

Major Air Traffic Flow Identification with Fractal-Based Graph Simplification

Chunyao Ma, TALAMUDIPI ADITYA ISHAAN, Huijuan Yang,
Thinh Hoang-Dinh, Sameer Alam, and Daniel Delahaye

SESAR Innovation Days 2025, Bled, Slovenia
1 - 4 December 2025



Background

What are major traffic flows? Why identify major flows?

Methodology

Bottom-up, local flows to global flows, stability

Experiment and results

Major flow identification using air traffic data on 14th July 2023

Conclusion and Discussion

Summary and further research

Major Flow Identification



Example of the SW-NW-axis air traffic:

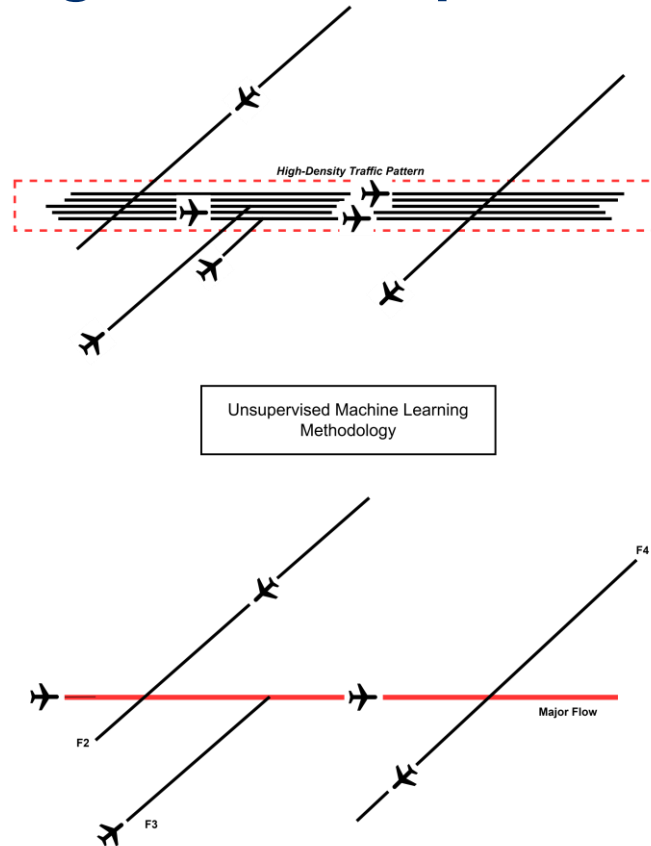
- EG**, EI** to/from LE**, LP**, GC**
- Number of Flights → 1300-1600 per day

- Among the overall air traffic in the airspace, there are **major air traffic flows** that are operationally important, playing key role in tasks such as **workload evaluation and DCB**;

- Major air traffic flows can be identified as **concentrated streams of flights** that follow similar trajectories between specific geographical regions or airport pairs [1].

[1] Y. Chen, Y. Zhao, and Y. Wu, "Recent progress in air traffic flow management: A review," Journal of Air Transport Management, vol. 116, p. 102573, 2024.

Challenge: the complex air traffic structure



Ideal scenario: air traffic is **concentrated** into well-defined flows (possible for TMA traffic),

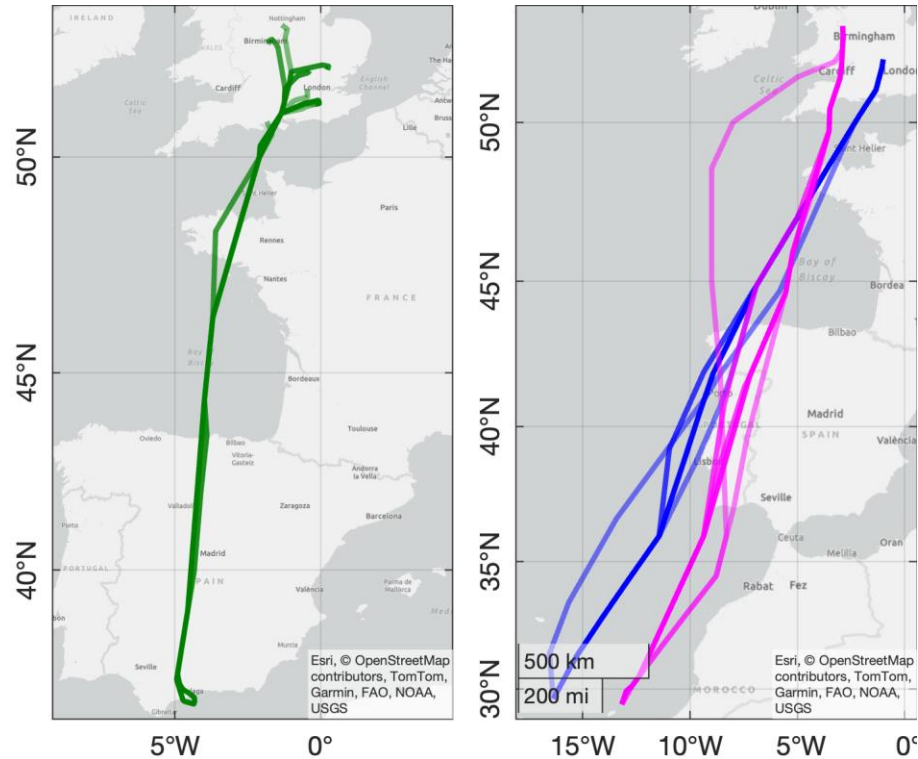
- following **similar directions, path lengths, same OD**,
- **clear distinctions** between different flows.



