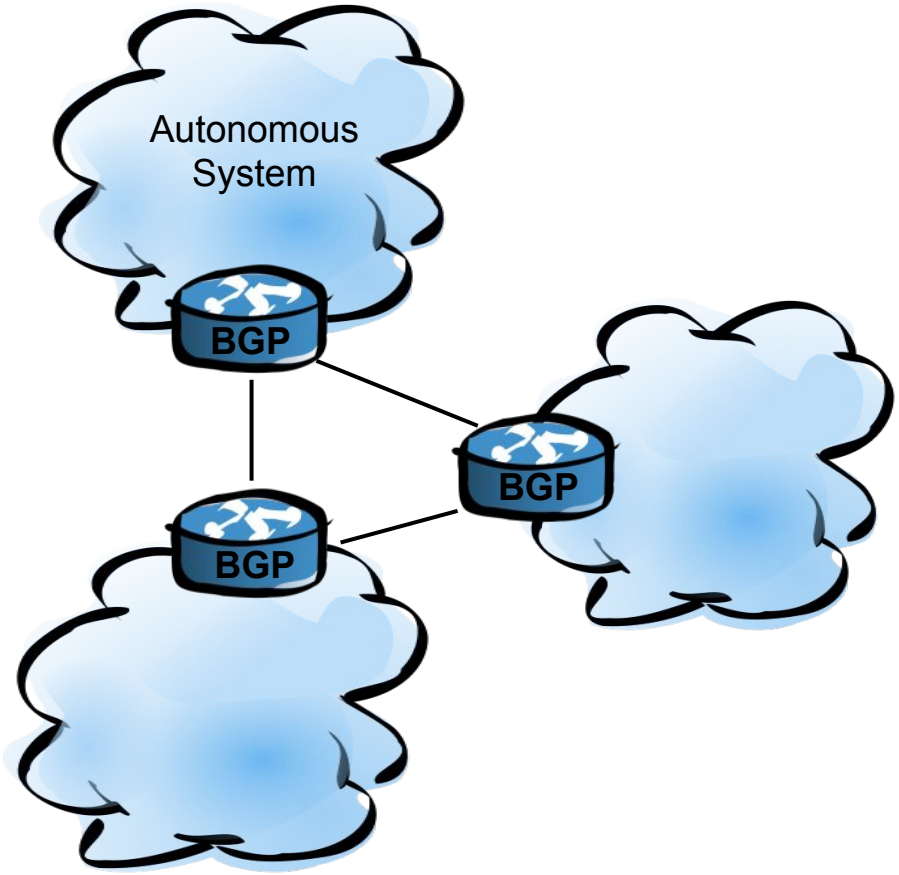
Automatic Inference of BGP Location Communities

Brivaldo A. Silva Jr Paulo Mol Osvaldo Fonseca
Ítalo Cunha Ronaldo A. Ferreira Ethan Katz-Bassett

SIGMETRICS 2022
Mumbai, India

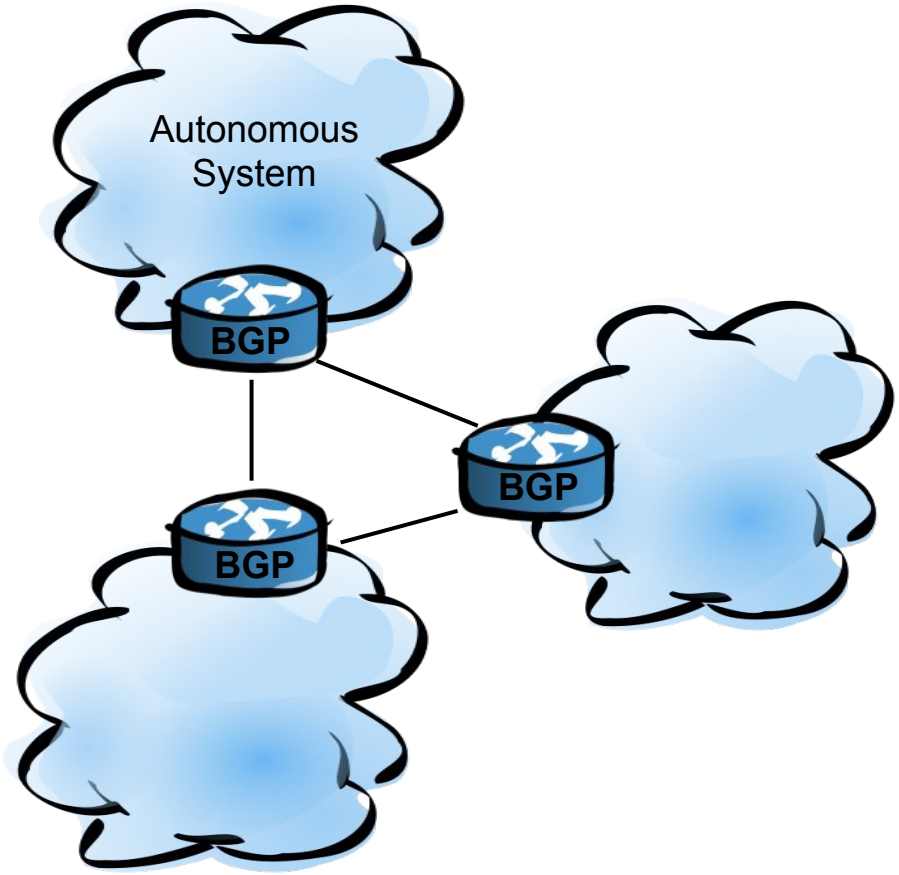
Internet Evolution

- BGP: the routing protocol used to exchange Internet routes and reachability information between autonomous systems



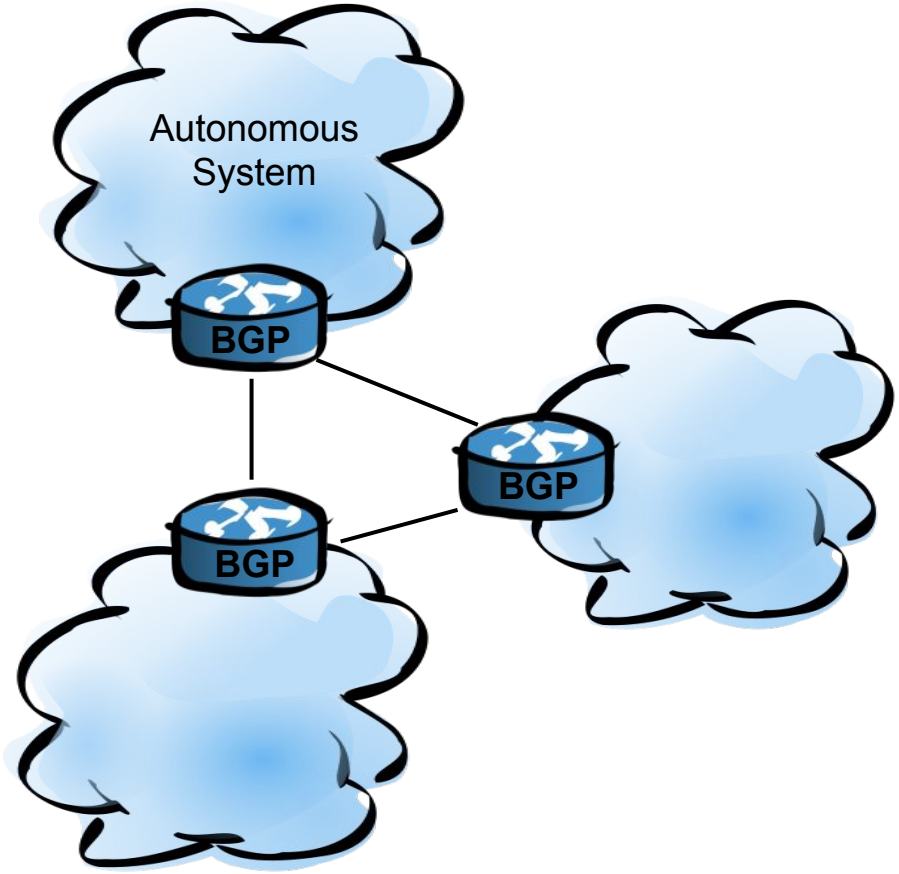
Internet Evolution

- BGP: the routing protocol used to exchange Internet routes and reachability information between autonomous systems
- BGP is an old protocol used to sustain the current need of stability, reliability, and complex policies.



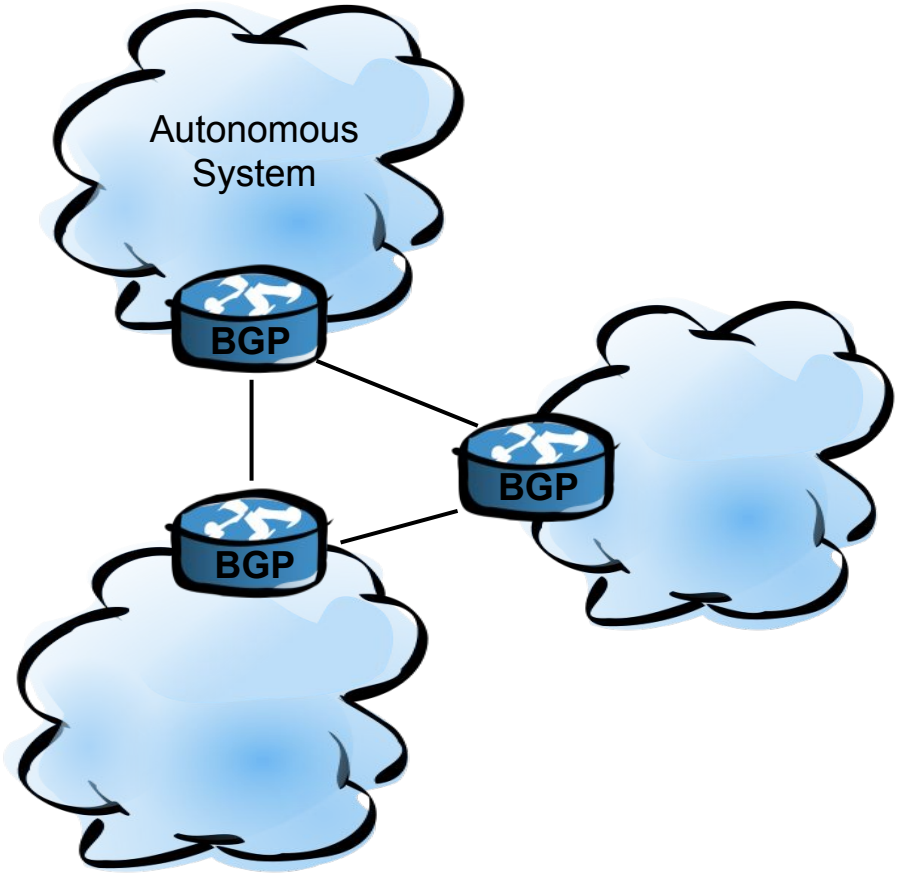
Internet Evolution

- BGP: the routing protocol used to exchange Internet routes and reachability information between autonomous systems
- BGP is an old protocol used to sustain the current need of stability, reliability, and complex policies.
- Operators explore all available options to improve the “user experience”.

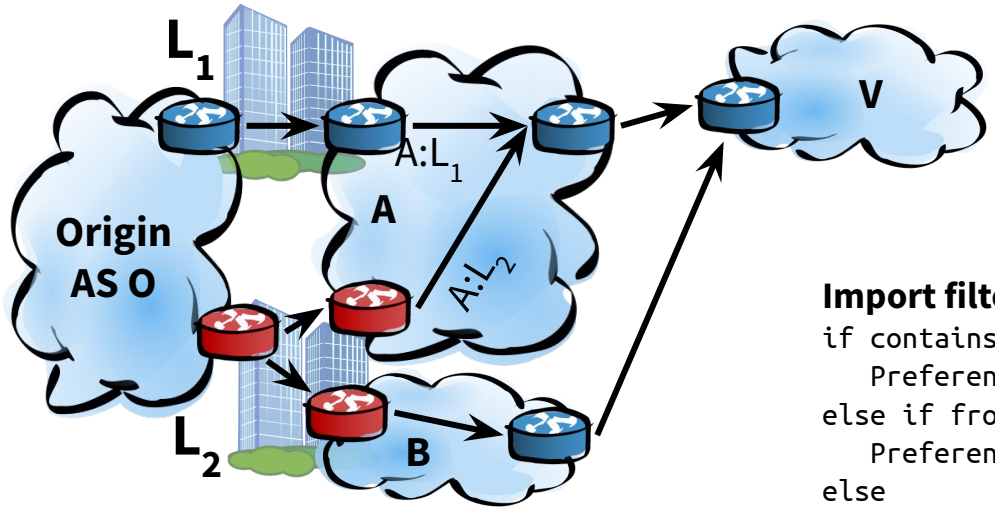


Internet Evolution

- BGP: the routing protocol used to exchange Internet routes and reachability information between autonomous systems
- BGP is an old protocol used to sustain the current need of stability, reliability, and complex policies.
- Operators explore all available options to improve the “user experience”.
- Network operators have increasingly relied on the communities attribute to instrument BGP.



