

De la réputation en matière de socio-économie des transports

La notoriété et l'influence

La notoriété est une notion qui renvoie fondamentalement à celle de visibilité et de présence dans l'espace public. Dans le domaine de l'économie transports, celle-ci est parfois confondue à l'influence institutionnelle et médiatique. Dès lors que l'on parle du monde de la recherche ou de l'université s'opposent alors deux logiques¹ : celle qui confère une « autorité » externe, institutionnelle et médiatique, et celle de la reconnaissance par les « pairs ». Cette opposition peut être parfois « radicale » en ce que la « pureté académique » percevrait la notoriété médiatique - qui est aussi une forme de vulgarisation - comme une tentative d'importer une légitimité « *hétéronome* » qui viendrait compenser un manque de rigueur ou de capital scientifique. Or, cette opposition structure en même temps les voies de communication. Aux uns la publication prioritaire dans des revues « scientifiques », aux autres la littérature institutionnelle et la vulgarisation. Ces deux univers - qui ne sont heureusement pas totalement clos et disjoints - vont donc à la fois communiquer de manière différente, mais s'attacher à légitimer (citer) les travaux rentrant dans leur propre catégorie.

L'influence, tirée de la notoriété et probablement de la qualité de la production, découle donc directement de « voies » distinctes, tout comme la « **reconnaissance** » par la communauté scientifique, d'expertise voire la *communauté épistémique*. Ce que Peter Haas décrit comme « *un réseau d'experts fondés sur le savoir — des communautés épistémiques — (jouant) un rôle en articulant les relations de cause à effet des problèmes complexes, en aidant les États à cerner leurs intérêts, en cernant les enjeux pour le débat collectif, en proposant des politiques précises et en désignant les points saillants de la négociation*² »

Tout cela ne fonctionne bien entendu pas de manière homogène y compris au sein de l'espace européen, même si, un temps, la Commission Européenne a tout fait pour rapprocher les équipes de recherche au sein de l'Union.

Que peut-on donc attendre de la recherche d'une mesure - sinon de l'influence - du moins de la notoriété des socio-économistes des transports ? Comment se traduisent leurs stratégies de production et de communication ? Comment fournir une image pour répondre à cette question ?

¹ Bourdieu, P. (1984). *Homo Academicus*. Les Éditions de Minuit.

² Haas, P. M. (1992). « Introduction: Epistemic Communities and International Policy Coordination ». *International Organization*, vol. 46, n° 1

A la recherche de « mesures »

Ceux qui ont eu l'occasion d'enseigner la production et l'utilisation des statistiques se sont souvent évertués à expliquer à leurs élèves que ce qu'on appelle désormais « la donnée » ou la « data », n'est jamais « donnée » !

Quand on parle de statistiques, il faut toujours rappeler qu'il y a derrière un processus de collecte, de traitement, d'analyse et de diffusion de données chiffrées qui peut répondre à des logiques diverses. On sait par exemple qu'on peut définir des enquêtes statistiques (mesure indirecte de grandeurs), mesurer des flux ou des stocks réels directement ou indirectement (comptages, recensements), ou encore utiliser des données issues d'autres processus (documents fiscaux, sociaux etc.).

En face de ces méthodes il y a (toujours) des biais que les statisticiens tentent de réduire par tous les moyens.

Nous savons tous que l'intention du collecteur de données, et la posture de l'enquête jouent un rôle tout comme l'instrument de mesure lorsqu'il s'agit de mesure « physiques » comme les comptages routiers ou la mesure de débits.

Dans le cas qui nous intéresse, le problème tient au fait qu'il n'y a pas d'outil d'enquête, mais des « données » que l'on peut tirer de différentes bases d'information dont la constitution même peut répondre à des objectifs très différents. Celui du référencement universitaire, comme pour OpenAlex, ou celui de la compilation des dépôts légaux comme pour la Bibliothèque Nationale. Pour reprendre les choses simplement : les biais existent et ne sont pas les mêmes, et parfois leur évaluation est quasi-impossible³.

Les sources

Ainsi, lorsque l'on cherche à éclairer la question de la notoriété des économistes des transports, on cherche logiquement ce qui peut bien « exister » comme « données » permettant de mesurer leur production. Cette production doit bien pouvoir être mesurée quantitativement avant d'être évaluée qualitativement. C'est en effet ce double critère qui doit, probablement, fonder la « notoriété » de nos collègues (et donc de nous-mêmes).

Me consacrant à l'examen quantitatif, j'ai dû donc solliciter plusieurs IA en les alimentant de ma connaissance des réseaux et des sites pouvant répertorier les « producteurs » pour leur faire rechercher ces fameux indicateurs constitutifs de la dimension quantitative de la production. Ce travail s'est avéré très itératif et m'a conduit à utiliser parallèlement plusieurs IA (Gemini, Claude, Chat-Gpt et surtout pour l'essentiel). Bien entendu il a fallu vérifier à plusieurs reprises les informations et interdire les facilités utilisées par l'intelligence artificielle qui parfois, utilise des algorithmes pour pallier le manque d'informations. En bref, parfois, elles inventent !

Il fallut donc partir des pratiques des chercheurs eux-mêmes.

Le monde numérique, comme celui des enseignants-chercheurs, se concentre généralement sur un indicateur **bibliométrique**.

Par exemple on s'appuie sur **OpenAlex** qui est un catalogue « du système de recherche mondial » (qui s'appuie sur d'autres sources) qui met en ligne une collection universelle indexée de publications scientifiques. On obtient ainsi pour chaque auteur un « nombre » de travaux

³ Nous avons pu expérimenter nous même que la simple validation d'un nom d'auteur préexistant et la « reconnaissance » de la paternité de certains écrits peut démultiplier la visibilité d'un auteur.

indexés (par an), mais aussi le nombre de citations reçues, et le chiffre magique «h-index⁴ » (ou indice de Hirsch) qui mesure l'impact d'un chercheur selon ses publications et ses citations.

OpenAlex est une base de données bibliographique gratuite gérée par l'organisation à but non lucratif OurResearch et offrant une alternative aux services payants (comme Scopus ou Web of Science). Elle regroupe plus de 250 millions de travaux.

Il faut noter ici que certaines disciplines publient et citent beaucoup, comme la médecine⁵.

Des sources clivantes et clivées

Ainsi pour prendre des collègues français connus économistes des transports et à la retraite, on trouve sur OpenAlex⁶ un nombre de publications compris entre 100 et 200 et un h-index entre 10 et 20. Quelques « jeunes » atteignent près de 400 publications et un h de plus de 50. Et l'amplitude réelle des francophones est de 900.

Ces chiffres comportent plusieurs biais fondamentaux.

LES PRATIQUES

La première tient aux pratiques de la recherche. En effet, et de manière variable selon les disciplines, les chercheurs privilégient les publications dans les revues dites « à comités de lecture », et pratiquent des citations très larges, et la majorité de ces revues sont de langue anglaise. L'objectif de certains est in fine de maximiser le « h-index ».

Par ailleurs on sait qu'il existe de véritables « cercles de citations », compréhensibles d'ailleurs en raison des proximités théoriques ou idéologiques. Sans vouloir considérer l'incidence de ces pratiques ou des algorithmes des sites comme OpenAlex, on ne peut que souligner le manque criant de données institutionnelles. L'indicateur « PIC » est composite (voir plus bas note 6).

Il s'agit de l'ensemble des références de la BnF de la base TRID de HAL et du répertoire RePEc. Les différents indicateurs « académiques » d'OpenAlex (publications, citations, h-index) ne donnent pas des résultats identiques ou comparables à ce que donne Google Scholar ou encore RePec (RePEc (Research Papers in Economics), leur fonctionnement étant différent.

