Massimo Rimondini

Dipartimento di Informatica e Automazione
Università degli Studi Roma Tre

For more information, please visit http://www.dia.uniroma3.it/~compunet

Statistics and comparisons about two solutions for computing the types of relationships between Autonomous Systems

--- WORK IN PROGRESS ---

October 2002

Document structure

This document is organized as follows:

- data, graphs and other similar information are placed on odd-numbered pages;
- descriptions, definitions, comments and other observations regarding the above mentioned stuff are reported on the even-numbered pages that face those containing the data they refer to.

Data sets

The following reports work on data that has been downloaded from the web site of Sharad Agarwal, Lakshminarayanan Subramanian, Jennifer Rexford, and Randy H. Katz (http://www.cs.berkeley.edu/~sagarwal/research/BGP-hierarchy). In particular, the SAT and VANTAGE graphs have been calculated using the data extracted on **April 18th, 2001**, and, in order to make correct comparisons, also the other statistics have been calculated using the same data. The information that have been used include BGP tables and AS paths lists from 10 different looking glasses (1, 1740, 3549, 3582, 3967, 4197, 5388, 7018, 8220, 8709).

AS graph

An *AS graph* is a graph representing logical adjacency relationships between couples of ASes. The term “logical” refers to the fact that two ASes are considered mutually adjacent if any of them exchanges routing information (i.e. a set of IP prefixes) with the other. Adjacency is not usually influenced by physical (geographical) contiguity properties.

The AS graph is here built on the basis of a list of AS paths that is the catenation of all the AS paths lists from the 10 looking glasses.

SAT graph

A *SAT graph* is a graph containing the same nodes and edges of the AS graph, but in which the latter are all oriented in compliance with a set of customer-provider relationships between ASes. Such relationships are inferred using algorithms and tools developed by G. Di Battista, M. Pizzonia and M. Patrignani (from now onwards, referenced as “bgpSat”), which work on a set of paths extracted from BGP tables and give as result a possible orientation of the AS graph. The orientation is always “from customer to provider”, meaning that an edge from AS x to AS y represents a situation in which x is a customer of y and y is a provider for x. The inferred graph contains no peering (i.e. unoriented) edges, but some edges (that existed in the AS graph) may be missing: this means that the algorithm could not infer an orientation for the corresponding couple of ASes.

The representation of the SAT graph chosen in this report is the following: SAT graph is a (purely) oriented graph, using an orientation that complies with the above described convention (from customer to provider), with no representation of missing edges (with respect to the AS graph), and that may miss some nodes (of the AS graph); that is, the SAT graph is here re-built (for statistical purposes) by exclusively parsing the output of the software that generated it (that contains only successfully oriented edges and the ASes that appear at their ends). Reports concerning missing edges are generated by subsequently processing an AS path list to check for missing adjacencies.

The original (inferred) SAT graph has been built by working on a complete list of AS paths from all the looking glasses; the latter has been obtained by simply catenating the AS paths lists belonging to each looking glass. The building process also included a phase of invalid paths reinsertion, to try to get a maximal set.

Number of unoriented path edges

This is the number of AS adjacencies (drawn from the complete AS paths list – i.e. the catenation of the AS paths lists belonging to all the looking glasses –) that don’t have a corresponding edge (orientation) on the SAT graph. This number may include repetitions: if a couple of ASes whose edge could not be given an orientation is repeated more than one time, the value is increased for each instance.

Number of candidate peerings

This is the number of possible peerings produced by bgpSat. It just counts the number of lines starting with `candidate peering`.

Path covering distribution

This value is taken directly from bgpSat’s output, and shows how many paths traverse each of the graph’s edges, regardless of the orientations (both of the paths and of the edges). More precisely, each path causes all the edges it traverses to increase by 1 their path covering.

Advertised space distribution

This is computed by parsing a catenation of the BGP tables of all the looking glasses (after making sure that each of them includes the looking glass that receives the route as first AS in all the AS paths). Each prefix that is advertised over a certain edge (i.e. from an AS to another) causes that edge’s advertised space to increase by a quantity that corresponds to the number of IP addresses that have been communicated. Like in the case of path covering distribution, no distinction is made as far as the advertisement’s direction is concerned.

Since the data is taken from a catenation of BGP tables, in order to avoid eventual repetitions in taking account of each prefix, the shown value is calculated as explained above but then (before computing the distribution) divided by 10 (the number of looking glasses), in order to attempt to report a mean value over all the looking glasses.

The smaller italic values show the same values when the division by 10 is not performed.

Strongly connected components (scc’s)

The reported sizes only consider the number of nodes.

AS graph

- NUMBER OF NODES: 10916
- NUMBER OF EDGES: 23761

- CONNECTED COMPONENTS:
  - total number: 1
  - cc's to edges ratio: 0,0000042
  - size distribution:
    - 1 cc of size 10916

SAT graph

- NUMBER OF NODES: 10897
- NUMBER OF EDGES: 23690
  - number of unoriented path edges: 643
  - number of candidate peerings: 4148
- PEERING EDGES:
  - number: 0
  - percentage of total: 0 %
- OTHER EDGES:
  - number: 23690
  - percentage of total: 100 %

- PATH COVERING DISTRIBUTION:
  - minimum: 1
  - average: 80
  - maximum: 19918
- ADVERTISED SPACE DISTRIBUTION:
  - minimum: 0 *8 (~ /29)*
  - average: 1128501 (~ /12) *11285020 (~ /9)*
  - maximum: 382845158 (~ /4) *3828451584 (~ /1)*

- STRONGLY CONNECTED COMPONENTS:
  - total number: 10312
  - scc's to edges ratio: 0,4352892
  - size distribution (nodes):
    - 10311 scc's of size 1
    - 1 scc of size 586

VANTAGE graph

A *VANTAGE graph* is a graph which provides a possible orientation of the edges of the AS graph by using the algorithms and tools developed by S. Agarwal, L. Subramanian, J. Rexford and R. H. Katz. This graph is made up of both oriented and unoriented edges. The first start from a node that is considered a customer and end on its supposed provider (the orientation is "from customer to provider", like for the SAT graph); the latter indicate the presence of a possible peering relationship between the ASes they connect. Again, like in the SAT graph, some edges may be missing (with respect to the AS graph), thus showing a failed attempt to give them an orientation.

The VANTAGE graph has been re-built (for statistical purposes), using a representation that is wholly similar to that of the SAT graph: oriented edges go from the customer to the provider, missing edges are not represented at all and peering edges are considered as unoriented edges. This fully corresponds to the definition of VANTAGE graph.

There is only one important (even if not so relevant on the computed values) difference in comparison with the SAT graph: the latter just contains all the nodes that appear as the ends of successfully oriented edges; in this case, the VANTAGE graph also considers all the nodes that appear at the ends of missing edges. This is due to the availability of a list of pairs of ASes whose relationship remains unknown. For example, if the relationship between AS *x* and AS *y* could not be inferred, AS *y* appears in the graph even if it is an isolated node (which has no inferred relationships with other nodes), because its adjacency with AS *x* is known by means of the mentioned list.

This could be done also in the case of the SAT graph (by analysing the paths list); anyway, the incidence of this choice is really limited: building a "pure" VANTAGE graph (which, of course, also considers some missing orientations) and then adding the isolated nodes requires the addition of just 8 nodes, which are really few in comparison with the total number of nodes.

Path covering distribution

This value is computed exactly as in the case of the SAT graph: for each path, the path covering distribution of the edges it traverses is increased by 1, regardless of orientations. Anyway, the value has been subsequently divided by 10 in order to avoid considering eventual repetitions and to try to get a mean value over all the looking glasses. Again, the AS paths file is the catenation of all the AS paths lists provided by Agarwal et al.

The smaller italic values show the same values if the division by 10 is not performed.

The value 0 as minimum path covering distribution is a clear hint (and this conjecture has been later verified) that some ASes do not appear in the paths lists provided by Agarwal. This means that they were filtered out for some reason (most probably, path incompleteness) while extracting them from the BGP tables. Nevertheless, the VANTAGE graph also contains these missing ASes (most probably, it was built directly on the basis of the BGP tables themselves).

Advertised space distribution

The computation of this value is quite the same as SAT graph's, including the division by 10, the source data set and the meaning of the smaller italic values.

Strongly connected components

The definition of "strongly connected components" remains exactly the same, with a little change: the paths leading from a node to another can traverse an unoriented (peering) edge in any direction; in other words, peering edges are considered as if they were doubly oriented.

Unoriented edges

This is the number of adjacencies on which VANTAGE failed to assign an orientation. It corresponds to the number of edges contained in the file of unknown relationships, even if it has been recomputed for better confidence.

VANTAGE graph

- **NUMBER OF NODES:** 10923
- **NUMBER OF EDGES:** 23757
  - number of unoriented path edges: 43242
  - number of candidate peerings: 0
- **PEERING EDGES:**
  - number: 1136
  - percentage of total: 5 %
- **OTHER EDGES:**
  - number: 22621
  - percentage of total: 95 %

- **PATH COVERING DISTRIBUTION:**
  - minimum: 0 *0*
  - average: 50 *506*
  - maximum: 22327 *223277*
- **ADVERTISED SPACE DISTRIBUTION:**
  - minimum: 0 *8 (~ /29)*
  - average: 1121286 (~ /12) *11212880 (~ /9)*
  - maximum: 382845158 (~ /4) *3828451584 (~ /1)*

- **STRONGLY CONNECTED COMPONENTS:**
  - total number: 10302
  - scc's to edges ratio: 0,4336406
  - size distribution:
    - 10096 scc's of size 1
    - 176 scc's of size 2
    - 21 scc's of size 3
    - 4 scc's of size 4
    - 1 scc of size 6
    - 1 scc of size 7
    - 1 scc of size 19
    - 1 scc of size 86
    - 1 scc of size 278

- **UNORIENTED EDGES:** 178

UNORIENTED graph for SAT graph

The *UNORIENTED graph* is a graph that is built in the following way: for each couple of nodes whose orientation could not be inferred by bgpSat, an unoriented edge is added on the UNORIENTED graph, and the ASes at its ends are added too, if necessary. In other words, it is a representation of all the missing edges on the SAT graph.

Path covering distribution

As always, this value is computed by parsing the complete list of paths (from all the looking glasses) and adding 1 to an edge each time a path traverses it. Like in the case of the SAT graph, no division by 10 has been performed.

Advertised space distribution

The computing process for this number is the usual one. Like in the case of the SAT graph, the advertised space of each edge has been divided by 10 before computing the distribution, while the values in small italic font do not include this operation.

UNORIENTED graph (for SAT graph)

- NUMBER OF NODES: 86
- NUMBER OF EDGES: 71

- PATH COVERING DISTRIBUTION:
  - minimum: 0
  - average: 8
  - maximum: 103
- ADVERTISED SPACE DISTRIBUTION:
  - minimum: 25 (~ /28)
    256 (~ /24)
  - average: 3812 (~ /21)
    38121 (~ /17)
  - maximum: 66892 (~ /16)
    668928 (~ /13)

- CONNECTED COMPONENTS:
  - total number: 19
  - cc's to edges ratio: 0,2676056
  - size distribution:
    - 15 cc's of size 2
    - 1 cc of size 3
    - 1 cc of size 4
    - 1 cc of size 23
    - 1 cc of size 26

Differences between SAT and VANTAGE

These analyses have been performed in order to show how many nodes/edges are missing in either graph; an even more important value is the one that counts how many oppositely oriented edges are there: they are quite few if compared to the total number of common edges, and this is a good hint of the quality of the orientation.

