

Net2Text: Query-Guided Summarization of Network Forwarding Behaviors



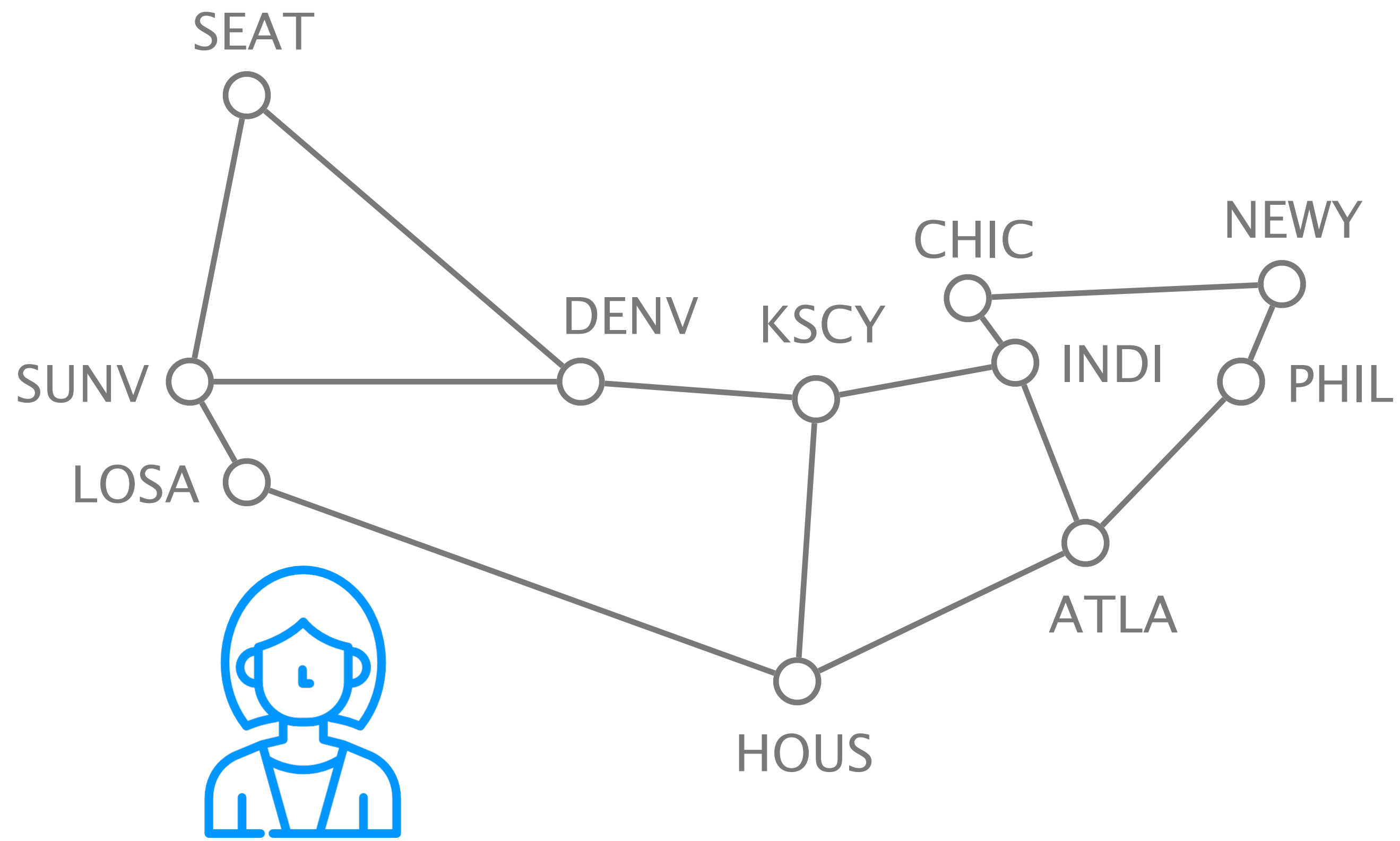
Rüdiger Birkner, Dana Drachsler-Cohen,
Martin Vechev, Laurent Vanbever

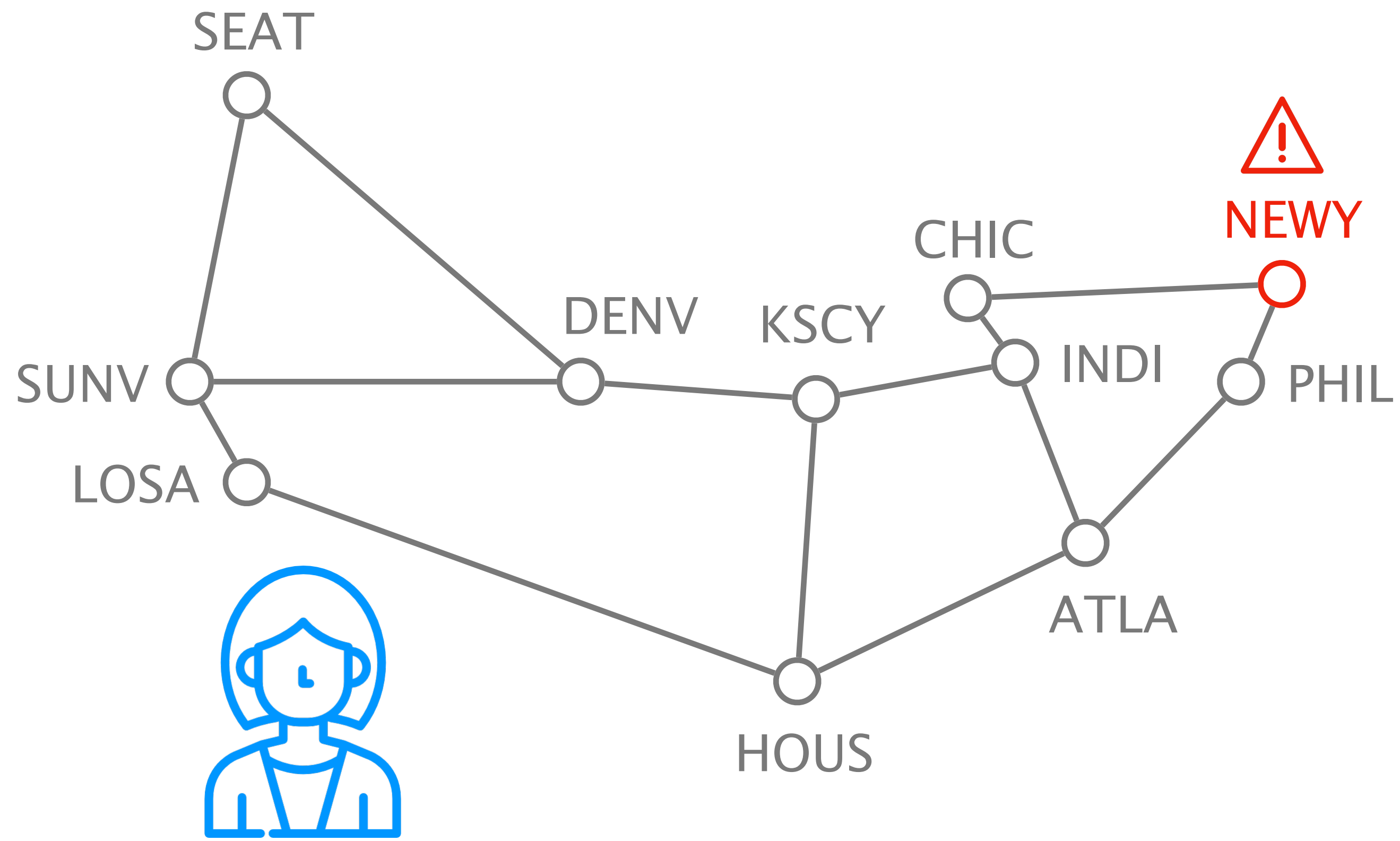
net2text.ethz.ch

NSDI '18

April, 11 2018

ETH zürich







Where is the traffic leaving in NEWY
coming from?

Where is the traffic leaving NEWY coming from?

Approach

Look at entire forwarding state
all the traffic statistics *to*

identify important destinations to reroute

Challenge

From a wealth of low-level data,
extract the high-level insights

Understanding how the network behaves,
can take hours

Fast reaction is required

Customer experience depends on it

Networks get more and more complex

New peerings, more routers, etc.

What if you could simply ask the questions...
and automatically get an answer?

Net **2** *Text*



|Type a message...

Net **2** *Text*



Where is the traffic...

Net2Text



question *in*
natural language

Where is the traffic leaving
in NEWY coming from?



Type a message...

Net2Text



Where is the traffic leaving
in NEWY coming from?



Type a message...

Net2Text



question *in*
natural language

Where is the traffic leaving
in NEWY coming from?



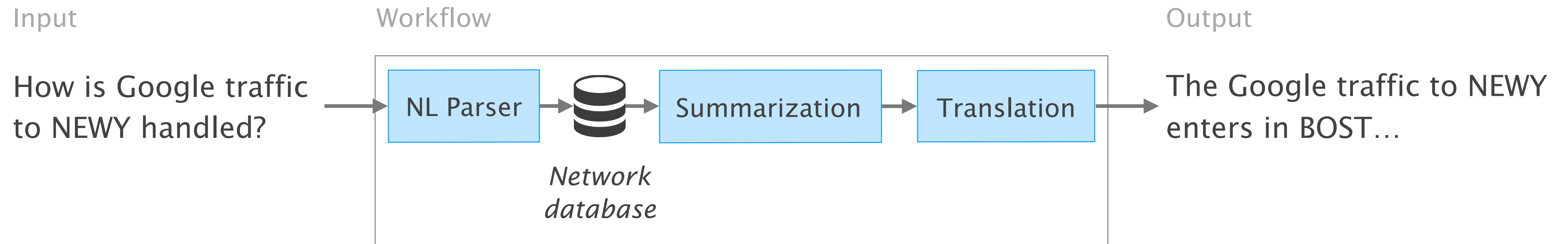
The traffic enters mostly in PHIL
and goes to Youtube and Netflix.



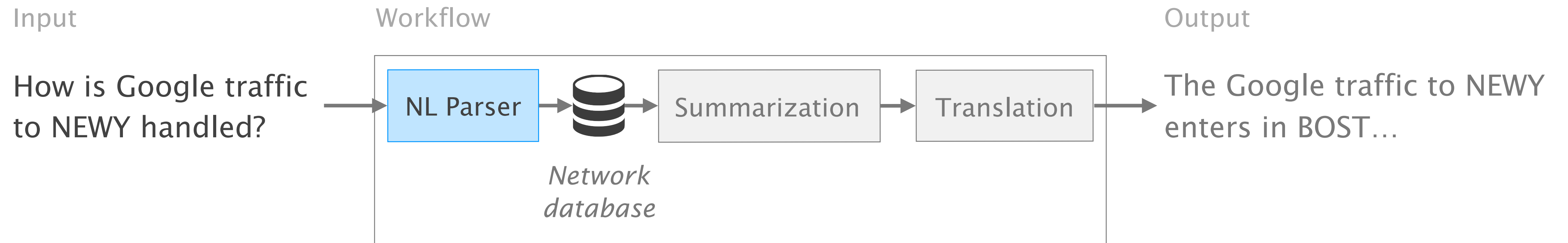
summary *in*
natural language

Type a message...

Net2Text has four stages:
parsing, data retrieval, summarization, translation



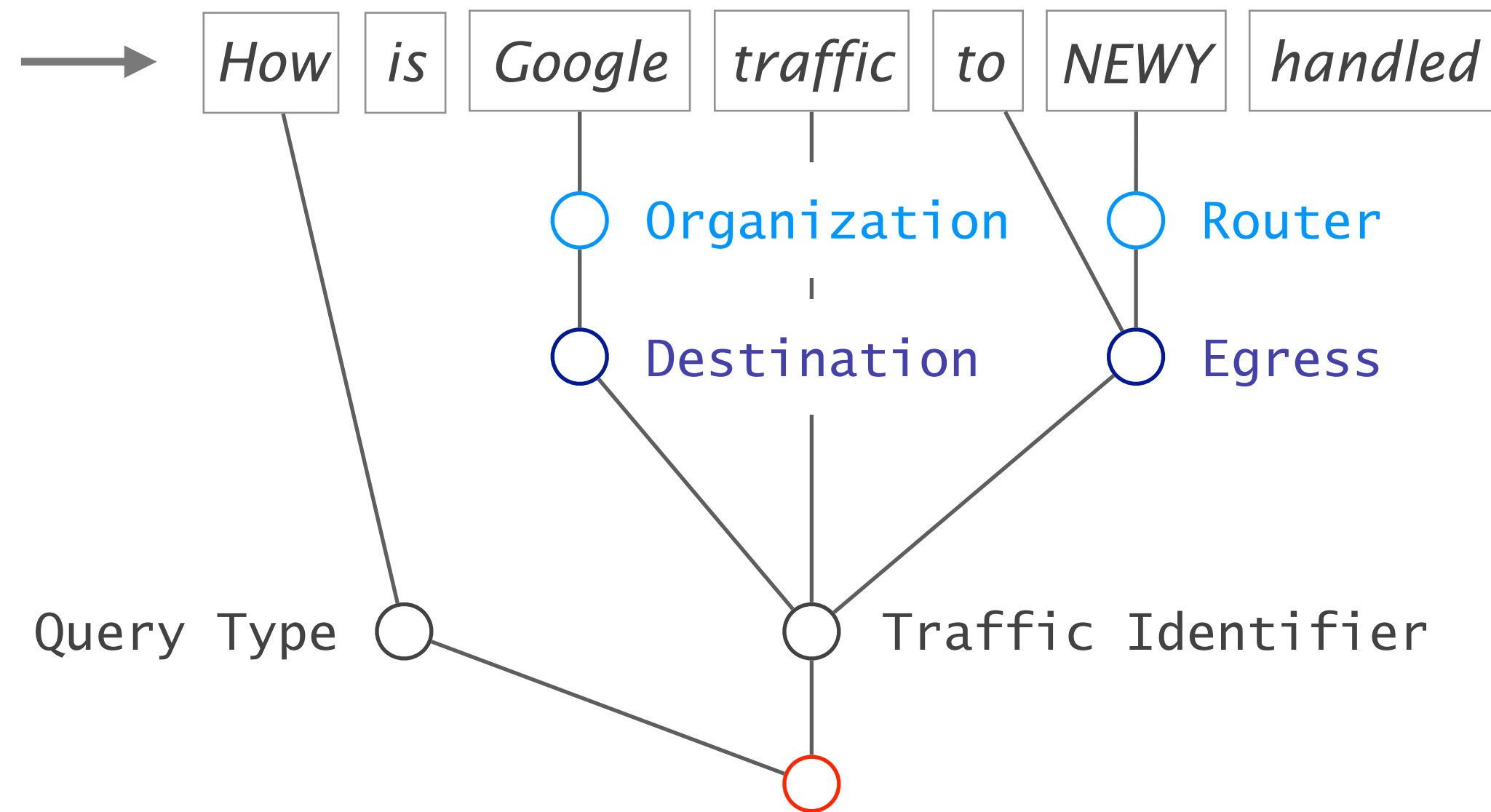
Net2Text has four stages:
parsing, data retrieval, summarization, translation



