BGPfuse: Using visual feature fusion for the detection and attribution of BGP anomalies

Mr. Stavros Papadopoulos
CERTH/ITI Greece

Presentation Outline

- Feature fusion
  - Current approaches
  - Visual vs Automated analysis
- BGP
  - Basic background
  - Feature definition
- BGPfuse: Visual feature fusion
  - 1) Parallel Coordinates User Interface
  - 2) Feature Graph view
  - 3) Combined Graph view
- Implementation in real life scenario

2. Feature fusion

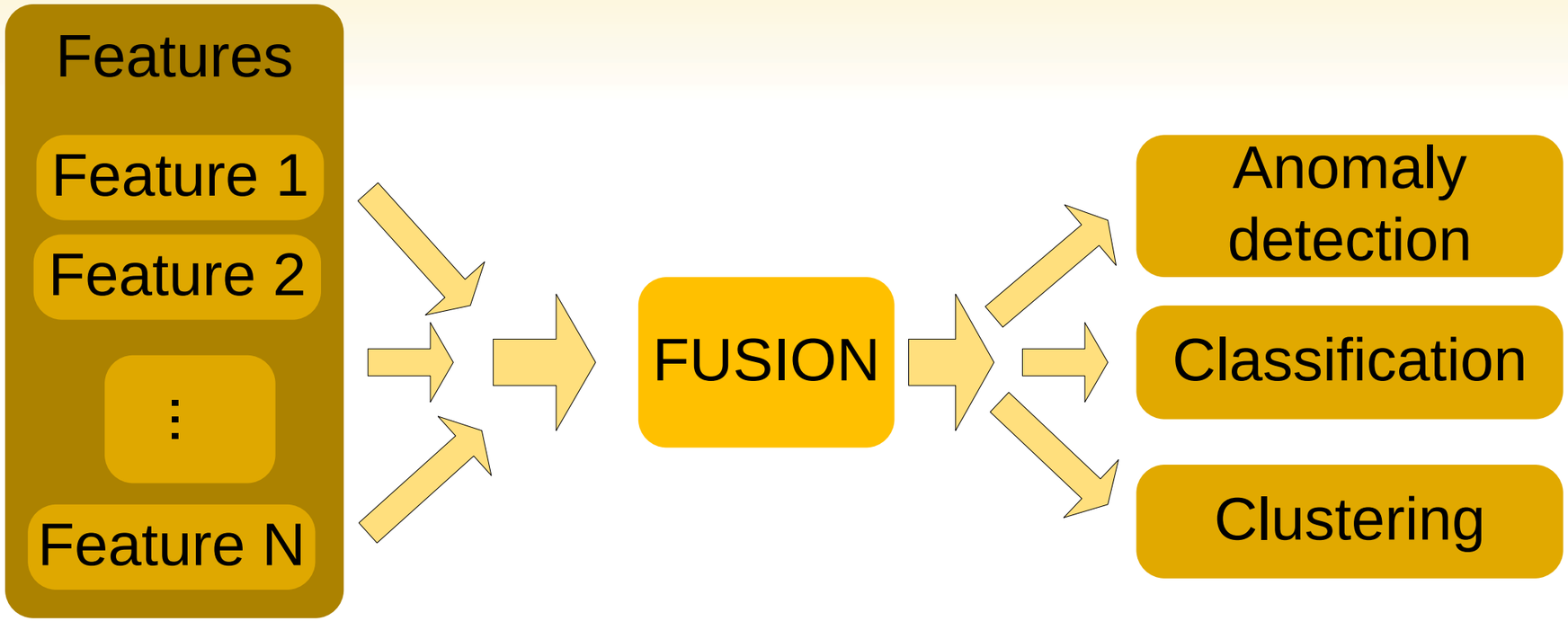


Fig. Feature Fusion procedure

Feature fusion background

- Appropriate combination of a set of features for classification, clustering or anomaly detection
- Examples of algorithmic fusion frequently used:
  - Weighted sum
  - Geometric/harmonic/generalized means
  - Support Vector machine (SVM)
  - Neural Networks (NN)
  - Combinations of Multiple classifiers

Visual vs Automated analysis

- Visual analysis
  - + uses power of human visual system
  - + user-guided analysis possible
  - + detect interesting features and parameter selections
  - + understand results in context
  - - limited dimensionality
  - - often only qualitative results
- Automated analysis
  - + hardly any interaction required (after setup)
  - + scales better in many dimensions
  - + precise results
  - - needs precise definition of goals
  - - result without explanation
  - - computationally expensive

2. BGP

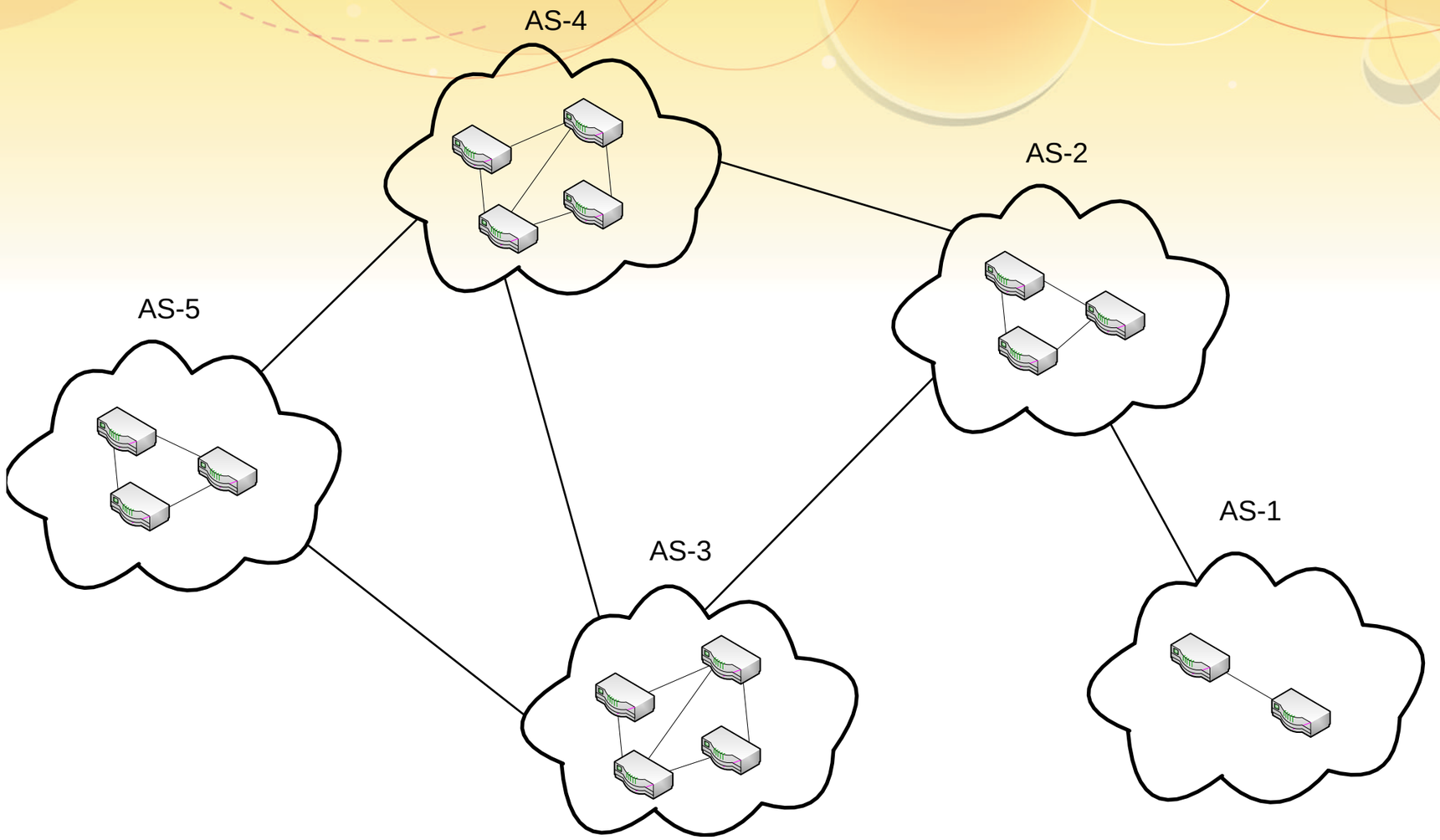


Fig. Interconnections between Autonomous Systems (AS)

