

La cartographie d'une sélection globale de flux, entre 'significativité' et 'densité'

*Mapping a global selection of flows, between values 'significance' and features
'density'*

Françoise Bahoken

**Édition électronique**URL : <http://netcom.revues.org/2565>

DOI : 10.4000/netcom.2565

ISSN : 2431-210X

Éditeur

Netcom Association

Édition imprimée

Date de publication : 16 décembre 2016

Pagination : 315-340

ISSN : 0987-6014

Référence électronique

Françoise Bahoken, « La cartographie d'une sélection globale de flux, entre 'significativité' et 'densité' », *Netcom* [En ligne], 30-3/4 | 2016, mis en ligne le 23 mars 2017, consulté le 03 avril 2017. URL : <http://netcom.revues.org/2565> ; DOI : 10.4000/netcom.2565



Netcom – Réseaux, communication et territoires est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

VARIA

LA CARTOGRAPHIE D'UNE SELECTION GLOBALE DE FLUX, ENTRE 'SIGNIFICATIVITE' ET 'DENSITE'

BAHOKEN FRANÇOISE¹

Résumé – *L'une des principales réponses à la réduction de la complexité graphique intrinsèque à la géo visualisation de flux (ou de mouvements) consiste à sélectionner avec « parcimonie » l'information à cartographier. Si plusieurs approches sont envisageables, l'une d'entre elles consiste à mobiliser un critère unique (a), endogène à la matrice de flux sous-jacente. Nous présentons la mise en œuvre de ce critère que l'on qualifie de « global », en suivant une suggestion de W. Tobler (1982), avant de proposer un (double) critère unique. Celui-ci résulte de l'articulation des deux principaux objectifs de la sélection à savoir : la significativité des valeurs des liens, donc des flux, ou la densité des liens, donc des individus. La validation empirique est réalisée sur une matrice de navettes intercommunales du département de l'Isère (France).*

Mots-clés - *matrice, géo visualisation, flux, liens, mouvements, sélection, critère global, Lorenz.*

Abstract – *One of the main responses for reducing the cartographic complexity of a flow (or a movements) map is to select the information with "parsimony". If several approaches remain available, one of them is to mobilize a single criterion (a), endogenous to the matrix. We present here the implementation of such unique selection criteria that we called "global criteria". From a solution initially proposed by Tobler (1982), we analyze the flow values (Fij) statistical distribution by mobilizing graphical tools. That leads us to propose a (double) single criterion so as to represent only these flow values responding to a significance or a density matrix objective. The empiric validation is made on a French commuters flow matrix.*

Keywords - *matrix, flow mapping, links, movements, selection, global criteria, Lorenz.*

¹ Géographe et cartographe, chargée de recherches à l'Université Paris-est, AME, SLOTT, IFSTTAR, associée à l'UMR 8504 GEOGRAPHIE-CITES / PARIS. E-mail : francoise.bahoken@ifsttar.fr

INTRODUCTION

Les données quantitatives numériques, en particulier celles dont la structure est relationnelle, connaissent un engouement renouvelé dans le contexte actuel. Ce nouvel intérêt émerge, en effet, à la faveur du développement de nouveaux dispositifs techniques de collecte, d'archivage mais aussi d'exploitation ; l'ensemble contribuant à étendre les possibilités de leur analyse sur le plan méthodologique, technique mais aussi thématique. La formalisation de ces données, dans le contexte de l'analyse géographique, est celle d'un réseau : les lieux sont assimilés à des acteurs localisés, mis en relation par un lien correspondant à un processus donné. Ce réseau est, par exemple, celui d'un ensemble de lieux touristiques pratiqués dans une ville par des individus qui n'y résident pas ; celui de zones d'emplois formées (par un agrégat de communes à l'échelle d'un département) par la polarisation de navetteurs provenant de communes davantage résidentielles.

L'information relative aux réseaux spatiaux...

L'information relative à ces réseaux spatiaux correspond soit à la trace d'une infrastructure matérielle, soit à celle d'un transfert immatériel entre deux lieux situés, en l'occurrence, dans un espace géographique. Elle est traditionnellement collectée² à l'échelle individuelle, de manière rétrospective, à partir d'un bulletin. Ce dispositif présente plusieurs limites : sa mise en œuvre est fastidieuse, onéreuse et le résultat revêt un caractère lacunaire. Normalisée de fait, l'information dont on dispose pour l'analyse est souvent agrégée, car elle résulte de différentes procédures d'anonymisation, de filtrages spatiaux et/ou temporels qui peuvent être appréhendés comme une dégradation, comme une sélection rendue nécessaire par l'impossibilité de pouvoir saisir l'exhaustivité des déplacements individuels (Bavoux et al., 2005). Pour remédier à ces limites, différents auteurs se tournent vers l'exploitation de données issues de méthodes alternatives de collecte de déplacements individuels. Les modèles qui en résultent fournissent non pas la trace des déplacements en eux-mêmes mais un marqueur de ces derniers. Par exemple, pour un territoire donné, la collecte peut porter sur l'ensemble des fichiers d'abonnements ou de désabonnements à des organismes fournisseur d'énergie, ou encore, pour un opérateur donné, sur l'ensemble des traces d'activités de téléphones portables. Ces données ne sauraient toutefois prétendre saisir l'exhaustivité des déplacements. En effet, seuls les abonnés aux services concernés sont comptabilisés et non l'ensemble de la population. Ces fichiers résultent donc d'un échantillonnage, non contrôlé, de la population enquêtée. C'est à partir de ces marqueurs que sont ensuite reconstituées des traces de déplacements individuels, enfin agrégées à différentes échelles spatio-temporelles (Oltéanu-Raimond et al. 2013). Leur principal intérêt tient dans l'exhaustivité de la population analysée, c'est-à-dire dans la quantité de données qu'elles représentent, à des niveaux de résolution spatiale et/ou temporelle assez fins, souvent intra communaux et en zone urbaine (les maillages des capteurs des télécommunications sont généralement plus lâches en zone rurale). Autrement dit, elles sont plus intéressantes pour ce qui est du nombre d'individus concerné (i.e. d'abonnés

² Le fonctionnement de ces méthodes de collecte est décrit dans l'Annexe A-1 (Bahoken, 2016).

au service de téléphonie mobile) que pour leur richesse sémantique (i.e. les caractéristiques socio-économiques des utilisateurs de mobiles). Si elles permettent effectivement la détection de communautés spatiales dans le cadre d'analyse de dynamiques urbaines (Blondel et al., 2008 ; Ratti et al., 2010), la prédiction de la mobilité générale (Calabrese et al., 2010) ou touristique (Oltéanu-Raimond et al. 2011), leur inférence dépend, avant tout, de la structure hétérogène du réseau de télécommunications. À l'inverse, les données collectées de manière plus traditionnelle, si elles sont moins massives (le nombre d'individus concernés fait souvent l'objet d'une sélection par échantillonnage), elles n'en demeurent pas moins intéressantes sur le plan qualitatif (quantité d'information décrite par l'échantillon de données).

... entre densité de données et significativité de l'information

Dans le contexte de l'analyse géographique, le principal inconvénient de ces nouvelles données réside dans leur pauvreté thématique. Généralement archivées dans une matrice de connexité (ou d'adjacence), leurs valeurs renseignent sur l'existence ou non d'une relation entre des couples de lieux. Si ce format binaire en fait un objet d'étude privilégié dans le cadre de la Théorie des graphes³, l'absence d'information sur les caractéristiques des liens limite, de fait, l'analyse à l'examen de leur seule existence (ou de traces de), de leur fréquence, du nombre de degrés entrants, sortants et de toutes autres mesures dédiées à l'analyse de réseaux d'acteurs ou à la détection d'agrégats.

Or, l'intérêt d'une analyse géographique réside bien dans la mise en évidence des motifs spatiaux du phénomène (social, environnemental, ...) dont il est question. Dans le cas particulier des mobilités spatiales (résidentielles, professionnelles, de loisirs,...), différents auteurs ont pu montrer que la distribution des lieux est une résultante des interactions spatiales qui se manifestent à différents niveaux d'intensité, aussi que ces interactions sont fonction à la fois des catégories de lieux (type, masse, rang ...) que de leurs positions géographiques relatives au sein de l'espace d'étude. De ce point de vue, si les données traditionnelles apparaissent lacunaires en termes de densité, elles n'en demeurent pas moins riches sur le plan thématique, ni moins complexes à mobiliser sur le plan méthodologique lorsqu'elles concernent de nombreux (couples de) lieux. Leur intérêt portant bien sur la quantité d'information qu'elles décrivent à un niveau élémentaire, un organisme tel que l'INSEE autorise désormais librement l'accès à différents fichiers de détails⁴ fort volumineux. Ainsi, en présence de corpus de données portant sur des réseaux spatiaux qui sont complexes, parce qu'ils concernent une population (ou une quantité d'information) importante, se pose la question de l'information à analyser. Cette question peut être appréhendée du point de vue de la réduction de la masse d'information disponible (par agrégation, décomposition, *clustering* ...) ou de sa pertinence, par le biais de la sélection des individus renseignant sur une information *significative*.

³ Rappelons que l'un des objectifs majeurs de la Théorie des graphes est de détecter des communautés au sein d'acteurs qui ne sont pas spatialisés, la composante spatiale en étant absente.