Differences between SAT and VANTAGE

- NODES ONLY IN SAT GRAPH: 0
- NODES ONLY IN VANTAGE GRAPH: 26
- NODES IN BOTH GRAPHS: 10897
- EDGES ONLY IN SAT GRAPH:
  - total: 178
  - peering edges:
    - number: 0
    - percentage of total: 0 %
  - other edges:
    - number: 178
    - percentage of total: 100 %
- EDGES ONLY IN VANTAGE GRAPH:
  - total: 245
  - peering edges:
    - number: 25
    - percentage of total: 10 %
  - other edges:
    - number: 220
    - percentage of total: 90 %
- EDGES IN BOTH GRAPHS:
  - total: 23512
  - consistently oriented:
    - number: 21253
    - percentage of total: 90 %
  - differently oriented:
    - number: 2259
    - percentage of total: 10 %
  - oppositely oriented:
    - number: 1148
    - percentage of total: 5 %
    - percentage of diff. oriented: 51 %
  - peers only in SAT graph:
    - number: 0
    - percentage of total: 0 %
    - percentage of diff. oriented: 0 %
  - peers only in VANTAGE graph:
    - number: 1111
    - percentage of total: 5 %
    - percentage of diff. oriented: 49 %
  - peers in both graphs:
    - number: 0
    - percentage of total: 0 %
    - percentage of cons. oriented: 0 %

DIFFERENTIAL graph

A *DIFFERENTIAL graph* is a graph which highlights different orientations between SAT and VANTAGE. In particular, for each of the edges that are common to the SAT and the VANTAGE graph (i.e. edges linking the same couples of ASes), their orientation in SAT and in VANTAGE is compared, and, if found to be different, an unoriented edge is inserted in the DIFFERENTIAL graph. The nodes at its ends are inserted too, if necessary.

This shows that the DIFFERENTIAL graph takes into account all kinds of differences in orientation: both oppositely oriented edges and edges that are peerings on only one graph.

Path covering distribution

These data follow the same rules as those for the other graphs: when derived from the VANTAGE graph, the path covering of each edge is preliminarily (before computing the distribution) divided by 10; when derived from the SAT graph, this operation is not performed. As always, small italic values represent the situation in which the division is not performed.

The choice to report data from both the two graphs is due to small gaps that have been pointed out between the information contained in each of them. These might be due to different methods used by bgpSat and by the tools that performed the reported analyses to compute the path covering and to approximation faults while computing the means of advertised spaces.

Advertised space distribution

The process is similar to that for path covering distribution, except that the division by 10 is performed both for data from SAT graph and for data from VANTAGE graph. Small italic values show the same quantities when the division is not performed.

Connected components

Because of the way in which the graph is built (edges are its “primary components”), the smallest connected component must be made up of at least 2 nodes.

DIFFERENTIAL graph’s edges distribution by Looking Glass

This report shows how many of the DIFFERENTIAL graph’s edges are visible from each looking glass.

DIFFERENTIAL graph

- NUMBER OF NODES: 1013
- NUMBER OF EDGES: 2259

- PATH COVERING DISTRIBUTION (data from SAT graph):
  - minimum: 1
  - average: 237
  - maximum: 19918
- PATH COVERING DISTR. (data from VANTAGE graph):
  - minimum: 0 *1*
  - average: 162 *1627*
  - maximum: 16615 *166153*
- ADVERTISED SPACE DISTR. (data from SAT graph):
  - minimum: 0 *8 (~ /29)*
  - average: 4027047 (~ /11) *40270400 (~ /7)*
  - maximum: 382845158 (~ /4) *3828451584 (~ /1)*
- ADVERTISED SPACE DISTR. (data from VANTAGE graph):
  - minimum: 0 *8 (~ /29)*
  - average: 4027048 (~ /11) *40270484 (~ /7)*
  - maximum: 382845158 (~ /4) *3828451584 (~ /1)*

- CONNECTED COMPONENTS:
  - total number: 169
  - cc's to edges ratio: 0,0748119
  - size distribution:
    - 143 cc's of size 2
    - 20 cc's of size 3
    - 4 cc's of size 4
    - 1 cc of size 5
    - 1 cc of size 646

DIFFERENTIAL graph’s edges distribution by Looking Glass

```
AS 1    sees 263  of the differently oriented edges
AS 1740 sees 293  of the differently oriented edges
AS 3549 sees 320  of the differently oriented edges
AS 3582 sees 1883 of the differently oriented edges
AS 3967 sees 580  of the differently oriented edges
AS 4197 sees 290  of the differently oriented edges
AS 5388 sees 299  of the differently oriented edges
AS 7018 sees 301  of the differently oriented edges
AS 8220 sees 408  of the differently oriented edges
AS 8709 sees 426  of the differently oriented edges
```

INCONSISTENCY graph

The *INCONSISTENCY graph* (or *graph of the incompatibilities*) is a subgraph of the differential graph, which only considers edges that are oppositely oriented in SAT and VANTAGE. In other words, every time an oppositely oriented edge is detected on the two graphs, an edge (and, if necessary, the nodes at its ends) is added to this graph.

Path covering distribution and Advertised space distribution

They both follow the criteria shown for the DIFFERENTIAL graph.

INCONSISTENCY graph's edges distribution by Looking Glass

This shows how many edges of the INCONSISTENCY graph can be seen from each looking glass.

INCONSISTENCY graph

- NUMBER OF NODES: 489
- NUMBER OF EDGES: 1148

- PATH COVERING DISTRIBUTION (data from SAT graph):
  - minimum: 1
  - average: 37
  - maximum: 3420
- PATH COVERING DISTR. (data from VANTAGE graph):
  - minimum: 0 *1*
  - average: 29 *296*
  - maximum: 3043 *30436*
- ADVERTISED SPACE DISTR. (data from SAT graph):
  - minimum: 0 *8 (~ /29)*
  - average: 682912 (~ /13) *6829110 (~ /10)*
  - maximum: 71274521 (~ /6) *712745216 (~ /3)*
- ADVERTISED SPACE DISTR. (data from VANTAGE graph):
  - minimum: 0 *8 (~ /29)*
  - average: 682911 (~ /13) *6829125 (~ /10)*
  - maximum: 71274521 (~ /6) *712745216 (~ /3)*

- CONNECTED COMPONENTS:
  - total number: 19
  - cc's to edges ratio: 0,0165505
  - size distribution:
    - 18 cc's of size 2
    - 1 cc of size 453

INCONSISTENCY graph’s edges distribution by Looking Glass

```
AS 1    sees 48  of the differently oriented edges
AS 1740 sees 53  of the differently oriented edges
AS 3549 sees 85  of the differently oriented edges
AS 3582 sees 918 of the differently oriented edges
AS 3967 sees 127 of the differently oriented edges
AS 4197 sees 68  of the differently oriented edges
AS 5388 sees 68  of the differently oriented edges
AS 7018 sees 57  of the differently oriented edges
AS 8220 sees 169 of the differently oriented edges
AS 8709 sees 107 of the differently oriented edges
```

Complement of the INCONSISTENCY graph

This graph is built on edges that have all but opposite orientations in SAT and in VANTAGE. (namely, edges that are peerings in VANTAGE and customer-provider in SAT). In this sense it is the complement (on the edges) of the INCONSISTENCY graph.

Path covering distribution and Advertised space distribution

They both follow the criteria shown for the DIFFERENTIAL graph.

Complement of the INCONSISTENCY graph's edges distribution by Looking Glass

This list shows how many edges of the complement of the INCONSISTENCY graph can be seen from each looking glass.

Complement of the INCONSISTENCY graph

- NUMBER OF NODES: 779
- NUMBER OF EDGES: 1111

- PATH COVERING DISTRIBUTION (data from SAT graph):
  - minimum: 1
  - average: 444
  - maximum: 19918
- PATH COVERING DISTR. (data from VANTAGE graph):
  - minimum: 0 *1*
  - average: 300 *3002*
  - maximum: 16615 *166153*
- ADVERTISED SPACE DISTR. (data from SAT graph):
  - minimum: 6 (~ /30) *64 (~ /26)*
  - average: 7482548 (~ /10) *74825480 (~ /6)*
  - maximum: 382845158 (~ /4) *3828451584 (~ /1)*
- ADVERTISED SPACE DISTR. (data from VANTAGE graph):
  - minimum: 6 (~ /30) *64 (~ /26)*
  - average: 7482542 (~ /10) *74825472 (~ /6)*
  - maximum: 382845158 (~ /4) *3828451584 (~ /1)*

- CONNECTED COMPONENTS:
  - total number: 197
  - cc's to edges ratio: 0,1773177
  - size distribution:
    - 168 cc's of size 2
    - 20 cc's of size 3
    - 4 cc's of size 4
    - 1 cc of size 5
    - 1 cc of size 7
    - 1 cc of size 19
    - 1 cc of size 81
    - 1 cc of size 255

Complement of the INCONSISTENCY graph's edges distribution by Looking Glass

AS 1 sees 215 of the differently oriented edges
AS 1740 sees 240 of the differently oriented edges
AS 3549 sees 235 of the differently oriented edges
AS 3582 sees 965 of the differently oriented edges
AS 3967 sees 453 of the differently oriented edges
AS 4197 sees 222 of the differently oriented edges
AS 5388 sees 231 of the differently oriented edges
AS 7018 sees 244 of the differently oriented edges
AS 8220 sees 239 of the differently oriented edges
AS 8709 sees 319 of the differently oriented edges

ISP hierarchy according to SAT

This hierarchical distribution is inferred on the basis of information provided by the SAT graph. The algorithm used is the following: an arbitrary node is labelled with an arbitrary layer (which will not influence the final distribution); then, the choice is spread around on its neighbouring nodes, in the following way: if the current node and the neighbouring one are linked by an edge that goes towards the neighbouring node, the latter is assigned a layer that is the one of the current node plus 1; conversely, if the edge is directed towards the current node, the neighbouring one is assigned a layer that is one unit smaller than the one of the current node. Each node that is assigned a new layer or that must change its old one spreads its status on its neighbours, until a stable assignment is reached (i.e. no more layer updates must be performed). Nodes belonging to the same strongly connected components are forced to have the same layer.

The so obtained hierarchy is really different from VANTAGE's.

ISP hierarchy according to VANTAGE

Agarwal et al. suggest not to build a hierarchy by assigning a layer to each node, but to point out those sets of nodes that satisfy a certain property (for example, inner core nodes should have a good connectivity each other).

Warning: in considering the two hierarchies you must be aware that they are computed on different data sets with different methods.