In reality, the cross-border air traffic is far more complex:

- flows can **emerge, disappear, split, merge, or intersect**,
- difficult to define stable and distinct flow structures as seen in the ideal scenario.

State-of-the-art



Grouping flights based on shared origin–destination (OD)

- Straightforward and easy to implement.
- **Selection of OD** requires experience (knowledge of the ODs of the major flows).
- Flights with **different OD** can be in the **same flow**, sharing substantial portions of their trajectories.

Clustering flights based on **trajectory similarity**

- Trajectory **distance**, Fréchet, DTW;
- Trajectory **embedding**, autoencoder,
- Works well in TMA (flows are distinctly structured);
- May **group spatially disconnected trajectories** that have similar shapes.
- May also group flights in the same flow into separate clusters due to local differences.

What to do?



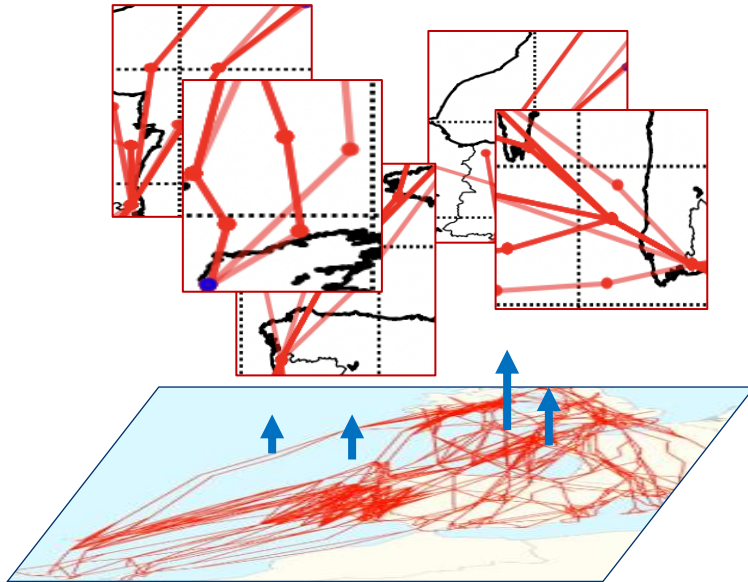
Major flow identification by OD or trajectory similarity;

- **Top-Down Perspective:** operates from the global structure of air traffic; to identify the underlying major air traffic flow segment.
- **Challenges with top-down methods:**
 - Air traffic flows exhibit **interconnectivity**, including:
 - Merging,
 - Overlapping,
 - Splitting.
 - These behaviors make **flow boundaries hard to distinguish**, leading to:
 - Reduced clustering accuracy,
 - Suboptimal identification results.

Life Philosophy: When a problem feels too complex to solve head-on, it often signals a need to look deeper and decompose.

