Constraint-based Pattern Mining in Dynamic Graphs

Céline Robardet

Univeristé de Lyon, INSA-Lyon, CNRS, LIRIS UMR 5205, F-69621 Villeurbanne, France

Introduction

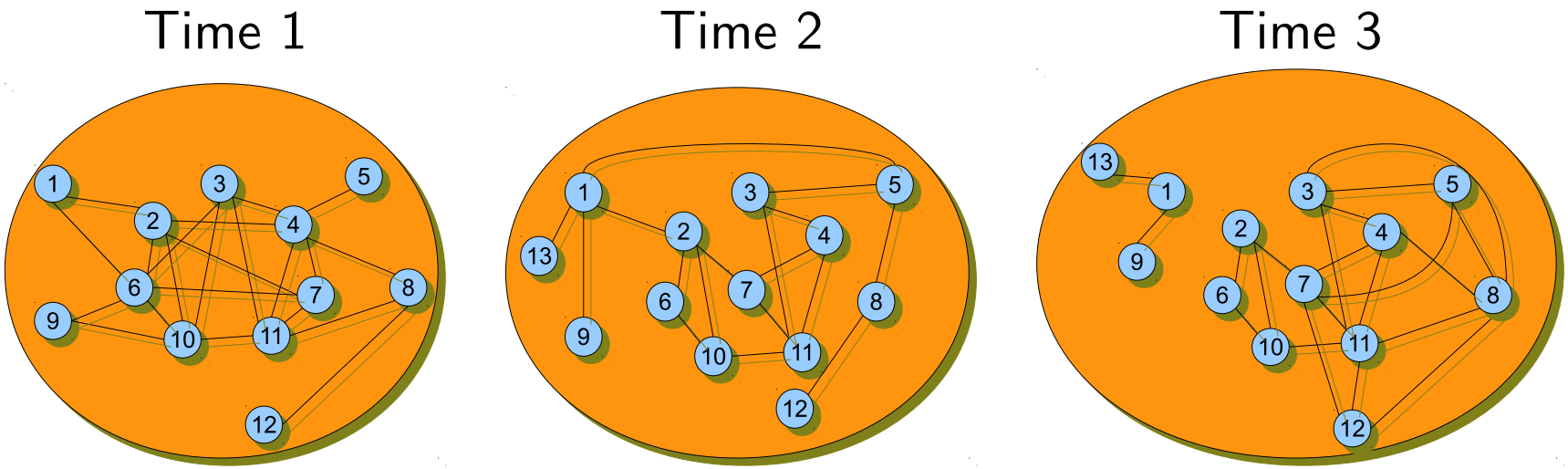
Dynamic Graphs

- Represent relationships between entities that evolve over time
- So far, much effort devoted to the analysis of graph properties (degree, diameter)

Objective

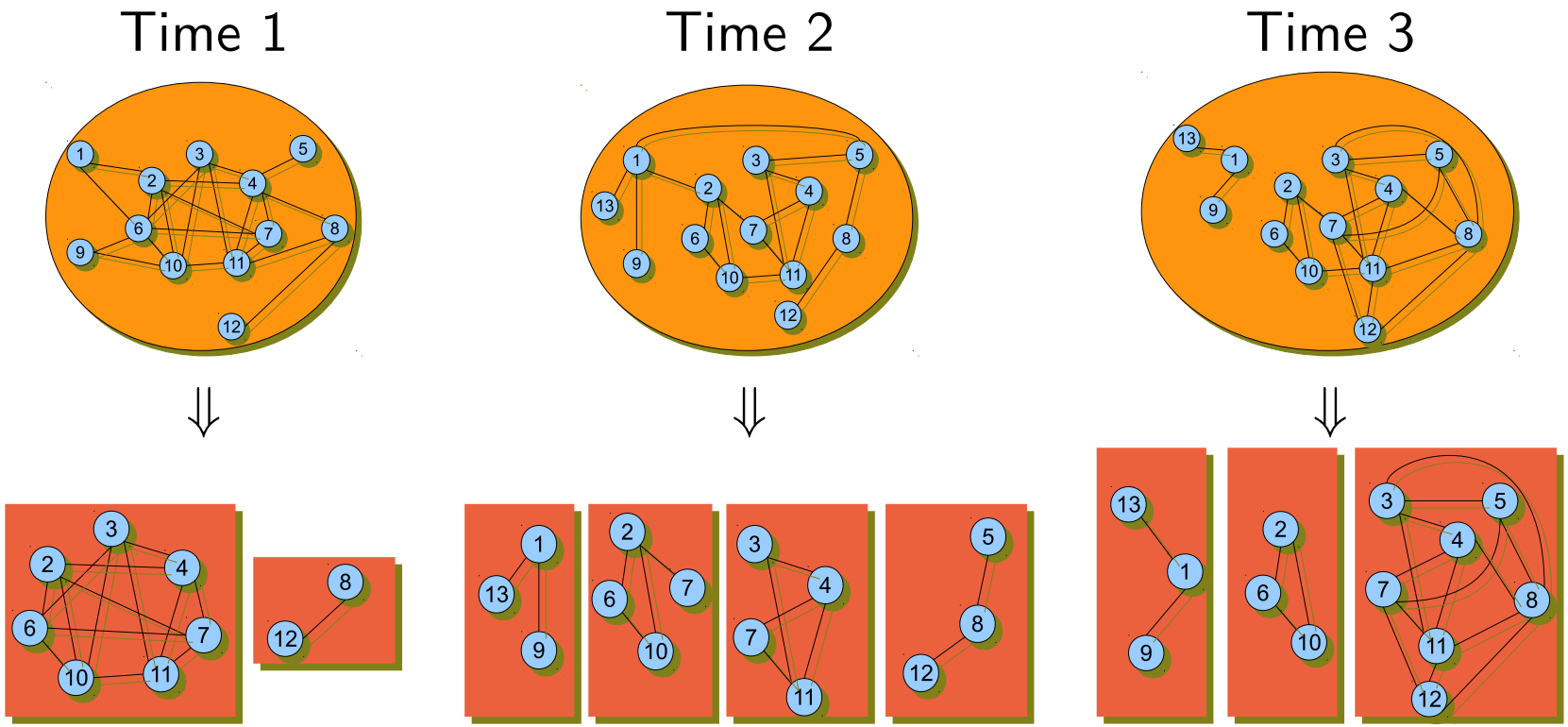
- Study the evolution of a graph over time
- Capture strong interactions in the graph and their evolution over time (evolving communities of social dynamic graphs)

