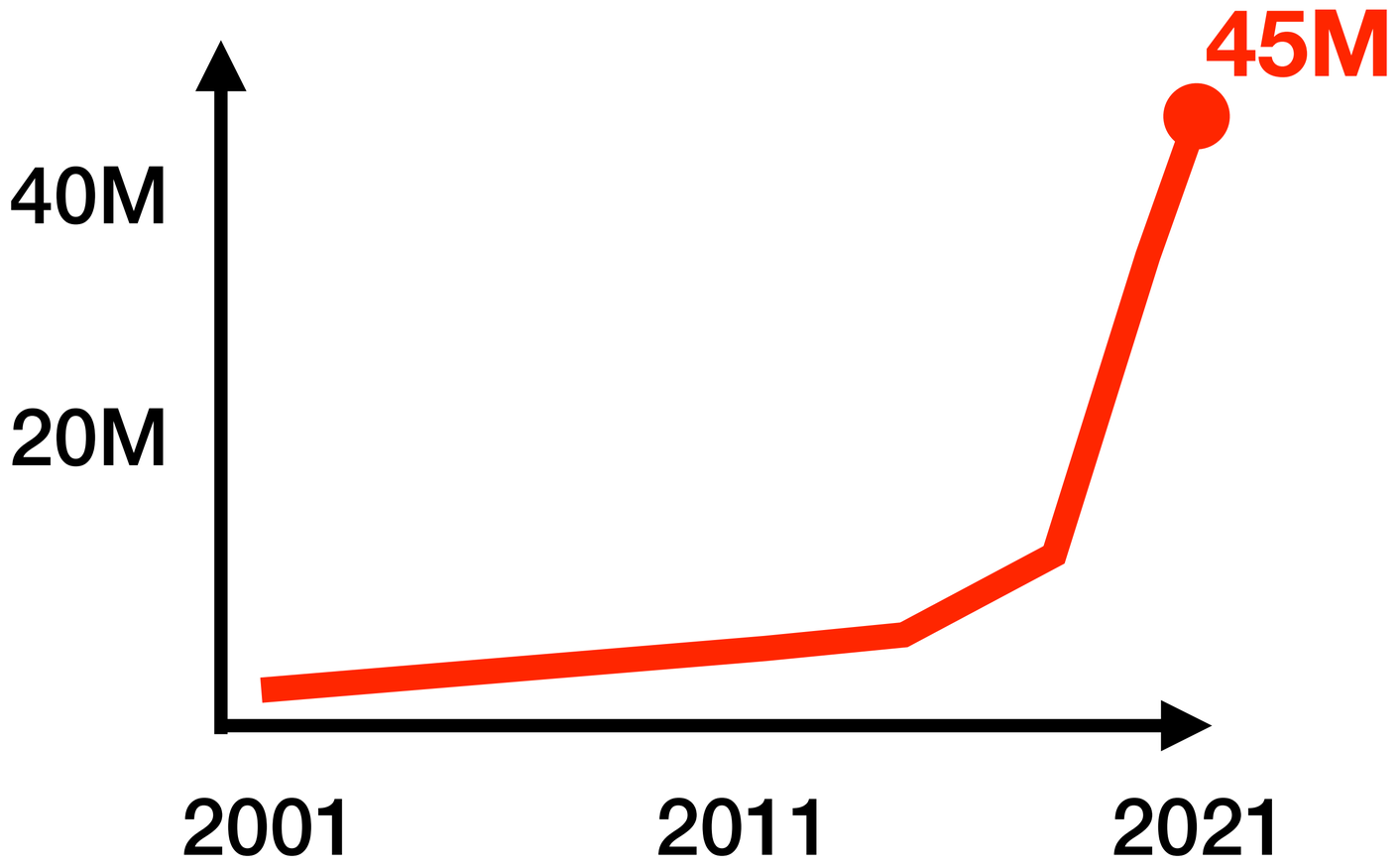
MVP : Measuring Internet Routing from the Most Valuable Points

Thomas Holterbach
University of Strasbourg

RIPE 85
Belgrade, Serbia
27 Oct 2022

Joint work with:
Thomas Alfroy
Cristel Pelsser

Median number of BGP routes collected every hour (RIPE RIS)