Flow decomposed: Bi-level flow Identification

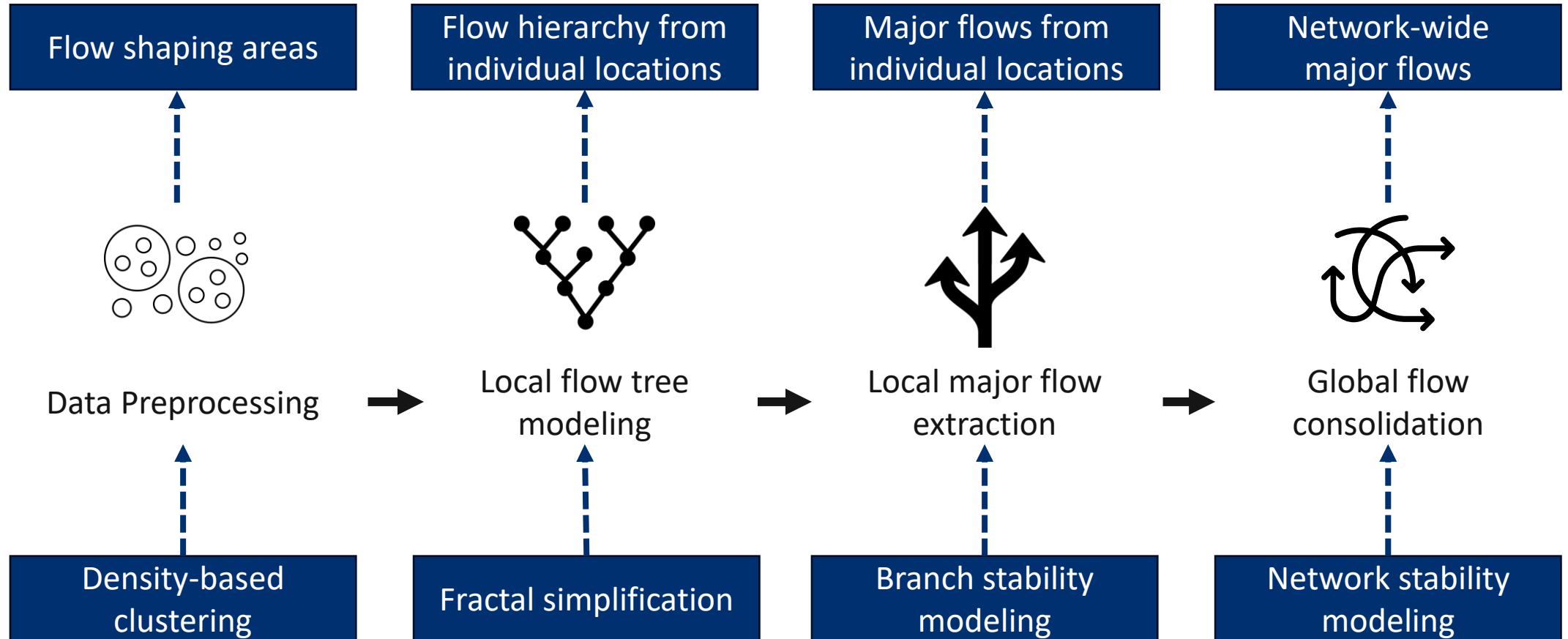
- **Local flow tree:** The outbound trajectories from an individual location often exhibit a hierarchical tree structure, where flights gradually branch out into finer sub-flows.



- **Air traffic flow network:** By integrating local flow trees from various locations, i.e., branches from different trees converge, overlap, or merge, a global air traffic flow network emerges.



Methodology Overview



Data Preprocessing – Flow Shaping Areas Identification

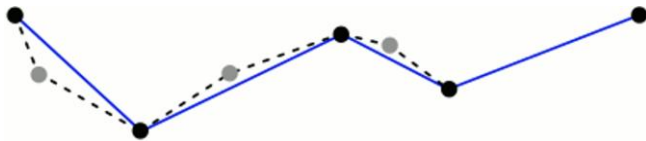
Flow-shaping areas:

- potential origins, terminations, and transitions of major flows;
- provide the foundation for downstream flow identification.

Trajectory-shaping waypoints extraction:

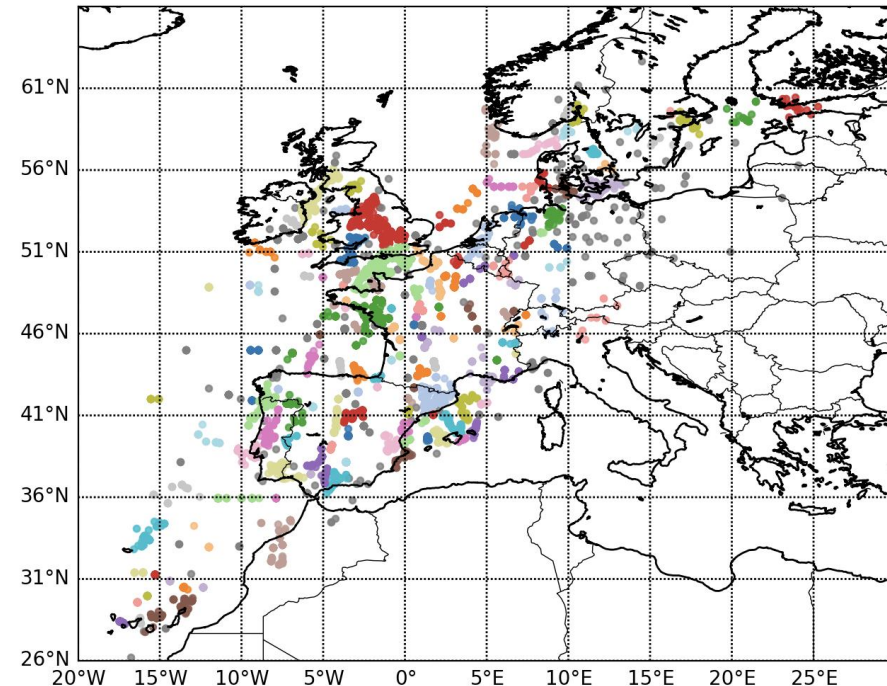
Ramer–Douglas–Peucker (RDP)

Algorithm;