⁴ <http://www.insee.fr/fr/bases-de-donnees/default.asp?page=open-data/open-data-utilisation.htm>

Cette problématique de la sélection de l'information significative n'étant pas nouvelle, nous proposons de l'examiner au regard des impératifs de l'analyse géographique de données, c'est-à-dire dans sa capacité à mettre en évidence des motifs spatiaux. Pour cela, nous allons généraliser le raisonnement usuel qui consiste à déterminer une valeur (α) unique de sélection des valeurs pertinentes, dans le cadre d'une approche que nous qualifions de *globale*⁵ (Bahoken, 2016a : 325-326). En repartant d'une suggestion de W. Tobler (1982), nous allons combiner – plutôt que les opposer – les deux visions existantes d'un même objet, en faisant l'hypothèse qu'elles sont complémentaires. Ainsi, pour un même couple de lieux (i, j), ces deux visions concernent, pour l'une, l'existence d'un lien (L_{ij}) telle qu'inscrite dans le champ de la Théorie des graphes et, pour l'autre, l'analyse de l'intensité (F_{ij}) de ce lien, donc du flux, telle que mobilisée en analyse spatiale. Cette posture va nous conduire à adopter un raisonnement mixte (densité*intensité), autrement dit à proposer l'introduction d'un double critère global de sélection des flux (Bahoken, 2016a : 325-345) que nous allons examiner dans le cadre d'une représentation cartographique de ces données.

Nous présentons, dans un premier temps, un état de la question de la sélection de l'information de flux, qui va nous conduire à présenter les données d'illustration : des navettes domicile-travail en Isère (2010). Dans un second temps, nous focalisons l'attention sur le critère de sélection global de l'information que nous allons étendre dans la troisième section, en proposant un double critère qui agit sur la part d'information significative (en termes d'intensité) en fonction de la part de liens (densité).

1. ETAT DE LA QUESTION DE LA SELECTION DES FLUX

La complexité des données qui portent sur la mise en relation des lieux constitue un réel problème pour leur analyse. Issues de l'exploitation de matrices origine-destination (OD) nécessairement denses, l'efficacité de la représentation qui en résulte, sous la forme de carte de flux, n'est plus assurée, en raison d'une complexité graphique avérée. Le raisonnement qui est traditionnellement adopté pour réduire cette complexité graphique consiste à agir lors du traitement numérique des données, c'est-à-dire en amont du processus de représentation ; il ne s'intéresse pas aux aspects cartographiques ou sémiologiques, c'est pourquoi ces derniers ne seront pas examinés ici.

⁵ L'approche *globale* signifie que le critère de sélection retenu agira à l'échelle de la matrice entière, sur l'ensemble des cellules du cœur de la matrice. Elle diffère de l'approche *locale* qui agit par ligne(s) et/ou par colonne(s), donc sur les lieux car elle présente un objectif différent. L'approche *locale* a pour objectif de représenter la structure hiérarchique des *localités* (des lieux de la zone d'étude). La plus connue est la méthode de Nystuen et Dacey (1961) qui conduit à ne cartographier que la sélection de liaisons (et non de flux) qui sont issues d'une analyse des flux dominants fondée sur une acception de l'attractivité des lieux (approximée par la somme marginale en colonne), de leur hiérarchie.

1.1. Les principales orientations méthodologiques

Quatre grandes orientations méthodologiques sont rencontrées dans la littérature. La première direction consiste à réduire la résolution de l'information, en utilisant des méthodes de regroupement liées à l'agrégation des données. Elles conduisent à réduire le nombre de lieux (changement de résolution) en agissant soit sur les attributs de ces lieux (agrégation territoriale ou spatiale) et/ou soit sur les liens (typologie par les flux). Assimilées à des méthodes de partitionnement, elles conduisent généralement à un changement conceptuel fort qui entraîne une modification de l'objet d'étude. La cartographie de flux intercommunaux, par exemple, diffère fondamentalement de celle de flux interrégionaux. Sur le plan géographique : l'échelle est locale / régionale ; au niveau territorial de la gouvernance, elle est communale / régionale ; du point de vue de l'interprétation de la figure : la vision est locale / nationale et du type de mobilités spatiales concernées : les flux s'expriment à courte / longue distance.

La seconde direction est purement graphique : introduite par Holten (2006), elle consiste à fusionner graphiquement les objets linéaires de façon à représenter des « routes principales ».

Les troisième et quatrième directions vers la réduction de la complexité graphique de la carte de flux sont suggérées par Beguin et Pumain (1994). Elles s'appuient sur des considérations qui sont davantage analytiques que graphiques. « *Les représentations cartographiques directes de tableaux d'échanges deviennent difficiles à construire et à lire lorsque le nombre des unités géographiques considérées est grand. Dans ce cas, on a intérêt à traiter l'information statistique avant de la représenter (par exemple en identifiant les flux principaux, ou en ajustant le tableau d'échanges par un modèle gravitaire pour faire apparaître par différence des flux préférentiels ou des effets de barrière)* » (Beguin, Pumain, 1994 : 118). Les auteures mentionnent, à titre d'exemples, les cadres d'analyse statistique descriptive et explicative des flux.

L'approche descriptive correspond à la troisième possibilité. Elle consiste à sélectionner les flux de manière à ne représenter que les « *principaux* », revenant à distinguer « *le bon grain de l'ivraie* » (Tobler, 1969)⁶. L'approche explicative, envisagée comme la quatrième possibilité, conduit à ne représenter qu'une sélection de flux « *préférentiels* » (à l'inverse de flux soumis à des « *effets de barrière* ») qui ont été estimés à la suite d'une étape de modélisation. Ces deux possibilités sont menées à un niveau global, elles s'opposent en ce sens à une sélection de l'information qui agit localement, comme dans le cas de l'analyse dite de flux dominants (Nystuen et Dacey, 1961).

La sélection dite de « *flux majeurs* », en référence au premier critère de sélection proposé par les auteurs correspond à une sélection de l'information qui serait menée à l'échelle locale (Bahoken, 2016a : 347-348). Elle consiste en effet à ne retenir que le plus gros flux émis par chacun des (i) lieux d'origine, le $\text{Max}_i(\text{F}_{ij})$, et non le plus gros flux

⁶ L'auteur considère que la sélection des flux répond à quatre objectifs : i) la « *distinction du bon grain de l'ivraie* » ; ii) l'identification de ce qui peut être expliqué de ce qui ne peut l'être ; iii) pour les séries chronologiques, la mise en évidence d'effets de saisonnalité ; iv) pour les séries spatiales, la représentation de flux qui s'expriment à l'échelle locale, régionale, nationale ou mondiale, dont les effets sont globaux (Tobler, 1969 : 234-253).

observé à l'échelle de la matrice : $(F_{ij}) > \alpha$. Sur le plan méthodologique, ce *critère local* conduit à conserver l'ensemble des lieux de la zone d'étude, à l'inverse d'un *critère global* qui ne retient que les lieux formant le sous-système de flux « *principaux* » ou « *préférentiels* », intéressant donc la troisième possibilité.

Quoi qu'il en soit, toutes ces solutions sont fondées sur le principe de l'identification d'une sélection de valeurs de flux considérées comme *significatives*, d'une part, au regard de critères préalablement définis (les flux sont « *principaux* ») et, d'autre part, compte tenu de leur pouvoir descriptif (explicatif ou prédictif) des motifs de flux analysés.

Dans ce registre, lorsqu'elle est réalisée dans cadre descriptif, l'approche la plus courante consiste à éliminer brutalement une partie de l'information, afin de ne représenter que les flux considérés comme *significatifs* par l'auteur de la carte. C'est le cas de l'essentiel des images des mobilités spatiales qui ne représentent que les $(\alpha = n)$ flux les plus importants $F_{ij} > (\alpha)$, exprimé en valeur, en tonnage, en volume... La mise en œuvre de critère s'inscrit dans une démarche que nous présentons dans la section suivante.

1.2. La sélection de l'information, selon un critère (α) unique

Il est important de noter au préalable que la part de flux *significatifs* sélectionnés est susceptible de varier fortement en fonction de l'existence ou non d'une présélection des flux réalisée dans le cadre d'un raisonnement selon la focale⁷ de la matrice. Dans notre cas, nous mobilisons donc le cœur d'une matrice (F_{ij}) complète qui s'exprime entre les N lieux d'une zone d'étude. Le nombre de cases renseignées est donc théoriquement de N^2 .

Sur le plan théorique, l'objectif d'estimer un critère (α) unique de sélection des valeurs de flux à représenter est un choix qui revêt une double conséquence. Il conduit à accepter l'idée selon laquelle, d'une part, certains lieux d'origine ou de destination des flux ne présenteront pas de lien (entrant ou sortant) sur la carte ; d'autre part, qu'une partie des couples de lieux n'a pas d'intérêt pour l'analyse.

L'approche *globale* de la sélection des flux, mobilisant un critère unique, vise en effet, à réduire la matrice aux seuls flux *significatifs* qui le sont, parce qu'ils répondent à un critère de représentativité de l'information disponible. C'est pourquoi elle conduit nécessairement à l'élimination d'une partie des valeurs de flux et, par conséquent, des lieux. Il en découle que les flux intéressant les couples de lieux jugés *non significatifs* dans le système cartographié seront éliminés de la représentation (comme la cartographie des flux « *préférentiels* » ne conserve que les couples de lieux concernés).

La question est donc de savoir comment s'assurer que l'on représente l'information de flux *significative*, en appliquant un critère unique de sélection qui agit à

⁷ La sélection des flux selon une focale de la matrice est une approche graphique et visuelle qui consiste à choisir la partie de la matrice à analyser. Elle s'appuie sur le fait qu'une matrice de flux peut être analysée en mobilisant : ses lignes et/ou ses colonnes et/ou son cœur ; ces trois parties correspondant respectivement à une analyse menée du point de vue des lieux d'origine et/ou de ceux de destination et/ou des valeurs de flux qui s'expriment entre les lieux d'origine et de destination.

l'échelle de l'ensemble de la matrice, en assurant la comparabilité (qui témoigne de la robustesse) du résultat obtenu.