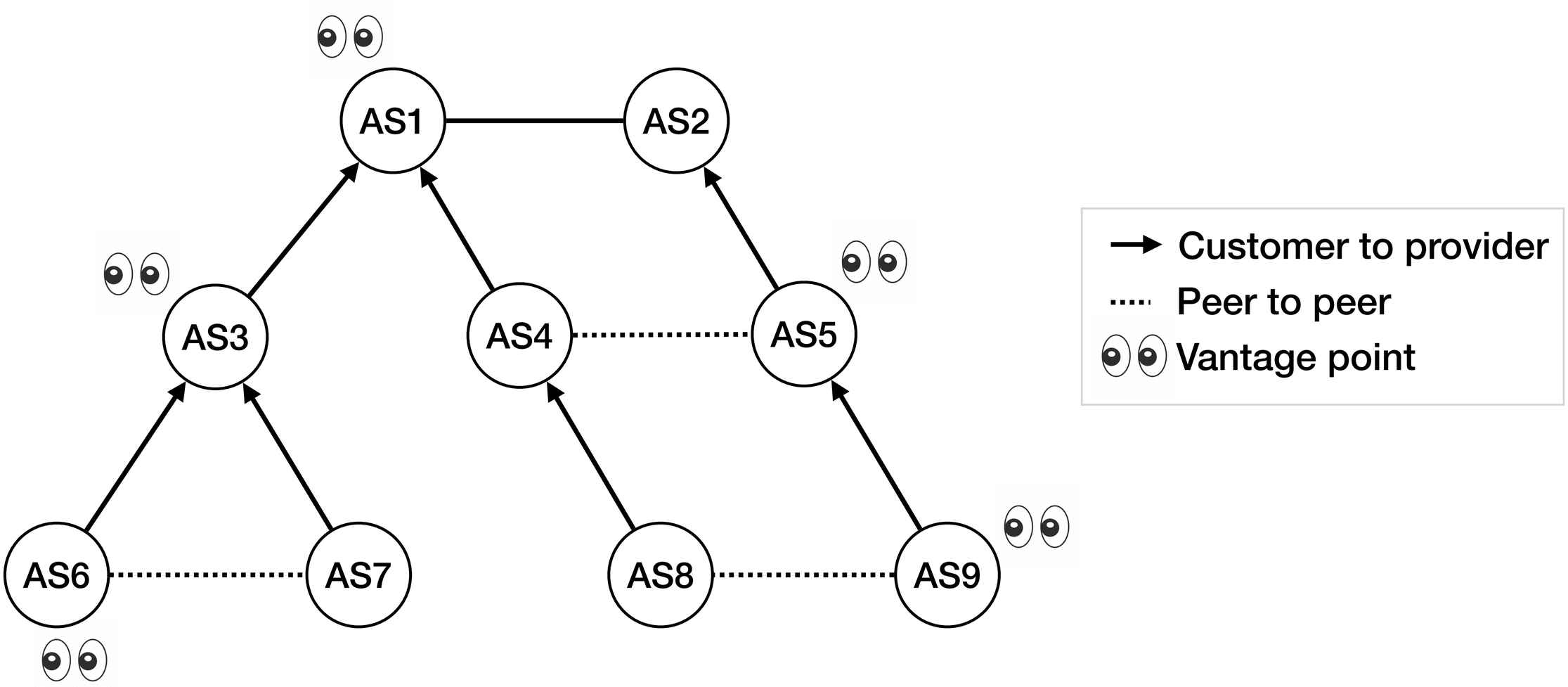


The number of collected BGP routes exponentially increases for two reasons

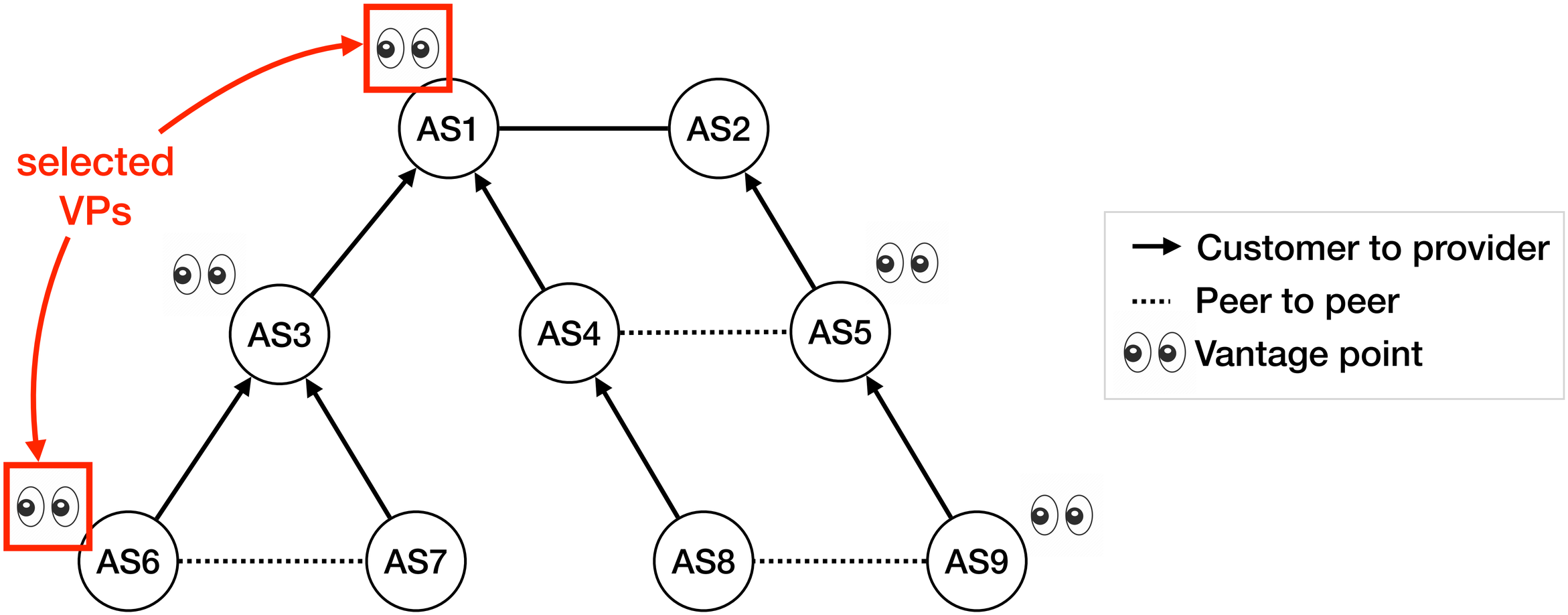
1. More and more vantage points are deployed

2. More and more prefixes are advertised

A common solution is to focus on a few vantage points



A common solution is to focus on a few vantage points

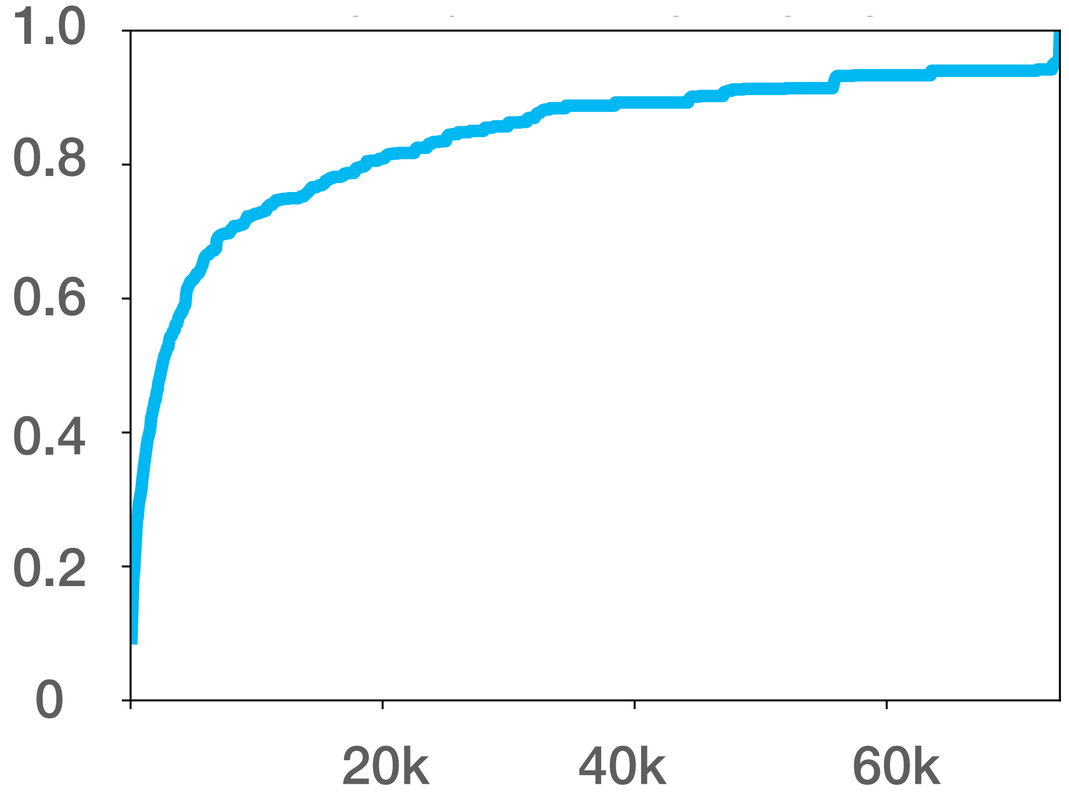


Selecting the right set of vantage points is hard because of two conflicting phenomena

High BGP routes redundancy

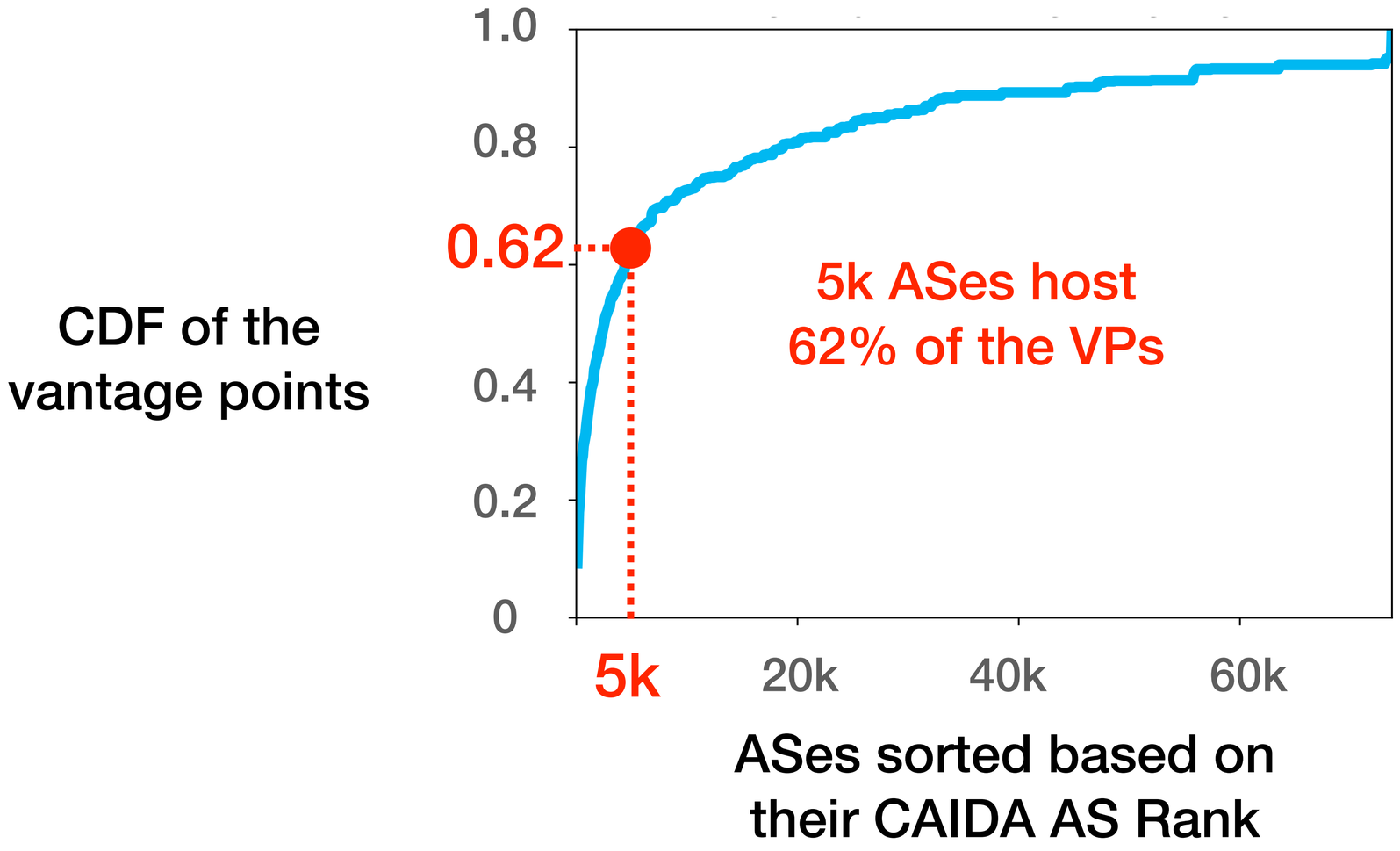
The skewed position of the vantage points amplifies the redundancy between the collected BGP routes

CDF of the vantage points



ASes sorted based on their CAIDA AS Rank

The skewed position of the vantage points amplifies the redundancy between the collected BGP routes



Selecting the right set of vantage points is hard because of two conflicting phenomena