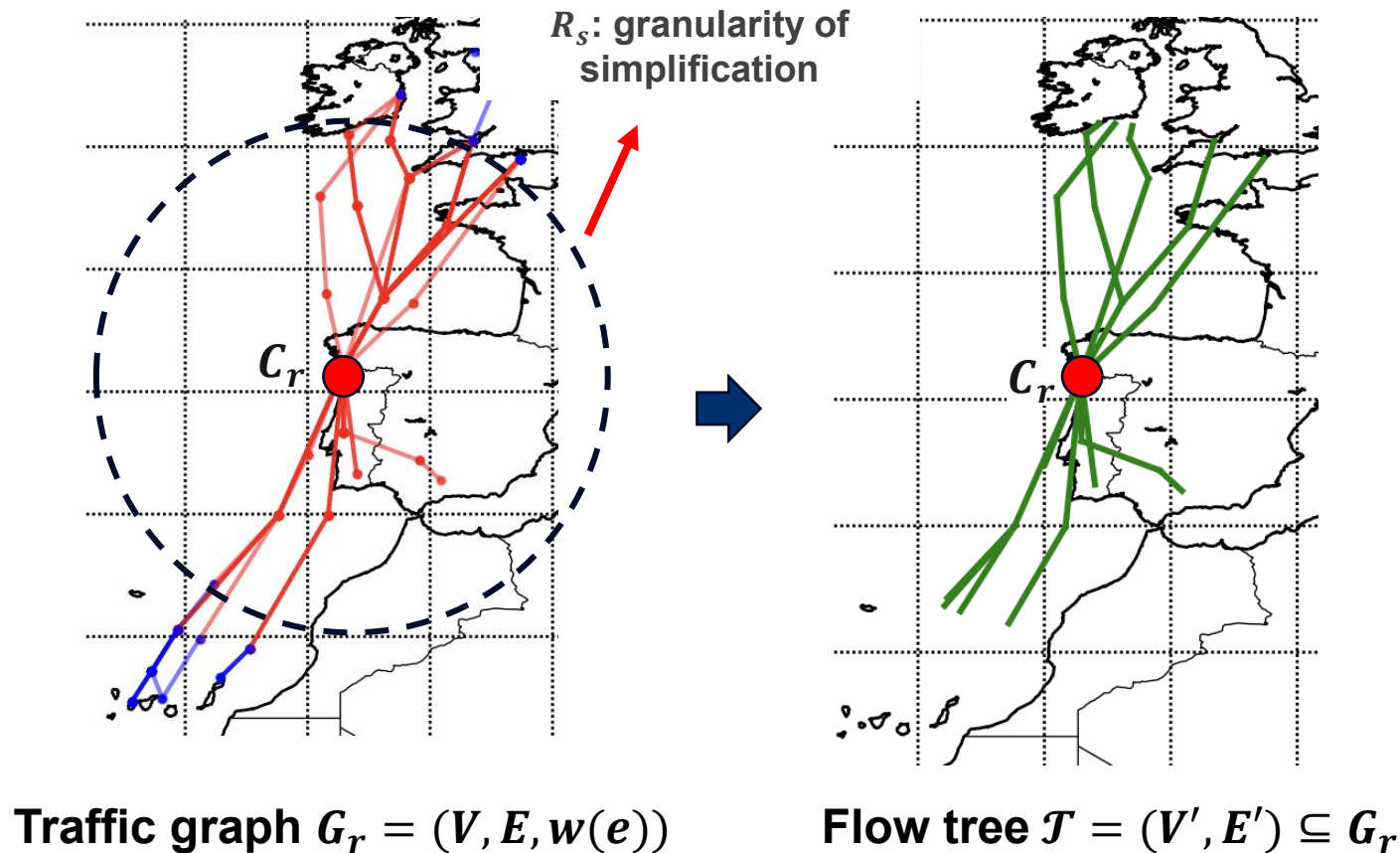


Flow-shaping areas identification:

Hierarchical clustering: HDBSCAN



Local flow tree modeling



Fractal-inspired Simplification:

- Smooths out smaller-scale deviations
- Preserves core flow structures at different levels of granularity

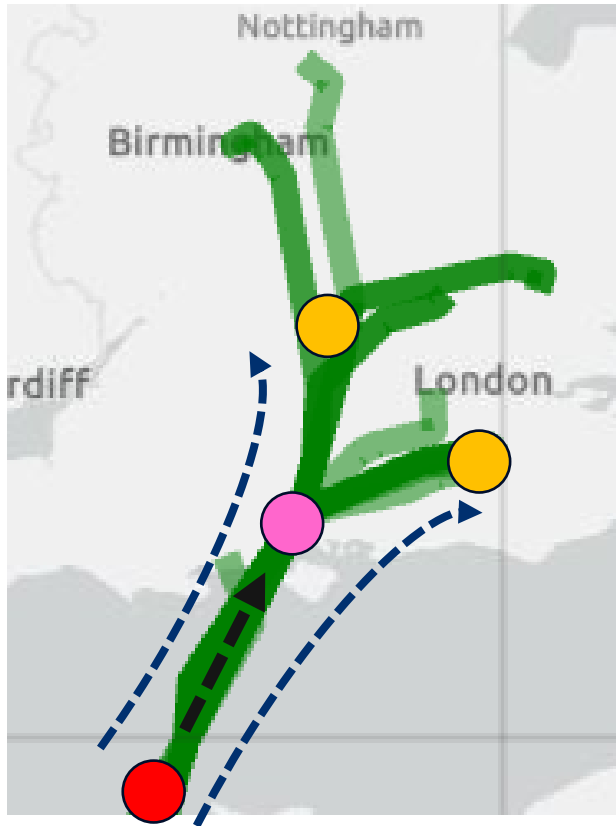
Shortest Path Preservation:

$$\text{Obj: } \min_{\mathcal{T} \subseteq G_r} \sum_{v \in \text{Leaves}(\mathcal{T}, r_s)} \text{Length}_{\mathcal{T}}(C_r, v)$$

Where:

- $\text{Length}_{\mathcal{T}}(C_r, v) = \sum_{e \in P_{\mathcal{T}}(C_r, v)} w(e)$
- $P_{\mathcal{T}}(C_r, v)$: path from root C_r to node v in $\mathcal{T}$

Local major flow extraction



Branch stability modeling

One major flow  Higher volume

or 

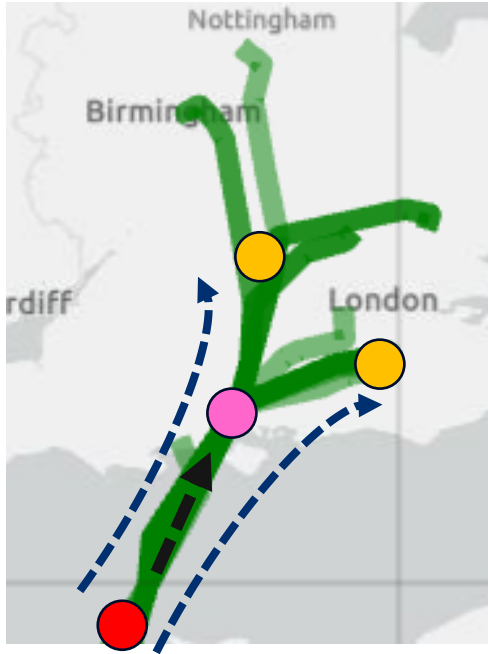
Two major flows  Longer flow span

How many flights are in flow?

Flow stability

How long does the flow exits?