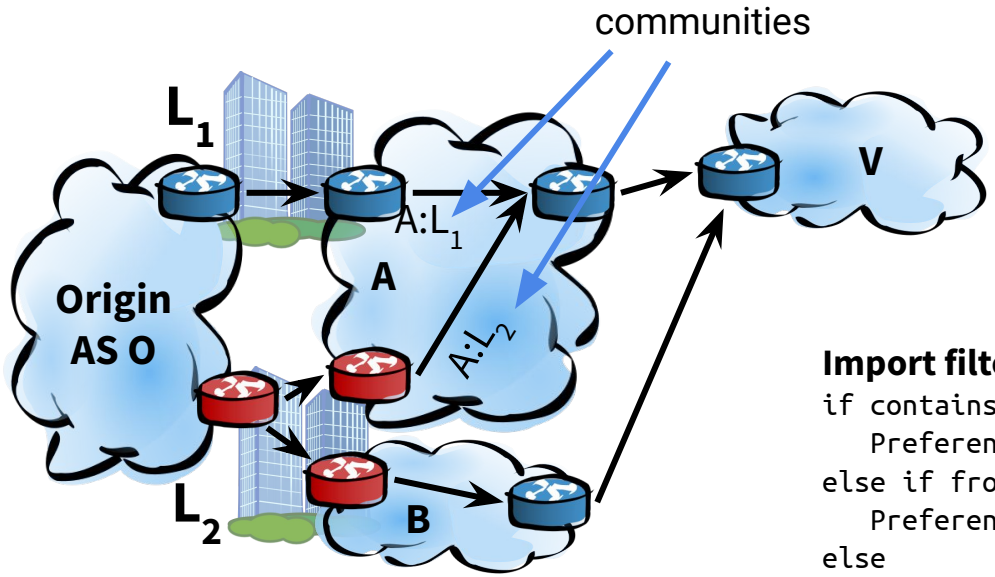
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

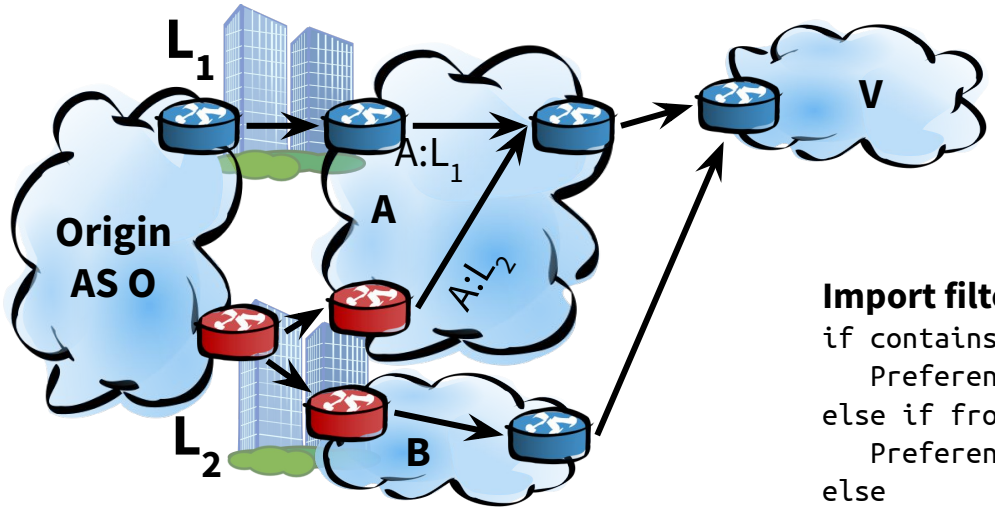
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

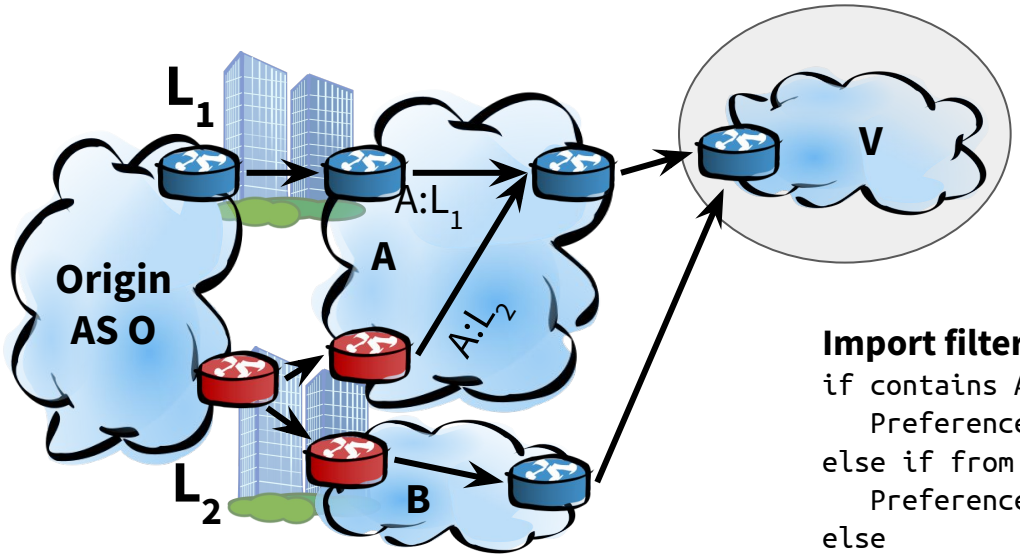
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

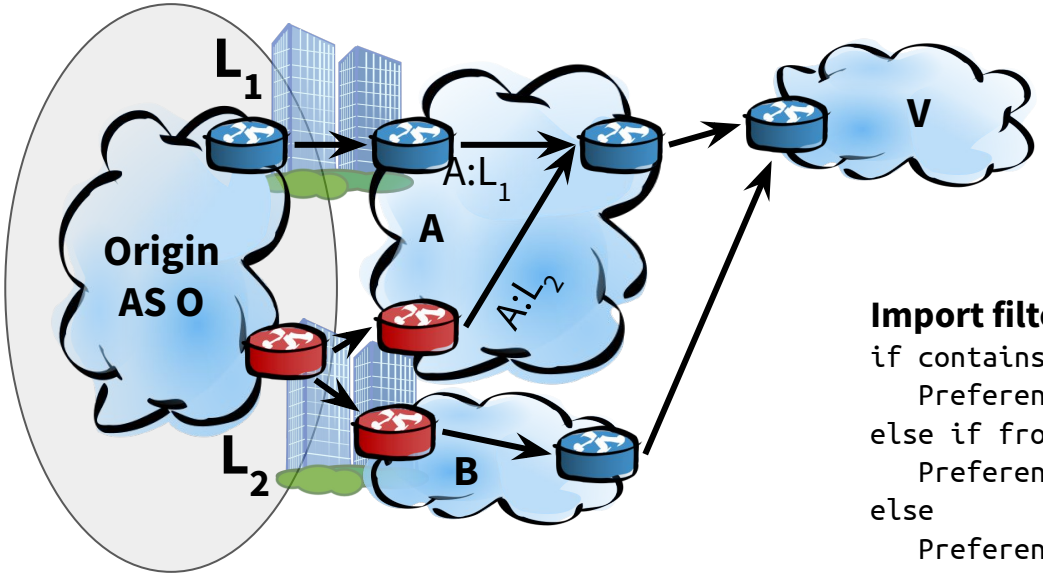
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

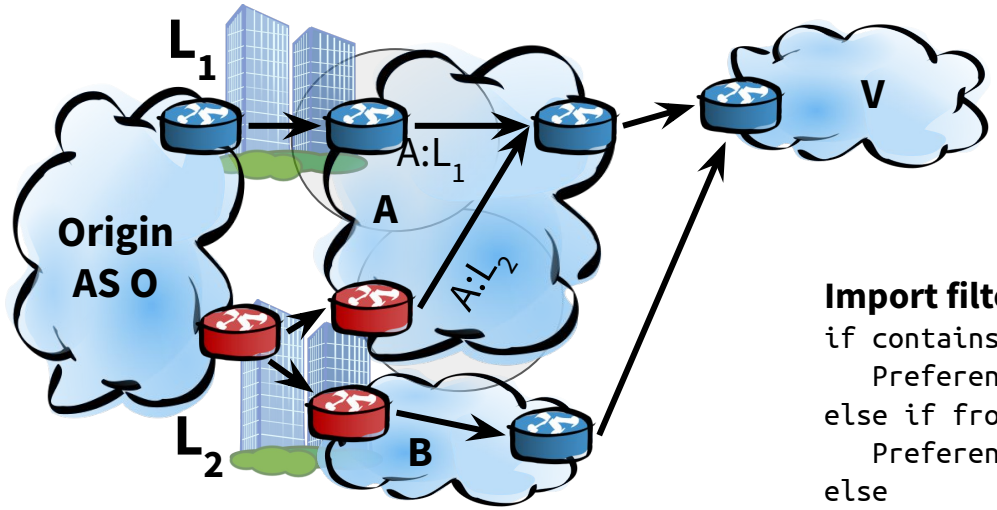
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

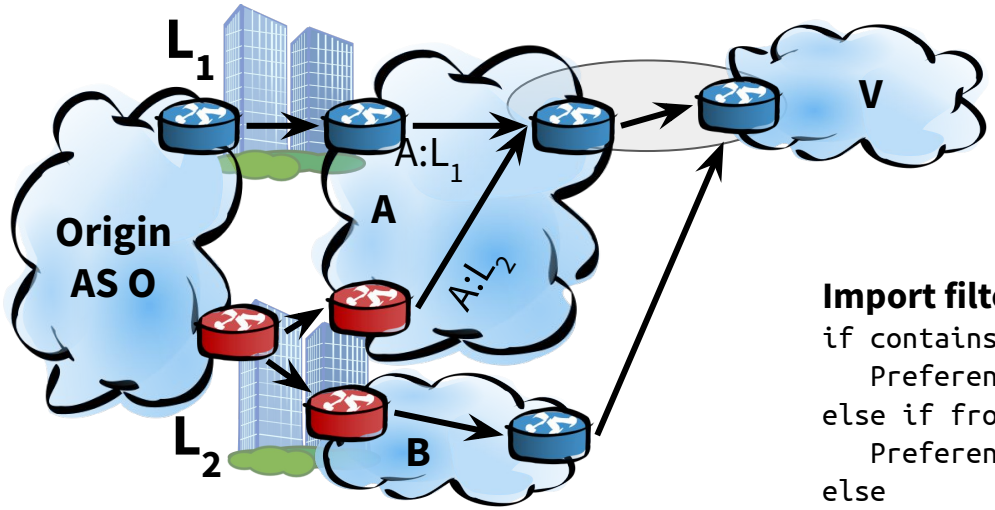
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

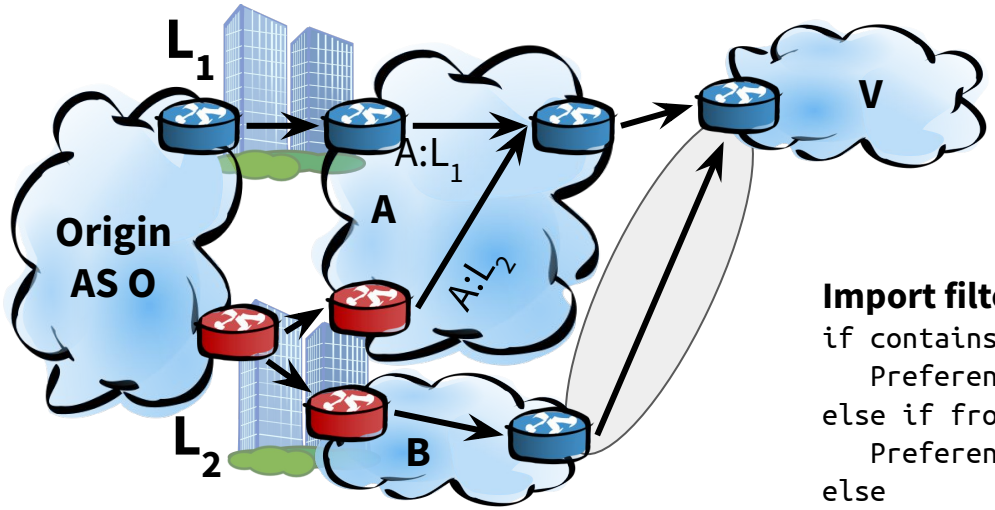
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