Basic background in BGP

- BGP stands for Border Gateway Protocol
  - De facto protocol used today for the exchange of routing informtion between Autonomous Systems (AS)
- AS is a is a collection of routers under the control of one network operator
  - Each AS is assigned a unique number
  - Each Internet AS has a hosting country, i.e. majority of its network infrastructure are located.

Basic background in BGP

- The basic component of BGP is the BGP message
  - BGP announcement
    - Contains the owner of the announced prefix and the AS path followed to reach that prefix.
    - {<prefix> : <AS-path>}, e.g. <164.25.48.8/24> : <AS-35 AS-2 AS-5>
  - BGP withdrawal
    - Provides information regarding the loss of visibility of a prefix from the routing tables

BGP vulnerabilities

- Lack of validity mechanisms
  - Vulnerable to attacks from ASes
- Hence, ASes can propagate false BGP information due to either misconfiguration or malicious intention.
- Basic types of BGP hijacks:
  - Prefix hijacking
  - Path hijacking (e.g. Man In the Middle)

BGP analysis - Definition of Features

- The sequence of the traversed ASes is highly dependent on their geographic presence.
- Analyzing the geographic coherence of the AS-paths could lead to anomaly detection
- Transform the AS-paths of the BGP announcements to Country-Paths:

AS-path

Counrty-path

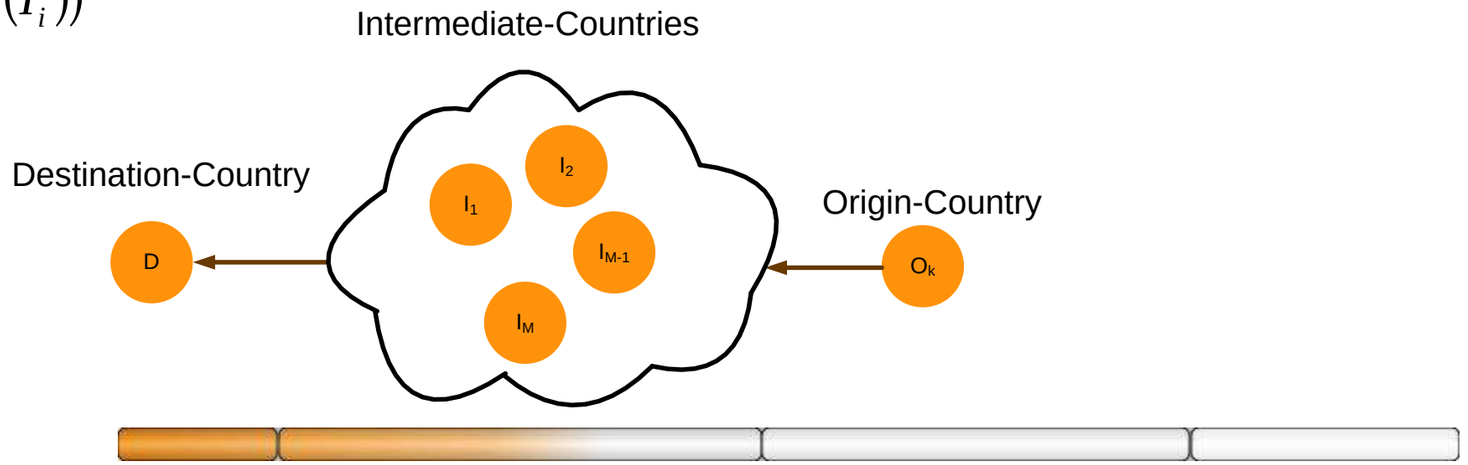
BGP analysis - Definition of Features

1. CAP: The probability of appearance of the *Intermediate-Country* within the AS-path towards the specific *Origin-Country*

$$CAP(I_i) = \frac{N(I_i)}{\sum_{i=1}^{N} N(I_i)}, \text{where } N(I_i) \text{the number of appearances of country } I_i$$

2. CAPZ: The Z-score of the aforementioned probability

$$CAPZ(I_i) = \frac{CAP(I_i) - E(CAP(I_i))}{\sigma(CAP(I_i))}, \text{where } E \text{ is the mean and } \sigma \text{ the standard deviation}$$



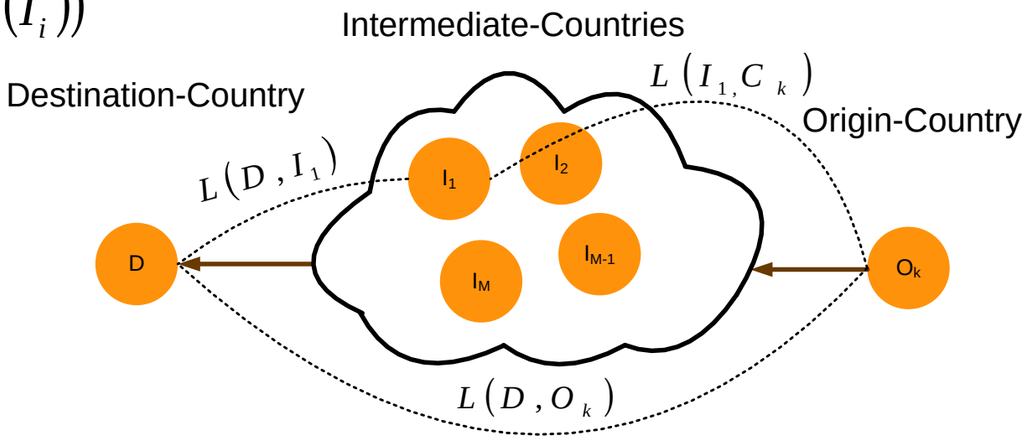
BGP analysis - Definition of Features

3. CGL: The geographic deviation introduced by the Intermediate- Country within the AS-path towards the specific Origin-Country

$$CGL(I_i) = \frac{L(D, I_i) + L(I_i + O_k)}{L(D, O_k)}, \text{where } L(X, Y) \text{the distance between countries } X \text{ and } Y$$

