Inferring AS relationships: Dead-End or Lively Beginning

Xenofontas Dimitropoulos (Fontas)
Dmitri Krioukov
Bradley Huffaker
kc claffy
George Riley

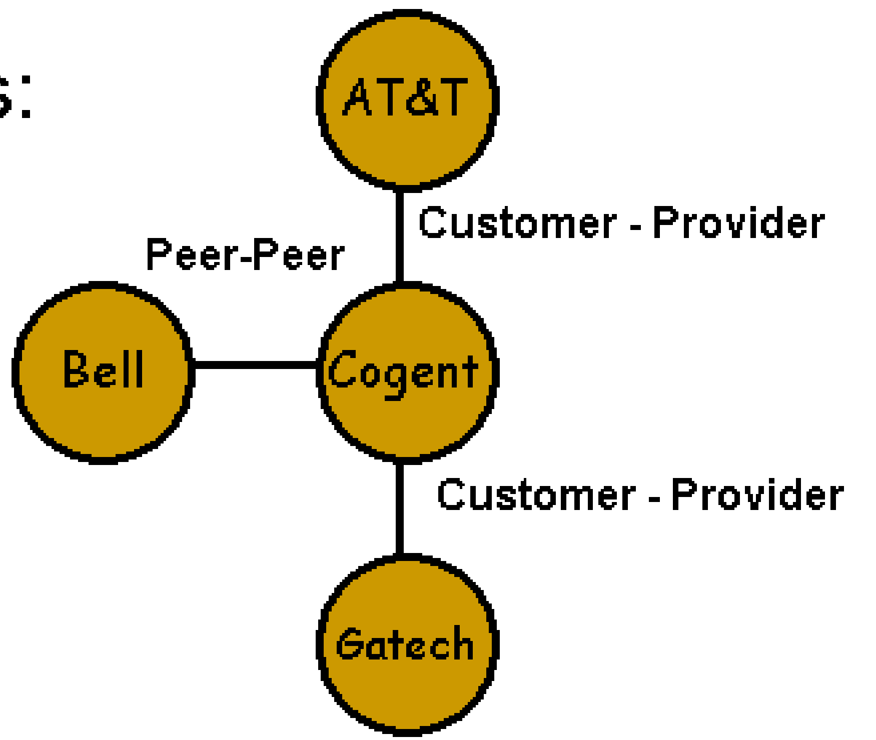
CAIDA/Geogia Tech

Outline

- Background on AS relationships.
- Previous work.
- New algorithm to infer AS relationships.
- Results.