Local major flow extraction



$\lambda(v)$

Flow span: $P_{\mathcal{T}}(C_r, v)$

High λ

More persistent flow

Low λ

Less persistent flow

$\lambda_b(F)$

$\lambda(v)$ when the flow F forms after the split off at v

$\lambda_d(f)$

$\lambda(v)$ when flight $f \in F$ leaves (flow split) or becomes noise at v

$Stability(F)$

How long F spans and how many flights F contains.

$$= \sum_{f \in F} (\lambda_d(f) - \lambda_b(F))$$

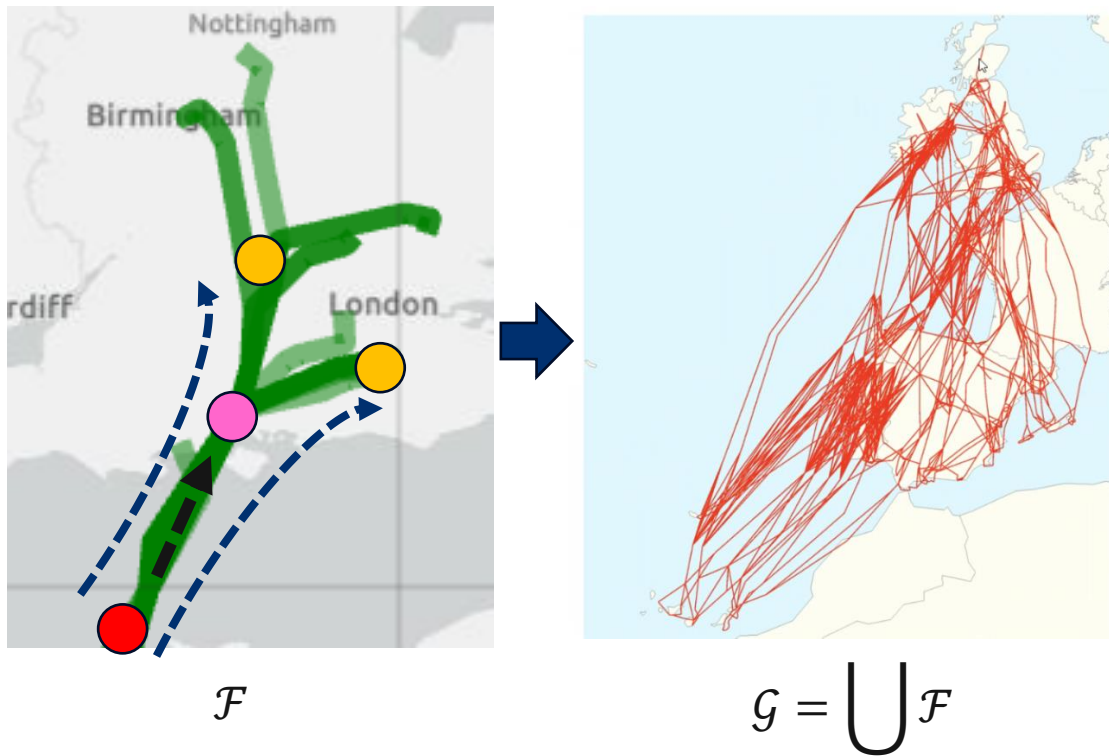
$$Obj: \max_{\mathcal{F} \subseteq \mathcal{H}} \sum_{F \in \mathcal{F}} Stability(F) \quad \left\{ \begin{array}{l} F_i \cap F_j = \emptyset, \text{ for all } F_i \neq F_j, F_i, F_j \in \mathcal{F} \\ |F_i| \geq N_{min}, \text{ for } F_i \in \mathcal{F} \\ \lambda_d(F_i) > \lambda_{min}, \text{ for } F_i \in \mathcal{F} \end{array} \right.$$

Flows are **non-overlapping**

Minimum volume of the flow

Minimum span of the flow

Global flow consolidation



Network stability modeling

Obj:
$$\max_{\mathbf{z}, \mathbf{a}} \sum_{F_i \in \mathcal{G}} \lambda_d(F_i) \cdot a_i \sum_{f_j \in \mathcal{D}} z_{i,j} b_{i,j}$$

s. t.: $b_{i,j} = 1 \text{ if } f_j \in F_i, \text{ otherwise } 0$

$$z_{i,j} \leq b_{i,j}, z_{i,j} \in \{0,1\}$$

$$\sum_{F_i \in \mathcal{G}} z_{i,j} \leq 1, \text{ for } f_j \text{ in } \mathcal{D}$$