The data



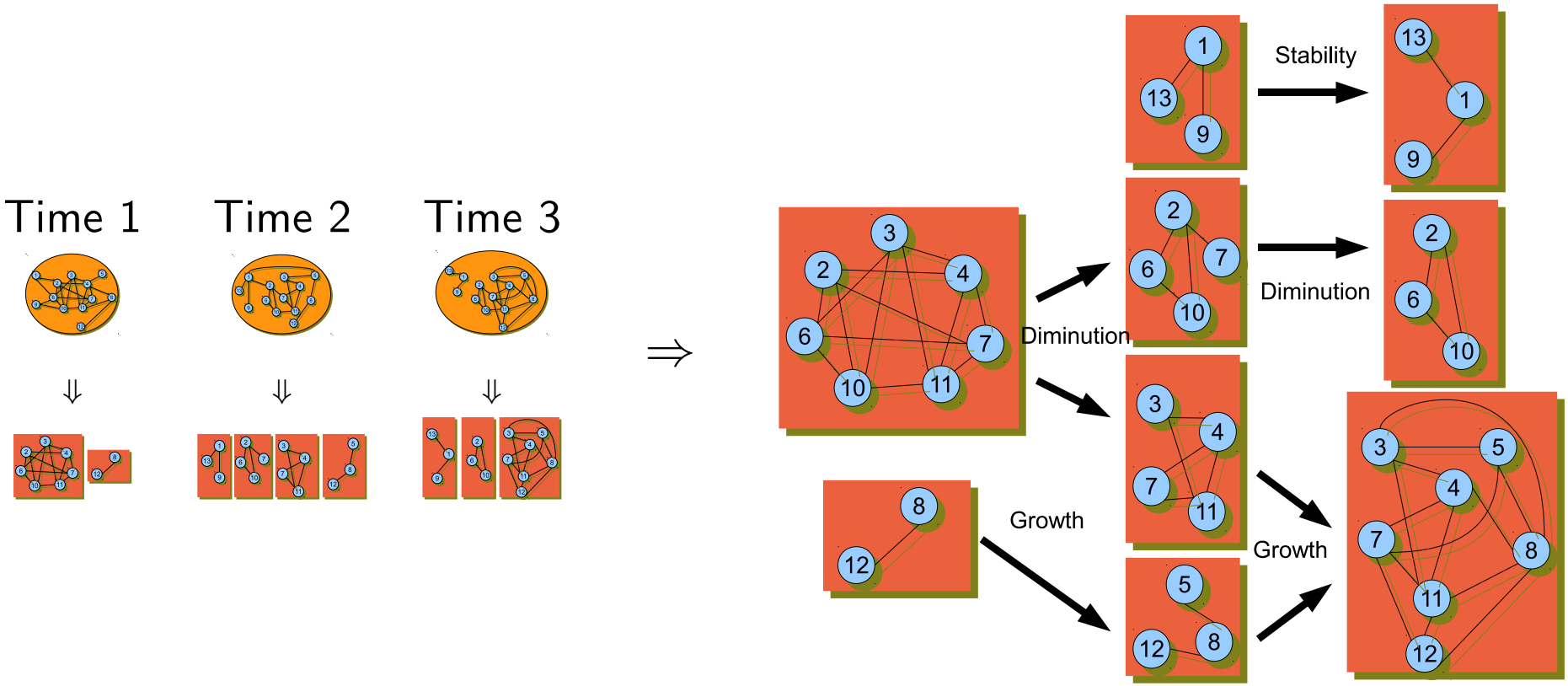
Focus on the **microscopic level** and propose a constraint-based mining approach to uncover **evolving patterns**.

