

Évaluation de la qualité des méthodes de sélection des “flux majeurs” et des “flux dominants” pour identifier des structures relationnelles préférentielles

Serge Lhomme

Maître de conférences en Géographie

<http://sergelhomme.fr/>
serge.lhomme@u-pec.fr

6 février 2019

1 Contexte de la recherche

2 Les données

3 Résultats

4 Conclusion

Contexte de la recherche

Objectif de la recherche : partitionner les territoires

En géographie, il est particulièrement intéressant d'analyser les relations existant entre les lieux afin par exemple d'identifier des groupes de lieux qui échangent beaucoup entre eux : des structures relationnelles préférentielles.

En effet ces groupes permettent de partitionner simplement le territoire en : différents systèmes territoriaux, différentes régions...

De par leur simplicité, les méthodes d'analyse de flux comme celles de sélection des flux dominants ou des flux majeurs sont très utilisées alors même qu'il n'existe pas de critères simples et explicites permettant d'en évaluer la qualité.

Cette recherche se donne donc pour objectif d'évaluer la qualité de ces méthodes lorsque celles-ci sont utilisées pour identifier des structures relationnelles préférentielles.

Contexte de la recherche

Les flux dominants et les flux majeurs

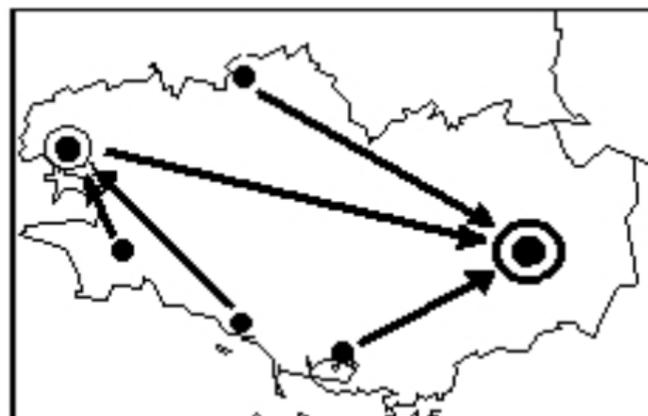
Développée par Nystuen et Dacey (1961) pour mettre en évidence des phénomènes de hiérarchie entre les lieux, la méthode des flux dominants s'appuie simplement sur la sélection du flux le plus important émis par chaque lieu. Pour chaque couple de lieux alors identifié, les relations de domination sont déterminées en comparant la somme des flux reçus.

La méthode des flux dominants peut être utilisée pour créer simplement des partitions des espaces étudiés. Néanmoins, cette méthode est très critiquée car elle ignore de nombreuses relations, dont certaines peuvent être considérées comme majeures.

C'est pourquoi d'autres options ont été développées afin d'assouplir ce critère de sélection. Ce sont les méthodes de sélection des flux majeurs. Dans ce cadre, les sélections peuvent être effectuées globalement ou localement ou encore de manière relative ou absolue...

Contexte de la recherche

Les flux dominants et les flux majeurs



Méthode de Nyusten
& Dacey

→ Flux dominant

⊙ Ville dominante

○ Ville relais

● Ville dominée

Contexte de la recherche

La détection de communautés

Définition usuelle

Les algorithmes de détection de communautés cherchent à identifier les sommets qui sont fortement connectés entre eux et peu connectés aux autres sommets et créent une partition du graphe étudié.

