



HAL
open science

Les sommets du palais. Analyser l'espace parlementaire avec des cartes auto-organisatrices

Julien Boelaert, Etienne Ollion

► To cite this version:

Julien Boelaert, Etienne Ollion. Les sommets du palais. Analyser l'espace parlementaire avec des cartes auto-organisatrices. *Revue Française de Science Politique*, 2020, 10.3917/rfsp.703.0373. hal-02930181

HAL Id: hal-02930181

<https://hal.science/hal-02930181>

Submitted on 4 Sep 2020

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.

LES SOMMETS DU PALAIS

Analyser l'espace parlementaire avec des cartes auto-organisatrices

Julien Boelaert , Étienne Ollion

À paraître dans la *Revue Française de Science Politique* en septembre 2020

[Version avant épreuves]

Résumé : Les élections législatives de 2017 ont été marquées par un renouvellement inédit sous la Vème République, avec l'arrivée d'élus aux profils rares. Comment se sont-ils adaptés à cet espace très codifié ? La réponse à cette question est difficile, car l'activité parlementaire a des facettes multiples que les méthodes classiques peinent à restituer ensemble. Afin de décrire ces formes d'engagement parlementaire relationnellement et dans leur diversité, cet article introduit une méthode de réduction de dimensionnalité : les cartes auto-organisatrices (*Self-Organizing Maps*, ou SOM). Il en expose les principes et en démontre l'intérêt en pratique, à partir du cas de l'engagement parlementaire des élus de 2017. Une comparaison avec l'analyse géométrique de données souligne l'intérêt des SOM, qui permet une analyse fine des types de *praxis* parlementaire. Plus généralement, l'article montre les SOM peuvent être un outil statistique utile, que ce soit pour réduire et visualiser une masse de données, ou pour représenter un espace social.

Julien Boelaert est maître de conférence en science politique à l'université de Lille (CERAPS). Ses travaux portent sur les transformations du personnel politique en France, et sur les méthodes numériques et quantitatives en sciences sociales. Il est l'auteur (avec S. Michon et É. Ollion) de *Métier : député. Enquête sur la professionnalisation de la politique*, Raisons d'agir, 2017 et (avec É. Ollion) de « The Great Regression. Machine Learning, Econometrics and the Future of Quantitative Social Sciences », *Revue française de sociologie*, 2018. julien.boelaert@univ-lille.fr

Étienne Ollion est chargé de recherche CNRS (CREST-ENSAE), et professeur associé à l'École polytechnique. Ses recherches portent sur le champ politique, ses acteurs et ses transformations, ainsi que sur les méthodes numériques. Il est l'auteur de plusieurs ouvrages, dont *Raison d'État. Histoire de la lutte contre les sectes en France* (La découverte, 2017). eollion@ensae.fr.

LES SOMMETS DU PALAIS¹

Analyser l'espace parlementaire avec des cartes auto-organisatrices

Il est un peu plus de 13 h, ce 20 juin 2017, à l'Assemblée nationale². En ce jour de rentrée, les députés élus ou réélus quelques jours auparavant viennent procéder aux formalités d'usage. Successivement, ils passent par les diverses étapes de l'enregistrement. D'un bureau à l'autre, ils déclinent leur identité et posent pour la photo officielle qui ornera leur page internet. Ils se voient remettre les identifiants de leurs différents comptes, collectent les insignes du parlementaire (l'écharpe tricolore, la cocarde pour la voiture, ou encore le « baromètre »).

Suite à l'élection d'Emmanuel Macron le mois précédent, c'est une législature profondément renouvelée qui prend possession des lieux ce jour. Avec 72 % de primos-députés (contre environ un tiers d'habitude), 28 % de parlementaires jamais élus à quelque mandat politique auparavant (contre 5 % lors de la mandature précédente), l'élection a été l'occasion d'un grand renouvellement des visages. Elle a non seulement vu nombre de responsables politiques historiques être éliminés, mais ils ont parfois été remplacés par de parfaits inconnus en politique. Plus qu'une alternance, la quinzième législature est marquée par le nombre record de novices dans ses rangs. Près d'un cinquième des nouveaux élus ont en effet une expérience politique nulle ou limitée.

Dans les couloirs du palais, ces tous nouveaux venus en politique sont aussi bien visibles. Amusés, émus, impressionnés, ils passent leur temps à prendre des photos, seuls ou en petits groupes. Ils sont aussi plus hésitants. L'un des auteurs de ce texte, présent dans la salle des Quatre-Colonnes, cette interface organisée entre le monde politique et celui des médias au sein de l'Assemblée, a parfois servi de guide improvisé. Une primo-députée sollicitée par un journaliste s'est aussi tournée vers lui pour demander « si elle a[vait] le droit de répondre ». Le tableau d'ensemble rappelle plus une visite organisée qu'une rentrée de parlementaires. Le contraste est alors d'autant plus net avec les autres parlementaires, plus habitués aux us du milieu, et clairement moins timides. Car au même moment, les plus rodés à la politique nationale sont déjà au travail.

¹En France, on ne dit pas : « cette recherche a été financée sur des fonds publics pérennes attribués équitablement entre chercheurs », mais « ce travail a été soutenu par l'Idex de l'Université de Strasbourg et le programme 'investissements d'avenir' » du gouvernement français » (et c'est regrettable).

²Extrait de journal de terrain, 20 juin 2017.

À deux pas, un petit groupe de trentenaires s'active au milieu des reporters. Dans ce moment de haute attention médiatique où se concentrent les journalistes, ils distribuent bons mots et cartes de visite, bien conscients de l'intérêt de figurer dans le carnet d'adresse soudainement dépeuplé des journalistes parlementaires³.

La scène cependant ressemble trop aux images d'Épinal pour qu'on ne s'interroge sur sa représentativité réelle, ou sur sa durée dans le temps. Passés ces premiers moments de découverte, les novices ne se sont-ils pas conformés à la réalité du travail parlementaire, comme leurs collègues avant eux ? Surtout, ce thème de la découverte plus ou moins aisée du champ politique ressemble à un trope commode pour commentateur pressé, un peu à l'image de cet angle sur les « bourdes » des nouveaux, aussi largement utilisé par les médias dans les premiers mois qu'il a rapidement disparu après.

Comment les nouveaux élus de 2017, portés par le succès du parti présidentiel, se sont-ils acclimatés au travail parlementaire ? Et les novices, célébrés lors de la campagne pour leur pouvoir transformateur sur l'institution et sur la politique en général, ne sont-ils finalement pas parvenus à accéder à des positions de pouvoir et à en faire bouger les lignes ? Apporter une réponse précise à cette question est moins simple qu'il n'y paraît. Les études de cas, aussi riches soient-elles, ne permettent pas d'avoir une vision synoptique. Et du point de vue quantitatif, la question de ce qui définirait le succès est complexe. Du fait du faible nombre de postes à responsabilité à l'Assemblée (une vingtaine au maximum⁴), s'intéresser aux positions institutionnelles auxquelles certains peuvent accéder fournit un tableau trop restrictif. Surtout, il existe des manières très différentes d'être parlementaire⁵, qu'une simple échelle unidimensionnelle ne saurait restituer. La remarque est devenue classique avec la critique récurrente des classements destinés à mesurer « l'activité » des parlementaires : les différentes facettes du métier de parlementaire ne peuvent être saisies de façon simplement linéaire.

Afin de pallier ces difficultés, cet article propose une approche nouvelle pour aborder cette

³L'opération est visiblement couronnée de succès, puisque dans les semaines suivantes, je les verrai régulièrement interviewés. Dans le courant de l'année suivante, deux des quatre membres de ce petit groupe se verront proposer une position de premier plan.

⁴Conventionnellement : présidence et vice-présidences, présidences de groupe, présidence de commission, questure, auxquelles on peut rajouter les quelques délégations.

⁵Comme ont pu le montrer les travaux sur la notion de « rôle parlementaire ». Voir Donald Searing, *Westminster's World. Understanding Political Roles*, Boston, Harvard University Press, 1994 et Olivier Rozenberg et Cécile Vigour, « Comment pensent les parlementaires ? Les rôles, valeurs et représentations des élus », in Olivier Rozenberg et Eric Thiers, *Traité d'études parlementaires*, Bruxelles, Larcier, 2018.

question. D'une part, il multiplie les sources de données, afin de proposer une approche plus riche du travail des députés (voir encadré). La principale innovation proposée dans cet article est toutefois méthodologique. La multiplicité des formes d'engagement des parlementaires évoquée ci-dessus invite en effet à ne pas restreindre l'étude à un seul indicateur. Pour rendre compte de la multiplicité des formes d'engagement tout en restituant les hiérarchies qui structurent l'espace parlementaire, il faut recourir à une méthode capable de représenter le parlement comme un espace multidimensionnel. Mais plutôt qu'une approche de type analyse géométrique de données (AGD⁶), on recourt ici à une technique statistique issue du *machine learning* : les cartes auto-organisatrices (ou *self-organizing maps*, SOM). Les SOM permettent en effet de mettre en évidence plusieurs traits saillants de l'activité parlementaire que les méthodes d'AGD laissent dans l'ombre.

À partir d'une étude de diverses activités menées par l'ensemble des élus lors de la première année de mandat, cet article montre comment se construit et s'interprète une telle « carte ». Il soutient qu'elle permet souvent une représentation plus réaliste que celles fournies par l'AGD. Elle conserve l'aspect multidimensionnel et hiérarchisé qui a fait la force de cette dernière méthode, mais elle permet aussi une attention fine aux sous-espaces de l'activité parlementaire et aux petites variations potentiellement significatives. Comparaison en pratique, l'article ne souligne pas moins les limites des SOM, qu'elles partagent avec de nombreuses autres approches d'exploration issues du *machine learning*.

L'article procède en quatre temps : il revient d'abord sur les débats autour de la mesure de l'activité parlementaire pour mieux présenter son approche et les données retenues. Il propose ensuite une rapide présentation théorique des cartes auto-organisatrices. La méthode est alors appliquée aux données parlementaires, et l'article expose en pratique l'analyse d'une carte auto-organisatrice. Il compare brièvement cette méthode à l'ACP (analyse en composantes principales), méthode standard de réduction de dimensionnalité. Enfin, on retourne à la question initiale : en localisant sur la carte les novices et les autres élus, on voit bien la position reléguée qu'ont occupé ceux-ci au cours de la première année.

La difficile mesure de l'activité parlementaire

La question de la quantification du travail des élus est ancienne. Dans les années 1970 déjà,

⁶Brigitte Le Roux et Henri Rouanet, *Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis*, London, Springer, 2004.

un *Institut pour la Mesure de l'Activité Parlementaire* (IMAP) proposait des classements de députés en fonction de leur activité. Déjà ses indicateurs étaient objet de discussion. Depuis, la forte publicisation d'indicateurs d'activité via différents sites internet a rendu le sujet plus sensible encore.