Ainsi l'image dite de notoriété sera pour les uns attachée à un indicateur strictement académique que l'on pourrait qualifier « *d'endogamique* », quand Google Scholar par exemple recense largement les publications dites institutionnelles, mais dont ne connaît pas la méthode d'extraction des données. Ceci explique la différence de visibilité de tel ou tel chercheur selon qu'on le cherche sur OpenAlex ou GoogleScholar. On retrouve la même différence fondamentale en considérant les publications enregistrées par la Bibliothèque Nationale de France (BNF) alimentées par le mécanisme du dépôt légal (voir infra le tableau).

⁴ Un auteur a un indice h de h s'il a publié h articles qui ont été cités au moins h fois chacun.

⁵ A titre d'exemple Didier Raoult a sur OpenAlex plus de 4000 travaux, et plus de 228000 citations ce qui lui donne un h-index de 199 ! A titre comparatif Philippe Aghion, récent prix Nobel d'économie, totalise 677 travaux, 78800 citations et un h-index de 118. Un autre Nobel d'économie, ayant travaillé sur les transports, Robert W. Fogel, totalise lui 280 travaux, 11340 citations et a un h-index de 50.

Quelques bases utilisées

Plateforme	Logique principale	Indexation	Traitement apparent de la littérature grise	Nombre total de références	Domaine
ResearchGate	Réseau social académique privé	Agrégation de publications déposées ou revendiquées par les chercheurs, complétée par exploration automatique de sources externes	Acceptée sans distinction de statut	160 millions environ	Multidisciplinaire, dominante sciences exactes/vie/ingénierie
OpenAlex	Infrastructure ouverte : agrégation automatique de métadonnées depuis de multiples sources ouvertes (Crossref, PubMed, DOAJ, ORCID, Unpaywall, dépôts institutionnels, DataCite)	Automatisée par exploration du Web (<i>web scraping</i>), algorithmes de TAL (NLP) et agrégation de sources variées (Crossref, PubMed, etc.).	Explicitement intégrée comme catégorie propre (preprints, jeux de données, logiciels, rapports) ; extension 2026 de +192M œuvres issues de DataCite et des dépôts	Estimations variables selon les sources allant de 245 millions de notices à 517 millions de works	Toutes disciplines
RePEc	Index bibliographique décentralisé et collaboratif, alimenté par des milliers d'archives (éditeurs, institutions, auteurs) via un protocole commun	Standardisée via des fichiers de métadonnées fournis par les institutions partenaires (format ReDIF).	Très favorable, notamment aux working papers, rapports économiques et littérature pré-publication.	Plus de 5,4 millions d'items indexés dont plus de 1,3 million de documents de travail ; plus de 100 millions de références extraites (CitEc)	Sciences économiques, gestion et disciplines connexes.
TRID	Base de données bibliographique de référence mondiale, fusionnant les bases de la TRB et de l'OCDE.	Indexation contrôlée par organismes de recherche transport ; notices de recherche publiée et en cours	Intégrée nativement et à égalité avec les articles : rapports techniques d'agences (DOT, FHWA...), actes de conférence et documents institutionnels forment une part volontaire et substantielle du corpus	1,5 million	Transports (tous modes) et disciplines connexes (ingénierie, économie des transports, urbanisme)
Google Scholar	Moteur de recherche généraliste appliqué au web académique ; algorithme propriétaire non documenté	Indexation automatique massive de pages web, dépôts, sites éditeurs ; aucune transparence sur les sources	Très favorable en découverte, mais très hétérogène : rapports, thèses, working papers, documents institutionnels et doublons peuvent coexister	~389 millions de documents (estimations variables, 160–389 M)	Multidisciplinaire
HAL	Archive ouverte nationale française ; auto-dépôt par chercheurs (voie verte) ; identifiants pérennes, validation CCSD	Dépôt par chercheurs et institutions ; métadonnées structurées, texte intégral, collections institutionnelles	Forte : thèses, rapports, working papers, prépublications, communications ; accès libre systématique	De 1 à 3 millions de documents	Multidisciplinaire (avec une forte représentation en sciences exactes, humaines et sociales).
BNF (catalogue général)	Catalogue de bibliothèque nationale fondé sur le dépôt légal ; notices bibliographiques + autorités	Catalogage professionnel normé (format InterMarc), notices bibliographiques distinctes des notices d'autorité	Inclut thèses, rapports, documents administratifs, ressources électroniques via le dépôt légal	> 13 M de notices bibliographiques + ~5 M de notices d'autorité	Multidisciplinaire

La description des données

CONSTRUIRE UN INDICATEUR DE PRODUCTION INSTITUTIONNELLE

LES AUTEURS

Le processus de constitution d'une base d'auteurs spécialisés en socio-économie des transports s'est fait selon un processus largement itératif. Nous avons en effet commencé avec un noyau de personnes que nous connaissions, puis utilisé plusieurs programmes d'Intelligence Artificielle (en l'espèce Mistral/vibe, Claude, Gemini, Chat-Gpt) pour élargir la base en s'appuyant sur des raisonnements assez simples. Il est facile en effet d'utiliser les rapports institutionnels pour en extraire des noms d'auteurs, de rechercher les co-auteurs des auteurs déjà recensés, de rechercher des chercheurs à partir de noms d'organismes, etc.. Ce processus itératif aboutit à conserver en l'état actuel un groupe de **165 auteurs**, à partir d'un groupe un peu plus large comprenant des personnes qui, manifestement n'ont pas publié ou n'ont laissé aucune trace visible. C'est précisément là que les sources d'information indiquées plus haut jouent un rôle important et que leurs biais peuvent avoir une incidence sur la composition de notre base d'auteurs.

LES DONNÉES

On l'a dit, les données regroupées dans notre « tableau des métriques » sont disparates aussi bien par leur sources, leur mode de constitution que par leur logique de production.

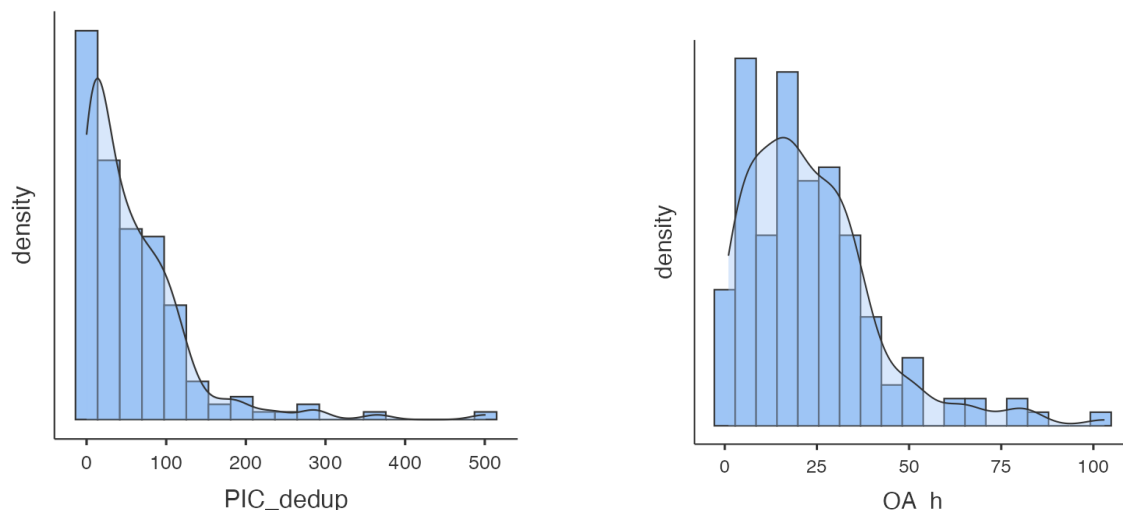
Les variables mesurant la production sont les suivantes, PIC étant une variable « calculée ». Le choix d'un indicateur de production calculé à partir d'extractions automatiques de données n'est pas évident. Nous avons éliminé la sommation pure et simple du nombre de références des bases BnF, HAL, RePEc et TRID, qui reviendrait à comptabiliser plusieurs fois les mêmes travaux. Trois formules ont été testées. L'une d'entre-elles, la plus simple calcule pour chaque auteur l'indice $PIC_{dedup} = BnF + \max(HAL, RePEc, TRID)$. Ainsi au pire l'indicateur est égal au double des publications recensées globalement. Cet indicateur se trouve par ailleurs faiblement corrélé positivement avec l'H-index d'OpenAlex OA_H ($r=+0.348$), ce qui est « logique ».