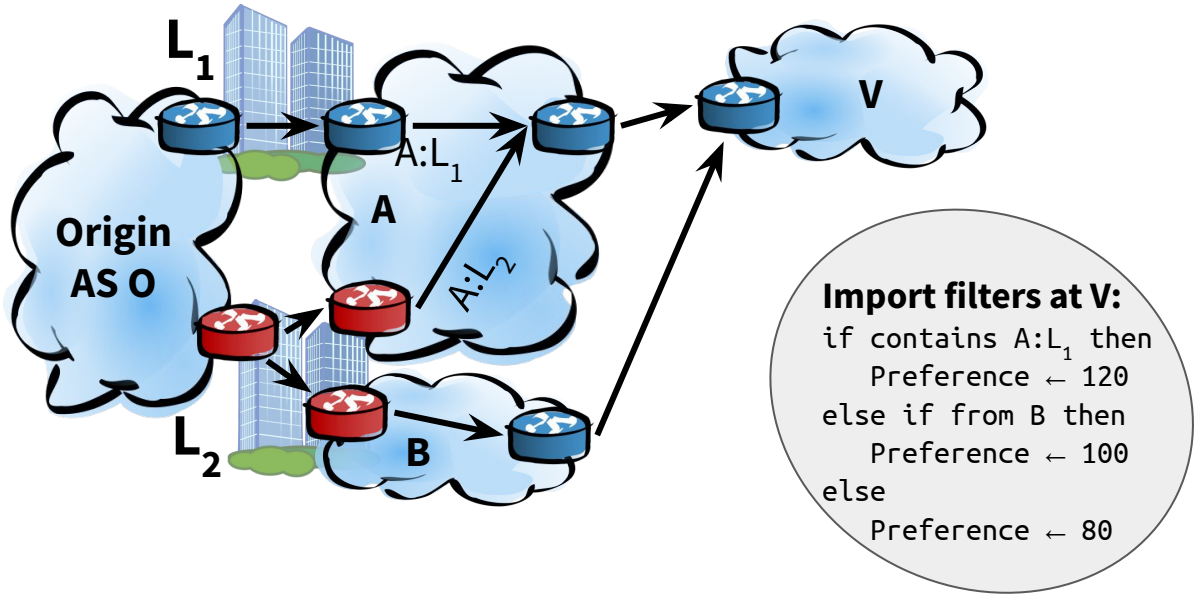
Location Communities



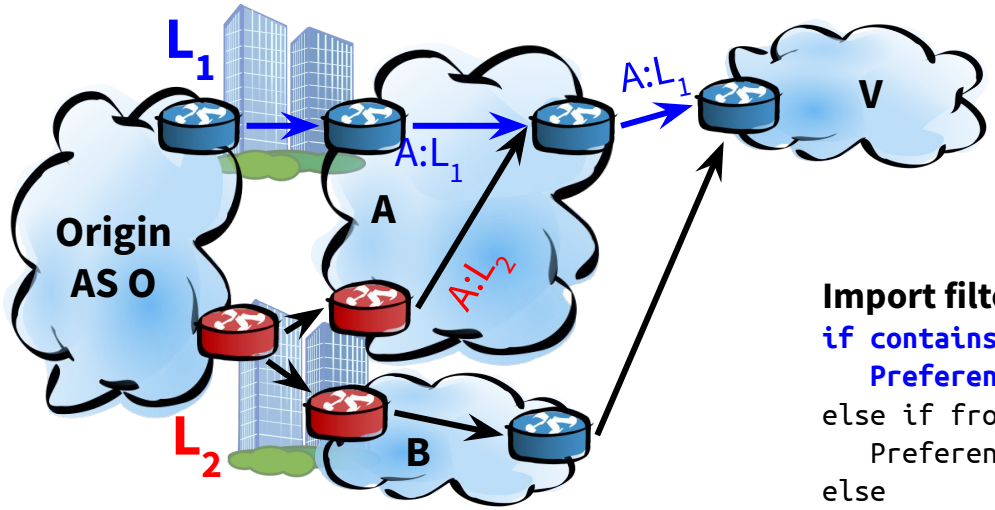
Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

Location Communities



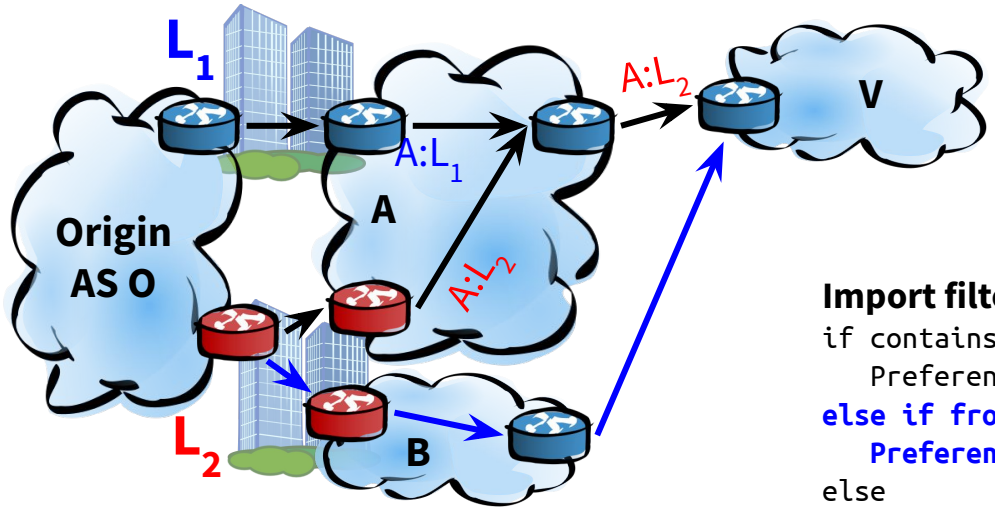
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

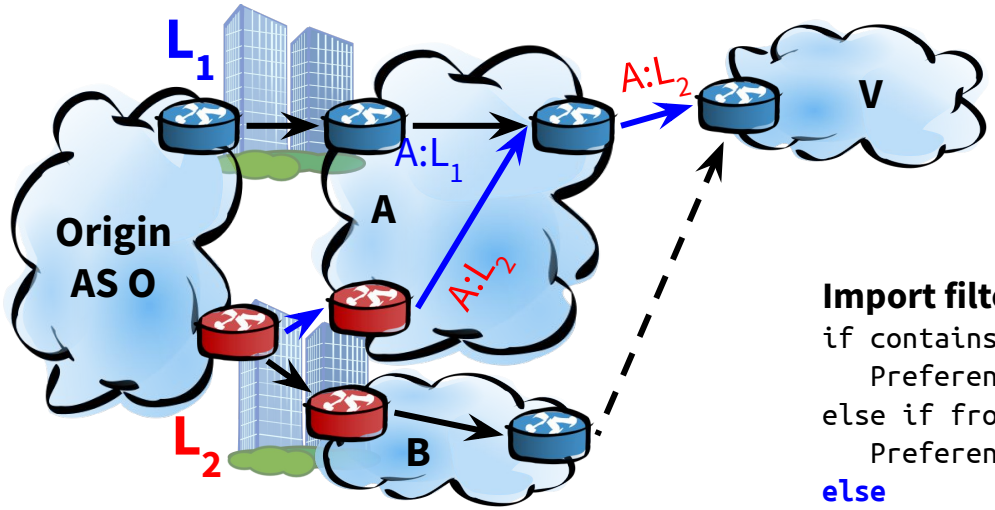
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

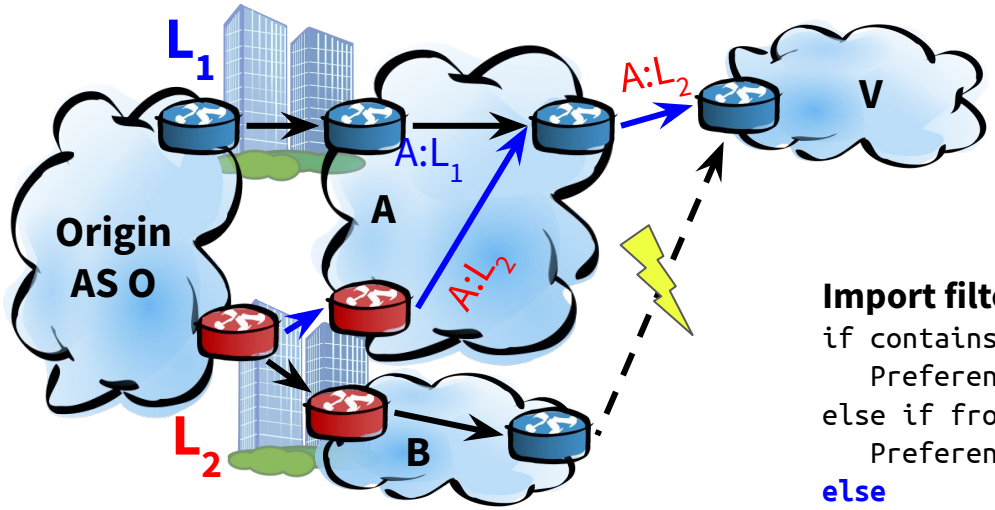
Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

Location Communities



Import filters at V:

```
if contains A:L1 then
   Preference ← 120
else if from B then
   Preference ← 100
else
   Preference ← 80
```

The problem is

- Community semantics are not standardized

The problem is

- ❏ Community semantics are not standardized
- ❏ Hard to know the semantics of a specific community

What information this community is related to?

The problem is

- ❏ Community semantics are not standardized
- ❏ Hard to know the semantics of a specific community
- ❏ Hard to know the community for some specific goal

Is the documentation reliable and up to date?

What information this community is related to?

The problem is

- ❏ Community semantics are not standardized
- ❏ Hard to know the semantics of a specific community
- ❏ Hard to know the community for some specific goal

Is the documentation reliable and up to date?

What information this community is related to?

How do I know if this route traverses LA or NY?

The problem is

- ❏ Community semantics are not standardized
- ❏ Hard to know the semantics of a specific community
- ❏ Hard to know the community for some specific goal

Is the documentation reliable and up to date?

What information this community is related to?

How do I know if this route traverses LA or NY?

Where is the documentation related to that community?

What we have so far?

- Current practical applications of BGP communities:
  - Help identify outages (Giotsas, *et al*, 2017)

What we have so far?

- Current practical applications of BGP communities:
  - Help identify outages (Giotsas, *et al*, 2017)
  - Determine the AS behaviour (remove or not communities) (Krenc, *et al*, 2021)

What we have so far?