D'une mesure de l'activité aux formes de l'investissement parlementaire

La question de la quantification du travail des élus est ancienne. Dans les années 1970 déjà, un *Institut pour la Mesure de l'Activité Parlementaire* (IMAP) proposait des classements de députés en fonction de leur activité. Déjà ses indicateurs étaient objet de discussion. Depuis, la forte publicisation d'indicateurs d'activité via différents sites internet a rendu le sujet plus sensible encore.

Les limites d'une approche quantifiée sont bien connues. La première est que les indicateurs habituellement mobilisés sont insuffisants pour restituer le travail des députés. Le nombre de présence en commission, le nombre de rapports produits ou d'amendements déposés sont trop partiels pour bien restituer les nombreuses facettes du travail des élus. De fait, même si on ne considère que la définition proprement constitutionnelle du rôle du parlement selon lequel celui-ci « vote la loi [...], contrôle l'action du Gouvernement [...], évalue les politiques publiques » (article 24), il est difficile de trouver des indicateurs pertinents pour chacune de ces activités. Le travail est en effet collectif, partagé et souvent délégué à une personne au nom du groupe. Par ailleurs, toute une partie des activités parlementaires a lieu en circonscription et est quasiment invisible dans les statistiques que l'on peut collecter via l'Assemblée. Or l'activité de circonscription, qui comprend un travail de représentation, la rencontre avec différents acteurs du territoire ou avec les administrés constitue une fraction parfois majoritaire du temps passé à l'exercice du mandat⁷. De même, l'écriture d'un rapport peut prendre plusieurs mois, quand le dépôt d'un amendement peut se faire en quelques secondes. L'agrégation irréfléchie des différents actes ne peut donc pas refléter l'investissement effectif des députés.

Une autre difficulté moins souvent évoquée mais pas moins réelle est que les élus n'ont

⁷Voir par exemple : Patrick Le Lidec, « Les députés, leurs assistants et les usages du crédit collaborateurs. Une sociologie du travail politique », *Sociologie du Travail*, 50 (2), 2008 ; Corentin Poyet et Sven T. Siefken, « Show or substance? The exchange between district and parliamentary activities », in Gabriel, Oscar, Kerrouche, Eric et Suzanne S. Schüttemeyer. *Political Representation in France and Germany: Activities of MPs and Views of Citizens*, London, Palgrave, 2018.

pas les mêmes opportunités pour réaliser les différentes activités qui sont mesurables. Les exemples sont nombreux : la majorité a, de droit, accès aux positions institutionnelles (la présidence, la plupart des vice-présidences et des présidences de commission, la questure, etc). Les députés du groupe majoritaire n'ont en revanche qu'un droit de tirage limité sur certaines pratiques (telles les questions au gouvernement, moment d'exposition médiatique recherché), qu'il faut encore partager entre un nombre de personnes parfois pléthorique. De même, les commissions permanentes ne se réunissent ni ne travaillent au même rythme, ce qui biaise en faveur des membres des commissions les plus actives tous les indicateurs qui s'appuient sur ces mesures.

Une dernière complication du point de vue de la recherche est que la publicisation de ces indicateurs chiffrés a pu avoir des effets sur la pratique des élus. C'était d'ailleurs la critique des questeurs de l'Assemblée face au site www.nosdeputes.fr de l'association *Regards Citoyens* : l'existence de mesures chiffrées du travail parlementaire pervertirait l'activité parlementaire en incitant à « faire du chiffre ». Les parlementaires déposeraient des amendements et interviendraient à tout-va pour faire augmenter leurs statistiques. À l'Assemblée comme dans tous les lieux où le benchmarking s'est développé, la loi de Goodhart fonctionne à plein : « quand un indicateur devient un objectif il cesse d'être un indicateur pertinent »⁸.

Si l'affirmation est rejetée par *Regards citoyens*, qui rappelle que l'inflation législative a commencé avant sa création en 2008, on trouve peu d'éléments pour réfuter les effets de ces décomptes à l'Assemblée. Chaque député rencontré au cours des années de terrain affirme y prêter attention. En entretien, les plus réservés déclarent « faire ce qu'il faut » pour ne pas mal figurer sur le site, quand d'autres reconnaissent clairement déposer amendements ou questions afin de rester dans le haut des classements. Leurs collaborateurs parlementaires ont en général moins de pudeur et racontent volontiers (ou montrent à l'enquêteur) leurs stratégies pour multiplier amendements et questions. L'argumentaire mobilisé en permanence pour se justifier est celui de la nécessité. Tous déclarent que les élus sont régulièrement confrontés à ces classements, le plus souvent via la presse locale, et que le stigmate d'un mauvais rang est trop lourd à porter.

Les remarques précédentes invitent à la prudence quant à la mesure de l'activité des élus. Il semble en effet impossible d'obtenir une mesure individuelle précise de cette dernière, du moins si par là on entend un indicateur synthétique. Mais si on aborde la question d'un point de vue global, et surtout si l'on parvient à restituer la diversité des formes d'engagement, une telle mesure

⁸Voir : Isabelle Bruno, Emmanuel Didier, *Benchmarking. L'État sous pression statistique*, Paris, La Découverte, 2013.

peut avoir un sens. À condition d'opérer un travail de correction des données (voir ci-dessous), on peut alors dresser un tableau des modalités d'investissement des parlementaires. Après tout, le fait que l'accès à certaines positions est inégal fait partie du quotidien des élus. Les députés exercent en effet leur activité sous diverses contraintes, dont celle de leur poids politique ou de la taille de leur groupe. De même, le fait que certains consacrent une partie de leur temps à « faire de la présence » ou « du chiffre » pendant que d'autres sont peu investis est révélateur de l'usage différencié qu'ils font de leur mandat⁹.

Les données de l'enquête : un an d'activités parlementaires

Pour saisir ces investissements différenciés, on s'est concentré sur la première année de la XVe législature (20 juin 2017 - 3 août 2018). On a alors collecté trois types de variables, qui chacun renvoie à une forme d'engagement dans l'activité parlementaire mise en avant dans la littérature (voir tableau 1)¹⁰. Le premier regroupe une série de variables qui doivent permettre de saisir la contribution à la production législative et au contrôle du gouvernement. Un autre ensemble cherche à saisir la visibilité médiatique des élus. Pour ce faire, on a collecté (via une procédure automatique) les invitations des députés dans différents médias audiovisuels et leurs mentions dans les quotidiens nationaux et locaux¹¹. Des informations relatives au profil Twitter des élus ont aussi été collectées, via l'interface de ce site. Enfin, une série d'activités parlementaires autres ont été intégrées, et ce afin de saisir autant que possible l'investissement en faveur de la circonscription, ou l'usage du mandat à des fins politiques de construction de carrière.

Au parlement, on construit en effet des carrières politiques, et toutes les activités ne sauraient être analysées sans saisir le travail de réunion et de création de coalition, ni sans voir le rôle crucial des médias dans la production et l'entretien d'un capital politique. De même, le nombre de salles réservées pour des fins personnelles donne à voir un usage particulier de l'Assemblée : que ce soit à des fins de création de coalitions (par exemple pour faire exister un courant), ou de spécialisation (organisation de colloques, rencontres avec des spécialistes), elle révèle un usage

⁹Ce faisant, on ne fait rien d'autre que de chercher à distinguer un ensemble de « rôles », tels que les a proposés Donald Searing dans *Westminster's World*, Cambridge, Harvard University Press, 1994. Voir aussi Olivier Rozenberg, « Le Parlement français et l'Union européenne (1993-2005) : l'Europe saisie par les rôles parlementaires », thèse pour l'obtention du doctorat de science politique, Institut d'études politiques de Paris, 2005.

¹⁰Voir les travaux cités précédemment, ainsi que ceux sur les « styles parlementaires » dans la lignée de l'ouvrage de Richard Fenno, *Home Styles. Representatives and their Districts*, Boston, Little and Brown, 1978.

¹¹Voir annexe 1 pour plus de détails.

non codifié, mais très fréquent, du mandat de parlementaire¹².

Comme indiqué précédemment, ces données ont été normalisées afin de prendre en compte les inégales chances d'accès aux activités. Les interventions et les productions des commissions ont été centrées et réduites par commission afin de ne pas favoriser indûment les membres des commissions les plus actives (Lois, Finances) par rapport à celles qui le sont moins (Affaires étrangères, Défense). De même, les questions au gouvernement ont été centrées et réduites par groupe politique.

Activités législatives	Rapports parlementaires ; Présences en commission ; Interventions en commission ; Interventions en hémicycle ; Nombre de votes ; Nombre d'amendements déposés
Visibilité médiatique	Invitations : chaînes d'information en continu ; Invitations : chaînes spécialisées (LCP, Public Sénat) ; Invitations : radios et télévisions nationales ; Mentions dans la presse nationale ; Mentions dans la presse locale ; Nombre de followers sur Twitter (logarithme)
Autres activités parlementaires	Responsabilités institutionnelles ¹³ ; Questions au gouvernement ; Questions écrites : nombre ; Questions écrites : diversité thématique ; Nombre total de salles réservées ; Nombre de salles distinctes réservées.

Tableau 1: Liste des variables actives retenues

Les cartes auto-organisatrices

Saisir ensemble les multiples façons d'investir le travail parlementaire, et donc les différents rôles de députés dans leur complexité, demande de prendre en compte simultanément

¹²Un aspect du travail parlementaire échappe assez largement à notre approche : l'investissement en circonscription. On verra toutefois qu'il est indirectement mesuré par d'autres indicateurs.

¹³Voir détail en annexe.

l'ensemble des indicateurs d'activité que nous avons retenus. Afin de rendre lisible cette masse de données, on recourt à une méthode statistique de réduction de dimensionnalité, les cartes auto-organisatrices, ou SOM (pour *self-organizing map*)¹⁴.

Les SOM, un algorithme ancien d'apprentissage statistiques

Les cartes auto-organisatrices font partie d'un ensemble plus large de méthodes regroupées aujourd'hui sous le terme d'apprentissage statistique (*machine learning*), qui ont en commun d'adopter une approche algorithmique pour effectuer des tâches classiques de la statistique. Par contraste avec la statistique mathématique¹⁵, les méthodes d'apprentissage ne posent que peu d'hypothèses sur la structure des données, et troquent l'exactitude mathématique des démonstrations classiques contre des heuristiques de calcul dont le principal mérite est de « fonctionner » en pratique¹⁶. Plus récentes (on date habituellement les premiers pas de l'intelligence artificielle à la fin des années 1950), elles ont également la particularité d'être plus gourmandes en capacités de calcul.

L'algorithme qu'on présente ici est un type de réseaux de neurones artificiels, relativement ancien puisque son invention remonte au début des années 1980¹⁷. Employé comme un moyen de simplification et de visualisation de jeux de données complexes, il a été utilisé dans de nombreux domaines théoriques et appliqués¹⁸, mais rarement dans les sciences sociales. Une particularité saillante des cartes auto-organisatrices est qu'elles fonctionnent à la fois comme un outil de clustering (une partition de l'espace de départ, comme le ferait une classification hiérarchique) et de réduction de dimensionnalité et de visualisation (comme l'analyse en composantes principales). Les cartes que produit l'algorithme se présentent sous la forme d'un ensemble de cases placées les

¹⁴Ou encore « carte de Kohonen », du nom de leur inventeur.

