



La simplification des réseaux en géographieL'évaluation des dépendances relationnelles induites par les chemins au sein du réseau et l'agrégation des sommets autour de plate-formes relationnelles

Jean-Francois Gleyze

► To cite this version:

Jean-Francois Gleyze. La simplification des réseaux en géographieL'évaluation des dépendances relationnelles induites par les chemins au sein du réseau et l'agrégation des sommets autour de plate-formes relationnelles. 2008. halshs-00224121

HAL Id: halshs-00224121

<https://shs.hal.science/halshs-00224121>

Preprint submitted on 30 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La simplification des réseaux en géographie

L'évaluation des dépendances relationnelles induites par les chemins au sein du réseau et l'agrégation des sommets autour de plate-formes relationnelles

Simplifying geographical networks

Assessment of relational dependences induced by the paths inside the network and nodes aggregation around transfer areas

J.-F. Gleyze

Institut Géographique National – Laboratoire COGIT

jean-francois.gleyze@ign.fr

Résumé

L'étude des réseaux en géographie est souvent entravée par la taille et la complexité de ces objets. La théorie des graphes et les développements réalisés dans le domaine de la sociologie offrent des outils formels pour constituer des sous-groupes cohésifs au sein des graphes et procéder à leur simplification. Pour autant, ces outils n'exploitent que la composante topologique des graphes, et ne prennent pas compte des spécificités spatiales des réseaux géographiques – et en particulier des réseaux d'infrastructure. En effet, la structure de tels réseaux n'est pas tant marquée par leur topologie que par l'organisation des chemins induite non seulement par la topologie, mais également par la métrique et la géométrie des composantes. Partant de ce constat, nous proposons d'évaluer la cohésion que les chemins créent entre les composantes du réseau grâce à l'indice de dépendance relationnelle. Dans un second temps, l'étude comparée des valeurs de dépendance relationnelle entre les différents sommets du réseau permet d'effectuer des rattachements entre ces sommets et de constituer des sous-groupes, correspondant aux plate-formes autour desquelles s'articulent les chemins. Cette méthode formelle de regroupement aboutit à un graphe simplifié, dont la structure est révélatrice de l'organisation des chemins, et sur laquelle peuvent être calculés et représentés tous les indicateurs construits à partir des informations sur ces chemins. Cette méthode peut enfin appliquée au réseau simplifié issu des premiers regroupements : par itérations successives, il est ainsi possible de mettre en évidence les directions privilégiées des chemins et les principaux sommets charnières articulant ces directions, à des niveaux d'étude emboîtés.

Abstract

Studying geographic network is hampered by the objects size and complexity. Graphs theory and progress done in sociology provide formal tools to constitute cohesive sub-groups within the graphs and as a consequence they can be simplified. Yet, these tools take only advantage of topological component of graphs. They do not take into account the spatial specificity of geographical networks, and in particular infrastructures networks. Indeed, the structure of such networks is affected by path organisation, induced not only by their topology but also by the components metric and geometry. Leading by this assessment, we propose to evaluate the cohesion that paths create between the network components, based on the relational dependence indicator. In a second step, the comparative study of relational dependence values between the various network nodes enables us to gather nodes and create sub-groups, which correspond to transfer areas around which pass some paths. This formal grouping method triggers off a simplified graph enhancing paths organisation. All the indicators can be then calculated and represented from the paths information. Finally this method can be applied on a simplified network coming from first grouping actions: by successive iterations, the favoured directions of paths and main nodes leading to these directions can be implemented at embedded study levels.

Mots-clés

réseau, graphe, sous-groupe, cohésion, agrégation, chemins, dépendance relationnelle, plate-forme relationnelle

Keywords

network, graph, subgroup, cohesiveness, agregation, paths, relational dependence, transfer area

Introduction

Qu'il s'agisse de réseaux de transport, d'énergie, de communication, de villes ou d'individus, l'étude des structures réticulaires en géographie se heurte souvent à des problèmes liés à la taille et à la complexité de ces structures. Ces problèmes se posent particulièrement dès lors qu'il s'agit d'évaluer les réseaux à des échelles inférieures à l'échelle de représentation. Dans ce cas, l'amas de composantes rend difficile la lecture et la synthèse des informations liées aux réseaux.

De la même manière que la cartographie possède des outils permettant de généraliser les cartes et d'en faciliter la lecture par des opérations de simplification lorsque l'échelle de représentation diminue, il existe des outils permettant d'agréger les composantes des réseaux et de simplifier leur évaluation et leur représentation. Ces méthodes sont issues de la théorie des graphes et leurs applications ont été particulièrement développées dans le domaine de la sociologie des réseaux. La première partie de cet article établit un panorama de ces méthodes, organisé selon les familles de critères régissant les regroupements. Ainsi que nous le verrons, ces critères sont exclusivement fondés sur des considérations d'ordre topologique, et n'intègrent en aucun cas les spécificités des réseaux géographiques – en particulier leur implantation spatiale et les propriétés géométriques qui en découlent.

Dans ce contexte, nous proposons d'aborder la cohésion au sein d'un réseau d'infrastructure en nous intéressant à la manière dont les chemins construits sur ce réseau créent des dépendances entre les composantes : pour évaluer cette forme de cohésion, nous développons l'indicateur de dépendance relationnelle en nous appuyant sur l'observation de tous les chemins construits sur le réseau. Dans un second temps, nous montrons que la connaissance de toutes les dépendances relationnelles entre paires de sommets permet d'opérer des rattachements entre ces sommets. Par construction, le rattachement d'un sommet à un autre est effectif si la majorité des chemins qui passent par le sommet 'amont' passent également par le sommet 'aval'. Cette méthode de rattachement est définie de manière formelle, et permet d'opérer des regroupements représentatifs de l'organisation des chemins au sein du réseau. Telle qu'elle est effectuée, chaque agrégation de sommets correspond à une plate-forme du réseau, collectant et distribuant des flux de chemins présentant des caractéristiques structurelles communes.

Les méthodes de regroupements au sein de graphes

Les opérations de regroupement au sein d'un graphe ont pour objet de rassembler des sommets selon des critères donnés (existence de liens directs / forts / intenses / positifs, etc.).

Ces opérations peuvent, dans un second temps, conduire à remplacer les regroupements par des noeuds agrégés, et aboutir à une représentation simplifiée du graphe. Cette simplification constitue néanmoins un deuxième travail de recherche, dans la mesure où elle implique de simplifier en parallèle l'ensemble des arcs du graphe ainsi que toutes les informations associées aux sommets et aux arcs (informations attributaires telles que les valuations ou les pondérations, informations structurelles telles que les chemins, etc.).

La théorie des graphes fournit plusieurs outils pour opérer des regroupements de sommets au sein d'un graphe. Nous allons étudier ces outils au travers du prisme de la sociologie :

- sur le plan théorique, cette discipline a largement contribué à leur formalisation et à leur développement ;
- sur le plan pratique, elle propose des cas d'études simples et concrets pour appréhender leur signification et leur portée.

En sociologie, les graphes sont souvent mobilisés pour modéliser les populations d'individus : les sommets et les arcs représentent alors les individus et les relations que ceux-ci entretiennent (relations d'un individu à un autre : par ex. « je suis parent de », relations de ressemblance : par ex. « j'appartiens au même groupe que », etc.). Les outils de regroupement issus de la théorie des graphes interviennent pour mettre en évidence des similarités entre les individus (on parle souvent de « cohésion ») selon différentes familles de critères, rattachées à des notions précises de théorie des graphes que nous développerons dans la suite.

L'état de l'art que nous présentons ici sur ces méthodes de regroupement s'appuie sur (Wasserman & Faust, 1994) et (Moody & White, 2003).

La notion de cohésion

Les méthodes de regroupements que nous allons présenter reposent sur la notion de « cohésion » entre les « membres », c'est-à-dire entre les individus d'un même sous-groupe.

Le *Petit Robert* définit la cohésion comme étant :

- la force qui unit les parties d'une substance matérielle ;
- le caractère d'un ensemble dont les parties sont unies, harmonisées.

Dans le domaine de la théorie des graphes appliquée à la sociologie, la notion de cohésion revêt une signification très proche, puisqu'elle caractérise les sous-

groupes d'individus fortement soudés, en regard des « connexions » effectives existant entre eux. Ces connexions matérialisent des relations d'ordre social¹, et se traduisent par des arêtes entre les sommets associés aux individus.

Il est important de bien distinguer la notion de « cohésion » de celle d'« adhésion », relative aux structures centralisées autour un sommet. De telles structures sont caractérisées par la convergence de nombreux liens vers un sommet « leader », et dépendent donc fortement de l'action ou du statut de ce sommet. Pour cette raison, les structures « adhésives » sont fragiles, car il suffit d'agir sur le sommet leader pour compromettre leur intégrité.

À l'inverse, les structures « cohésives » sont caractérisées par une multitude de liens indépendants, qui contribuent à en souder fortement les membres². La force de cohésion du groupe est alors mesurée en regard de la difficulté à « casser » la structure, à provoquer sa scission.

En définitive, la cohésion d'un groupe est une propriété résultant exclusivement de la configuration des relations à l'intérieur du groupe. Elle est d'autant plus forte que les liens à l'intérieur du groupe sont capables de maintenir sa structure.

Éléments qualitatifs pour identifier les structures cohésives

La structure d'un groupe est maintenue au mieux lorsque tous ses individus sont reliés deux à deux. Ce type de structure correspond à la notion de « clique » de la théorie des graphes, mais il reste peu observé dans la pratique, car l'absence d'une seule relation entre deux individus d'une clique met à mal cette structure. Pour autant, la contrainte d'« exhaustivité » des relations requise dans la construction des cliques peut être atténuée et ouvrir la voie à des méthodes de regroupement plus probantes d'un point de vue opérationnel.