High BGP routes redundancy

Sparse BGP event visibility

Many events are detected by a tiny fraction of the vantage points

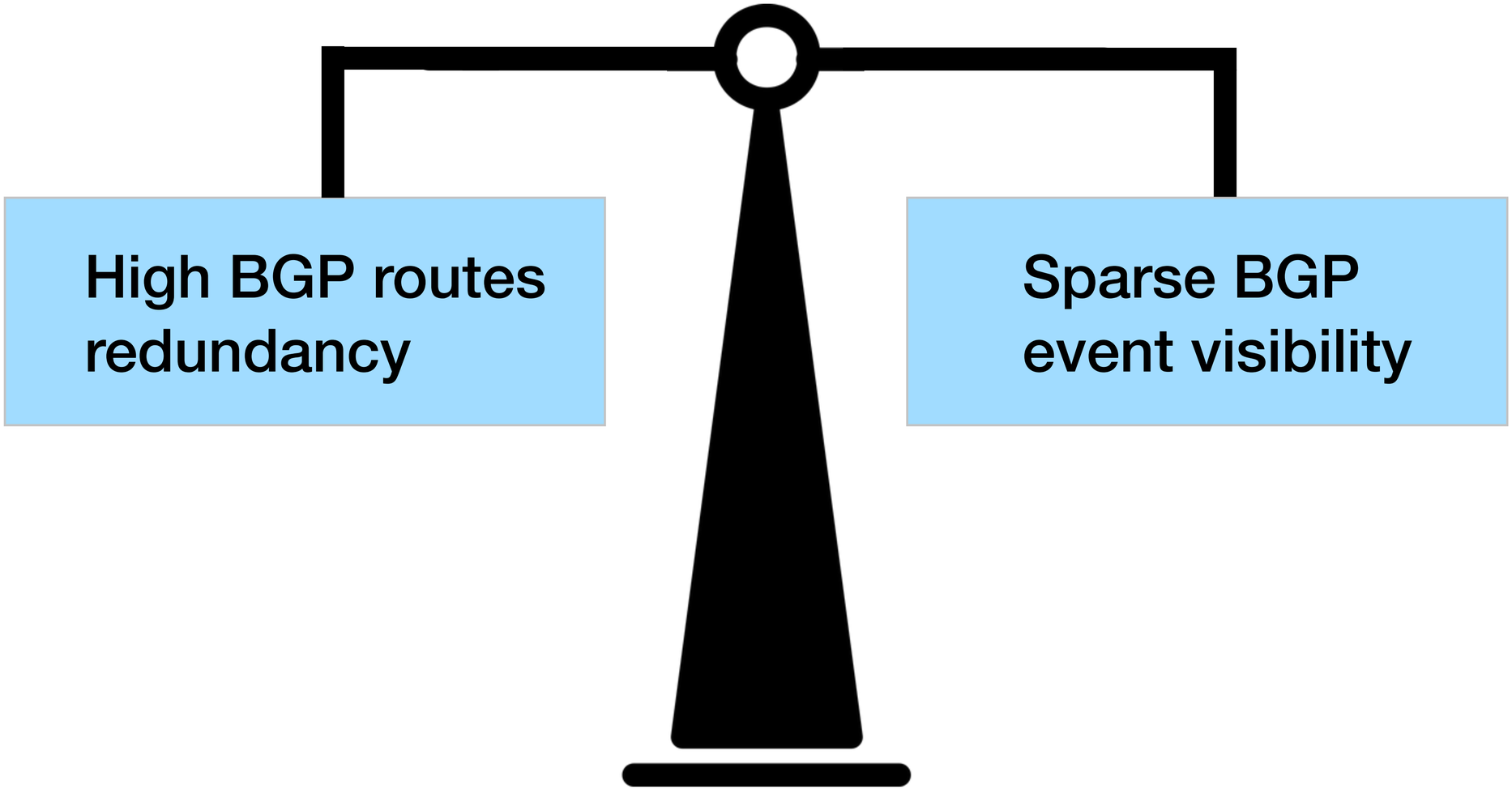
	Number of vantage points that detected the event		
	< 21	21 - 80	> 80
Proportion of the BGP Hijacks	65 %	19 %	16 %

MANRS blogpost: BGP Security in 2021

High BGP routes redundancy

Sparse BGP event visibility

MVP : Measuring Internet Routing from the Most Valuable Points



Goal: **Select BGP Vantage Points (VPs) that maximises utility and minimises volume of data**

Step #1: **Quantifying the observation of the VPs for past events**

Step #2: **Measuring similarity between VPs for every event**

Step #3: **Selecting a set of dissimilar VPs**

Goal: **Select BGP Vantage Points (VPs) that maximises utility and minimises volume of data**

Step #1: **Quantifying the observation of the VPs for past events**

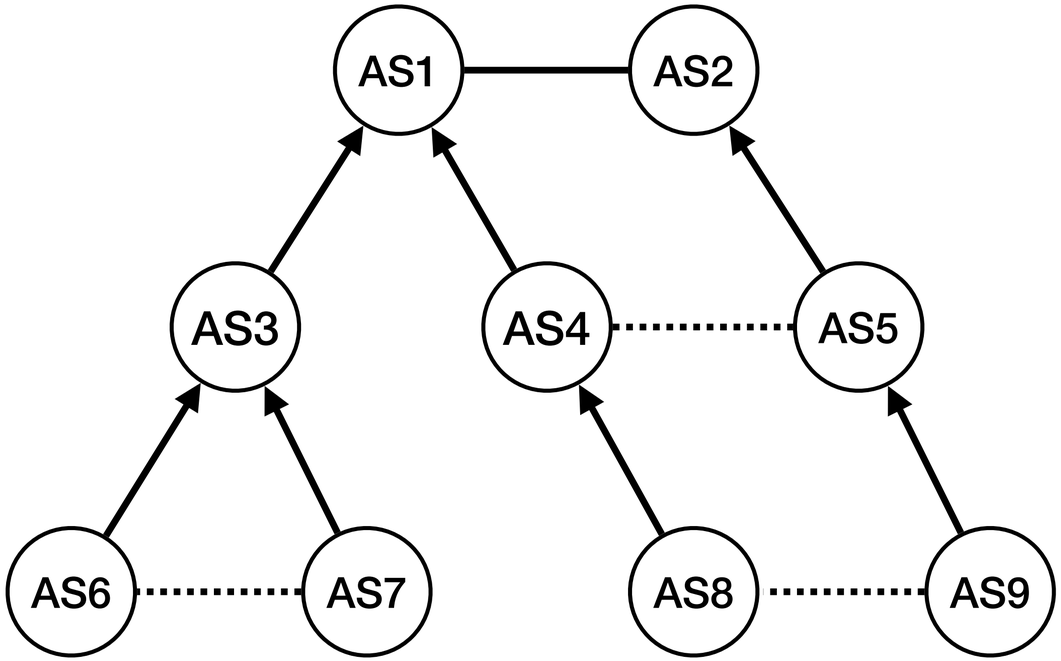
Step #2: **Measuring similarity between VPs for every event**

Step #3: **Selecting a set of dissimilar VPs**

MVP computes the change induced by a new AS link on topological features

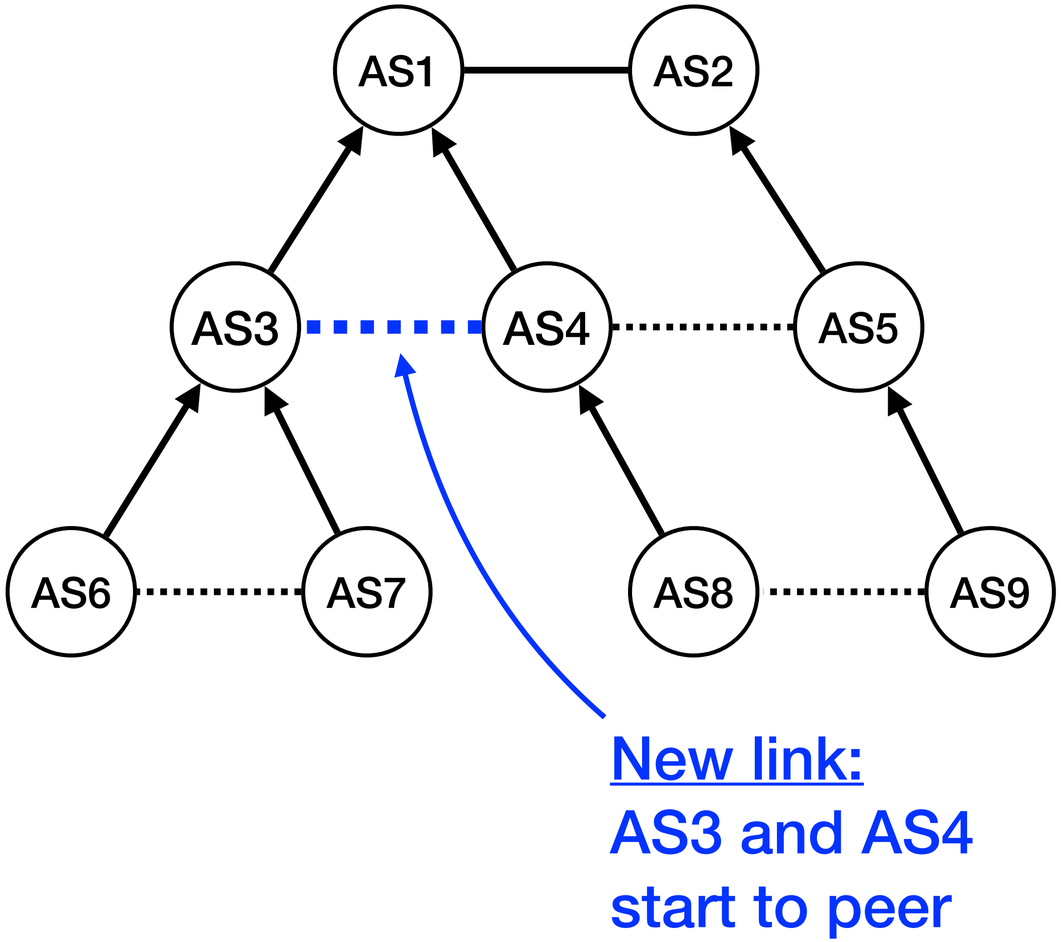
MVP computes the change induced by a new AS link on topological features