The parser maps the operator's query to the internal query language

Input

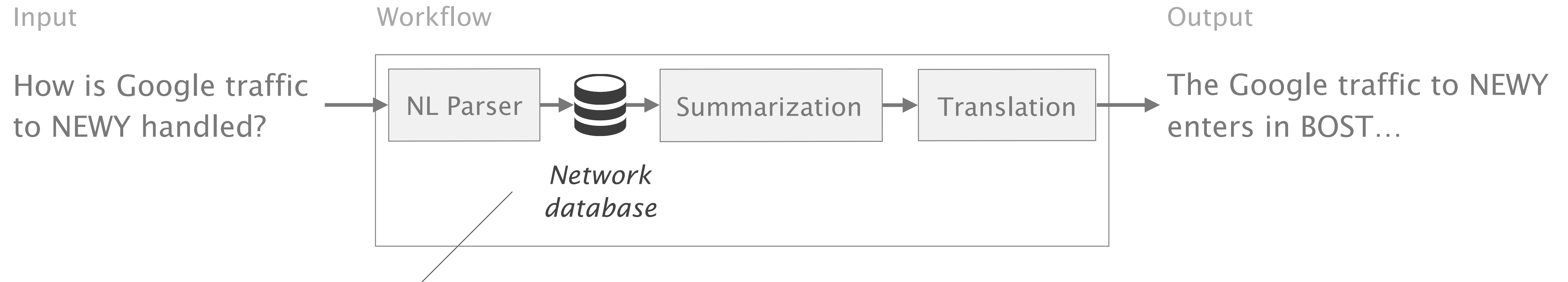
How is Google traffic to NEWY handled?



Output

```
SELECT * FROM paths  
WHERE egress=NEWY  
AND dest=Google
```

Based on the query, Net2Text retrieves all relevant data



```
SELECT * FROM paths
WHERE egress=NEWY
AND dest=Google
```

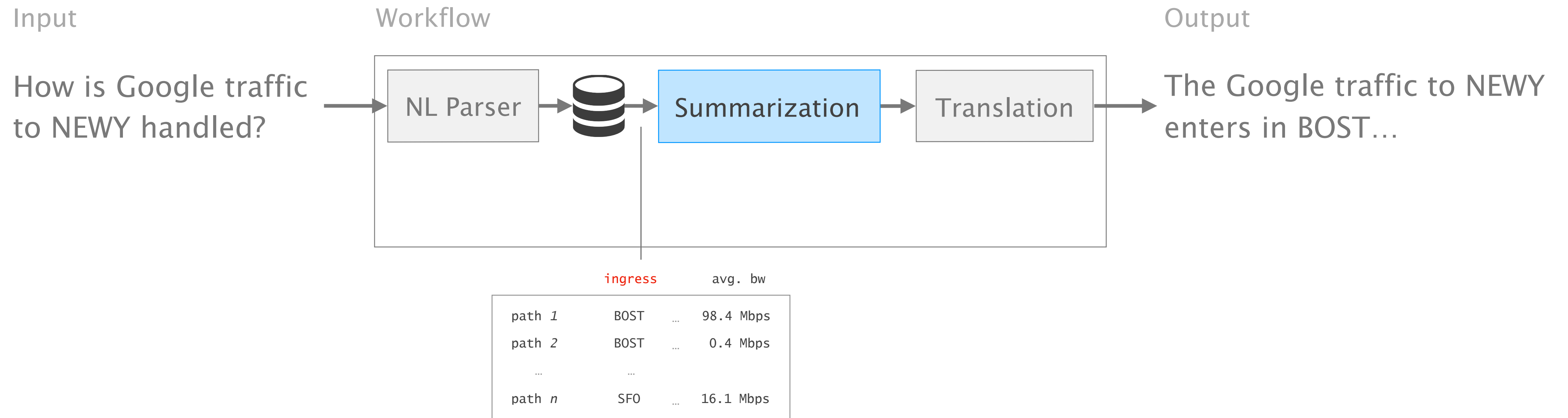

The database maintains
the forwarding state and traffic statistics

The database maintains the forwarding state and traffic statistics



	prefix	dest.	ingress	egress	...	avg. bw
path 1	8.8.8.0/24	Google	BOST	NEWY	...	98.4 Mbps
path 2	46.14.0.0/16	Swisscom	BOST	NEWY	...	0.4 Mbps
path 3	81.63.0.0/17	Swisscom	ATLA	NEWY	...	25.0 Mbps
...
path n	8.8.178.0/24	Yahoo	ATLA	HOUS	...	1.0 Mbps

All the data is summarized by identifying a few clusters



All the data is summarized by identifying a few clusters

Input pertaining to Google traffic leaving in NEWY

Output

	prefix	ingress	short. path	...	avg. bw
path 1	8.8.8.0/24	BOST	T	...	98.4 Mbps
path 2	8.8.4.0/24	BOST	T	...	0.4 Mbps
path 3	66.102.0.0/20	BOST	F	...	25.0 Mbps
path 4	35.184.0.0/19	HOUS	F	...	25.0 Mbps
...
path n	35.184.0.0/19	BOST	T	...	1.0 Mbps

All the data is summarized by identifying a few clusters

Input pertaining to Google traffic leaving in NEWY

	prefix	ingress	short. path	...	avg. bw
path 1	8.8.8.0/24	BOST	T	...	98.4 Mbps
path 2	8.8.4.0/24	BOST	T	...	0.4 Mbps
path 3	66.102.0.0/20	BOST	F	...	25.0 Mbps
path 4	35.184.0.0/19	HOUS	F	...	25.0 Mbps
...
path <i>n</i>	35.184.0.0/19	BOST	T	...	1.0 Mbps

Output



{BOST_{*i*}},

Each cluster represents a path specification

A summary consists of multiple path specifications

Input pertaining to Google traffic leaving in NEWY

	prefix	ingress	short. path	...	avg. bw
path 1	8.8.8.0/24	BOST	T	...	98.4 Mbps
path 2	8.8.4.0/24	BOST	T	...	0.4 Mbps
path 3	66.102.0.0/20	BOST	F	...	25.0 Mbps
path 4	35.184.0.0/19	HOUS	F	...	25.0 Mbps
...
path n	35.184.0.0/19	BOST	T	...	1.0 Mbps



Output

{BOST_i},

All the data is summarized by identifying a few clusters

Input pertaining to Google traffic leaving in NEWY

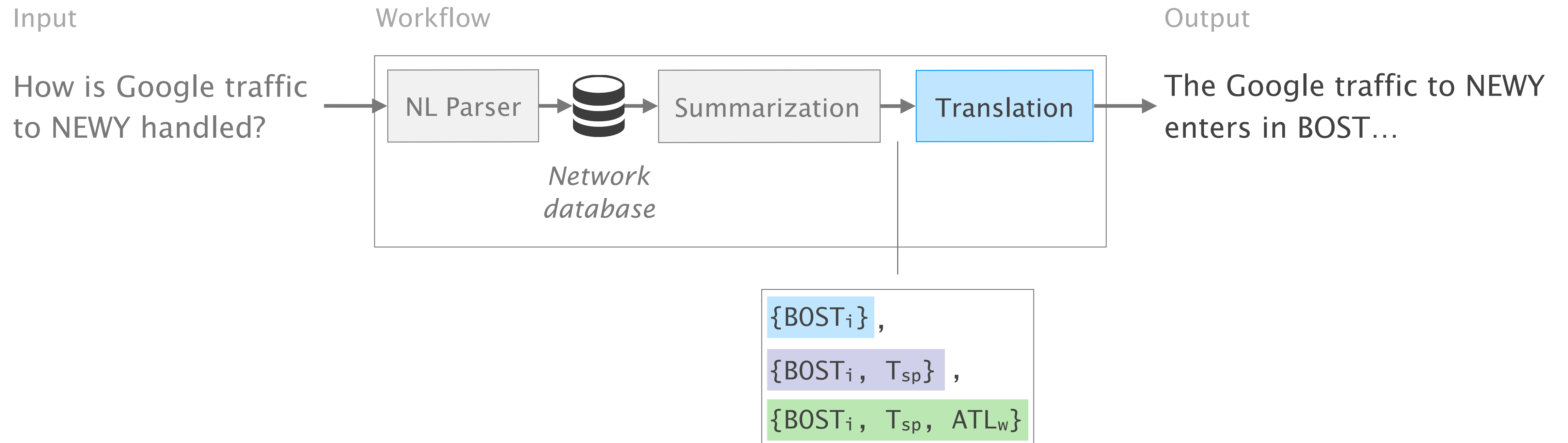
	prefix	ingress	short. path	...	avg. bw
path 1	8.8.8.0/24	BOST	T	...	98.4 Mbps
path 2	8.8.4.0/24	BOST	T	...	0.4 Mbps
path 3	66.102.0.0/20	BOST	F	...	25.0 Mbps
path 4	35.184.0.0/19	HOUS	F	...	25.0 Mbps
...
path n	35.184.0.0/19	BOST	T	...	1.0 Mbps

Output

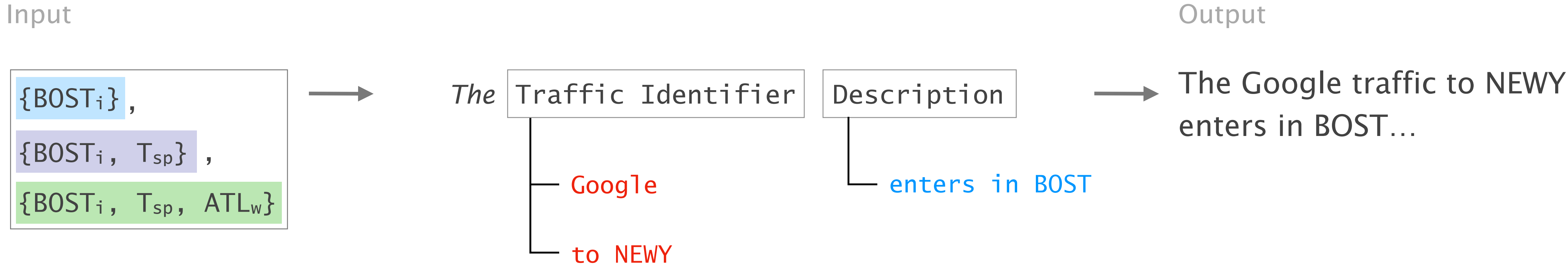


{BOST _{i} }	,
{BOST _{i} , T _{sp} }	,
{BOST _{i} , T _{sp} , ATL _{w} }	

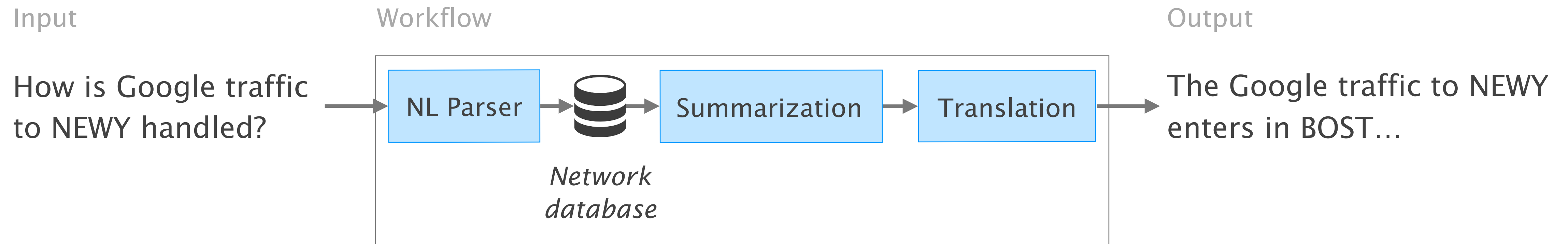
Path specifications are translated back to natural language



The translation uses templates to obtain natural language from path specifications



Net2Text has four stages:
parsing, data retrieval, summarization, translation



Net2Text

- 1 Summarization
from question to succinct answer
- 2 Scaling
summarizing fast
- 3 Performance & operator interviews
summaries within a few seconds

Net2Text

- 1 Summarization
from question to succinct answer
- 2 Scaling
summarizing fast
- 3 Performance & operator interviews
summaries within a few seconds