- Current practical applications of BGP communities:
  - Help identify outages (Giotsas, *et al*, 2017)
  - Determine the AS behaviour (remove or not communities) (Krenc, *et al*, 2021)
  - Correct values for blackhole communities (Giotsas, *et al*, 2017)

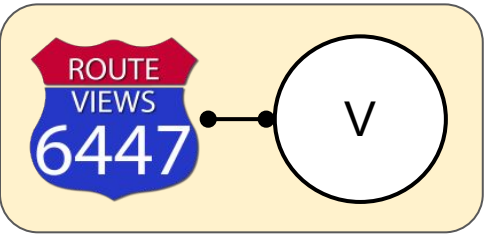
Algorithm

Inference of Location Communities

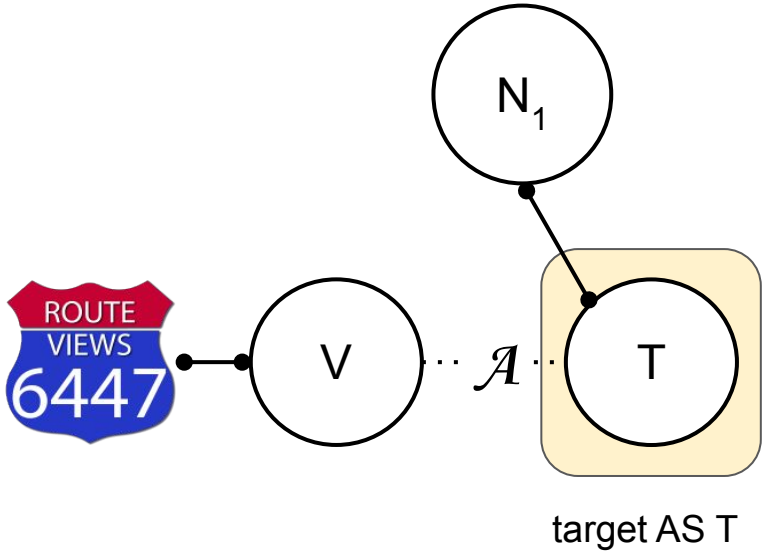
How we capture Location Communities

Collector

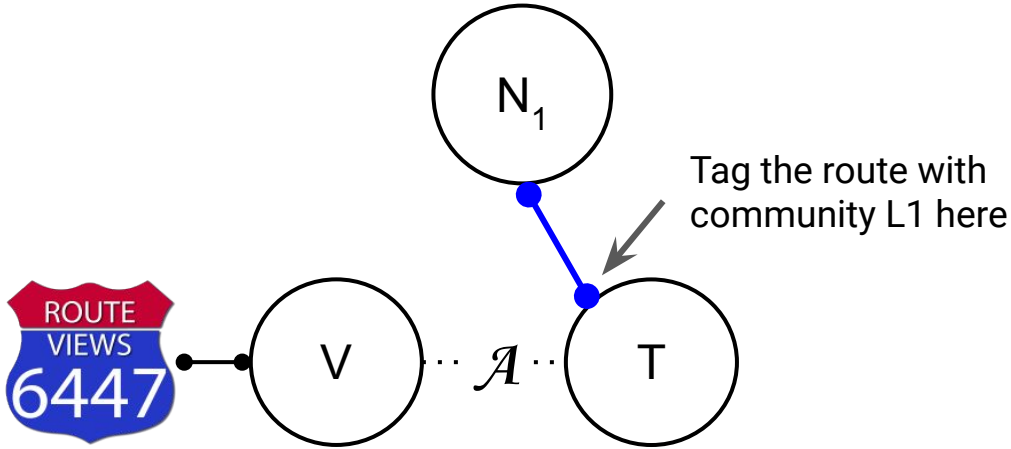
How we capture Location Communities



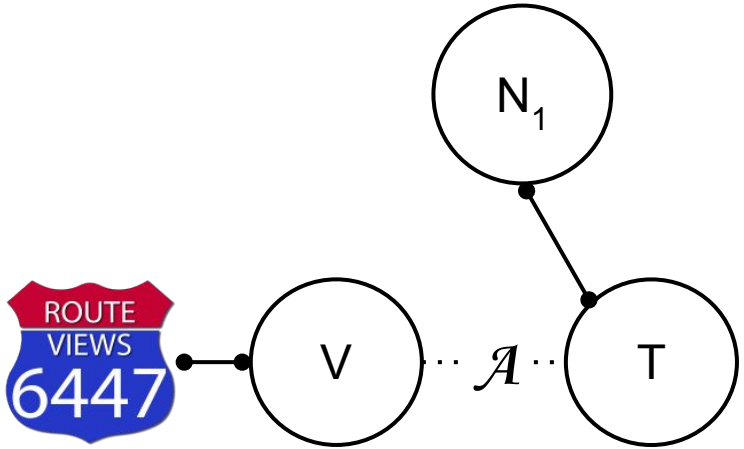
How we capture Location Communities



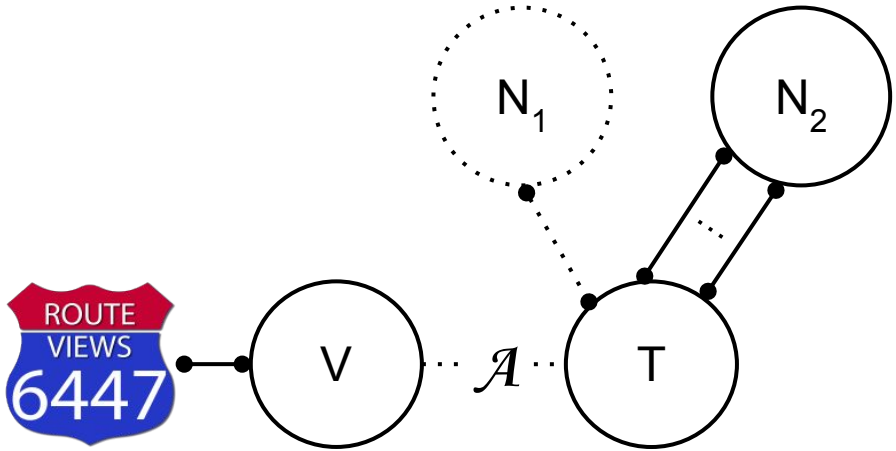
How we capture Location Communities



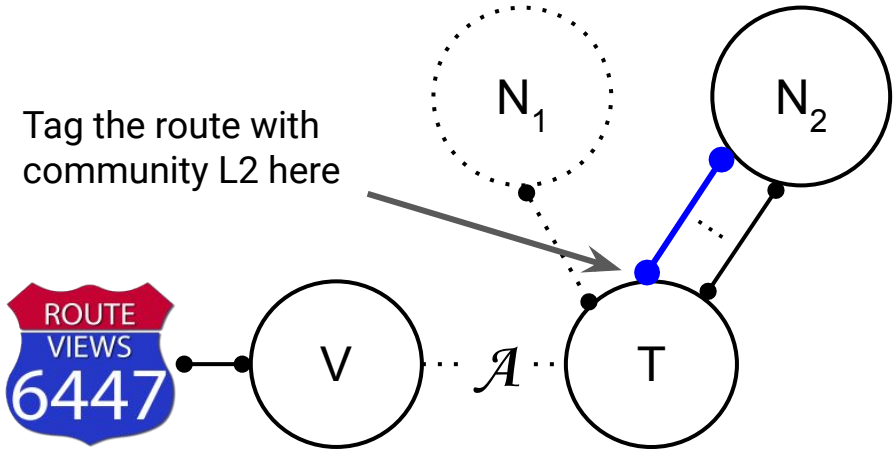
How we capture Location Communities



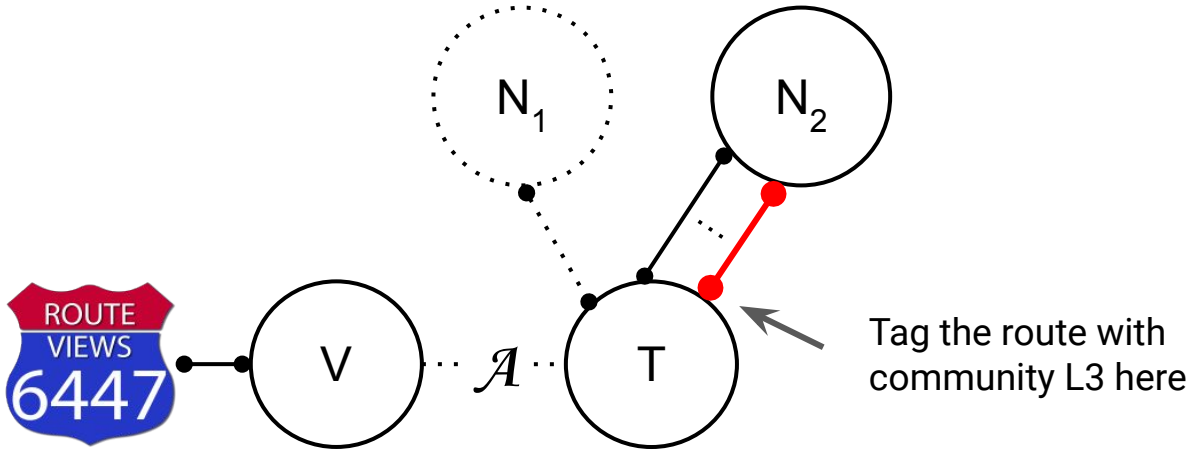
How we capture Location Communities



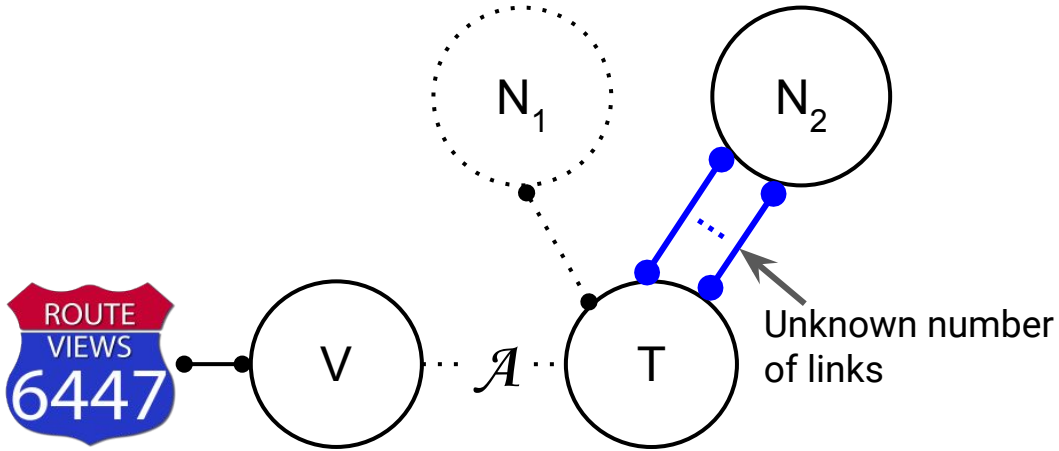
How we capture Location Communities



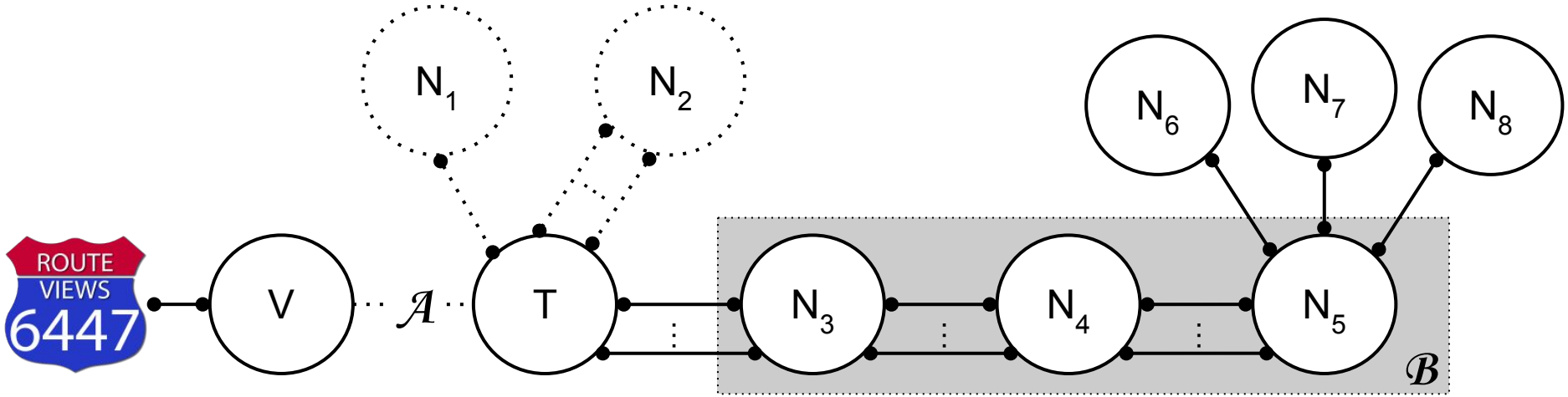
How we capture Location Communities



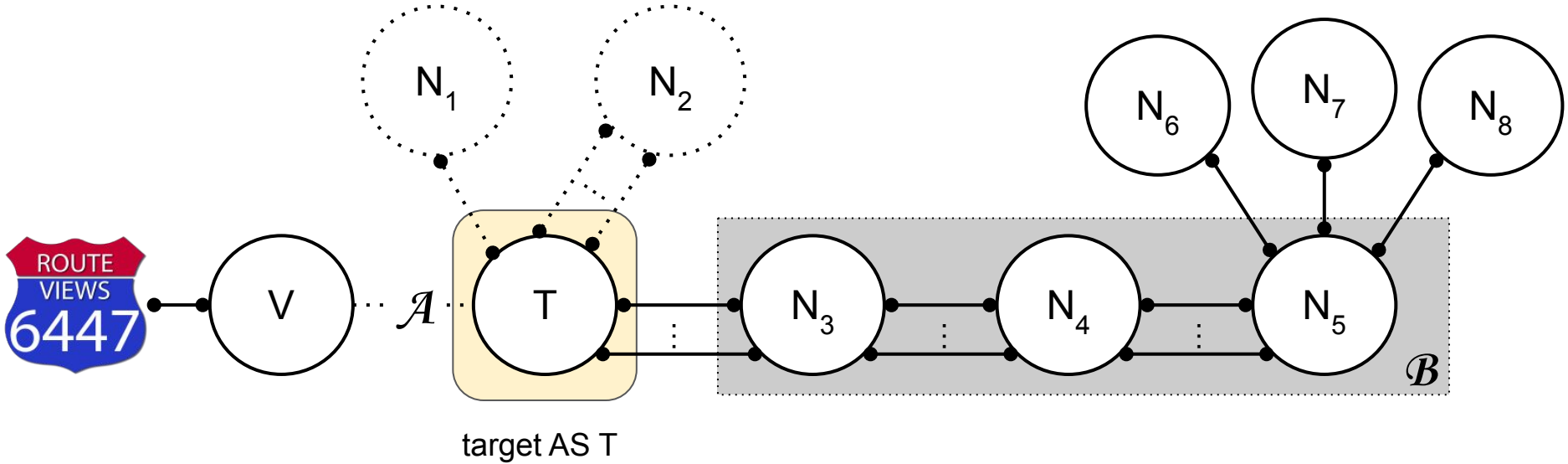
How we capture Location Communities



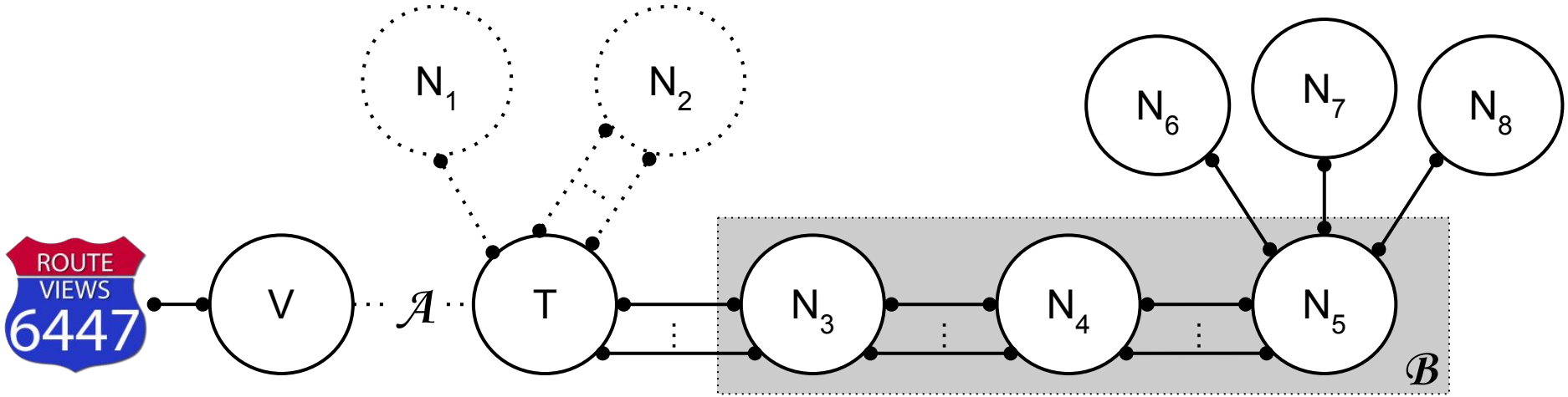
How we capture Location Communities



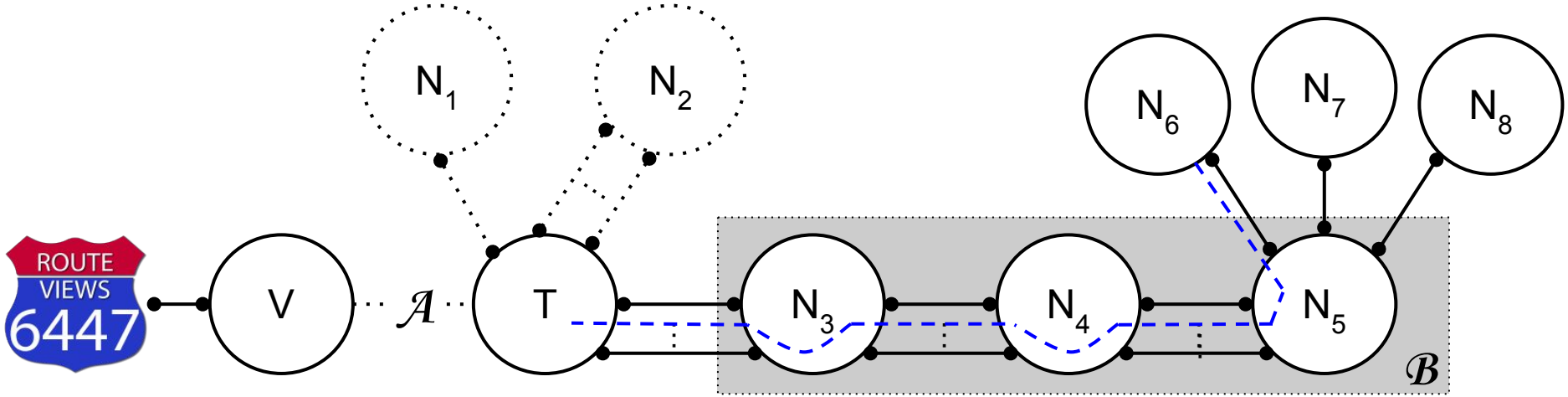
How we capture Location Communities



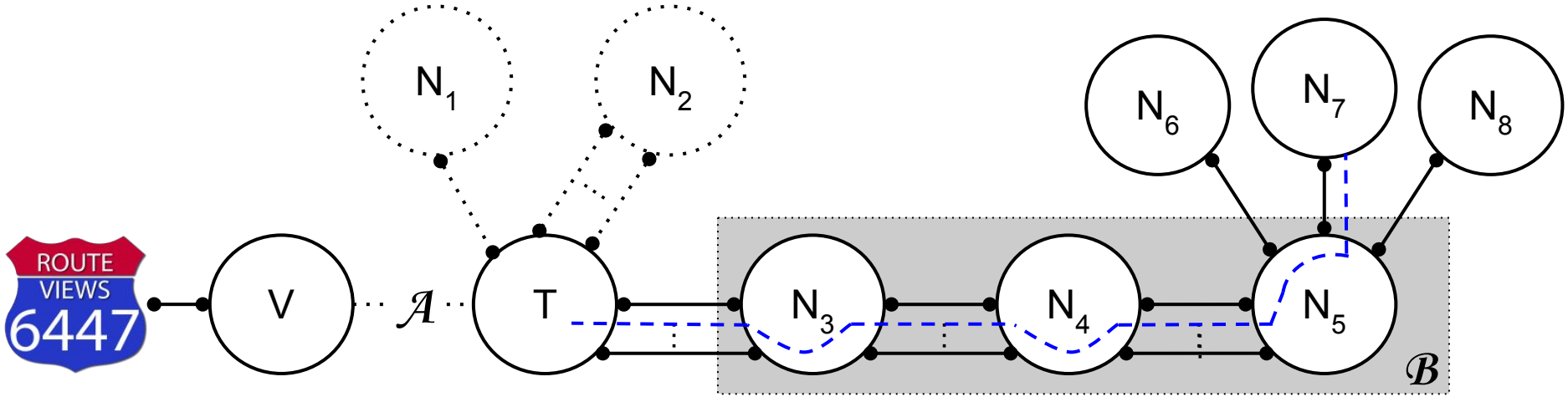