Step 1: Mining constrained subgraphs in each static graph



Compute **highly connected** subgraphs that are also **isolated**.

Step 2: Mining evolving subgraphs



Combine patterns from **consecutive time stamps** to construct a **global model** of the dynamic of the graphs.

Step 1: Mining constrained subgraphs in static graphs

Subgraphs of interest are usually those made of vertices that have a high density of edges.

$\Rightarrow$ Cliques are subgraphs with maximal density

$\Rightarrow$ Pseudo-cliques relax this strong property using a user-defined threshold $\sigma \in [0, 1]$:

$$S \text{ is a pseudo-clique iff } \frac{2|E_S|}{|S|(|S|-1)} \geq \sigma$$

Properties

The pseudo-clique constraint is not anti-monotonic:
Expanding a subgraph by adding a vertex could make the density increase or decrease.

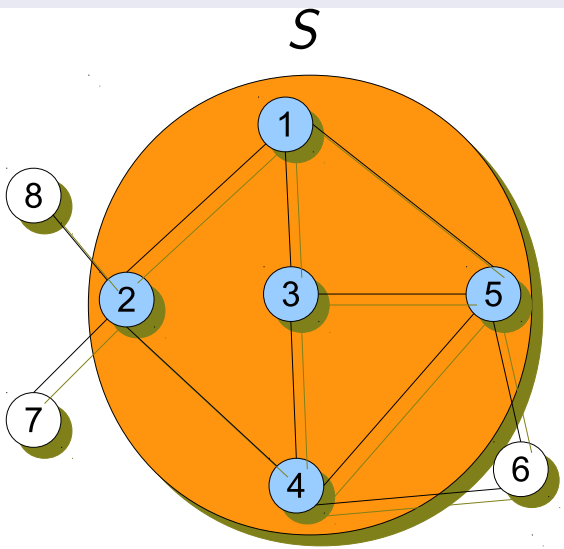
How to enumerate pseudo-cliques?

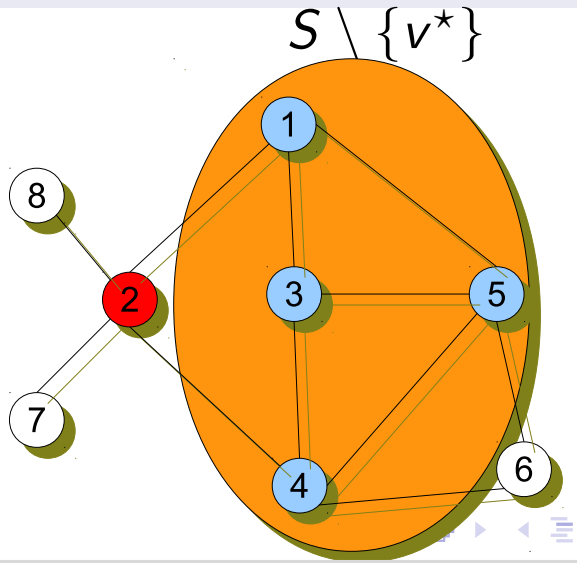
Loose anti-monotonicity of pseudo-cliques

Pseudo-cliques can always be grown from a smaller pseudo-clique with one vertex less [F. Zhu et al. (*PAKDD 2007*)]:

- Let S be a pseudo-clique
- Let $v^\star$ be a vertex of S having the smallest degree on S
- Thus $S \setminus \{v^\star\}$ is also a pseudo-clique

$\sigma = \frac{2}{3}$





How to enumerate pseudo-cliques?

PCE [T. Uno (ISAAC 2007)]

Given a pseudo-clique S, find v such that:

$$\deg_{S\cup\{v\}}(v) = \min_{u\in S\cup\{v\}} \deg_{S\cup\{v\}}(u) \quad (1)$$

$$\frac{\sum_{u\in S\cup\{v\}} \deg_{S\cup\{v\}}(u)}{n(n-1)} \geq \sigma \quad (2)$$

Other useful constraints

Maximality and isolated constraint

Not all the pseudo-cliques of a graph are of importance:

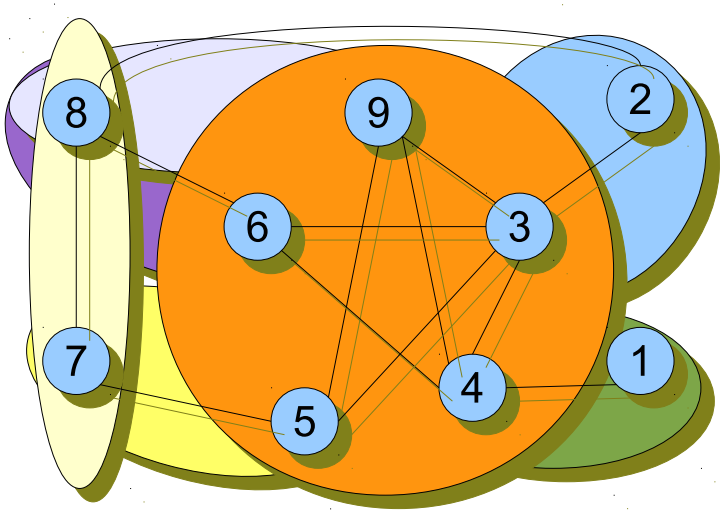
- Some are redundant (because non maximal)
- Others have many links to external vertices

The isolation constraint imposes a maximum to the average number of external links per vertex:

$$\frac{\sum_{u \in S} (\deg(u) - \deg_S(u))}{|S|} \leq \gamma$$

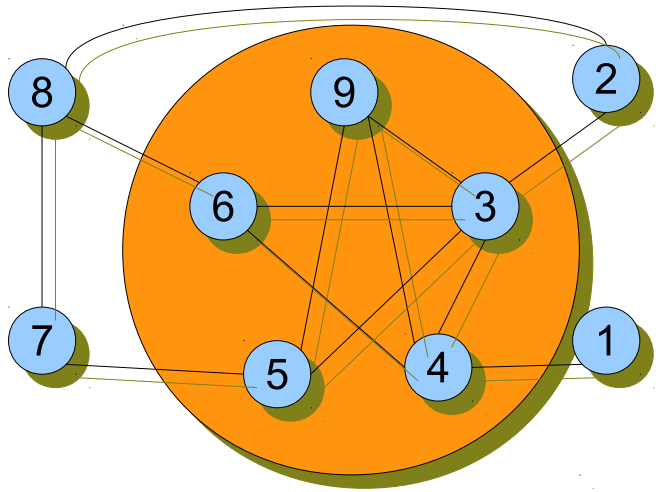
Other useful constraints

$\sigma = 0.7$ Without Isolated constraint



9 patterns

With $\gamma = 1$



1 pattern

Step 2: Global model of dynamic graph construction

Objective

Structured the numerous patterns obtained in step 1 to answer the following questions:

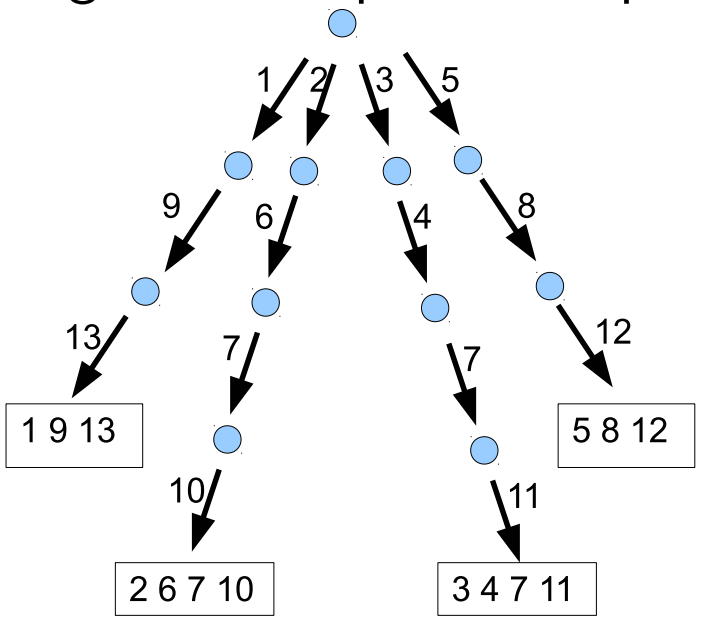
- Do the strong interactions growth, diminish or remain stable over time?
- When do the change occur?

Temporal relationships between time consecutive subgraphs

- **Stability:** S remains the same between $t-1$ and t
- **Growth:** S enlarges at time t
- **Diminution:** S shrinks at time t
- **Extinction:** S disappears at time t
- **Emergence:** S appears at time t