¹⁵Par statistique mathématique, ou encore statistique classique, on désigne ici l'ensemble des méthodes qui reposent sur une démonstration mathématique rigoureuse, et dont les propriétés statistiques (pour un estimateur : le biais, la variance, la convergence, l'efficacité...) sont connues d'après les hypothèses prises sur les données et les mécaniques du calcul probabiliste. Les différences entre cette approche nouvellement baptisée classique et l'apprentissage automatique sont parfois exagérées, et la frontière entre les deux familles de méthodes est en réalité assez floue. Pour autant, la distinction est utile pour mettre en avant les particularités du *machine learning*.

¹⁶Les outils de la statistique classique se présentent généralement sous la forme d'un problème mathématique dont on sait calculer la solution exacte à partir des données (les coefficients d'une régression linéaire, les matrices de rotation d'une ACP, etc.). En apprentissage statistique, les modèles mathématiques sont généralement beaucoup plus complexes, et les solutions exactes hors d'atteinte. C'est là qu'interviennent les heuristiques de calcul, qui sont des astuces mathématiques pour tenter d'approcher une hypothétique solution optimale.

¹⁷Teuvo Kohonen, *Self-organizing maps*, Londres, Springer, 1997.

¹⁸Dubravko Miljković, « Brief Review of Self-Organizing Maps », *MIPRO*, 2017.

unes à côté des autres. Chacune de ces cases représente une petite partie de l'espace des données (ici l'espace des activités parlementaires), et contient les observations (les députés) qui se trouvent dans cette partie de l'espace.

Par construction, chaque case contient donc un groupe d'élus aux pratiques semblables. L'objectif de l'algorithme est alors d'organiser la carte afin que des cases proches (sur la carte) correspondent à des groupes d'élus dont les pratiques sont proches (dans l'espace initial des pratiques). Cela signifie qu'on doit trouver des députés aux profils relativement proches dans les cases voisines de la carte, et des députés aux profils différents dans les cases éloignées. Une fois la SOM calculée, elle permet une représentation graphique simplifiée du jeu de données, qui bien qu'aplanie sur deux dimensions en conserve les propriétés topologiques.

Heuristique et évaluation de l'algorithme

Une carte auto-organisatrice se présente abstraitement sous la forme d'un réseau, c'est-à-dire d'un ensemble de sommets reliés entre eux par des arêtes. Tous les sommets ne sont pas reliés entre eux : un SOM ressemble généralement à une grille en deux dimensions, semblable à un filet déplié, où chaque sommet n'est relié qu'à un petit nombre d'autres, qu'on appelle ses « voisins ». Cette notion de voisinage est importante, puisque l'objectif principal de l'algorithme sera de faire correspondre les distances entre sommets sur la carte (deux sommets reliés directement ont une distance de 1, deux sommets reliés indirectement par un troisième ont une distance de 2, etc.) à des distances dans l'espace des données¹⁹. Le nombre de nœuds du filet (sommets du réseau) ainsi que leur disposition sont choisis dès le départ par l'utilisatrice. C'est elle qui décide de la structure du réseau, en fonction d'une série de critères (nombre d'individus, distribution des données, niveau de granularité des résultats souhaité).

Le lien entre l'espace abstrait de la carte et celui, concret, des données que l'on souhaite résumer, est fait au moyen de « prototypes ». À chaque cellule de la carte correspond en effet par construction un prototype, un point artificiel dans l'espace des données de départ. Dans notre cas, chaque case de la carte correspond ainsi à un député fictif, représenté par une combinaison particulière d'activités parlementaires. La finalité de l'algorithme SOM est de placer ces

¹⁹Pour une description plus rigoureuse de l'algorithme voir Trevor Hastie, Robert Tibshirani et Jerome Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, New York, Springer, 2009 (chapitre 14.4).

prototypes au milieu de poches d'individus semblables entre eux, tout en faisant en sorte que des cellules voisines sur la carte d'origine soient aussi voisines dans l'espace analysé. Les prototypes sont initialement placés à des positions arbitraires dans l'espace, et sont ensuite déplacés petit à petit pour améliorer l'organisation de la carte²⁰. La structure de la carte (le nombre et la disposition des cellules) ne change pas.

L'algorithme procède de façon itérative, en répétant un grand nombre de fois une suite d'opérations relativement simples. À chaque itération, une observation (un député) est tirée au sort. L'algorithme calcule alors la distance (euclidienne) de cette observation à chacun des prototypes de la carte, et associe l'observation à la cellule dont le prototype est le plus proche. Ce prototype gagnant est alors déplacé, sur une petite distance \square , en direction de l'observation ; on déplace également en direction de l'observation, sur une distance plus petite, les prototypes des cellules voisines de la gagnante. La métaphore du filet prend ici tout son sens, si tant est que l'on considère que les mailles du filet sont élastiques : lorsque l'on tire un des nœuds du filet, on tire avec lui, un peu moins fort, les nœuds auquel il est relié.

Au fil des itérations, on réduit progressivement²¹ la distance de déplacement \square , ainsi que le nombre de voisins du gagnant qui seront eux aussi déplacés. Ainsi, au cours des premières itérations les prototypes gagnants ainsi que de nombreux prototypes voisins sont déplacés sur de longues distances, de sorte que le filet se déplie rapidement pour couvrir grossièrement l'ensemble des données. Les dernières itérations correspondent au contraire à un moment d'ajustement, où seul le prototype gagnant est déplacé, et ce sur une très petite distance.

À la fin de ce processus d'apprentissage, chaque observation est assignée à la cellule de son prototype gagnant. On obtient donc une classification des observations, mais également -- c'est tout l'intérêt des SOM -- une représentation spatiale des différentes classes, puisque chaque groupe d'observations prend une place sur le canevas prédéfini de la carte. C'est cette organisation spatiale qui rend possible les nombreuses représentations graphiques d'interprétation des SOM, qui aident à comprendre la structure du jeu de données. Elle se traduit par la forme en mosaïque que prennent

²⁰De ce point de vue, le fonctionnement des SOM est similaire à celui des algorithmes de classification par « nuée dynamique » (*k-means*), qui passent eux aussi par la création et l'ajustement itératif de prototypes pour représenter les centres de classes. L'originalité des SOM consiste en ce que chaque prototype est attaché à une cellule sur la carte en deux dimensions, et que des cellules proches sur la carte doivent correspondre à des prototypes proches dans l'espace. On peut donc voir les SOM comme une variante contrainte des nuées dynamiques, qui permet d'opérer une réduction de dimensionnalité en même temps qu'une classification.

²¹Le rythme de décroissance d' \square est défini de façon à garantir que l'algorithme converge vers une solution stable.

généralement ces graphiques, par juxtaposition de petits graphiques qui synthétisent chacun les caractéristiques de la sous-population associée à une cellule.

La qualité de la carte obtenue est finalement évaluée via différentes métriques. L'une consiste, assez classiquement pour une méthode de classification, à mesurer la quantité d'information perdue en résumant le jeu de données aux seuls prototypes de la carte (*i.e.* la distance moyenne des observations à leur prototype gagnant) ; c'est l'erreur dite de « quantification », qui permet également de calculer la proportion de variance conservée par le *clustering*. Informative, cette mesure ne concerne que la qualité de la classification, et non la qualité de la correspondance topologique entre le jeu de données initial et sa représentation sous forme de carte. On mesure à cette fin l'erreur « topographique », définie comme la proportion d'observations dont le deuxième prototype le plus proche ne correspond pas à un nœud directement voisin du nœud gagnant. Enfin, l'erreur dite de Kaski-Lagus²² combine de manière astucieuse en une seule métrique ces deux dimensions de l'approximation que constitue la carte par rapport au jeu de données de départ.

Une fois les individus placés dans les cellules, des regroupements sont possibles. En particulier, il est courant d'avoir recours à un algorithme de clustering (sur la matrice de distance entre cellules) afin d'isoler des groupes. Ces grandes zones de la carte sont appelées des super-classes. Elles permettent une meilleure lisibilité de la carte, en particulier lorsque le nombre de cellules est important. Elles reflètent bien la philosophie de l'algorithme SOM, qui plutôt qu'en axes comme dans le cas de l'AGD invite à un raisonnement sous forme de zones, hiérarchisées par des dimensions plus générales, mais pas forcément linéaires. Ces possibilités sont illustrées sur le cas concret des pratiques des élus de la XV^{ème} législature dans les pages qui suivent.

Enfin, si l'algorithme original de Kohonen était conçu uniquement pour traiter des données numériques, pour lesquelles il est possible de calculer une distance euclidienne entre individus²³, de nombreuses extensions ont vu le jour depuis qui permettent par exemple de traiter des données catégorielles²⁴, ou même des données relationnelles, uniquement définies par une matrice de distances entre individus²⁵. Il est par ailleurs possible de transformer des données catégorielles en

²²Samuel Kaski, Krista Lagus, « Comparing Self-Organizing Maps », in *International Conference on Artificial Neural Networks*, 1996.

²³Comme en ACP, les choix initiaux de normalisation des variables, et éventuellement de leur pondération, vont donc déterminer l'importance relative qui sera donnée aux différentes variables dans le résultat final.

²⁴Cottrell *et al.*, 2003, *art. cit.*

²⁵Madalina Olteanu, Nathalie Villa-Vialaneix et Marie Cottrell, « On-Line Relational SOM for Dissimilarity Data », *Advances in Intelligent Systems and Computing*, 198, 2012.

données euclidiennes au moyen d'une analyse des correspondances multiples pour appliquer une SOM à l'ensemble des coordonnées produites.

Représenter l'espace parlementaire

Les SOM sont donc une méthode qui regroupe des individus sur la base d'une série de caractéristiques partagées. Le principal résultat est une carte en deux dimensions, faite d'un certain nombre de cellules. Chacune d'entre elles regroupe les individus les plus proches du « prototype » qui les a capturés.

L'interprétation des cartes auto-organisatrices

Conventionnellement, l'interprétation se fait de manière visuelle - en tout cas en première approche. On cherche alors à étudier la répartition des individus, et plus encore à déterminer les « pôles » ou « zones » de pratiques communes sur la carte. La figure 1 est une « carte de population », qui représente la répartition des individus au sein de chacune des cellules. Certaines cellules concentrent des dizaines d'individus, quand la cellule au coin nord-ouest n'en a que six, et que celles qui entourent cette dernière n'en ont au maximum qu'une petite douzaine. Sur cette figure, on a aussi donné à voir le sexe des individus, à l'aide de symboles. Le cadran nord-ouest, dont on verra qu'il est le pôle dominant de l'Assemblée, est très largement masculin.