1. ***MVP*** takes an AS link that appears



MVP computes the change induced by a new AS link on topological features

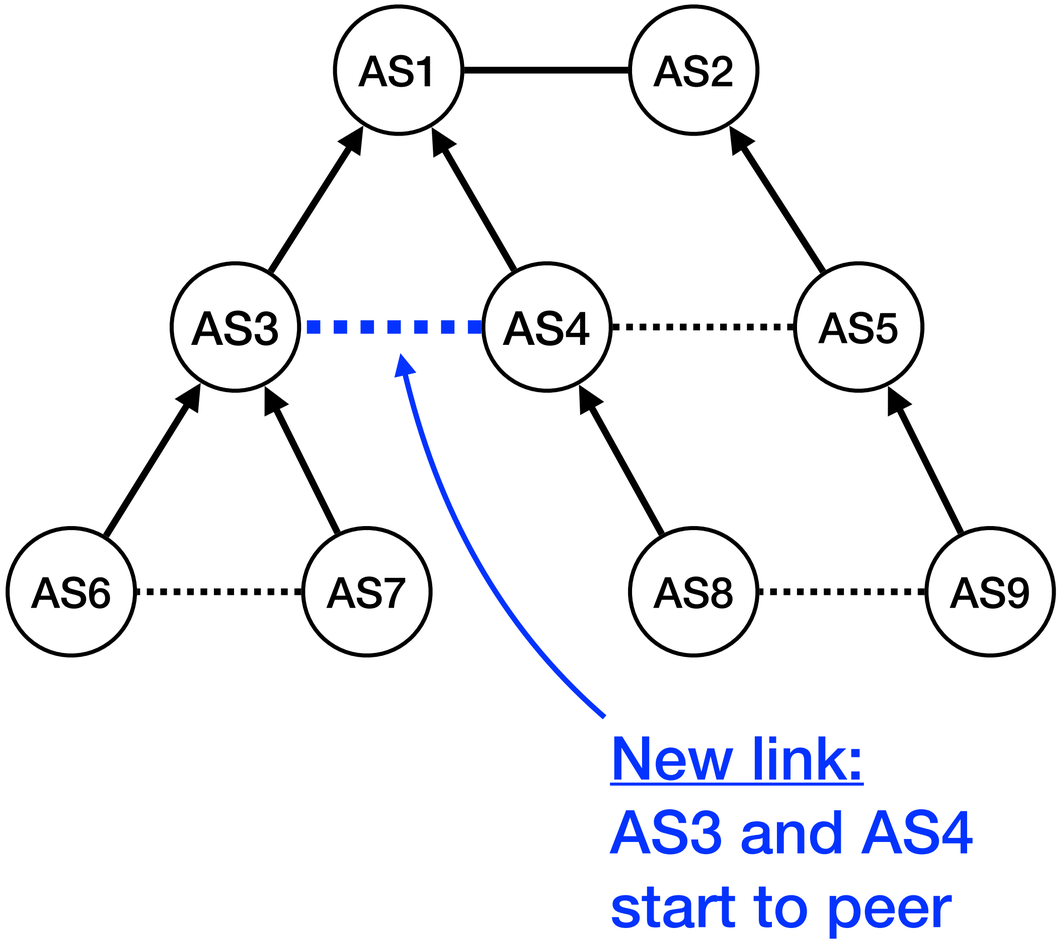
1. ***MVP*** takes an AS link that appears



MVP computes the change induced by a new AS link on topological features

1. ***MVP*** takes an AS link that appears

2. ***MVP*** computes the change induced by the new link on the feature values

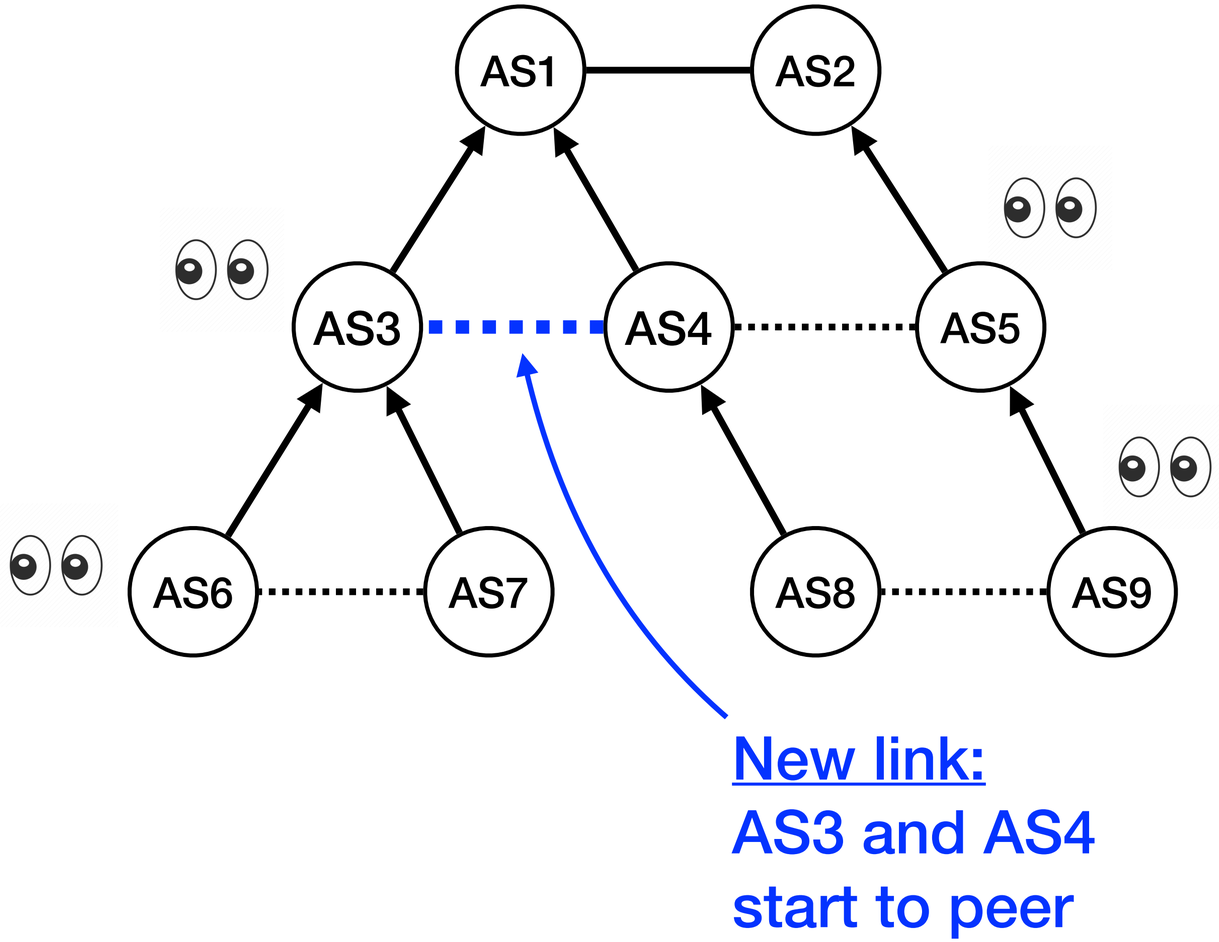


MVP computes the change induced by a new AS link on topological features

1. ***MVP*** takes an AS link that appears

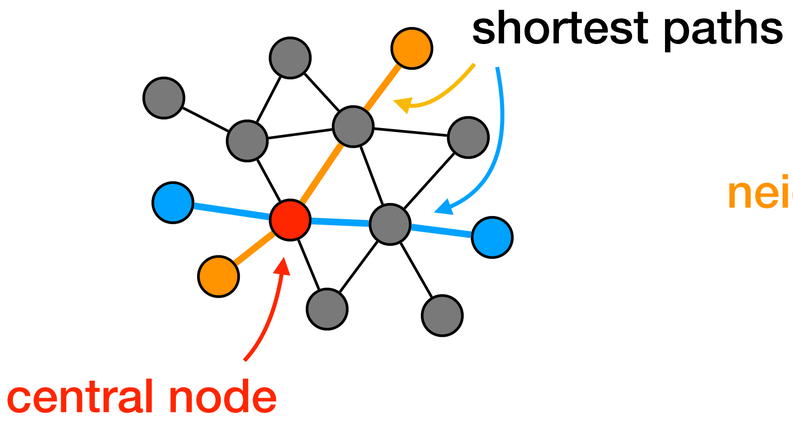
2. ***MVP*** computes the change induced by the new link on the feature values