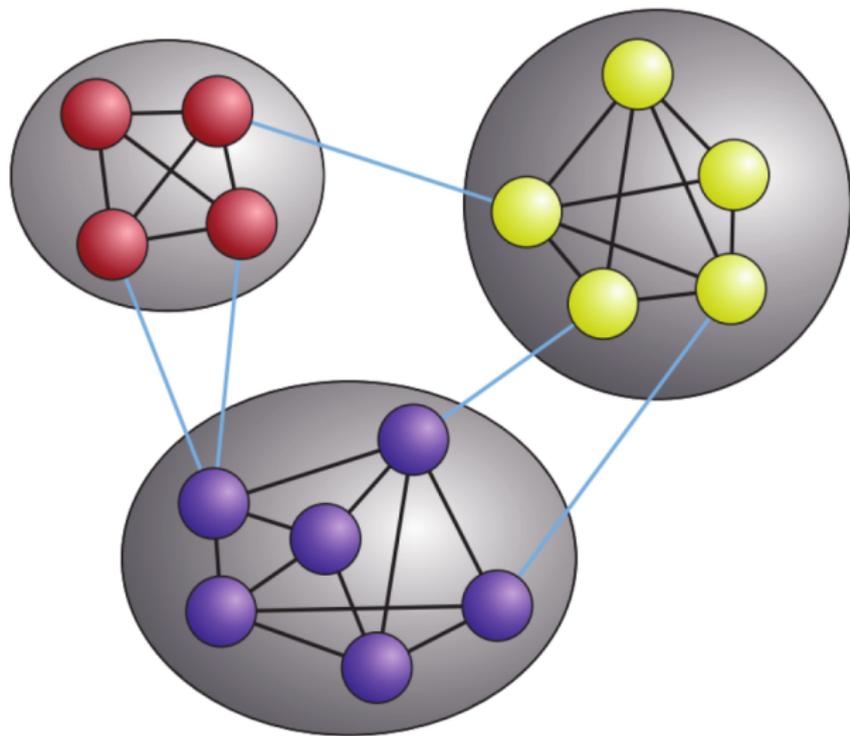
Optimiser la qualité de la partition

Pour évaluer la qualité d'une partition, on utilise une fonction de qualité donnant un score à la partition étudiée. La plupart des algorithmes de détection tente alors d'obtenir le meilleur score possible et cherche ainsi la partition qui maximise cette fonction de qualité.

Il existe plusieurs fonctions de qualité, mais il semble que la fonction la plus utilisée actuellement soit celle de Newman et Girvan (2004) : la modularité.

Contexte de la recherche

La détection de communautés



Contexte de la recherche

La modularité

La modularité

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

La modularité est une mesure comprise entre -1 et 1.

La partition triviale regroupant l'ensemble des sommets étudiés dans la même communauté a une valeur nulle. En théorie, les partitions obtenant des valeurs positives de modularité peuvent être considérées comme intéressantes.

Dans les faits, les valeurs obtenues par les algorithmes de détection de communautés sont souvent comprises entre 0,3 et 0,7. Une valeur inférieure à 0,3 peut donc être considérée comme faible.

Les données

Pour cette recherche, les données de mobilité professionnelle intercommunale (les flux domicile-travail entre les communes françaises) publiées par l'INSEE (les données MobPro) ont été utilisées.

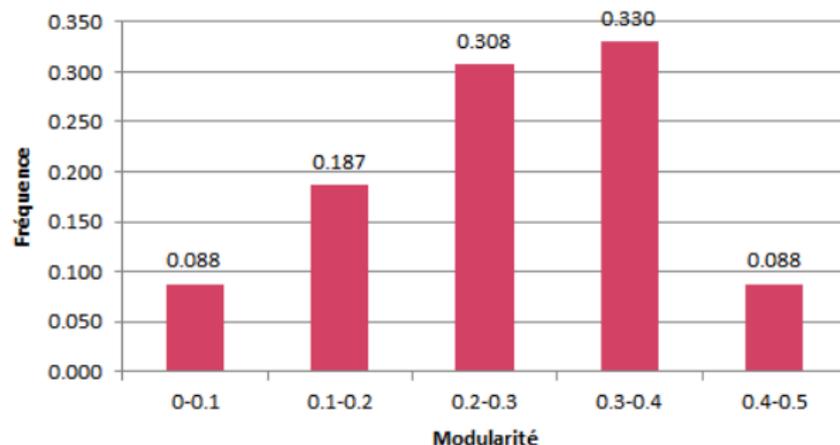
En effet, ces données de déplacements fréquents et réguliers sont de plus en plus mobilisées en géographie pour étudier des territoires.

91 départements de France métropolitaine ont alors été étudiés de manière indépendante afin de produire autant d'analyses indépendantes.

Ces analyses concernent des territoires très différents, que ce soit en termes de nombre de communes (de 36 à 895), de superficie (de 176 à 10 725 km²) ou encore d'organisation spatiale. Les réseaux étudiés offrent donc des configurations variées.