À défaut d'être reliés deux à deux, les sommets d'un groupe peuvent en effet former une structure cohésive lorsque les relations offertes par les arêtes sont suffisamment variées : en ce sens, la cohésion est reliée avec la capacité des membres à être facilement atteints, c'est-à-dire leur « atteignabilité ».

¹ Ces liens relationnels sont effectifs, et doivent être distingués des liens « idéationnels », construits sur les sentiments d'appartenance des individus, mais ne reflétant pas de réalité tangible. Les liens relationnels débouchent sur la notion de « cohésion » (que nous développerons dans la suite), tandis que les liens idéationnels débouchent sur la notion distincte de « solidarité ».

² L'opposition entre structures cohésives et adhésives se retrouve dans l'étymologie des mots. Aux connotations d'égalité et de connexions multiples associées aux termes de cohésion, collègue ou compagnon, s'oppose la notion de centralité et de référence, présente dans les termes d'adhésion, d'adversaire ou d'administrateur.

Une autre piste consiste à établir un lien entre la cohésion d'un groupe et la « densité » (ou le volume) des arêtes observée à l'intérieur de ce groupe. Bien qu'intuitivement satisfaisante, cette caractérisation est insuffisante : un coup d'œil à la structure linéaire et à sa version légèrement modifiée sur la Figure 1 permet de s'apercevoir qu'en dépit de l'égalité des densités des deux structures, l'atteignabilité – et donc la cohésion – des points est modifiée.

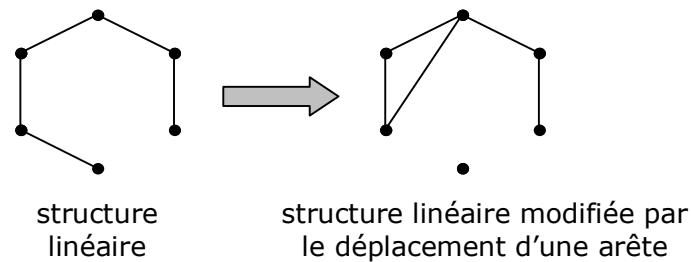


Figure 1 : la densité des arêtes ne suffit pas pour décrire la cohésion d'un sous-groupe - d'après : (Moody & White, 2003)

Si la densité des relations ne constitue pas un indicateur exclusif de la cohésion, il n'en demeure pas moins que l'uniformité de la distribution des liens est requise pour prévenir l'éclatement du groupe. Dans le cas contraire, les sous-groupes caractérisés par la rareté de certaines relations apparaissent plus vulnérables : leur modification provoquera plus facilement la scission du groupe.

Dans cette logique, une troisième famille de méthodes de construction exige que des critères sur l'adjacence des sommets soient vérifiés à l'intérieur des sous-groupes : de tels critères sont définis à partir des « degrés » des sommets.

En termes d'adjacences, il peut être pertinent de prendre en compte non seulement les arêtes à l'intérieur du groupe que l'on constitue (ces arêtes « intra-groupes » doivent avoir la plus forte densité possible), mais également les arêtes reliant les sommets de ce groupe aux autres sommets du graphe (ces arêtes « inter-groupes » doivent avoir la plus faible densité possible).

De manière plus formelle, il est enfin possible de construire des sous-groupes en essayant non pas d'opérer des regroupements de sommets, mais en identifiant les familles d'arêtes inter-groupes qui les séparent. Pour cela, on peut s'appuyer sur la fragilité structurelle que présentent ces familles, compte-tenu de leur faible densité. Ces familles correspondent alors aux petits ensembles d'arêtes, dont la suppression entraîne la déconnexion du graphe. Ce type d'approche recoupe tout un pan de la théorie des graphes, consacré aux « ensembles de coupure ».

Sur la base de ces premiers éléments qualitatifs, nous présenterons donc les méthodes de regroupement des sommets autour des cinq notions suivantes :

- l'exhaustivité des relations ;
- la proximité ou l'atteignabilité des sommets à l'intérieur du sous-groupe ;
- le degré des sommets du sous-groupe ;
- la comparaison des relations intra-groupes et inter-groupes ;
- l'identification des ensembles de coupure.

Méthodes d'identification de sous-groupes au sein d'un graphe

Rappels de théorie des graphes – Définitions et notations

Un sous-graphe correspond à la réunion d'un sous-ensemble de sommets du graphe et des arêtes existant entre eux.

Un sous-graphe est dit « maximal » relativement à une propriété si :

- il vérifie cette propriété,
- toute extension du sous-graphe par un ou plusieurs sommets extérieurs conduit à ne plus vérifier cette propriété.

Symétriquement, un sous-graphe est dit « minimal » relativement à une propriété si :

- il vérifie cette propriété,
- toute restriction du sous-graphe par suppression de l'une ou de plusieurs de ses composantes conduit à ne plus vérifier cette propriété.

Un chemin entre deux sommets est une séquence de sommets et d'arêtes distincts reliant ces deux sommets. La longueur d'un chemin est donnée par son nombre d'arêtes. L'existence d'au moins un chemin entre deux sommets garantit que ceux-ci sont atteignables l'un par rapport à l'autre. Le (un) plus court chemin entre deux sommets est appelé « géodésique » : sa longueur caractérise la distance entre les deux sommets. Par la suite, on notera $d(i, j)$ la longueur de plus court chemin entre les sommets i et j .

Le diamètre d'un graphe (resp. d'un sous-graphe) est égale à la plus grande des longueurs parmi toutes les géodésiques observées sur le graphe (resp. le sous-graphe).

Deux sommets sont dits « connectés » s'il existe au moins un chemin les reliant. Le graphe est dit « connecté » si toutes les paires de sommets sont connectées. Deux chemins de i à j sont indépendants si et seulement si ils ne partagent aucun sommet en commun, à l'exception bien sûr des sommets i et j .

Les sous-groupes construits à partir de l'exhaustivité des arêtes

Les sous-groupes pour lesquels les sommets sont reliés deux à deux correspondent formellement à la notion de « clique » de la théorie des graphes.

Une clique est en effet un sous-graphe complet maximal d'au moins trois sommets : en d'autres termes, toutes les paires de sommets au sein de la clique

sont connectées, et il n'existe aucun sommet extérieur à la clique qui soit connecté à tous les sommets de la clique.

Dans un graphe, un sommet peut appartenir à plusieurs cliques.

Par définition, la notion de clique est très restrictive pour définir un sous-groupe. L'absence d'une seule relation entre deux sommets d'une clique potentielle rend l'existence de celle-ci impossible. En pratique, dans un réseau étendu, il existe peu de cliques et elles sont généralement de petite taille.

Les sous-groupes définis à partir de l'atteignabilité et du diamètre

La notion d'atteignabilité est le point de départ pour étendre la notion de clique, dont nous avons vu qu'elle était très contraignante. L'atteignabilité est notamment pertinente dans l'étude des réseaux d'information. Il s'agit ainsi de définir les sous-groupes autour de la rapidité de communication entre les sommets qui les composent. Les sommets ne sont donc pas obligés d'être adjacents comme c'était le cas pour les cliques.

La notion de sous-groupe relative au concept d'atteignabilité exige des sommets constitutifs des sous-groupes d'être proches les uns des autres.

Sur cette base, on définit la n -clique comme tout sous-graphe maximal de diamètre inférieur ou égal à n . En d'autres termes, les longueurs des géodésiques entre paires de sommets de ce sous-graphe n'excèdent pas n , et on ne peut pas trouver de sommet extérieur au sous-graphe qui soit situé à une distance inférieure ou égale à n de tous les sommets du sous-graphe.

Telles qu'elles sont définies, les n -cliques peuvent avoir des diamètres supérieurs à n , car la définition n'exige pas des chemins inter-sommets d'être inclus dans la n -clique.

Les sous-groupes définis à partir du degré des sommets

Dans un sous-groupe défini à partir de critères sur les degrés, les membres doivent être adjacents à un nombre « relativement grand » d'autres membres du sous-groupe. En ce sens, on construit les sous-groupes selon le même principe que les cliques (tout le monde doit être relié à tout le monde), mais avec des contraintes moins fortes (tout le monde doit être relié à un nombre minimum d'individus).

Cette notion se révèle intéressante pour aborder les problèmes de vulnérabilité des sous-groupes : la vulnérabilité intervient lorsque les relations entre les individus ne sont pas assez redondantes, de sorte que la suppression de certaines arêtes au sein du sous-groupe compromet les relations entre les

individus. En l'occurrence, les n -cliques ne sont pas assez robustes, dans le sens où leur structure peut facilement être cassée suite à la rupture d'un seul de leur sommets (ainsi, une structure en étoile autour d'un sommet constitue une 2-clique et est vulnérable à la rupture du sommet central).

A contrario, les k -plexes et les k -cores sont des sous-groupes plus robustes, car ils sont construits de telle sorte que leurs sommets présentent un nombre d'adjacences relativement grand.

Un k -plex est un sous-groupe maximal contenant n sommets, tels que chacun de ces sommets est adjacent à au moins $n - k$ autres sommets du sous-groupe. En d'autres termes, chaque sommet est relié à tous les autres, à k sommets près.

En s'intéressant – non plus au nombre maximum de liaisons *absentes* – mais au nombre minimum de liaisons *présentes* depuis chaque sommet, on définit le k -core : un k -core est ainsi défini comme un sous-graphe maximal contenant g_s sommets, tels que chacun de ces sommets est adjacent à au moins k autres sommets du sous-groupe.

Les sous-groupes définis par comparaison des relations intra-groupes et inter-groupes

Les définitions introduites précédemment sont fondées sur les propriétés des relations à l'intérieur des sous-groupes. De fait, les groupes cohésifs dans les réseaux sociaux sont tels que les individus qui les composent sont plus proches les uns des autres que des individus extérieurs.

Ce constat rappelle les deux aspects généraux qui régissent le concept de sous-groupe, à savoir :

- la concentration de relations au sein du sous-groupe,
- la comparaison de l'intensité et de la fréquence des relations intérieures et extérieures au groupe (resp. relations « intra-groupes » et « inter-groupes »).