4. CGLZ: The Z-score of the aforementioned CGL feature

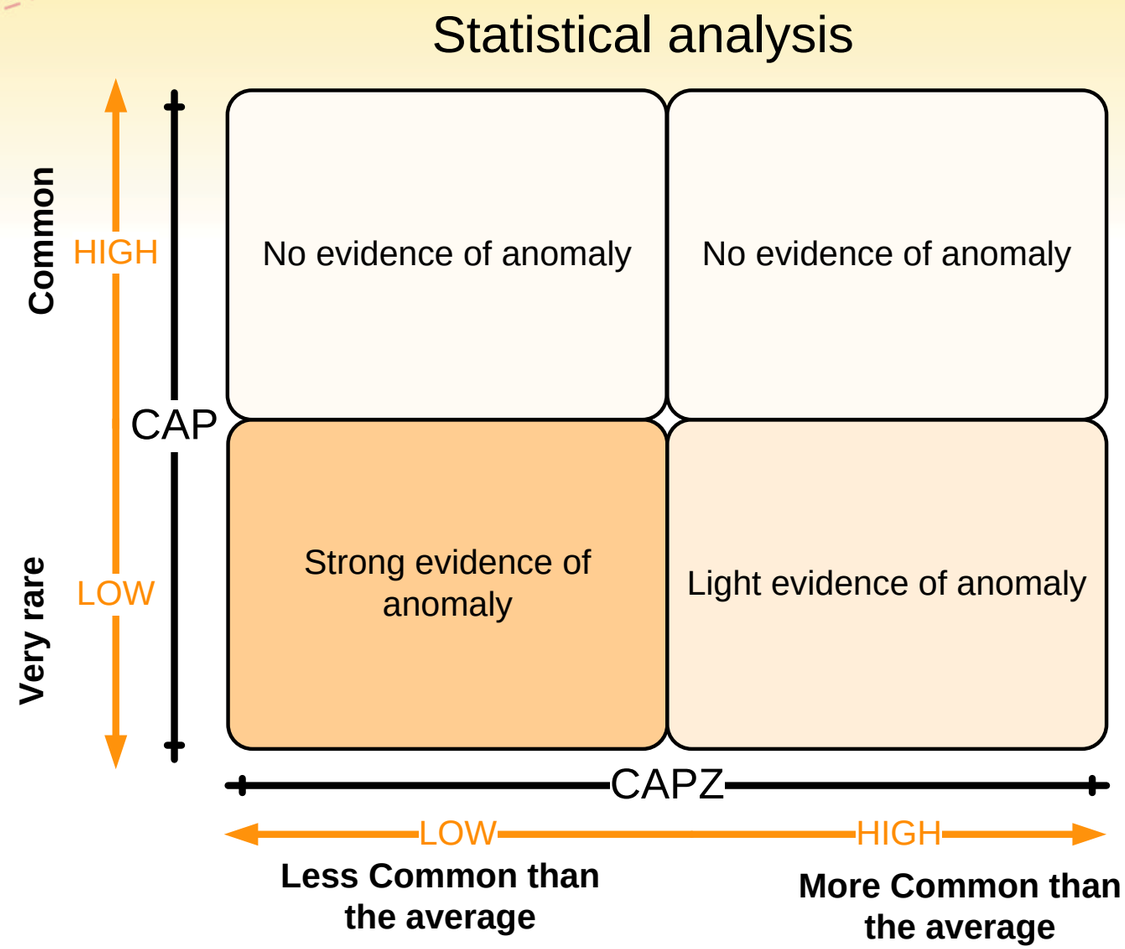
$$CGLZ(I_i) = \frac{CGL(I_i) - E(CGL(I_i))}{\sigma(CGL(I_i))}, \text{where } E \text{ is the mean and } \sigma \text{ the standard deviation}$$

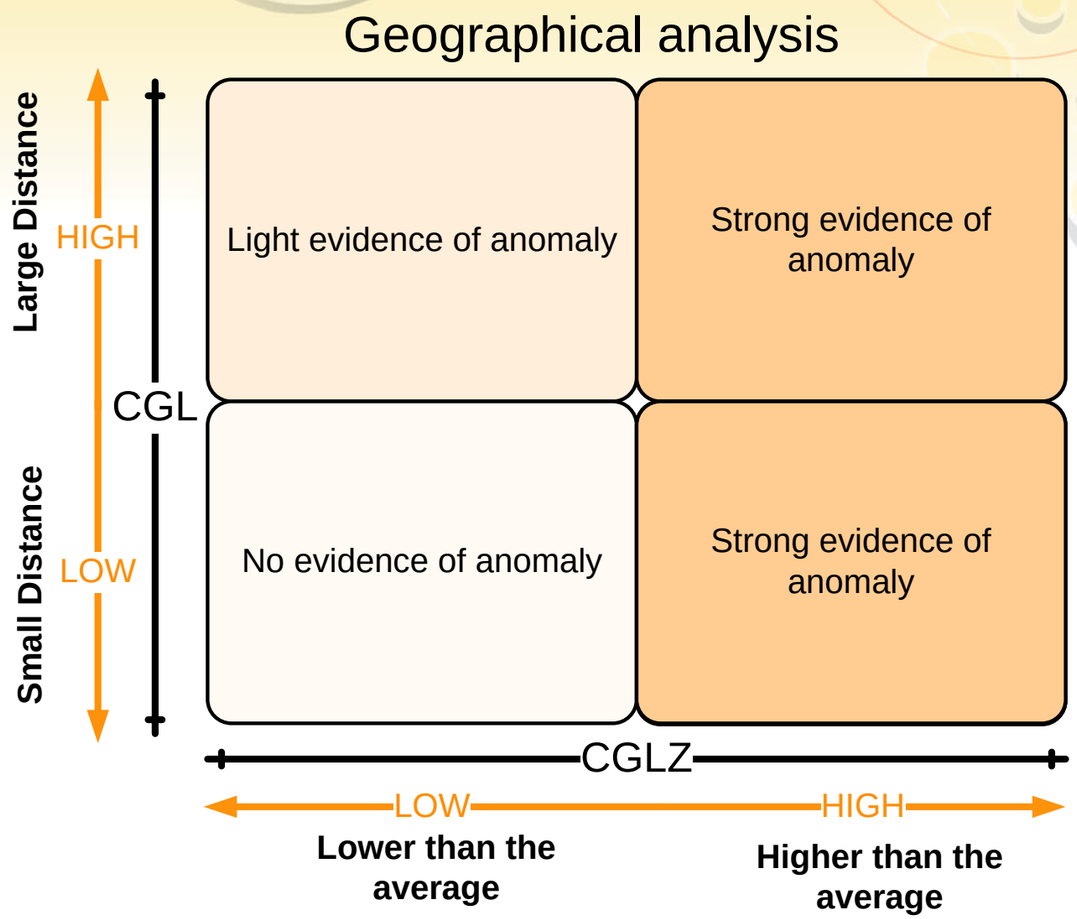


BGP analysis - Definition of Features

- The values of the overall BGP path-change event are equal to the corresponding values of the **less probable** and **more deviating** Intermediate-Country.
- The most suspicious ASes in the path are the ones that are hosted in this **outlying** Intermediate-Country.
- Thus, the aforementioned features can be eventually **defined on per Intermediate-AS basis, for each Origin-Country** appearing in the BGP announcements (Destination-Country is static)