3. ***MVP*** computes this change on the AS topology observed by every VPs

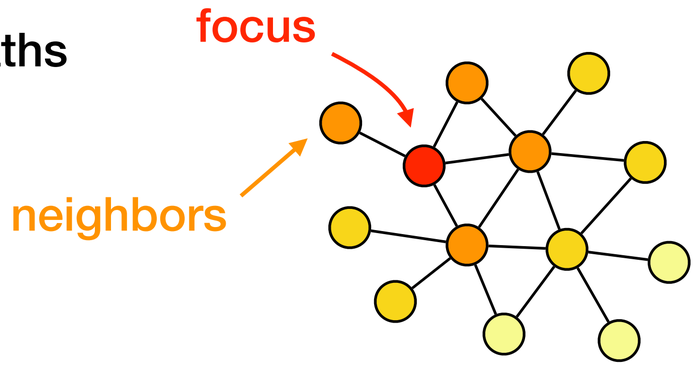


MVP uses a total of 20 topological features that can be divided into three categories

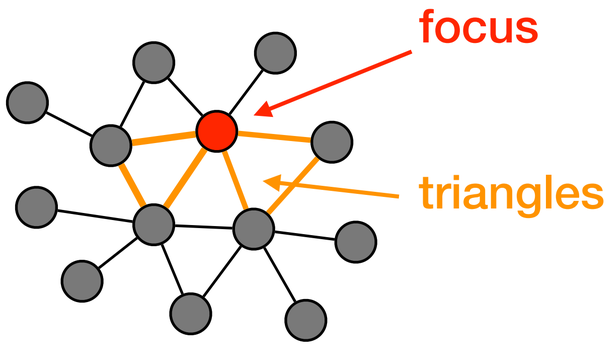
Node centrality



Neighborhood richness



Topological patterns



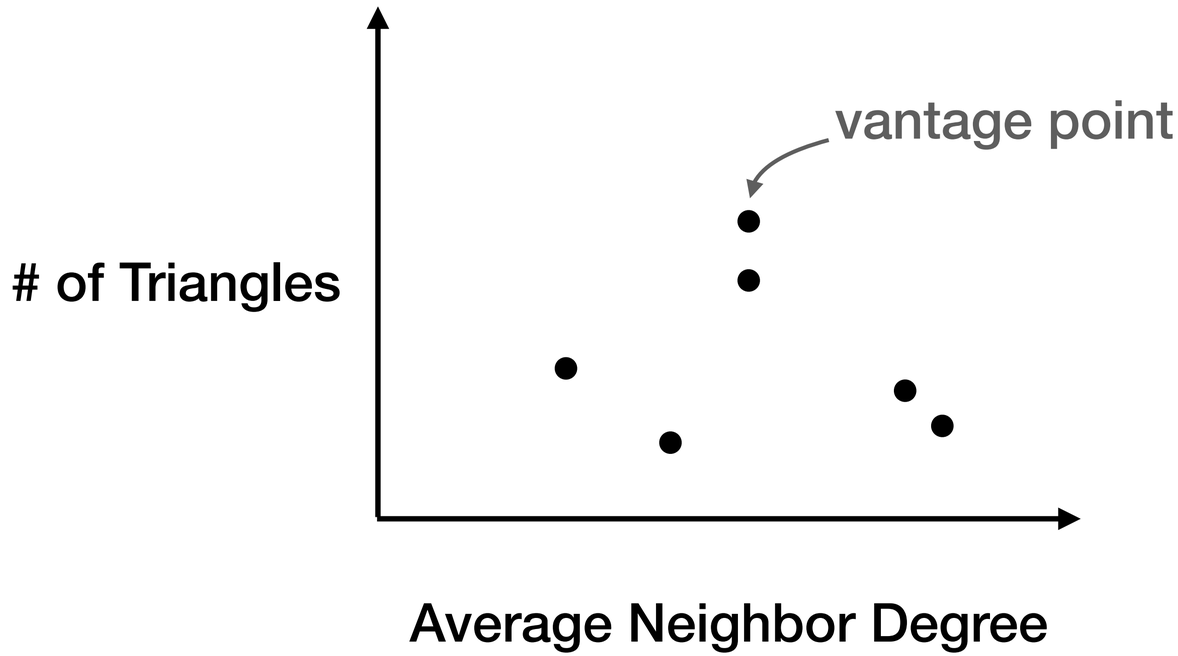
Goal: Select BGP Vantage Points (VPs) that
maximises utility and minimises volume of data

Step #1: Quantifying the observation of the VPs for past events

Step #2: **Measuring similarity between VPs for every event**

Step #3: Selecting a set of dissimilar VPs

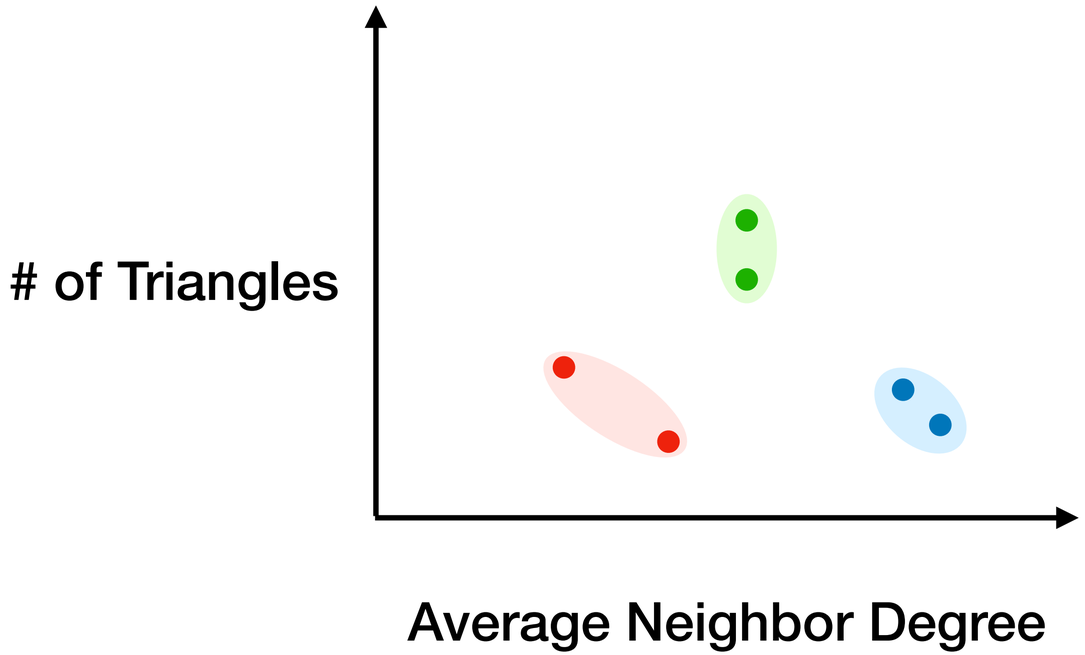
MVP clusters the vantage points for each event based on their feature values



We only depict two dimensions for clarity

MVP clusters the vantage points for each event based on their feature values

We use K-means to compute clusters



We only depict two dimensions for clarity

Goal: **Select BGP Vantage Points (VPs) that maximises utility and minimises volume of data**

Step #1: **Quantifying the observation of the VPs for past events**

Step #2: **Measuring similarity between VPs for every event**

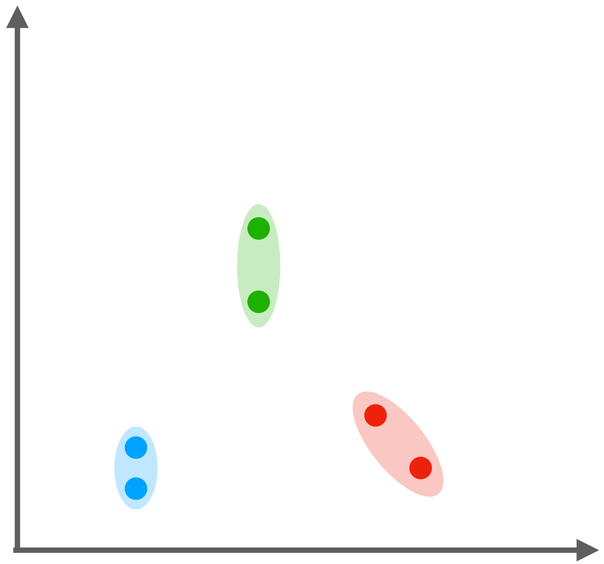
Step #3: **Selecting a set of dissimilar VPs**