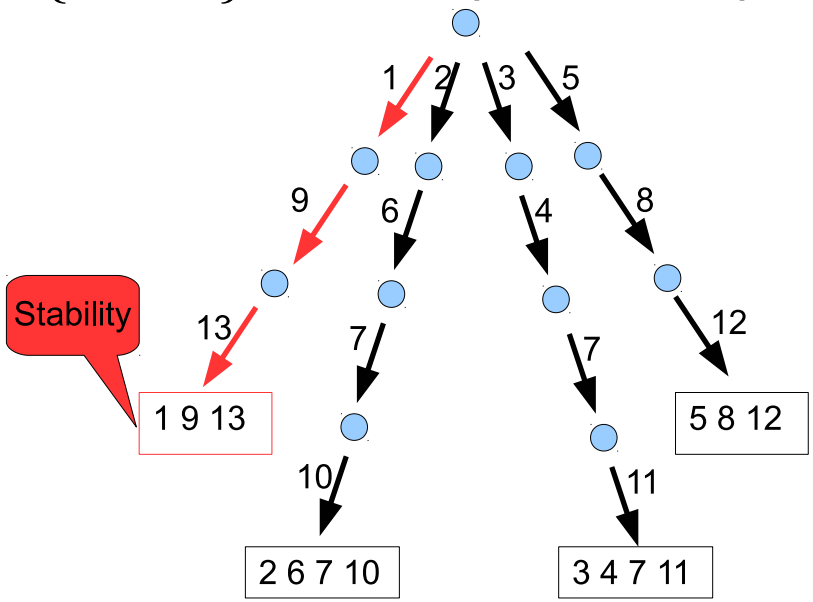
Step 2: Global model of dynamic graph construction

Trie containing the valid pseudo-cliques of time 2.



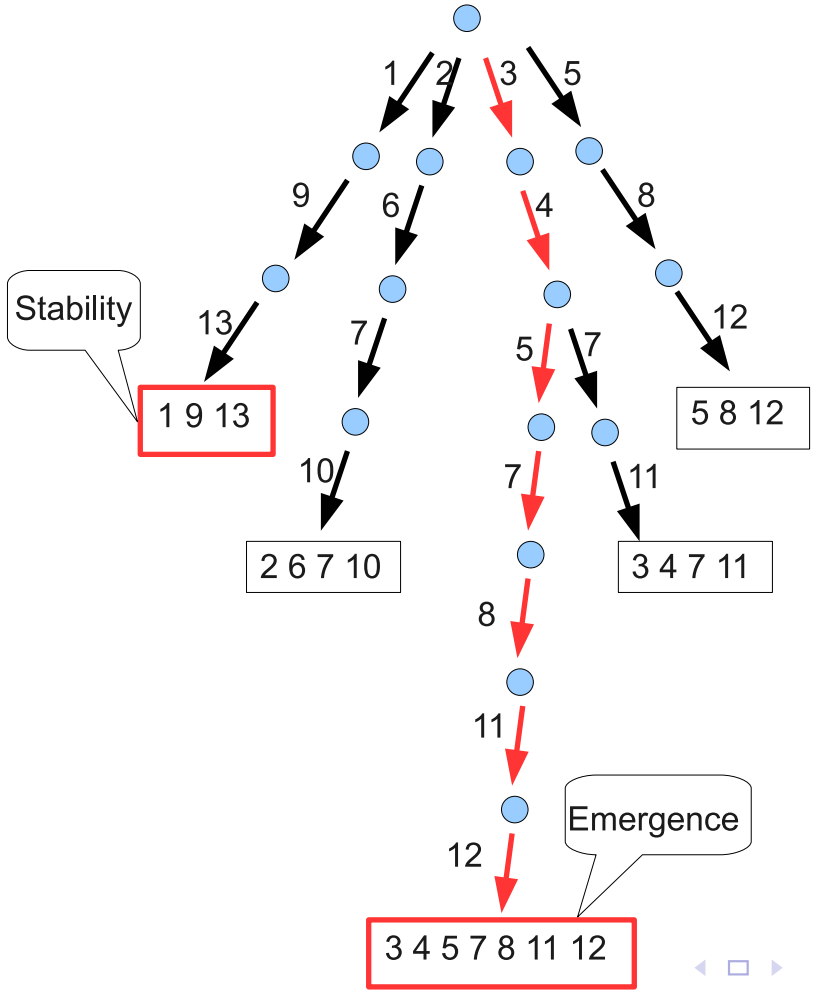
Step 2: Global model of dynamic graph construction

Insertion of $\{1\ 9\ 13\}$, a valid pseudo-clique of time 3.