Le raisonnement sous-jacent à cette interrogation est en fait identique à celui qui conduit à la réalisation de discrétisations statistiques pour des données de stocks. En effet, l'application d'un seuil unique correspond formellement à une dichotomie, autrement dit à une segmentation de la série de valeurs en deux classes distinctes : celles qui répondent au critère de sélection, par exemple : $F_{ij} > (\alpha)$, et celles qui n'y répondent pas : $F_{ij} \leq (\alpha)$.

Existe-t-il en conséquence une méthode permettant d'estimer la valeur du seuil (α) de manière à ce que le résultat cartographique obtenu puisse être comparable entre deux matrices ? Ce seuil est-il unique ? Pour contribuer à la réflexion, nous mobilisons une matrice de navettes intercommunales.

1.3. Présentation des données

Pour examiner la question de la sélection des flux à représenter, nous mobilisons une matrice de flux portant sur des migrations domicile-travail entre les communes de l'Isère (2010), elle correspond formellement à un réseau immatériel spatial valué.

Les données sont issues du fichier de détail MOBilités PROfessionnelles des individus (MOBPRO) réalisé par l'INSEE pour l'année 2010. Ce fichier propose un ensemble de variables sur un échantillon d'individus actifs de 15 ans ou plus ayant un emploi, se déplaçant quotidiennement entre leur lieu de résidence et leur lieu de travail. Chaque enregistrement correspond à un individu décrit selon ses principales caractéristiques sociodémographiques, celles du ménage auquel il appartient et celles de son déplacement. La localisation des lieux de résidence et travail est la commune, hormis pour les communes de Paris, de Lyon et Marseille renseignées au niveau de l'arrondissement municipal. Seuls les actifs ayant un emploi et résidant dans une commune française sont recensés. Ces emplois sont à temps complet ou à temps partiel, rémunérés ou non ; ils incluent les individus ayant une « activité réduite », les chômeurs, les étudiants, retraités salariés ainsi que les militaires du contingent.

De ce fichier MOBPRO nous en avons extrait les seuls flux internes au département de l'Isère (38). Ces flux élémentaires sont notés (F_{ij}). L'Isère comptant 533 communes, cette matrice est par essence fortement soumise à l'*effet spaghetti* (Bahoken, 2015) puisqu'elle compte (533*532) cases hors diagonale, soit 283 556 valeurs théoriques à représenter. Cette matrice pourtant fermée (elle est limitée aux flux échangés entre les communes du département) ne compte pas le même nombre de lieux d'origine et de destination (i.e. de lignes et de colonnes, ce qui signifie qu'elle n'est pas « carrée ») : les flux concernés s'expriment réellement entre (532*498) communes. C'est pourquoi nous l'avons transformée de sorte que le nombre de lignes et de colonnes soit identique (533*533), diagonale principale incluse. Nous justifions cette transformation sur le plan statistique : une matrice carrée permet de mobiliser sa propriété générale de symétrie que nous allons mobiliser pour calculer un volume bilatéral de navettes ; aussi sur le plan cartographique par le besoin de ne pas représenter un figuré orienté (des têtes de flèches, par exemple), mais un lien ou une bande, comme présenté dans Bahoken (2016b). Pour cela, il est nécessaire que la matrice soit symétrique, ce qui n'est

pas le cas de (F_{ij}) . C'est pourquoi faisons le choix de décomposer cette information numérique, pour examiner plutôt le volume bilatéral de flux, noté (F_{ij}^+) , sa cartographie est présentée sur la figure 1.

Sur le plan sémiologique, nous mobilisons des bandes dont la largeur est proportionnelle à la valeur de (F_{ij}^+) , en application des variables visuelles Taille et Transparence. Elle consiste à ré-échantillonner l'intensité de teinte noire des figurés en fonction de leur valeur de sorte que les plus petits flux (les plus fins ou encore les plus longs) soient moins visibles que les autres. Par ailleurs, pour renforcer la perception des différentes catégories apparentes de figurés, en particulier des plus gros, nous avons suivi la suggestion de W. Tobler (1987) de disposer ceux qui présentent les valeurs les plus importantes au premier plan (il s'agit aussi des figurés qui sont les plus courts et, par conséquent, des plus opaques).

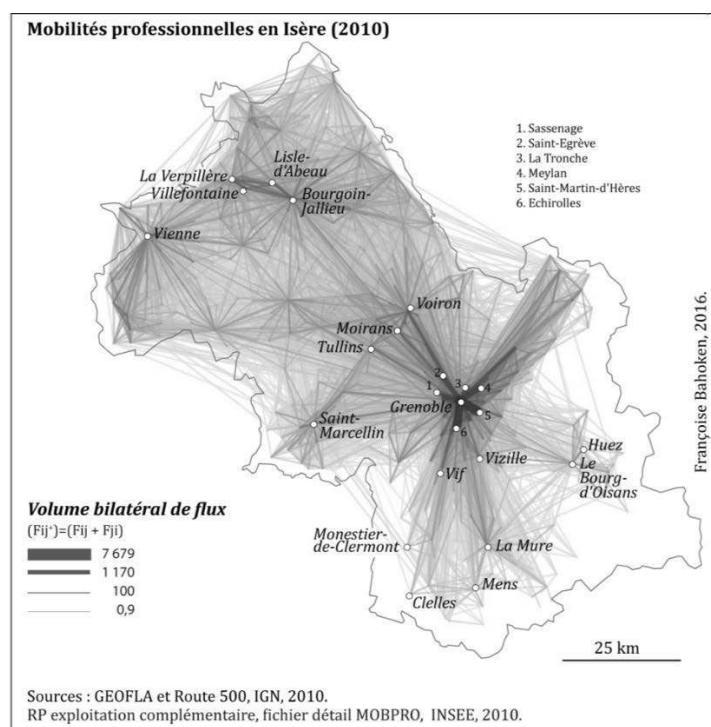


Figure 1: Cartographie de l'ensemble des volumes bilatéraux de flux: (F_{ij}^+) de l'Isère.

Ces navettes s'inscrivant nécessairement dans un territoire, nous mobilisons en outre, pour les besoins de la démonstration, une troisième matrice formée par le produit de la valeur des flux par une acceptation de la distance géographique parcourue (D_{ij}) . Cette troisième matrice correspond formellement à une matrice de *mouvement*⁸ (M_{ij}) , que

⁸ Il s'agit d'une approche numérique d'estimation de ce mouvement, ce dernier pouvant également être approximé sur les plans graphiques et/ou cartographiques.

nous avons également transformée en (Mij^+) . Sa cartographie est présentée sur la figure 2, suivant les mêmes règles de sémiologie graphique que précédemment.

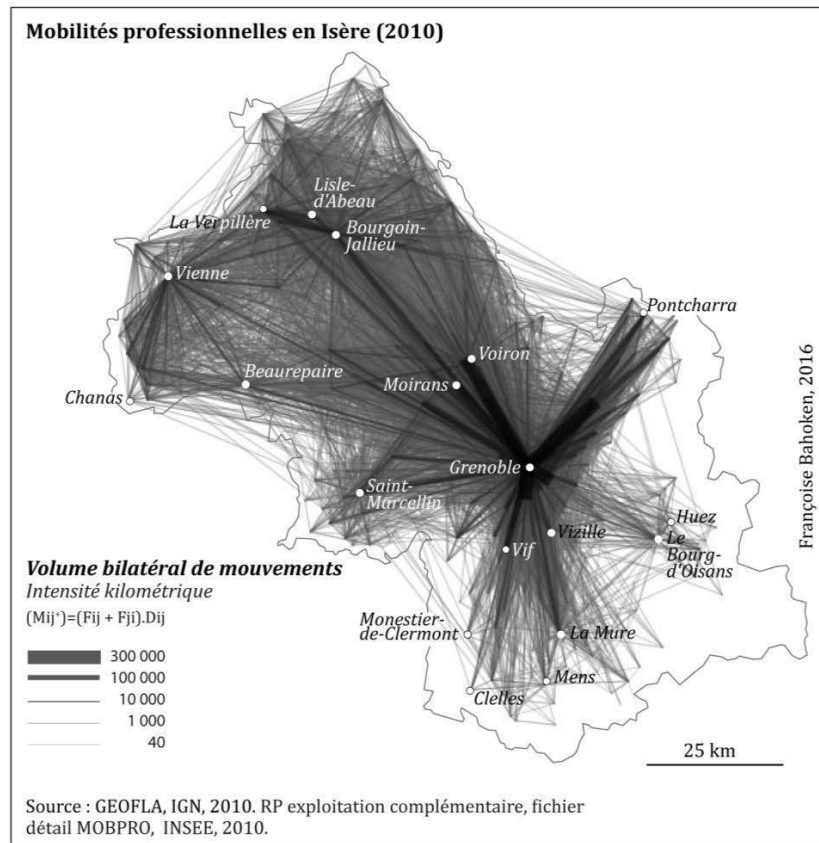


Figure 2 : Cartographie de l'ensemble des volumes bilatéraux de mouvements (Mij^+) de l'Isère.

L'objectif est donc de réduire la complexité des figures 1 et 2, soit la quantité de figurés représentée, tout en assurant la significativité de l'information représentée, en appliquant un *critère unique de sélection globale*. Toutes les images présentées par la suite pourront être comparées avec la figure 1 et la figure 2.

Sur le plan thématique, les données portant sur des navettes, l'objectif de la sélection est de mettre en évidence les pôles d'emplois du département, leur distribution spatiale et l'extension éventuelle de leurs bassins de recrutement, inversement de mettre en évidence des communes davantage résidentielles, voire peu attractives sur le plan économique.