MVP clusters the VPs for 750 new AS link events

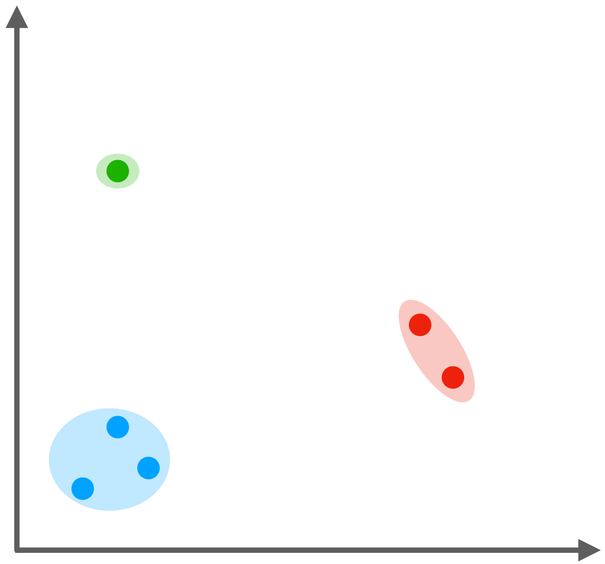
1. ***MVP*** takes new AS links that are visible by at least ten VPs and scattered on the Internet

2. ***MVP*** runs the clustering for every of the 750 new AS links independently

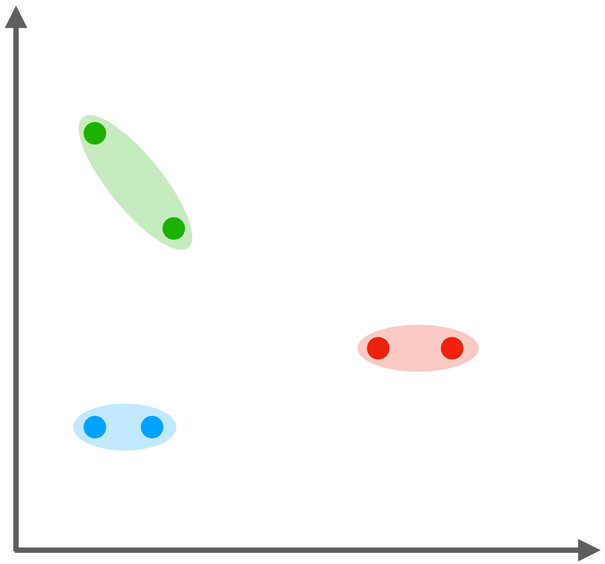
MVP uses a pair-wise similarity score that estimates similarity across all events for a pair of VPs



Event 1



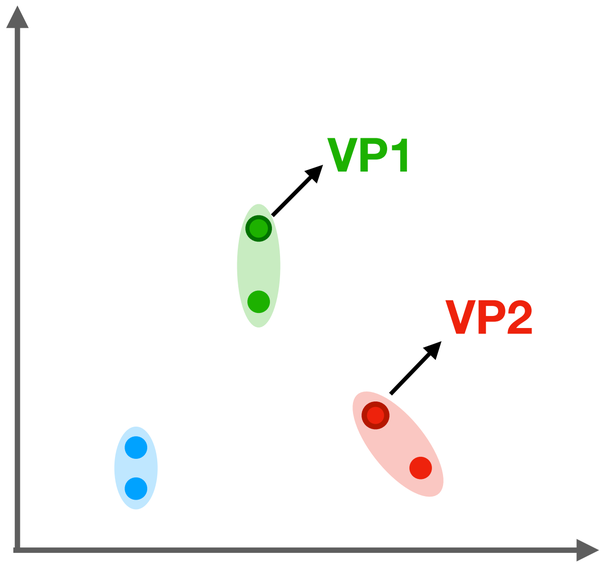
Event 2



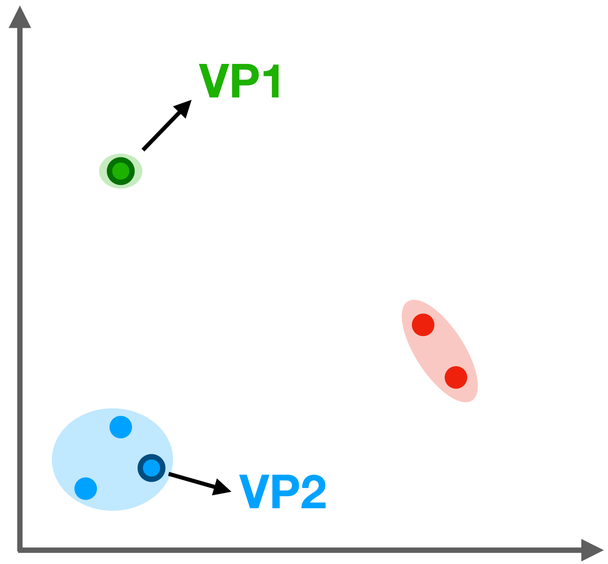
Event 3

MVP uses a pair-wise similarity score that estimates similarity across all events for a pair of VPs

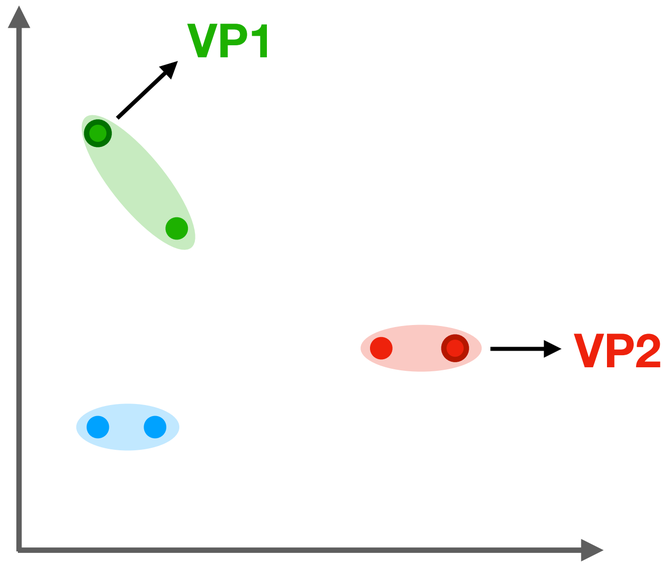
Similarity score (VP1 , VP2) : 0



Event 1



Event 2

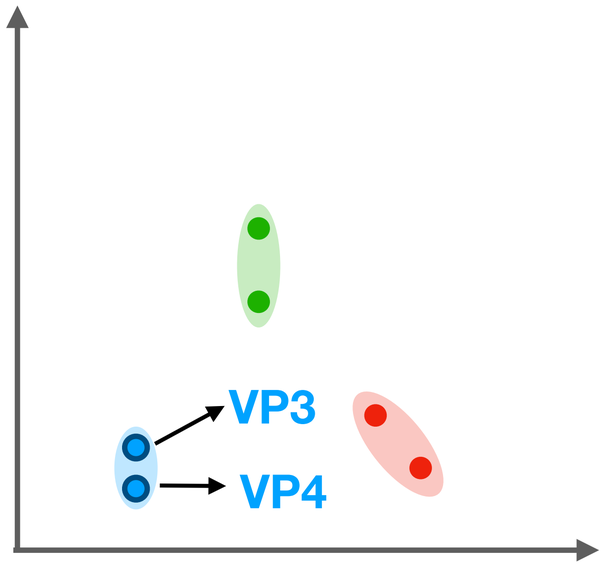


Event 3

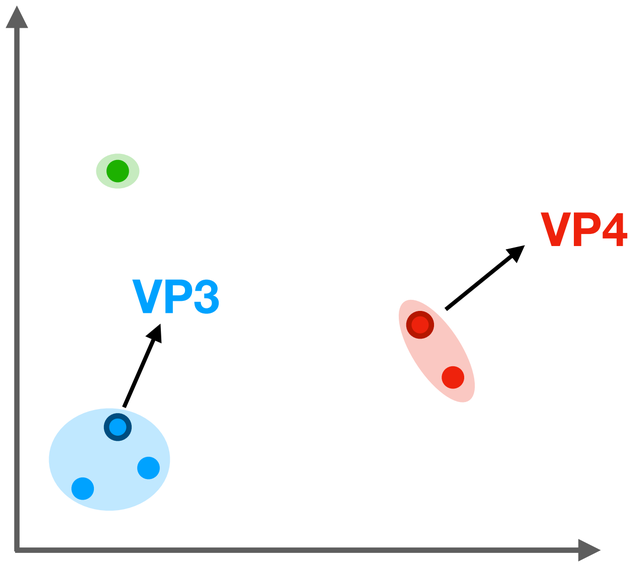
MVP uses a pair-wise similarity score that estimates similarity across all events for a pair of VPs