ISP hierarchy according to SAT

```
   1 = highest layer (large ISPs, "dense core")
  12 = lowest layer  (small ISPs, "customers")

   9 ASes in layer 1
       (12095, 4553, 8526, 8627, 12315, 4589, 8356, 8652, 8914)
 586 ASes in layer 2
   5 ASes in layer 3 (13126, 8513, 4005, 18477, 12976)
 155 ASes in layer 4
7333 ASes in layer 5
1905 ASes in layer 6
 629 ASes in layer 7
 215 ASes in layer 8
  48 ASes in layer 9
   9 ASes in layer 10
       (6880, 17147, 16780, 16990, 13161, 12591, 5467, 8530,
       23004)
   2 ASes in layer 11 (5461, 9056)
   1 ASes in layer 12 (2643)
```

ISP hierarchy according to VANTAGE

```
   1 = highest layer (large ISPs, “dense core”)
   5 = lowest layer  (small ISPs, “customers”)

  20 ASes in layer 1
       (1, 1239, 174, 1755, 1833, 209, 2548, 2828, 2914, 3356,
       3549, 3561, 3967, 4006, 4200, 5511, 6453, 701, 7018,
       8918)
 129 ASes in layer 2
 897 ASes in layes 3
 971 ASes in layer 4
8898 ASes in layer 5
```

DIFFERENTIAL graph's hierarchical node distribution according to VANTAGE

This is the distribution of node layers in the classification provided by Agarwal et al.

Numbers between brackets show the distribution only for the nodes of the largest connected component of the DIFFERENTIAL graph.

INCONSISTENCY graph's hierarchical node distribution according to VANTAGE

Exactly the same as DIFFERENTIAL graph's.

VANTAGE peering hierarchical distribution according to VANTAGE

This shows how many nodes in each layer have at least one peering relationship with any other node, where the layers are those provided by Agarwal et al.

VANTAGE's inner core

These values are reported just to show that no core AS has a non-peering relationship with other core ASes (conversely, all the core ASes only have peering relationships among themselves), and that core ASes almost form a clique. It is also shown how many of the peerings involve at least one core AS.

DIFFERENTIAL graph's hierarchical node distribution according to VANTAGE

```
 20  (20) ASes in layer 1
127 (127) ASes in layer 2
708 (358) ASes in layer 3
113 (102) ASes in layer 4
 45  (39) ASes in layer 5
```

INCONSISTENCY graph's hierarchical node distribution according to VANTAGE

```
 18  (18) ASes in layer 1
 96  (91) ASes in layer 2
217 (204) ASes in layer 3
113 (101) ASes in layer 4
 45  (39) ASes in layer 5
```

VANTAGE peering hierarchical distribution according to VANTAGE

```
 20 ASes in layer 1
124 ASes in layer 2
663 ASes in layer 3
  0 ASes in layer 4
  0 ASes in layer 5
```

VANTAGE'S inner core (layer 1 according to VANTAGE)

- PEERING EDGES: 156
- OTHER EDGES: 0

- PEERING EDGES IN WHOLE VANTAGE GRAPH:
  - total: 1136
  - with at least one core AS: 368
  - others: 768

Largest SCC's of SAT and VANTAGE

This report shows what happens inside the largest strongly connected components both of SAT and of VANTAGE.

Common zone

These values represent statistics about nodes and edges that appear in the largest strongly connected components of both the SAT and the VANTAGE graph.

Largest SCC's of SAT and VANTAGE

- SAT LARGEST SCC:
  - number of nodes: 586
  - number of edges: 4148

- VANTAGE LARGEST SCC:
  - number of nodes: 278
  - number of edges: 2345

- COMMON ZONE:
  - number of nodes: 249
  - number of edges:
    - total: 2236
    - consistently oriented:
      - number: 917
      - percentage of total: 41 %
    - oppositely oriented:
      - number: 587
      - percentage of total: 26 %
    - peers only in VANTAGE:
      - number: 732
      - percentage of total: 33 %

Total number of distinct paths

This is the number of distinct paths taken from the full list of AS paths (the catenation of the AS paths lists from all the 10 looking glasses).

Paths traversing >x edges of the INCONSISTENCY graph

This value is computed considering only distinct paths.

Number of subsequent crossings

This considers the number of times that a path traverses an edge of the INCONSISTENCY graph after having already traversed one. Therefore, this value is not a number of paths, but a number of crossings. The value is computed as follows: for each path (taken from the full list of distinct paths), if it traverses an edge of the INCONSISTENCY graph for the 2nd, 3rd time and so forth, the value is increased by 1 for each crossing subsequent to the first.

Crossing point distribution

This report shows, for each path, the position in which it traverses an edge of the INCONSISTENCY graph. The value is calculated as follows: for each path, a list of positions in which it traverses edges of the INCONSISTENCY graph is built; for example, path A B C D E traverses edge A B (if any) in position 0, edge B C (if any) in position 1, edge C D (if any) in position 2 and edge D E (if any) in position 3; then, all the values are normalized between 0 and 10 (in the example, position 0 remains 0, position 1 becomes 3, position 2 becomes 7 and position 3 becomes 10) and the distribution is built. The normalization is performed in order to make the positions be independent from path length.

Paths of length 2 are explicitly reported, because it is difficult to determine if the crossing takes place at their beginning or at their ending.

The first report (1st crossing only) only considers the first time that a path traverses an edge of the INCONSISTENCY graph (disregarding any further crossing); the second one also takes into account crossings that are subsequent to the first, and, as such, the values that are shown are number of crossings instead of number of paths.

Paths traversing Ig’s largest cc’s core ASes

The paths considered here are those that traverse at least one core AS (according to the hierarchy provided by Agarwal et al.) belonging to the largest connected component of the INCONSISTENCY graph. Values are computed on the basis of the full list of distinct paths.

“Ig” stands for INCONSISTENCY graph.

Paths traversing Ig’s edges

These are paths that traverse at least one edge of the INCONSISTENCY graph.

Also traversing unoriented edges

These values show how many missing (not oriented) adjacencies in the SAT and in the VANTAGE graph are crossed by paths traversing edges of the INCONSISTENCY graph. The value 0 for SAT means that none of the paths that traverse at least one edge of the INCONSISTENCY graph also runs across missing edges of the SAT graph. The value 38 for VANTAGE shows that 38 missing adjacencies are hit (possibly more than one time) by the paths that traverse at least one edge of the INCONSISTENCY graph.

Paths traversing edges of the INCONSISTENCY graph

- TOTAL NUMBER OF DISTINCT PATHS: 502517

- PATHS TRAVERSING >x EDGES OF THE INCONSISTENCY GRAPH:
  - x = 0: 43884
  - x = 1: 917
  - x = 2: 67
  - x = 3: 0
- NUMBER OF SUBSEQUENT CROSSINGS: 984

- CROSSING POINT DISTRIBUTION (1st crossing only)
  - 8324 paths in position 0
  - 248 paths in position 1
  - 4187 paths in position 2
  - 7223 paths in position 3
  - 1384 paths in position 4
  - 14041 paths in position 5
  - 5193 paths in position 6
  - 661 paths in position 7
  - 275 paths in position 8
  - 0 paths in position 9
  - 2094 paths in position 10
  - 254 paths of length 2
- CROSSING POINT DISTRIBUTION (all crossings)
  - 8431 crossings in position 0
  - 253 crossings in position 1
  - 4358 crossings in position 2
  - 7460 crossings in position 3
  - 1529 crossings in position 4
  - 14266 crossings in position 5
  - 5273 crossings in position 6
  - 671 crossings in position 7
  - 279 crossings in position 8
  - 0 crossings in position 9
  - 2094 crossings in position 10
  - 254 paths of length 2

Mixed analyses

- PATHS TRAVERSING IG’S LARGEST CC’S CORE ASES:
  - number: 454523
  - not traversing any Ig edge: 421093

- PATHS TRAVERSING IG’S EDGES:
  - number: 43884
  - not traversing any Ig’s largest cc’s core AS: 10454
  - not traversing any core AS: 10433
  - also traversing unoriented edges:
    - on SAT: 0
    - on VANTAGE: 38

Paths slope on SAT graph

An *ascending leg* of a path is a (possibly empty) subset of its edges (considered in the order in which they appear in the path) that includes the first node of the path and considers all edges that are assigned an ascending orientation by SAT. In other words, the ascending leg of a path AS_1, AS_2, ..., AS_n is a subset S of edges AS_1-AS_2, AS_2-AS_3, ..., AS_{x-1}-AS_x such that AS_1 appears in S, all the edges in S are contiguous and correspond to customer-provider edges of the SAT graph (SAT orientation is $AS_1 \rightarrow AS_2$, $AS_2 \rightarrow AS_3$, etc.).

A *descending leg* of a path is a (possibly empty) subset of its edges (considered in the order in which they appear in the path) that includes the last node of the path and considers all edges that are assigned a descending orientation by SAT. In other words, the descending leg of a path AS_1, AS_2, ..., AS_n is a subset S of edges AS_{n-k}-AS_{n-k+1}, ..., AS_{n-2}-AS_{n-1}, AS_{n-1}-AS_n such that AS_n appears in S, all the edges in S are contiguous and correspond to provider-customer edges of the SAT graph (SAT orientation is $AS_{n-2} \leftarrow AS_{n-1}$, $AS_{n-1} \leftarrow AS_n$, etc.).

The considered paths are those from the full list of paths (the catenation of all the AS paths lists from the 10 looking glasses), without repetitions.

Paths slope on SAT graph

- DISTINCT PATHS:
  - total: 502517
  - ascending leg longer than descending:
    - number: 167059
    - percentage of total: 33 %
  - same ascending and descending leg length:
    - number: 140364
    - percentage of total: 28 %
  - descending leg longer than ascending:
    - number: 195094
    - percentage of total: 39 %

Samples

These drawings show various situations of different orientations assigned by SAT and by VANTAGE. Circled and underlined AS numbers identify core ASes according to the hierarchy provided by Agarwal et al.

The choice of the paths is made as follows: a first group of paths is taken from the list of those that traverse at least one core AS; the remaining ones also transit over edges of the INCONSISTENCY graph (i.e. they are considered in order to show what happens when the assigned orientations are opposing).

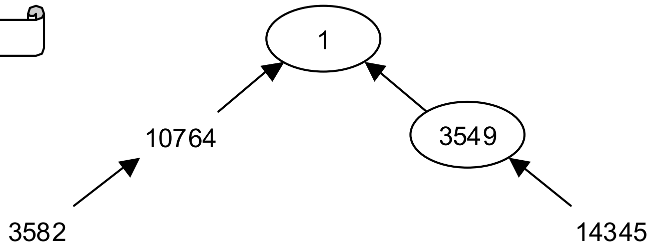
These examples also consider invalid paths (those that have a “valley”, or a chain of peerings, or a peering that is not located at “top level”). There is also one instance of missing orientation (that is represented with a missing edge).

In some cases it is difficult to say whether one solution is better than the other, but there are situations that clearly enhance the quality of either orientation. Of course, these are only a few samples of what happens all over the graphs, but they can provide an idea about the “behaviour” of the two inference systems (SAT’s and VANTAGE’s).