2. LE CRITERE DE SELECTION GLOBALE DES FLUX

La sélection de l'information semble impérative, dans le cadre d'analyse des données relationnelles, qu'elles soient d'ailleurs valuées ou non. Déjà en son temps, E. Ravenstein (1885) n'avait figuré qu'une sélection de flux locaux dans sa fameuse carte sur les courants migratoires. J. Bertin a également suggéré, dans ses principes de sémiologie cartographique, les deux points suivants.

« 2°) *Ne pas représenter les mouvements au-dessous d'une certaine quantité, ce qui allège l'information d'un cortège d'impondérables généralement non significatifs.* 3°) *Ne pas représenter les mouvements au-dessous d'un certain pourcentage. Mais un très faible pourcentage peut cependant donner une quantité importante si la région est très peuplée* ». (Bertin, 1973).

2.1. L'impératif de la sélection des flux

La sélection de l'information qui porte sur des mobilités spatialisées de lieux à lieux est rendue nécessaire pour plusieurs raisons.

D'abord, parce que la densité de la matrice, c'est-à-dire le nombre de lieux impliqués dans le système de mobilités ou, autrement dit, le taux de remplissage des cellules peut être tel que l'image directe des flux qui en résulte est illisible. Cela signifie *a priori* qu'il n'est pas envisageable de représenter toute l'information disponible dans une matrice de flux dense. S. Holland et D. A. Plane (2001) l'ont d'ailleurs bien montré dans leur article dédié aux méthodes de cartographie de flux, mobilisant différents critères statistiques : « *Puisqu'il n'est pas possible de représenter tous les flux, l'approche consiste à les sélectionner, ce qui signifie qu'un certain nombre de flux – souvent la majorité – seront réputés moins saillants quel que soit le critère utilisé pour sélectionner les flux.* »⁹ (Holland, Plane, 2001).

Ensuite, parce que l'objectif de la représentation des flux en question ne doit pas nécessairement conduire à changer d'objet d'étude (d'échelle géographique, de niveau thématique...) ou à mobiliser des données exogènes. En effet, sélectionner des données de flux oblige à l'inverse à focaliser l'attention sur les seuls éléments saillants de la matrice, sur les principaux motifs (*patterns*) des flux et/ou des lieux, en fonction de la thématique considérée et de l'objectif de la figure.

Si la sélection des données de flux répond théoriquement à différents objectifs, nous allons d'emblée restreindre notre propos à deux d'entre eux. Le premier tient au besoin de réduire la complexité de la figure, *l'effet de surcharge graphique* de la figure (Bahoken, 2015). Le second objectif vise à sélectionner l'information de manière robuste, afin que les cartes réalisées puissent être comparées avec d'autres soit parce qu'elles portent sur des catégories de flux relevant de la même zone d'étude, soit parce qu'elles concernent le même type de flux mesuré à différentes dates.

⁹ Cette affirmation de S. Holland et D. Plane, énoncée il y a un une petite quinzaine d'années peut apparaître datée, car elle ne tient pas compte des nouvelles possibilités d'action au niveau sémiologique.

2.2. L'approche classique

Une pratique courante de sélection des flux à représenter consiste à ne retenir que les *Top-n* flux, soit les *n plus grosses valeurs de flux*. Cette procédure étant largement répandue, nous n'allons pas nous y attarder et la présentons succinctement.

Cette méthode est souvent mise en œuvre pour cartographier les flux qui s'expriment à l'échelle mondiale. Le critère utilisé est alors celui des 50 (*Top-50*) voire des 100 (*Top-100*) plus gros flux, supposés concerner les lieux qui présentent une attractivité à l'échelle mondiale. Le critère est d'ailleurs souvent appliqué pour sélectionner les flux, indépendamment du fait qu'ils soient observés du point de vue des liens ou des lieux. N. Cattani (2004) avait ainsi retenu le top-100 (les 100 premières plus grosses valeurs) des flux internationaux de passagers par voie aérienne et C. Grataloup (2011), les 50 plus gros flux mondiaux de marchandises.

Les cartes de flux ainsi réalisées présentent plusieurs caractéristiques : elles décrivent une part réduite de l'information et cela, quelle que soit l'échelle à laquelle la procédure est appliquée et la thématique. Elles éludent également de la représentation un certain nombre de lieux d'autant plus important que la valeur du seuil est élevée. Enfin, leur pertinence n'est pas avérée, surtout à l'échelle locale. En effet, la valeur du critère devant être réalisée en lien avec la résolution des données et la thématique du flux afin de produire, à peu de frais, des cartes qui soient efficaces.

Indépendamment de cela, le principal problème de ce critère réside dans la production de cartes qui ne sont ni reproductibles, ni comparables entre elles, de la même façon qu'une discrétisation statistique réalisée selon un seuillage naturel produit des cartes très intéressantes, mais hélas non comparables.

Pour y remédier, W. Tobler (1982) a suggéré de visualiser la distribution des valeurs de flux au préalable de leur exploitation. Aussi d'adopter à minima « *une stratégie optimale [qui] consiste à retirer tous les flux dont la valeur est inférieure au flux moyen total* » (Tobler, 1982), plus généralement à un paramètre statistique. En suivant la première suggestion de l'auteur, il a pu être démontré une déviation des valeurs des distributions de (F_{ij}^+) et (M_{ij}^+) au regard de la loi normale, y compris lors d'ajustements réalisés par la fonction logarithme (Bahoken, 2016a : 328-331). La figure 3 présente les histogrammes correspondants.

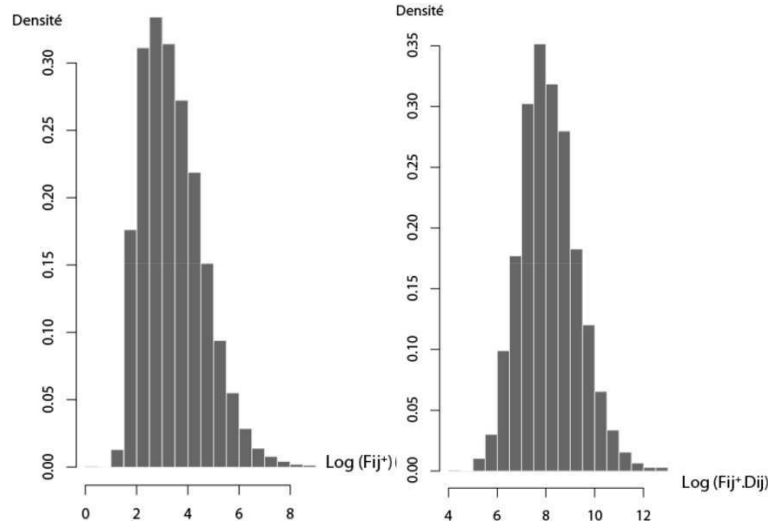


Figure 3 : Histogrammes des valeurs transformées de (F_{ij}^+) et (M_{ij}^+) .

Les déviations visibles sur la figure 3, en particulier pour la distribution de flux (à gauche) devraient interdire de recourir à des méthodes d'analyse qui postulent la normalité de ces valeurs, les résultats ne sauraient être considérés comme totalement satisfaisants.

Deux possibilités d'investigation sont dès lors envisageables. La première d'entre elles consiste à utiliser d'autres fonctions d'ajustement qui seraient mieux adaptées à la forme de la distribution des valeurs de flux. L'investigation serait alors menée à un niveau statistique qui dépasse le cadre de notre analyse, c'est pourquoi elle ne sera pas engagée. La seconde possibilité consiste à examiner la courbe de fréquences cumulées des valeurs de flux (F_{ij}) , au lieu de celle des fréquences simples. Nous proposons d'examiner cette seconde possibilité en mobilisant un troisième outil graphique : la courbe de concentration (ou courbe de Lorenz).

2.3. Le recours à une courbe de Lorenz

Nous mobilisons la courbe de Lorenz, car elle permet d'articuler sur une même figure, la représentation de la répartition des fréquences cumulées des individus et de leurs mesures, en pourcentage de l'information totale qu'ils représentent respectivement. Dans notre cadre d'analyse des flux, les individus correspondent aux liens que nous notons (L_{ij}) . Ces liens étant toujours valués, les valeurs sont celles du cœur de la matrice, initialement des volumes bilatéraux de flux (F_{ij}^+) ou de mouvements (M_{ij}^+) .

La courbe de Lorenz est traditionnellement utilisée pour évaluer l'inégalité d'une distribution statistique qui, lorsqu'elle est appliquée au cas des flux, correspond à l'existence d'une dissymétrie entre la part en pourcentage de liens (L_{ij}) et la part de l'information totale de flux (F_{ij}) ou de mouvement (M_{ij}) qu'ils représentent. Cet outil permet de faire le lien entre les aspects analytiques (l'information) et graphiques (la densité) de la carte de flux : autrement dit entre la part d'information qui est représentée

sur la carte et la part de liens équivalente. Sur le plan cartographique, elle permet d'ajuster la *surcharge graphique* au regard de la perception des motifs.

La figure 4 présente les courbes de Lorenz de (F_{ij}^+) et de (M_{ij}^+) de l'Isère, exprimées en fonction du (%) cumulé de liens (L_{ij}), sa construction est bien détaillée dans Bretagnolle (1996).