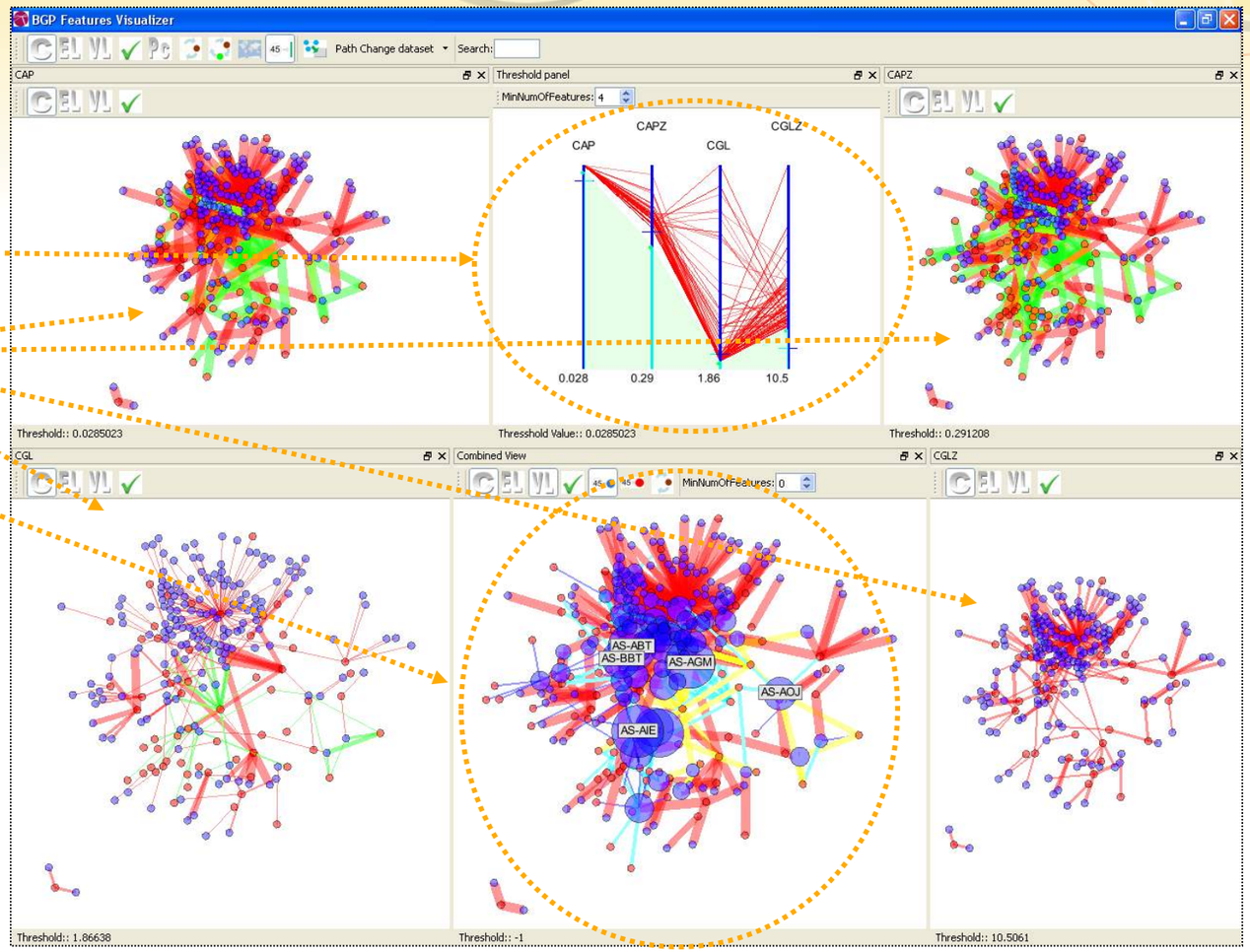
BGP analysis - Definition of Features





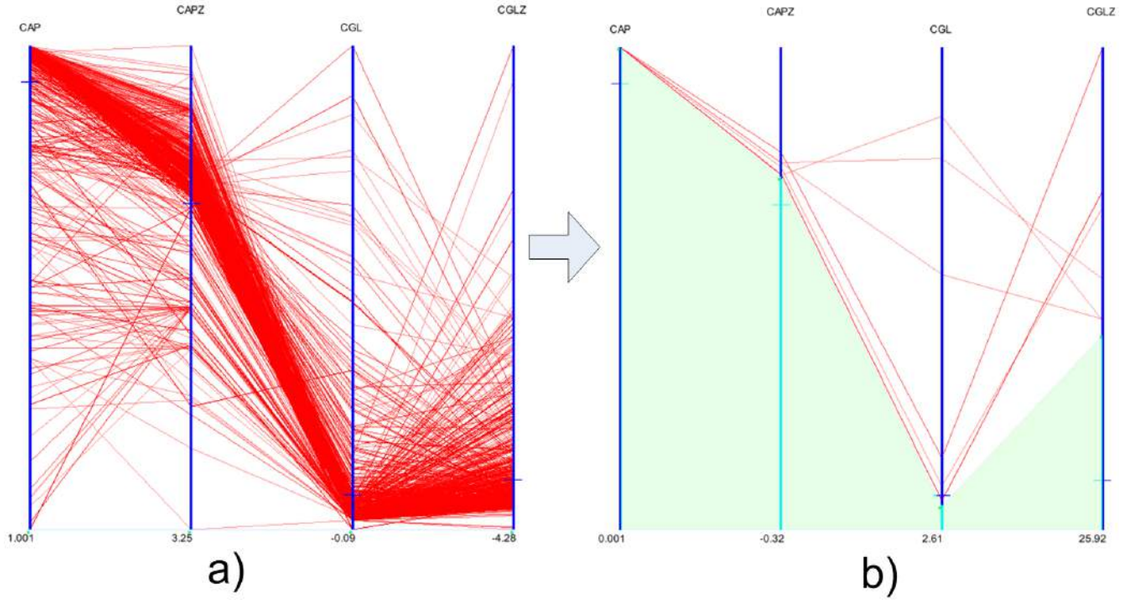
3. BGPfuse: Visual feature fusion

- Three main components:
  1. Parallel Coordinates User Interface
  2. Feature Graph view
  3. Combined Graph view



Parallel Coordinates User Interface

- Parallel coordinates visualization
  - Enhanced with filtering capabilities using sliders
- Important values (e.g low CAP and high CGL) are positioned at the upper part of the view
- The events whose at least one feature value is below the predefined thresholds are omitted from the visualization



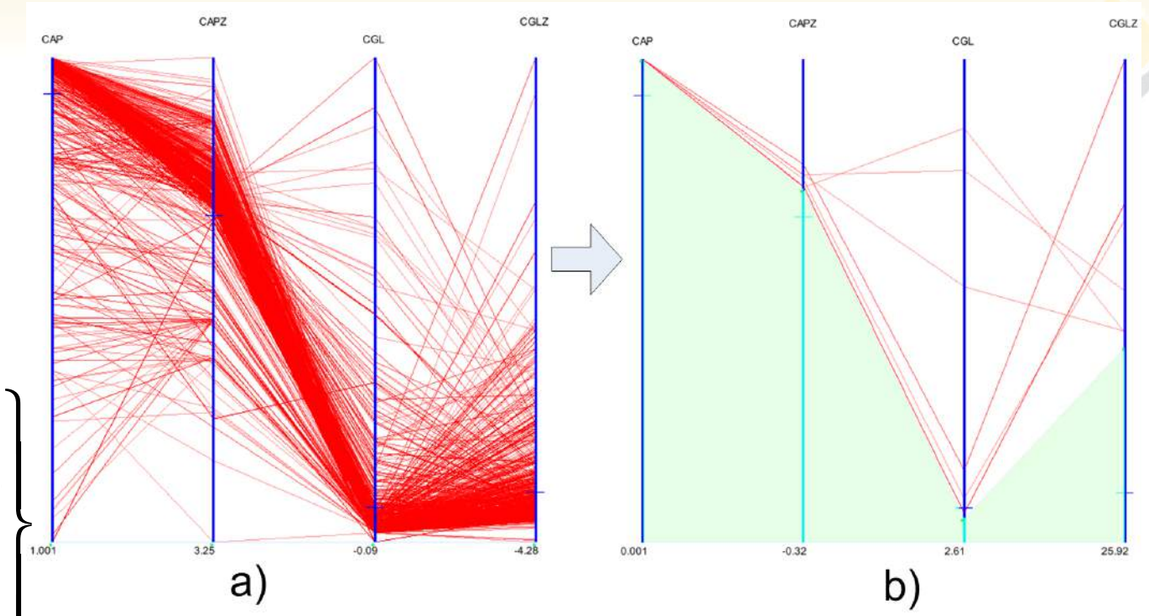
Parallel Coordinates User Interface

- Slider position represents feature weights
- Slider threshold values:

$$T = \{ t_i \mid i \in \{CAP, CAPZ, CGL, CGLZ\} \}$$

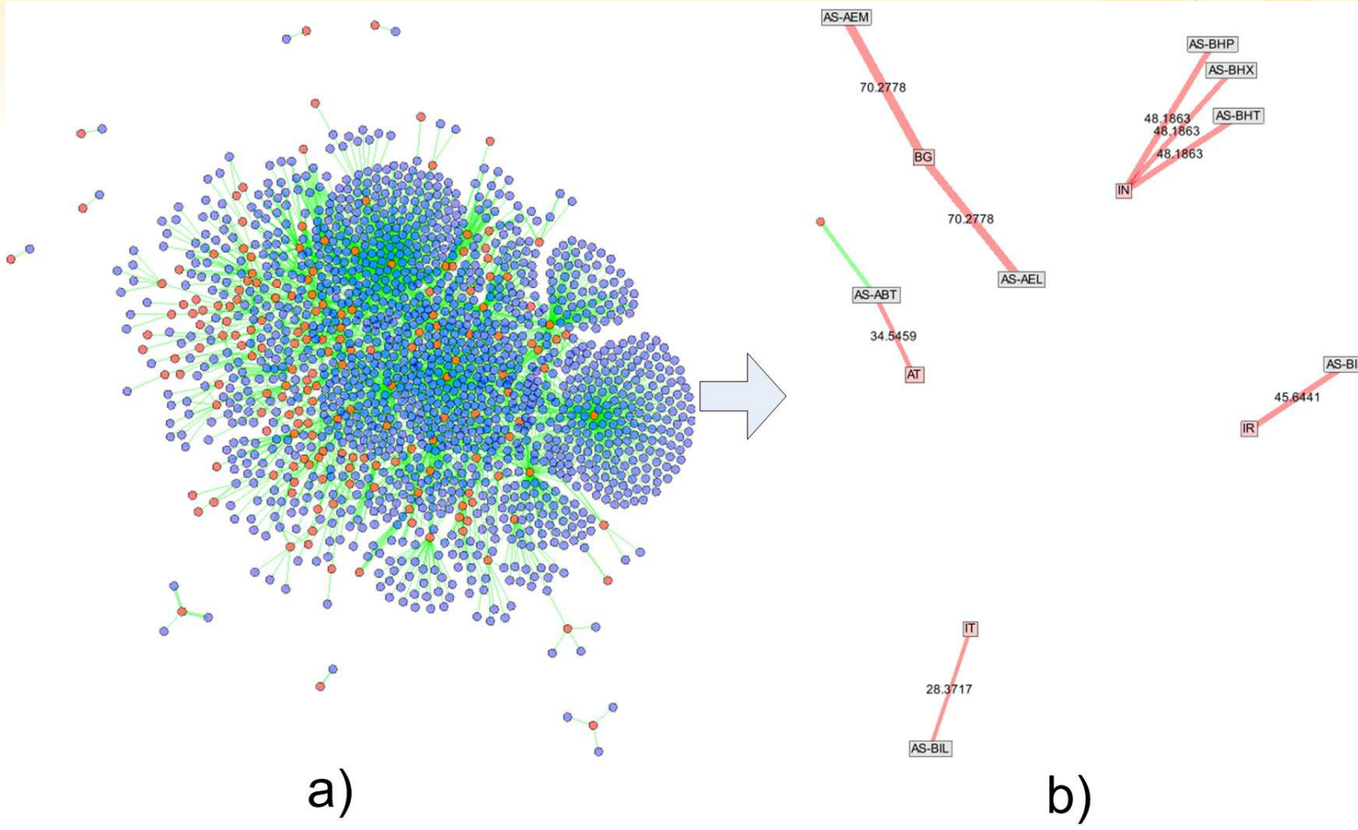
- Filtering function:

$$f_T(t_i, e_i^j) = \begin{cases} 1 & \begin{cases} e_i^j(w) > t_i, \forall i \in \{CGL, CGLZ\} \\ e_i^j(w) < t_i, \forall i \in \{CAP, CAPZ\} \end{cases} \\ 0 & , \text{for all other cases} \end{cases}$$



Feature Graph view

- Graph based visualization of each feature
- Edge = Path change event
- Red vertices = Origin Countries
- Blue vertices = Intermediate ASes
- Visualization of:
  - Intermediate ASes and source Countries involved in suspicious events
  - relationships that may exist between actors



Feature Graph view

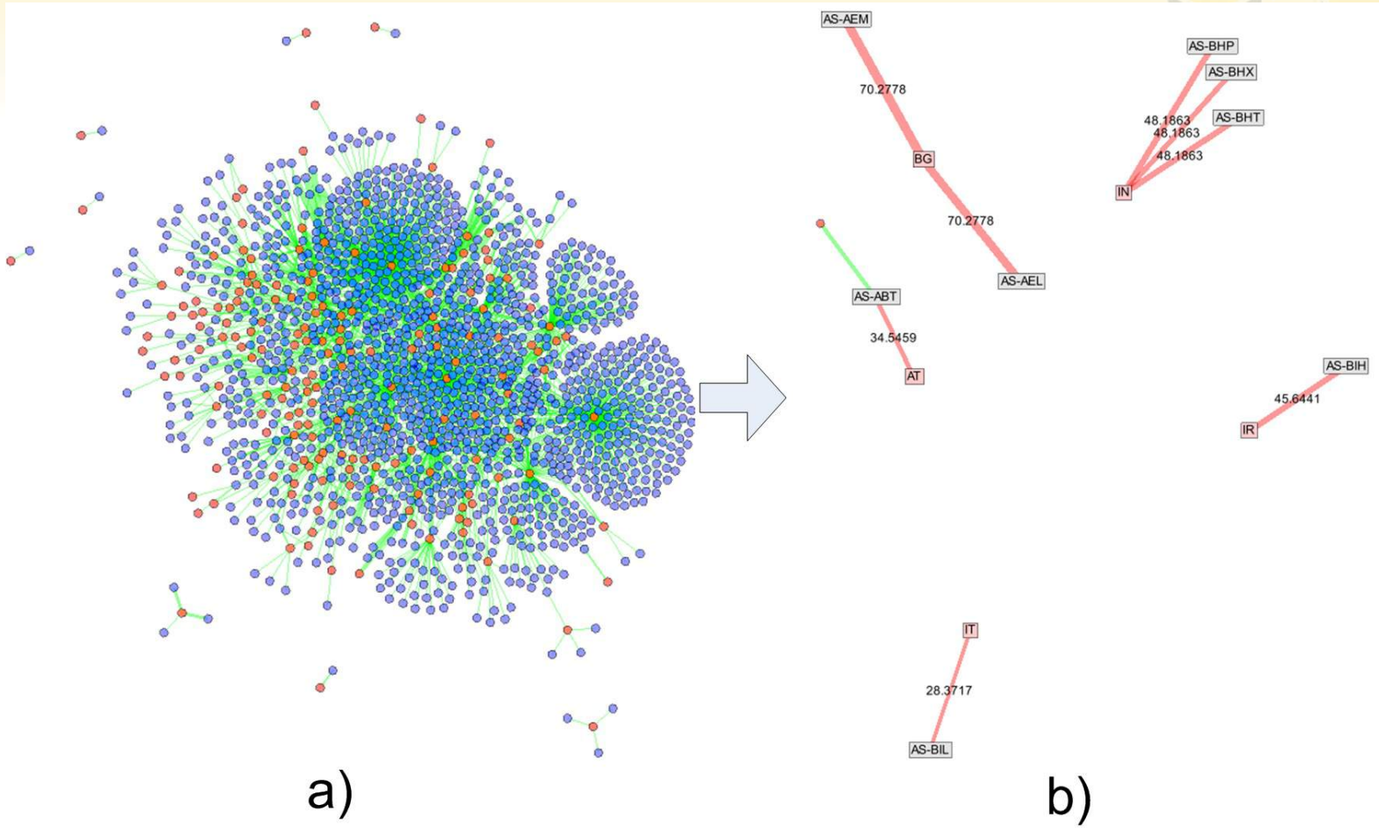
- Width of edges proportional to the importance of the feature value
- Set of edges:

$$E_i = \left\{ e_i^j \mid \forall f_T\left(t_i, e_i^j\right) = 1 \right\}$$

- Set of vertices:

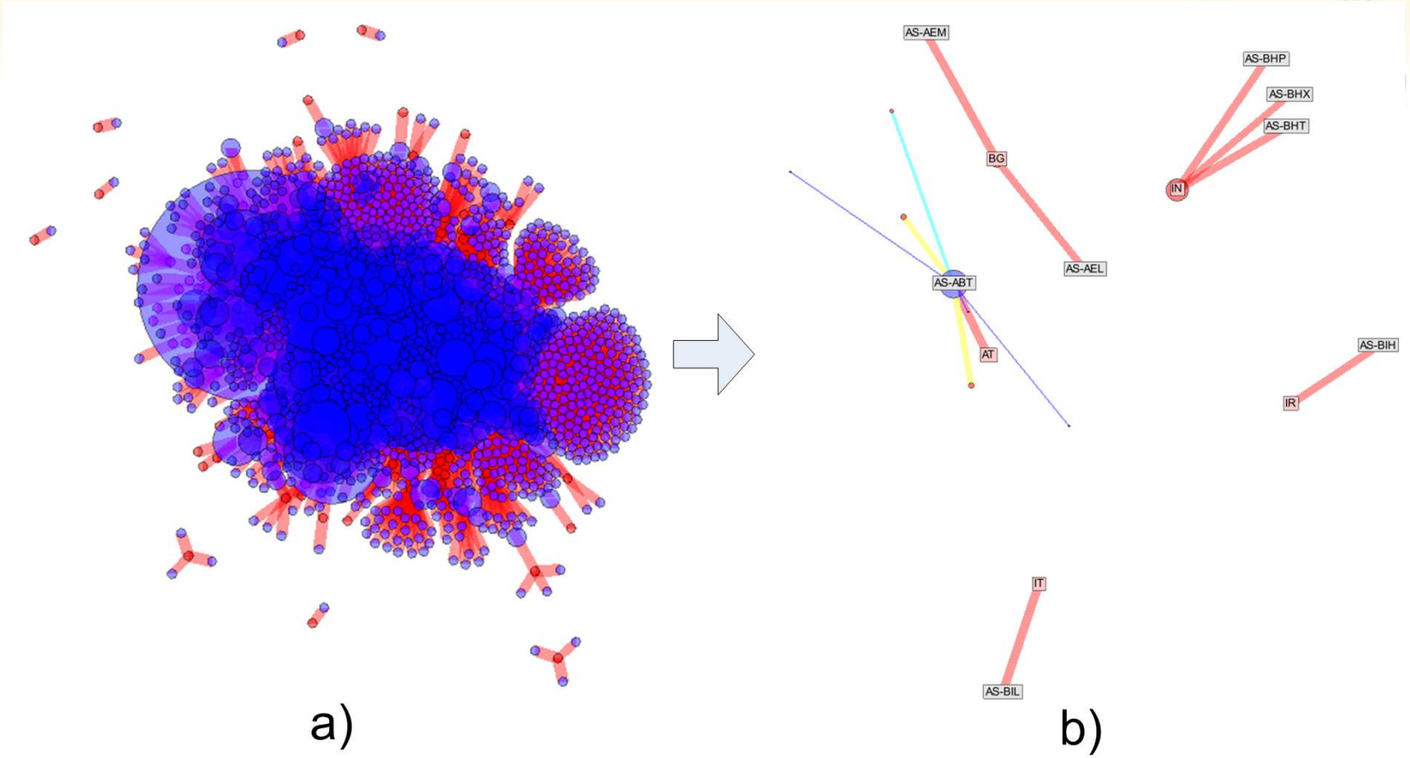
$$V_i = \left\{ v_i^j \mid v_i^j \in e_i^k, \forall e_i^k \in E_i \right\}$$

Where e_i^j is a path change event caused by an intermediate AS



Combined Graph view

- The Combined Graph view is a fused graph of all the individual feature graphs
- It highlights structural similarities between the individual feature graphs so as to:
  - Highlight the suspicious BGP path change events across any number of features
  - Reveal possible participation of an actor in multiple events, visible from multiple features.