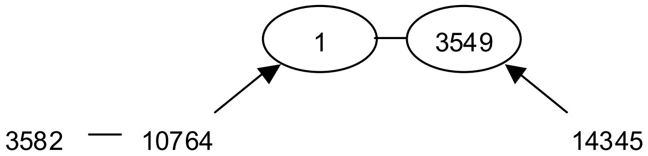
Samples

▶ PATH: 3582 10764 1 3549 14345

SAT

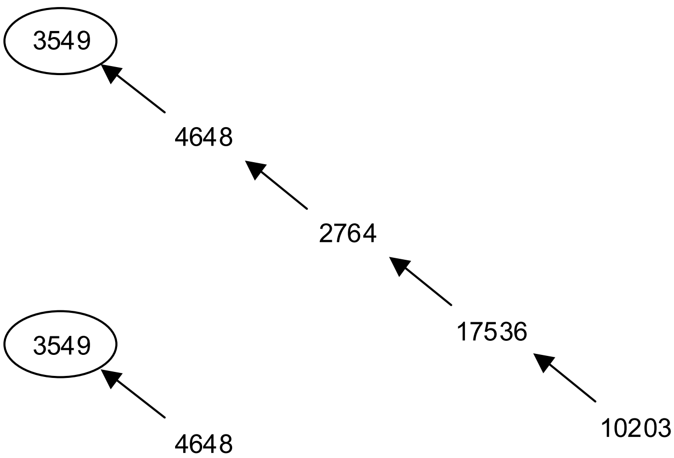


VANTAGE

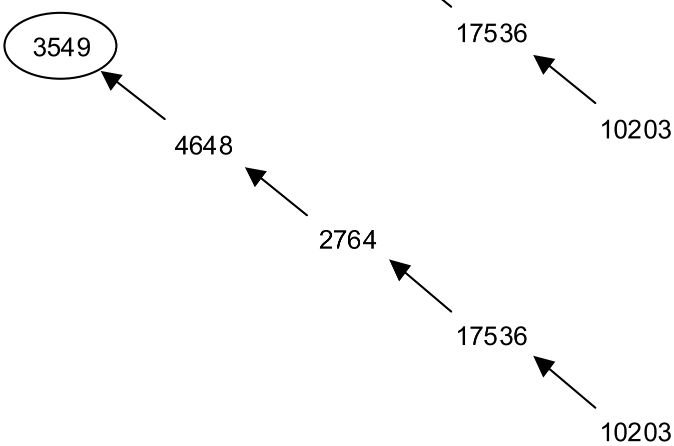


▶ PATH: 3549 4648 2764 17536 10203

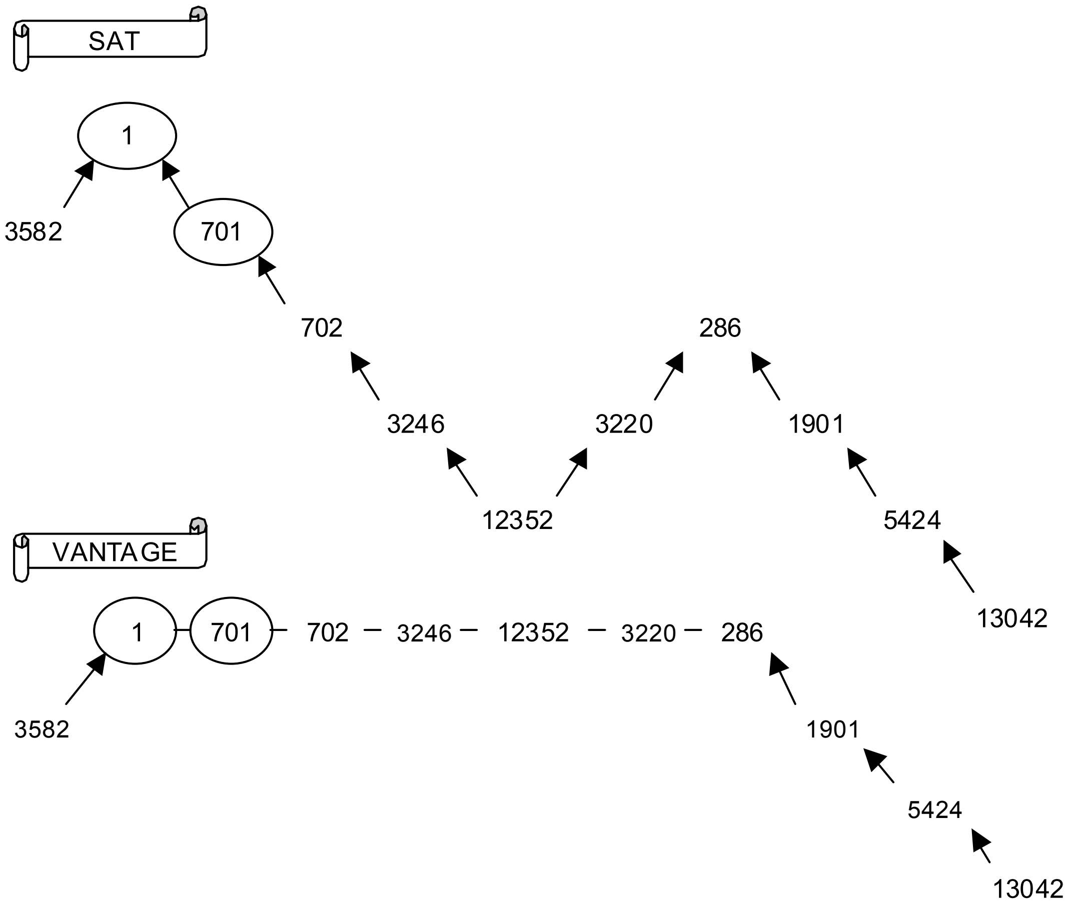
SAT



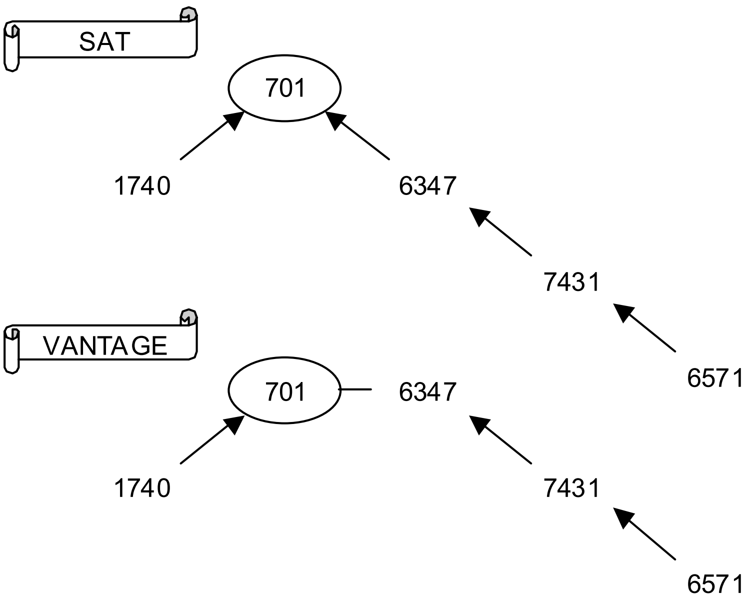
VANTAGE



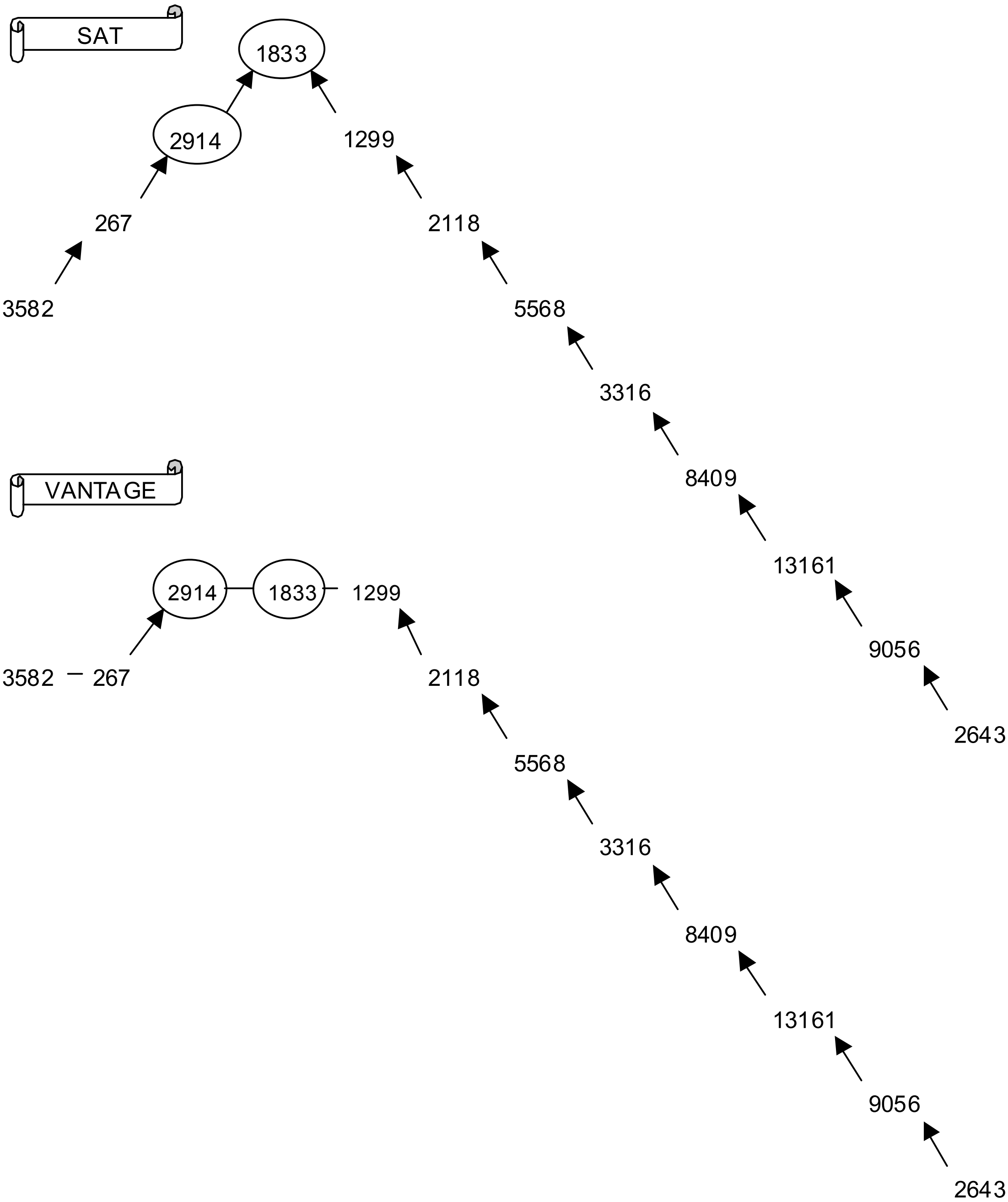
▶ PATH: 3582 <u>1</u> <u>701</u> 702 3246 12352 3220 286 1901 5424 13042



