

Développement d'un algorithme de détection de communautés spatiales

Serge Lhomme

Université Paris-Est Créteil (UPEC), Lab'Urba EA 3482

8 novembre 2017

La détection de communautés

Définitions

Définition usuelle mais pouvant être trompeuse

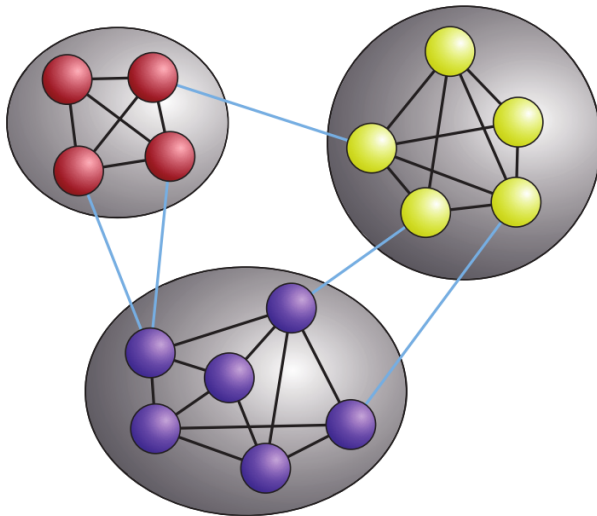
Une communauté regroupe des éléments fortement connectés entre eux et peu connectés aux autres communautés.

Définition personnelle préférée

Une communauté regroupe des éléments qui échangent (partagent) plus que ce qui pouvait être attendu.

La détection de communautés

Image d'Epinal



La détection de communautés

Mesure de la modularité

La modularité consiste à comparer la quantité de relations observées (A_{ij}) au sein d'une communauté avec celle estimée par un modèle de référence (P_{ij}). Toutes les valeurs obtenues par les communautés C de la partition P étudiée sont additionnées.

Modularité

$$Q = \frac{1}{2m} \sum_{C \in P} \sum_{i,j \in C} (A_{ij} - P_{ij})$$

Estimation

$$P_{ij} = k_i \times k_j$$

La détection de communautés

Avantages et limites pour la géographie

Avantages

Si les données utilisées peuvent être formalisées sous la forme d'un graphe, ces méthodes semblent pertinentes pour effectuer des partitions fondées sur les relations entre les lieux et ainsi identifier des sous-systèmes territoriaux.

Limites

Le modèle de référence ne tient pas compte de l'effet probable de l'espace sur les relations. Les partitions obtenues peuvent donc être considérées comme biaisées ou produisant des résultats triviaux.

La détection de communautés spatiales

La modularité spatiale

La modularité spatiale compare les valeurs observées (A_{ij}) pour chaque relation d'une communauté avec leurs valeurs estimées (P_{ij}^{spa}). Les valeurs estimées tiennent compte de l'effet de l'espace sur l'intensité des relations.

Modularité spatiale

$$Q^{spa} = \sum_{C \in P} \sum_{i,j \in C} \left(\frac{A_{ij}}{m} - \frac{P_{ij}^{spa}}{m^{spa}} \right)$$

Estimation

$$P_{ij}^{spa} = k_i \times k_j \times f(d_{ij})$$

La détection de communautés spatiales

Les travaux pionniers

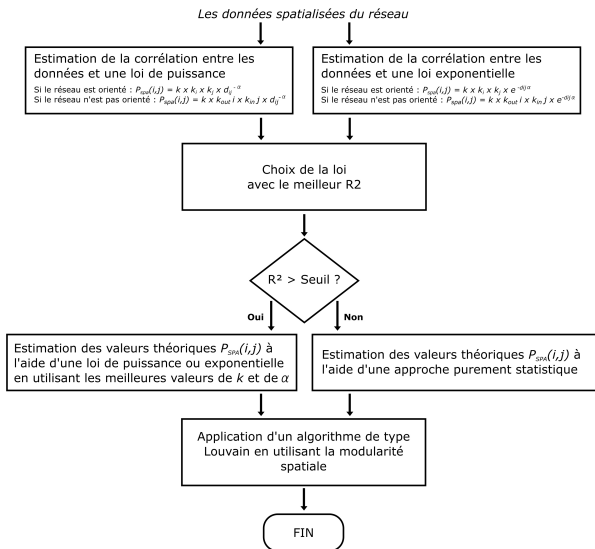
- Expert et al., 2011
 - ▶ Première proposition appliquée aux données de téléphonie mobile
 - ▶ Calibration des valeurs estimées par rapport au volume total
 - ▶ Algorithme de Louvain pour maximiser la modularité spatiale
- Cazabet et al., 2017
 - ▶ Application aux données Vélo'v
 - ▶ Calibration des valeurs estimées par rapport aux marges
 - ▶ Algorithme de Louvain pour maximiser la modularité spatiale