Statistiques descriptives des métriques						
	OA_h	BnF	RePEc	HAL	GS_h	TRID
Nombre d'auteurs présents	152	22	108	71	24	134
Manquants	13	143	57	94	141	31
Moyenne	23.2	32.9	58.3	25.3	44.1	15.0
Médiane	19.0	19.5	61.5	4	36.5	4.50
Déviatiion standard	17.9	35.1	38.6	57.7	28.9	52.1
Minimum	1	1	1	1	5	1
Maximum	103	137	129	280	120	501

Nota : GS_h (H-index de google Scholar n'a pas été utilisée en raison de son trop faible champ)

L'examen des deux séries H-index et notre indicateur PIC - dont la nature est très différente - fait ressortir effectivement une amplitude bien plus grande pour le « PIC »

Statistiques descriptives des métriques retenues		
	PIC_dedup	OA_h
N	165	152
Manquants	0	13
Moyenne	58.5	23.2
Médiane	40	19.0
Mode	5.00	8.00
Déviatiion standard	69.6	17.9
Minimum	0	1
Maximum	501	103
Coefficient d'asymétrie	2.83	1.50
Asymétrie de l'erreur-standard	0.189	0.197



Les **variables qualitatives**, de leur côté sont les suivantes :

Tableau des variables qualitatives utilisées

Variable	Type	Ce qu'elle mesure
M1	F/M (0/1)	Production d'ouvrages et articles structurants
M2	F/M (0/1)	Apport méthodologique ou conceptuel
M3	F/M (0/1)	Influence institutionnelle (régulation, politiques)
M4	F/M (0/1)	Rayonnement (école, disciples, manuels)

Leur constitution découle de l'analyse des productions et de la visibilité des auteurs. Bien entendu on peut penser que cette perception est biaisée par notre expérience et notre inscription institutionnelle et des référentiels⁷, mais elle dépend essentiellement des traces laissées par les auteurs, et leur nature, puisque nous avons eu largement recours à la recherche automatique de références.

L'analyse croisée des familles et de leurs statistiques est possible, mais elle fournit un très grand nombre de courbes de distribution.

⁷ Comme par exemple de produire ou non des travaux quantitatifs (économétrie, modélisation)...

La méthode

A partir des ces données disparates, qui reflètent probablement tout autant les stratégies de communication des auteurs, que les pratiques ou sociologies de la recherche et des études propres aux différents pays européens nous avons choisi de nous livrer à une analyse exploratoire en ayant recours à l'analyse de données. Notre choix s'est porté sur une articulation entre 3 outils : la classification hiérarchique ascendante (CHA) - qui permet de définir des clusters - l'Analyse en composante principale (ACP), et l'Analyse Factorielle de Données Mixtes (AFDM).

In fine, dans nos analyses nous allons utiliser les variables quantitatives suivantes : H-index d'OpenAlex (OA_H) et l'indicateur PIC qui figure la Production Institutionnelle Comptée. Par ailleurs nous utilisons dans l'analyse factorielle les variables qualitatives M1 à M4 qui décrivent le type d'apport de l'auteur.

Les analyses globales

L'analyse qui suit consiste donc à mieux explorer notre base de socio-économistes européens⁸ spécialisés dans notre domaine. Cette élaboration est itérative. Au total notre base de données sans être exhaustive reflète bien la diversité des producteurs de contenu.

Les grands biais tenant à l'inscription institutionnelle des auteurs et les choix de publication se retrouvent parfaitement dans les différentes bases utilisées.

Il est pour autant intéressant d'essayer de tirer de l'ensemble des informations compilées une image plus synthétique, et une réponse à notre interrogation sur la « notoriété ».

Que dit en effet cette collection de données relatives aux travaux des socio-économistes des transports ?

Deux méthodes sont alors possibles.

La première consiste à mener une **analyse en composantes principales** sur les données quantitatives recueillies.

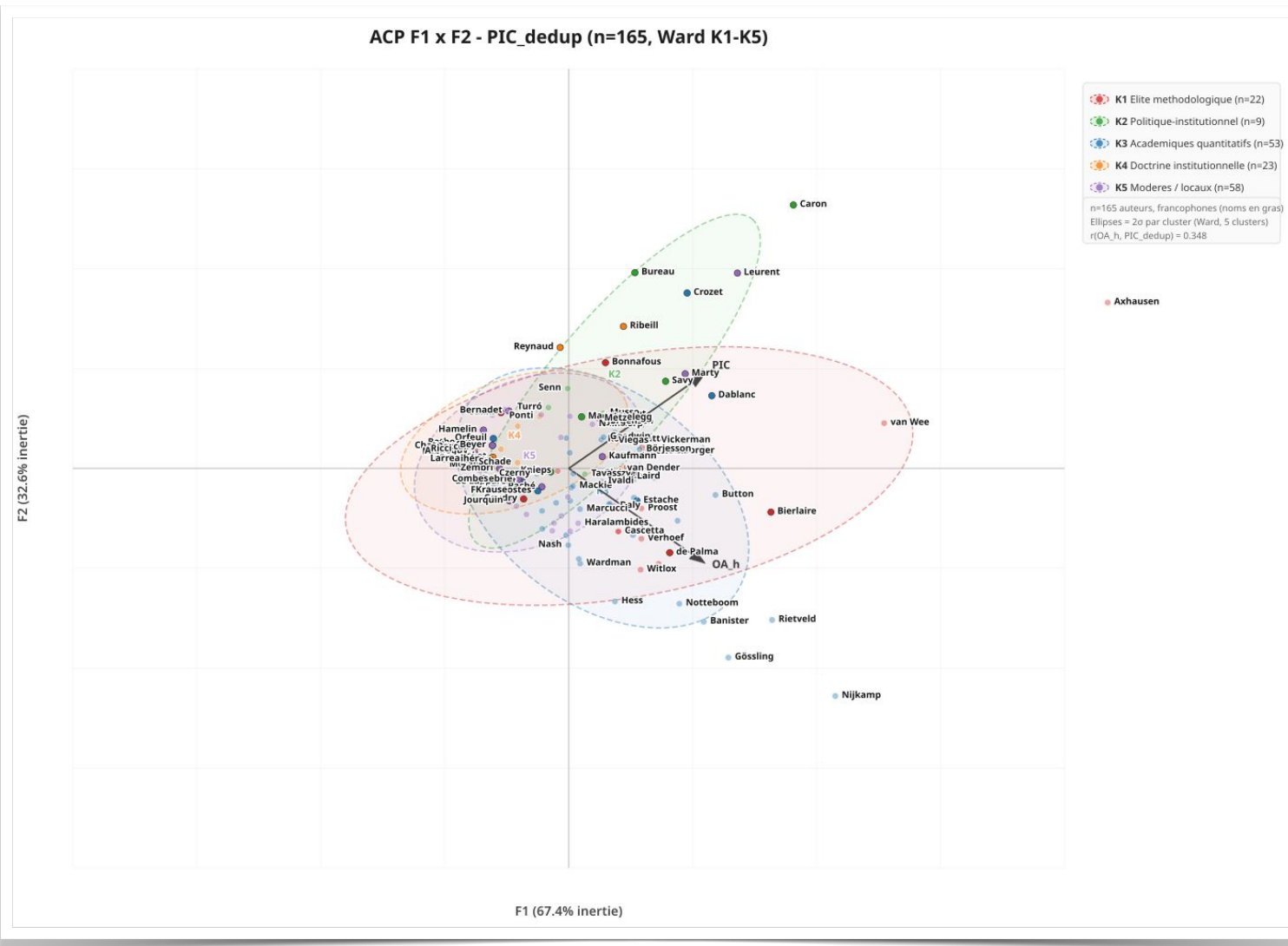
La seconde consiste à mener une **Analyse Factorielle de Données Mixtes** (FAMD en anglais et AFDM en français).