Finding a summary of the
network-wide forwarding state is simple

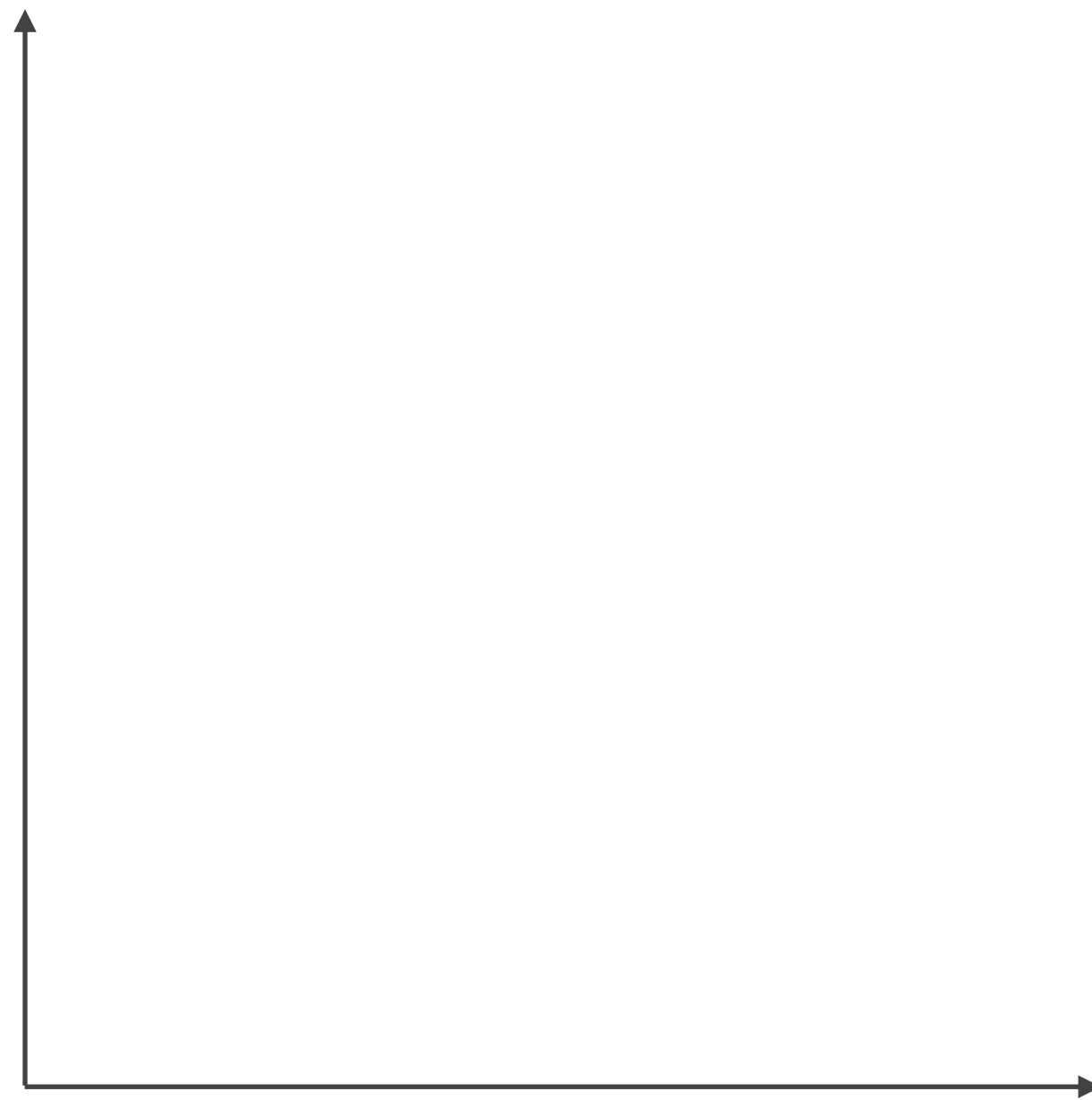
Traffic is being forwarded.

Finding a summary of the
network-wide forwarding state is simple

Traffic from LOSA to 35.184.0.0/19,
which is owned by Google,
is leaving the network in CHIC
and takes the path
SUNV, DENV, KSCY, INDI to CHIC.

Coverage

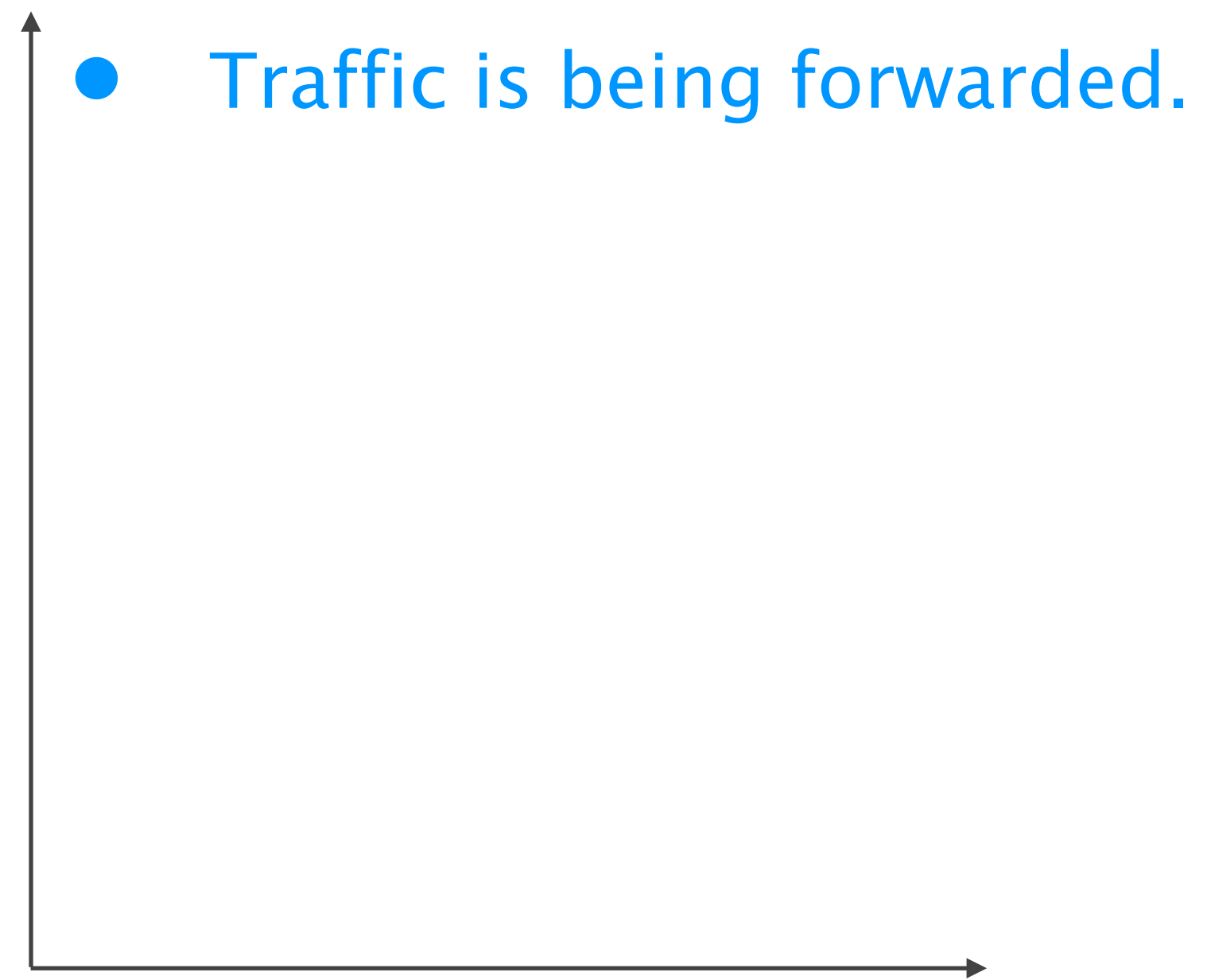
amount of data
described by the summary



Explainability

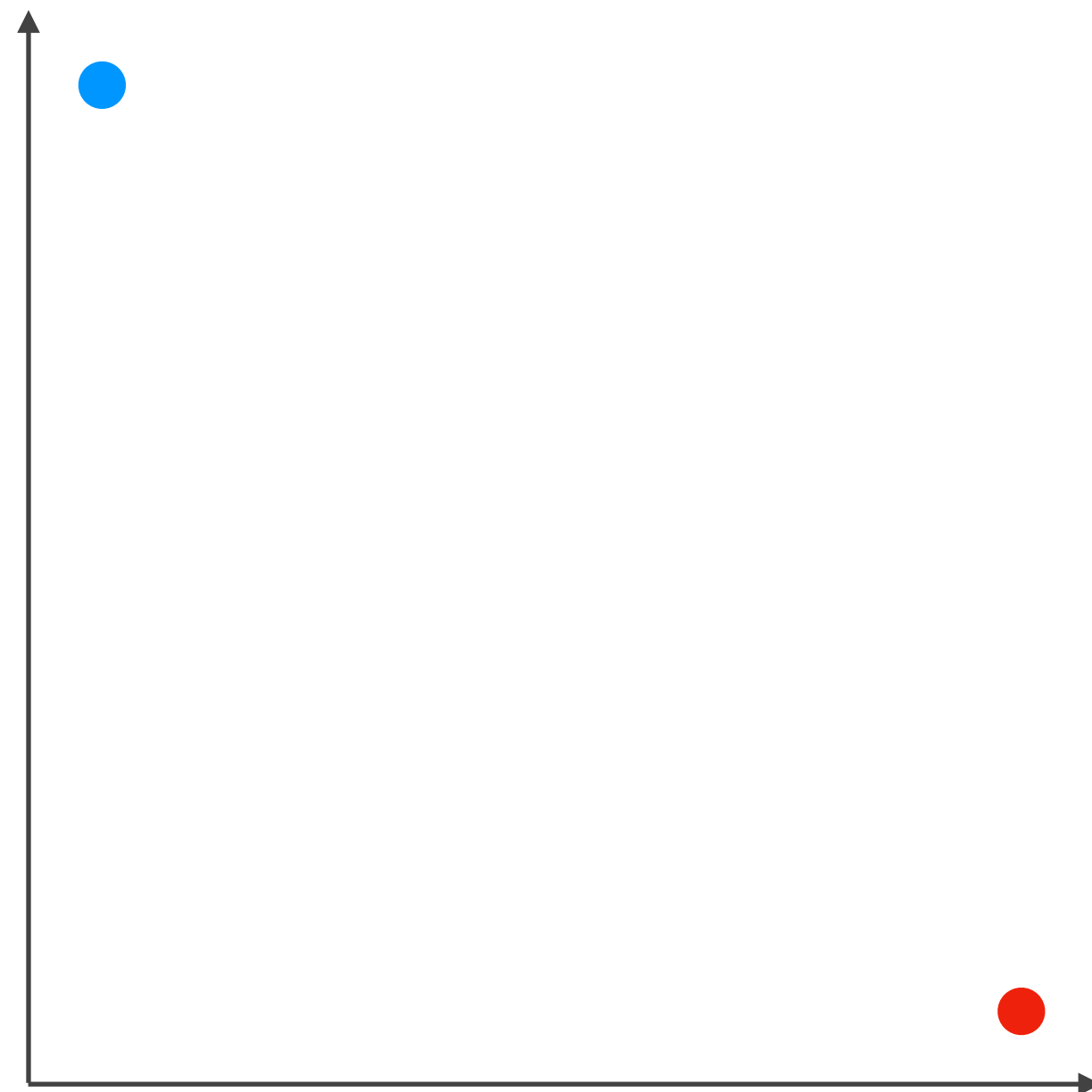
amount of detail
provided by the summary

Coverage



Explainability

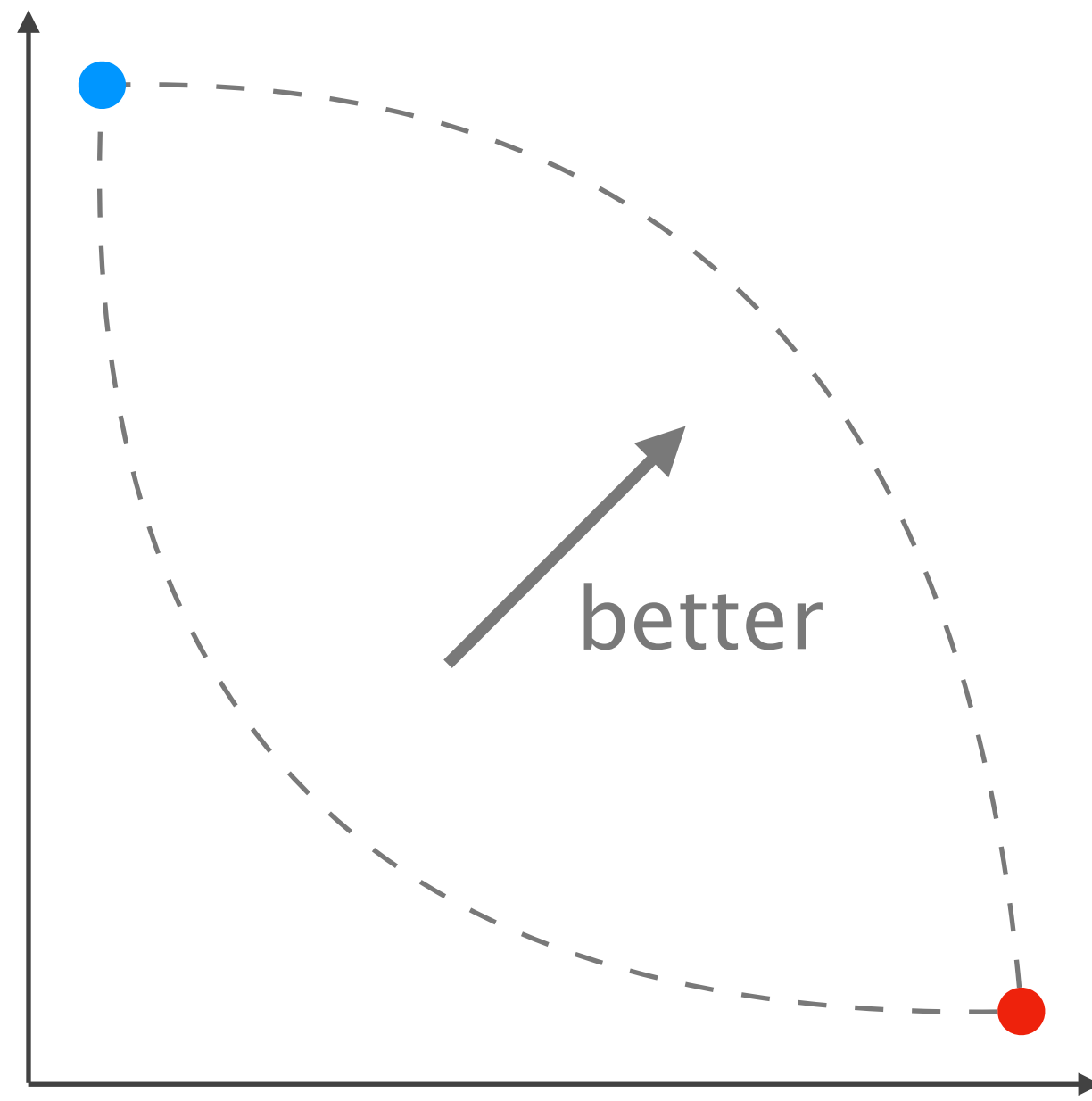
Coverage



Traffic from LOSA to 35.184.0.0/19,
which is owned by Google, ...