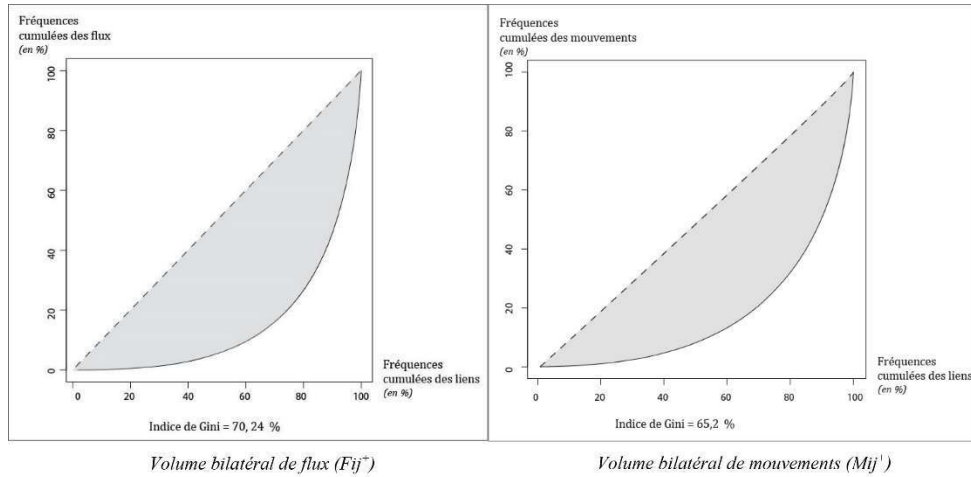


Figure 4 : Courbes de Lorenz des volumes de flux (F_{ij}^+) ou de mouvements estimés (M_{ij}^+) de l'Isère.

Cette figure 4 décrit l'inégalité éventuelle de la fréquence des valeurs de flux, par ailleurs mesurée par l'indice de concentration de Gini. Cet indice correspond, sur le plan graphique, à la mesure de la partie du graphique portée en gris sur cette figure 4. Il traduit le degré de concentration des valeurs sur les figurés représentés : 70,4% pour (F_{ij}^+) et 65,2% pour (M_{ij}^+) . L'indice varie entre 0 et 1, lorsqu'un seul individu détient toute l'information : dans ce cas, l'égalité de la fréquence des valeurs est parfaite. Plus la courbe de Lorenz s'éloigne de la diagonale et plus l'indice de Gini est élevé, et la distribution concentrée sur quelques liens qui contribuent fortement à l'information.

Mobilisé dans le cadre d'analyse des flux, l'indice de Gini permet donc de mesurer la concentration des valeurs fortes de la matrice sur quelques individus. Cela revient à appliquer, simultanément, les deux raisonnements¹⁰ distincts suivants.

Le premier porte sur l'intensité des liens, il mobilise leur valeur dans le cadre d'une *logique de flux* (ou de *mouvements*). Le second raisonnement s'appuie sur l'existence des individus (couples de lieux), soit sur une *logique de liens* intéressant le nombre de figurés décrivant ces valeurs. Ainsi, l'application d'un *critère unique de sélection* tient d'une double *logique de flux* et de *liens* (F_{ij} , L_{ij}) : plus l'indice de Gini d'une matrice est élevé et plus on peut espérer visualiser un maximum d'information avec un minimum de liens,

¹⁰ Il convient de noter que le processus de construction cartographique est invariable, quel que soit la *logique* (de flux ou de liens) appliquée : toutes mobilisent *a minima* la variable visuelle taille pour agir sur la *largeur* des liens qui sera proportionnelle à l'éventuelle valeur qu'ils symbolisent.

réduisant alors d'autant la complexité graphique de la figure, tout en révélant ses principaux motifs.

Sur le plan thématique, l'intérêt de ce double raisonnement permet de révéler une forme d'attractivité économique de quelques communes (ou d'un agrégat), en l'occurrence en termes d'emploi. Elle permet de poser des hypothèses quant au caractère plus ou moins égalitaire de l'offre d'emploi du département, sur quelques zones idéalement situées par rapport aux axes de transport, compte tenu de la topographie parfois compliquée des lieux. L'emploi est-il équitablement réparti, dispersé ou concentré sur quelques communes formant des pôles d'emplois ? Pour y répondre, examinons la manière dont ce double raisonnement est appliqué.

3. L'ESTIMATION D'UN DOUBLE CRITERE (Fij, Lij) DE SELECTION DES FLUX

Une manière simple d'estimer ce double critère consiste à adopter une approche visuelle, en observant son fonctionnement sur la courbe de Lorenz. L'examen de cette courbe serait d'ailleurs grandement facilité si l'on pouvait disposer d'une interface interactive qui aiderait l'utilisateur à la détermination de ce critère de sélection, de manière visuelle (directement sur le graphique) ou *a priori*, en indiquant la valeur du *critère unique de sélection*, mais c'est un autre projet. Dans l'immédiat, examinons le fonctionnement de ce double critère et ses conséquences sur le plan thématique.

3.1. Le fonctionnement du critère de sélection de la part d'information en fonction de sa densité, et inversement

La figure 5 présente les résultats obtenus à la suite de l'application d'un critère qui agit simultanément sur la part d'information de flux et la quantité de figurés présentés sur la matrice (M_{ij}^+) de l'Isère. Il s'agit d'une expérimentation cartographique réalisée pour montrer le rôle joué par ce double critère dans la mise en évidence des motifs spatiaux de ces mobilités professionnelles. Les différentes images révèlent la morphologie du réseau de pôles d'emplois isérois, à différents niveaux de densité. Elles laissent deviner un système de navettes quotidiennes qui s'expriment à plusieurs échelles territoriales, posant l'hypothèse d'une organisation hiérarchique sous-jacente.

En dessous de chacune des images est portée la valeur (α) du critère de sélection des flux : $(M_{ij}^+) \geq (\alpha)$, puis la part de liens (L_{ij}) représentée et la part de flux, en % du total que ces liens représentent. Seuls les liens qui présentent une valeur supérieure ou égale au critère (α) sont représentés. Les points correspondent aux chefs-lieux des communes de l'Isère. Ceux qui ne présentent pas de liens, donc qui ne sont pas concernés par des flux supérieurs au critère (α) , sont donc éliminés de l'analyse. Ainsi, la première image (en haut à gauche) correspond à la représentation de 20% de l'information totale, par conséquent aux valeurs du volume bilatéral de mouvements supérieures ou égales à 50 348 $(M_{ij}^+) \geq 50\,348$, par 0,4% des liens.

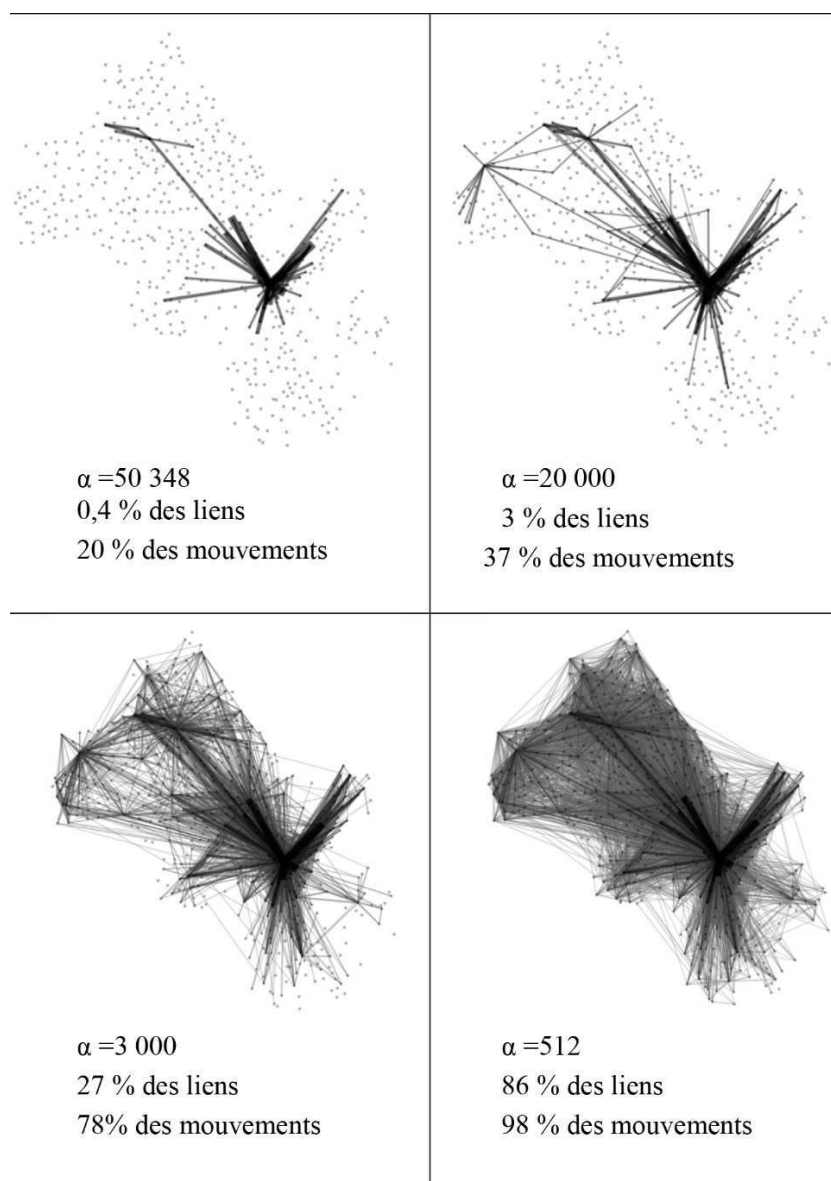


Figure 5 : Illustration du fonctionnement d'un double critère de sélection globale sur la matrice (M_{ij}^+) de l'Isère.

Source : fichier détail MOBPRO, INSEE (2010). GEOFLA, IGN (2010).

Sur le plan thématique, il est important de noter que ces images illustrent la masse des petits flux qui parcourent de longues distances et non des mobilités locales. Ce résultat est normal puisque la matrice (M_{ij}^+) est à interpréter ici selon une *logique de mouvement*. Il explique aussi le fait que la première image mette en évidence le premier pôle d'emploi, à savoir celui de l'agglomération de Grenoble (car elle polarise à elle seule

20% des mouvements) ; de même que son bassin de recrutement à la fois local et subrégional, à la faveur des axes de transport structurant le Y grenoblois. L'exploration interactive résumée sur la figure 5 met rapidement en évidence les flux inter pôles – les mobilités iséroises qui connectent des pôles d'emplois distants et donc différents, toujours à la faveur des axes de transport – et qui font fi de la topographie contrainte des lieux. Les volumes bilatéraux de mouvements les plus importants concernent des actifs qui vont travailler dans les communes situées au cœur de l'agglomération, sur le replat de la cuvette grenobloise, ou dans les deux autres pôles d'emplois du département.