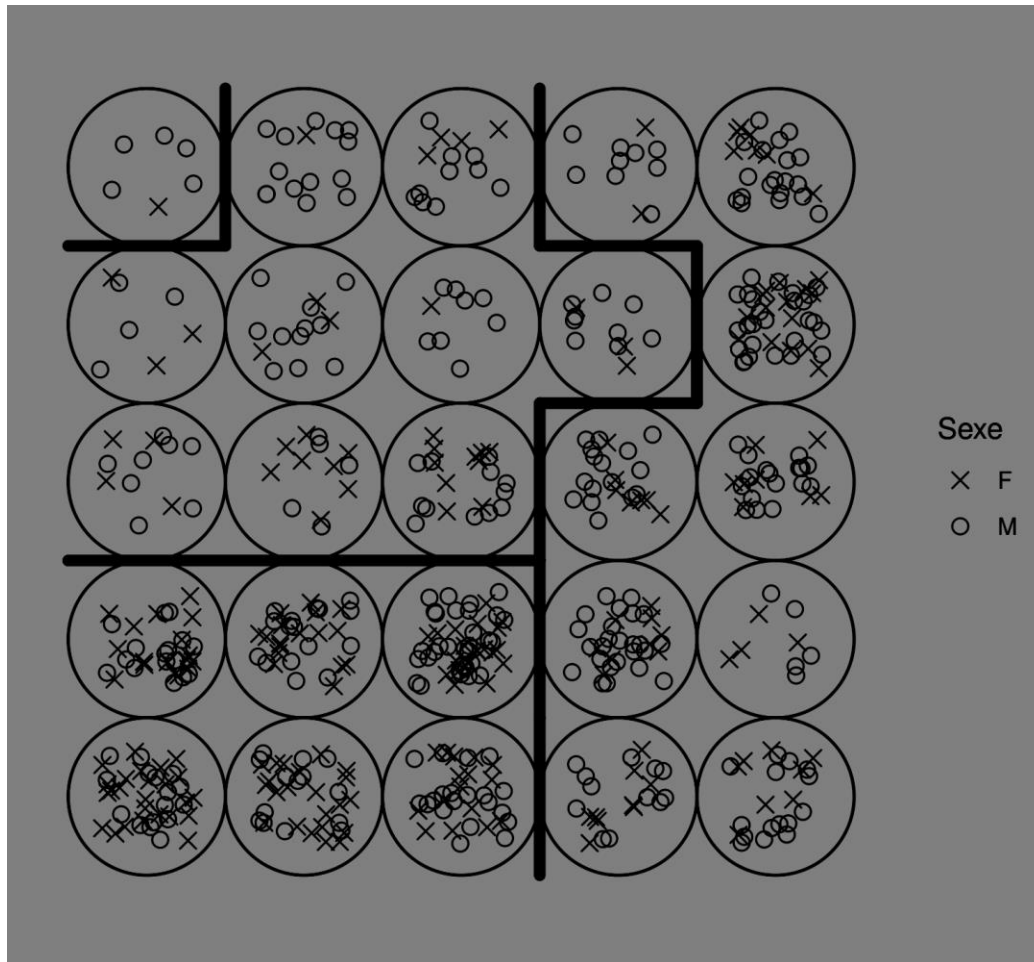


Figure 1 : Carte de population

La carte représente le nombre d'individus par cellule.

La position exacte des points à l'intérieur des cellules est aléatoire et ne doit pas être interprétée.

Avec les SOM, c'est sur ces individus ainsi regroupés que sont effectuées les interprétations. Au stade de l'exploration, il est assez fréquent de passer par l'une des nombreuses options de visualisation qu'offrent les logiciels qui implémentent les cartes auto-organisatrices²⁶. L'approche classique consiste à visualiser, variable par variable, l'intensité des pratiques, puis à partir de ces informations à produire une interprétation générale. Une représentation conventionnelle des cartes auto-organisatrice est alors un graphique « radar », où chaque variable

²⁶De nombreuses implémentations des SOM et de leurs extensions sont disponibles les principaux logiciels de traitement statistique et langages de programmation. L'ensemble des statistiques de cet article ont été produites sous R ; nous avons employé l'outil aweSOM pour l'exploration des cartes, et la bibliothèque kohonen pour la production des graphiques.

est représentée par un segment de disque dont la longueur représente l'intensité moyenne de la pratique dans chaque cellule (voir annexe). Ce graphique est utile pour offrir une vision synthétique des résultats, mais le détail est mieux présenté - et plus simple à publier - avec une carte de fréquentation qui donne à voir l'intensité des pratiques²⁷.

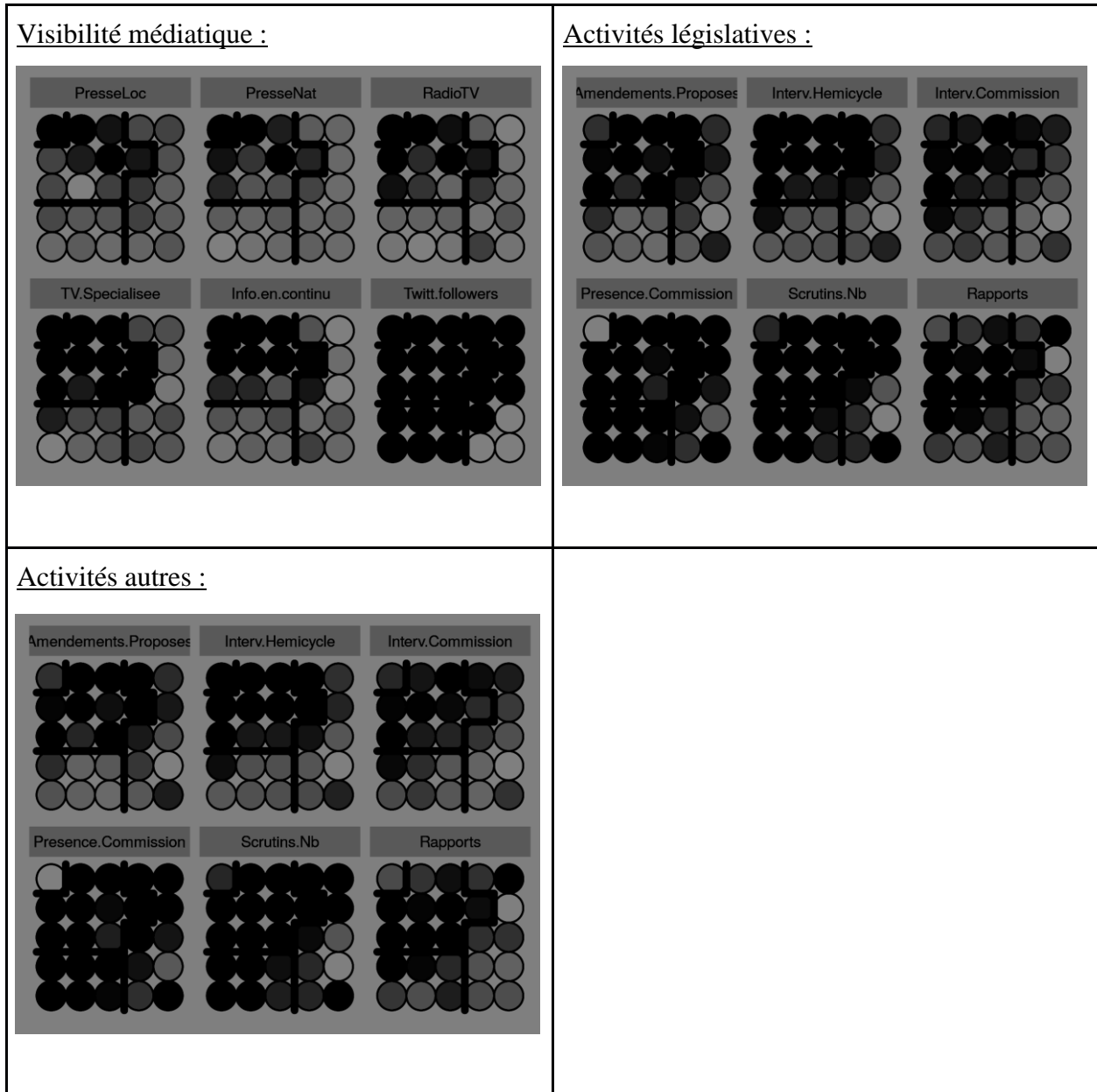


Figure 2: Cartes de fréquentation (*heatmap*)

²⁷Le noir indique une pratique moyenne intense pour la cellule, le blanc une pratique faible.

Valeur de chaque activité sur la carte de l'espace parlementaire (blanc pour une activité nulle, noir pour l'activité maximale).

L'analyse relative aux apparitions médiatiques est la plus évidente. Elle met clairement en évidence le fait qu'une petite minorité de parlementaires située dans le coin nord-ouest concentre la majorité des invitations et mentions, quand ceux situés dans le quart sud-est n'en ont presque aucune. Une ligne se dessine entre ces deux points, au long de laquelle la visibilité médiatique décroît. Dans le détail, des variations apparaissent. L'asymétrie entre élus est très prononcée pour la presse nationale (*PresseNat*), tout comme pour les radios et TV nationales (*RadioTv*). Dans un espace limité, seuls quelques-uns parviennent à exister.

La situation se reproduit, quoique de façon moins marquée, sur les télévisions spécialisées (*TV.Specialisee*) que sont *La Chaîne Parlementaire* et *Public Sénat*. On voit en effet une activité toujours concentrée dans le quadrant nord-est, mais qui déborde largement le seul coin de la carte. Ces deux chaînes donnent un accès à un plus grand nombre d'élus. Cette différence s'explique largement par leur mission : financées par les institutions parlementaires, l'une et l'autre ont pour but de donner à voir le travail des élus. Elles restent toutefois inégalitaires elles aussi, car une majorité des députés n'a pas été invité une seule fois dans leurs émissions au cours de cette première année de mandat.

Entre la presse nationale et les chaînes d'informations spécialisées se trouvent les chaînes d'information en continu (*Info.en.Continu*) que sont BFM, LCI, CNews, etc.. Cette position intermédiaire s'explique par leur grille de programmes : une part importante de leur temps d'antenne est dédiée à la politique nationale, le plus souvent via un débat sur l'actualité. Largement organisées autour du débat entre ces acteurs, elles sont fortement demandeuses et moins sélectives que les chaînes nationales hertziennes, ce qui explique cette situation d'entre-deux.