Similarity score (VP1 , VP2) : 0

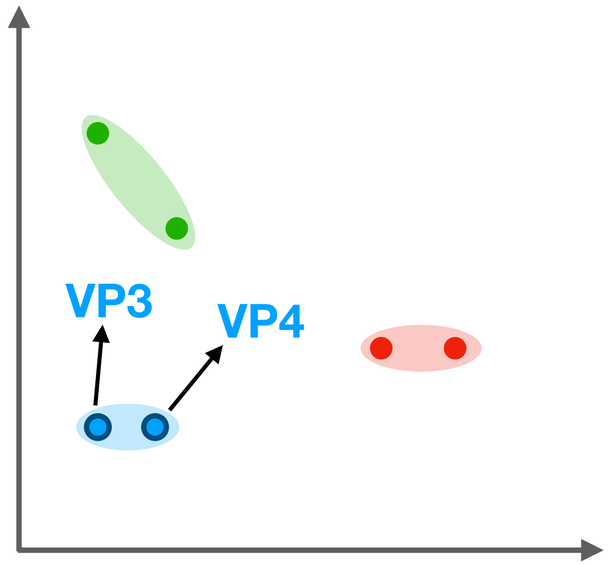
Similarity score (VP3 , VP4) : 0.67



Event 1



Event 2



Event 3

MVP uses the similarity score
to greedily build a set of dissimilar vantage points

MVP uses the similarity score
to greedily build a set of dissimilar vantage points

1. ***MVP*** first selects the most similar VP

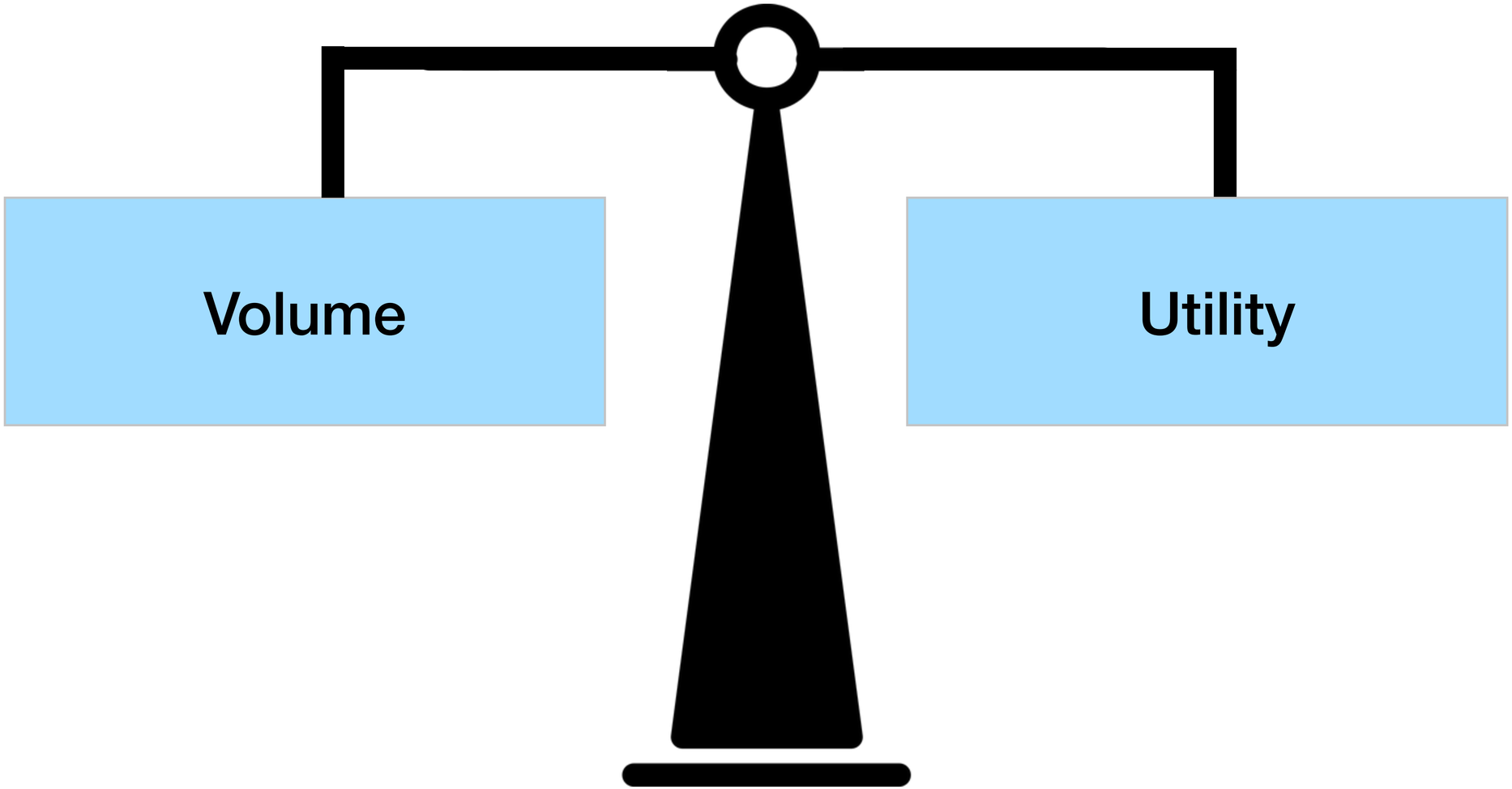
MVP uses the similarity score
to greedily build a set of dissimilar vantage points

1. ***MVP*** first selects the most similar VP

2. ***MVP*** selects the VP that is the most dissimilar
with the ones already selected


repeat

We evaluate *MVP* on the tradeoff between volume and utility



We evaluate *MVP* on three use cases

discovered AS links

% detected hijacks and new AS links

% detected transient paths

We evaluate *MVP* on three use cases and compare it to three baselines

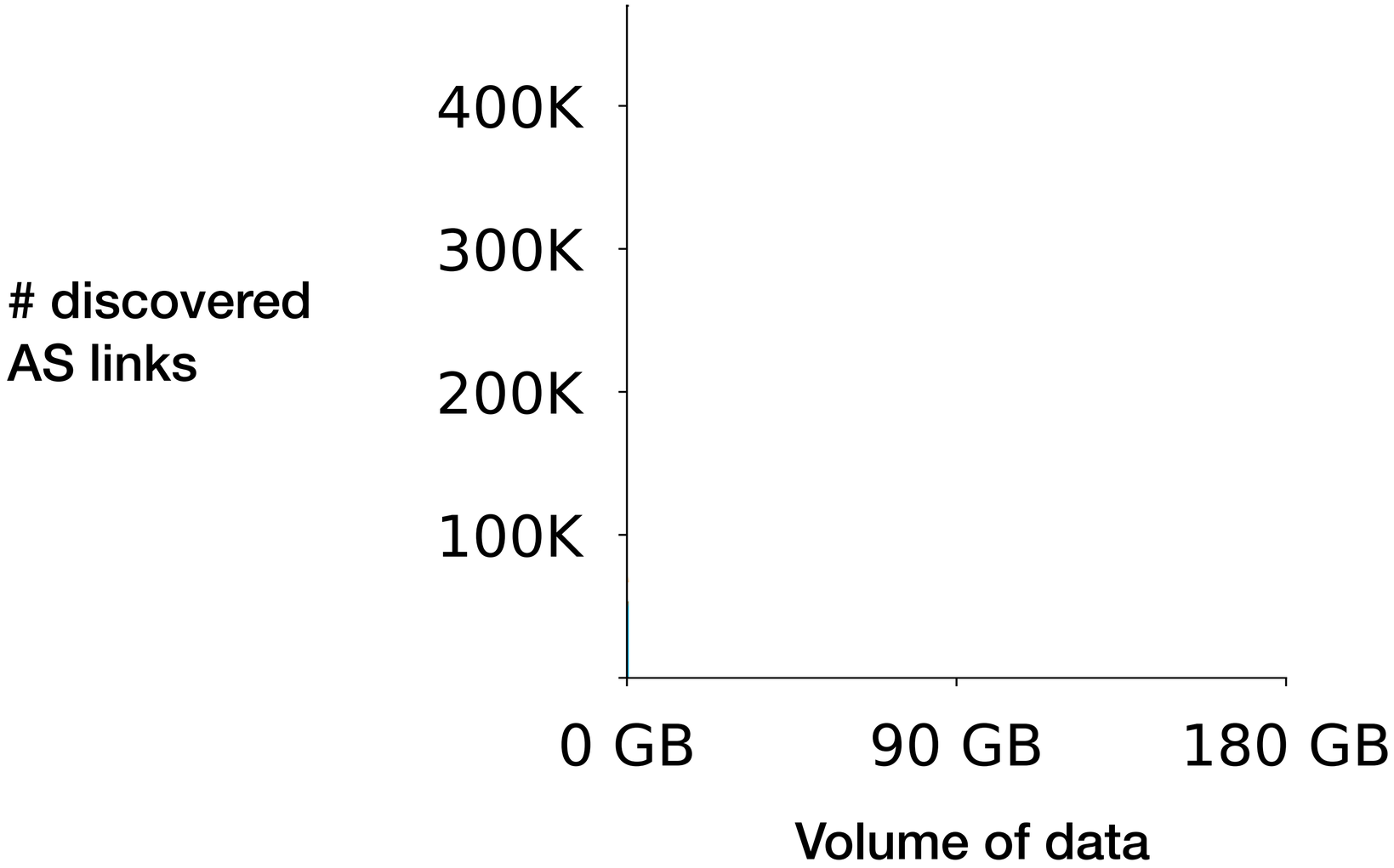
	# discovered AS links	% detected hijacks and new AS links	% detected transient paths
MVP			
Random			
Distance-based			
Max AS links			