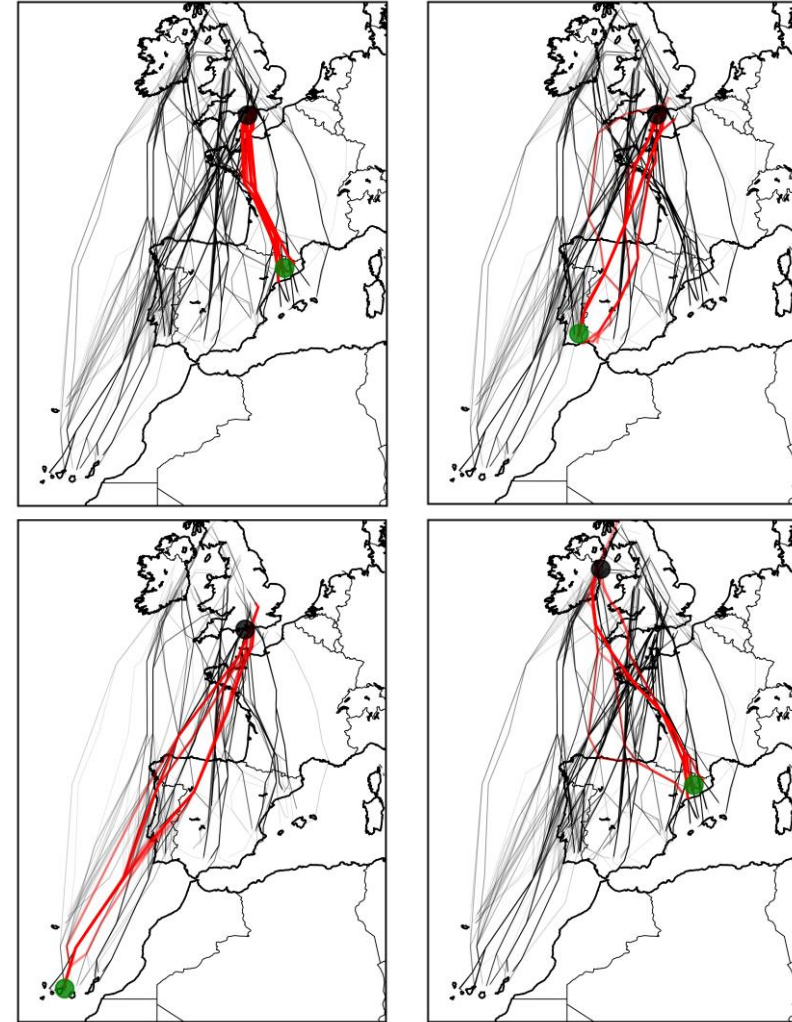
$$a_i = 1 \text{ if } \sum_{f_j \in \mathcal{D}} z_{i,j} b_{i,j} \geq N_{\min}, \text{ otherwise } 0$$

$N_{\min}$: Minimum #flights required for a flow

Traffic: 14th July 2023

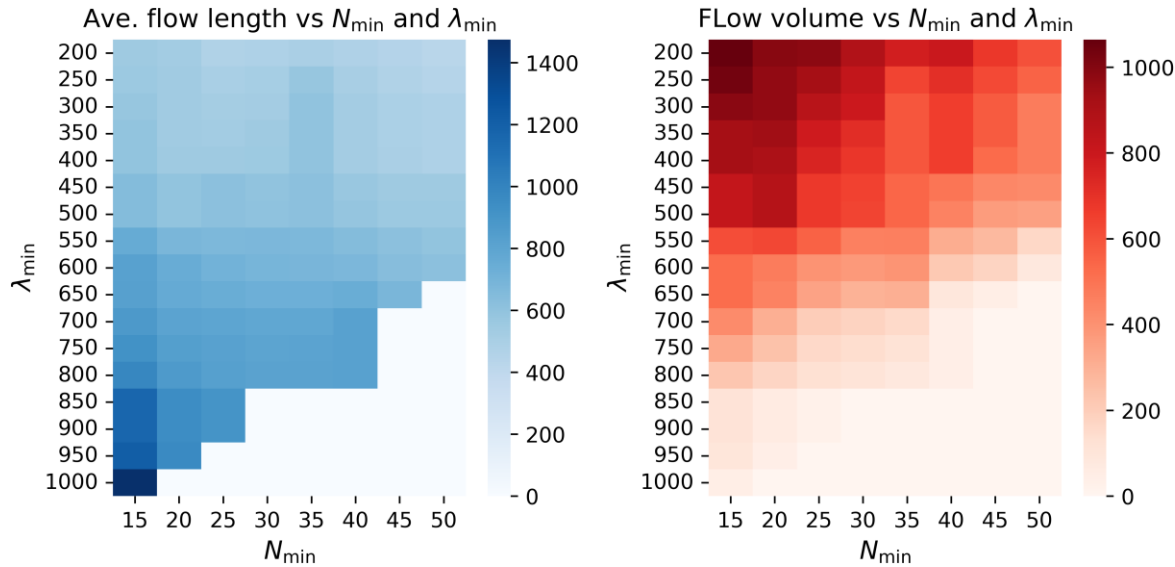
Northwest–southwest (NW–SW) traffic.
Maximum length of flight segments: 1943 nm $\rightarrow R_s$
 $N_{min} = 15$ and $\lambda_{min} = 500$ nm