La presse locale, enfin, donne à voir un modèle légèrement différent. D'une part, les élus de premier plan, situés dans le quadrant nord-ouest, sont très souvent cités. La cellule qui se situe dans ce coin regroupe deux anciens candidats à l'élection présidentielle de 2017 (Marine Le Pen, Jean-Luc Mélenchon), un ancien Premier Ministre (Manuel Valls), le président du groupe majoritaire (Richard Ferrand), et celui du groupe Socialiste (Olivier Faure). Elle est d'ailleurs bien plus foncée que toutes les autres, tout comme le sont les cellules adjacentes, signe que la politique nationale domine les reportages qui traitent de parlementaires dans ces médias. Mais on voit aussi

des cellules légèrement grisées vers le centre de la carte, signe que certains élus dans ces points de l'espace parlementaire sont eux aussi régulièrement mentionnés. L'analyse des résultats détaillés révèle qu'il s'agit là de parlementaires pour la plupart déjà élus lors de la législature précédente. Souvent investis dans leur territoire d'élection où ils sont particulièrement actifs, ils sont régulièrement mentionnés par la presse de leur circonscription.

Du nord-ouest au sud-est se dessine ainsi une ligne qui va des élus les plus visibles à ceux qui le sont le moins, avec des spécificités entre eux selon les supports d'invitation. Le nombre de personnes abonnées au compte Twitter des élus confirme cette logique, avec une augmentation le long de cet axe, qui part des quelques élus sans compte pour relier les anciens candidats à la présidentielle.

D'autres lignes de partage apparaissent aussi sur cette carte. Si on s'intéresse cette fois aux activités législatives, on retrouve d'une part la distinction entre les élus particulièrement actifs (à gauche de la diagonale sud-est/nord-ouest) et d'autres qui le sont moins. Quelques élus sont même très peu actifs. On les trouve dans les quelques cellules qui se situent sur la frontière droite de la carte. Ces cellules ne sont presque jamais coloriées, pas même pour la présence aux votes, aux questions au gouvernement, ou en commission (même si la différence y est moindre, les absences répétées étant financièrement sanctionnées). Ces élus ne sont pas complètement inactifs, et on trouve dans cette zone des élus sur-investis dans la rédaction de question écrites. Comme on l'a dit précédemment, celles-ci sont assez peu valorisées car leur impact est considéré comme mineur, mais elles peuvent servir de signal pour démontrer son intérêt pour sa circonscription. Cette stratégie est parfois couronnée de succès, car ces élus peuvent être relativement cités par la presse locale, ce qui semble indiquer un fort investissement en direction de ses administrés.

Des cartes à la carte : produire une représentation stylisée

Afin de résumer les nombreuses informations présentées contenues dans les cartes individuelles, la figure 3 propose une représentation stylisée de l'espace d'activité parlementaire obtenu par les cartes auto-organisatrices. Elle décrit les différents formes d'engagement des élus lors de cette première année de mandat. La diagonale qui va du nord-ouest au sud-est indique l'opposition entre élus très visibles et ceux qui sont restés invisibles, que ce soit dans les divers médias ou au palais Bourbon. Les SOM restituent alors une dimension fondamentale de l'institution, à savoir la forte hiérarchie qui la traverse.

La proximité avec une représentation d'analyse géométrique de données ne doit toutefois pas faire oublier que la construction des SOM est différente (voir la comparaison ci-dessous). Une carte auto-organisatrice se lit en effet différemment - plutôt sous forme de zones de plus ou moins grande intensité de pratique. Elles peuvent être typifiées par des îlots, où l'intensité est maximale au centre et décroît à mesure qu'on s'en éloigne, ainsi que le montraient les *heatmaps* (cartes de fréquentation) ci-dessus.

En formalisant un peu plus, on peut en repérer quelques-uns, saillants. L'îlot « contribution législative » se centre sur les élus qui participent le plus au travail de la loi et à son vote. La consultation des noms (non représentée) montre qu'il est centré sur les présidents de commission. Ce n'est pas une surprise, puisqu'ils sont ceux qui organisent le travail législatif en commission, puis en hémicycle. Autour d'eux, on trouve les élus de la majorité, avec les rapporteurs de textes ayant donné lieu à un long travail, les responsables de textes plus petits, et enfin tous les élus LREM qui participent aux votes sans forcément intervenir autrement sur le texte.

D'autres élus situés sont eux aussi très actifs au parlement, quoique de manière différente. Ceux situés au nord de la carte sont ainsi particulièrement actifs en termes de pratiques législatives. Ils interviennent en hémicycle, ils déposent massivement des amendements. Avec les présidents de commissions, ce sont même eux qui interviennent le plus en commission. Cette forme d'engagement est toutefois différente, car comme on le verra plus bas, la plupart des individus dans cette zone sont des élus de l'opposition. Actifs au parlement, ils ont peu de chances de voir leur action traduite dans la loi. Par contre, ils utilisent leur présence à la chambre comme une caisse de résonance. Ils se servent de l'Assemblée comme d'une tribune, un lieu où il est possible de faire entendre sa position politique par les outils de procédure parlementaire.

Encore une autre modalité d'engagement, là encore souvent évoquée dans la littérature sur l'activité parlementaire, consiste pour certains élus à s'investir plus massivement dans leur circonscription. Elle est moins directement perceptible sur ces cartes, du fait de l'absence d'informations précises relatives à la circonscription. Deux indicateurs indirects permettent cependant de saisir des signes de cette pratique. Le fait d'être mentionné dans la presse locale pour des élus qui ne sont pas des ténors de l'Assemblée en est un. Le fait de déposer de très nombreuses questions écrites, sur des sujets très divers, en est un autre. Largement déconsidérées par les élus qui y voient une pratique invisible (les demandes sont de fait traitées par les services du ministère, pas par le cabinet), elles sont cependant un moyen de démontrer l'intérêt du parlementaire pour sa

circonscription. Une fois déposées, ces questions sont très souvent publicisées sur ses réseaux sociaux. Là encore, cette pratique est bien plus fréquente parmi les membres de l'opposition.

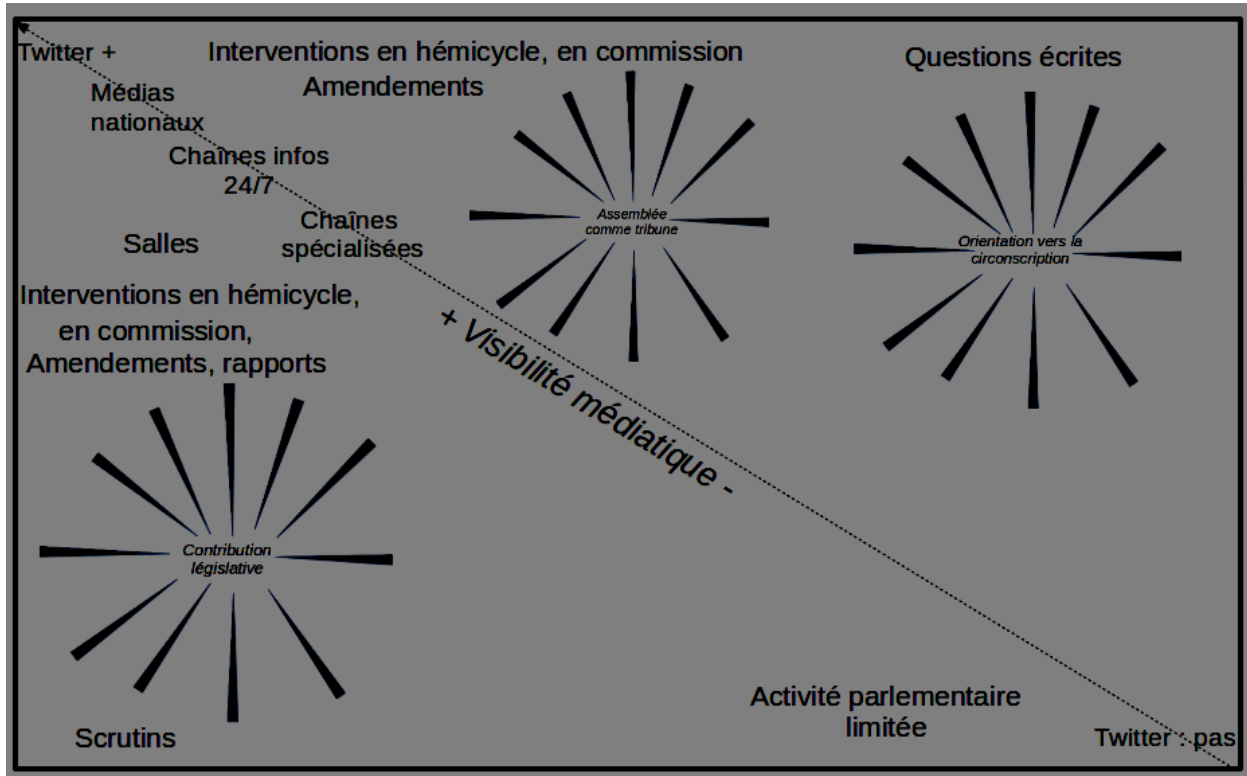


Figure 3 : Représentation stylisée des pratiques à l'Assemblée

Le graphique synthétise les informations obtenues dans la carte de fréquentation (figure 2). Les variables indiquées le sont à titre indicatif parmi celles qui sont largement sur-représentées.

SOM et analyse de géométrie de données

La réduction de dimensionnalité est un problème classique en sciences sociales, et les SOM ne sont pas, loin s'en faut, la méthode la plus classique pour le résoudre. Pourquoi alors y recourir plutôt, par exemple, qu'aux techniques plus courantes de l'analyse géométrique de données²⁸ que sont l'analyse en composantes principales, l'analyse des correspondances multiples, etc. ?

²⁸Brigitte Le Roux, *Analyse géométrique des données multidimensionnelles*, Paris, Dunod, 2016.

Le principal avantage de la réduction que proposent les SOM réside dans le fait qu'il s'agit d'une méthode non linéaire²⁹. En analyse géométrique de données, chaque composante est une combinaison linéaire des variables de départ, soit forcément une droite qui traverse l'espace de départ, et ce quelle que soit la forme véritable du nuage de points. Si cette forme est fortement non-linéaire, sa représentation peut nécessiter un grand nombre d'axes pour rendre compte de la complexité des données. Les SOM, en s'affranchissant de cette contrainte, permettent au contraire de représenter en une seule représentation graphique (deux dimensions) des données dont la structure peut être fortement non-linéaire³⁰.

Un avantage clair se dessine dans le cas de données à distribution asymétrique, c'est-à-dire dans lesquelles quelques individus prennent des valeurs extrêmes qui contrastent avec les valeurs moyennes de la masse de la population. Il suffit pour s'en convaincre de comparer les cartes auto-organisatrices avec l'analyse en composantes principales produite sur le même jeu de données et représentée par la figure 4a. L'axe 1 de cette ACP oppose, de façon prévisible, un grand nombre d'élus (à droite) d'un petit nombre d'autres (gauche). L'analyse des composantes montre que cet axe distingue entre individus fortement dotés par une sur-activité législative et médiatique, et les autres. Si cette opposition est bien capturée, et que l'axe 2 distingue lui ceux qui participent activement à l'activité législative et les autres, voilà à peu près tout ce qu'on peut tirer de cette représentation - surtout que l'axe 3 reste peu interprétable³¹.