MVP always selects VPs that exhibit the best tradeoff between volume and utility of the data

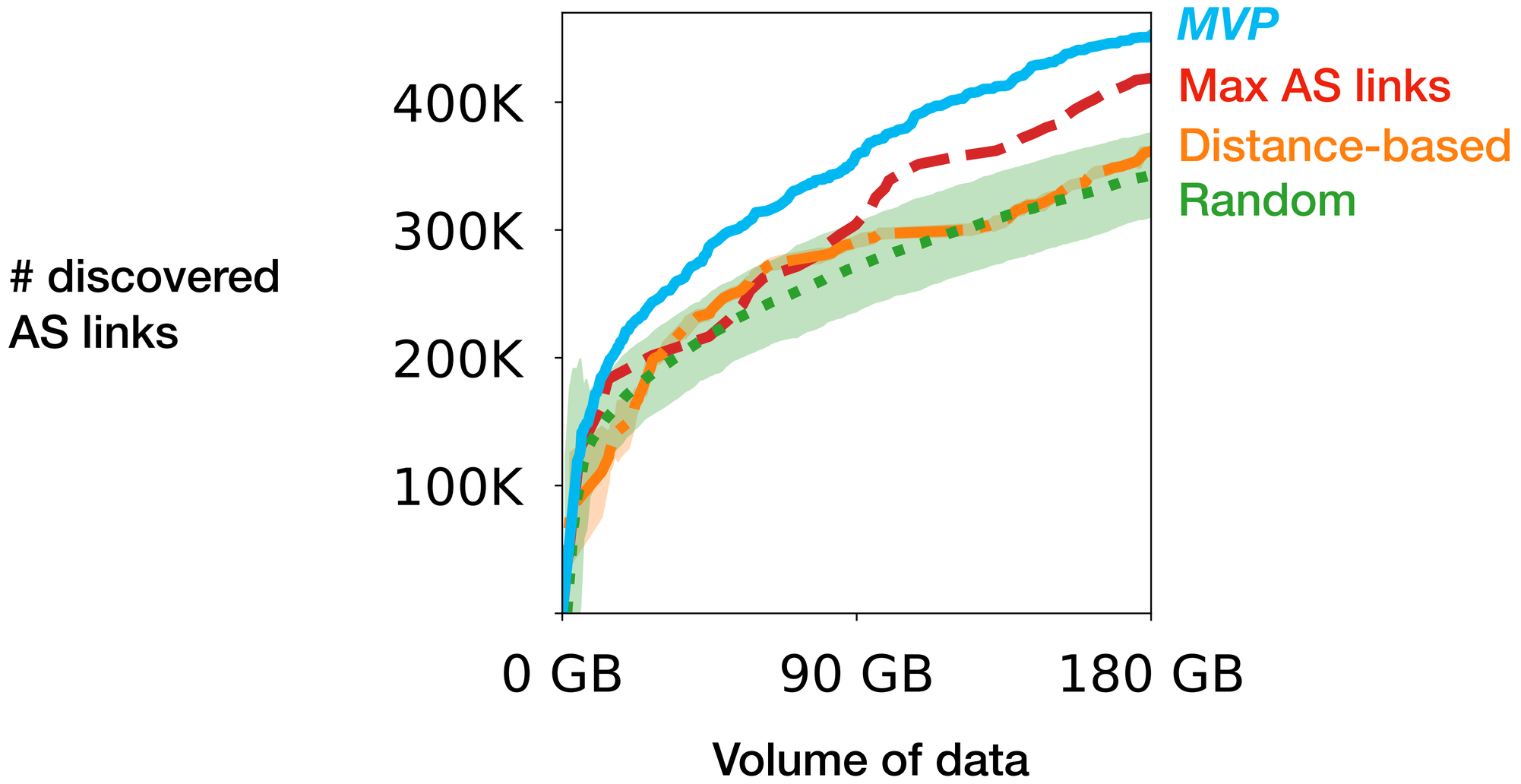
	# discovered AS links	% detected hijacks and new AS links	% detected transient paths
MVP			
Random			
Distance-based			
Max AS links			

MVP always selects VPs that exhibit the best tradeoff between volume and utility of the data

	# discovered AS links	% detected hijacks and new AS links	% detected transient paths
MVP			
Random			
Distance-based			
Max AS links			


400K
300K
200K
100K
discovered
AS links
0 GB
90 GB
180 GB
Volume of data

MVP **selects vantage points that see 300k AS links with 58% less BGP routes compared to a random selection**



MVP always selects VPs that exhibit the best tradeoff between volume and utility of the data

	# discovered links	% detected hijacks and new links	% detected transient paths
MVP			
Random			
Distance-based			
Max AS links			

See our poster at IMC'22

Try it out!

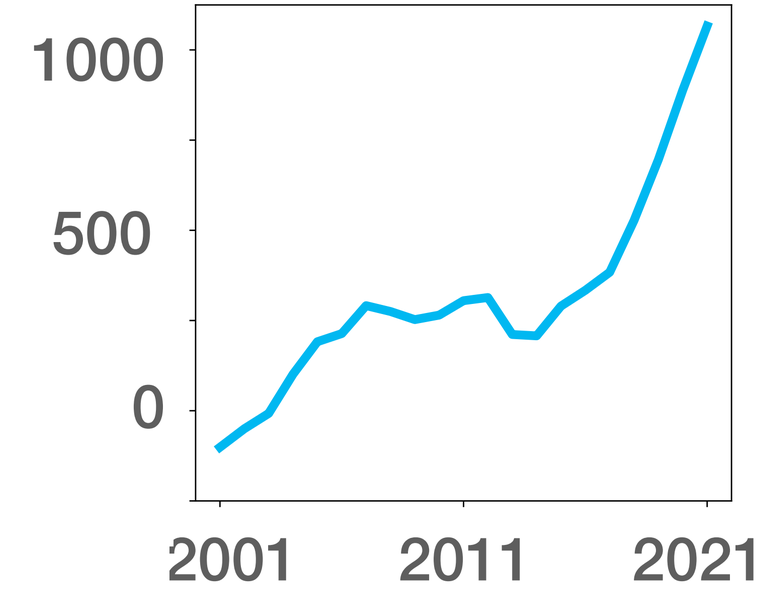
An alpha version of ***MVP*** is running and waits for requests

curl -d ‘volume=5GB’ http://5.161.124.63/mvp

This exponential increase is the consequence of

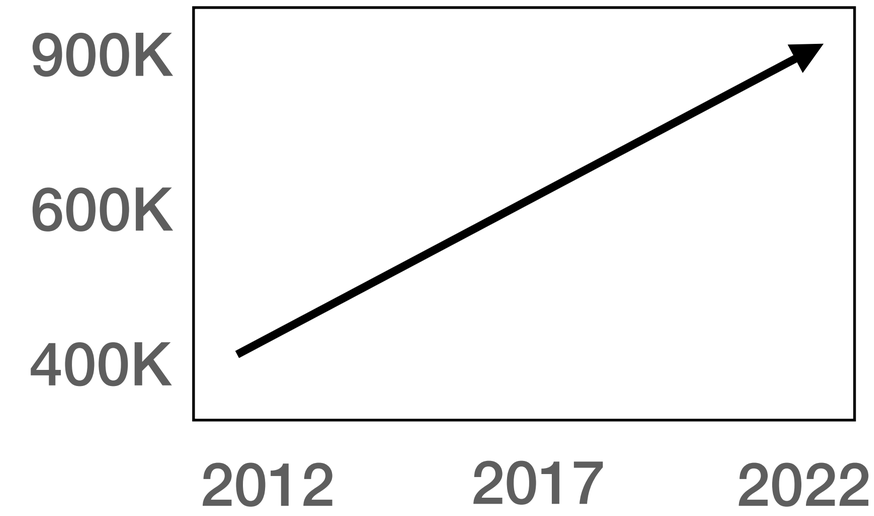
1. More vantage points (VPs) being deployed

Number of VPs (RIPE RIS)



2. More prefixes being advertised

Number of prefixes



Distance between VPs is not a good indicator of their similarity

AS links observed from only one of the two VPs