

Découverte de motifs sur les usages du réseau Vélo'v

Loïc Cerf

Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205, F-69621, France

Plan

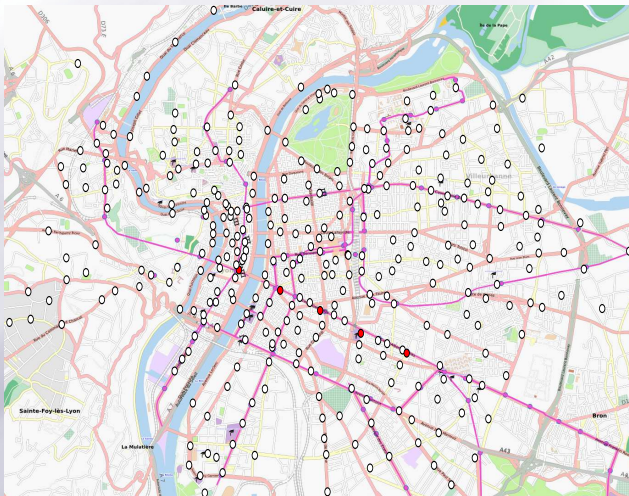
- 1 Motifs locaux
- 2 Des propriétés intéressantes
- 3 Perspectives

Plan

- 1 Motifs locaux
- 2 Des propriétés intéressantes
- 3 Perspectives

Exemple de motif

De midi à minuit :



Exemple de motif

Analyse dynamique de l'usage du réseau Vélo'v

338 stations Vélo'v sont réparties à Lyon et sa proche banlieue. À n'importe laquelle de ces stations, l'utilisateur peut emprunter un vélo et le ramener à n'importe quelle autre station. L'activité moyenne du réseau sur une journée de 2006/2007 est obtenue par agrégation par pas de temps de trois heures. À chaque pas de temps un trajet est fréquent ou non (*données Booléennes*).

Pré-traiter autrement les données permet de découvrir autre chose.

Exemple de motif

Analyse dynamique de l'usage du réseau Vélo'v

338 stations Vélo'v sont réparties à Lyon et sa proche banlieue. À n'importe laquelle de ces stations, l'usager peut emprunter un vélo et le ramener à n'importe quelle autre station. L'activité moyenne du réseau sur une journée de 2006/2007 est obtenue par agrégation par pas de temps de trois heures. À chaque pas de temps un trajet est fréquent ou non (*données Booléennes*).

Pré-traiter autrement les données permet de découvrir autre chose.

Extraction complète de motif locaux

Un motif *local* ne renseigne que sur les éléments (dans notre cas les stations et les périodes de temps) qu'il implique.

Les motifs locaux que nous extrayons :

- s'inscrivent dans des données à deux dimensions ou plus (on peut, par exemple, ajouter une dimension « jours de la semaine » à l'analyse) ;
- sont maximaux (en ajoutant un élément, ils ne sont plus valables) ;
- peuvent être contraints à présenter certaines propriétés.

Une extraction *complète* de motifs locaux renvoie *tous* ceux qui satisfont les propriétés requises (c'est-à-dire de *requête*).

Extraction complète de motif locaux

Un motif *local* ne renseigne que sur les éléments (dans notre cas les stations et les périodes de temps) qu'il implique.

Les motifs locaux que nous extrayons :

- ≡ s'inscrivent dans des données à deux dimensions ou plus (on peut, par exemple, ajouter une dimension « jours de la semaine » à l'analyse) ;
- ≡ sont maximaux (en ajoutant un élément, ils ne sont plus valables) ;
- ≡ peuvent être contraints à présenter certaines propriétés.

Une extraction *complète* de motifs locaux renvoie *tous* ceux qui satisfont les propriétés demandées (concept de *requête*).

Extraction complète de motif locaux

Un motif *local* ne renseigne que sur les éléments (dans notre cas les stations et les périodes de temps) qu'il implique.

Les motifs locaux que nous extrayons :

- ≡ s'inscrivent dans des données à deux dimensions ou plus (on peut, par exemple, ajouter une dimension « jours de la semaine » à l'analyse) ;
- ≡ sont maximaux (en ajoutant un élément, ils ne sont plus valables) ;
- ≡ peuvent être contraints à présenter **certaines propriétés**.

Une extraction *complète* de motifs locaux renvoie *tous* ceux qui satisfont les propriétés demandées (concept de *requête*).

Extraction complète de motif locaux

Un motif *local* ne renseigne que sur les éléments (dans notre cas les stations et les périodes de temps) qu'il implique.

Les motifs locaux que nous extrayons :

- ≡ s'inscrivent dans des données à deux dimensions ou plus (on peut, par exemple, ajouter une dimension « jours de la semaine » à l'analyse) ;
- ≡ sont maximaux (en ajoutant un élément, ils ne sont plus valables) ;
- ≡ peuvent être contraints à présenter **certaines propriétés**.

Une extraction *complète* de motifs locaux renvoie *tous* ceux qui satisfont les propriétés demandées (concept de *requête*).

Plan

- 1 Motifs locaux
- 2 Des propriétés intéressantes
- 3 Perspectives

Des propriétés implémentées

- ≡ Nombre minimal/maximal d'éléments dans une dimension : « au moins 6 stations de départ et 4 périodes de temps ».
- ≡ Grosseur *globale* minimale/maximale : « le produit du nombre de stations de départ et de périodes de temps excède 24 ».
- ≡ Symétrie entre deux dimensions : « les stations de départ et d'arrivées sont les mêmes ».
- ≡ Persistance temporelle : « les périodes de temps impliquées se suivent (ou *presque*) ».