Les lambda-sets sont des sous-groupes construits à partir de ce second critère. Pour ces sous-groupes, la cohésion est liée à celle de connectivité : les sous-groupes identifiés doivent être difficiles à déconnecter à la suite de suppression d'arêtes. Formellement, les lambda-sets correspondent aux sous-ensembles maximaux de nœuds tels que le nombre minimum d'arêtes à couper pour déconnecter toute paire de nœuds de cet ensemble est strictement supérieur à celui nécessaire pour déconnecter toute paire formée d'un nœud de cet ensemble et d'un nœud extérieur.

Par construction, les lambda-sets ne se recouvrent pas, à moins d'entretenir des relations d'inclusion. Notons enfin que les lambda-sets ne dépendent pas de critères sur le degré ou sur l'atteignabilité : de fait, les sommets qui les constituent ne sont pas forcément adjacents ni proches.

Une autre manière de prendre en compte les densités comparées intra- et intergroupes consiste à partitionner le graphe en un nombre k de sous-groupes prédéfini, tels que la densité intra-groupes soit maximale et la densité intergroupes soit minimale : les ensembles issus de ce partitionnement sont appelés « factions » ou « k -groupes ».

Par construction, les factions considèrent explicitement les fortes valeurs de densité comme des garanties de forte cohésion.

Les sous-groupes définis à partir des ensembles de coupure

Une coupure (ou un ensemble de coupure) entre deux sommets i et j est un ensemble S de sommets dont la suppression déconnecte i et j .

Le cardinal $N(i, j)$ de la coupure minimale entre les sommets i et j correspond par ailleurs au nombre maximal de chemins indépendants que l'on peut tracer entre i et j (théorème « max-flow – min-cut »). Ce nombre est la valeur de séparation nodale entre les sommets i et j .

La plus petite valeur $N(i, j)$ de séparation nodale observée sur un graphe (parmi l'ensemble de toutes ses paires de sommets (i, j)) correspond à la connectivité nodale k du graphe. On parle alors de graphe k -connecté : concrètement, toute paire de sommets du graphe peuvent être reliés par k sommets indépendants (au moins). En particulier, un graphe 1-connecté est un composant, un graphe 2-connecté (resp. 3-connecté) est un bicomposant (resp. un tricomposant), etc., et plus généralement, un graphe k -connecté est un k -composant.

Par construction, la connectivité nodale est définie à partir de la plus petite valeur de séparation nodale observée sur le graphe, aussi est-il probable que certains sous-graphes présentent des valeurs de connectivité nodale plus élevées. Par conséquent, l'identification des ensembles de coupure du graphe permet d'isoler, de part et d'autre de ces coupures, des sous-graphes de meilleure connectivité.

Cette méthode dite de « cohesive blocking » (Moody & White, 2003) permet d'identifier hiérarchiquement des sous-groupes de plus en plus cohésifs (au sens de la k -connectivité) au sein des graphes.

Illustration

(Moody & White, 2003) illustrent ces différentes méthodes de regroupement sur le graphe suivant :

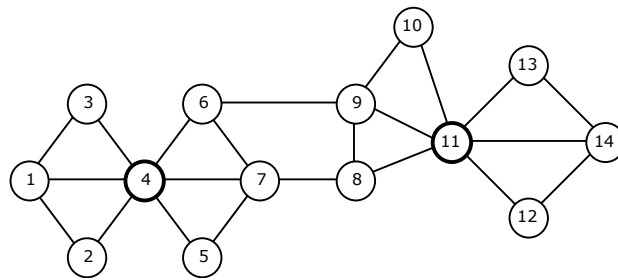


Figure 2 : exemple de graphe pour la comparaison de méthodes d'identification de sous-groupes cohésifs - d'après : (Moody & White, 2003)

Sur ce graphe, les sommets 4 et 11 ont été mis en valeur, car ils sont deux noeuds de coupure, et jouent donc un rôle particulier dans l'arrangement des relations au sein du graphe. Il faudra par conséquent être particulièrement attentifs à la manière dont ils seront traités lors de la constitution des sous-groupes.

Type de structure	Paramètre	Définition	Sous-groupes identifiés sur l'exemple de la Figure 2
clique	$\emptyset$	Sous-graphe complet	$\{1\ 2\ 4\}$ $\{1\ 3\ 4\}$ $\{4\ 5\ 7\}$ $\{4\ 6\ 7\}$ $\{8\ 9\ 11\}$ $\{9\ 10\ 11\}$ $\{11\ 12\ 14\}$ $\{11\ 13\ 14\}$
N-clique	2-cliques	Ensemble maximal de sommets au sein duquel le plus court chemin entre toute paire de sommets n'excède pas N .	$\{1\ \text{à}\ 7\}$ $\{4\ \text{à}\ 8\}$ $\{4\ 6\ \text{à}\ 9\}$ $\{6\ \text{à}\ 9\ 11\}$ $\{6\ 8\ \text{à}\ 11\}$ $\{8\ \text{à}\ 14\}$
	3-cliques		$\{1\ \text{à}\ 9\}$ $\{4\ \text{à}\ 9\ 11\}$ $\{4\ 6\ \text{à}\ 11\}$ $\{6\ \text{à}\ 14\}$
	4-cliques		$\{1\ \text{à}\ 11\}$ $\{4\ \text{à}\ 14\}$
k-plex	$k=1$ taillemini=3	Tout membre d'un k -plex d'ordre n est connecté directement à au moins $n-k$ autres membres. Nota : le degré du graphe étudié étant 2, l'ordre d'un k -plex doit être au moins égal à $k+2$.	Toutes les cliques
	$k=2$ taillemini=4		$\{1\ 2\ 3\ 4\}$ $\{4\ 5\ 6\ 7\}$ $\{6\ 7\ 8\ 9\}$ $\{8\ 9\ 10\ 11\}$ $\{11\ 12\ 13\ 14\}$
	$k=3$ taillemini=5		$\{1\ 2\ 4\ 5\ 7\}$ $\{1\ 2\ 4\ 6\ 7\}$ $\{1\ 3\ 4\ 5\ 7\}$ $\{1\ 3\ 4\ 6\ 7\}$ $\{4\ 6\ 7\ 8\ 9\}$ $\{6\ 7\ 8\ 9\ 11\}$ $\{8\ 9\ 11\ 12\ 14\}$ $\{8\ 9\ 11\ 13\ 14\}$ $\{9\ 10\ 11\ 12\ 14\}$ $\{9\ 10\ 11\ 13\ 14\}$
k-core		Sous-graphe connecté maximal tel que tout sommet est de degré supérieur ou égal à k .	Le graphe en entier est un 2-core.
Lambda-set	$\lambda=3$	Sous-ensemble maximal de sommets tel que le nombre minimum d'arêtes à couper pour déconnecter toute paire de sommets de cet ensemble est strictement supérieur à celui nécessaire pour déconnecter toute paire formée d'un sommet de cet ensemble et d'un sommet extérieur.	$\{1\ 4\ 6\ 7\}$ $\{8\ 9\ 11\ 14\}$
Faction	2-groupes	Partition du graphe en un nombre de groupes donné, obtenue par maximisation des densités intra-groupes et minimisation des densités inter-groupes.	$\{1\ \text{à}\ 7\}$ $\{8\ \text{à}\ 14\}$
	3-groupes		$\{1\ \text{à}\ 4\}$ $\{5\ \text{à}\ 9\}$ $\{10\ \text{à}\ 14\}$
	4-groupes		$\{1\ \text{à}\ 3\}$ $\{4\ \text{à}\ 7\}$ $\{8\ \text{à}\ 11\}$ $\{12\ \text{à}\ 14\}$
Bicomposant		Sous-graphes biconnectés	$\{1\ \text{à}\ 4\}$ $\{4\ \text{à}\ 11\}$ $\{11\ \text{à}\ 14\}$

Cet exemple montre que le choix du critère de regroupement conduit à des sous-groupes de tailles et de natures variées. En particulier, certains critères conduisent à centrer les sous-groupes autour des noeuds de coupure, tandis que d'autres critères placent ces mêmes noeuds en marge des sous-groupes.

Quelle que soit la méthode choisie, les critères intervenant dans la construction des sous-groupes restent exclusivement d'ordre topologique, car ils reposent sur le voisinage local des sommets (existence / absence d'arêtes, degré), sur les longueurs de plus court chemin définies en nombre d'arêtes ou encore sur la notion de coupe.

En ce sens, ces critères auront une portée limitée pour la constitution de sous-groupes sur les graphes valués, car les questions de voisinage ou de longueurs de chemins sont particulièrement sensibles à la valuation des arêtes.

Afin de prendre en compte les déformations induites par la valuation des arêtes sur la structure du graphe et le positionnement relatif des sommets au sein de cette structure, nous proposons de travailler sur les plus courts chemins observés sur le graphe valué.

En considérant les successions de sommets et d'arêtes visités le long des plus courts chemins, nous allons montrer dans un premier temps qu'il est possible de définir une forme de cohésion relationnelle entre les composantes du graphe. Selon cette notion, une composante i du graphe est d'autant plus dépendante relationnellement d'une autre composante j que la majorité des plus courts chemins qui transitent par i transitent également par j .

Sur la base des dépendances croisées entre toutes les composantes du graphe – et en particulier entre toutes les paires de sommets adjacents – nous verrons dans un second temps qu'il est possible d'effectuer des regroupements de sommets compte-tenu des principales dépendances relationnelles observées sur le graphe. En l'état actuel des recherches, ces regroupements peuvent conduire à une représentation simplifiée du graphe, caractéristique – par construction – de l'offre relationnelle du graphe en termes de plus courts chemins. Pour cette raison, nous verrons que :

- tous les indicateurs calculés à partir des informations (structurelles ou attributaires) sur les plus courts chemins peuvent être résumés sur cette représentation simplifiée du graphe ;
- le graphe simplifié peut, à son tour, faire l'objet d'une nouvelle simplification selon la même procédure et, par itérations successives, aboutir à une représentation emboîtée de sous-groupes.