How we capture Location Communities



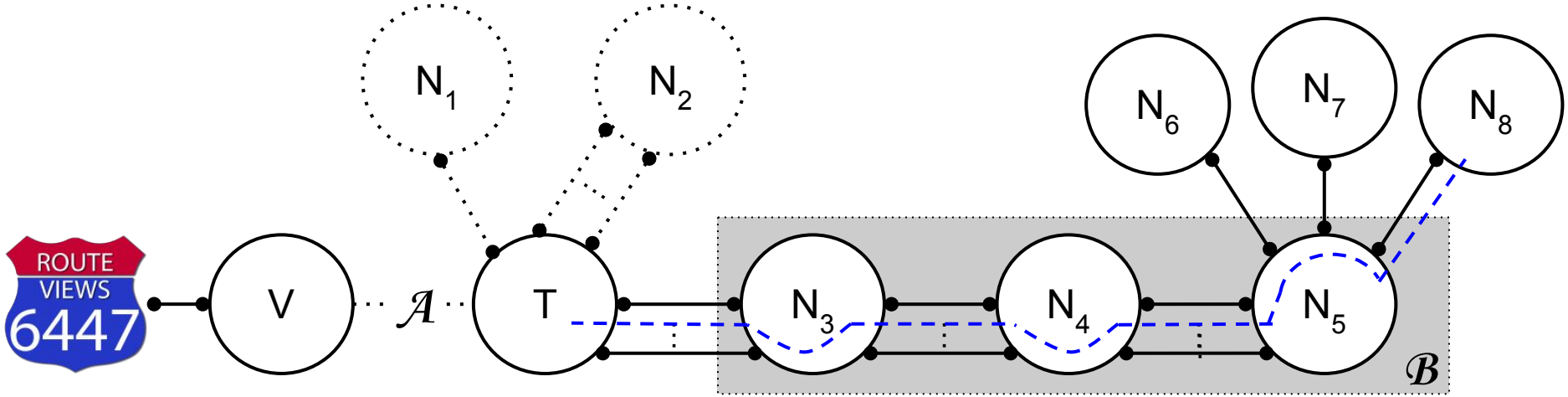
How we capture Location Communities



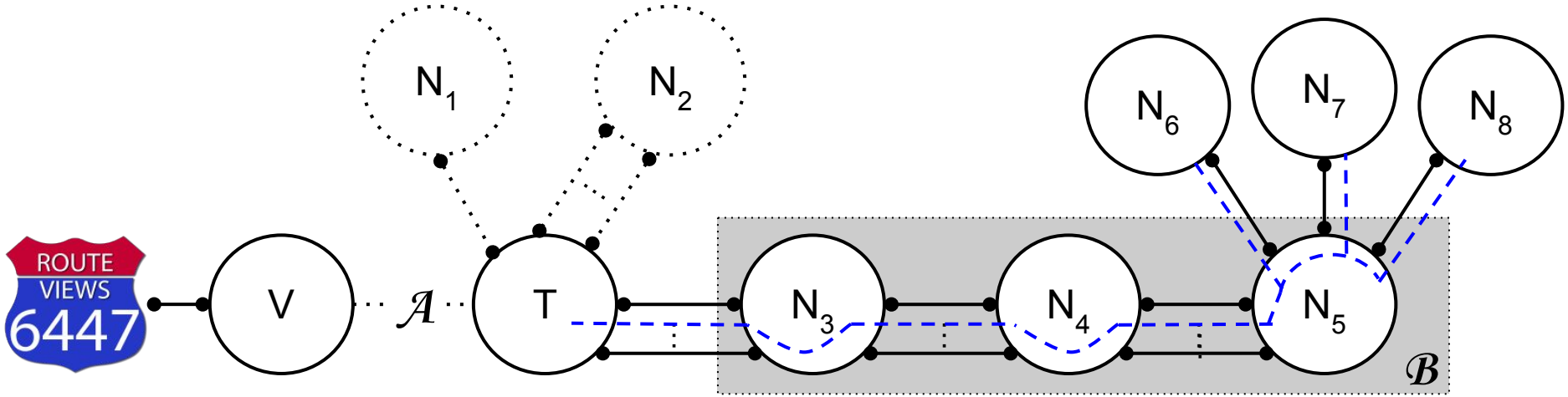
How we capture Location Communities



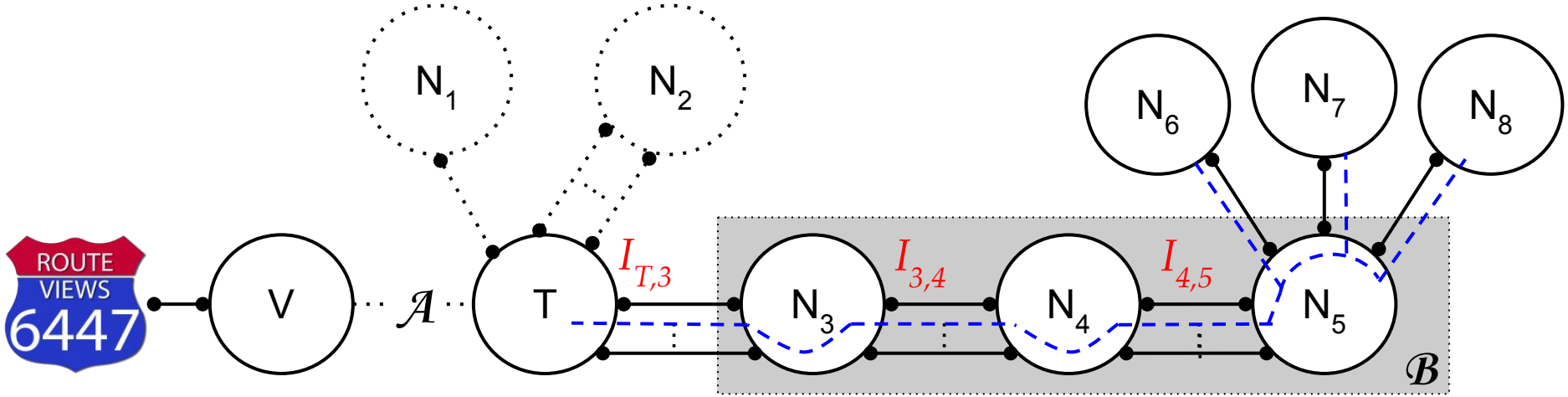
How we capture Location Communities



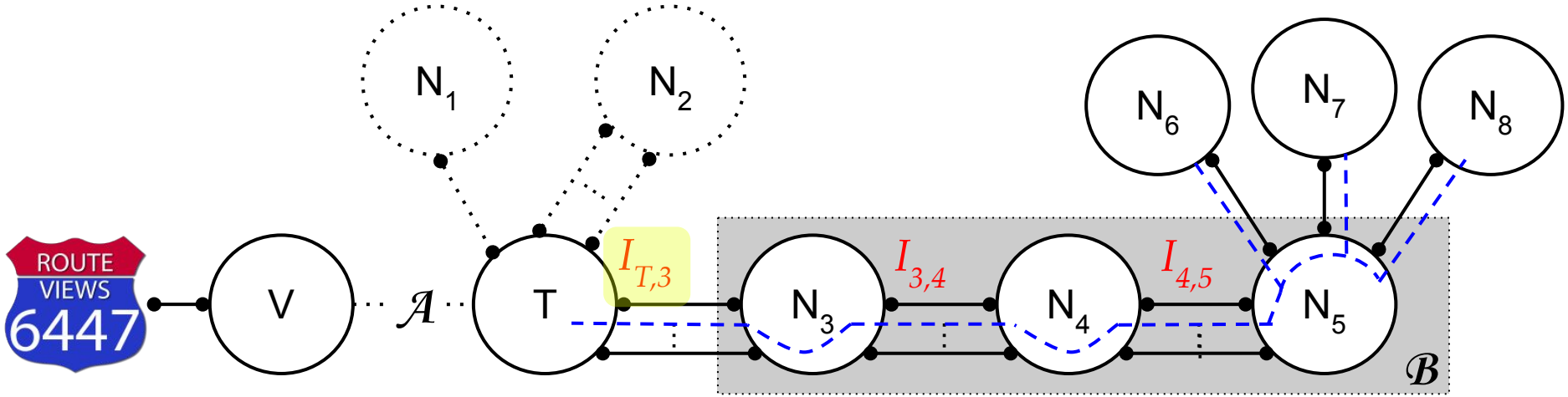
How we capture Location Communities



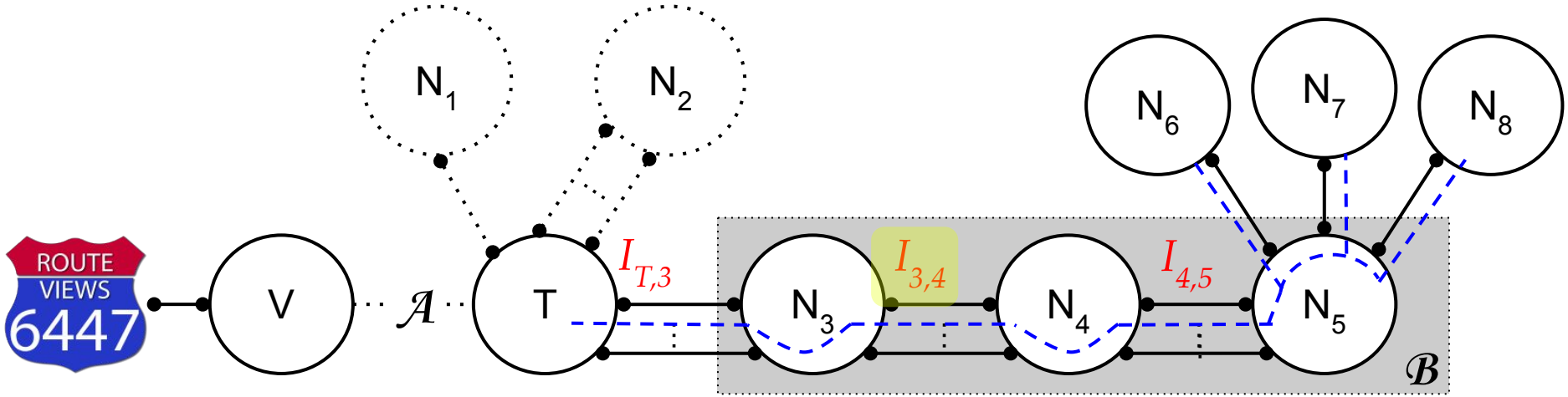
How we capture Location Communities



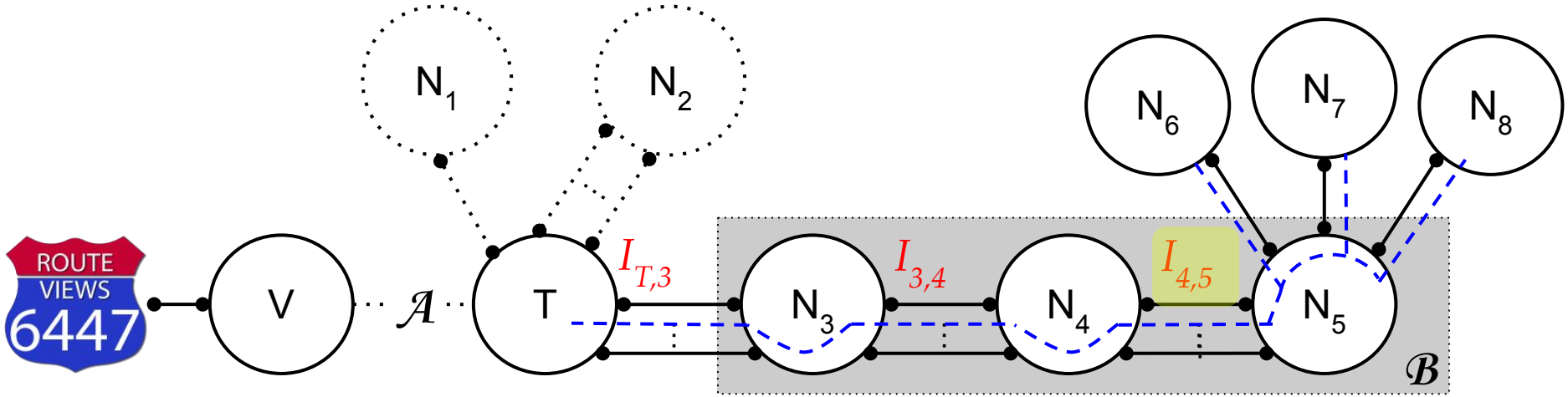
How we capture Location Communities



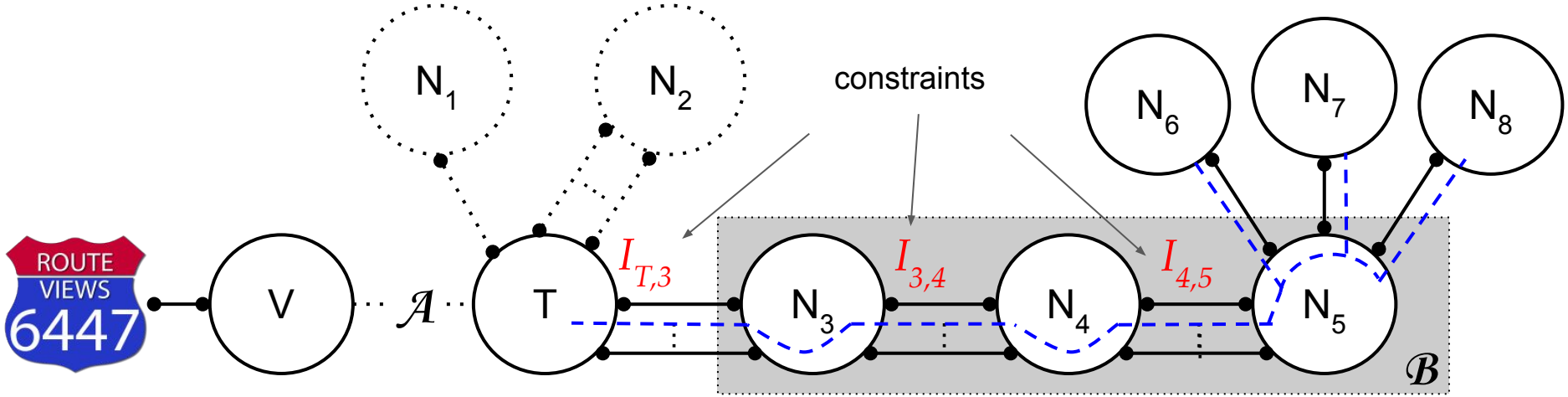
How we capture Location Communities



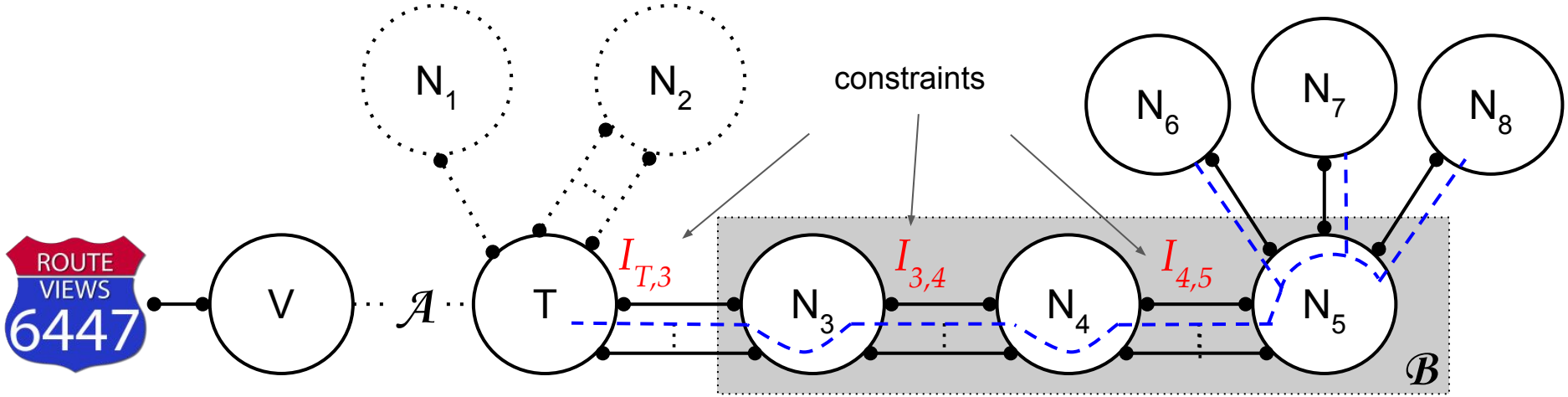
How we capture Location Communities



Evaluate Location Communities (limitation)



Evaluate Location Communities (summary)



Results and Dataset

Ground Truth (Tier 1)

Network (AS)	Community Type				CAIDA
	Geo	Dev/link	Relation	Action	
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Ground Truth (Tier 1)

Network (AS)	Community Type				CAIDA
	Geo	Dev/link	Relation	Action	
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Ground Truth (Tier 1)

Network (AS)	Community Type: Geo	Community Type: Dev/link	Community Type: Relation	Community Type: Action	CAIDA
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Ground Truth (Tier 1)