Explainability

Coverage



Explainability

Summarization is an optimization problem
guided by the summary score

Score

Weighted sum of the amount of traffic covered by each path specification in the summary.

Summarization is an optimization problem
guided by the summary score

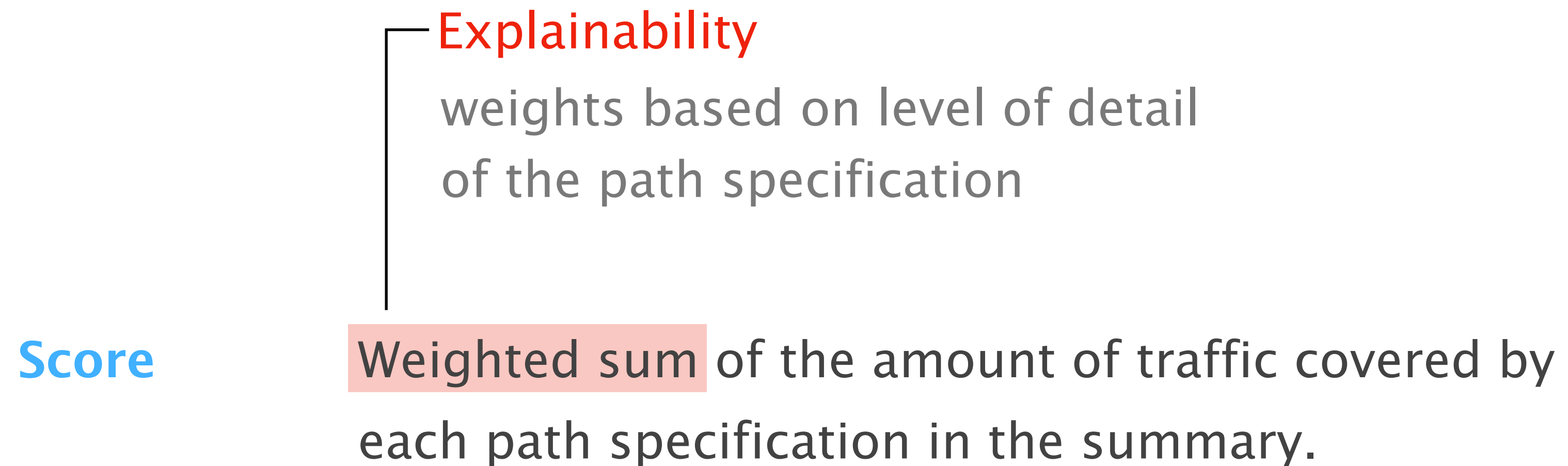
Score

Weighted sum of the amount of traffic covered by
each path specification in the summary.

Coverage



Summarization is an optimization problem
guided by the summary score



Summarization is an optimization problem
guided by the summary score

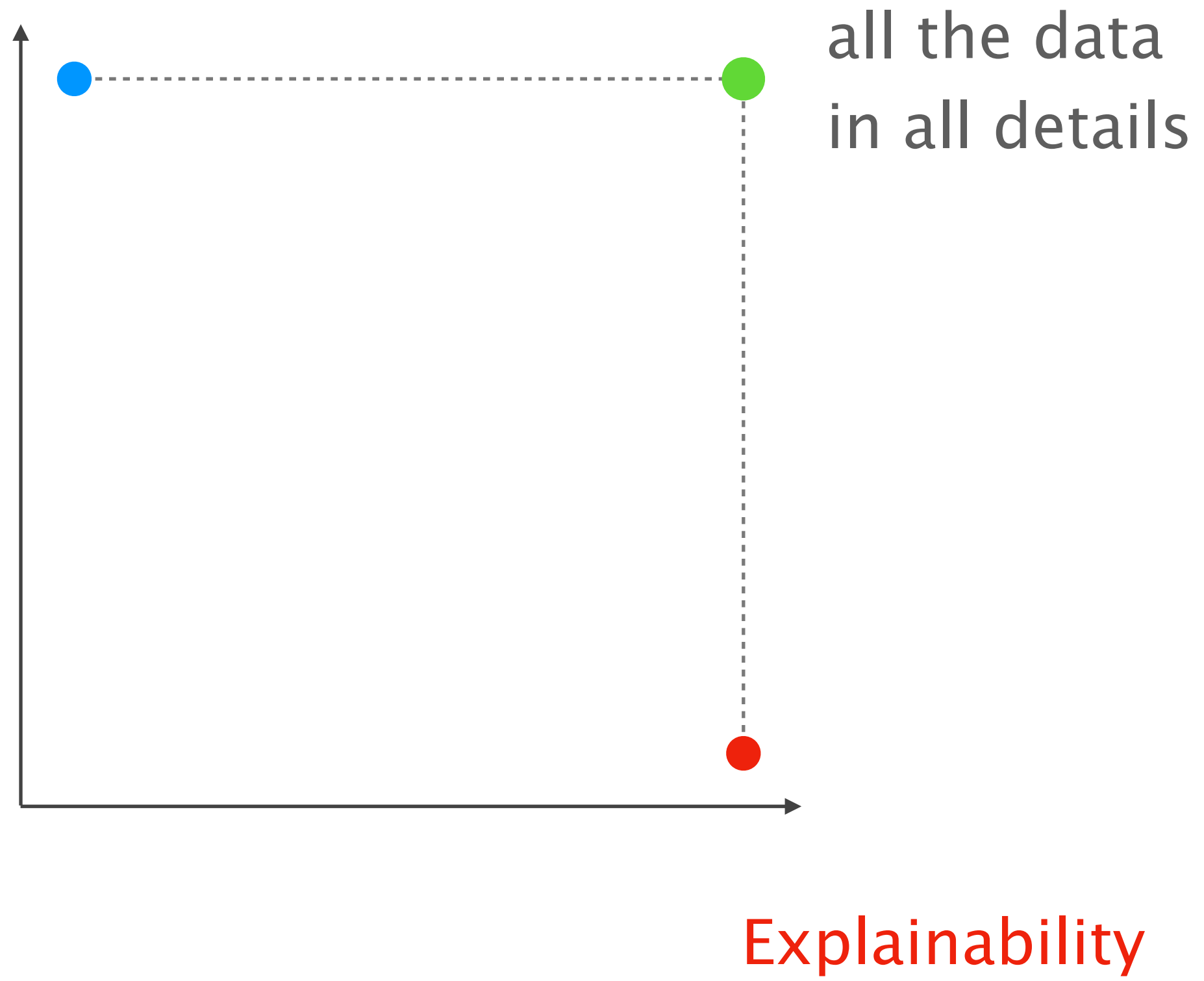
Score

Weighted sum of the amount of traffic covered by each path specification in the summary.

Goal

Find path specifications that maximize the score.

Coverage



Summarization is an optimization problem
guided by the summary score and a size restriction

Score

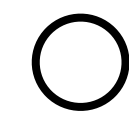
Weighted sum of the amount of traffic covered by each path specification in the summary.

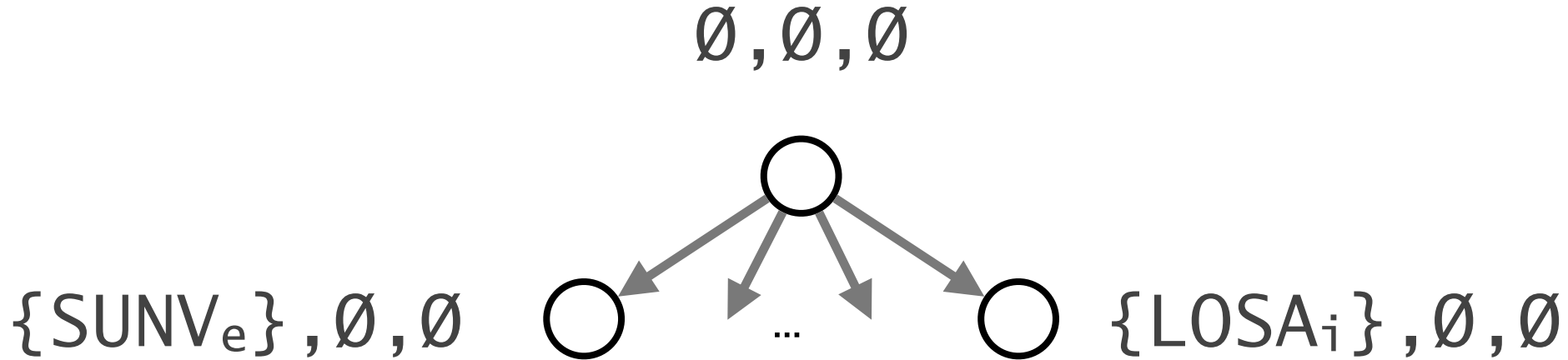
Goal

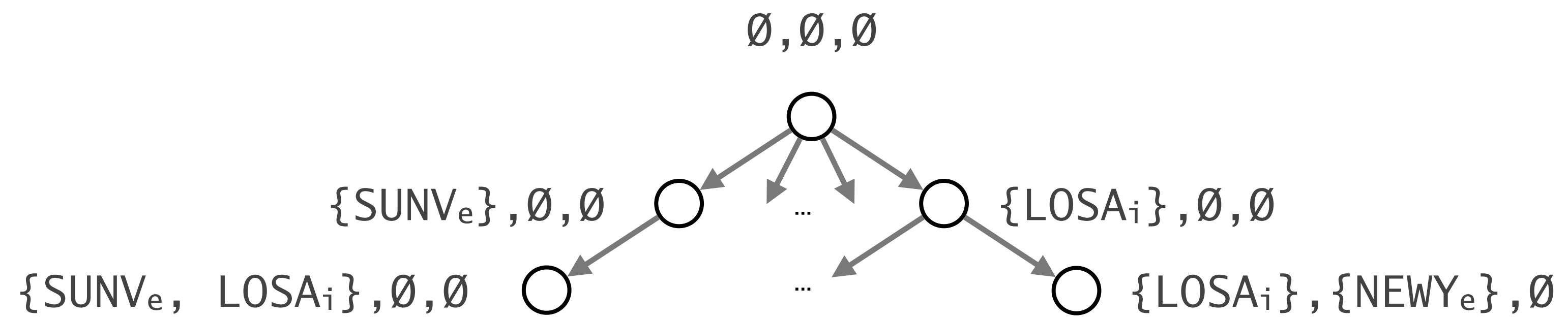
Find k path specifications each of size at most t that maximize the score.

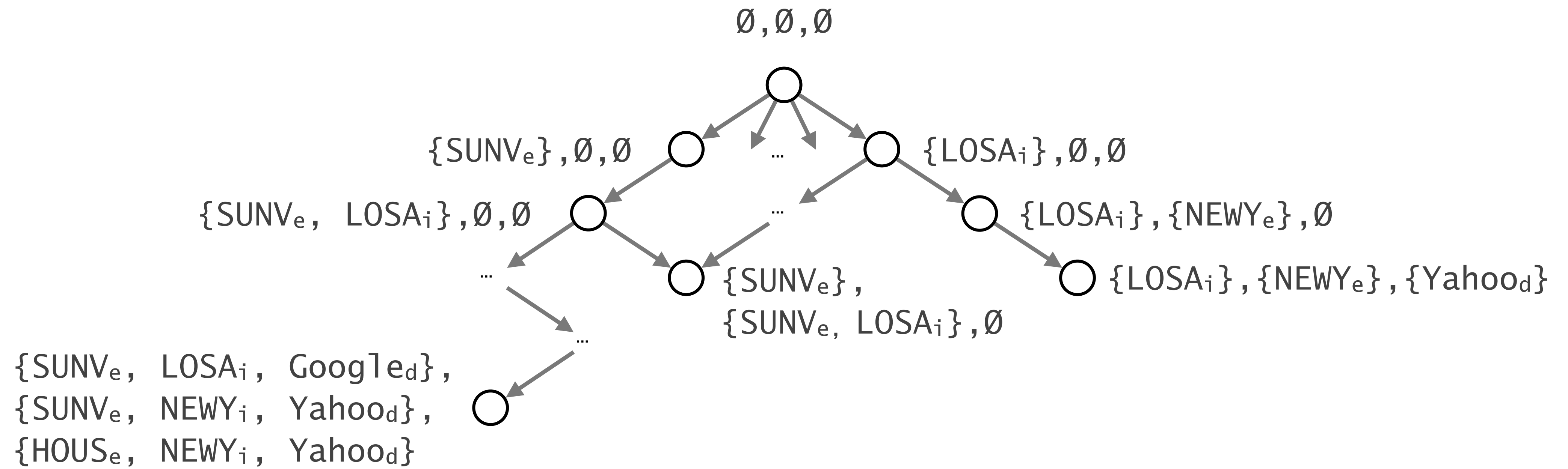
$$k = 3, t = 3$$

$$\emptyset, \emptyset, \emptyset$$

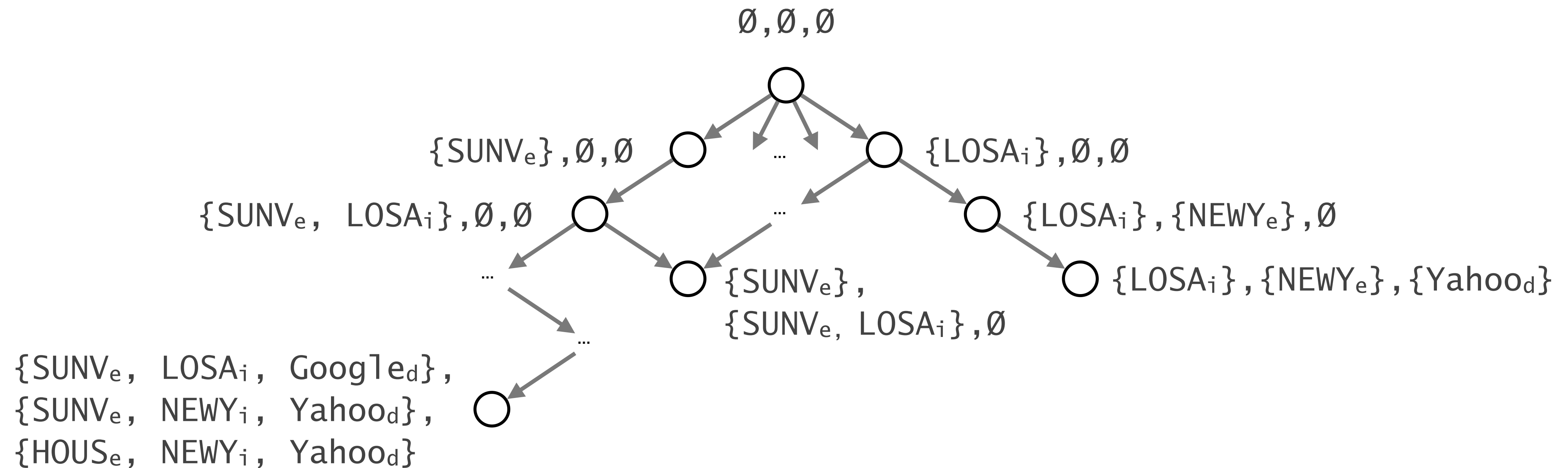








The search space is exponential
 in the number of path specifications and feature values