Appréhender la cohésion des composantes au travers des dépendances relationnelles induites par les plus courts chemins

Dans un contexte géographique, les réseaux de transport matérialisent des infrastructures réelles. En ce sens, les réseaux ne sont pas invoqués pour étudier la topologie des relations entre paires de sommets, mais étudier la mise en relation des sommets distants, selon des chemins ancrés sur les infrastructures.

De ce fait, lorsque l'on pose la question de la cohésion des composantes du réseau, l'existence de liaisons directes entre les sommets n'est pas tant importante que la combinaison des composantes structurant les chemins. Selon ce schéma, deux composantes non voisines topologiquement peuvent participer conjointement à plusieurs chemins : cette implication commune crée une

dépendance entre les deux composantes (que nous qualifierons de « relationnelle », car elle est construite sur l'observation des chemins) qui peut être quantifiée, et déboucher sur la simplification des réseaux par constitution de sous-groupes en regard de cette propriété.

Définir et quantifier la dépendance relationnelle des composantes d'un réseau

Dans le contexte de réseaux d'infrastructures, les composantes du réseau sont solidarisées au sein des chemins reliant les différents couples de sommets origine-destination. La dépendance relationnelle entre deux composantes (sommets et arêtes) est d'autant plus forte qu'il existe de chemins articulés conjointement sur ces deux composantes.

Pour quantifier cette dépendance, nous proposons donc de caractériser les composantes du réseau par la liste des chemins auxquels elles participent³. Pour cela, nous considérons la matrice d'implication du graphe représentatif du réseau, listant en colonnes les relations origine-destination et en lignes les composantes du réseau. Cette matrice décrit les composantes impliquées dans le(s) plus court(s) chemin(s) de chacune des relations origine-destination indiquée en colonne : pour une relation k donnée, le coefficient (i, k) indique la proportion des chemins transitant par la composante i au titre de cette relation⁴ (cf Figure 3).

³ La construction de la mesure de dépendance relationnelle que nous présentons ici est extraite de notre travail de thèse sur la vulnérabilité des réseaux – cf (Gleyze, 2005).

⁴ Nous travaillons ici avec les plus courts chemins mais, par construction, cette modélisation n'exclut pas de travailler avec des chemins (ou des flux de chemins) construits sur d'autres critères : de la même manière, il suffit de renseigner la proportion des chemins transitant par les différentes composantes du réseau pour chacune des relations.

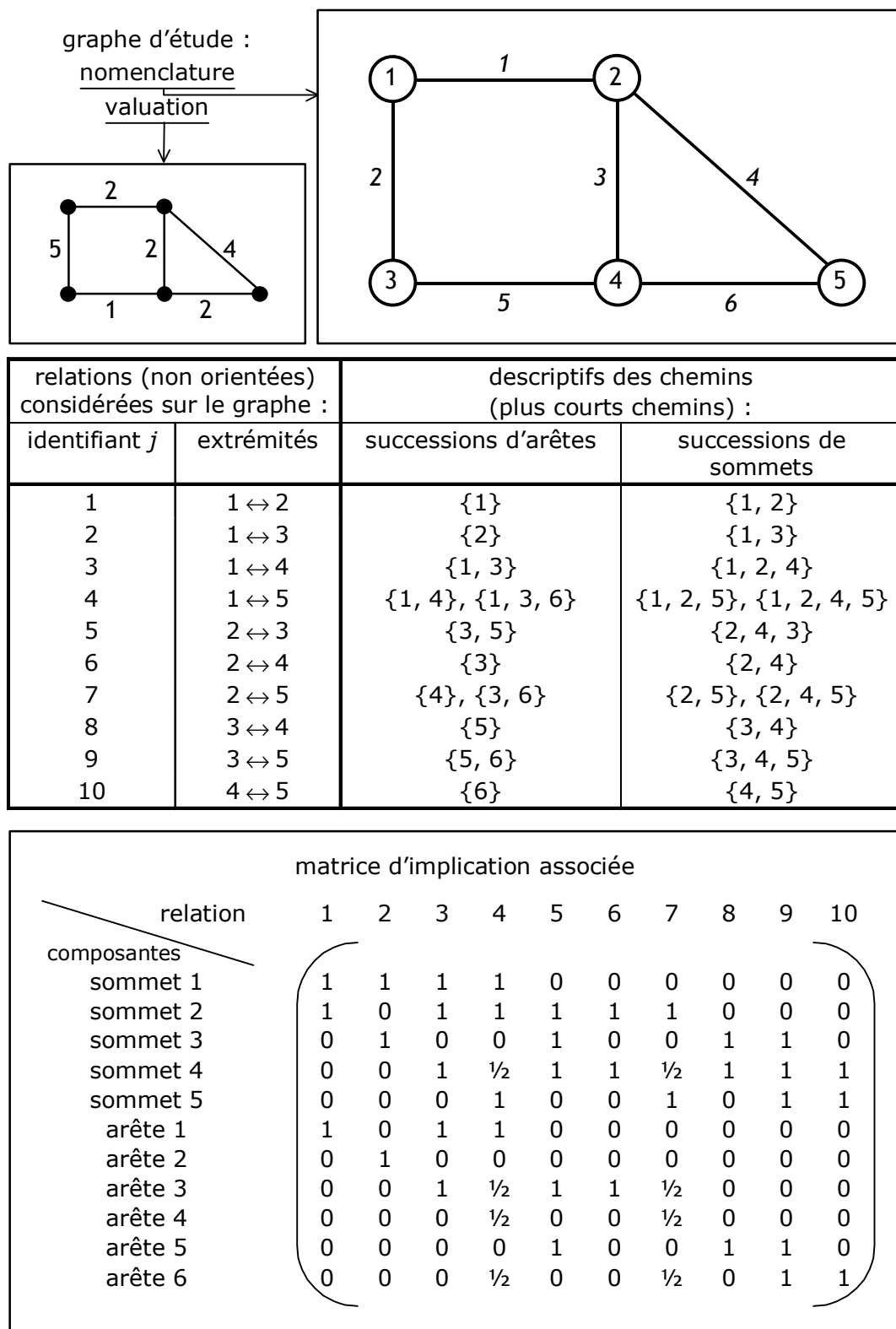


Figure 3 : exemple de calcul d'une matrice d'implication - d'après : (Gleyze, 2005)

Si on lit la matrice en lignes, il apparaît que la ligne i décrit le taux d'implication de la composante i au sein des différents chemins origine-destination construits sur le réseau : dans la suite, nous appellerons ce vecteur « signature relationnelle » de la composante i (Gleyze, 2005).

→ Ainsi, la quantification de la dépendance relationnelle entre deux composantes i et j revient donc à mesurer une 'proximité' ou une 'ressemblance' entre leurs signatures relationnelles.

Telle que nous l'avons définie qualitativement, la dépendance relationnelle entre ces deux composantes est d'autant plus forte qu'elles participent conjointement à des chemins, c'est-à-dire que les coefficients non nuls de leurs signatures relationnelles sont localisés dans les mêmes colonnes. Sur l'exemple de la Figure 3, nous notons que la proportion de chemins partagés par deux composantes n'est pas la même selon que l'on se place du point de vue de l'une ou de l'autre des deux composantes. Ainsi, l'arête 2 et le sommet 3 sont solidarisés par la relation 2 (chemin reliant les sommets 1 et 3) – mais celle-ci constitue la seule relation dans laquelle l'arête 2 est impliquée, tandis que le sommet 3 est impliqué dans trois autres relations.

→ La mesure de la dépendance relationnelle d'une composante i à une composante j doit donc rendre compte de l'écart de la signature relationnelle i à la signature relationnelle j : elle constitue à ce titre une mesure asymétrique.

En pratique, il existe dans la littérature de nombreuses méthodes pour comparer des vecteurs.

Les statistiques fournissent ainsi des distances – en particulier la distance euclidienne ou la distance du χ^2 (Dervin, 1996) – mais nous les avons écartées d'emblée, car elles sont associées à des vecteurs de variables continues et ne représentent par conséquent aucune réalité physique pour le problème qui nous intéresse : que penser en effet de la distance entre deux composantes dans l'espace multidimensionnel des relations envisageables sur le réseau ?

Outre ces outils statistiques, les mathématiques appliquées proposent par ailleurs des distances spécifiquement conçues pour comparer deux vecteurs de booléens, signalant la présence (codage 1 ou « true ») ou l'absence (codage 0 ou « false ») des différents attributs listés pour les deux individus statistiques associés. Ceci est pertinent pour notre étude, car il s'agit bien de comparer deux composantes d'un réseau en fonction des relations auxquelles elles participent toutes deux (« true ») ou non (« false »)⁵.

⁵ Par construction, les signatures relationnelles ne sont pas forcément des vecteurs de booléens, car il peut arriver que certaines relations soient réalisées par plusieurs chemins, et que certains coefficients soient par conséquent égaux à des fractions. En pratique, ce point relève du détail, et nous apporterons en temps voulu les modifications nécessaires à la mesure que nous aurons retenue.

La distance classiquement utilisée à cet effet est la distance de Jaccard. Pour deux vecteurs A et B comptant respectivement a et b attributs égaux à 1 et en possédant un nombre n en commun⁶, cette distance normalisée vaut :

$$d_{AB}^{\text{Jaccard}} = 1 - \frac{n}{a + b - n}$$

Nous avons testé cette mesure pour comparer un vecteur A à différents vecteurs B (cf Figure 4). Le choix des vecteurs B est tel que nous décrivons la gamme des distances allant de 0 (vecteur A lui-même) à 1 (vecteur B_4 correspondant au négatif de A).