Cette figure illustre l'intérêt du recours à une courbe de concentration, pour sélectionner visuellement un pourcentage de liens par rapport à un pourcentage de flux par classe correspondant à l'estimation du seuil. L'analyse étant visuelle, elle nous conduit à la conclusion selon laquelle le raisonnement qui préside à la détermination du critère (α) de sélection globale des flux à représenter résulte d'un arbitrage entre la part de liens et la part d'information de flux représentée. C'est pourquoi il n'est pas exactement unique, mais double.

Si l'estimation proposée est visuelle, c'était pour les besoins de la démonstration. En fait, ce double critère est en pratique toujours objectivé puisque nous pouvons démontrer qu'il correspond à un raisonnement fondé soit sur la sélection des flux (ou des mouvements), soit sur celle des liens. Ces deux logiques agissent de la même manière sur le même critère : elles sont mises en œuvre dans un objectif similaire de cartographie de flux *représentatifs* de l'information disponible, de ceux qui sont considérés comme *significatifs*.

La question qui se pose alors est celle de savoir comment estimer la valeur du critère (α) : en suivant une *logique de liens* ou bien de *flux* / de *mouvements* ? Les résultats obtenus sont-ils invariants quelle que soit la logique du raisonnement adoptée ? Quel est l'intervalle plausible de valeurs prises par (α) ?

3.2. L'estimation raisonnée du double critère de sélection des flux

L'enjeu de la détermination de ce *double critère*, pour les matrices de flux denses qui présentent un indice de Gini élevé, réside dans la représentation (graphique) optimale du maximum d'information de (F_{ij}^+) ou de (M_{ij}^+) par un minimum de liens (L_{ij}), afin de garantir clarté, efficacité visuelle et analytique. Cette question intéresse celle du seuillage d'une série statistique, un problème ancien et bien connu dans le champ du traitement de données qui se pose avec acuité pour les flux depuis plusieurs années, sans toutefois avoir été résolu (Becker, Eick, Miller, Wilks, 1990 : 93) cités par (Dorling, 1991 : 112), à notre connaissance.

Nous pouvons avancer deux raisons pour expliquer l'absence de critère de sélection unique des flux. La première tient probablement à la mise en œuvre d'approches statistiques qui sont inadaptées. Les statistiques paramétriques classiques sont mobilisées pour déterminer la valeur du critère de sélection, alors que leur normalité n'est pas avérée (voir figure 3). La seconde raison tient au caractère relatif du critère de sélection des flux qui ne correspondrait pas à une valeur exacte, il s'inscrirait dans un intervalle de valeurs non nécessairement borné. La recherche d'une valeur optimale et unique de seuillage des flux serait-elle vaine ? A. Bretagnolle nous a récemment suggéré une possibilité intéressante d'investigation, qui consisterait en

l'examen de différents points remarquables de la courbe de Lorenz, telle que mise en œuvre par T. Louail et al. (2014). Les auteurs ont effectivement mobilisé la courbe de Lorenz pour analyser une matrice domicile-travail, dans l'objectif de déterminer les points chauds d'origine et de destination des flux. Leur analyse est donc menée dans le cadre d'un raisonnement porté sur les *lieux* (et non sur les liens ou sur les flux).

En fait, nous pensons initialement que la question de l'estimation du critère de sélection est plutôt soit celle de l'estimation d'un intervalle¹¹ de valeurs de flux les plus fortes, c'est-à-dire concentrées sur l'histogramme, soit inversement, celle de la dispersion des liens ; tout dépend de la logique du raisonnement adopté.

Ces deux visions d'une même information, selon qu'elle est appréhendée du point de son intensité ou de sa densité traduisent deux raisonnements. Le premier est fondé sur la réalisation de cartes représentant de l'ordre de *80% des flux* (c'est-à-dire 80% de l'information totale disponible dans la matrice) et la seconde, celle des *2% des flux* les plus importants (autrement dit : 2% des navettes dont la représentation renseigne sur la part d'information de flux la plus importante). Elles correspondent respectivement à des objectifs de significativité et de densité de l'information.

3.2.1 L'objectif de significativité de l'information sélectionnée ($\alpha \approx 80\%$)

La détermination du critère de sélection des flux fondée sur un objectif de *significativité* des valeurs vise à représenter la part maximale d'information de manière lisible, elle s'inscrit dans le raisonnement géographique traditionnel. L'exemple précédent de 80% de l'information totale signifie que les 20% restants sont considérés comme résiduels. Raisonner ainsi revient à mobiliser la règle de répartition des 80/20 spécifique aux distributions statistiques de forme Pareto.

Le transfert de ce principe à la cartographie d'une matrice de flux conduit en fait à l'hypothèse plausible selon laquelle une faible part de liens contribue à décrire une part importante de l'information de flux. Sur le plan thématique, qu'une faible part de communes présente des emplois suffisamment variés, en quantité et sont raisonnablement accessibles pour pouvoir attirer une population importante. C'est probablement la raison pour laquelle différents auteurs (Tobler, 1982 ; Dorling, 1991 ; Bopda et al, 2000) se réfèrent à la règle des 80/20 pour sélectionner la part de flux qu'ils représentent (Voir Encart 1).

¹¹ L'exploration de l'effet cartographique de l'application d'un double critère de sélection nous a conduits à la conclusion selon laquelle la carte de flux ne variait pas dans un certain intervalle de valeurs de flux ou de liens. Ce résultat nous a conduit à suggérer l'adoption d'une logique floue dont une première analyse est présentée dans (Bahoken, 2014 : 237).

W. Tobler (1982) considère précisément la règle des 85-25% des valeurs de flux à représenter, arguant du principe que 85% du total des flux sont bien décrits par 25% des figurés. D. Dorling (1991) propose de ne représenter que de l'ordre de 80% de l'information qu'il considère *significative*, afin que « [...] plus des trois-quarts des flux soient représentés sur la carte, les plus typiques étant ceux des banlieusards. Cette technique est particulièrement utile lorsque tous les flux inter districts de la Grande-Bretagne sont représentés, car il y a trop peu d'espace [sur la carte] pour utiliser efficacement la technique de l'ordonnement visuel des figurés, les largeurs des lignes ne pourront pas être discriminées autrement » (Dorling, 1991 : 111). A. Bopda et al. (2000), après avoir présélectionné les flux dits « majeurs » représentent les « [...] échanges migratoires [...] qui totalisent dans chaque cas plus de la moitié des échanges migratoires observés » (Bopda, Grasland, Poulain, 2000 : 113), ce qui correspond environ à 60% de l'information disponible représentée par 5% des flux. N. Cattan suit une démarche similaire, en ne représentant que le top-100 des plus gros flux aériens internationaux, ce qui revient à ne représenter que 15% de l'information totale, par 1% des liens, soit les « [...] principaux flux de passagers qui s'effectuent entre les trois ensembles territoriaux que sont les Etats-Unis, l'Europe occidentale et les trois dragons du sud-est asiatique, [...] » (Cattan, 2004).

Encart 1 : Exemples de mise en œuvre de la règle des 80/20.

La figure 6 applique ce principe des 80/20 aux matrices (F_{ij}^+) et (M_{ij}^+) de l'Isère. Sur la partie haute de la figure, les flux sont sélectionnés en fonction d'un objectif de représentativité de 80% de l'information totale, tandis que sur la partie basse, l'objectif est fixé à 20% des liens.

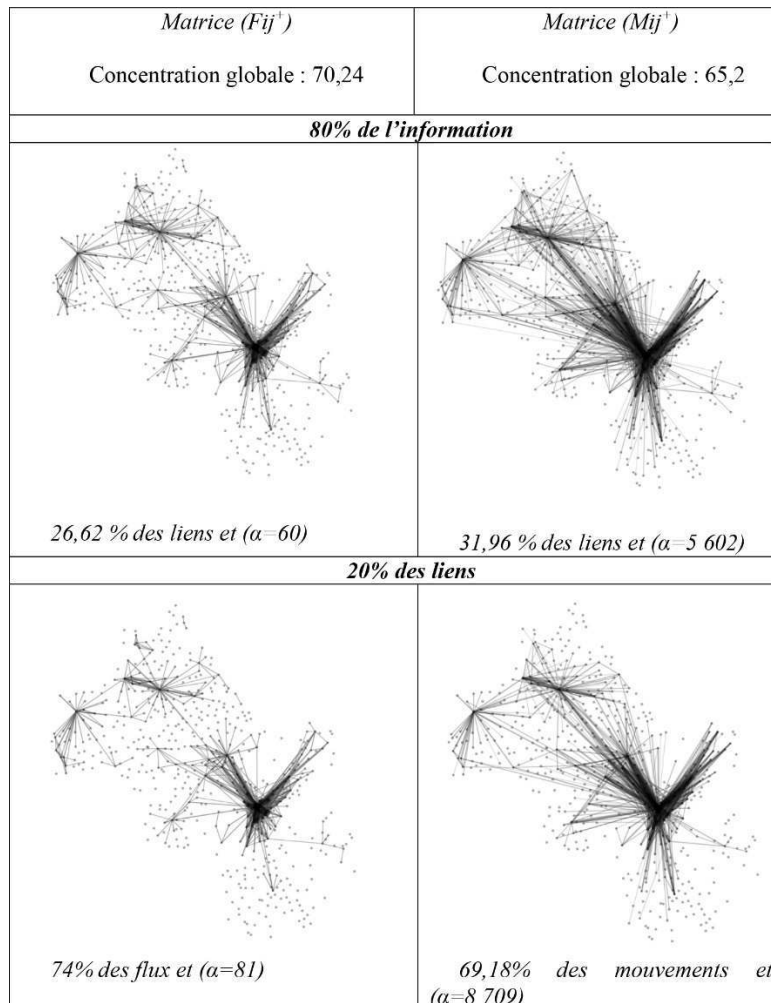


Figure 6 : Comparaison de la cartographie des 80% de flux significatifs des matrices (Fij⁺) et (Mij⁺) de l'Isère et de 20% des liens.