Dans les différentes représentations, tous les points sont figurés, mais seuls les auteurs francophones sont signalés par leur nom, et sont également indiqués par leur patronyme les « auteurs singulier ».

⁸ Nous y avons ajouté arbitrairement quelques québécois comme Marc Gaudry, à la fois francophone et impliqué dans de nombreuses études en France et en Europe.

L'ANALYSE EN COMPOSANTES PRINCIPALES (ACP)

Cette analyse est extrêmement simple. Elle repose en effet sur les deux variables retenues plus haut (OA_h et PIC_dedup), et donc nous avons vu qu'elles étaient intéressantes.



Le premier plan restitue donc 100% de l'inertie ($F1 \approx 67\%$, et $F2 \approx 33\%$).

Le premier axe F1 est un axe de **dimension**. Dans sa direction, OA_h et PIC varient ensemble. Plus on va à droite, plus l'auteur est à la fois visible internationalement (OA_h élevé) et est un producteur « institutionnel » (PIC élevé). Les deux flèches OA_h et PIC pointent vers la droite avec un angle aigu, traduisant une covariance positive ($\text{Cov} = r = +0,35$, les variables étant standardisées).

Le second axe est au contraire un axe de **discordance**. Il oppose les auteurs dont le PIC est fort relativement à leur OA_h (en haut) à ceux dont l'OA_h domine (en bas). Les deux flèches OA_h et PIC sont symétriques de part et d'autre de l'axe horizontal. C'est pour nous l'un des éléments forts de l'analyse : si les deux « visibilité » sont corrélées (faiblement), les auteurs se différencient fortement par leur « communication », et sans doute par la stratégie qui la sous-tend.

Nous avons donc très explicitement **deux critères nettement distincts de « notoriété » : la reconnaissance académique et l'insertion/influence institutionnelle**. En effet, on peut supposer que les publications institutionnelles - baptisées « littérature grise » - relèvent d'un processus production, voire de choix des rédacteurs, fondé sur leur visibilité et leur compétence (procédures d'appel, d'offre, recrutements des socio-économistes, constitution des équipes, ex-ante, etc.), et de processus d'évaluation des travaux ex-post. Ce type de production ne répond pas aux mêmes contraintes et mode d'écriture que la production « d'articles scientifiques ».

On peut penser qu'il existe - de plus en plus - une tendance à la normalisation des articles d'économie s'appuyant sur la structure dite IMRaD-like, c'est à dire : Introduction, Méthodes, Résultats, et Discussion (souvent complété par Résumé/Abstract, Revue de littérature, Conclusion, Références). Cette construction conduit indubitablement à amplifier la dimension bibliographique des travaux, contrairement aux études et recherches d'économie appliquée dites "institutionnelles", ou "littérature grise". Pour autant, une partie de la littérature grise (notes du FMI, de l'OCDE, working papers NBER/CEPR qui sont eux-mêmes de la littérature grise avant publication) importe de plus en plus l'apparat bibliographique académique, précisément parce que ces institutions cherchent la même légitimité scientifique. En outre, l'inflation des citations, découle aussi (surtout ?) des logiques d'évaluation de la recherche (comptage de citations, exigences des « référés », construction de légitimité par affiliation à une littérature), ce qui n'est pas dans le cahier des charges de l'écrasante majorité des productions institutionnelles.

Certains auteurs sont singuliers - au sens où ils ont des métriques incohérentes par rapport à celles de leur groupe. Par exemple Nijkamp et Caron dépassent leur cluster⁹, à l'inverse certains sont singuliers par « discordance » ou par défaut.

L'ANALYSE FACTORIELLE ANALYSE FACTORIELLE DE DONNÉES MIXTES (AFDM)

L'élargissement de l'analyse des données à des variables qualitatives est permise par l'AFDM. Cette analyse combine en effet deux métriques quantitatives et des variables qualitatives.

L'AFDM combine ainsi 4 variables binaires (M1–M4) dites qualitatives, ce sont nos « familles », et 2 variables numériques continues (OA_h, PIC, voir plus haut).

LES CLUSTERS

Statistiques descriptives

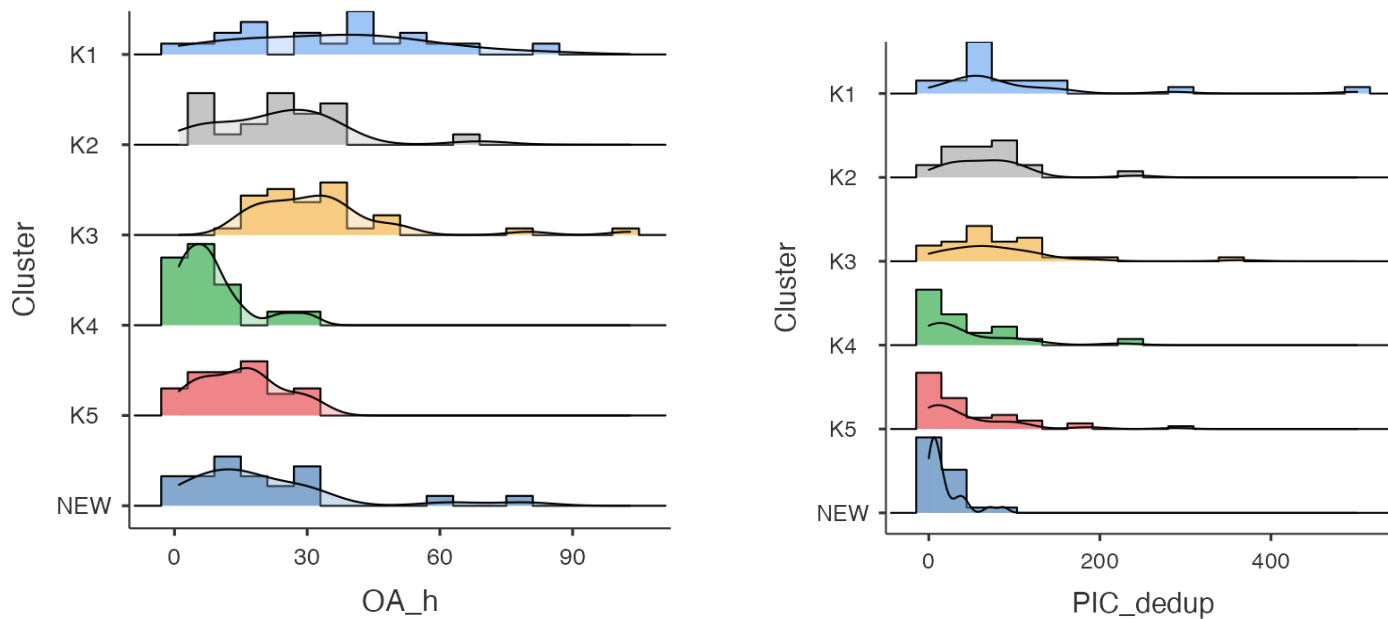
	Nom	Nombre d'auteurs du cluster
Élite méthodologues	K1	22
Institutionnels	K2	9
Académiques quantitatifs anglophones	K3	53
Doctrine institutionnelle	K4	23
Modérés locaux	K5	58

Les 165 auteurs sont répartis par ailleurs en **5 clusters** (Ward). La méthode consiste en une classification hiérarchique ascendante qui, à chaque étape, fusionne les deux groupes