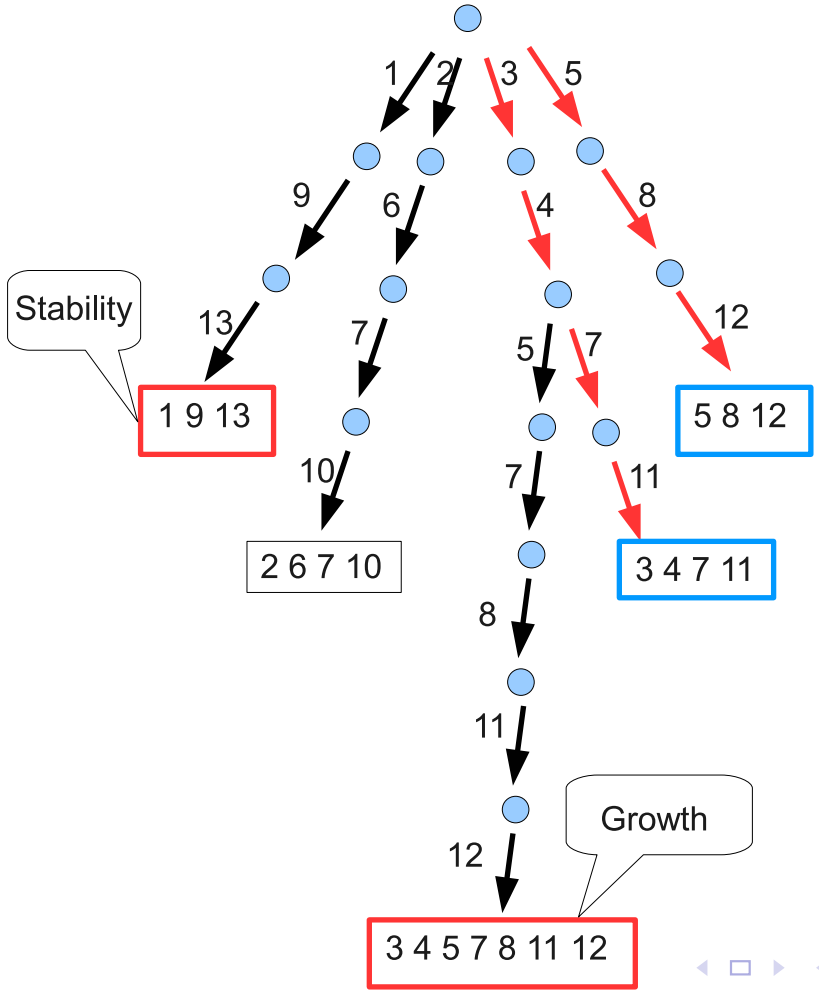
Step 2: Global model of dynamic graph construction

Insertion of {3 4 5 7 8 11 12}, a valid pseudo-clique of time 3.



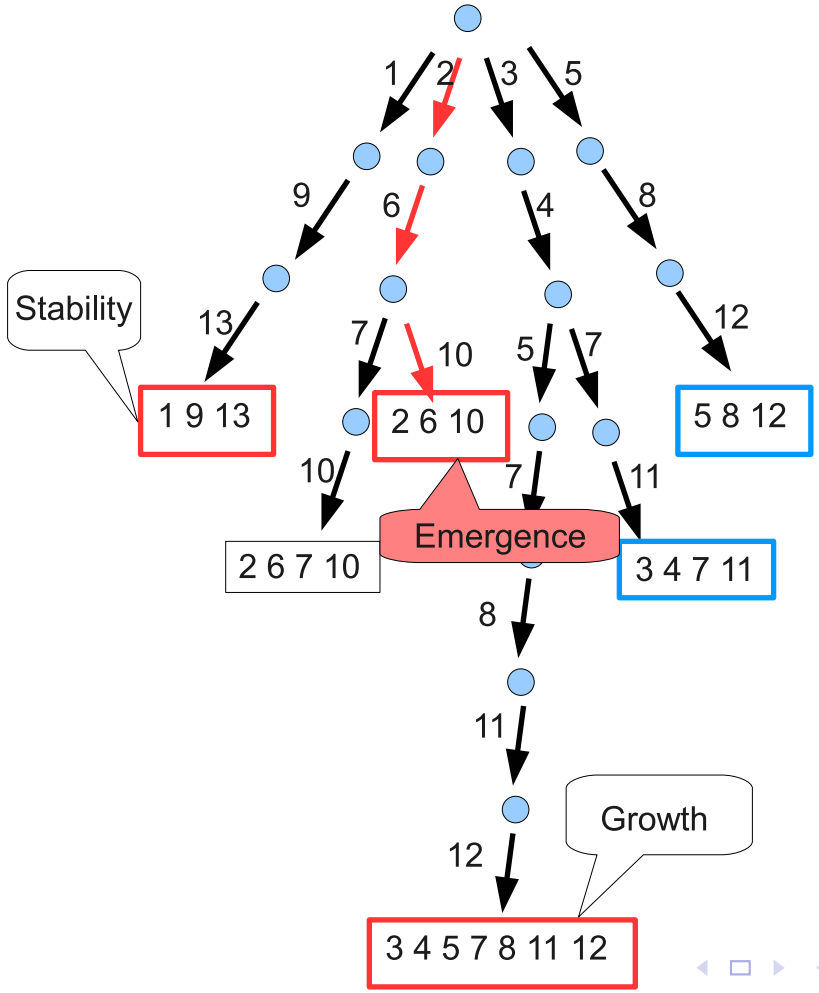
Step 2: Global model of dynamic graph construction

Does the subgraph $\{3\,4\,5\,7\,8\,11\,12\}$ form a growth evolving pattern?