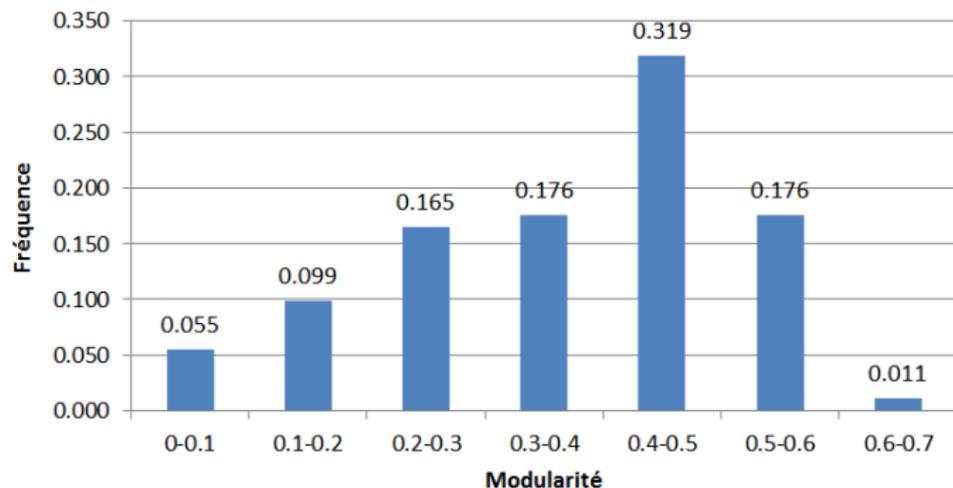
Résultats

Des valeurs particulièrement faibles de modularité pour les flux majeurs



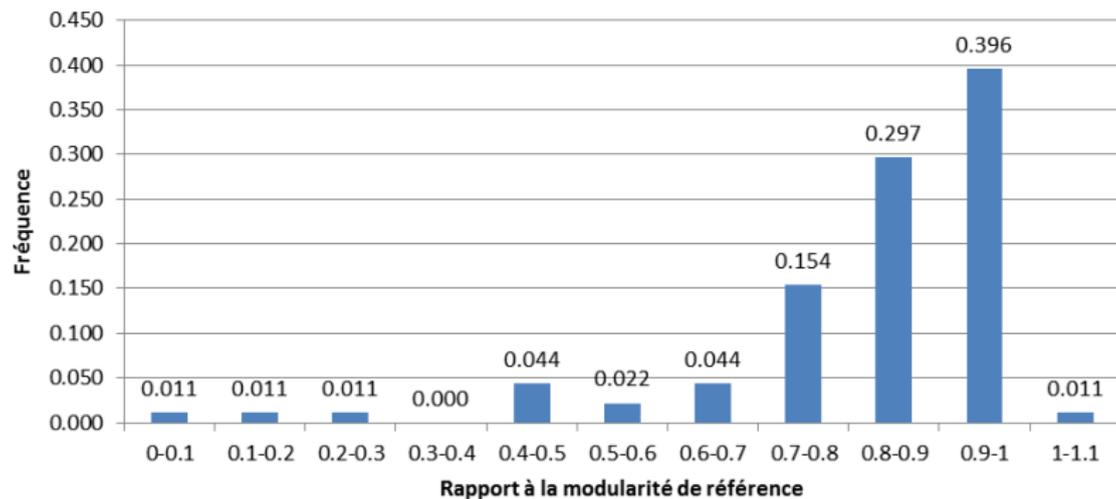
Résultats

Des résultats meilleurs pour les flux dominants



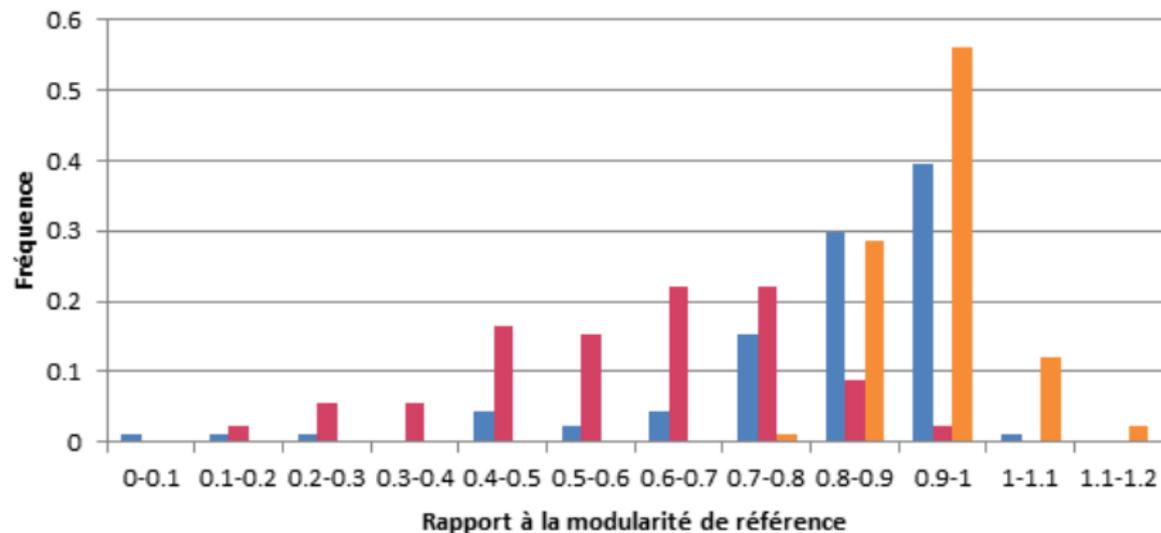
Résultats

Des résultats pas si mauvais à première vue



Résultats

Des résultats inquiétants en comparaison à un simple du clustering spatial



Conclusion

Les résultats obtenus par les méthodes de sélection des flux majeurs ou des flux dominants ne peuvent pas être considérés comme satisfaisants lorsque celles-ci sont utilisées pour identifier des structures relationnelles préférentielles.

Cela ne présage en rien sur la qualité de ces méthodes pour identifier des phénomènes de hiérarchie entre les lieux. Ces résultats mettent en exergue l'importance d'utiliser ces méthodes avec précaution et de bien délimiter leurs champs d'application.

Les résultats obtenus par l'algorithme de clustering spatial interrogent la pertinence des algorithmes de détection de communautés lorsque ceux-ci sont appliqués sur des données relationnelles brutes.

En effet, lorsque la modularité obtenue est inférieure à 0,35 l'algorithme des k-moyennes donne des résultats meilleurs que l'algorithme « walktrap » dans 50% des cas.