22 major flows are identified.
Length: 647nm
Total number of flights: 826:
Total Daily flights: 1368



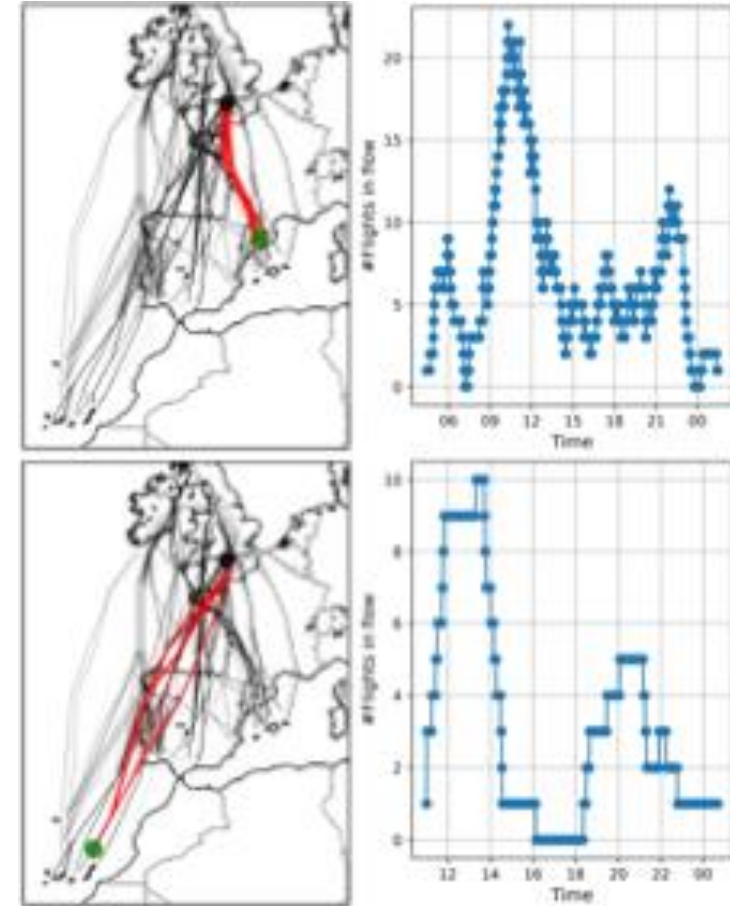
Examples of major flows identified

Traffic: 14th July 2023



Different parameters yield different flows.

- $\uparrow \lambda_{min}$ yields longer major flows but reduces the #flights.
- $\uparrow N_{min}$ shortens the flows and excludes flows with low volume.



Monitoring the number of flights within each flow

- emerge and disappear
- peak and low-demand

The proposed **bottom-up major flow identification** methodology provides a **robust and interpretable** means of identifying major air traffic flows through:

- Detection of flow-shaping areas through density-based clustering,
- Construction of local flow trees with fractal-inspired simplification;
- Introduction of a flow stability metric balancing traffic volume and flow persistence;
- Optimization-based consolidation of local flows into a coherent global representation.
- Varying parameters such as N_{min} , λ_{min} , and R_s allows the level of granularity to be tailored to different operational requirements.

The proposed **bottom-up major flow identification** methodology opens several avenues for further research:

- Coupling with predictive models would enable proactive congestion management and support collaborative decision-making processes.
- The identified major flows can facilitate the analysis of congestion correlations and propagation at the network level. The major flow outputs enable the manager to identify which flows are likely to generate high demand in specific regions and apply targeted rerouting strategies.

Thank You!

Chunyao Ma
ATMRI, NTU, Singapore
chunyao.ma@ntu.edu.sg