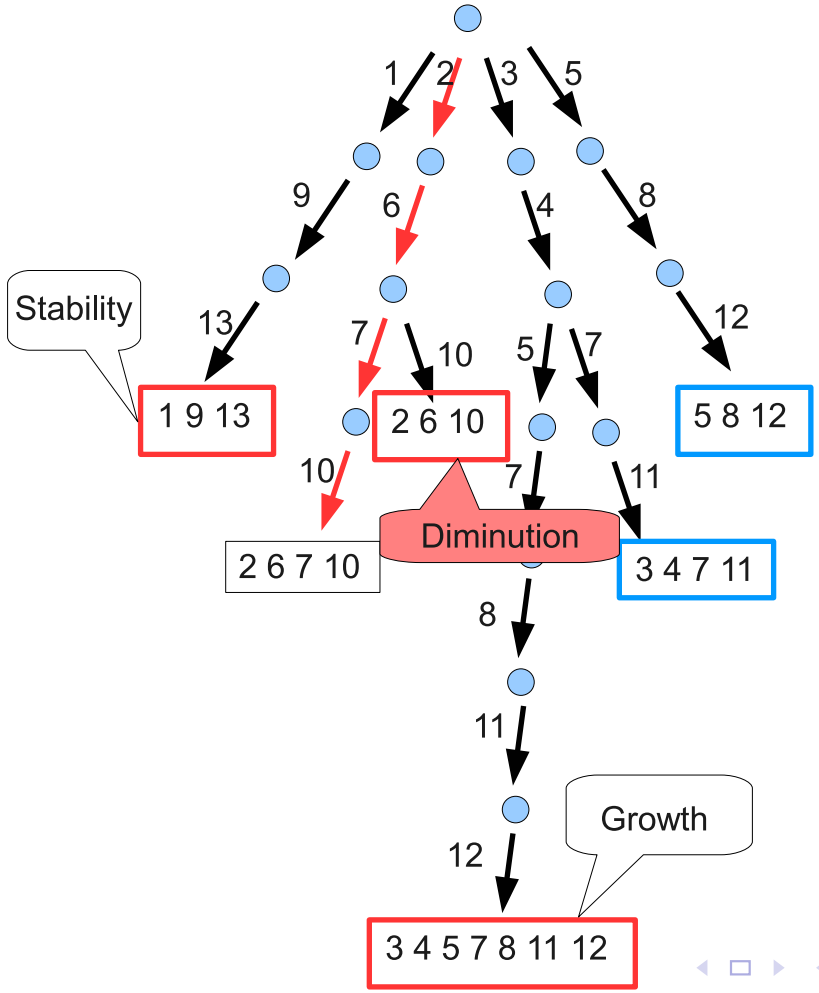
Step 2: Global model of dynamic graph construction

Insertion of $\{2\,6\,10\}$, a valid pseudo-clique of time 3.