Source : fichier détail MOBPRO, INSEE (2010). GEOFLA, IGN (2010).

Sur le plan thématique, la représentation de 80% de l'information est fructueuse. On observe que les cartes de flux (Fij⁺) mettent en évidence des navettes locales là où celles du mouvement estimé (Mij⁺) révèlent des navettes à plus longue distance, de même que le rôle structurant des axes de transport. Il est par ailleurs intéressant de noter que les premières décrivent des navettes qui sont polarisées par l'un des trois principaux pôles d'emplois du département à savoir, au sens de l'INSEE (2010) : les zones de Grenoble, de Vienne-Roussillon et de Bourgoin-Jallieu. Elle décrit les motifs inverses de celles des mouvements qui s'expriment en dehors des bassins locaux de recrutement, à plus longue distance.

Ce résultat témoigne de l'efficacité du double critère proposé (la sélection de 80% de l'information significative). Si l'on considère le cas de celle de Grenoble, on observe qu'elle attire des navetteurs locaux et que son aire d'influence s'étend jusqu'aux confins du département, en englobant des sous-bassins d'emplois locaux, à la faveur du réseau routier principal : de Pontcharra à l'est à Saint-Marcellin à l'ouest, à La Mure au sud. Plus au Nord, les motifs de ces navettes grenobloises s'étendent jusqu'à l'Isle-d'Abeau et la Verpillère, située dans le bassin d'emploi de Bourgoin-Jallieu.

Sur le plan méthodologique, les motifs de la figure 6 varient peu en fonction des logiques de raisonnement (flux, mouvement) : l'indice de Gini est de 70,24 pour les flux et de 65,2 pour les mouvements, ce qui traduit la robustesse de la méthode. Cela valide aussi l'hypothèse selon laquelle il n'existe pas de différence notable, au niveau du rendu cartographique, selon que la règle des 80/20 est mise en œuvre du point de vue des flux (ou des mouvements) ou des liens, les cartes ne variant pas sensiblement. Ce principe fonctionne, en effet, quel que soit le degré de concentration de la matrice. Ce résultat valide aussi le fait que le critère de sélection s'inscrit dans un intervalle de valeurs à l'intérieur desquelles l'image varie peu¹².

Le second type de raisonnement qui peut être appliqué pour déterminer le critère de sélection s'appuie sur un *objectif de densité* de la matrice. Il aboutit également à la représentation des quelques flux que l'on considère comme les plus *significatifs*. De ce fait, le raisonnement examiné dans la sous-section suivante est l'inverse du précédent, tout en étant lié.

3.2.2 L'objectif de densité de la matrice (2-5%)

L'idée qui consiste à déterminer la valeur du critère de sélection des flux en fonction d'un *objectif de densité* de la matrice est fondée sur un raisonnement purement graphique qui découle de la théorie des graphes. Elle s'appuie sur le nombre de cases de la matrice renseignées, c'est-à-dire sur le nombre de liens disponible pour l'analyse par rapport au nombre de liens théorique. Sur le plan théorique, cette notion de densité est empruntée à l'analyse des réseaux sociaux, issue de la théorie des graphes.

La densité est l'un des indicateurs classiques permettant de caractériser un réseau de relations. L'indice (γ) gamma, souvent utilisé pour les réseaux spatialisés, exprime, par exemple, le rapport entre le nombre de liens – qui seront représentés – sur le nombre de liens possibles. On peut donc considérer qu'il correspond au degré de remplissage de la matrice. Cet indice varie entre 0 et 1 lorsque le réseau d'acteurs est complet, c'est-à-dire en présence de l'ensemble des liaisons théoriquement disponibles, soit $N*(N-1)$. Plusieurs formulations de la densité d'un réseau spatial, éventuellement valué, existent dans la littérature et toutes peuvent être interprétées comme une forme de sélection des valeurs qui s'applique, par extension, aux matrices de flux origine-destination puisqu'il s'agit formellement du même tableau.

Pour les flux, correspondant à des réseaux spatiaux valués, nous avons observé que le critère (γ) généralement admis est de l'ordre de 5% de la densité de la matrice, lorsqu'elle est complète. Cette affirmation correspond aussi au résultat des

¹² En effet, pour (F_{ij}^+) , par exemple, le résultat est similaire selon que $(\alpha=60)$ ou $(\alpha=81)$ qui correspondent respectivement à la sélection des 80% de flux ou des 20% de liens.

expérimentations menées par différents auteurs (Tobler, 1982 ; Bopda et al., 2000 ; Cattan, 2004 ; Dorling, 1991) conduisant à une estimation de la densité de valeurs de flux à représenter comprise entre 3 et 5 % du total des liens (voir Encart 2).

J. Bertin (1973) indique avoir appliqué les deux restrictions suivantes pour réaliser les cartes des migrations interdépartementales de 1954 : seules les valeurs de la « *tendance migratoire* » supérieures à 2% et plus de la population d'origine sont représentées ; pour les autres, elles seront représentées dès lors qu'elles concernent plus de 10 000 migrants. W. Tobler estime à 5%, la part de liens à représenter sur une carte de flux, illustrant l'efficacité de ce seuil sur une matrice de migrations internes réalisées entre 1965 et 1970, entre les 3 141 comtés américains. La sélection de 5% des flux suggérée revient alors à ne représenter *in fine* que 27 003 figurés au lieu des 9 862 740 disponibles. Cette valeur de densité de 5% correspond également à la suggestion de C. Grasland (2010), obtenue sur deux matrices de flux commerciaux : l'une à l'échelle mondiale (181*181 pays) en 2007 et l'autre à l'échelle des 29 pays de l'OCDE, en 2005. Le raisonnement mené par D. Dorling (1991) conduit à un seuil de 4%, obtenu par l'analyse d'une matrice de navettes interdistricts britanniques réalisées en 1981. A noter, que ce dernier seuil a été estimé sur une présélection des flux qui sont supérieurs à 1 000 individus ($F_{ij} > 1\ 000$), exprimés en termes de probabilités de fréquences conditionnelles qui correspondent à la pondération des valeurs de flux par les populations des lieux d'origine et de destination.

Encart 2 : Exemples de détermination du critère de densité de la matrice.

Les résultats d'une sélection des flux de l'Isère selon un objectif de densité des liens sont présentés sur la figure 7. Sur le plan méthodologique, le raisonnement mené sur la part de liens conduit à une élimination plus importante des valeurs, que lorsqu'il est mené sur la part de flux. On notera que les volumes de flux (F_{ij}^+) ainsi décrits sont également polarisés par les trois principales zones d'emploi de l'Isère déjà mentionnées. Il est possible de les comparer aux seuils de 3, de 4 et de 5% du total des liens disponibles. Ainsi, au seuil de 3%, l'image décrit 40,7% des navettes qui correspondent aux flux de plus de 399 actifs de quinze ans et plus ($\alpha = 399$). On observe que ces motifs diffèrent peu lorsque la part de liens est portée à 4 ou à 5%. Ce résultat n'est pas anodin. Il suggère, d'une part, comme le précédent, le fait que le critère de sélection s'inscrit dans un continuum de valeurs, ce qui a pour conséquence qu'il n'est pas nécessaire qu'il soit défini de manière exacte. D'autre part, le fait que l'analyse thématique des résultats s'appuyant sur un objectif de densité de la matrice de 3, de 4 ou de 5% (donc sur une notion issue de la Théorie de graphes) ne change pas fondamentalement l'interprétation thématique de la carte, à l'inverse de la logique de raisonnement sous-jacente. Dit autrement, une analyse menée sur les valeurs des liens, qui intéresse également leur

sémantique (flux et mouvement), produit des cartes plus riches sur le plan qualitatif (figure 6).

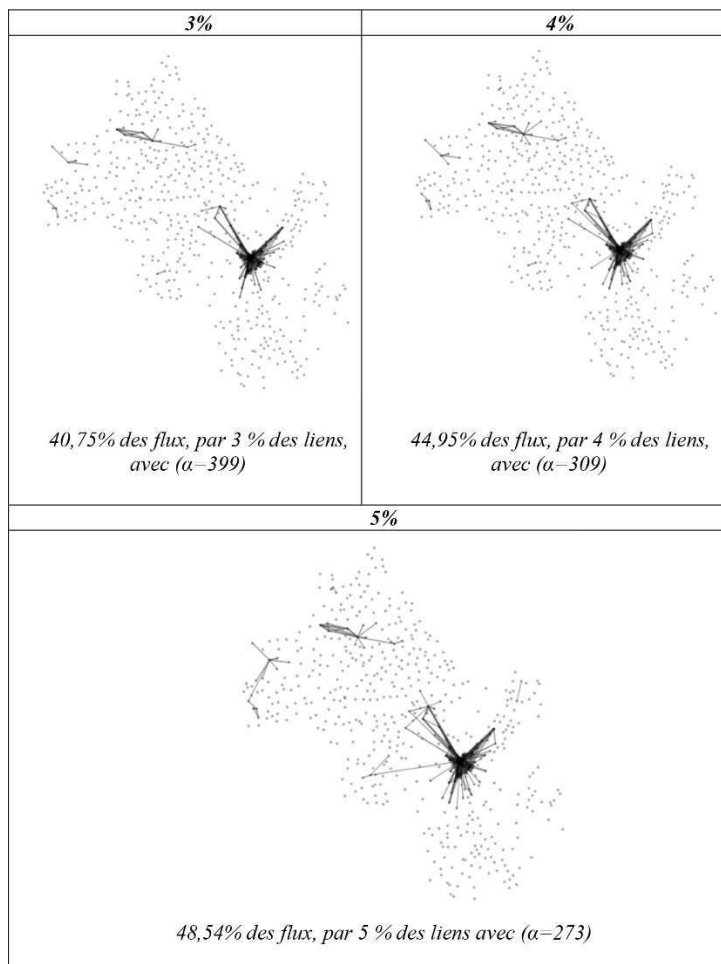


Figure 7 : Comparaison de la cartographie des volumes de flux bilatéraux de l'Isère selon différents objectifs de densité de la matrice (F_{ij+}).