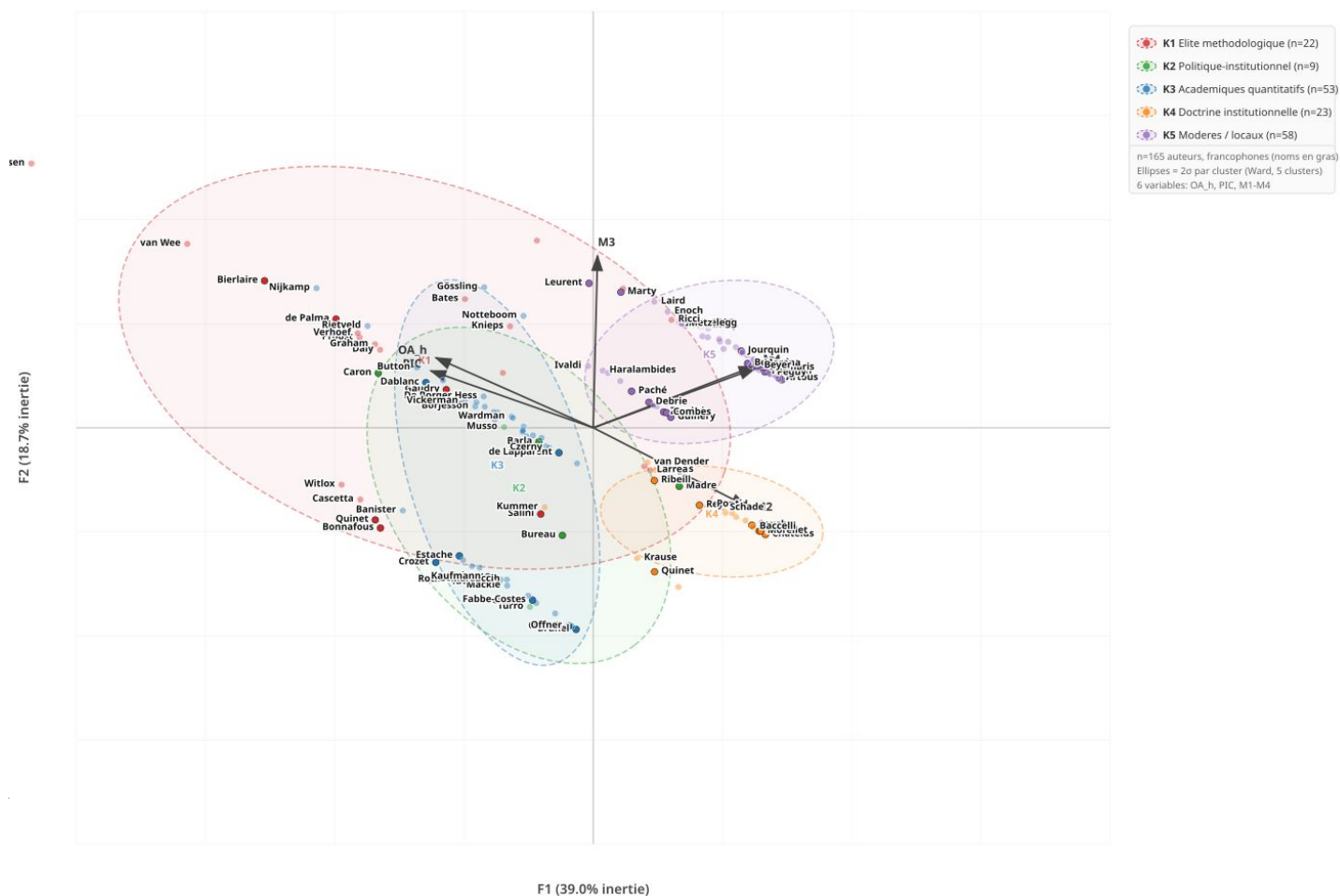
⁹ Nijkamp (K3, OA_h = 103, soit 3× la moyenne K3), Caron (K2, PIC = 365, soit 2,5× la moyenne K2)

minimisant l'augmentation de l'inertie intra-classe (en pratique on mesure l'augmentation de l'inertie intra-classe due à la fusion). Elle produit ainsi des clusters compacts et homogènes. Elle ne mesure pas la distance directe entre centres, mais le « coût de la fusion ». On fusionne la paire ayant le coût minimum. Appliquée ici aux coordonnées factorielles de l'ACP (OA_h, PIC_dedup standardisés), elle répartit l'ensemble des 165 auteurs en 5 clusters K1 à K5, sans reste ni groupe résiduel.

Distribution des métriques selon les clusters



AFDM F1 x F2 - PIC_dedup (n=165, Ward K1-K5)

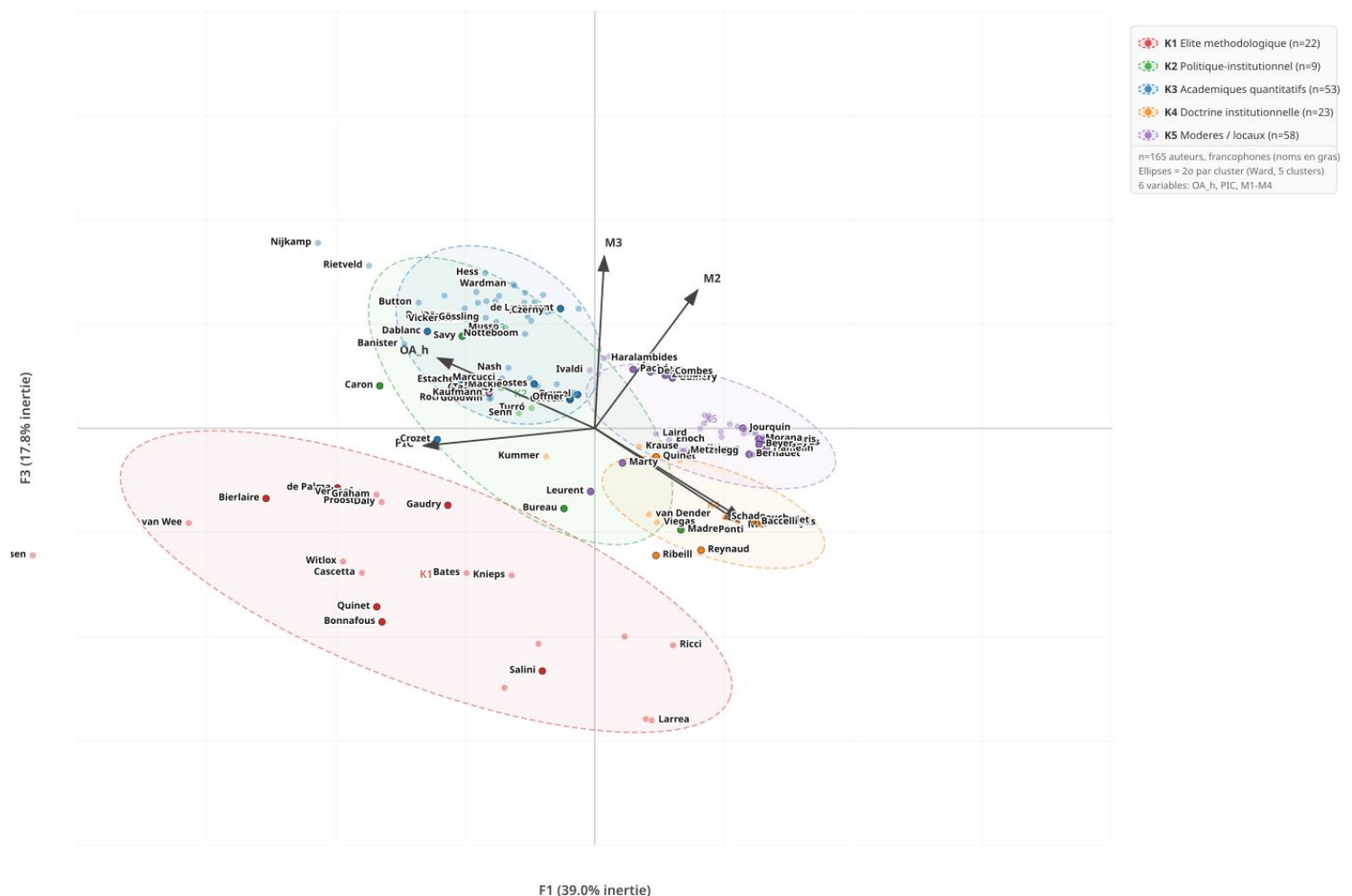


L'inertie cumulée captée par les trois premiers axes (F1+F2+F3) est d'environ 73 %. L'analyse semble bien révéler trois dimensions structurantes : la production académique (F1, 40%), la modélisation (et les techniques quantitatives) (F2, 19%), et la dimension plus politique (F3, 16%). Les 5 clusters se séparent nettement selon ces axes.