Network (AS)	Community Type				CAIDA
	Geo	Dev/link	Relation	Action	
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Ground Truth (Tier 1)

Network (AS)	Community Type				CAIDA
	Geo	Dev/link	Relation	Action	
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Ground Truth (Tier 1)

Network (AS)	Community Type				CAIDA
	Geo	Dev/link	Relation	Action	
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Ground Truth (Tier 1)

Network (AS)	Community Type: Geo	Community Type: Dev/link	Community Type: Relation	Community Type: Action	CAIDA
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Ground Truth (Tier 1)

Network (AS)	Community Type				CAIDA
	Geo	Dev/link	Relation	Action	
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Ground Truth (Tier 1)

Network (AS)	Community Type: Geo	Community Type: Dev/link	Community Type: Relation	Community Type: Action	CAIDA
Verizon (701)	0	0	0	11	0
NTT (2914)	93	0	2	44	39
GTT (3257)	10,000*	11,000*	1,783*	13,023*	68
Deutsche Telekom (3320)	24	0	3	0	17
Level 3 (3356)	178	0	2	5	82
PCCW Global (3491)	44	0	0	21	24
Lumen (3549)	239	239	239	87	28
Orange (5511)	46	0	0	55	11
Zayo (6461)	804*	0	6	152	0
Telecom Italia (6762)	51	0	1	133	42

Data Set

Our data set is built with:

- BGP feeds from RouteViews, RIPE RIS and Isolario.