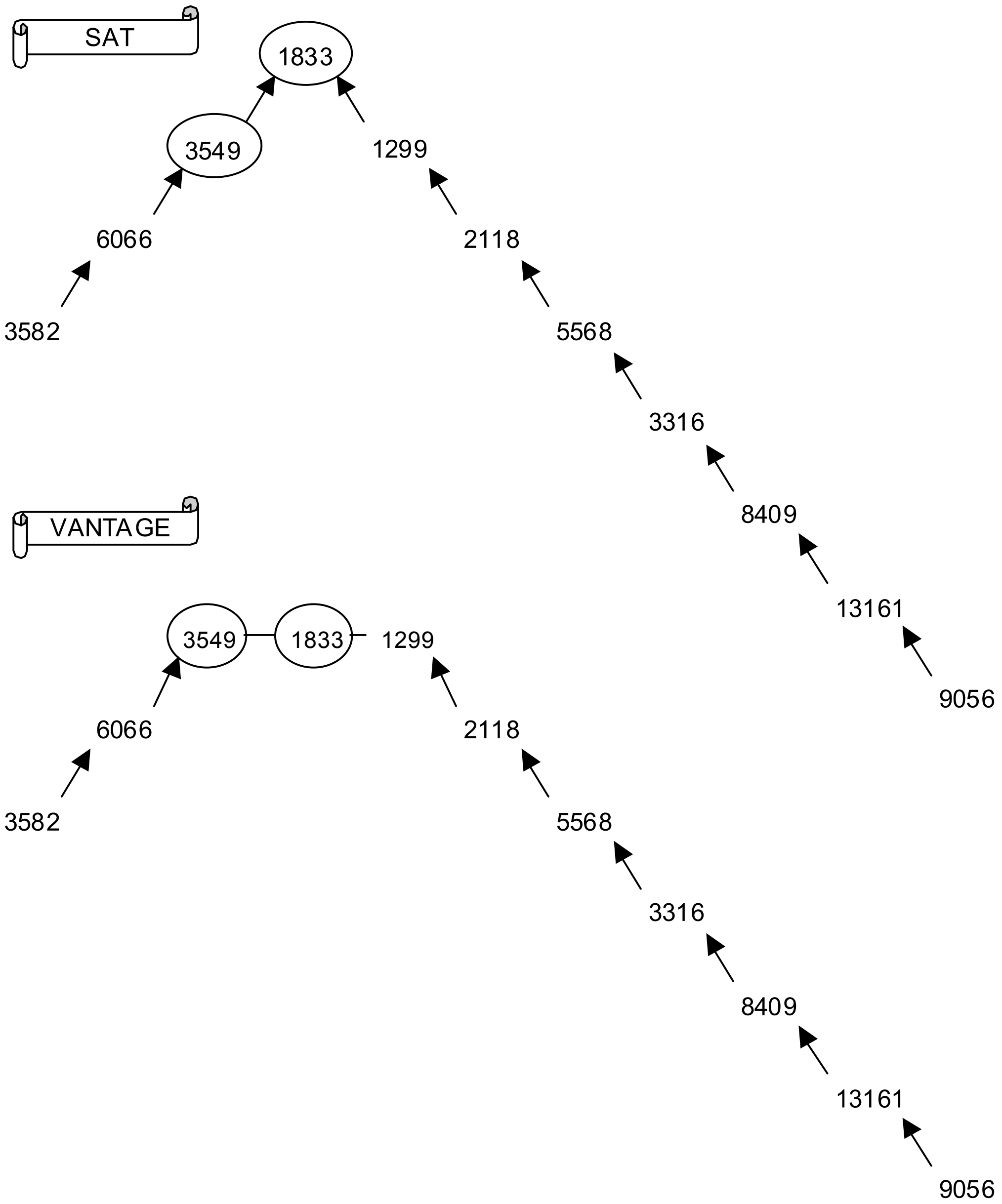
▶ PATH: 1740 <u>701</u> 6347 7431 6571



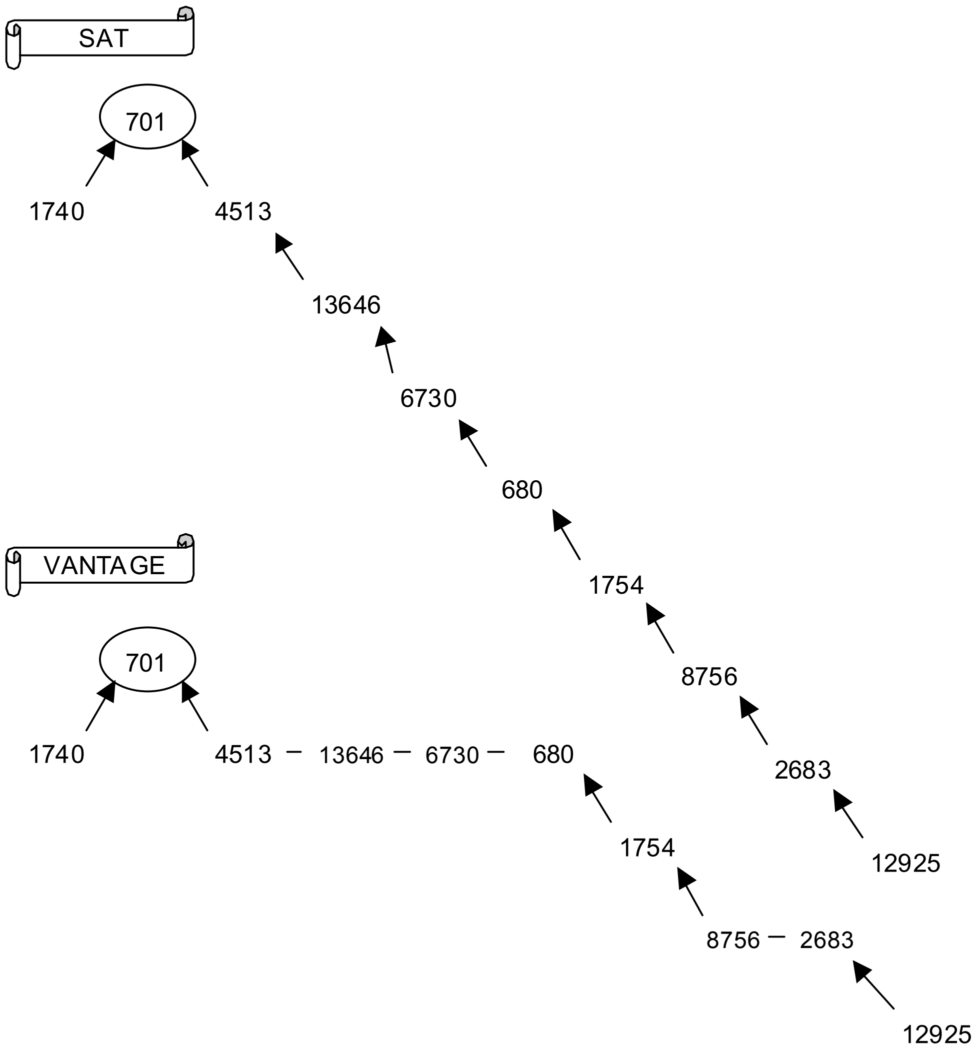
▶ PATH： 3582 267 2914 1833 1299 2118 5568 3316 8409 13161 9056 2643



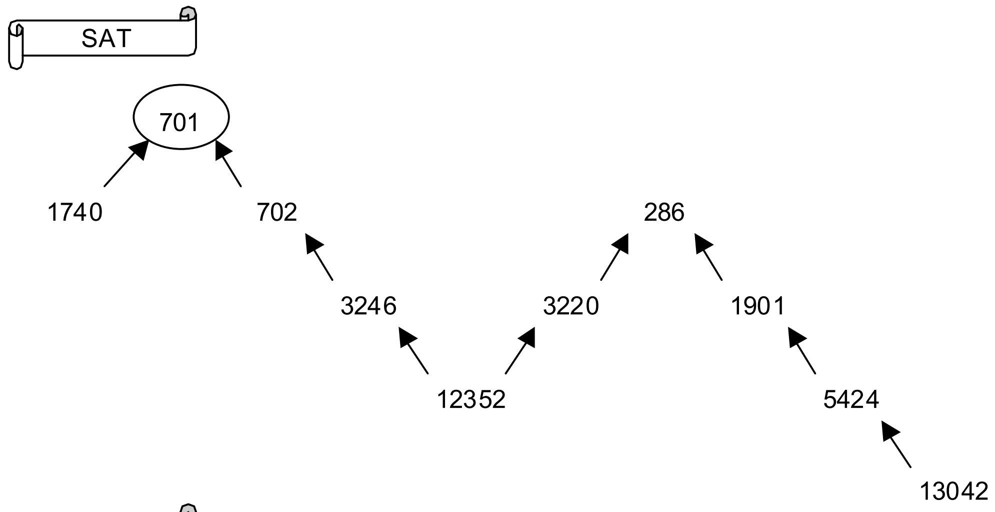
▶ PATH: 3582 6066 3549 1833 1299 2118 5568 3316 8409 13161 9056