CONCLUSION

L'application d'un *critère global unique* de sélection des l'information de flux à cartographier s'inscrit dans le cadre d'un raisonnement mixte, qui consiste à estimer la part *d'information significative* de flux à représenter par rapport à la part de liens qu'elle représente. La particularité de cette approche est de ne mobiliser qu'un et un seul critère, qui agit sur l'ensemble de la matrice. Elle conduit à la réalisation d'une analyse où ce qui prime, c'est la thématique. Dans l'exemple retenu, la figure 6 met en évidence les pôles d'emplois qui sont significatifs à l'échelle du département de l'Isère, elle les différencie selon qu'ils concernent des mobilités réalisées à courte distance (*logique de flux*) ou à plus

longue distance (*logique de mouvement*). De ce point de vue, la figure 6 nous apparaît la plus importante, car elle permet de valider, d'une part, l'intérêt du double critère proposé sur des matrices différentes (au niveau sémantique et géographique), en comparant la cartographie de 80% de l'information qu'elles contiennent. Elle montre l'intérêt d'une analyse menée sur l'intensité des interactions spatiales, dans un objectif de *significativité*. Elle s'oppose ainsi à la figure 7, dont le raisonnement mené selon un objectif de *densité* plutôt que de *significativité*. Ici, les images ne semblent mettre en évidence que les flux qui se produisent à l'intérieur du périmètre de zones d'emplois définies par l'INSEE, sans pour autant les connecter. La validité de cette hypothèse mériterait d'être confirmée, en utilisant des méthodes de sélection qui considèrent la notion d'appartenance territoriale à une même entité.

Sur le plan méthodologique, le recours à l'indice de Gini a pu être considérée comme une mesure désuète voire insuffisante par certains auteurs (Bretagnolle, 1996) dans la mesure où il restreint à l'analyse la représentation des zones de concentration (ou de dispersion) des emplois. Il ne permet pas de définir à lui seul la valeur d'un critère significatif de sélection.

Sur le plan de la construction cartographique, la figure 6 compare la représentation des volumes de navettes locales (Fij⁺) et subrégionales (Mij⁺). Compte tenu de la topographie de l'Isère, il pourrait être intéressant de mobiliser en sus leur orientation en degrés. L'une des hypothèses qui pourrait être intéressante est la dépendance de ces navettes au réseau routier principal, en raison de l'échelle (locale) mais aussi du contexte géographique fortement contraint. Dit autrement, à l'échelle locale, les motifs spatiaux de la demande de transport pour des motifs professionnels correspondent peu ou prou à ceux de l'offre locale.

Enfin, pour ce qui est de l'impératif de sélection des flux à cartographier, nous avons vu que l'une des particularités de ce *critère unique de sélection* tient à l'élimination de certains lieux de l'analyse. Il s'agit des couples de localités dont l'intensité du flux ne répond pas au critère de sélection. Cependant, le fait de considérer certains lieux comme inintéressants peut être appréhendé comme gênant pour l'analyse. Si tel est le cas, il est possible de sélectionner les flux en conservant l'ensemble des lieux de la zone d'étude, en agissant au niveau *local*. Pour ce faire, il convient de mobiliser un critère qui agit sur les *localités*, mobilisant obligatoirement les (i) lignes (et/ou les (j) colonnes) de la matrice (en complément éventuel des valeurs de flux). Cela revient à adopter une *logique de lieux* à la *logique de flux* (ou de *mouvements*) classique.

BIBLIOGRAPHIE

- BAHOKEN F. (2014), « L'intérêt du raisonnement logique en cartographie de flux. L'exemple de migrations internes », *European Journal of GIS and Spatial Analysis | Revue Internationale de Géomatique (RIG)*, Lavoisier, vol. 24, n° 2, pp. 231-250.
- BAHOKEN F. (2015), « Eléments pour une représentation (carto)graphique des matrices de flux », *M@ppemonde*, n° 115, 2014/3, 16 p. URL : <http://mappemonde.mgm.fr/num43/articles/art14304.html>
- BAHOKEN F. (2016a), Propositions de solutions liées au choix d'un critère de sélection global : la cartographie de *flux significatifs*. Chapitre 8, « *Contribution à la cartographie d'une matrice de flux* ». Thèse de doctorat en Géographie – Sciences des Territoires, Université Denis Diderot (Paris 7), pp. 325-345. HAL Id : [tel-01273776](https://tel.archives-ouvertes.fr/tel-01273776)
- BAHOKEN F. (2016b), « L'approche cartographique de la décomposition des matrices de flux », *M@ppemonde*, n° 116, URL : <http://mappemonde.mgm.fr/num44/articles/art14404.html>
- BEGUIN M., PUMAIN D. (1994, 2003), *La représentation des données géographiques. Statistique et cartographie*, Armand Colin, Coll. Cours Géographie, 2^{de} édition, 192 p.
- BERTIN J. (1973), *Sémiologie graphique*, Paris, Mouton-Gauthier-Villars, 2^{ème} édition, Les réimpressions de l'EHESS (1998), 431 p.
- BOPDA A., GRASLAND C., POULAIN M. (2000), « Evaluation comparative de l'influence des limites linguistiques sur les comportements migratoires : applications aux cas de la Belgique, du Cameroun et de la Tchécoslovaquie », in : « Régimes démographiques et territoires : les frontières en question », *Actes du colloque de l'AIDELF Les frontières en question*, 1988, La Rochelle, AIDELF, 20 p.
- BRETAGNOLLE A. (1996), « Etude des indices de concentration d'une population », *L'espace Géographique*, n° 2, pp. 145-157.
- CATTAN N., GRASLAND C. (1994), « Migrations et effets de barrière en Tchécoslovaquie (1960-1989) », in : Rey V. (éd.), *La Tchécoslovaquie en 1992 : transition, fragmentation, recomposition*, Presses de l'ENS Fontenay-Saint-Cloud, coll. Sociétés, Espaces, Temps, pp. 97-120.
- CATTAN N. (2004), « Le monde au prisme des réseaux aériens », *Flux*, vol. 4, N° 58, pp. 32-43. [En ligne] <http://www.cairn.info/revue-flux-2004-4-page-32.htm> [lien vérifié le 02/08/2015]
- DORLING D. F. L. (1991), "Chapter 6: "Cobweb of Flows"", in: *The Visualization of Spatial Social Structure*, Phd. Thesis, University of Newcastle upon Tyne, 577 p.
- GRASLAND C. (2010), *Transformation of flow matrix to symmetric Boolean graph: A proposal form comparative analysis*, EuroBroadMap. Vision of Europe in the World. Small or medium scale focused project FP7-SSH-2007-1, Flows and Network package, Unpublished methodological paper n° 3, 7 p.
- GRATALOUP C. (2011), *Représenter le monde*, La Documentation Française, La Documentation photographique – Les dossiers, n° 8084, 64 p.

- HOLLAND S. C., PLANE D. A. (2001), "Methods of mapping migration flow patterns", *Southeastern Geographer*, vol. 41, n° 1, pp. 89-104.
- HOLTEN D. (2006), "Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data", *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, vol. 12, n° 5, pp. 741-748.
- LOUAIL T., LENORMAND M., CANTUROS O. G., PICORNELL M., HERRANZ R., FRIAS-MARTINEZ E., RAMASCO J. J., BARTHELEMY M. (2014), "From mobile phone data to the spatial structure of cities", *Nature*, Article number: 5276. URL : <http://www.nature.com/articles/srep05276> [Lien vérifié le 2 juin 2016 2015].
- NYUSTEN J. D., DACEY M. F. (1961), "A graph theory interpretation of nodal regions", *Papers and proceeding of the Regional Science Association*, vol. 7, n° 1, pp. 29-42. [En ligne] http://deepblue.lib.umich.edu/bitstream/handle/2027.42/45977/10110_2005_Article_BF01969070.pdf?sequence=1[Lien vérifié le 27 octobre 2015].
- POULAIN M. (1981), *Contribution à l'analyse d'une matrice de migration interne*, Département de Démographie UCL, Cabay, Louvain-la-Neuve, 225 p.
- RAVENSTEIN E. G. (1885), "The Laws of Migration", *Journal of the Statistical Society of London*, n° 52, pp. 167-235.
- TOBLER W. R. (1969), "Geographical filters and their inverse", *Geographical Analysis*, vol. 3, n° 1, pp. 234-253
- TOBLER W. R. (1979), "A geographical flow mapping program", *Geographical Analysis*, vol. 1, n° 13, pp. 1-20.
- TOBLER W. R. (1982), Cartographic study of movement tables, Presentation in the *National Computer Graphic Assembly* - session on Statistical Graphics - Mapping, Anaheim, 17-06-1982.