Des propriétés à implémenter ?

- Le motif n'est valide que certains jours et jamais d'autres : « motif présent qu'en fin de semaine ».
- Faible/Forte dispersion géographique : « les deux stations les plus éloignées le sont d'au plus 1km ».
- Propriétés associées aux éléments : « au moins la moitié des stations sont à moins de 100 mètres d'une bouche de métro ».
- ...

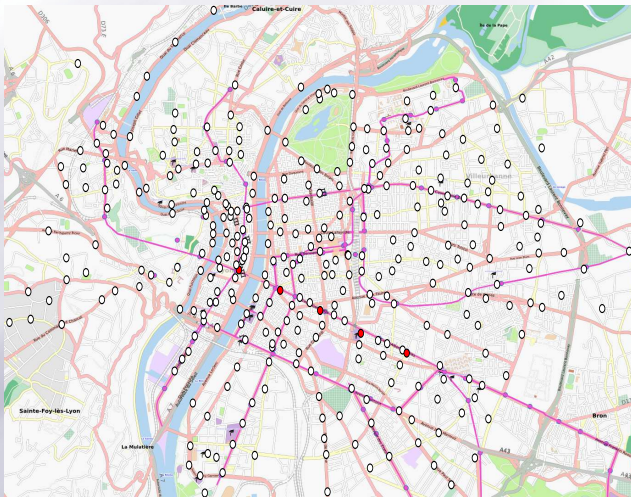
Une conjonction de plusieurs contraintes pour améliorer la pertinence

Contraintes pour l'extraction de motifs tolérant pour l'analyse dynamique de l'usage du réseau Vélo'v

- Au moins 6 stations impliquées
- Au moins 4 périodes de temps
- Symétrie entre les stations de départ et d'arrivée
- Presque-persistance temporelle

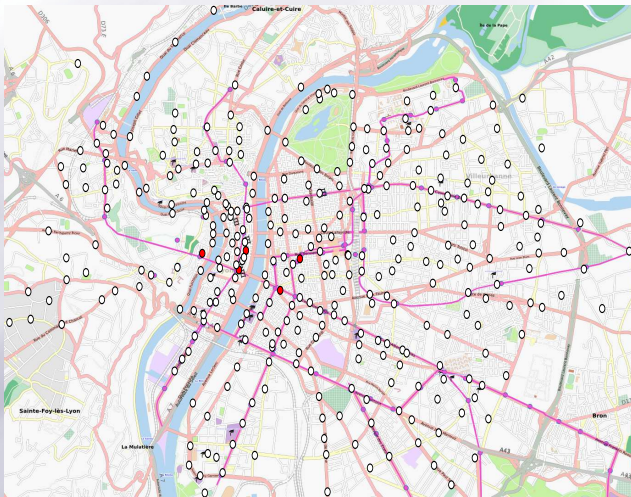
Les trois motifs extraits

De midi à minuit :



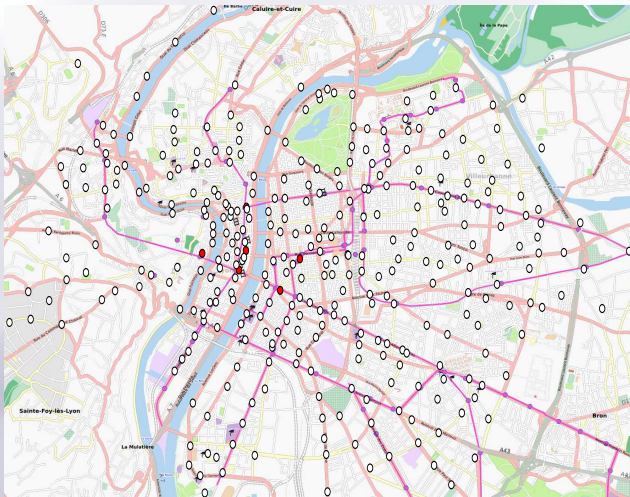
Les trois motifs extraits

De 9h à 18h et de 21 à 24h :



Les trois motifs extraits

De 9h à 18h et de 21 à 24h :



Plan

- 1 Motifs locaux
- 2 Des propriétés intéressantes
- 3 Perspectives

Perspectives

Agglomérer les motifs

À partir d'une grande collection de petits motifs locaux, des motifs tolérant plus d'exceptions, et donc *plus gros*, sont obtenus par *agglomération* (problème de conservations des propriétés voulues).

Règles d'association impliquant plusieurs dimensions

Une règle comme « Le samedi et le dimanche, les usagers Vélo'v vont au Parc de la Tête d'Or » a une pertinence qui peut être quantifiée. Plutôt que d'émettre des hypothèses, *toutes* les règles suffisamment pertinentes sont générées depuis les données.

Perspectives

Agglomérer les motifs

À partir d'une grande collection de petits motifs locaux, des motifs tolérant plus d'exceptions, et donc *plus gros*, sont obtenus par *agglomération* (problème de conservations des propriétés voulues).

Règles d'association impliquant plusieurs dimensions

Une règle comme « Le samedi et le dimanche, les usagers Vélo'v vont au Parc de la Tête d'Or » a une pertinence qui peut être quantifiée. Plutôt que d'émettre des hypothèses, *toutes* les règles suffisamment pertinentes sont générées depuis les données.