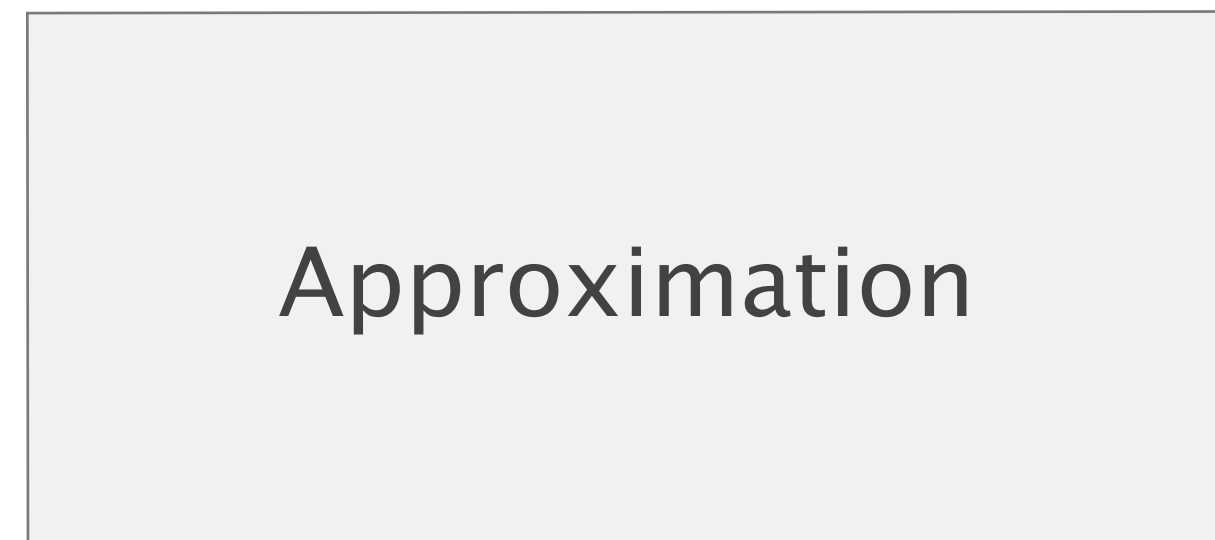
Due to the size of the search space,
exhaustive exploration is not feasible

Net2Text

- 1 Summarization
from question to succinct answer
- 2 **Scaling**
summarizing quickly
- 3 Performance & operator interviews
summaries within a few seconds