Combined Graph view

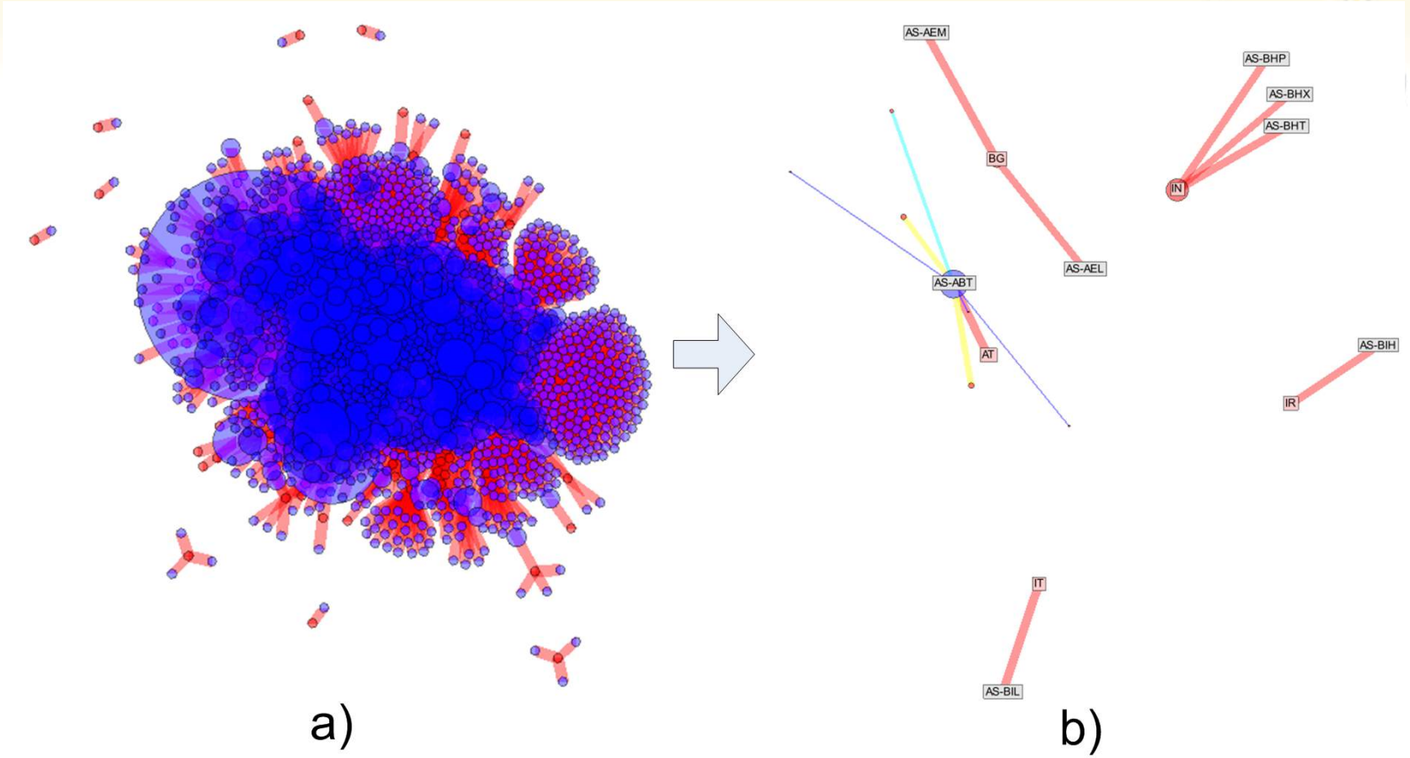
- The Combined Graph view is a graph $GC = \{VC, EC\}$ where:
  - The set of edges:

$$EC = \bigcup_i E_i$$

  - The set of vertices

$$VC = \bigcup_i V_i$$

- Thus showing the events that remain in all features after the application of filtering



Combined Graph view

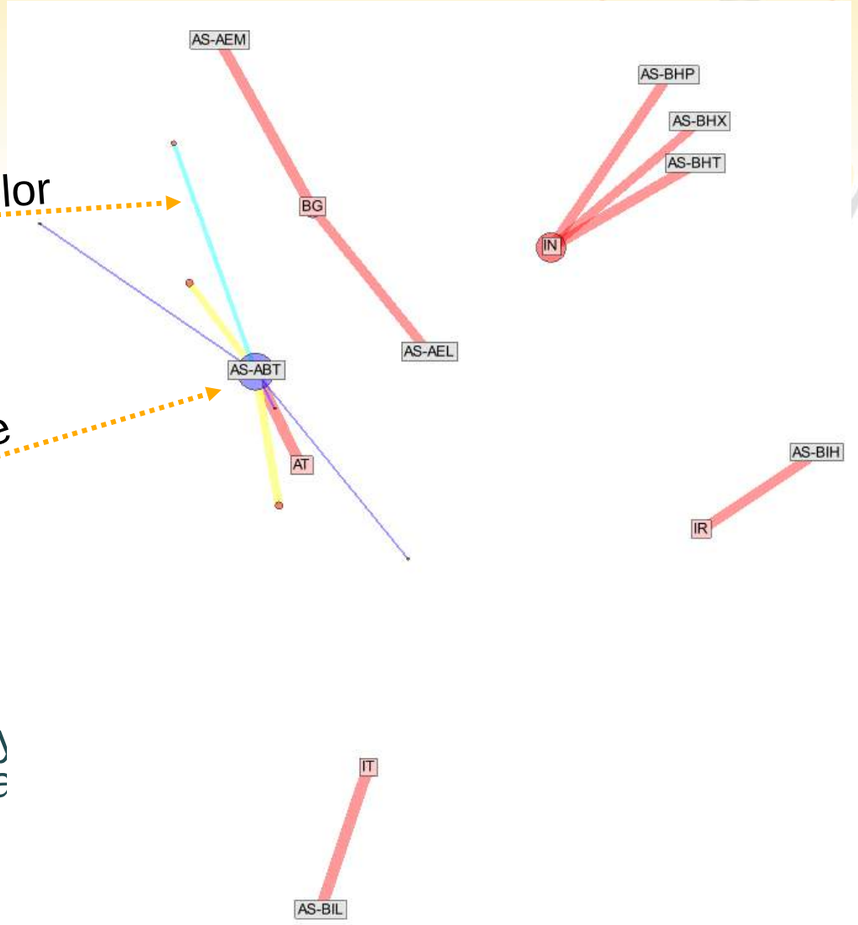
- Additional features are defined in order to highlight interesting events:
  - ***Degree of existence*** of each edge (path change event)

$$D_e\left(ec^k\right)=\sum_i f_T\left(t_i,e_i^j\right),\forall e_i^j=ec^k$$

    - Measures how many features the corresponding event is visible from, after the application of filtering.
  - ***Degree of anomaly*** of a vertex (AS or Country)

$$D_v\left(vc^k\right)=\sum_j D_e\left(ec^j\right),vc^k\in ec^j,\forall ec^j\in EC$$

    - High Degree of anomaly of a vertex implies that it is involved in many BGP path change events visible from many features after the applica of filtering