²⁹Certaines méthodes liées à l'AGD, comme l'analyse des correspondances multiples ou autres formes d'*optimal scaling*, sont parfois qualifiées de non-linéaires, alors même qu'elles reposent sur des modèles mathématiques linéaires (voir Jacqueline Meulman, « Prediction and classification in nonlinear data analysis: Something old, something new, something borrowed, something blue », *Psychometrika*, vol. 68(4), pages 493-517, 2003). Elles consistent en effet à opérer une transformation non-linéaire des variables de départ avant de les traiter par les méthodes habituelles de factorisation de matrices. Par exemple, on peut traiter des variables continues par ACM (plutôt que par ACP) en les transformant d'abord en variables catégorielles, par quantiles ou autres ; le résultat est bien une réduction non-linéaire des données de départ. Si la frontière entre méthodes linéaires ou non est donc plus floue qu'il n'y paraît, les méthodes que nous qualifions ici de non-linéaires sont celles qui, comme les SOM, se départissent entièrement des calculs d'algèbre linéaire.

³⁰Il est à cet égard significatif que, dans certaines implémentations, l'initialisation des prototypes de la carte se fait en les disposant suivant une grille sur le premier plan factoriel d'une analyse en composantes principales. Ceci illustre l'ambition des SOM de donner une représentation des données plus précise que celle d'une analyse géométrique, puisqu'elle commence par celle-ci en vue de l'améliorer.

³¹La forme des points renvoie aux clusters que produisent les cartes auto-organisatrices, afin de montrer les correspondances entre les méthodes (et la plus grande granularité des SOM). La construction de ces clusters est évoquée plus bas.

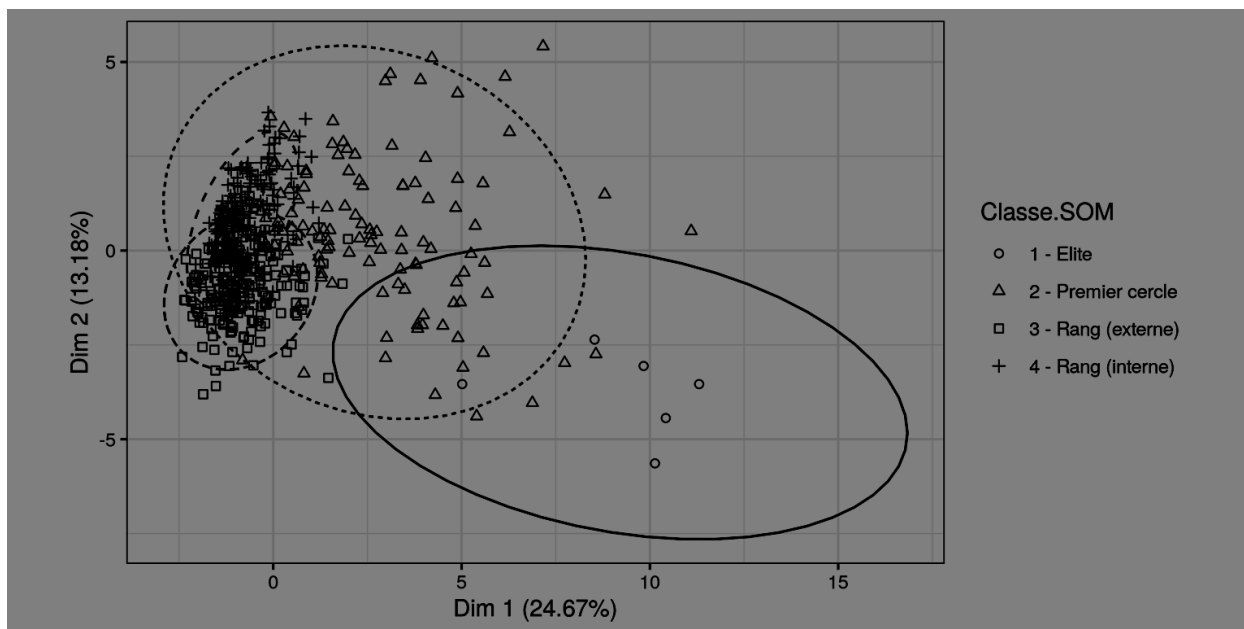


FIGURE 4a: Projection des députés sur le premier plan factoriel de l'ACP des données de pratiques parlementaires. Les ellipses (gaussiennes, seuil 95%) représentent les quatre classes déterminées construites à partir de la carte auto-organisatrice.

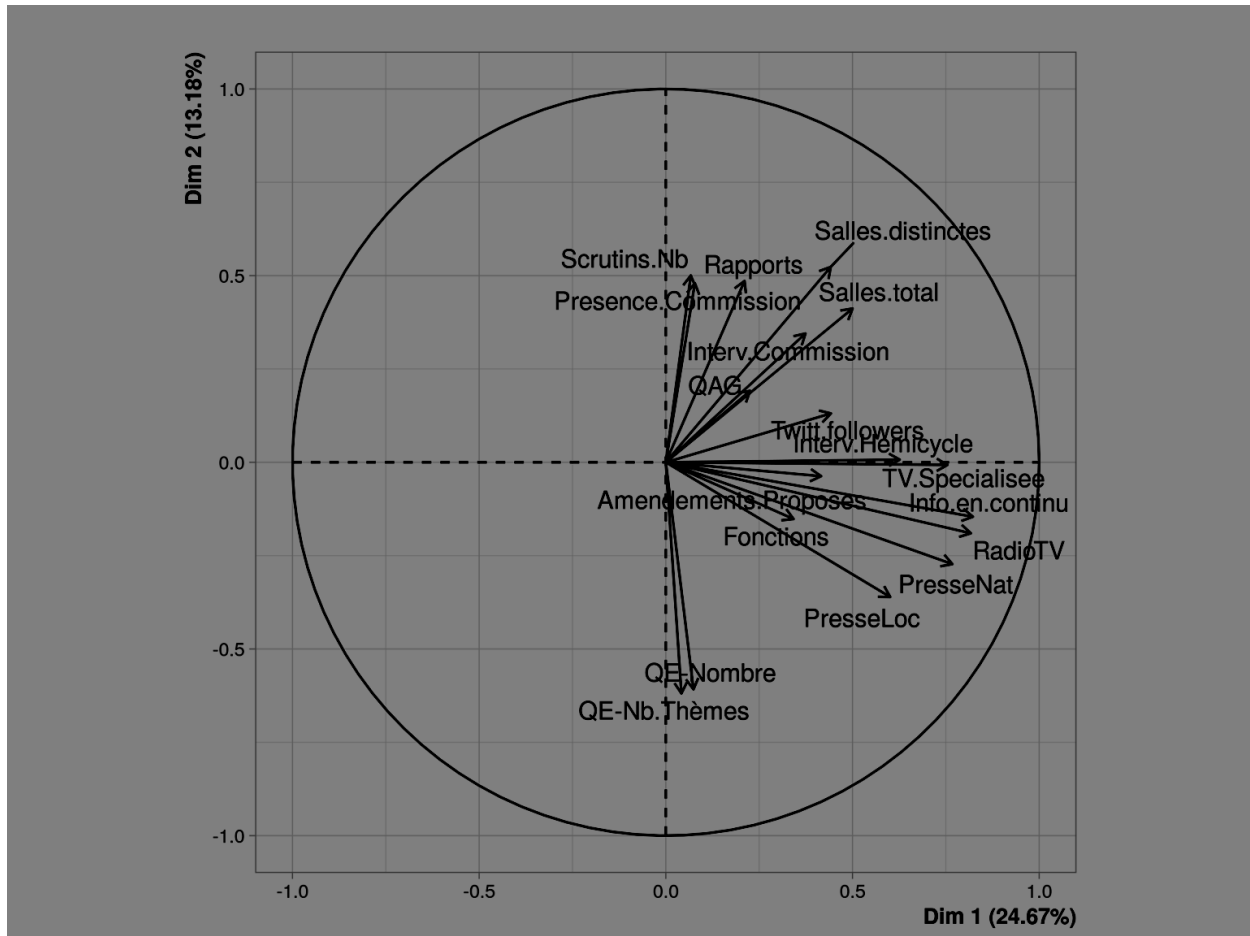


FIGURE 4b: Cercle des corrélations de l'ACP

Du fait de la concentration de la plupart des individus à la gauche du graphique, il est difficile de saisir toute différence moins marquée que l'opposition entre les ténors et tous les autres. Il n'est pas non plus possible de saisir certains des rôles particuliers que mettaient clairement en avant les SOM. Étudier la structure interne de la majorité de la population nécessite alors de multiplier les représentations des axes suivants, qui représentent une fraction de variance plus faible, ou d'exclure au préalable les individus extrêmes.

Cette pratique classique en AGD n'est pas nécessaire avec les cartes auto-organisatrices, qui permettent de saisir en une seule étape une grande partie de ces différences. Dans le cas des députés, les principales dimensions qui structurent le nuage de points entre la masse des élus et quelques-uns plus centraux sont bien visibles, mais elles n'occultent pas les nombreuses

différences entre députés moyens qui, quoique plus petites, n'en sont pas moins significatives.

L'abandon de l'hypothèse de linéarité ne se fait toutefois pas sans coûts. Le modèle ne fonctionnant pas par construction d'axes, il n'est plus possible de rapporter les corrélations ou contributions des différentes variables aux axes du graphique, ni de connaître la qualité de la transformation réalisée (part de variance conservée par les axes successifs). Ces indicateurs essentiels à l'interprétation d'une AGD n'ont pas de sens dans le cas des SOM. L'interprétation se fait d'abord de manière graphique, et sur la base des résultats de statistique descriptive issus de la carte.

Une autre difficulté, ou du moins propriété déstabilisante pour qui est habitué aux résultats exacts de la statistique classique, tient à la nature stochastique des SOM. En effet, puisque l'algorithme fait intervenir le hasard, deux exécutions successives donnent deux cartes différentes - même si leur interprétation qualitative reste la même dans la grande majorité des cas. Comme souvent avec les méthodes de clustering, le choix de la carte « optimale » peut se faire sur la base des mesures de qualité des cartes obtenues, mais peut également faire intervenir la préférence subjective du chercheur. Le lecteur est alors prié de croire en la bonne foi de ce dernier... ce qui ajoute encore un peu aux soupçons de manipulation des résultats auxquelles sont toujours sujettes les méthodes de réduction de dimensionnalité³².