L'observation des résultats fait valoir deux défauts rédhibitoires :

- la mesure n'est pas capable de discriminer les deux vecteurs B_1 et B_2 , alors qu'ils ont été construits de manières radicalement différentes : le premier est inclus dans A , tandis que le second recouvre A de moitié et possède tous les attributs manquant à A ;
- la mesure est symétrique, et elle ne peut pas être scindée en deux parties asymétriques : en d'autres termes, la distance entre A et B ne peut pas s'exprimer comme l'addition d'un écart de A à B et d'un écart de B à A , ce qui va à l'encontre de notre souhait initial d'isoler la dépendance relationnelle d'une composante à une autre.

⁶ Cette mesure – comme celle qui sera présentée ensuite – effectue la comparaison sur les attributs partagés (la présence de « 1 » aux mêmes endroits), mais pas sur l'absence partagée d'attributs (la présence de « 0 » aux mêmes endroits). De fait, la prise en compte de ces « dégoûts partagés » (pour reprendre l'expression de (Brunet, 2004)) bruerait trop fortement la mesure dans notre application, car les densités d'attributs dans les vecteurs (*i.e.* la proportion de relations au sein desquelles les composantes sont impliquées) est généralement très faible.

A	0	1	1	0	0	0	1	1	nombre d'attributs : $a = 4$	$\rightarrow d_{AA}^{\text{Jaccard}} = 0$
B_1	0	0	0	0	0	0	0	1	nombre d'attributs : $b = 1$, dont $n = 1$ communs avec A	$\rightarrow d_{AB_1}^{\text{Jaccard}} = 0,75$
B_2	1	0	1	1	1	1	0	1	$b = 6$, $n = 2$	$\rightarrow d_{AB_2}^{\text{Jaccard}} = 0,75$
B_3	1	0	0	1	1	1	0	1	$b = 5$, $n = 1$	$\rightarrow d_{AB_3}^{\text{Jaccard}} = 0,875$
B_4	1	0	0	1	1	1	0	0	$b = 4$, $n = 0$	$\rightarrow d_{AB_4}^{\text{Jaccard}} = 1$

Figure 4 : exemples de distances de Jaccard sur des vecteurs de longueur 8

Partant de ce constat, nous avons approfondi nos recherches, et la solution à notre problème nous vient contre toute attente de la lexicographie. Dans ce domaine de recherche, la comparaison des styles d'écrivains consiste en effet à confronter la présence ou l'absence des différents mots de la langue au sein de leurs champs lexicaux respectifs, et se ramène donc bien à un problème de distance entre deux vecteurs de booléens. Dans ses travaux consacrés aux auteurs français, (Brunet, 2004) propose de remplacer la distance de Jaccard par une distance plus intuitive (les notations sont les mêmes que celles utilisées pour la distance de Jaccard) :

$$d_{AB}^{\text{Brunet}} = \frac{a-n}{a} + \frac{b-n}{b}$$

Par construction, le premier terme représente la proportion d'attributs de A que l'on ne retrouve pas dans B . Symétriquement, le second terme représente la proportion d'attributs de B que l'on ne retrouve pas dans A .

Aussi, nous proposons de décomposer cette distance entre A et B selon un écart de A à B et un écart de B à A :

$$d_{AB}^{\text{Brunet}} = d_{A \rightarrow B} + d_{B \rightarrow A} \text{ avec } d_{A \rightarrow B} = \frac{a-n}{a} \text{ et } d_{B \rightarrow A} = \frac{b-n}{b}$$

En cela, la distance proposée par (Brunet, 2004) nous permet d'élaborer la mesure asymétrique d'un écart entre deux vecteurs, conformément à nos souhaits.

Si l'on reprend les exemples étudiés en Figure 4, nous nous apercevons à présent que les défauts précédemment relevés avec la distance de Jaccard sont

ici palliés par la nouvelle mesure $d_{B \rightarrow A}$ dérivée de la distance proposée par (Brunet, 2004), à savoir (cf Figure 5) :

- les vecteurs B_1 et B_2 sont à présent correctement discriminés : l'écart de B_1 à A est nul car $B_1 \subset A$, tandis que l'écart de B_2 à A vaut $2/3$ car $2/3$ des attributs de B_2 ne sont pas présents dans A ,
- par construction, la mesure est asymétrique : en particulier, l'écart de B_1 à A est nul, alors que l'écart de A à B_1 est élevé ($= 3/4$) compte-tenu de l'existence de nombreux attributs de A non partagés par B_1 .

A	<table><tr><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	1	0	0	0	1	1	$a = 4 \rightarrow d_{A \rightarrow A} = 0$
0	1	1	0	0	0	1	1			
B_1	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	0	0	0	0	0	0	0	1	$b = 1, n = 1 \rightarrow d_{B_1 \rightarrow A} = 0 \text{ et } d_{A \rightarrow B_1} = 0,75$
0	0	0	0	0	0	0	1			
B_2	<table><tr><td>1</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td></tr></table>	1	0	1	1	1	1	0	1	$b = 6, n = 2 \rightarrow d_{B_2 \rightarrow A} = 0,667 \text{ et } d_{A \rightarrow B_2} = 0,5$
1	0	1	1	1	1	0	1			
B_3	<table><tr><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td></tr></table>	1	0	0	1	1	1	0	1	$b = 5, n = 1 \rightarrow d_{B_3 \rightarrow A} = 0,8 \text{ et } d_{A \rightarrow B_3} = 0,75$
1	0	0	1	1	1	0	1			
B_4	<table><tr><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td></tr></table>	1	0	0	1	1	1	0	0	$b = 4, n = 0 \rightarrow d_{B_4 \rightarrow A} = 1 \text{ et } d_{A \rightarrow B_4} = 1$
1	0	0	1	1	1	0	0			

Figure 5 : exemples de décomposition de la distance proposée par (Brunet, 2004) sur des vecteurs de longueur 8

Transposée aux signatures relationnelles, la quantité normalisée $1 - d_{B \rightarrow A}$ est une mesure pertinente de la dépendance relationnelle de la composante B à la composante A , car elle correspond à la proportion de chemins empruntant la composante A parmi l'ensemble de tous les chemins empruntant la composante B .

→ La dépendance relationnelle de B à A est donc mesurée par le taux de relations articulées conjointement sur les deux composantes, parmi l'ensemble de toutes les relations impliquant B .

Afin d'adapter cette mesure aux signatures relationnelles remplies non seulement de 1 et de 0, mais également de fractions, nous proposons de remplacer les quantités a et b par les marges des vecteurs associés, et la quantité n par la somme des différences POSITIVES relevées deux à deux entre les cellules du vecteur B et celles du vecteur A (cette transposition est naturelle et conduit aux mêmes résultats que ceux obtenus sur des vecteurs booléens lorsque les proportions manipulées sont réduites à des 0 et des 1).

En notant $\rho_k(A)$ (resp. $\rho_k(B)$) représente la proportion de chemins empruntant A (resp. B) au titre de la relation k , et en notant p_k le poids attribué à la relation k , la dépendance relationnelle de la composante B à la composante A vaut :

$$D(B, A) = \text{dépendance relationnelle de } B \text{ à } A$$

$$= 1 - \frac{\sum_{\text{relation } k} p_k \cdot \max [(\rho_k(B) - \rho_k(A)); 0]}{\sum_{\text{relation } k} p_k \cdot \rho_k(B)} = \frac{\sum_{\text{relation } k} p_k \cdot \min [\rho_k(B); \rho_k(A)]}{\sum_{\text{relation } k} p_k \cdot \rho_k(B)}$$

Il est important de noter que, par construction :

- cet indice correspond au pourcentage de relations pondérées empruntant à la fois A et B parmi toutes celles qui empruntent B : il est donc normalisé et peut donc s'exprimer comme un pourcentage ;
- cet indice est défini pour une composante du réseau vis-à-vis d'une autre, mais rien n'empêche de l'étendre à des familles de composantes pour peu que l'on redéfinisse ce que l'on entend par « emprunter une famille de composantes » ;
- en effectuant la somme pondérée des dépendances relationnelles de toutes les composantes du réseau à une composante donnée A , il est enfin possible d'obtenir un indice de dépendance relationnelle globale à A ⁷.

Outre les applications que nous développerons dans la suite pour effectuer des regroupements de sommets, l'indice de dépendance relationnelle est intéressant pour identifier les combinaisons récurrentes de composantes au sein des chemins. Cette propriété est particulièrement pertinente dans un contexte de risques, car elle permet d'anticiper les zones du réseau qui seront les plus affectées par l'endommagement d'une composante donnée (de fait, il s'agit des zones qui sont le plus dépendantes relationnellement de cette composante).

La Figure 6 représente l'indice de dépendance relationnelle à la station de métro République, calculé pour l'ensemble des composantes (stations et tronçons ferrés) du réseau de métro parisien.

⁷ En pratique, nous avons constaté sur nos différents réseaux d'étude que la dépendance relationnelle à une composante est extrêmement proche de sa centralité intermédiaire – c'est-à-dire du poids total des chemins empruntant la composante (bien qu'aucune relation mathématique ne relie explicitement ces indices), aussi n'avons-nous pas entrepris d'approfondir l'indice agrégé de dépendance relationnelle à une composante.