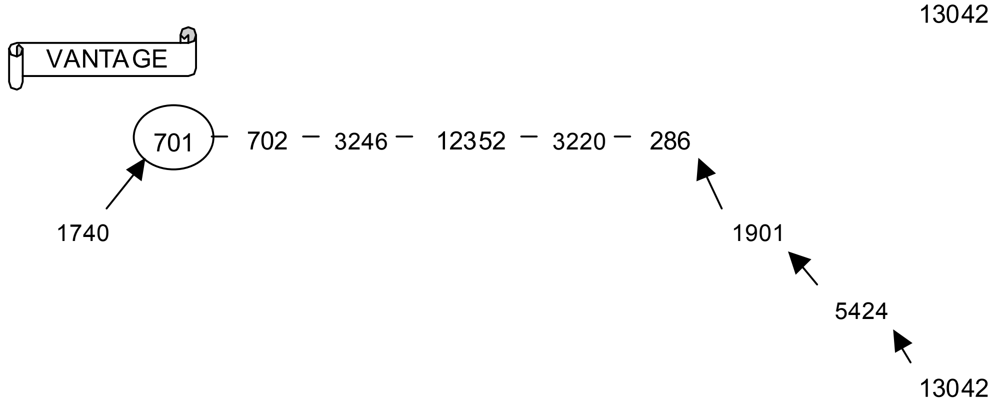


▶ PATH: 1740 701 4513 13646 6730 680 1754 8756 2683 12925

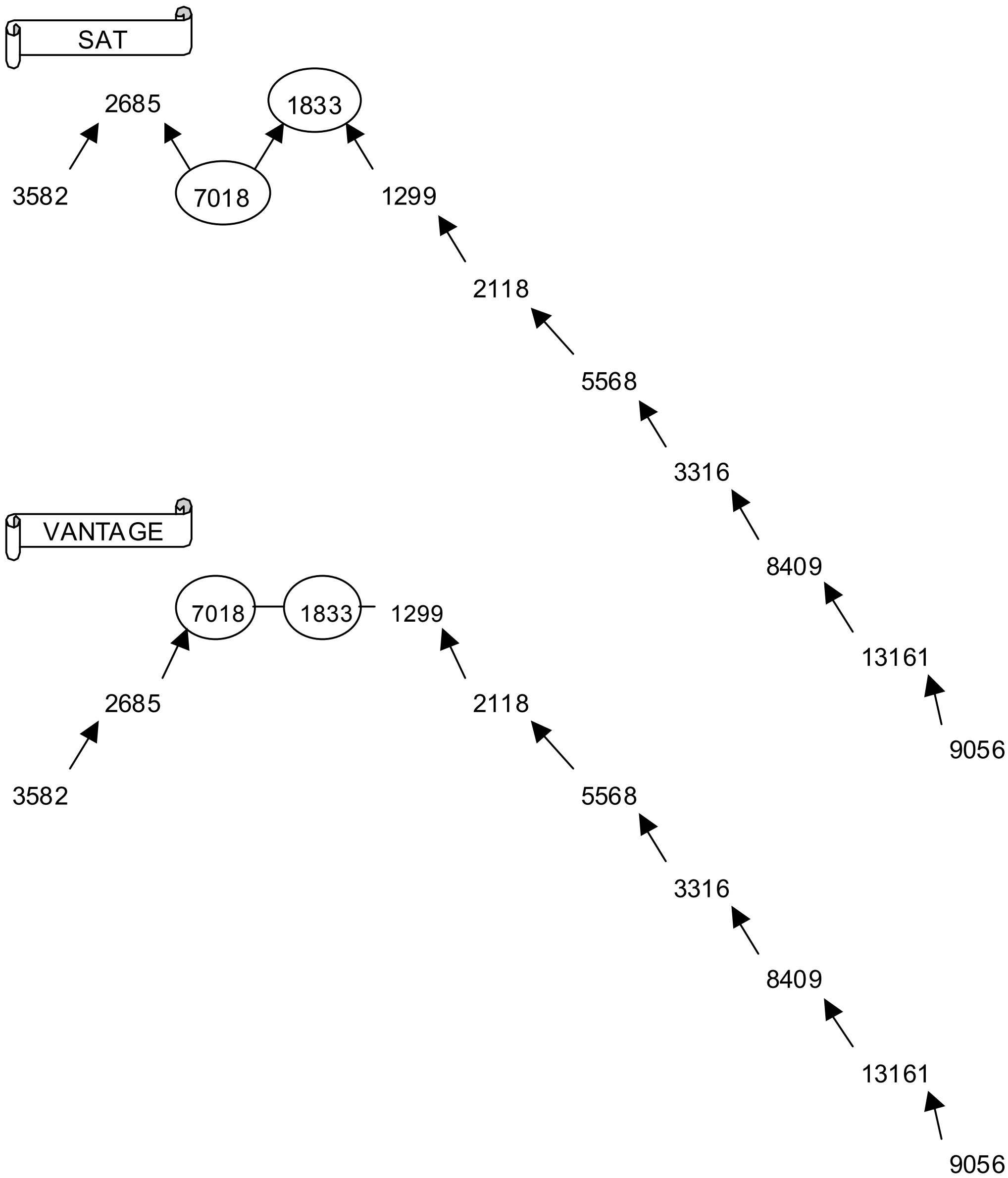


▶ PATH: 1740 701 702 3246 12352 3220 286 1901 5424 13042

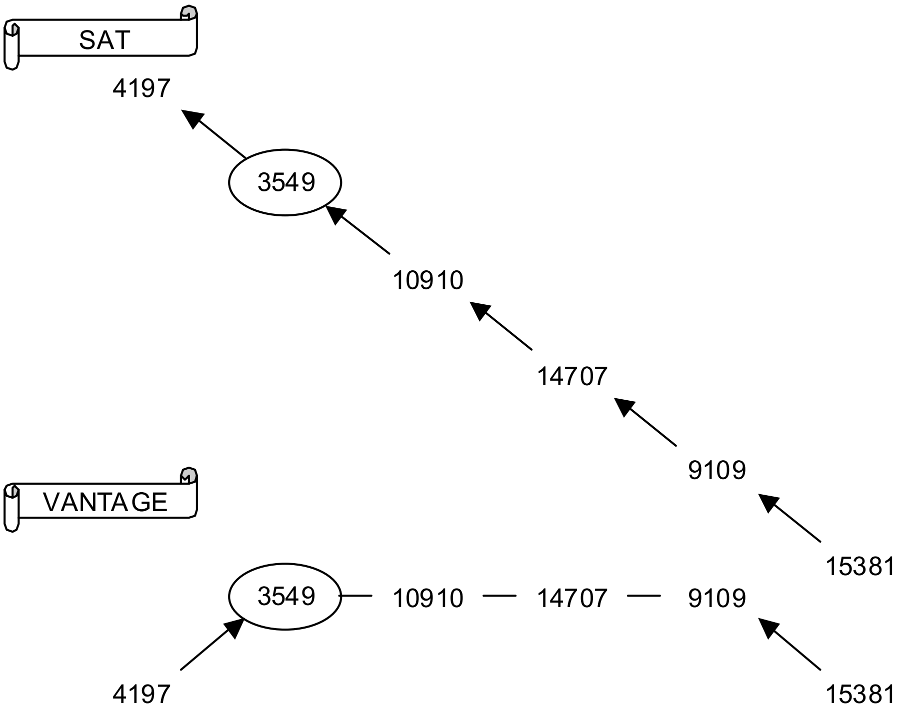




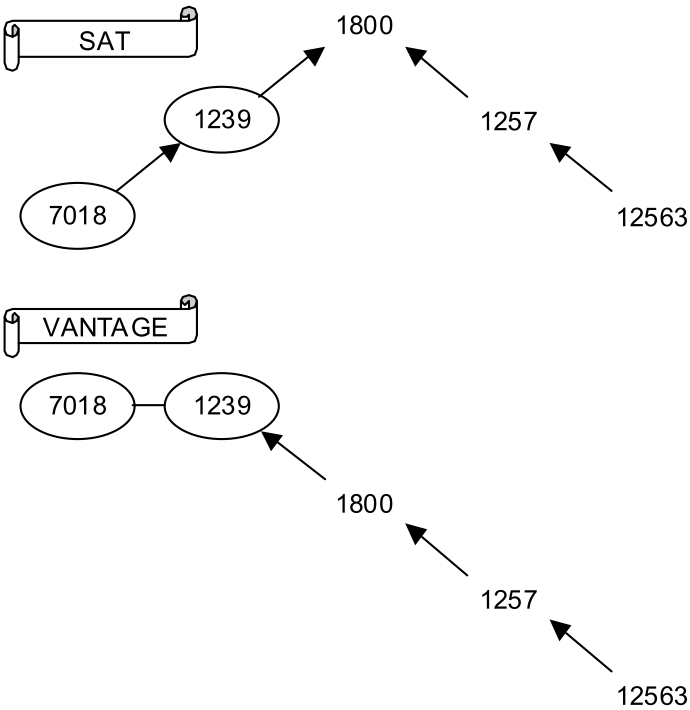
▶ PATH: 3582 2685 7018 1833 1299 2118 5568 3316 8409 13161 9056



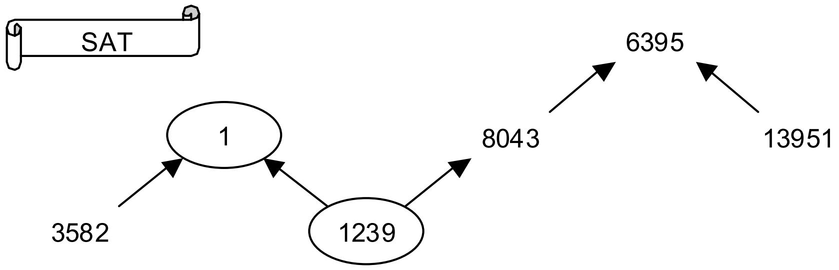
▶ PATH: 4197 3549 10910 14707 9109 15381