4. Implementation in real life scenario

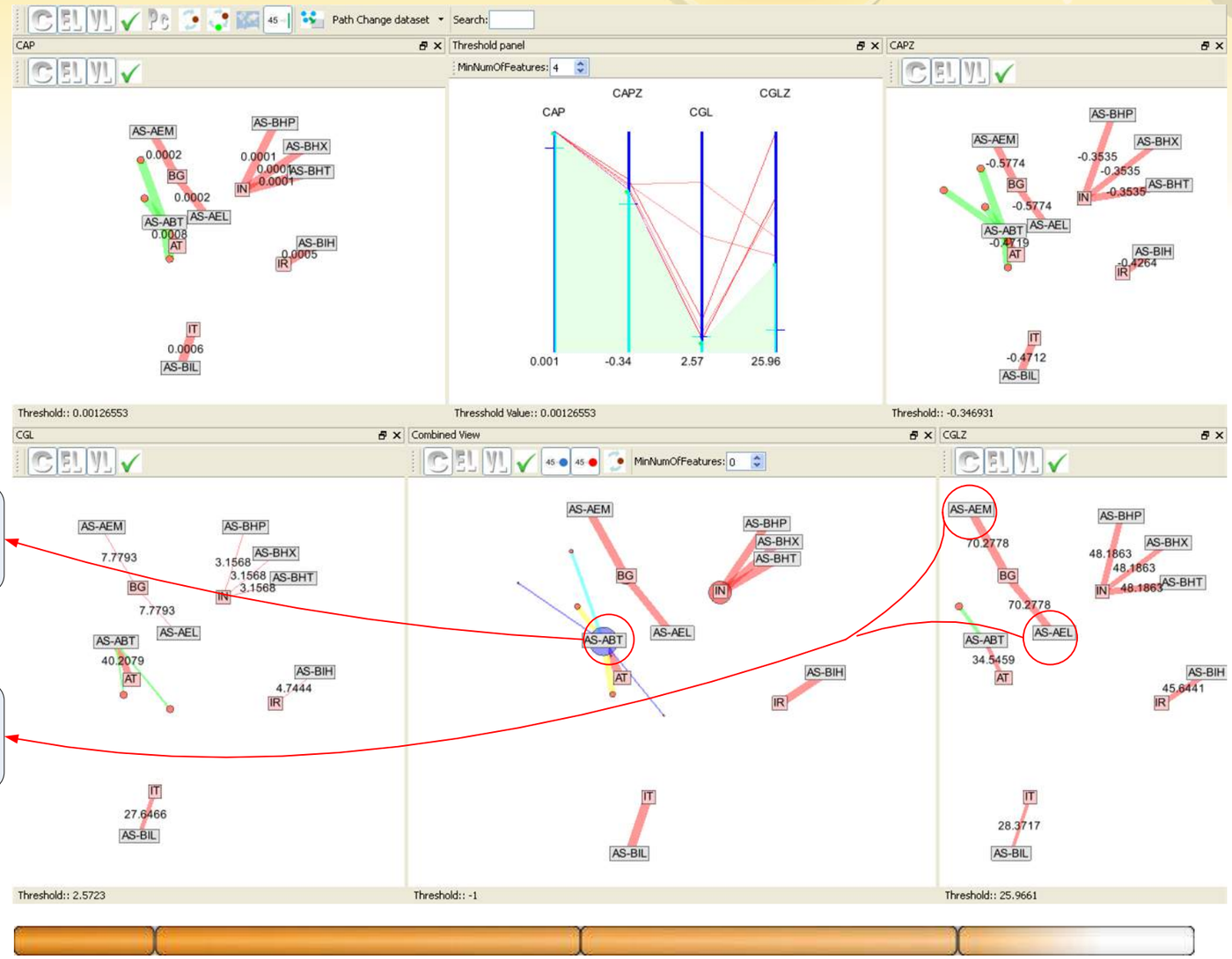
- Hijacking:
  - On August 20, 2011, a Russian telecommunication company (**Victim-AS**), reported to the North American Network Operators Group (NANOG) that five of its **prefixes had been hijacked**.
  - **False routes** were injected for the purpose of diverting Internet traffic through the **Hijacking-AS** located in US.
- Countermeasure:
  - The Victim-AS responded on August 24, by **announcing longer subprefixes** with the correct paths.

Note: the actual **AS-numbers** on the figures of BGPfuse **are not presented** due to privacy concerns.

Datasets Used

- For the analysis procedure, there are two possibilities:
  1. Take into account all the BGP events (W_{all}) that refer to different paths, despite the fact that a subset of them might not have caused an actual path change
  2. Filter the events by taking into account only the BGP events that have caused a successful path change event (W_{pc})
- It is worth noting that $W_{pc} \subseteq W_{all}$

Visualize all the BGP events that refer to different paths (W_{all})



AS-ABT is involved in many events

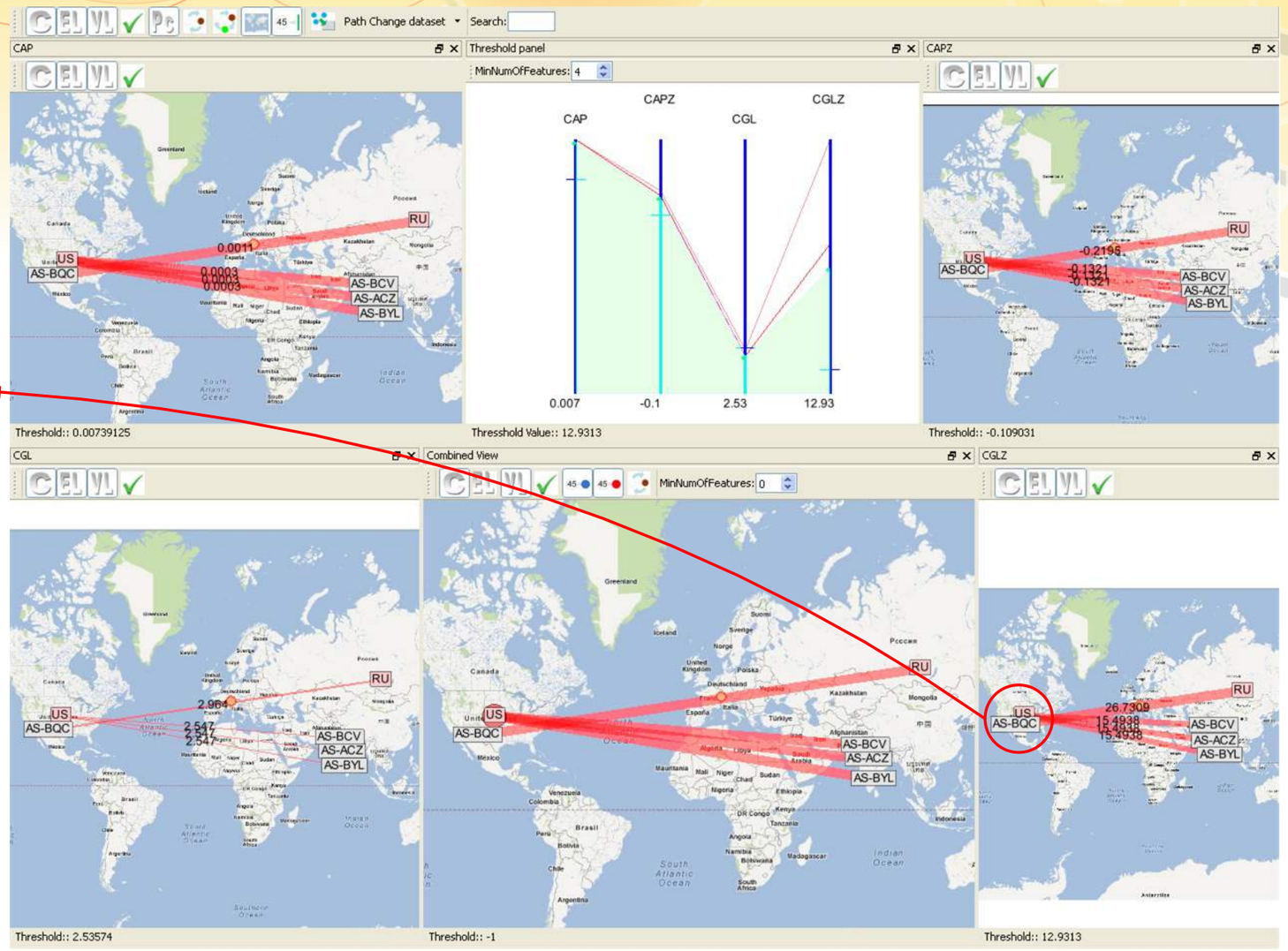
AS-AEM and AS-AEL have extremely high CGLZ score

VAST · INFOVIS · SCIVIS · BIOVIS · LDAV VIS 2013

Visualize the successful BGP path change events (W_{pc})

- AS-BQC is the Hijacking AS of the aforementioned event,
- Hijacking an AS located in Russia

AS-BQC has very high CGLZ score



Conclusions

- Use of visual feature fusion for BGP attack detection
- Allow the user to change the importance of each feature on the fly based on the feedback provided by the visual display
- Graph based visualizations to highlight relationships between different actors, as well as underline important actors
- Scalable approach to multiple features

Thank you