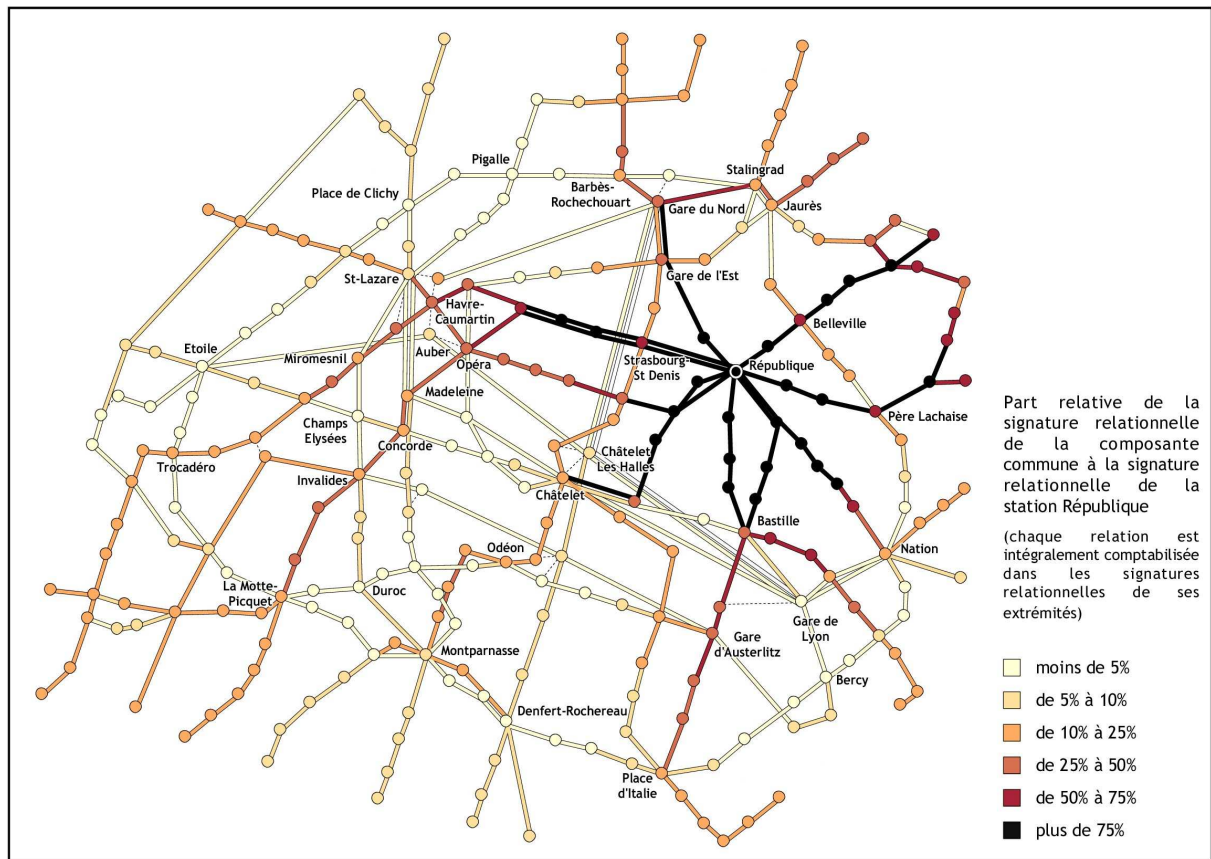


Figure 6 : dépendance relationnelle des stations et tronçons ferrés du réseau de métro parisien à la station République - source : (Gleyze, 2005)

Sur cette figure, nous repérons les composantes solidaires de République : il s'agit des composantes situées sur les lignes passant par République, mais également certaines portions de lignes qui en dépendent par le jeu des correspondances (ainsi en est-il du tronçon de la ligne 4, reliant la Gare du Nord à Montparnasse, sans passer par République). *A contrario*, cet indicateur permet de repérer les composantes complémentaires de République, c'est-à-dire les composantes peu tributaires relationnellement de cette station, et donc probablement peu affectées par un éventuel endommagement de celle-ci (c'est notamment le cas de la ligne 1, reliant d'est en ouest Nation à Étoile).

Au-delà de la cartographie des dépendances relationnelles relatives à une composante donnée, l'indice de dépendance relationnelle peut être calculé pour toutes les paires de composantes et être recensé dans une matrice carrée (asymétrique, car la dépendance de *A* à *B* est, par construction, différente de la dépendance de *B* à *A*). La Figure 7 présente à ce titre la matrice des dépendances relationnelles calculées sur le graphe d'exemple de la Figure 3, en considérant que les relations sont équipondérées.

matrice des dépendances relationnelles											
composantes	s1	s2	s3	s4	s5	a1	a2	a3	a4	a5	a6
sommet 1	1	0,75	0,25	0,38	0,25	0,75	0,25	0,38	0,13	0	0,13
sommet 2	0,50	1	0,17	0,67	0,33	0,50	0	0,67	0,17	0,17	0,17
sommet 3	0,25	0,25	1	0,75	0,25	0	0,25	0,25	0	0,75	0,25
sommet 4	0,21	0,57	0,43	1	0,43	0,21	0	0,57	0,14	0,43	0,43
sommet 5	0,25	0,50	0,25	0,75	1	0,25	0	0,25	0,25	0,25	0,75
arête 1	1	1	0	0,50	0,33	1	0	0,50	0,17	0	0,17
arête 2	1	0	1	0	0	0	1	0	0	0	0
arête 3	0,38	1	0,25	1	0,25	0,38	0	1	0,25	0,25	0,25
arête 4	0,50	1	0	1	1	0,50	0	1	1	0	1
arête 5	0	0,33	1	1	0,33	0	0	0,33	0	1	0,33
arête 6	0,17	0,33	0,33	1	1	0,17	0	0,33	0,33	0,33	1

Figure 7 : matrice des dépendances relationnelles associée au graphe de la Figure 3

La matrice des dépendances relationnelles va nous permettre à présent d'effectuer des regroupements de sommets, selon une structure arborescente révélant les dépendances les plus significatives.

Le regroupement des sommets du réseau à partir de leurs dépendances relationnelles - La constitution de « plate-formes relationnelles »

À cette étape, nous considérons que nous disposons de la matrice des dépendances relationnelles des composantes du réseau que l'on étudie.

Notre objectif est d'opérer des regroupements sur les sommets, afin de proposer une simplification du réseau qui soit représentative des combinaisons de composantes les plus fréquentes au sein des chemins.

Ces combinaisons étant révélées par l'indice de dépendance relationnelle, nous allons effectuer les regroupements de sommets en se fondant sur les valeurs les plus élevées de la matrice des dépendances relationnelles.

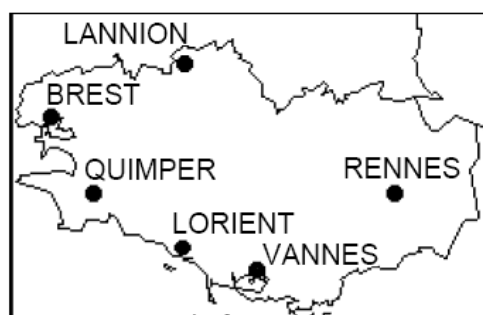
La procédure que nous adoptons est la suivante :

- s'agissant d'effectuer des regroupements de sommets, nous ne considérons que les dépendances relationnelles de sommet à sommet ;
- au sein de la signature relationnelle d'un sommet, nous ne considérons que les dépendances qui concernent les sommets voisins (c'est-à-dire reliés directement par une arête) – nous justifions ce choix par la structure même des chemins, qui exige qu'avant d'atteindre un sommet voisin de degré n , un chemin transite au préalable par un sommet voisin de degré $n-1$.

Le rattachement éventuel d'un sommet à un autre s'effectue donc à partir de l'observation de ses dépendances relationnelles immédiates (*i.e.* à destination de ses voisins directs).

Le chevelu des chemins articulé sur un sommet A étant homogène à un flux, les dépendances relationnelles du sommet A à ses voisins B , calculés à partir de ces chemins, quantifient les 'dominations' des voisins B sur A en termes de flux. Cette analogie nous permet d'utiliser la méthode élaborée par (Nyustén & Dacey, 1968) pour mettre en évidence les relations dominantes sur un ensemble de sommets qui s'échangent des flux.

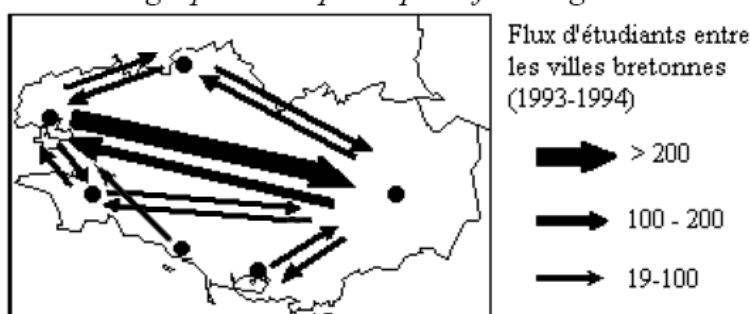
Selon cette méthode, on considère chaque sommet A , et on identifie le sommet B vers lequel A envoie le flux le plus important. Si la somme des flux reçus par B est supérieure à la somme des flux reçus par A , alors on considère que A est 'dominé' par B (cf exemple en Figure 8).



matrice des flux : migrations d'étudiants

		vers (j)						
Fij		BRE	LAN	LOR	QUI	REN	VAN	Oi
de (i)	BRE	-	19	10	54	369	14	466
	LAN	40	-	4	0	48	6	98
	LOR	19	0	-	2	6	2	29
	QUI	67	0	0	-	29	9	105
	REN	190	62	5	21	-	70	348
	VAN	8	1	0	0	56	-	65
	Dj	324	82	19	77	508	101	1111

Cartographie des principaux flux migratoires



METHODE NYUSTEN & DACEY : i est dominé par j si :

- (1) i envoie son flux le plus important vers j
- (2) La somme des arrivées de j est plus importante que la somme des arrivées de i



Figure 8 : exemple d'application de la méthode des flux dominants de (Nyusten & Dacey, 1968) sur les migrations d'étudiants au sein des villes bretonnes en 1993-1994 - d'après : (Grasland, 2007), sur la base d'une enquête effectuée par la Direction des Etudes et de la Prospective du Ministère de l'Education Nationale et de l'Enseignement Supérieur

Cette méthode peut être appliquée au problème qui nous intéresse, en effectuant les transpositions suivantes :

- le flux d'un sommet A vers un sommet B est assimilé à la dépendance relationnelle de A à B , pour peu que A soit directement relié à B ⁸ ;
- le sommet B dont A dépend le plus domine effectivement A si le sommet B est impliqué dans plus de chemins que le sommet A ⁹.