L'axe F1 ($\approx 39\%$), oppose assez bien la production académique au profil dit doctrinal.

L'axe F2 ($\approx 19\%$) rendrait compte de l'importance des travaux de modélisation.

AFDM F1 x F3 - PIC_dedup (n=165, Ward K1-K5)



Enfin, l'axe F3 ($\approx 16\%$) réfère plus directement au caractère « politique » des travaux.

L'axe F1 (37,2 %) figure l'opposition entre la visibilité académique (OA_h) et les familles « productives » et à fort rayonnement. De manière grossière on pourrait considérer que les modélisateurs purs sont moins visibles dans les classements bibliométriques.

L'axe F2 (20,5%) oppose l'indicateur PIC à la famille M3 (les institutionnels) ce qui reviendrait à une opposition entre le politique et le qualitatif (méthodologie).

Enfin l'axe F3 (16%) est quasi exclusivement M2 un axe « économétrie pure » qui isole le cluster K1 (baptisé « élite »).

Cette analyse confirme ou renforce cette image du métier de socio-économiste des transports, parfois écartelé entre deux exigences (voir Bourdieu Et P. Haas, op. cit.), et sans doute trois, avec l'émergence d'une demande locale d'expertise et d'études de plus en plus forte.

Il en découle une considérable différence - qui se traduit parfois en opposition relative - des modes de réputation/reconnaissance de ces métiers de l'économie des transports.

Si l'on préfère cela caractérise le côté multiforme de ce métier. Pour autant, le glissement vers l'académisme peut être un risque majeur de perte d'influence et de visibilité de la discipline. La désarticulation est réelle, elle n'est pas interprétative : c'est ce qu'oppose l'axe F1. Pour autant, une partie importante de notre « groupe d'auteurs » se retrouve peu visibles (52 sur 162). Le risque de l'académisme est vérifié : Le cluster le plus politiquement influent (K2) est aussi le plus petit (9 auteurs) et le moins visible académiquement. Le cluster le plus grand et le plus visible (K3, 52 auteurs, OA_h = 32) n'apparaît presque pas dans « l'espace politique » (PIC = 7,8). Il y aurait de manière plus réelle que latente une tendance de la pluralité du métier à se hiérarchiser au détriment de l'influence.

Conclusions ?

Finalement, ce long détour par la compilation de données, et la recherche d'un traitement de nature à livrer une image - une représentation - de la notoriété de nos socio-économistes européens des transports, nous a conduit à une évidence dont l'analyse a pu confirmer la matérialité. L'aide apportée au fil de ce travail par les différentes Intelligences Artificielles a été précieuse pour le traitement de masse, l'identification de profils ignorés, grâce à ses capacités d'investigation rationnelle.

En revanche elle a nécessité un nombre considérable d'*interactions* et de *vérifications*, et a permis de mettre en évidence les limites d'un traitement de masse automatisé. Pour autant, même essentiellement perfectible, ce travail nous a permis de nous interroger sur la trace que laisse notre activité - travail spécifique, spécialisé - désormais souvent baptisé d'expertise dans la presse. Cette trace est celle sur laquelle, sans doute, on pourrait mener - travail beaucoup plus complexe - une analyse des influences et des apports théoriques et pratiques. En effet, comme nous l'avons souligné la « une tendance de la pluralité du métier à se hiérarchiser au détriment de l'influence » peut être une évolution inquiétante.

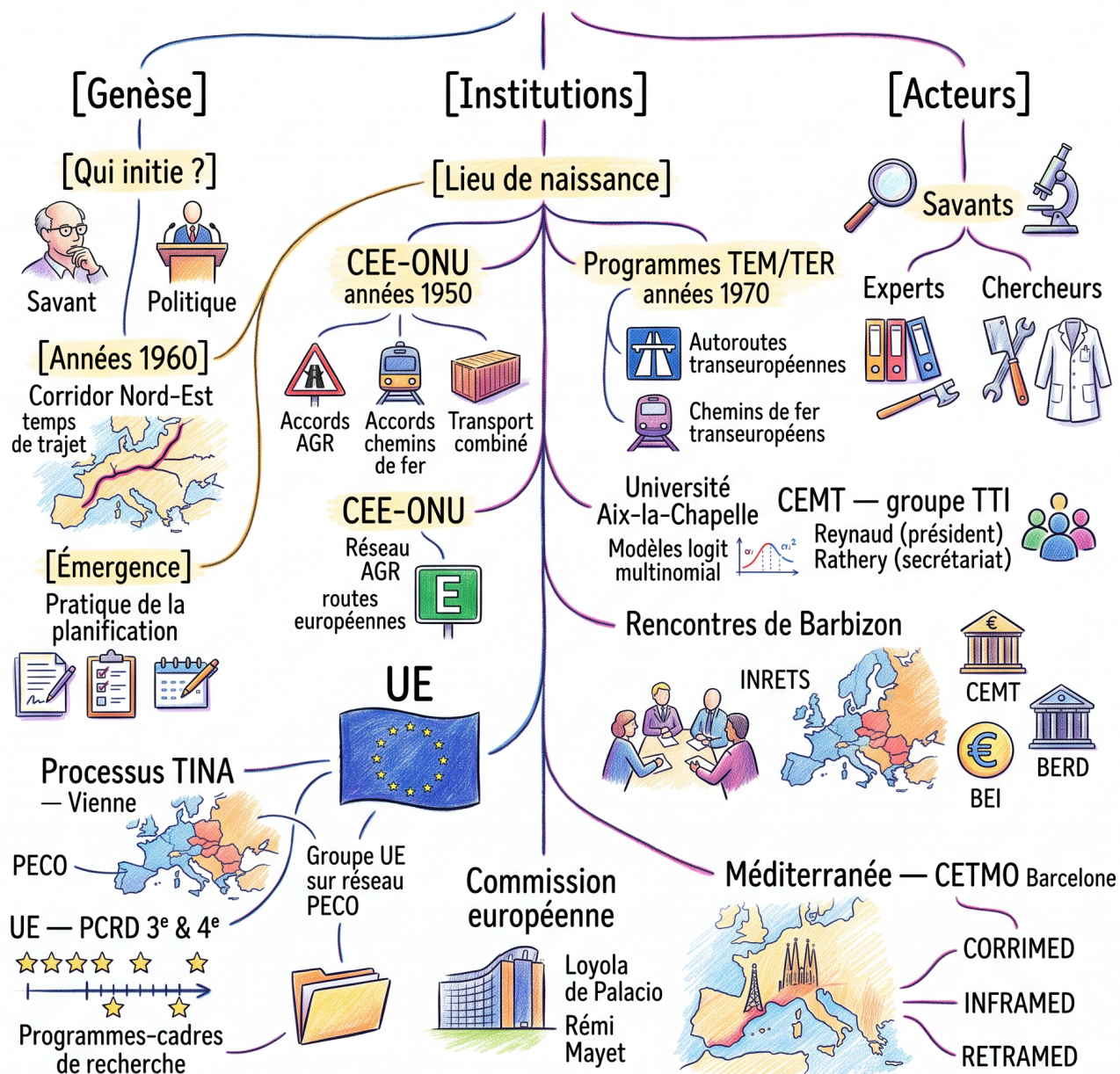
L'approche concrète des mécanismes d'influence est sans doute la plus délicate.