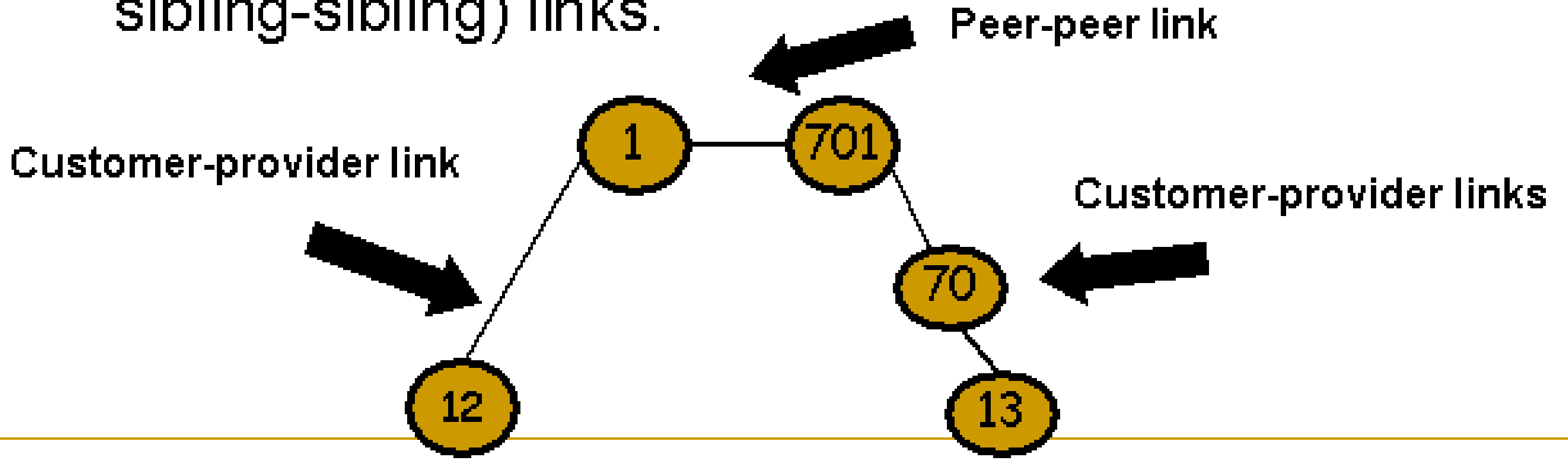
AS relationships

- AS relationships types:
  - Customer - Provider
  - Peer – Peer
  - Sibling -Sibling



Structure of AS paths (Valley Free Model)

- AS paths have the following hierarchical structure:
  - zero or more customer-provider (or sibling-sibling) links.
  - followed by zero or one peer-peer link.
  - followed by zero or more customer-provider (or sibling-sibling) links.



Previous Work

- ***GAO: "On Inferring Autonomous Systems relationships in the Internet" L. Gao ToN 2001.***
  - *Introduce AS relationships classification, Valley Free Model, and heuristic solution.*
- ***SARK: "Characterizing the Internet hierarchy from multiple vantage points" L. Subramanian et. al INFOCOM 2002.***
  - Introduce the Type-of-Relationships (ToR) problem: *Given an undirected graph G derived from a set of BGP paths P, assign a direction (reflecting a customer-provider or peer-peer relationship) to every edge in G such that the total number of valid paths in P is maximized.*
- **DPP: "*Computing the Types of the Relationships between Autonomous Systems*", *G. Di Battista et al. INFOCOM 2003.***
- **EHS: "*Classifying customer-provider relationships in the Internet*", *T. Erlebach et al. CCN 2002***
  - Find that no peer-peer relationships can be inferred in ToR formulation.
  - Prove that ToR is NP- and APX-complete.
  - Introduce a rigorous approximation to ToR to compute customer-provider relationships only.

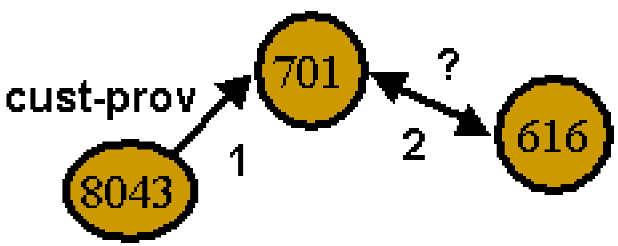
Outline

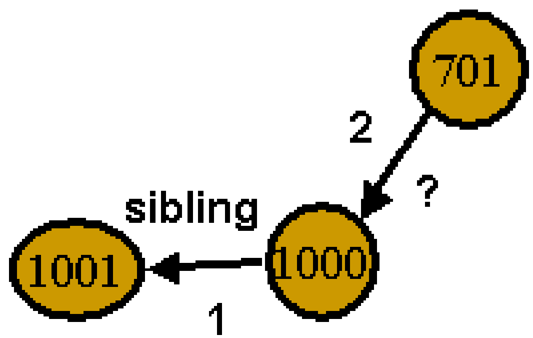
- Background on AS relationships.
- Previous Work.
- New algorithm to infer AS relationships.
- Results.

ToR limitations.

- **Case 1:** some edges can be directed either way without causing invalid paths
- **Fix:** introduce additional incentive to direct edges along the node degree gradient

- **Case 2:** trying to direct sibling links proliferates errors
- **Fix:** discover sibling links using the WHOIS database





New Formulation

- Introduce a new formulation of the AS relationships inference problem as a multi-objective optimization:
  - Objective 1: Maximize number of valid paths (like EHS/DPP/SARK).
  - Objective 2: Direct links along node degree gradient (like GAO).
- Intuition: paths are now colored by the their nodes’ degrees which allows to exploit the structure of a path in detecting anomalous paths.