Il est enfin utile de rappeler que si les deux méthodes ont une même ambition générale (la réduction de dimensionnalité), elles diffèrent dans l'application. L'ACP et les méthodes d'AGD cherchent à faire ressortir ce qu'on pourrait appeler la structure globale du nuage de points, c'est-à-dire à déterminer les (deux ou plus) dimensions les plus saillantes. Les méthodes non linéaires ont chacune un objectif différent, mais leur attention à la complexité fait qu'elles cherchent plutôt à préserver un élément de la topographie initiale. C'est particulièrement clair dans le cas d'outils comme t-SNE, un algorithme de *machine learning* non supervisé plus récent, qui cherche surtout à produire une représentation à deux dimensions où des points proches dans l'espace de départ

³²Que ce soit le choix des individus à inclure, le choix des variables et de leur pré-traitement (normalisations, transformations...), ou encore dans certains cas le choix des méta-paramètres de l'algorithme, les étapes qui conduisent aux résultats finalement publiés sont nombreuses, et sont autant d'embranchements de la recherche (invisibles pour le lecteur) qui auraient pu conduire à des résultats qualitativement différents. De ce fait, il existe toujours en réduction de dimensionnalité, probablement à plus forte raison encore qu'en régression, des soupçons de manipulation des résultats. De ce point de vue, on pourrait défendre que la stochasticité des SOM n'ajoute que marginalement au doute que devrait toujours susciter une analyse géométrique ou une classification hiérarchique. À cet égard, une bonne pratique consiste aussi à rendre ses données disponibles pour permettre des revisites.

seraient proches sur le graphique final³³. L'objectif est alors de représenter au mieux la structure locale des sous-populations cohérentes, sans se préoccuper des dimensions qui pourraient structurer le nuage de point. Dans le cas des SOM, on se situe dans un entre-deux puisque la non-linéarité permet de bien représenter des individus aux pratiques potentiellement très différentes sans que la représentation finale en soit bouleversée. Les SOM cherchent toutefois aussi à conserver une représentation de la structure globale.

Analyser l'espace parlementaire avec des variables supplémentaires

Comme c'est le cas en analyse l'analyse géométrique, il est possible avec les SOM de représenter, outre les variables actives, toute une série d'informations complémentaires. Il peut s'agir de variables supplémentaires (comme le genre, ou le parti d'affiliation), ou il peut s'agir des individus supplémentaires qu'on superpose sur la carte. Là encore, les SOM disposent d'un avantage : là où la variable supplémentaire est projetée en ACP, nécessitant une mesure de qualité de projection et rapprochant sur le papier des variables pas forcément proches dans l'espace de départ, ce n'est pas le cas avec les SOM. Les valeurs prises par les variables supplémentaires sont ici directement calculées pour chaque cellule de la carte, offrant un tableau bien plus précis. Elles ne sont pas non plus agrégées comme dans le cas de l'AGD, ce qui peut masquer des variations fortes³⁴. Afin d'illustrer ces possibilités, on revient à la question initiale à partir de deux exemples : le lien entre la place sur la carte et l'affiliation partisane, puis la position des novices dans cet espace parlementaire reconstitué.

Des opportunités structurées par les partis

L'interprétation peut ainsi être enrichie via la représentation de variables supplémentaires. La position des partis est particulièrement instructive. Comme le montre la figure 5, les élus du parti majoritaire, La République en Marche (LREM) sont présents presque partout. Ils sont

³³Laurens van der Maaten et Geoffrey Hinton, « Visualizing High-Dimensional Data Using t-SNE », *Journal of Machine Learning Research* 9 (Nov):2579-2605, 2008.

³⁴À l'inverse, la lecture d'une projection de variable est plus difficile sur une SOM que sur une AGD car elle se fait en au moins autant de points qu'il y a de cellules sur la carte (voir figures 2 et 4), plutôt qu'en un seul point par plan factoriel. Pour observer les corrélations entre deux variables, il faut donc se pencher de près sur la carte en analysant leurs variations croisées de cellule en cellule. On retrouve ici une caractéristique commune à la plupart des méthodes flexibles d'apprentissage : plus nuancés, leurs résultats sont aussi moins synthétiques.

cependant très sur-représentés dans le sud-ouest de la carte, dont on a vu qu'il regroupait des personnes très présentes aux votes. Cette sur-représentation est la conséquence directe de l'appartenance au parti majoritaire à l'Assemblée. Ce dernier doit en effet en permanence mobiliser ses troupes afin de s'assurer de l'obtention d'une majorité numérique lors des scrutins.

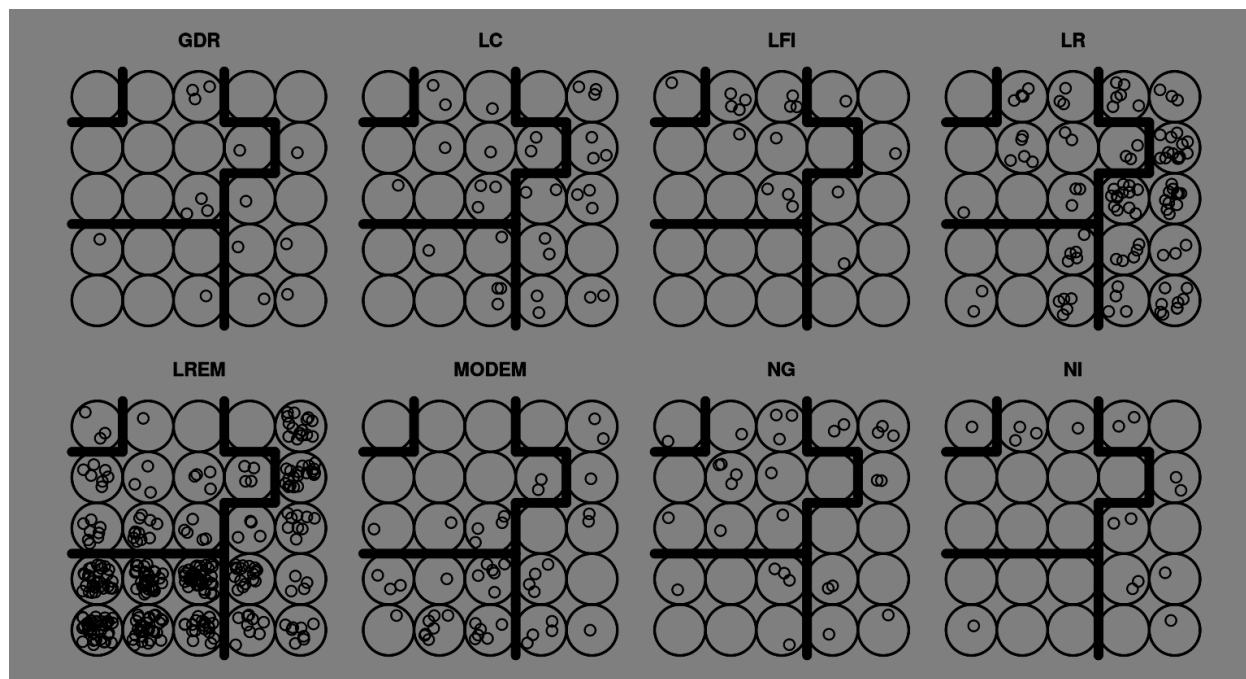


Figure 5: Distribution des élus par parti sur la carte de l'Assemblée

Les Républicains (LR), Nouvelle Gauche/Parti Socialiste (NG), La République en Marche (LREM), Les Constructifs / UDI (LC), Mouvement Démocrate (MODEM), La France Insoumise (LFI), Gauche Démocrate et Républicaine / Parti communiste (GDR), Non inscrits (NI)

Le contraste est alors très net avec les élus du parti de droite *Les Républicains (LR)*.

Premier parti d'opposition à l'Assemblée avec une centaine de membres, on retrouve ses membres sont là aussi situés partout. Plusieurs « ténors » de l'opposition, anciens présidents de parti, anciens ministres et députés chevronnés, sont placés dans le coin nord-ouest. On trouve aussi nombre d'entre eux dans la moitié est de la carte, là où les possibilités de contribution à la production législative sont plus faibles. C'est aussi parmi ce groupe politique que se retrouvent une bonne partie des élus fortement investis dans leur circonscription, mais moins actifs à l'Assemblée (bord nord-est de la carte). Élus depuis longtemps pour la plupart, ils se sont repliés sur leur territoire local après la défaite de 2017, une position classique pour les groupes d'opposition dont les marges de manœuvre sont faibles au palais Bourbon.

Le groupe *La France Insoumise* (LFI) présente un troisième type d'engagement parlementaire. Alors qu'il est particulièrement difficile d'accéder aux positions proches du coin nord-ouest, ses membres s'y concentrent massivement. On voit ici la traduction concrète d'une stratégie affirmée dès le début de la législature : utiliser l'Assemblée comme une tribune politique. Extrêmement présents dans les médias, ils sont aussi très investis dans de nombreuses activités parlementaires.

La place des novices et les sous-espaces de la carte

On l'a dit, une carte auto-organisatrice peut aussi être découpée en une série de zones, qu'indiquent les traits pleins sur la figure 5 ci-dessus. Le découpage se fait via un algorithme de classification³⁵, qui regroupe les cellules qui se ressemblent le plus. Dans le cas présent, on a opté pour une division en 4 zones. L'une fait ressortir une seule cellule, qui représente l'élite politique. Située au quadrant nord-ouest, elle comprend les quelques leaders cités ci-dessus. À côté d'eux, on trouve un groupe d'élus à la fois visible et particulièrement investi dans les activités mesurées par les différentes variables. On les appelle le « premier cercle » car c'est ici que se retrouvent les

³⁵Pour le regroupement des cellules de la carte nous nous sommes appuyés sur une classification hiérarchique, en utilisant le critère de regroupement de Ward. Cette deuxième étape de *clustering*, qui revient à créer des super-classes à partir des nœuds de la carte, est une pratique courante dans l'utilisation des SOM (voir par exemple Marie Cottrell, Patrick Letrémy, Patrick Rousset et Smail Ibbou, « Cartes auto-organisées pour l'analyse exploratoire de données et la visualisation », *Journal de la Société Française de Statistique*, 144, 2003). Dans le cas présent, la classification séparait en plusieurs classes les cellules du groupe « premier cercle » avant de séparer les députés « du rang » en « rang investissement Assemblée » et « rang investissement externe ». Nous avons choisi d'opérer un regroupement manuel des cellules du premier cercle, afin de simplifier la lecture de la carte. Les distances euclidiennes entre prototypes du « premier cercle » sont certes grandes, puisqu'ils représentent des députés très actifs selon des modalités différentes, mais la séparation de ces cellules est moins utile à l'interprétation que celle des députés du rang, moins actifs mais beaucoup plus nombreux.

principaux responsables politiques nationaux. Le reste des élus constitue ce qu'on pourrait appeler des « députés du rang ». La partie située au sud regroupe des élus très investis dans le quotidien du travail parlementaire. La partie située à l'est rassemble des individus moins actifs à l'Assemblée, principalement du fait du manque d'opportunités lié à leur appartenance à l'opposition.