Attention

Ces approches constituent une généralisation des travaux de Newman et Girvan. L'absence de modèles spatiaux de référence connus peut alors conduire à des aberrations : elles donnent des solutions identiques aux approches classiques lorsque l'espace ne joue aucun rôle sur l'intensité des relations.

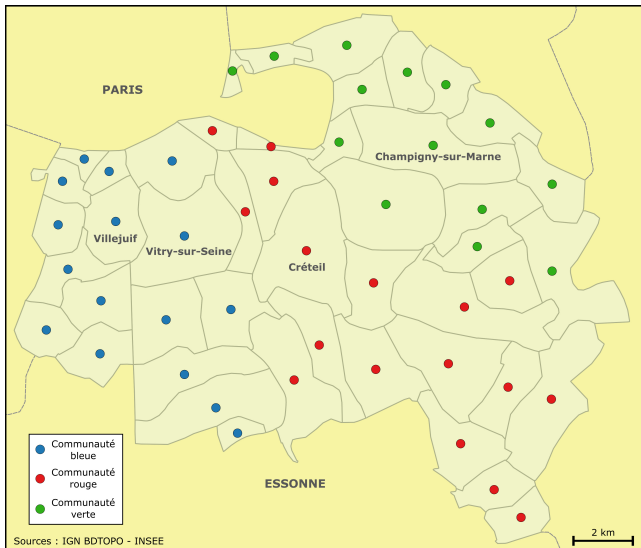
La détection de communautés spatiales

Une proposition alternative : Travailler avec des modèles spatiaux explicites



Résultats

Les flux domicile-travail dans le Val-de-Marne

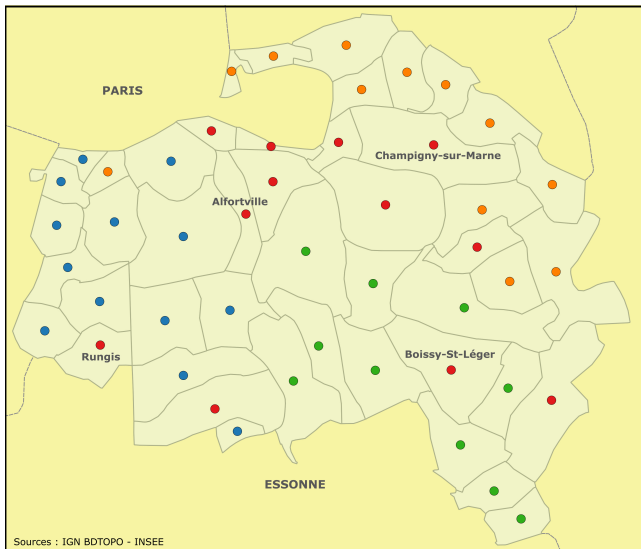


Application de Louvain (modularité classique) : pas de surprise.



Résultats

Les flux domicile-travail dans le Val-de-Marne



Application de l'algorithme développé.

Conclusion

Comme les travaux pionniers portant sur la détection de communautés spatiales, la solution développée permet de faire face à certains biais. Les communautés apparaissent ainsi moins contraintes spatialement.

Néanmoins, malgré l'utilisation de modèles spatiaux explicites, l'interprétation reste compliquée.

De manière générale, les différents résultats obtenus en fonction des choix de modélisation interrogent la sensibilité des méthodes de détection de communautés.

La perspective principale sera d'adapter la notion de communautés spatiales à des méthodes de détection de communautés recouvrantes.