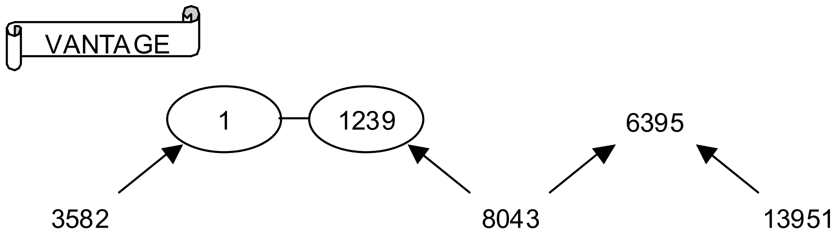


▶ PATH: 7018 1239 1800 1257 12563

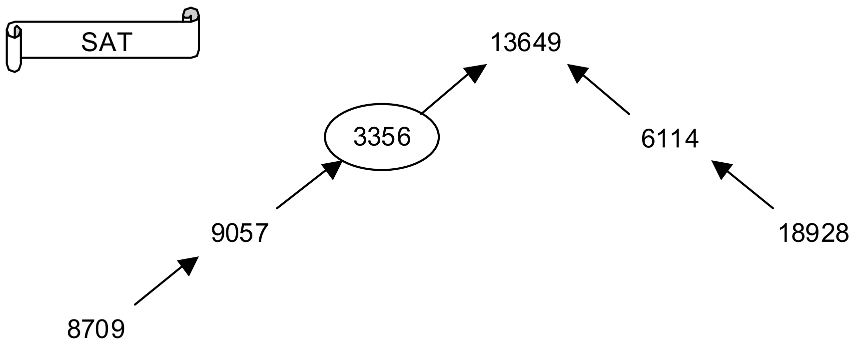


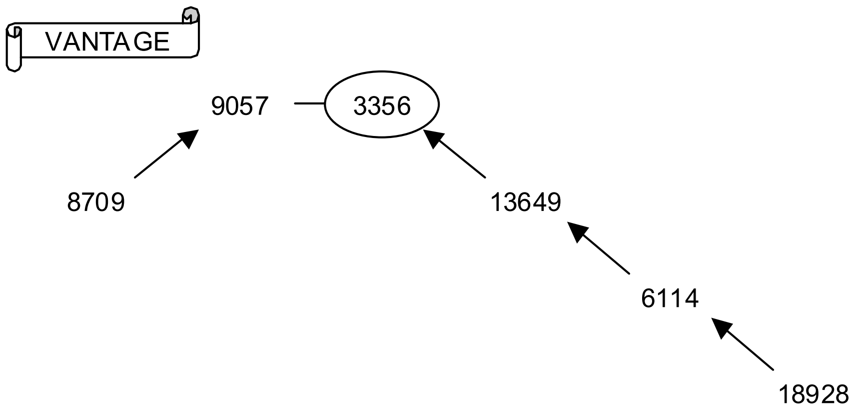
▶ PATH: 3582 1 1239 8043 6395 13951



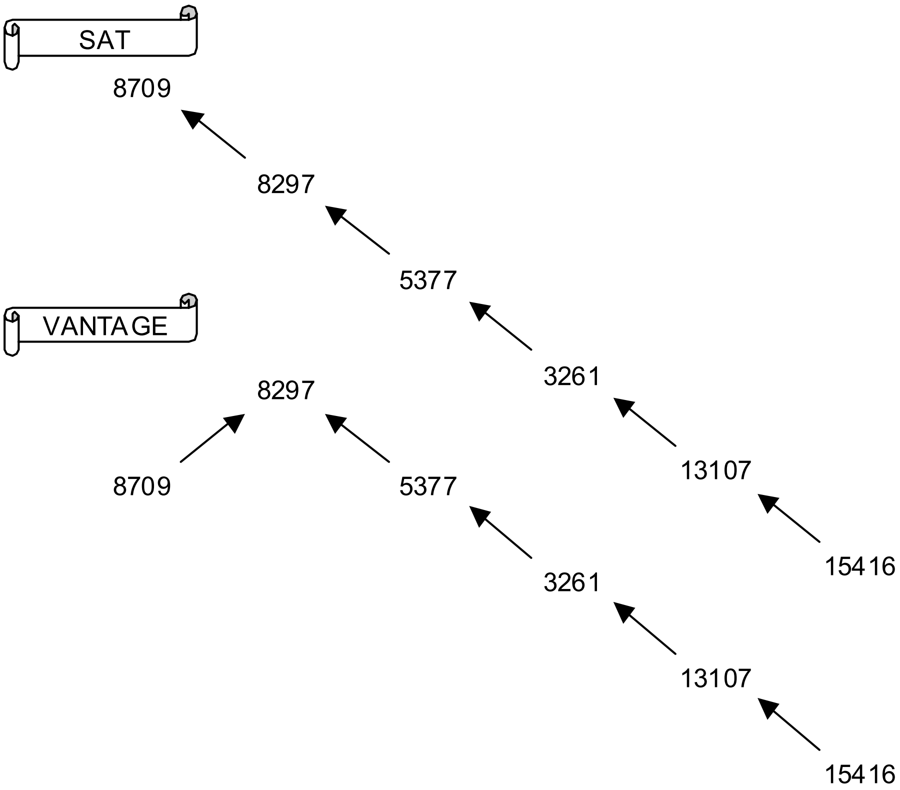


▶ PATH: 8709 9057 3356 13649 6114 18928

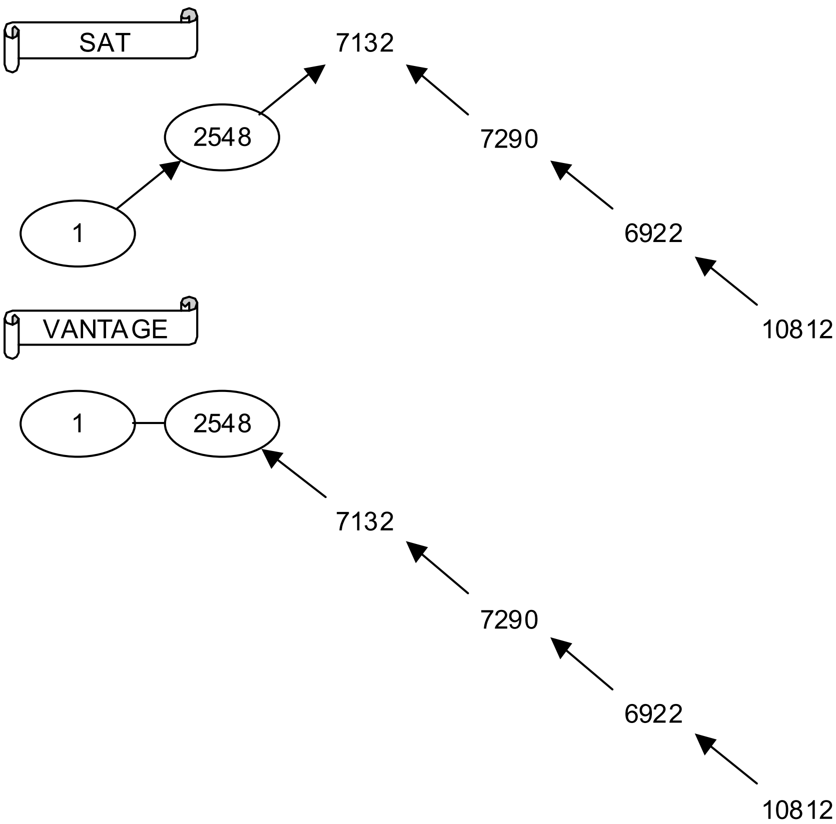




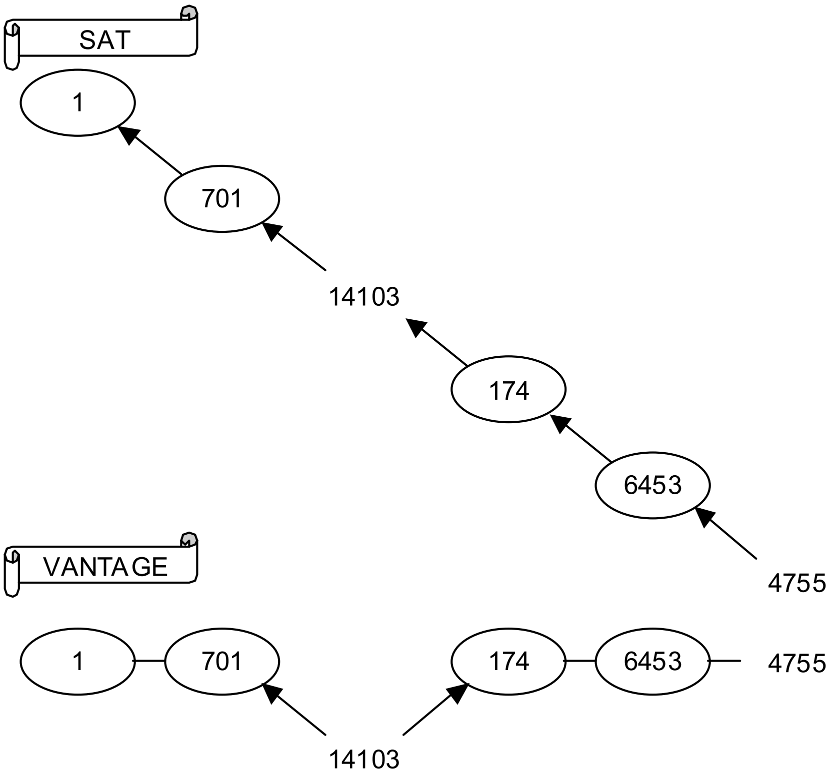
▶ PATH: 8709 8297 5377 3261 13107 15416



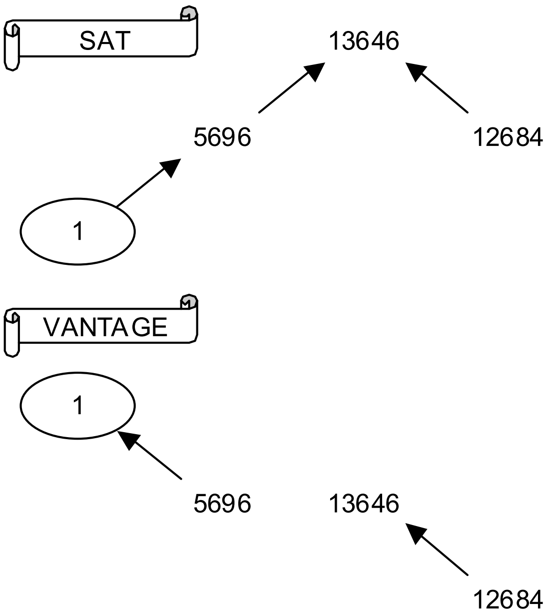
▶ PATH: 1 2548 7132 7290 6922 10812



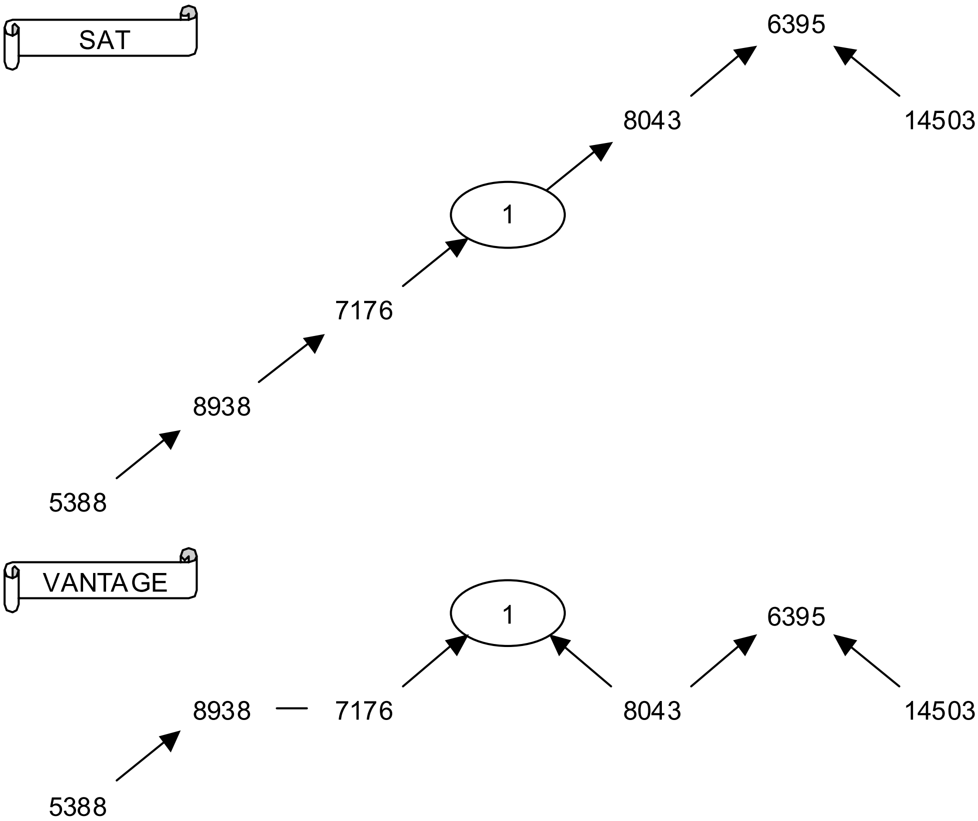
▶ PATH: 1 701 14103 174 6453 4755



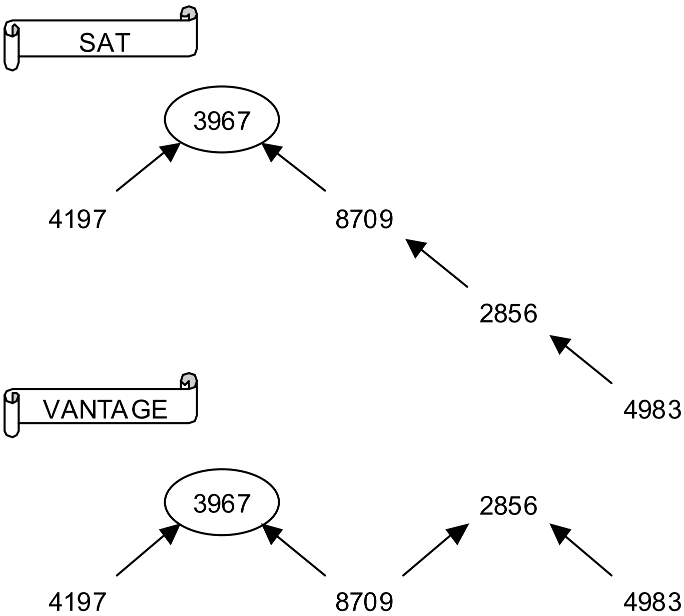
▶ PATH: 1 5696 13646 12684



▶ PATH: 5388 8938 7176 1 8043 6395 14503