Reduce to MAX2SAT

- Objective 1 (O1): Maximize number of valid paths
  - Introduces clauses $w_{kl}\alpha(x_k \vee x_l)$ in the MAX2SAT instance.
- Objective 2 (O2): Direct along the node degree gradient
  - Introduces clauses $w_{kk}(1-\alpha)(x_k \vee x_k)$ in the MAX2SAT instance.
- Tune parameter α to adjust relative weight on the two objectives.

Solution of the problem

- Solve MAX2SAT using the SDP approximation (94 approximation ratio).
- Use BGP paths from RouteViews and 18 other route servers (1,025,775 paths) for input set of paths.
- Find sibling links using WHOIS database and remove them from the graph.

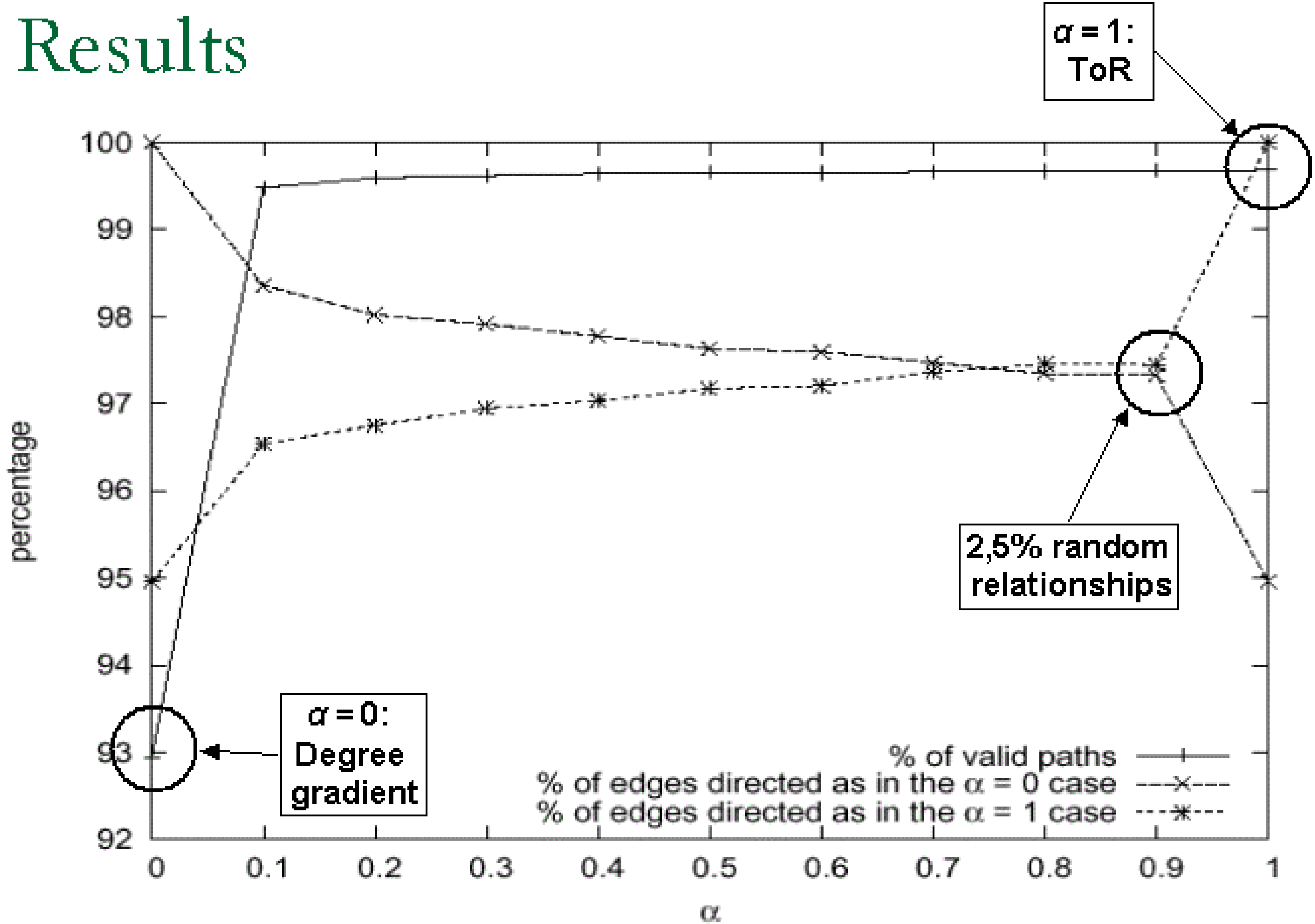
AS hierarchy

- Rank ASs by their *reachability:* size of customer cone of an AS.

Table 2: Ranking of ASs induced by our inference algorithm. The ranks of the top five ASs for $\alpha = 0$ and $\alpha = 1$ are shown for different values of α. The AS numbers are matched to an AS name using the WHOIS databases.

AS #	AS name	AS outdegree	$\alpha = 0.0$	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1.0$
701	UUNET	2373	0-1	0-173	0-217	1-242	1-252	17-476
1239	Sprint	1787	1-1	0-173	0-217	1-242	1-252	17-476
7018	AT&T	1723	2-1	0-173	0-217	1-242	1-252	17-476
3356	Level 3	1085	3-1	0-173	0-217	1-242	1-252	17-476
209	Qwest	1072	4-1	0-173	0-217	1-242	1-252	17-476
11551	Pressroom Services	2	1742-941	1419-398	1435-391	1449-390	1457-386	0-4
6721	Nextra Czech Net	3	1742-941	833-88	853-90	874-90	884-89	0-4
3643	Sprint Australia	17	194-1	222-1	233-1	261-1	268-1	0-4
1243	Army Info. Systems	2	2683-62853	2753-14655	1435-391	1449-390	1457-386	0-4