Autant l'influence académique peut-être mesurée - avec de forts biais - via les H-index, autant l'influence institutionnelle, idéologique ou politique réelle est délicate à estimer. La réalité de l'influence - qui va au-delà de l'influence scientifique - est difficilement mesurable et analysable sans se fonder sur une étude précise et détaillée des décisions ou de l'élaboration des politiques. On peut essayer de l'approcher sur des cas précis, isolés, sur des grands mouvements - comme on en a connu en Europe - mais imputer une influence à 10, 20 et a fortiori 140 économistes est impossible.

Nous avons tenté, à partir de documents de témoignage, de reconstituer une sorte d'analyse morphologique des décisions. Le résultat est extrêmement décevant.. l'identification des rôles étant complexe sauf à reconstituer soi même une nouvelle histoire. Nous avons fait un test avec le texte que nous avons co-rédigé avec Christian Reynaud où, pour l'essentiel, ce dernier est interrogé sur sa contribution à l'émergence et la mise en oeuvre d'un concept - celui de **corridor de transport** - qui s'est mué en outil de planification.

L'image complexe qu'on en tire peut effectivement souligner le rôle de tel ou tel. Mais elle provient directement d'un des acteurs. Et soyons honnêtes, la probabilité qu'un sociologue ou un historien traite (et puisse traiter) ce sujet en profondeur est extrêmement faible voire nulle. Alors l'influence restera sans doute longtemps un jugement très subjectif, alimenté par des représentations biaisées, et alimentées pour le demeurer. Mais la prolifération numérique nous donne l'opportunité de mieux la comprendre.

CONCEPT DE CORRIDOR



Annexe

Cette annexe restitue un travail sommaire réalisé à partir d'une liste enrichie par itération d'auteurs ayant traité des transports au cours des XVIIIème et XIXème siècles.

Le passé (XVIIIème et XIXème siècles) et les « traces » laissées par les socio-économistes des transports

L'analyse exploratoire a été faite à partir d'une liste d'un fichier constitué de **102 auteurs**. Sur ces auteurs certains ont peu ou pas d'informations bibliométriques apparentes.

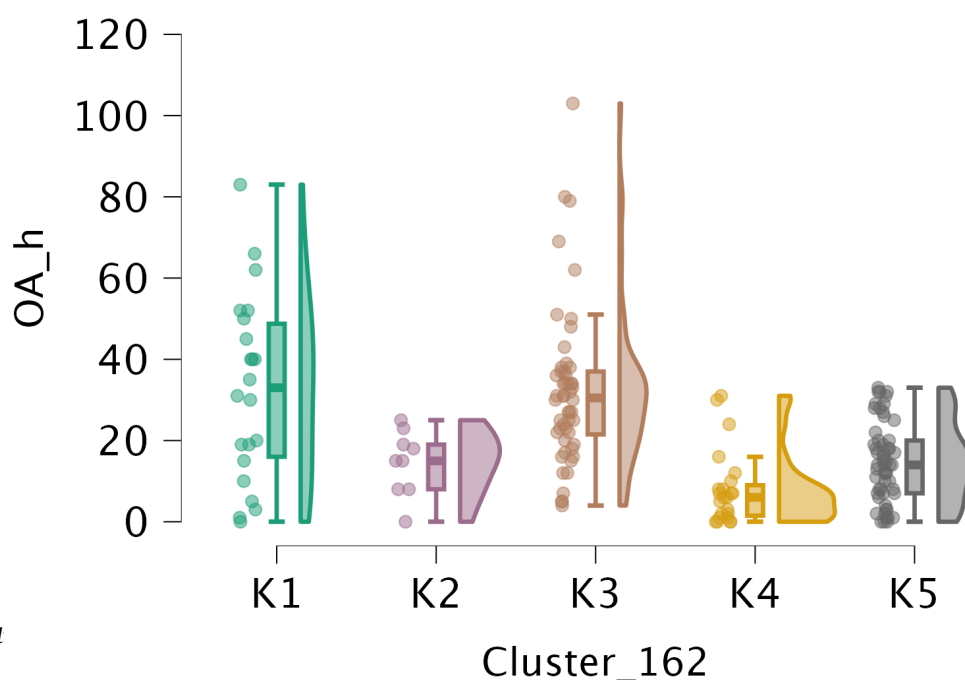
On y trouve des spécialistes des transports (souvent ingénieurs) aussi bien que des économistes ayant consacré une part de leur travail à la question des transports ou de la circulation. Les uns et les autres pouvant avoir des positions institutionnelles plus ou moins fortes.

Du coup, leur présence aussi bien à la BNF que dans les bases bibliométriques est extrêmement variable. De la diversité des auteurs d'école une extrême diversité des traces bibliométriques. Ainsi pour les travaux répertoriés par **OpenAlex**, on a un intervalle de 0 (sic) à 1549, avec une moyenne de 119,5.

Pour les citations l'étendue passe de 0 à plus de 45710. A la **BNF** le nombre de références varie de 0 à 100, deux auteurs n'étant pas présents. Les H-index d'OpenAlex - qui reflètent bien le caractère académique des travaux - ont une moyenne de 12 et varient entre 1 et 78. L'indicateur de production de Semantic Scholar varie de 0 à 1886.

Une analyse plus détaillée est très révélatrice. Un auteur essentiel comme Jules Dupuit, n'a que 9 travaux indexés dans OpenAlex et un h-index de 6. Bien entendu des « généralistes » comme Marx (1 549 et un h-index de 78) ou Adam Smith (298 travaux et h-index de 56), le rendent « invisible ». Les auteurs qui sont à la fois praticiens, vulgarisateurs et théoriciens du chemin de fer (Flachat, Pereire, Enfantin, Dutens) sont proches du zéro bibliométrique : 0 à 11 travaux, 0 à 10 citations. Quelques-uns sortent du lot comme Michel Chevalier (64 travaux, 1 231 citations, 42 notices BnF). A la fois académique et militant saint-simonien il a sans doute une partie significative de sa notoriété liée autant au traité de libre échange avec le Royaume uni, qu'aux transports. Clément Colson - dont la notoriété a été importante et qui a suscité la naissance d'un groupe baptisé « l'anti-Colson » (X-Crise) est complètement absent d'OpenAlex.

Cette singularité - conduit à limiter l'analyse à un nombre d'auteurs réduit (**74 auteurs** retenus (sur 102) : ceux ayant ≥ 3 valeurs non nulles) ce qui revient à exclure de fait les moins visibles.



LES CLUSTERS

La classification hiérarchique ascendante nous permet d'obtenir les clusters suivants :

C1 (n=19) Auteurs marquants mais à visibilité bibliographique modérée et homogène. On trouve là des auteurs importants ayant marqué l'histoire de la discipline comme Dupuit, List, Lardner ou Launhardt, mais peu de généralistes sauf Bastiat. Moyennes brutes : ~50 œuvres OA, ~250 citations OA, BnF ≈ 2 , S² cit. ≈ 167 , S² h ≈ 4 .