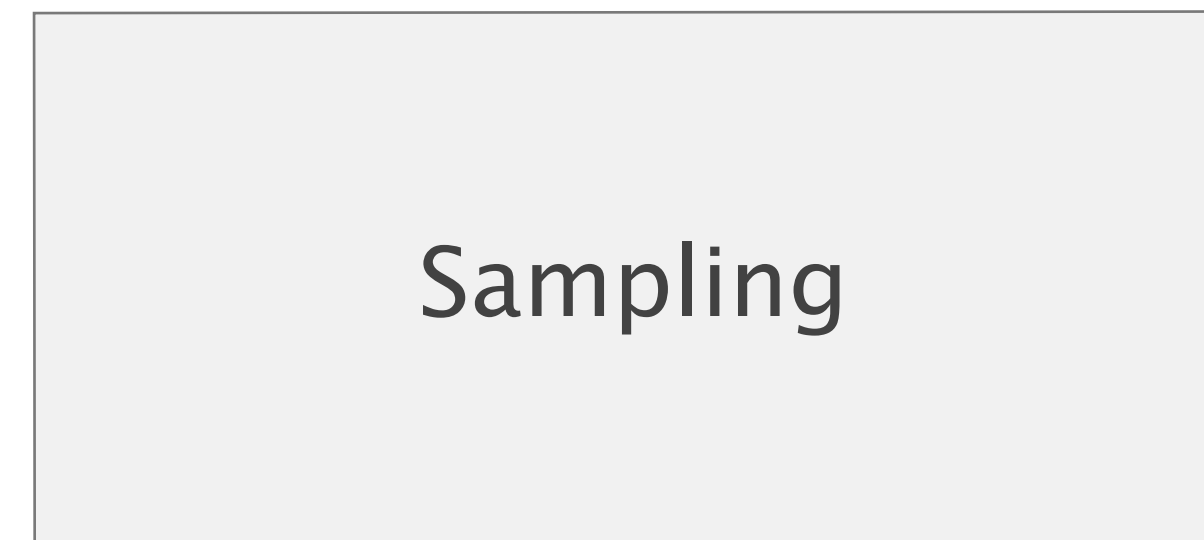
Net2Text relies on two optimizations

Optimization 1



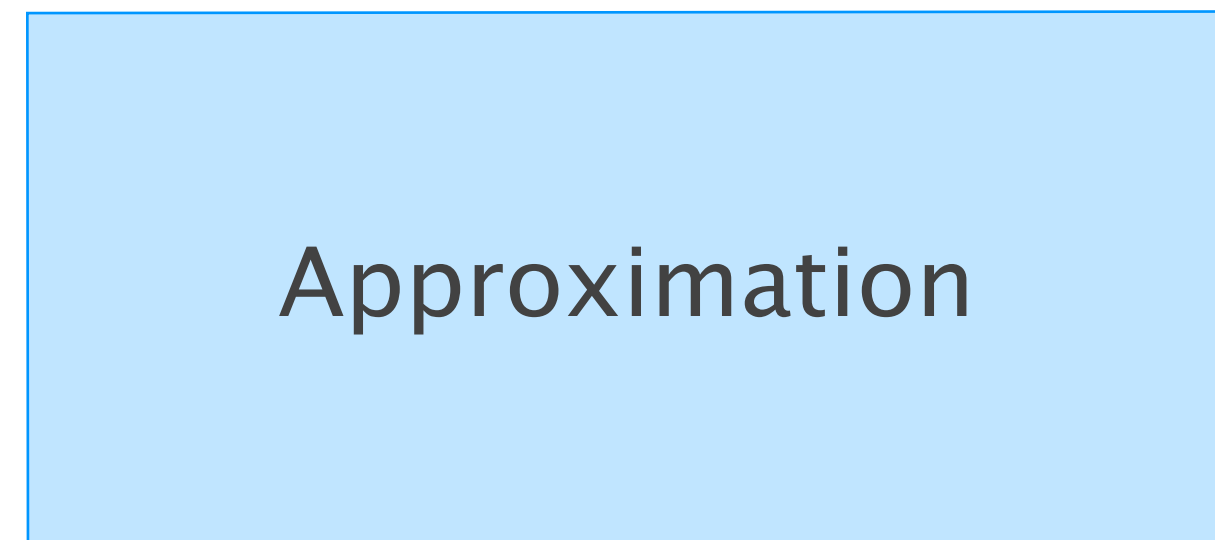
Reduce the search space

Optimization 2

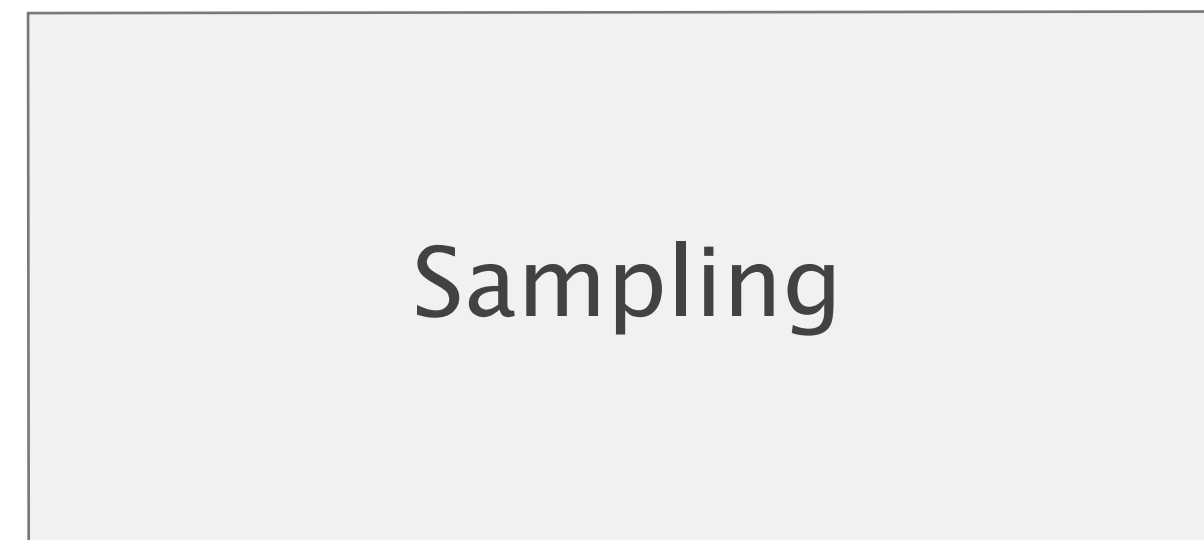


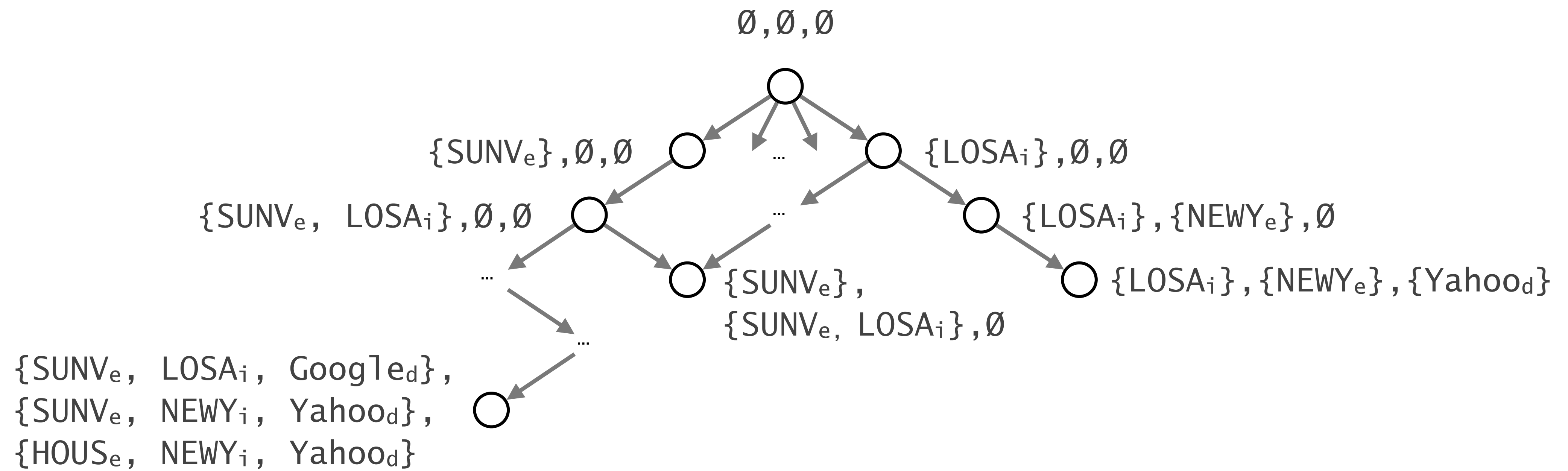
Reduce the input data

Optimization 1

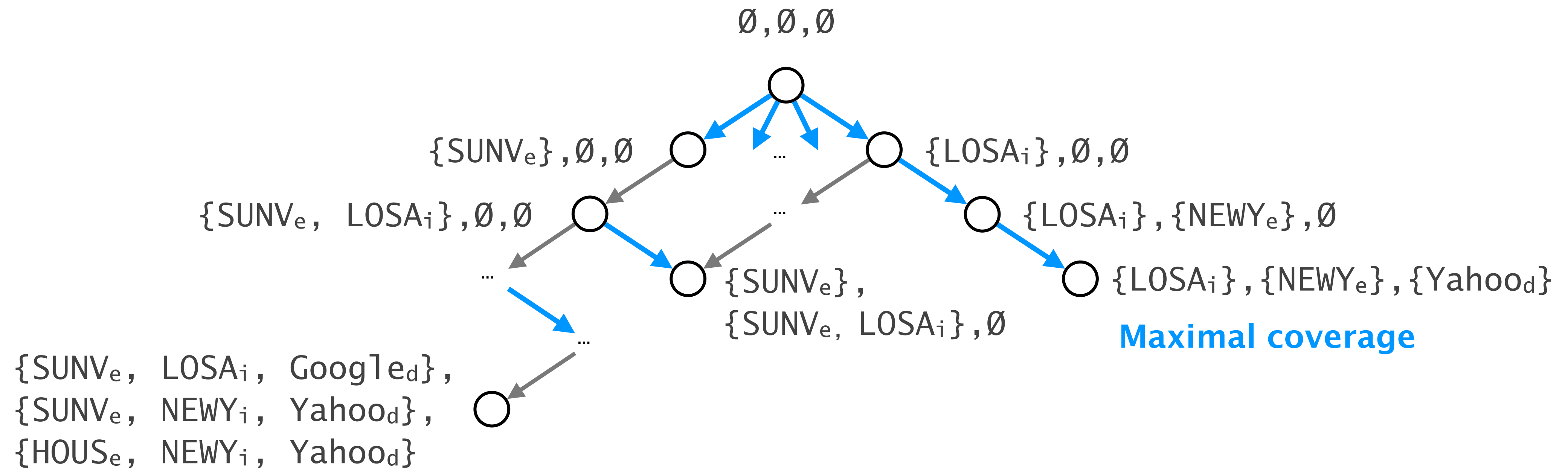


Reduce the search space

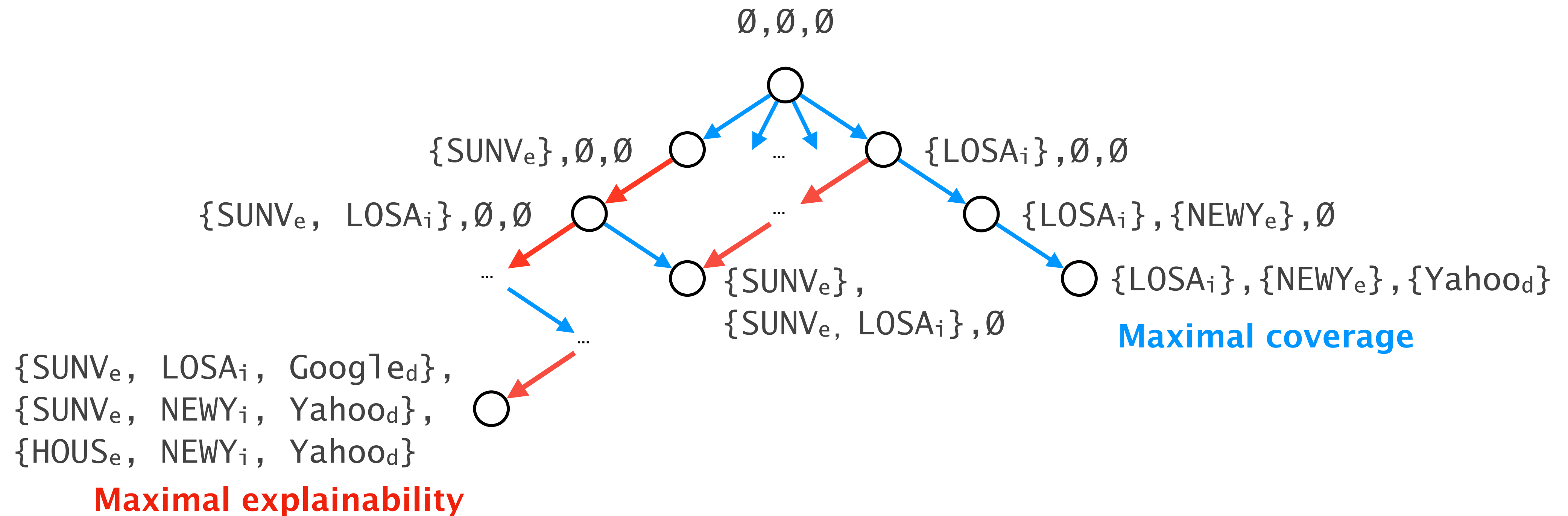




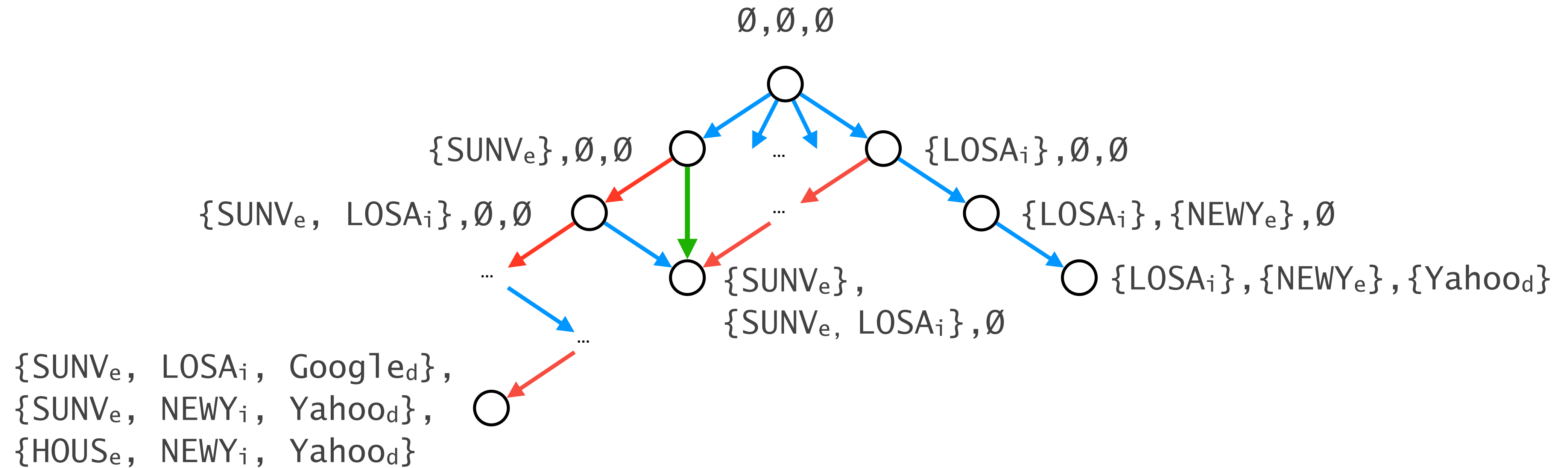
The search space contains two types of edges:
blue edges that increase coverage



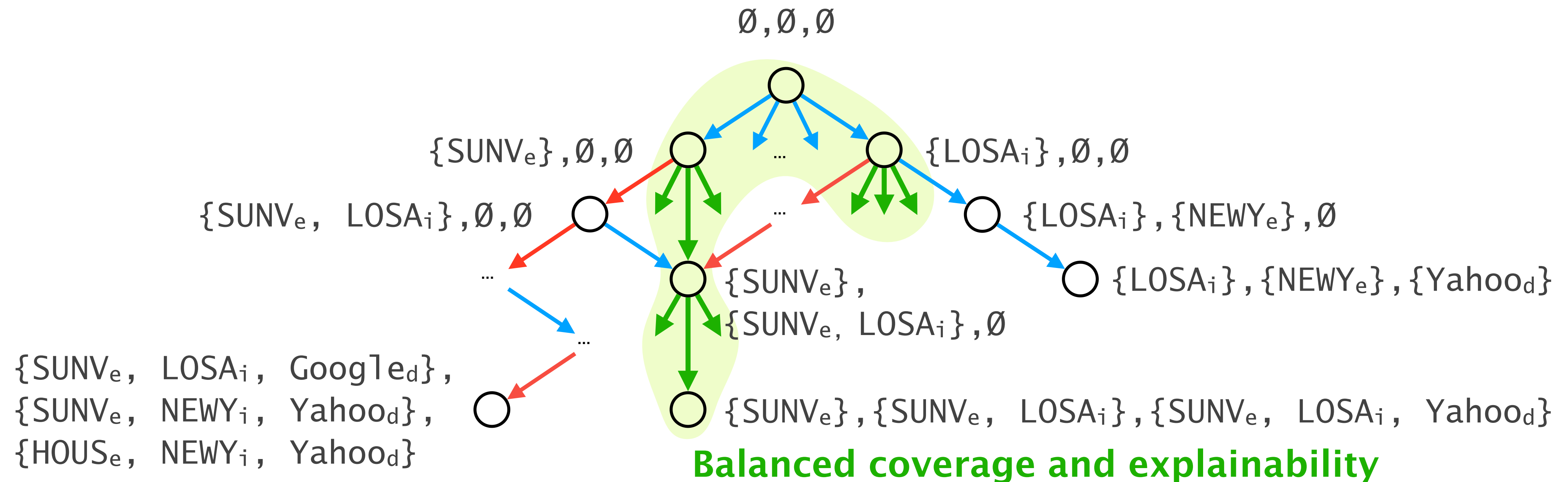
The search space contains two types of edges:
red edges that increase explainability



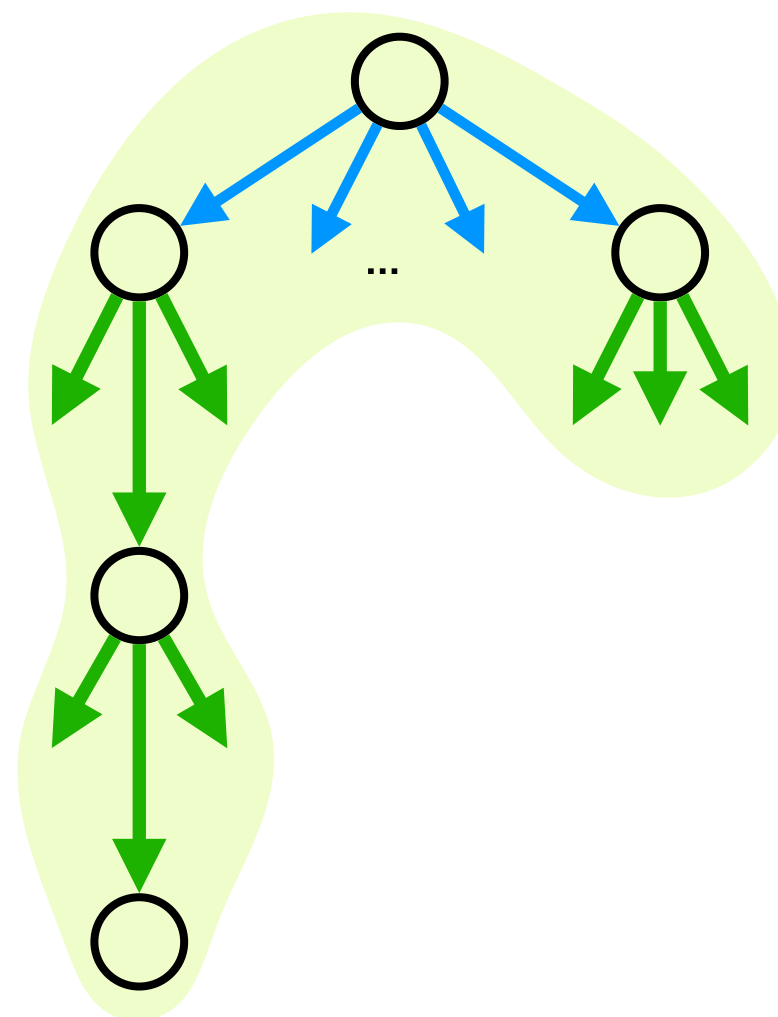
Net2Text reduces the search space
to solutions that balance coverage and explainability



Net2Text reduces the search space
to solutions that balance coverage and explainability



Net2Text reduces the search space to solutions that balance coverage and explainability



Graph has a monotonicity property

The child's score is always higher

Net2Text greedily explores the graph

Always follow most promising path

Solution is not far off from best solution

Guaranteed lower bound on the score

Approximation

Optimization 2

Sampling

Reduce the input data

Network traffic is highly skewed across multiple levels

Level 1

Traffic distribution

Few destinations carry most of the traffic

Level 2

Routing and network topology

Repetitive forwarding patterns

Insight

Network traffic is repetitive and redundant

Net2Text uses redundancy in the data
to speed up summarization by sampling

Problem Net2Text iterates over all entries at least once

Insight Summary is resilient to loss of redundant information

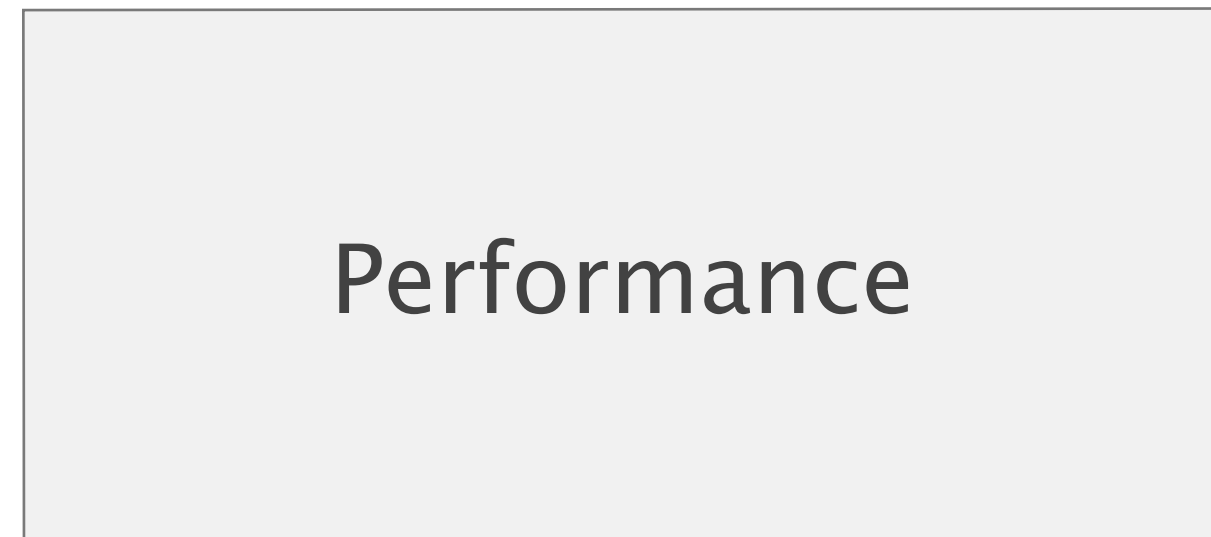
Solution Reduce input data by sampling

Net2Text

- 1 Summarization
from question to succinct answer
- 2 Scaling
summarizing fast
- 3 Performance & operator interviews
summaries within a few seconds