En reprenant les notations utilisées précédemment pour définir la dépendance relationnelle d'une composante à une autre, nous proposons d'effectuer le rattachement du sommet A par dépendance relationnelle au sommet B tel que :

$$B = \underset{\{S, \text{voisin direct de } A\}}{\operatorname{argmax}} \{D(A, S)\}$$

si et seulement si la condition de domination suivante est vérifiée :

$$\sum_{\text{relation } k} p_k \cdot \rho_k(B) > \sum_{\text{relation } k} p_k \cdot \rho_k(A),$$

ou, de manière équivalente⁹ : $D(A, B) > D(B, A)$

Par construction, la condition de domination implique que les rattachements ne peuvent être réciproques (B ne peut être rattaché à A si A est rattaché à B) ni former des cycles (A ne peut être rattaché à C si B est rattaché à A et C est

⁸ dans la mesure où nous ne considérons que les dépendances relationnelles aux voisins immédiats (cf discussion en début de sous-partie)

⁹ La condition exacte énoncée par Nyusten et Dacey exige que le flux total reçu par j soit supérieur au flux total reçu par i . Néanmoins, cette condition n'a pas de sens si on considère la somme des dépendances relationnelles 'reçues' des sommets voisins, car celles-ci sont homogènes des taux (et ne se rapportent donc pas à des quantités comparables, puisque les dénominateurs de ces taux correspondent au poids total des chemins dans lesquels ces sommets sont impliqués et varient donc d'un sommet à un autre). Partant du fait que la dépendance relationnelle d'un sommet i à un sommet j est proportionnelle au poids total des chemins qui impliquent conjointement i et j , il nous a semblé pertinent de transposer la condition de domination en comparant les poids totaux des chemins qui impliquent i d'une part, et j d'autre part.

Par ailleurs, compte-tenu de l'expression de la dépendance relationnelle d'une composante à une autre, on montre facilement que :

$$\begin{aligned} \text{poids total des chemins impliquant } j &\geq \text{poids total des chemins impliquant } i \\ \Leftrightarrow \text{dépendance relationnelle } i \rightarrow j &\geq \text{dépendance relationnelle } j \rightarrow i \end{aligned}$$

Dans ces conditions, la condition de rattachement stipule, de manière équivalente, qu'un rattachement est effectif si la dépendance associée n'est pas plus faible que la dépendance réciproque.

rattaché à B). Pour cette raison, le graphe construit sur les sommets du réseau et sur les arcs réalisant les rattachements par dépendance relationnelle présente une structure de forêt (*i.e.* il est composé de composantes connexes présentant des structures d'arbres).

Les rattachements ne pouvant s'effectuer que selon les valeurs croissantes d'implication (champ $\sum_{\text{relation } k} p_k \cdot \rho_k(\cdot)$), chacun des arbres de cette forêt présente

une structure convergeant sur le sommet possédant la plus forte valeur d'implication. Par construction, ce sommet est singulier au sein de cet arbre car, de proche en proche, il est celui avec lequel les autres sommets de l'arbre partagent le plus de chemins : au sein de cet arbre, ce sommet singulier joue le rôle de 'collecteur' et de 'distributeur' des chemins, à destination des autres zones du graphe. Pour cette raison, nous proposons de baptiser ces arbres « plate-formes relationnelles » et de les identifier grâce aux sommets « charnières » autour desquels s'articulent ces relations.

En définitive, l'indice de dépendance relationnelle permet de partitionner le réseau en plusieurs zones, construites sur des sous-graphes structurés en arbres orientés (les plate-formes relationnelles), articulés autour de sommets collectant et distribuant la majorité des chemins impliqués dans ces zones (les sommets charnières).

La Figure 9 présente le découpage en plate-formes relationnelles obtenu sur le graphe du réseau routier français, à partir de l'observation des plus courts chemins pour l'ensemble de toutes les relations origine-destination envisageables sur le réseau et pondérées surfaciquement (pour plus de détails sur la structure de ce réseau et sur la méthode de pondération, voir (Gleyze, 2005)).

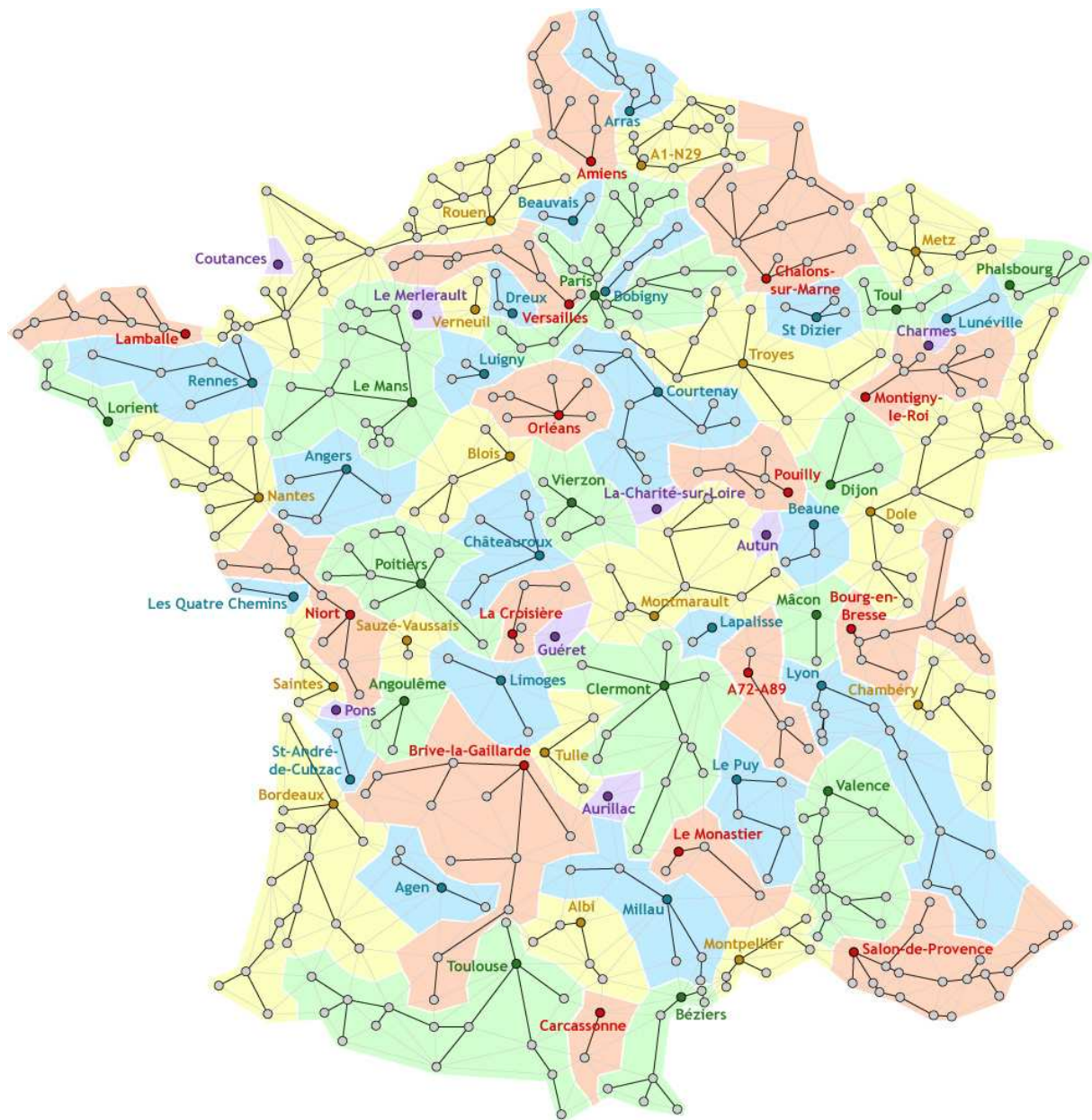


Figure 9 : plate-formes relationnelles du réseau routier français, déduites de l'observation de tous les plus courts chemins envisageables sur le réseau et des dépendances relationnelles qui en découlent - les sommets charnières associés aux plate-formes sont indiqués en couleurs, tandis que les arêtes de rattachement entre les sommets sont surlignées en noir

Ce découpage permet de comprendre les directions privilégiées des plus courts chemins.

Nous notons en particulier que :

- les zones en bordure de réseau acheminent les chemins selon des corridors, vers des secteurs du réseau plus denses et mieux connectés : Salon-de-Provence, Toulouse, Bordeaux, Lamballe, Dole, Rouen, etc. ;
- à l'intérieur de l'espace d'implantation du réseau, les zones adoptent des formes plus ou moins régulières (selon que le réseau distribue les chemins de manière isotrope ou non) et plus ou moins vastes (selon que les chemins se dispersent rapidement ou non) : ainsi, Paris 'capte' les chemins sur une zone

très étendue, mais néanmoins irrégulière compte-tenu des influences des zones voisines (notamment Versailles et Bobigny) – à l'inverse, une zone comme Vierzon est relativement petite et régulière, ce qui dénote un probable éclatement des chemins autour de cette ville (sans pour autant minimiser son rôle de plate-forme, bien au contraire).

Ce regroupement en plate-formes relationnelles peut, dans un second temps, déboucher sur une représentation simplifiée du graphe :

- en agrégeant tous les sommets d'une même zone ;
- en agrégeant toutes les arêtes à cheval entre deux zones voisines.

Cette simplification peut s'accompagner d'une simplification de la matrice d'implication construite sur le réseau initial. Cette simplification ne pose pas de problèmes formels dans la mesure où la notion d'« implication » se transpose immédiatement aux groupes de sommets et aux groupes d'arêtes :

- l'implication d'un sommet agrégé au sein d'une relation est égale à la proportion de chemins transitant par l'ensemble des sommets de la zone associée, au titre de cette relation : en première approximation, on peut évaluer cette proportion à partir des proportions de chemins observées sur les arêtes situées sur la bordure de la zone ;
- l'implication d'une arête agrégée au sein d'une relation est égale à la proportion de chemins transitant par l'ensemble des arêtes reliant les zones situées de part et d'autre de cette arête agrégée, au titre de cette relation.