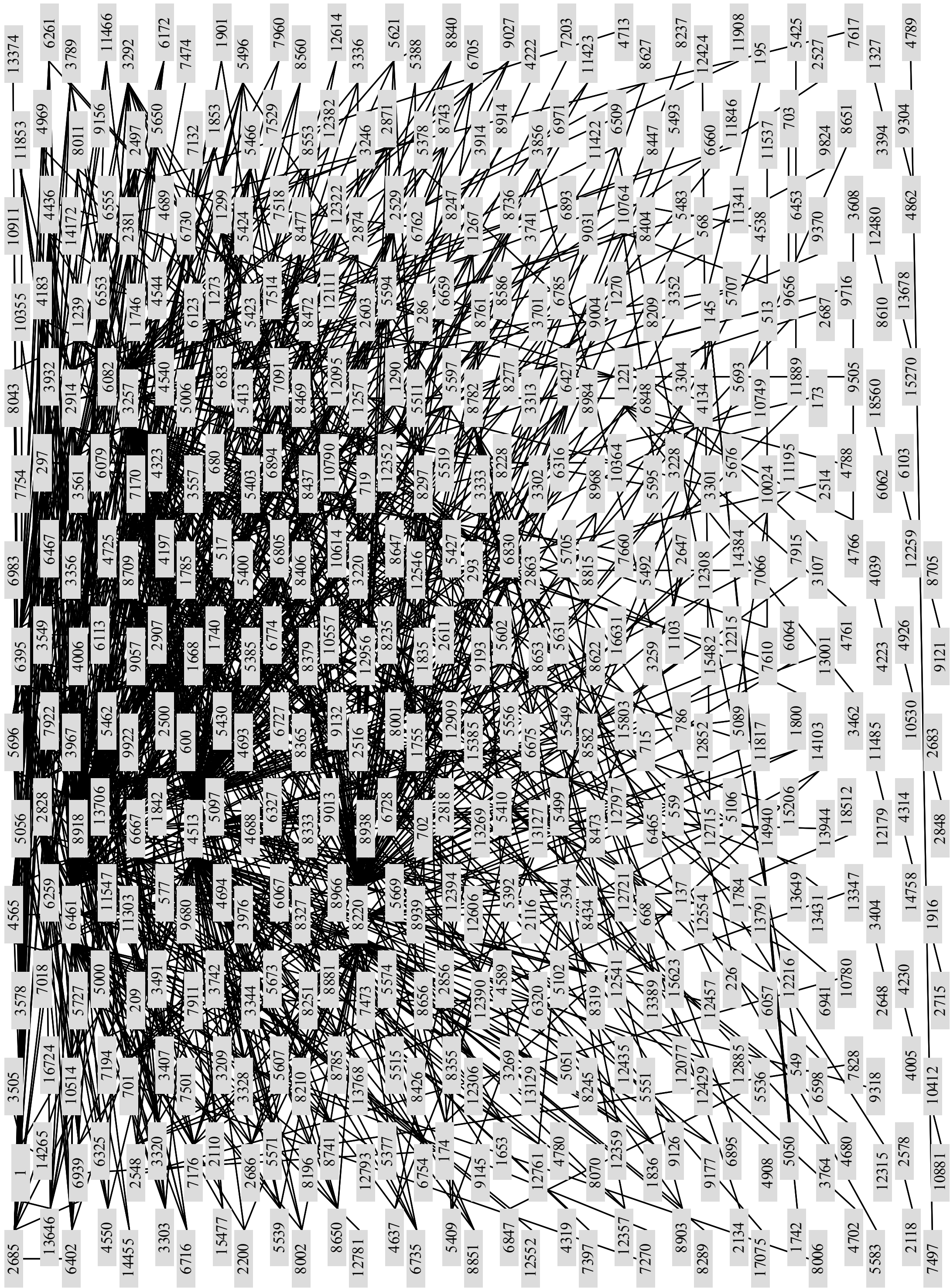
▶ PATH: 4197 3967 8709 2856 4983



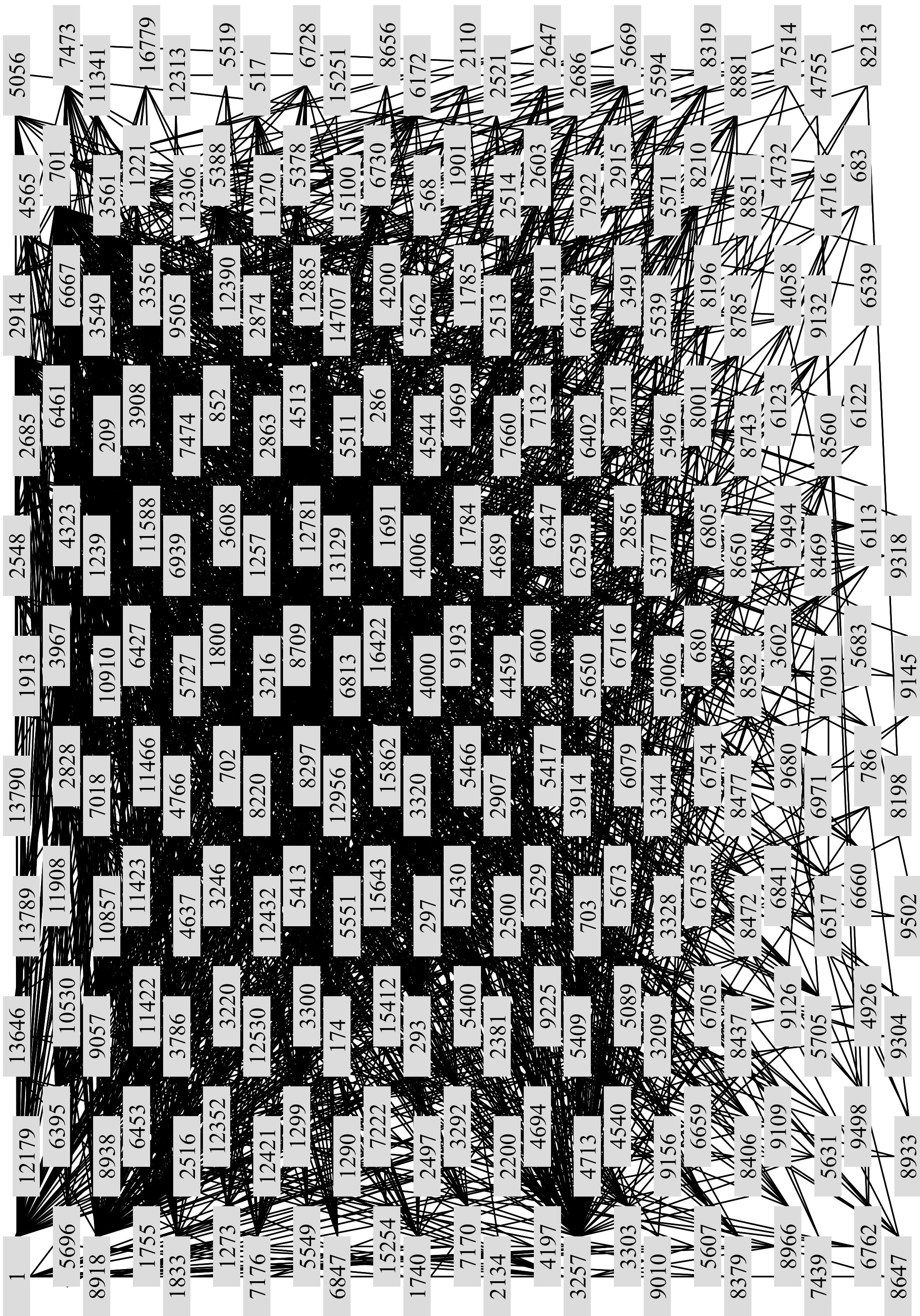
Sketches

These drawings are reported with the purpose of providing a raw idea of how some graphs would appear. Nodes displacement does not follow a particular order.

▶ INCONSISTENCY graph



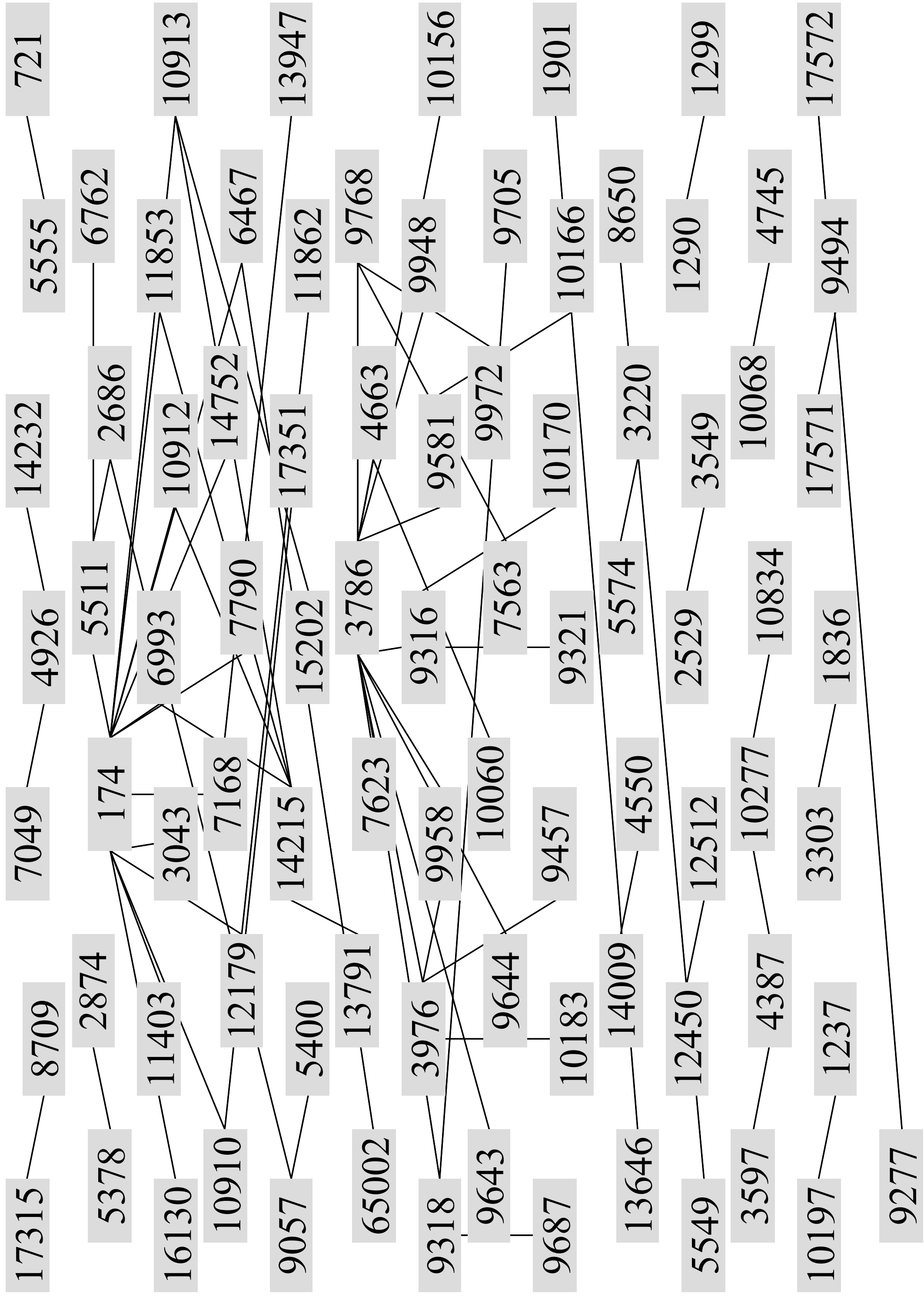
▶ COMMON EDGES IN SAT/VANTAGE'S LARGEST SCC



Missing (unoriented) edges

This graph, together with the following one, shows (in a much more readable way than the preceding ones) which are the ASes whose adjacencies are most difficult to be oriented.

▶ MISSING (UNORIENTED) EDGES ON SAT GRAPH



▶ MISSING (UNORIENTED) EDGES ON VANTAGE GRAPH