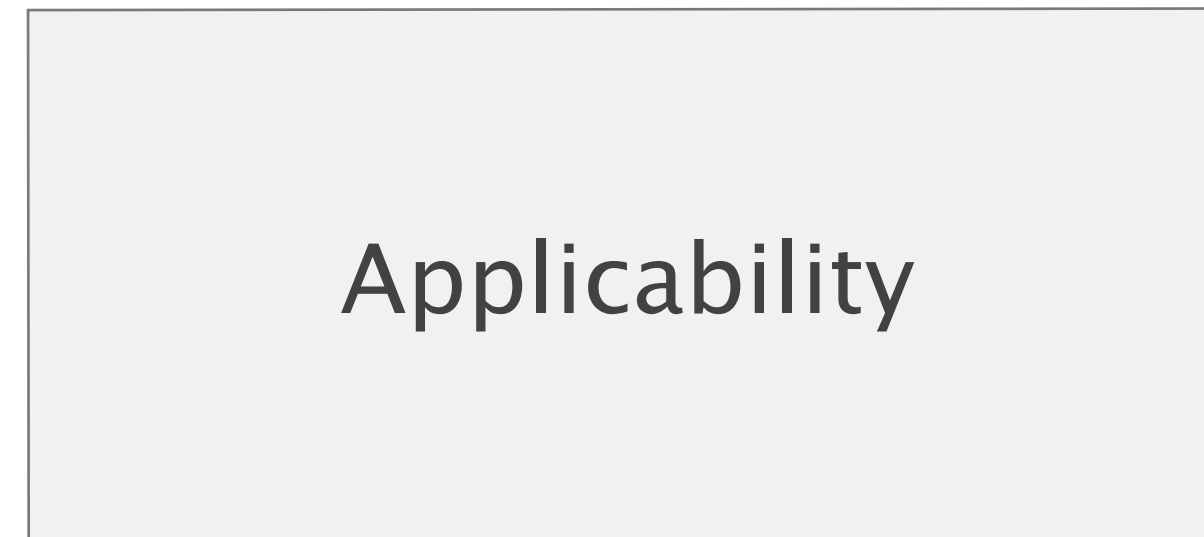
Net2Text needs to be quick and applicable

Aspect 1



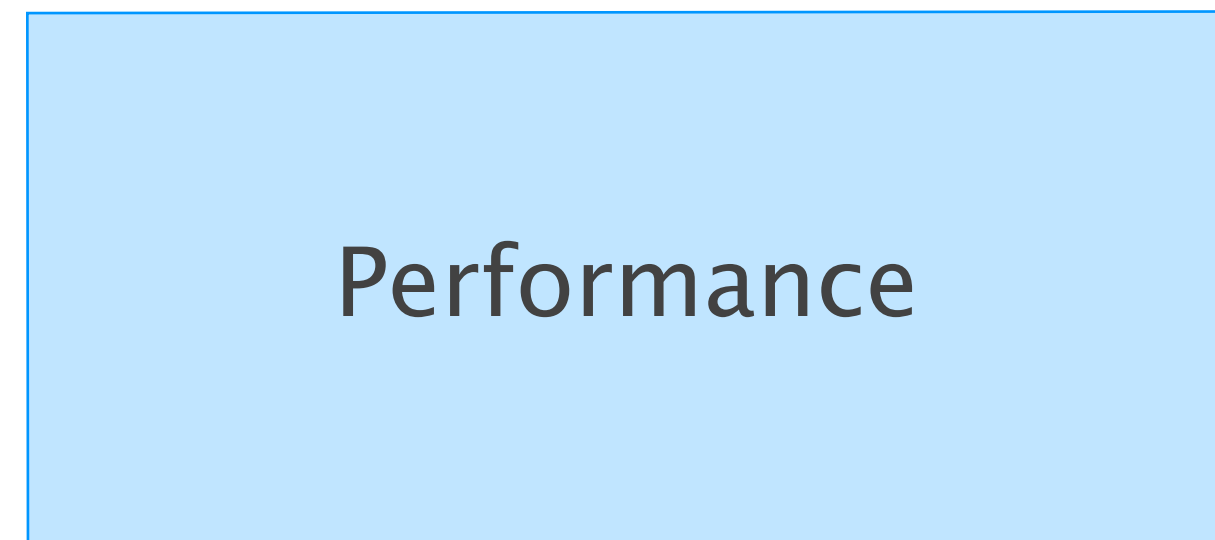
End-to-end timing

Aspect 2

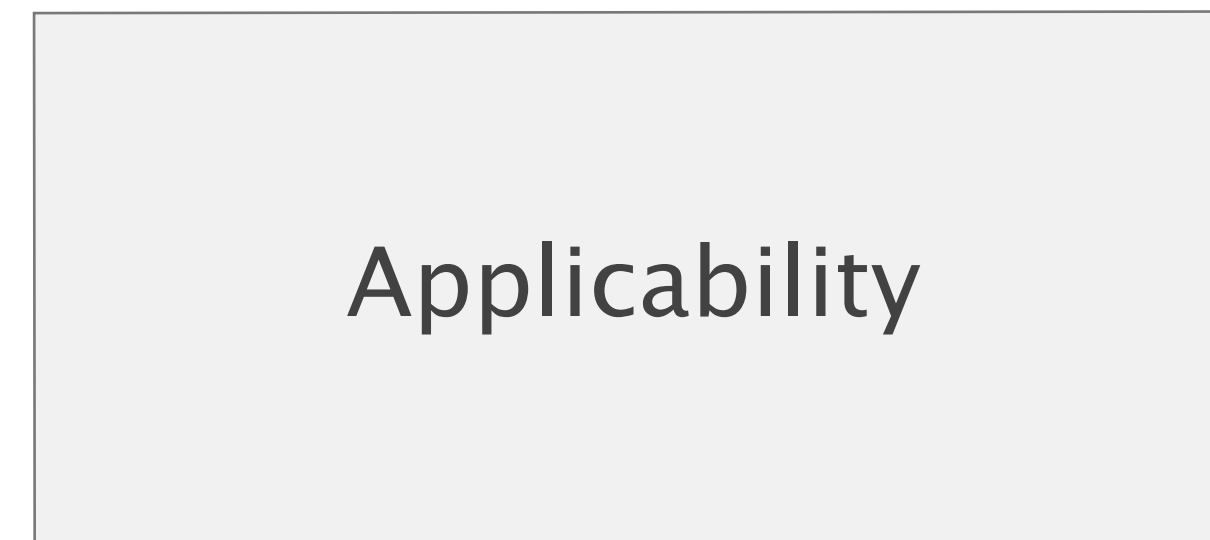


Operator interviews

Aspect 1



End-to-end timing



Pushing Net2Text to its limits by summarizing the entire forwarding state

Question How is traffic being forwarded?

Setup ATT North America from Topology Zoo
25 nodes, 10 of them egresses

Full routing tables (~650k prefixes)