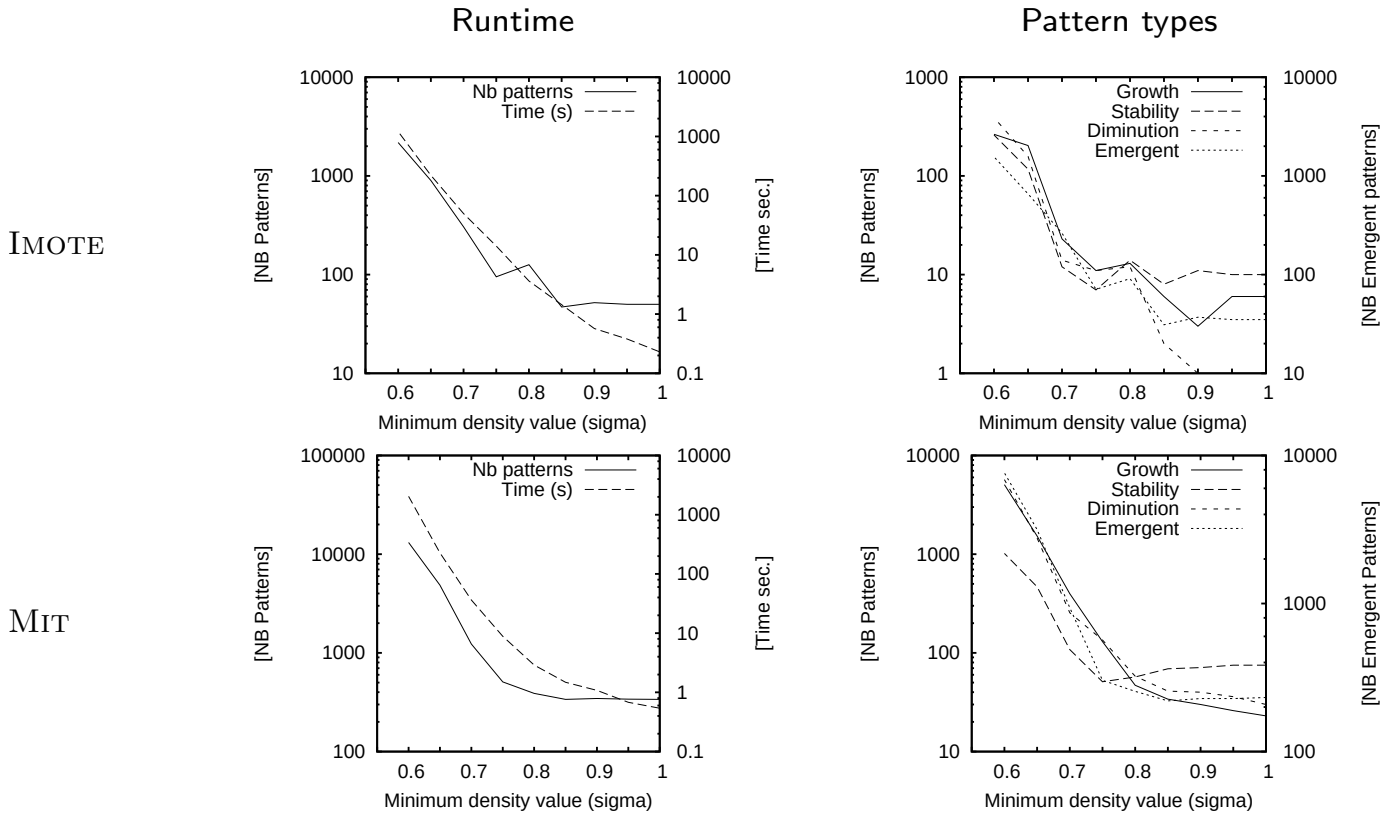


Step 2: Global model of dynamic graph construction

Is $\{2\,6\,10\}$ a subset of a pattern from time stamp 2?



Results for IMOTE and MIT evolving graphs



Runtime and number of extracted patterns (logarithmic scales) (left) and number of patterns of each type (right) for different density threshold σ.

Lyon's shared bicycle system VELO'V

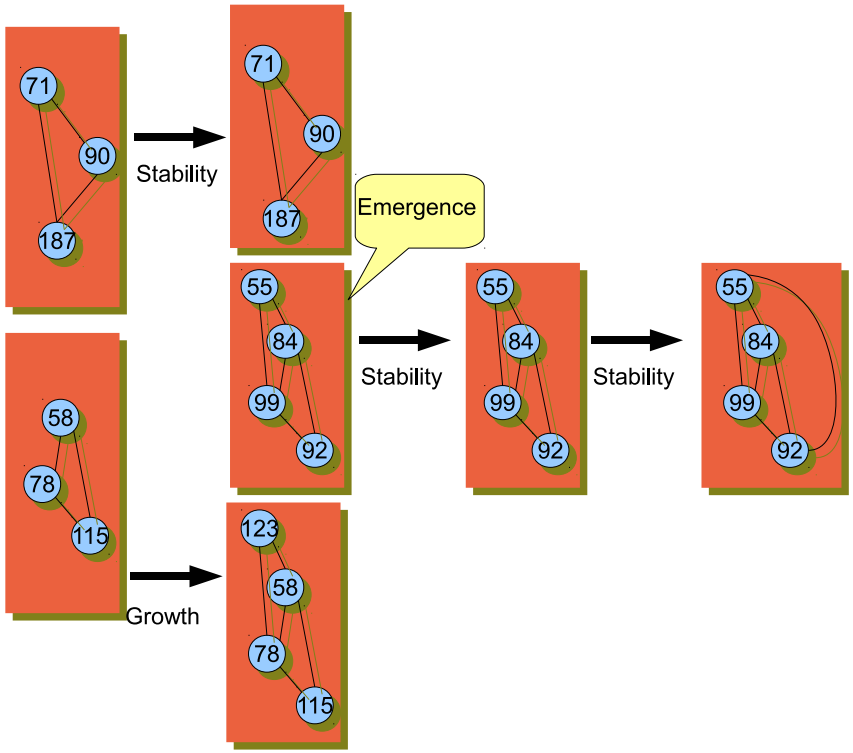
VELO'V system

- 340 stations spread in Lyon
- 4000 bikes available in those stations
- rental at any station, return it at any other one

VELO'V data

- More than 13 millions of bicycle trips
  - Time-stamps of the trip
  - Rent and return station IDs

Results for VELO'V evolving graph



Evolving patterns ($\sigma = 0.8$ and $\gamma = 5$) of Velo'v stations and their localization in Lyon. Patterns are shadowed on the map.

Conclusion

Evolving-pattern mining problem in dynamic graph

- A local-to-global framework:
  1. Local patterns are mined in static graphs;
  2. They are combined with the one extracted in the previous graph to form evolving patterns.
- EVOLVING-SUBGRAPHS algorithm efficiently mines all evolving patterns that satisfy the constraints.
- Our experiments on real life datasets show that our approach produces high quality patterns that are useful to understand the graph dynamics.