6712	France Transpac	2	2683-62853	2753-14655	2774-14634	298-2	1-252	4-13

Results


α = 1:
ToR
100
99
98
97
96
95
94
93
92
percentage
2,5% random
relationships
α = 0:
Degree
gradient
% of valid paths
% of edges directed as in the α = 0 case
% of edges directed as in the α = 1 case
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
α

Conclusions and Future work

- What's done?
  - Find that ToR solutions do not yield correct AS relationships.
  - Identify ToR problem and introduce a natural generalization of the previous AS relationships inference algorithms using a standard multiobjective optimization method.
- What's in progress?
  - How to determine optimal α.
  - Infer p2p links.
  - Validation.

Reachability based rank of AS: http://as-rank.caida.org/

Inferring AS relationships: Dead-End or Lively Beginning

Xenofontas Dimitropoulos (Fontas)
Dmitri Krioukov
Bradley Huffaker
kc claffy
George Riley

CAIDA/Geogia Tech

Questions

?

What is Semi-Definite Programming?

- SDP is a variation of ordinary linear programming (LP), where the the nonnegativity constraint is replaced by a semidefinite constraint on matrix variables.

$$\min C \bullet X \quad \text{subject to} \quad A_k \bullet X = b_k \ (k = 1, \ldots, m); \quad X \succeq 0,$$

Weight function

- Weights:

$$w_{kl}(\alpha) = \begin{cases} c_2\alpha & \text{if} \quad \{kl\} \in P, \\ c_1(1-\alpha) f(d_k^-, d_k^+) & \text{if} \quad k = l \leqslant m_1, \\ 0 & \text{otherwise.} \end{cases}$$

- *f*: a function of degrees of adjacent nodes.

$$f(d_i^-, d_i^+) = \frac{d_i^+ - d_i^-}{d_i^+ + d_i^-} \log(d_i^+ + d_i^-).$$

- α: tuning parameter.
- c1,c1: normalization coefficients.