Project	Collectors		VPs		Total ASes (thousands)		Routes (millions)	
Year	2017	2020	2017	2020	2017	2020	2017	2020
RV	17	20	192	232	61	72	96	184
RIPE	20	20	330	510	61	72	115	311
Isolario	4	5	83	145	60	72	66	209
Total (uniq)	41	45	529	738	61	73	277	704

Data Set

Our data set are built with:

- BGP feeds from RouteViews, RIPE RIS and Isolario.
- The first available route table dump (RIB) from each BGP route collector on December 2017, 2018, 2019, and 2020.

Project	Collectors		VPs		Total ASes (thousands)		Routes (millions)	
Year	2017	2020	2017	2020	2017	2020	2017	2020
RV	17	20	192	232	61	72	96	184
RIPE	20	20	330	510	61	72	115	311
Isolario	4	5	83	145	60	72	66	209
Total (uniq)	41	45	529	738	61	73	277	704

Data Set

Our data set are built with:

- BGP feeds from RouteViews, RIPE RIS and Isolario.
- The first available route table dump (RIB) from each BGP route collector on December 2017, 2018, 2019, and 2020.

We use BGP RIBs to process stable routes.

Project	Collectors		VPs		Total ASes (thousands)		Routes (millions)	
Year	2017	2020	2017	2020	2017	2020	2017	2020
RV	17	20	192	232	61	72	96	184
RIPE	20	20	330	510	61	72	115	311
Isolario	4	5	83	145	60	72	66	209
Total (uniq)	41	45	529	738	61	73	277	704

Data Set

Our data set are built with:

- BGP feeds from RouteViews, RIPE RIS and Isolario.
- The first available route table dump (RIB) from each BGP route collector on December 2017, 2018, 2019, and 2020.

We use BGP RIBs to process stable routes.

Project	Collectors		VPs		Total ASes (thousands)		Routes (millions)	
Year	2017	2020	2017	2020	2017	2020	2017	2020
RV	17	20	192	232	61	72	96	184
RIPE	20	20	330	510	61	72	115	311
Isolario	4	5	83	145	60	72	66	209
Total (uniq)	41	45	529	738	61	73	277	704

Data Set

Our data set are built with:

- BGP feeds from RouteViews, RIPE RIS and Isolario.
- The first available route table dump (RIB) from each BGP route collector on December 2017, 2018, 2019, and 2020.

We use BGP RIBs to process stable routes.

Project	Collectors		VPs		Total ASes (thousands)		Routes (millions)	
Year	2017	2020	2017	2020	2017	2020	2017	2020
RV	17	20	192	232	61	72	96	184
RIPE	20	20	330	510	61	72	115	311
Isolario	4	5	83	145	60	72	66	209
Total (uniq)	41	45	529	738	61	73	277	704

Data Set

Our data set are built with:

- BGP feeds from RouteViews, RIPE RIS and Isolario.
- The first available route table dump (RIB) from each BGP route collector on December 2017, 2018, 2019, and 2020.

We use BGP RIBs to process stable routes.

Project	Collectors		VPs		Total ASes (thousands)		Routes (millions)	
Year	2017	2020	2017	2020	2017	2020	2017	2020
RV	17	20	192	232	61	72	96	184
RIPE	20	20	330	510	61	72	115	311
Isolario	4	5	83	145	60	72	66	209
Total (uniq)	41	45	529	738	61	73	277	704

Results of the Evaluation of Location Communities

Configuration	Precision	Recall	Inf. Recall
Prioritize precision	0.93	0.72	0.89
Default configuration	0.91	0.80	0.87
Prioritize recall	0.87	0.81	0.89

Results of the Evaluation of Location Communities

Configuration	Precision	Recall	Inf. Recall
Prioritize precision	0.93	0.72	0.89
Default configuration	0.91	0.80	0.87
Prioritize recall	0.87	0.81	0.89

Results of the Evaluation of Location Communities

Configuration	Precision	Recall	Inf. Recall
Prioritize precision	0.93	0.72	0.89
Default configuration	0.91	0.80	0.87
Prioritize recall	0.87	0.81	0.89

Results of the Evaluation of Location Communities

Configuration	Precision	Recall	Inf. Recall
Prioritize precision	0.93	0.72	0.89
Default configuration	0.91	0.80	0.87
Prioritize recall	0.87	0.81	0.89

Results of the Evaluation of Location Communities

Configuration	Precision	Recall	Inf. Recall
Prioritize precision	0.93	0.72	0.89
Default configuration	0.91	0.80	0.87
Prioritize recall	0.87	0.81	0.89

Automatic Inference vs CAIDA

Community Type	Database	Recall	Precision	Communities Total	Communities Correct
Geolocation	CAIDA	0.21	0.86	303	261
	Inferences	0.77	—	—	—
Location	Inferences	0.80	0.91	1081	983

Automatic Inference vs CAIDA

Community Type	Database	Recall	Precision	Communities Total	Communities Correct
Geolocation	CAIDA	0.21	0.86	303	261
	Inferences	0.77	—	—	—
Location	Inferences	0.80	0.91	1081	983

Automatic Inference vs CAIDA

Community Type	Database	Recall	Precision	Communities Total	Communities Correct
Geolocation	CAIDA	0.21	0.86	303	261
	Inferences	0.77	—	—	—
Location	Inferences	0.80	0.91	1081	983

Automatic Inference vs CAIDA

Community Type	Database	Recall	Precision	Communities Total	Communities Correct
Geolocation	CAIDA	0.21	0.86	303	261
	Inferences	0.77	—	—	—
Location	Inferences	0.80	0.91	1081	983

Automatic Inference vs CAIDA

Community Type	Database	Recall	Precision	Communities Total	Communities Correct
Geolocation	CAIDA	0.21	0.86	303	261
	Inferences	0.77	—	—	—
Location	Inferences	0.80	0.91	1081	983

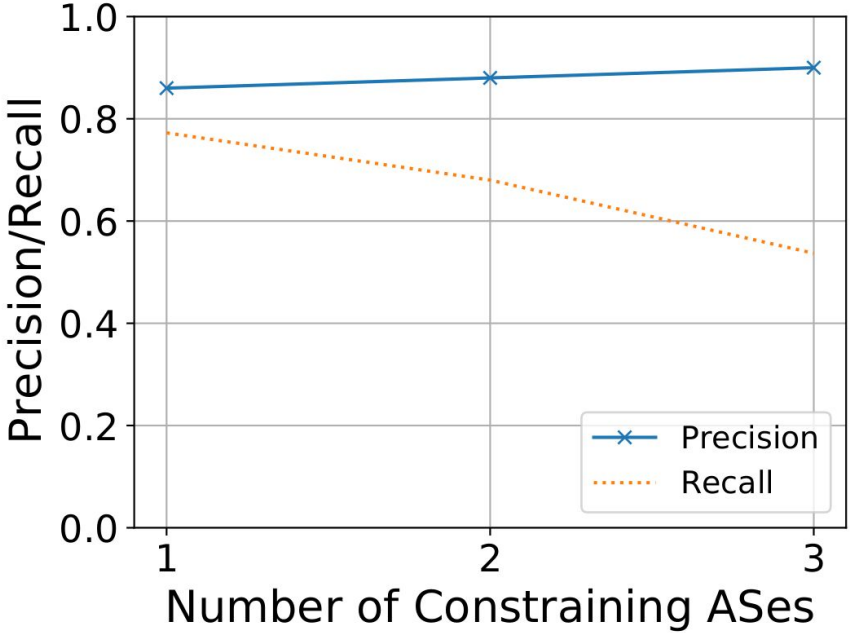
Automatic Inference vs CAIDA

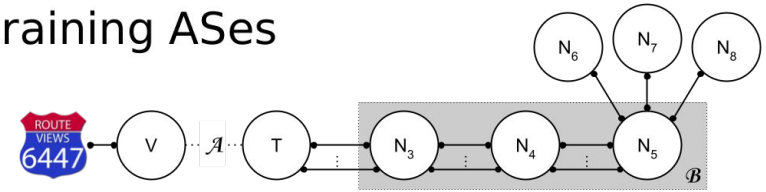
Community Type	Database	Recall	Precision	Communities Total	Communities Correct
Geolocation	CAIDA	0.21	0.86	303	261
	Inferences	0.77	—	—	—
Location	Inferences	0.80	0.91	1081	983

Automatic Inference vs CAIDA

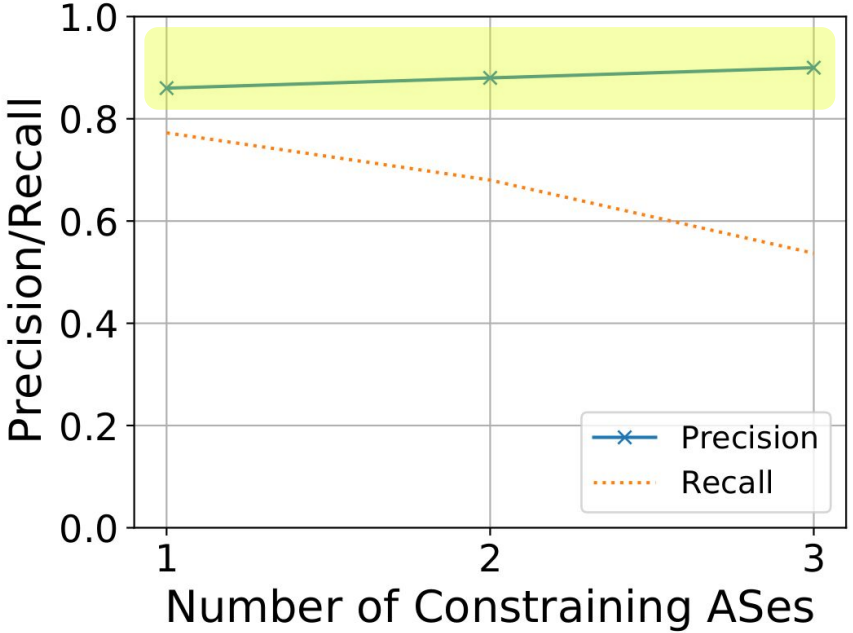
Community Type	Database	Recall	Precision	Communities Total	Communities Correct
Geolocation	CAIDA	0.21	0.86	303	261
	Inferences	0.77	—	—	—
Location	Inferences	0.80	0.91	1081	983