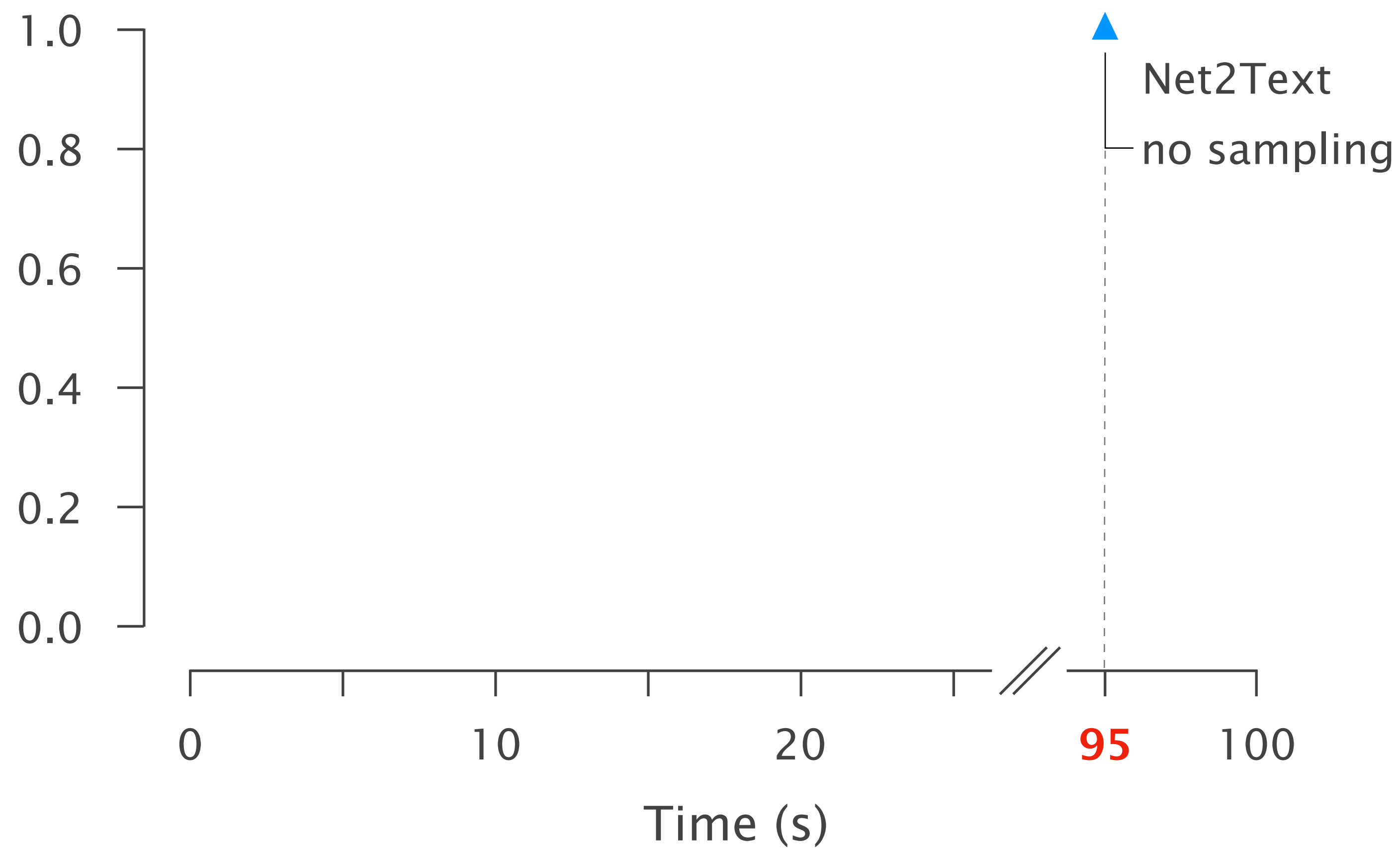
Four features ingress

egress

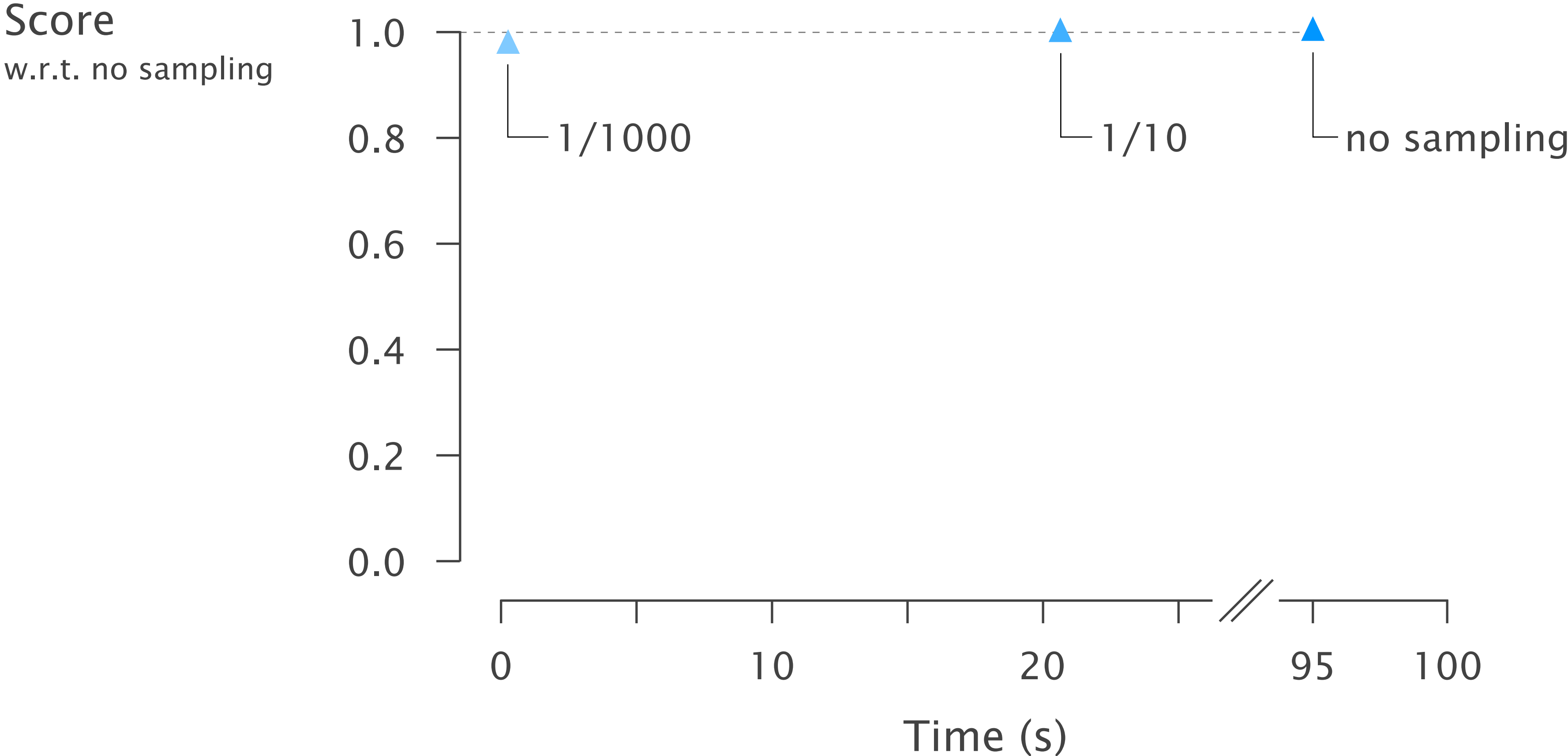
destination

shortest path

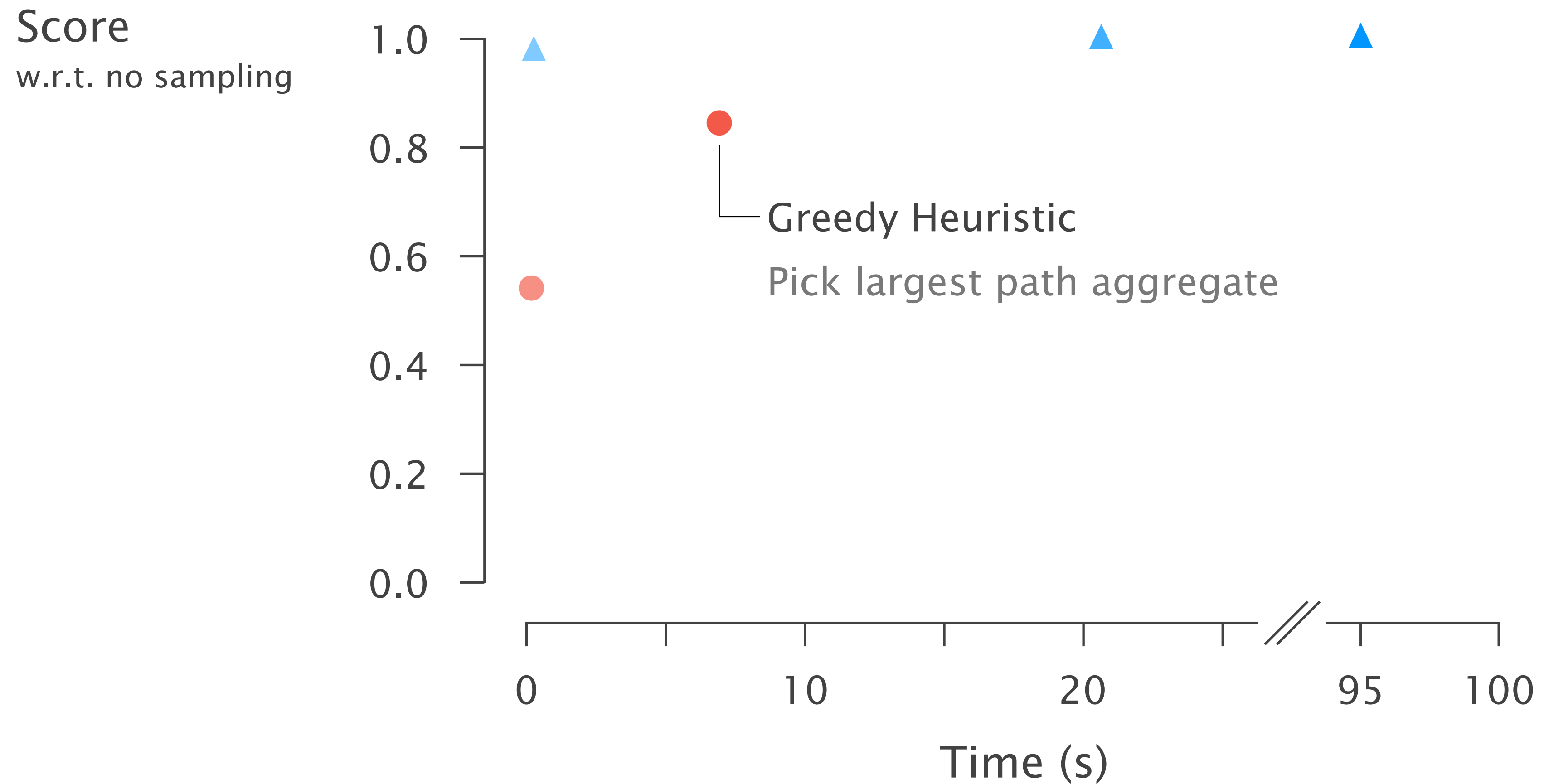
Score
w.r.t. no sampling



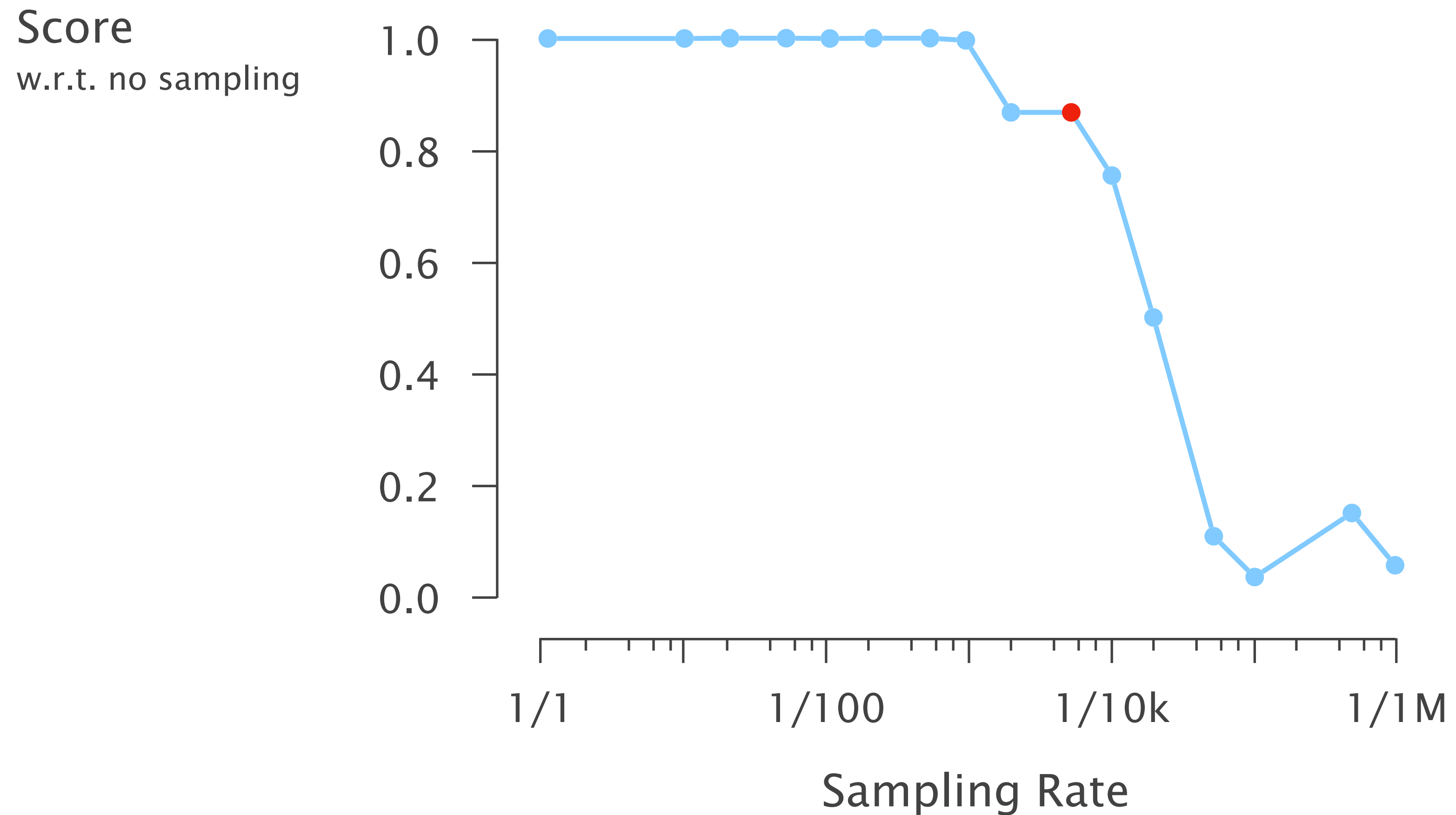
Net2Text finds good summaries within seconds thanks to sampling



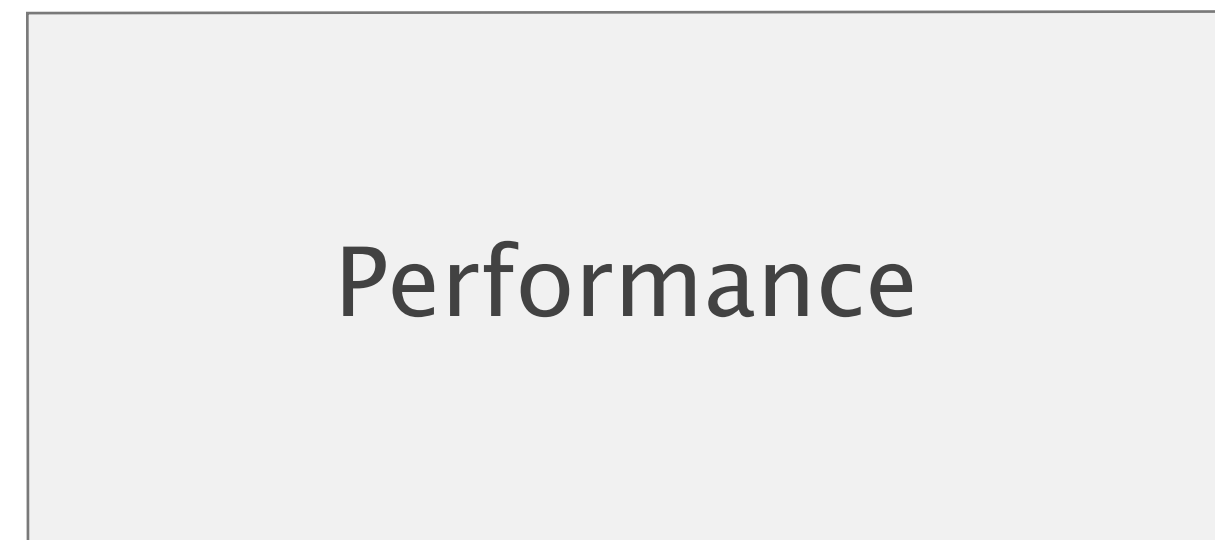
Baseline is slightly faster than Net2Text,
but not as resilient to sampling



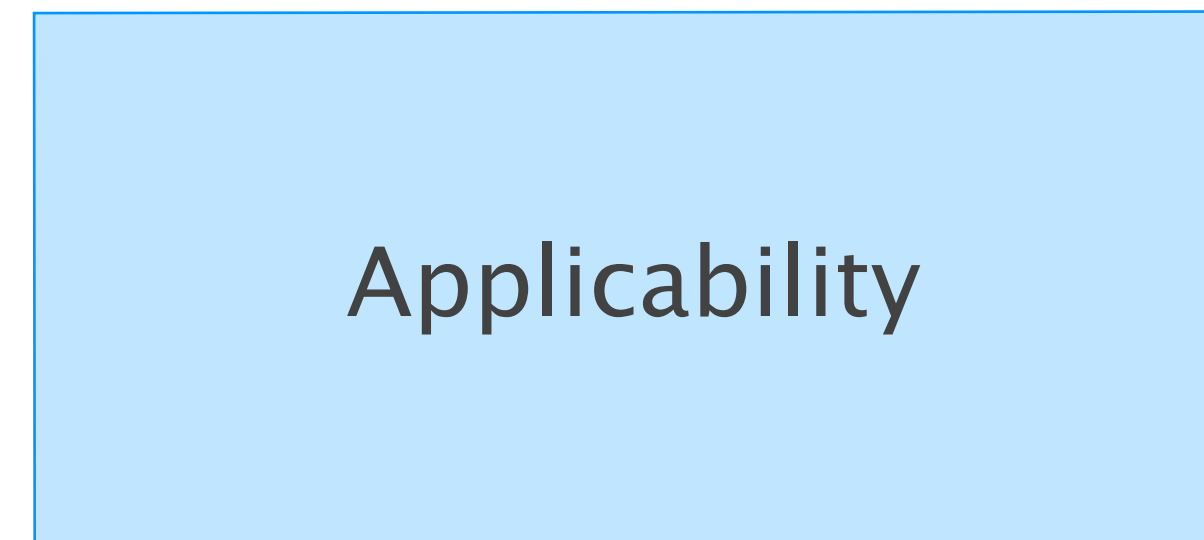
Only sampling higher than 1/5k
has a significant effect on the score



Net2Text needs to be quick and applicable



Aspect 2



Operator interviews

We asked various operators about Net2Text,
they found it useful

- Assistants** Operators see value of assistants in their daily tasks
Support in all time consuming tasks
- NL I/O** NL is useful, especially for less technical people
Operators don't mind to use query languages
- Questions** Supported questions are relevant
Especially "Where is the traffic coming from?"

Net2Text assists network operators by summarizing the forwarding state



Net2Text answers questions in natural language with a succinct summary in natural language



Net2Text presents a summary that balances coverage and explainability



Net2Text responds in a timely manner and the supported queries are relevant

Net **2**Text



Query-Guided Network Captioning

net2text.ethz.ch



Rüdiger Birkner



Dana Drachsler-Cohen



Martin Vechev



Laurent Vanbever