Auteurs : von Justi, Campomanes, Turgot, List, Lardner, Bastiat, Dupuit, Madoz, Brunel, Laveleye, Launhardt, Fawcett, Cheysson, Guyot, Picard, Costa, Pantaleoni, De Viti, Barone

C2 (n=19) — Les grands généralistes

On retrouve ici les « grands auteurs » (grande notoriété) qui ont consacré une part de leur œuvre aux transports. Forte présence dans les trois bases utilisées. 282 œuvres OA, ~7 769 citations OA, h ≈ 26 , BnF ≈ 12 , S² cit. ≈ 541 , S² h ≈ 6 .

Auteurs : Smith, Beccaria, Say J.B., Thünen, Cattaneo, Mill, Ferrara, Marx, Engels, Boiteau, Walras, Schmoller, Marshall, Gide, Böhm-Bawerk, Weber, Einaudi, Marlio, Divisia

Les métriques

	OA Works	OA Citations	OA H-Index	BNF	S ² Semantic Scholar	H Index S2
Présents	80	75	71	100	62	102
Manquants	22	27	31	2	40	0
Moyenne	120	2528	12	5	224	
Ecart Type	270	7036	16	15	395	
Minimum	1	0	1	0	0	
Maximum	1549	45710	78	99	1886	

C3 (n=16) — Les auteurs « discrets » à faible empreinte numérique

Ingénieurs, administrateurs et réformateurs dont la contribution aux transports est importante (concessions, chemins de fer, canaux) mais dont les écrits sont peu indexés dans les bases scientifiques contemporaines. ~7 œuvres OA, ~7 citations, BnF ≈ 2 , S² cit. ≈ 14 .

Auteurs : Floridablanca, Flórez Estrada, Brisson, von Gerstner, Enfantin, Pereire É., Pereire I., Lalanne, Dussard, Figuerola, Jacini, Freycinet, Bloch, Colson, Maristany, Thévenez

C4 (n=18) — Les auteurs renommés « reconnus sur OpenAlex, et ignorés de Semantic Scholar »

Economistes théoriciens bien indexés dans OpenAlex (qui couvre largement l'histoire des idées) mais mal couverts par Semantic Scholar (qui privilégie les publications récentes en sciences dures). Fort sur OpenAlex (~175 œuvres, ~2 070 citations, h ≈ 13) mais très faible sur S² (cit. ≈ 61 , h ≈ 2). BnF modéré (≈ 5). On trouve là des auteurs de grande renommée. les fondateurs de l'économie théorique — sont ici parce que leur héritage scientifique est référencé par OpenAlex mais peu cité dans les articles S² modernes.

Auteurs : Galiani, Jovellanos, Ricardo, Melnikov, Chevalier, Proudhon, Spencer, Jullien, Say L., Echegaray, Jevons, Wagner, Menger, Leroy-Beaulieu, Pareto, Witte, Wieser, Wicksell

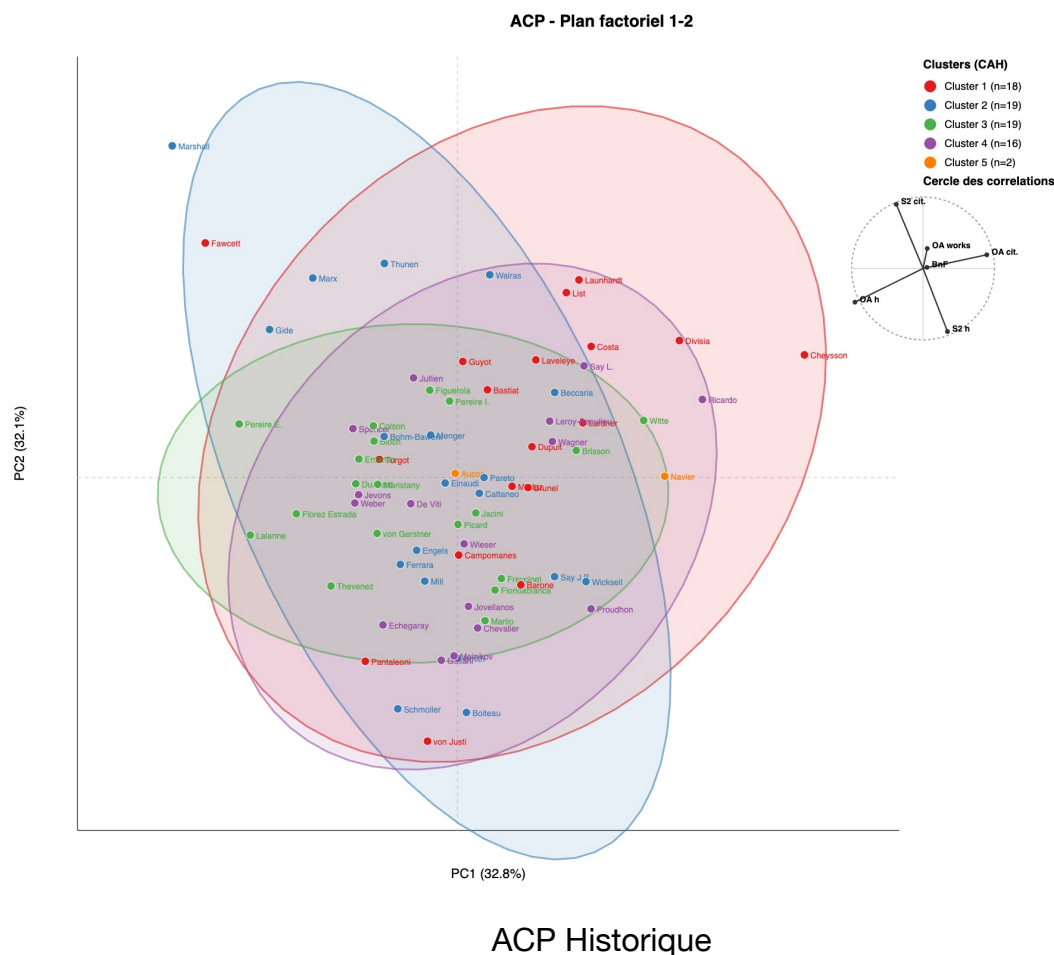
C5 (n=2) — Navier et Auroc

Il s'agit ici d'un groupe très spécifique avec une empreinte OpenAlex faible (~6 œuvres), mais BnF exceptionnellement élevé (≈ 71 , soit 14x la moyenne globale de 7). S² moyen. Ce sont des cas atypiques de reconnaissance patrimoniale nationale non traduite en visibilité scientifique internationale.

Auteurs : Navier, Auroc

L'ANALYSE EN COMPOSANTES PRINCIPALES

Le plan ci-dessous capture **83,4 %** de l'information — ce qui est important. Le passage à PC3 n'ajoute que 12 %.



L'axe PC1 est un axe de visibilité bibliométrique. On y trouve à gauche les grands auteurs, et de grands spécialistes des transports comme Floridablanca, Brissson, Lalanne, Enfantin, Freycinet, sont pratiquement des bases scientifiques modernes. Autrement dit il reflète probablement directement le référencement a posteriori des auteurs.

L'axe PC2 reflète directement la « divergence entre bases ». Cet axe sépare les trois indicateurs OpenAlex (OA works, OA cit., OA h) et BnF (−0,10) des indicateurs Semantic Scholar (S² cit. et S² h). En ce sens il opposerait une reconnaissance historique d'une reconnaissance plus récente. Il s'agit donc d'un « biais de couverture entre bases ».

Ainsi Paul Boiteau et Gustav von Schmoller, sont des auteurs aux opinions très différentes. Boiteau s'opposa au Plan Freycinet, quand Von Schmoller théorise le concept de mercantilisme d'État appliqué aux transports. Johann Heinrich Gottlob von Justi (proche sur le plan des deux autres) se retrouve sur les mêmes thèses que Schmoller.

A l'opposé on trouve des auteurs comme Marshall, von Thunen ou encore Walras.

Cette analyse confirme largement l'opposition de nature des « réputations » des économistes ayant traité de transport, y compris dans les bases bibliométriques modernes.