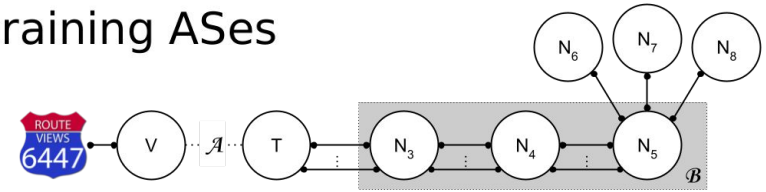
Number of Constraining ASes



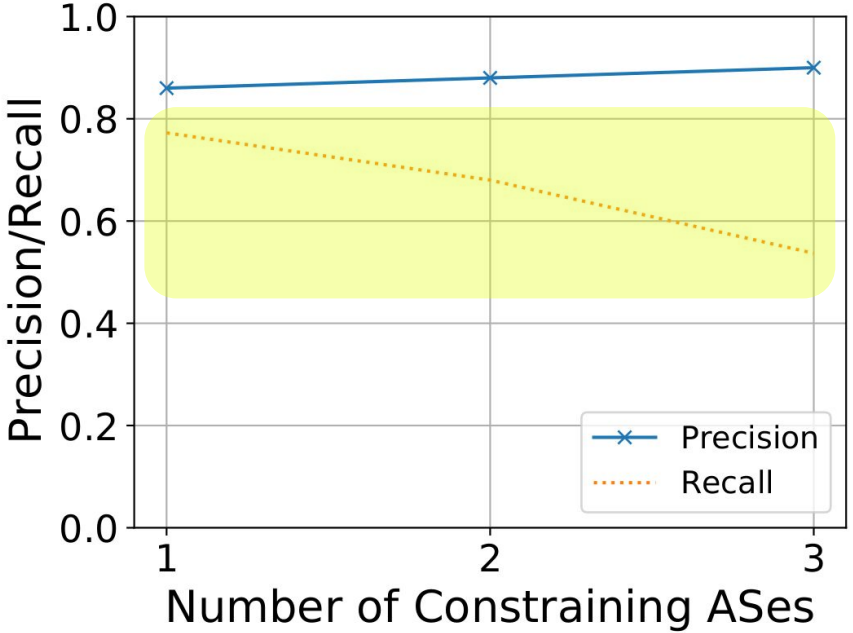


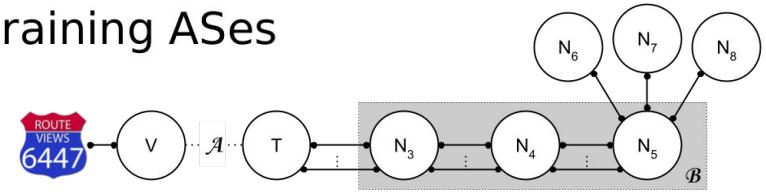
Number of Constraining ASes



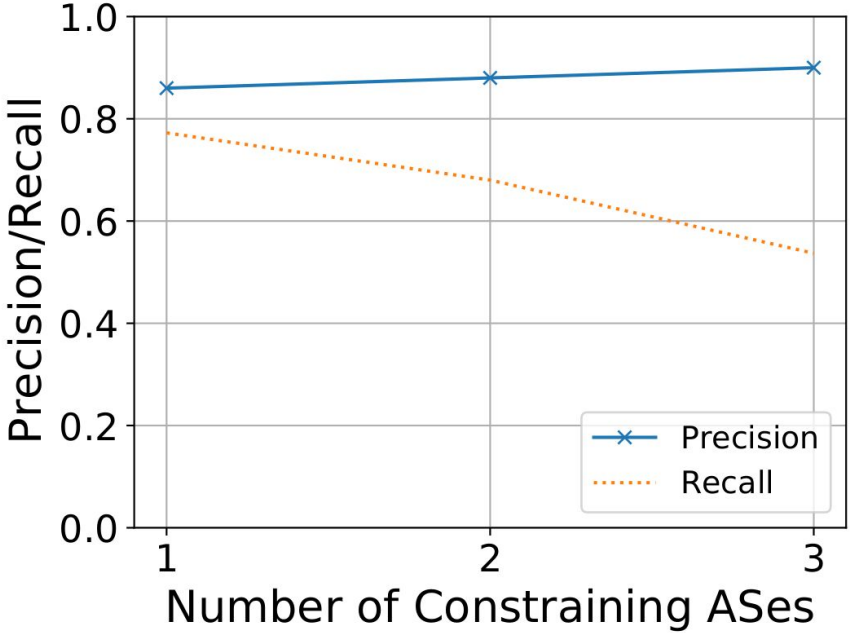


Number of Constraining ASes

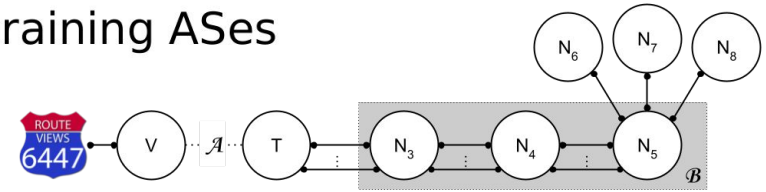




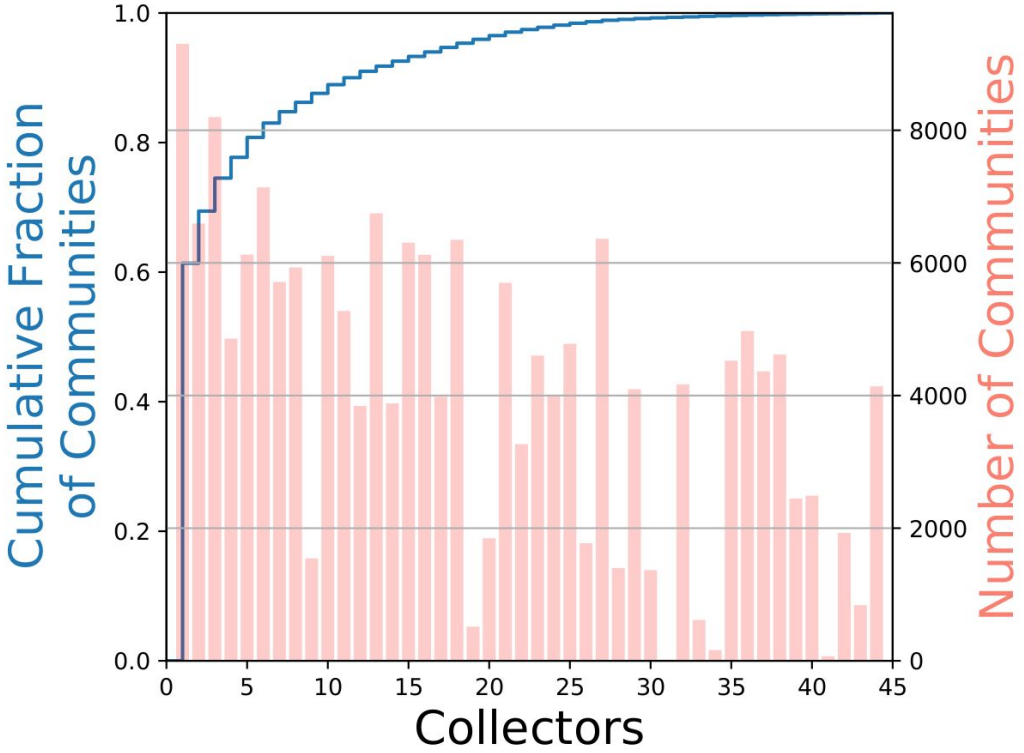
Number of Constraining ASes



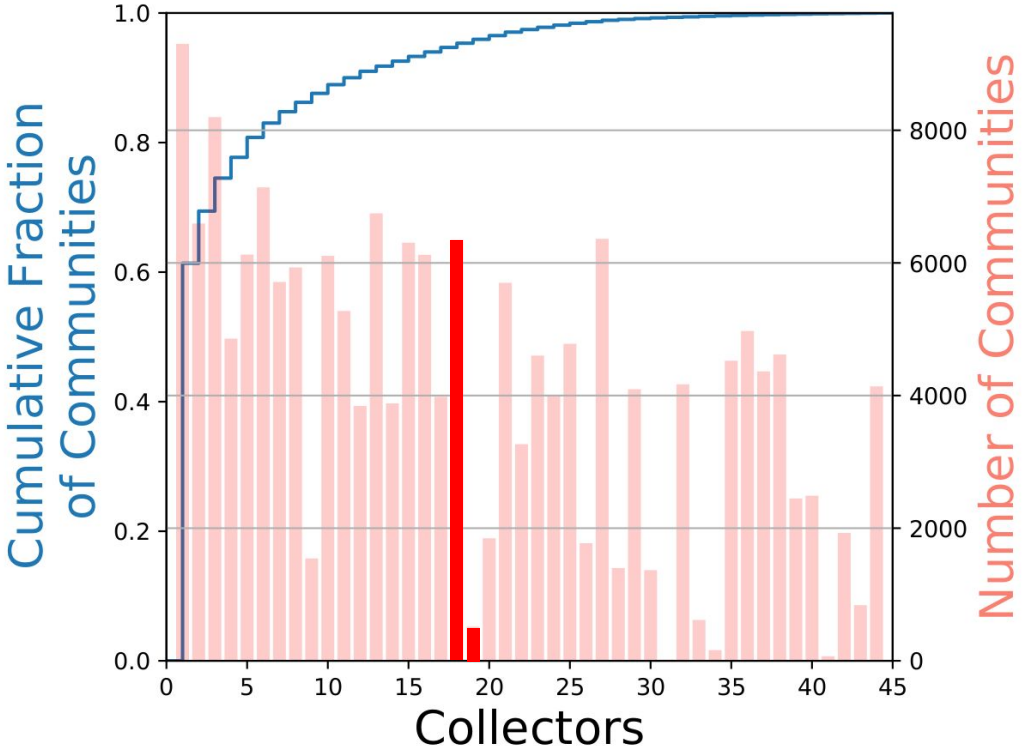
- More constraining ASes lead to higher precision at the cost of recall.



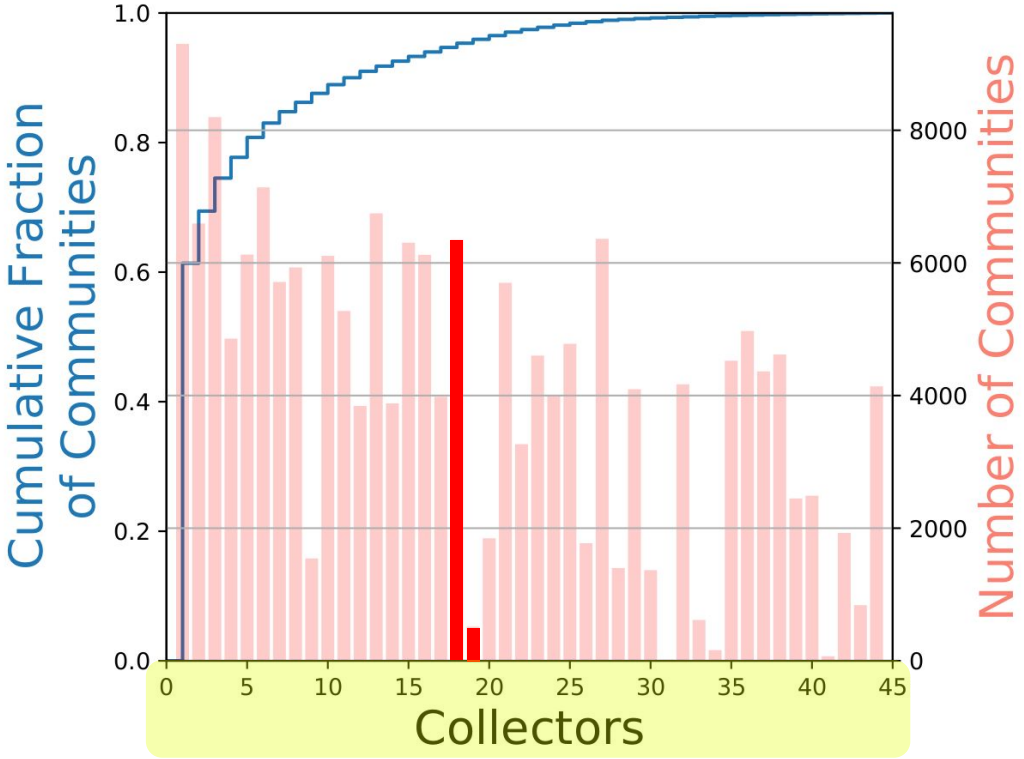
Contribution on Inference per Collector



Contribution on Inference per Collector



Contribution on Inference per Collector



- Additional collectors would support more inferences.

Conclusion

- In recent years, the use of BGP communities has increased significantly.

Conclusion

- In recent years, the use of BGP communities has increased significantly.
- Our work is the first we are aware of to use routing announcements to infer the semantics of BGP communities.

Conclusion

- In recent years, the use of BGP communities has increased significantly.
- Our work is the first we are aware of to use routing announcements to infer the semantics of BGP communities.
- Our algorithm automatically infers location communities and achieves high precision (93%) and recall (81%).

Conclusion

- In recent years, the use of BGP communities has increased significantly.
- Our work is the first we are aware of to use routing announcements to infer the semantics of BGP communities.
- Our algorithm automatically infers location communities and achieves high precision (93%) and recall (81%).
- Our manually-built ground truth DB, as well as the code to generate the location DB are publicly available.

Thank you!