En première approximation, on peut évaluer ces proportions par un simple bilan entrée-sortie sur les proportions de chemins observées sur les arêtes situées sur la bordure de la zone¹⁰.

Disposant d'une matrice d'implication pour ce réseau simplifié, nous sommes donc en mesure de calculer les dépendances relationnelles entre les sommets agrégés, et d'établir de nouveaux regroupements, à un ordre supérieur.

La Figure 10 présente à ce titre les regroupements d'ordre 2 obtenus sur la simplification du réseau routier français, déduite elle-même du premier regroupement présenté en Figure 9.

¹⁰ Cette méthode de calcul est très simple à mettre en oeuvre, mais elle n'est pas tout à fait exacte, notamment lorsque les chemins impliqués entrent et sortent plusieurs fois dans la zone considérée (auxquels cas ils peuvent être comptabilisés plusieurs fois). Néanmoins, ce cas est rare, et une confrontation des résultats obtenus par cette méthode et des résultats issus de calculs exacts montre que les erreurs d'évaluation sont négligeables.

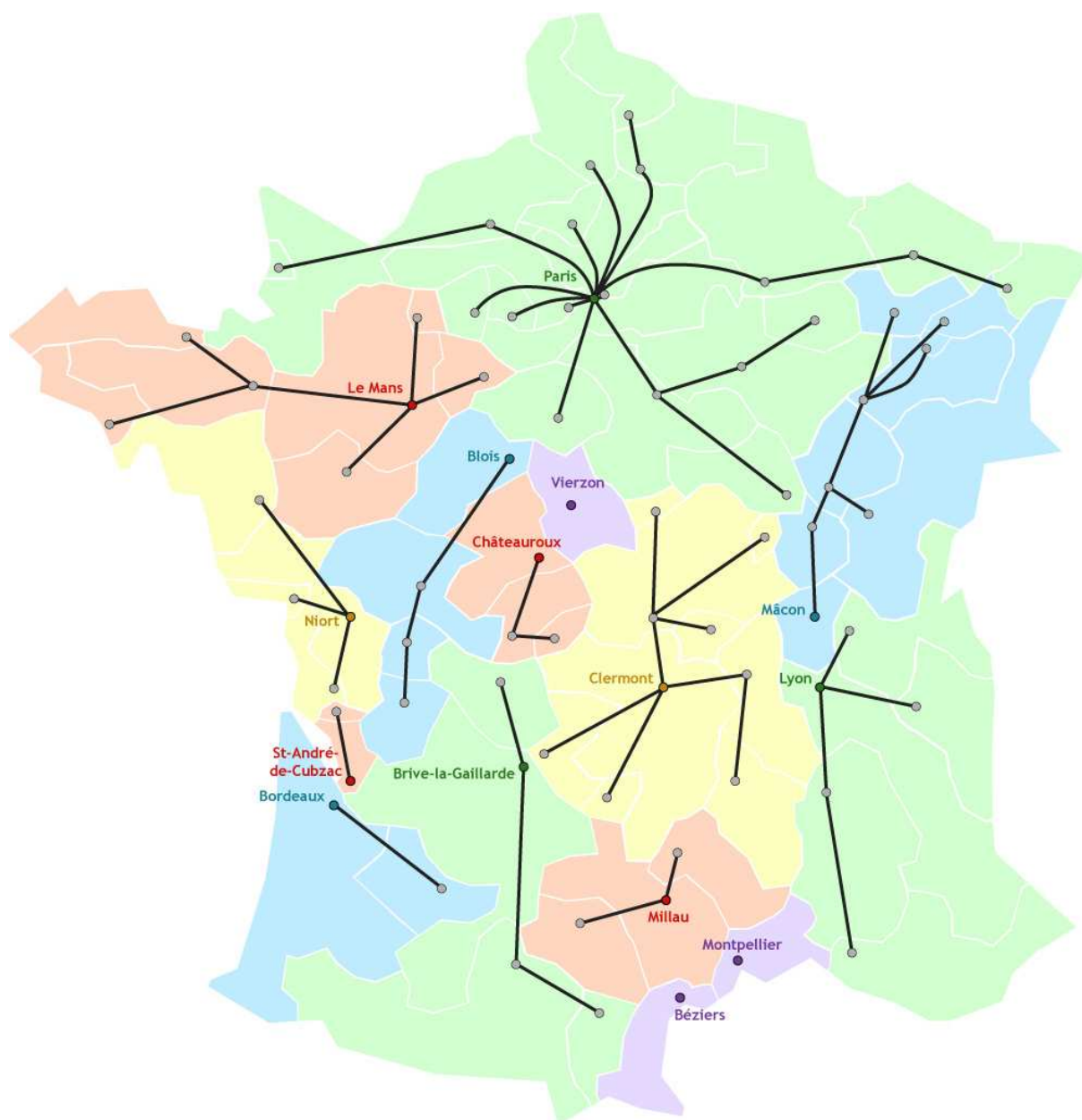


Figure 10 : plate-formes relationnelles du réseau routier français simplifié déduit de la Figure 9

Ce regroupement d'ordre 2 permet de comprendre la manière dont les chemins se structurent à un niveau supérieur. Sur cette figure, on relève notamment le rôle fédérateur de la plate-forme parisienne, captant les chemins sur un large tiers nord du pays. Dans une moindre mesure, la plate-forme clermontoise joue un rôle analogue au centre du pays. Outre ces deux zones, les grandes zones du réseau mettent en évidence des couloirs privilégiés, la plupart du temps selon des axes nord-sud (flancs est, axe Angoulême-Blois, axe Toulouse-Limoges). Notons enfin la persistance de petites zones compactes et isolées, autour desquelles les chemins s'articulent sans privilégier telle ou telle direction en particulier (c'est notamment le cas de la plate-forme de Vierzon, au carrefour des grands couloirs est et ouest et de la zone parisienne).

Au-delà des structures simplifiées que cette méthode permet d'obtenir successivement, l'utilisation des matrices d'implication et leur simplification à chaque étape du processus garantit de pouvoir calculer et représenter tout indicateur calculé à partir des chemins.

Conclusion

La méthode de regroupement des sommets à partir des dépendances relationnelles est pertinente dans le domaine des réseaux d'infrastructure, car elle repose sur l'observation des chemins, qui constituent des objets révélateurs de la structure du réseau et de l'articulation de ses composantes.

Cette méthode consiste à rattacher un sommet B à un sommet A dès lors que la dépendance relationnelle de B à A est suffisamment forte. Cette dépendance correspond à la proportion de chemins impliquant conjointement B et A parmi l'ensemble des chemins impliquant B . Elle est considérée comme 'suffisamment forte' dès lors que A est le sommet dont B est le plus dépendant (parmi ses voisins topologiques directs) et que A est globalement plus impliqué que B dans les chemins observés sur le réseau.

En ce sens, les résultats de ces rattachements renseignent sur la cohésion des composantes induite par le jeu des combinaisons dont elles font l'objet au sein des chemins.

Cette méthode présente deux avantages majeurs :

- elle permet d'aboutir à une représentation simplifiée du réseau initial sur laquelle la méthode peut à nouveau être appliquée : de cette manière, il est possible d'observer à différentes échelles la manière dont les chemins s'articulent et les directions privilégiées qu'ils empruntent ;
- étant construite sur l'observation des chemins, elle permet d'assurer le suivi de tous les indicateurs quantitatifs construits à partir des chemins, en les recalculant sur les agrégations successives du réseau, compte-tenu des matrices d'implication associées (car celles-ci résument, par construction, les informations liées à ces chemins sur les composantes agrégées) : ainsi peut-on envisager le calcul des indicateurs d'accessibilité et de centralité (entre autre) sur les versions agrégées du réseau initial.

Au-delà des réponses qu'elle apporte à la simplification et à l'évaluation des réseaux d'infrastructure en géographie, cette méthode pose les fondements de l'étude formelle de la cohésion des composantes, induite par la structure des chemins observés sur le réseau. En ce sens, elle ouvre la voie à des recherches théoriques sur l'évaluation des dépendances relationnelles (étude comparative

des signatures relationnelles des composantes, mesure de leur écart), sur les critères de regroupement (variations sur la méthode de Nyusten-Dacey appliquée à la matrice des dépendances relationnelles) et sur la représentation simplifiée des réseaux (construction de structures simplifiées à partir des sommets agrégés).

Bibliographie

- Brunet É., 2004, Où l'on mesure la distance entre les distances. Revue en ligne *Texte !*, Rubrique « Dits et inédits », 17 p., Site Internet : <www.revue-texte.net/Inedits/Brunet/Brunet_Distance.html>
- Dervin C., 1996, *Classification automatique – Comment interpréter les résultats ?* Collection STAT-ITCF, 74 p.
- Gleyze J.-F., 2005, *La vulnérabilité structurelle des réseaux de transport*. Thèse de doctorat de Géographie de l'Université de Paris 7, 848 p.
- Grasland C., 2007, *Analyse spatiale : formes et processus – Interactions spatiales : régions, flux, barrières*. Support de cours du master Carthagéo, Université Paris 7, Site Internet : <<http://grasland.script.univ-paris-diderot.fr/ANSPAM1/cours.html>>
- Moody J. & White D., 2003, Social cohesion and embeddedness: a hierarchical conception of social groups. *American Sociological Review*, 68(1): p. 103-127, Site Internet : <eclectic.ss.uci.edu/~drwhite/soc_con17.pdf>
- Nyusten J., Dacey M., 1968, A graph theory interpretatives of nodal regions. *Spatial Analysis, a reader in statistical Geography*, Éd. Berry & Marble, Englewood Cliff, NJ: Prentice Hall
- Wasserman S. & Faust K., 1994, *Social Network Analysis – Methods and applications*. Cambridge University Press, 857 p.