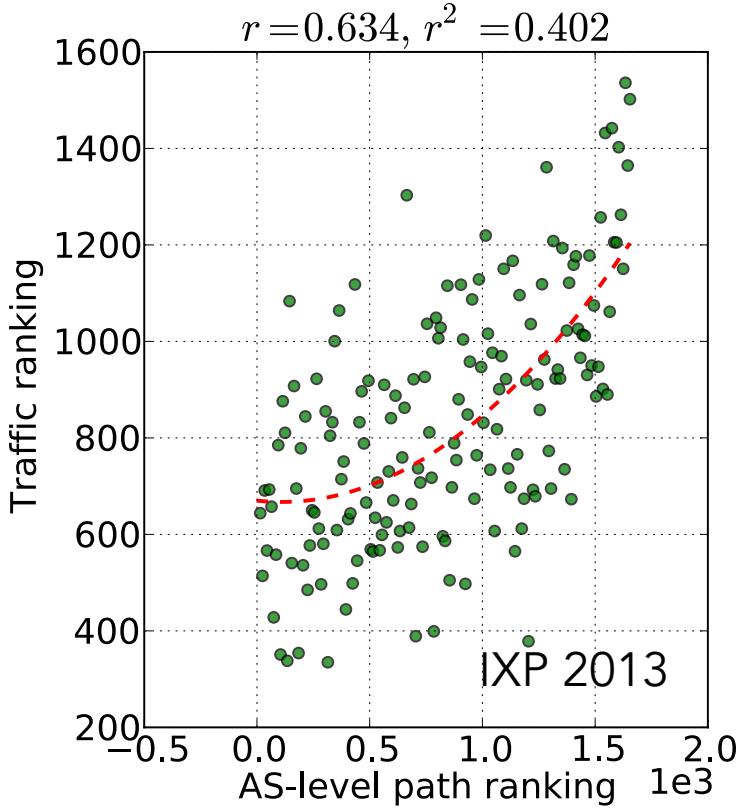
# CAIDA: ALTP-frequency vs. Traffic Volume



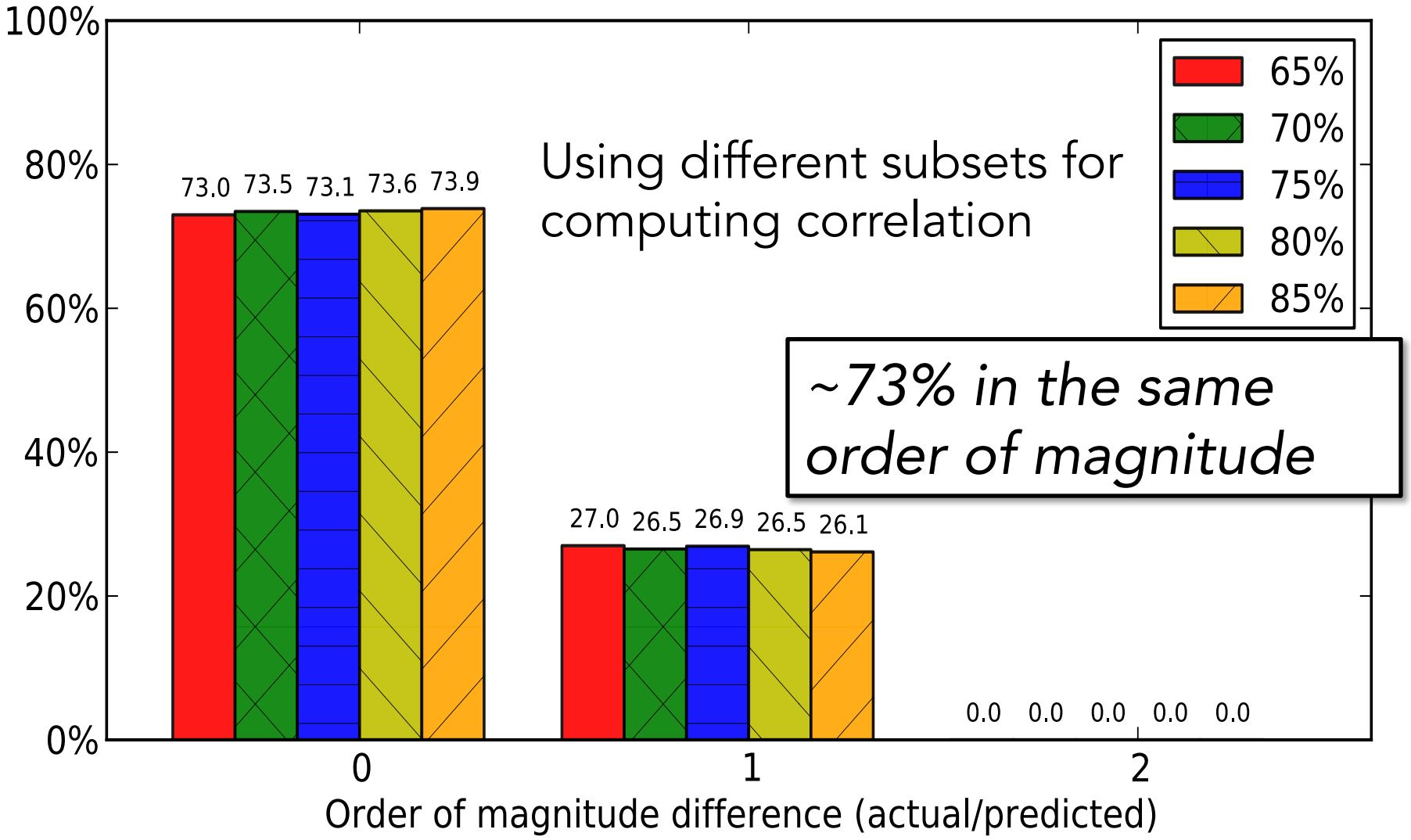



# CAIDA: ALTP-frequency vs. Traffic Volume



# Traffic Prediction with ALTP-frequency



# Path Ranking for ISP 2011/2013