Comment se situent, sur cette carte ainsi segmentée, les novices ? Dans un article portant sur la composition de l'Assemblée, on avait proposé de distinguer les nouveaux élus de 2017 en fonction de leur trajectoire politique antérieure³⁶. En prenant en compte la part de la vie active passée activement (de manière rémunérée) en politique et l'exercice de ces fonctions (part du temps passé dans des fonctions nationales vs locales), on aboutissait à une partition des nouveaux députés en quatre groupes. On distinguait alors ceux qui avaient passé la plupart de leur vie sans toucher de rémunération d'une activité politique (« société civile »), ceux qui avaient eu une seconde carrière puis s'étaient engagés en politique pour quelques années avant cette élection (« seconde carrière »), ainsi que deux autres groupes : les élus qui accédaient à l'Assemblée après une longue période en politique passée pour la plupart en mandat local (« ancrage local »), et finalement ceux qui avaient aussi passé une part importante de leur vie en politique, mais avaient eux assez rapidement accédé à une position au cœur du pouvoir (« carrière au centre »).

Pour notre propos, cette quadripartition effectuée sur les primo-députés doit être enrichie. Il faut d'une part ajouter les députés sortants. Minoritaires sous cette législature, ils représentent tout de même 28% de la population des élus de l'Assemblée.

Par ailleurs, si la catégorie « société civile » est heuristique dans le cadre d'une interrogation sur la professionnalisation politique, car elle rassemble des personnes qui n'ont pas (ou presque) vécu financièrement de la politique, elle mérite d'être retravaillée dans le cadre d'une réflexion sur les effets de l'expérience politique. En effet, ce cluster rassemble des profils différents, avec d'un côté des personnes qui ont connu une longue socialisation en politique, et de l'autre celles qui découvrent la politique avec leur entrée à l'Assemblée. C'est d'ailleurs là une ligne de partage entre deux groupes comme *La France Insoumise* et *La République en Marche*. L'un et l'autre concentrent la majorité des élus qu'on retrouve dans le groupe « société civile », mais leurs profils varient toutefois, car la totalité des élus LFI sont des élus aux trajectoires

³⁶Pour plus de détails sur la construction, voir Julien Boelaert, Sébastien Michon, Étienne Ollion, « Le temps des élites », *Revue française de science politique*, 68 (5), octobre 2018, p. 795 et sq.

militantes affirmées, certains ayant même exercé des responsabilités importantes. Ce n'est pas le cas pour une partie des élus LREM, et plusieurs dizaines d'entre eux ne s'étaient jamais engagés avant 2017.

Afin de rendre compte de cette différence, on a donc divisé la catégorie société civile en deux sous-catégories : « expérience politique » et « novices », selon qu'elles et ils ont eu un engagement politique actif ou non. La première catégorie reste peu cohérente, car elle rassemble des personnes qui ont été militants pendant deux décennies et des néo-députés qui ont pris part à une seule campagne. Elle prend toutefois sens par rapport à celle des novices. Le tableau 2 résume ces différents profils³⁷.

+ Temps passé en fonctions locales -	« Société civile » : <i>Novices / Exp. politique</i> (113)(129)	<i>Seconde carrière</i> (51)	<i>Ancrage local</i> (71)	<i>Parlementaires réélus</i> (152)
			<i>Carrière au centre</i> (52)	
-Part de la vie passée rémunéré en politique+				

Tableau 2: Représentation stylisée des carrières politiques des élus de 2017

La figure 6 montre où se positionne chacun de ces groupes sur la carte. Il démontre que, pour la première année de mandat au moins, le passé politique joue un rôle central dans le placement des élus. Plutôt que par cellule, on considère leur placement dans chacune des super-classes précédemment distinguées. La première (« élites ») et la deuxième (« 1er cercle ») voient une forte sur-représentation des élus appartenant au groupe « carrière au centre ». Ils y sont même légèrement mieux représentés que les députés réélus, qui pourtant avaient pendant cinq ans au

³⁷Ce tableau reprend, tout en le précisant, le tableau présenté dans l'article de Boelaert *et al.*, 2018, *art. cit.*, p. 796. Il en diffère principalement par sa meilleure description de la dimension temporelle des carrières des différents types d'élus. Il précise aussi le nom d'une catégorie, hier appelée « entourage », et ici « carrière au centre ». Car si les anciens membres des entourages d'élus sont bien représentés dans cette catégorie, ils ne sont pas les seuls.

moins eu l'occasion de se familiariser avec l'institution et ses codes. Ces deux groupes se distinguent clairement des autres, et en premier lieu des novices, quasiment absents des places de pouvoir. Ces derniers sont en revanche largement sur-représentés parmi les députés du « rang » à investissement interne.

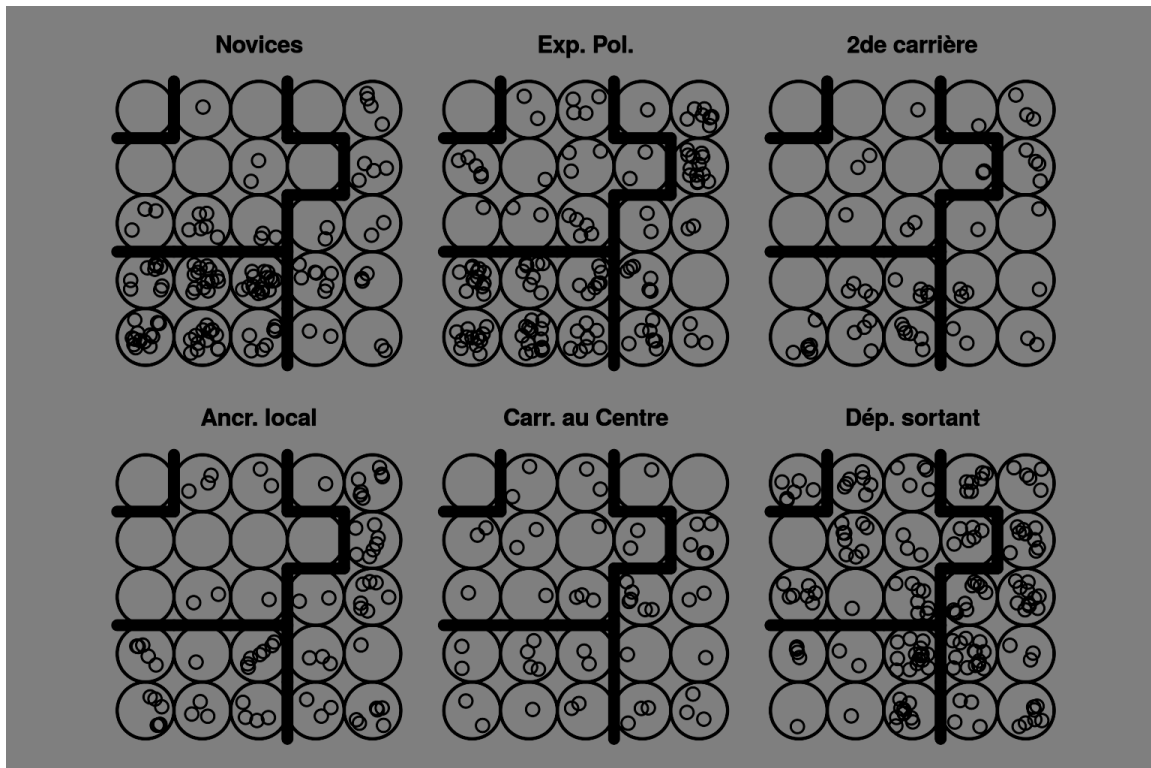


Figure 6: Distribution des élus selon le parcours politique antérieur

Le tableau 2 vient quantifier ces résultats : les novices n'accèdent presque jamais aux positions de premier plan, alors que les députés réélus, ainsi que les primo-députés qui ont fait une carrière au centre, s'y trouvent très bien représentés. Ce résultat confirme des éléments déjà connus, à savoir que les anciens collaborateurs parlementaires (surtout s'ils étaient basés à Paris) et les anciens membres de cabinet empruntent une voie royale. Arrivés plus jeunes à l'Assemblée, ils y accèdent plus rapidement aux responsabilités. Vraie au cours des décennies passées, cette situation n'a finalement pas été modifiée par l'élection de 2017, qui a pourtant vu le bouleversement de nombre de paramètres politiques. Et qui s'était faite sur la promesse d'un renouvellement dont on voit que dans les faits, il n'a pas bouleversé les hiérarchies internes au champ politique.

	Rang (investissement Assemblée)	Rang (investissement externe)	1^{er} cercle	Élites	Total (effectifs)
Novices	63,7	23,0	13,3	0,0	100 (113)
Expérience politique	48,8	31,8	19,4	0,0	100 (129)
Seconde carrière	49,0	35,3	15,7	0,0	100 (51)
Ancrage local	36,6	52,1	11,3	0,0	100 (71)
Carrière au centre	25,5	43,1	31,4	0,0	100 (52)
Députés	20,8	45,5	29,9	3,9	100

sortants³⁸					(152)
Total	40,6	37,6	20,7	1,1	100 (568)

Tableau 3 : Répartition des profils d'activité des députés selon leur type de carrière politique (En %, sauf indication contraire)

Conclusion

L'objectif de ce texte est de montrer les potentialités d'une méthode statistique de réduction de dimensionnalité et de clustering, les cartes auto-organisatrices (*self-organizing maps*). Ce que l'exemple des députés de la XVème législature montre, c'est l'intérêt de cette méthode pour produire une représentation réaliste du parlement français lors de sa première année. Les différents rôles souvent mentionnés dans la littérature sont alors clairement distingués par la carte, qui permet de représenter à la fois les grandes tendances du nuage de points, mais aussi de rendre justice à des variations plus minimes, mais dans bien des cas pas moins importantes. En offrant la possibilité de distinguer entre députés du rang, en donnant à voir des formes d'investissement au sein de la majorité entre élus actifs dans le travail législatif et élus orientés vers les médias, les cartes auto-organisatrices permettent d'opérationnaliser nombre de réflexions que les chercheuses avaient mises en avant via des méthodes qualitatives.

Au-delà du cas des parlementaires, les SOM sont un outil utile pour les sciences sociales en ce qu'elles permettent plus généralement de représenter et catégoriser un espace social dans sa diversité. Sans forcément remiser les méthodes classiques, l'analyse géométrique de données en premier lieu, les SOM offrent des potentialités d'exploration, de catégorisation et de représentation souvent supérieures, et en tout cas bien différentes par rapport aux méthodes plus classiques. C'est particulièrement le cas lorsque les données ont une distribution asymétrique de leurs variables.

³⁸Marine Le Pen et Jean-Luc Mélenchon ont été placés dans cette catégorie, et non dans celle où ils se trouvaient via la catégorisation, afin de refléter le poids de leur expérience politique (en tant qu'ancienne parlementaire européenne, qu'ancien sénateur).

Cette situation oblige souvent à de nombreuses opérations en AGD, quand la flexibilité des SOM leur permet d'intégrer cette caractéristique. Leur intérêt est aussi plus évident à une époque où l'abondance des sources données numériques³⁹ fait que le nombre d'indicateurs s'accroît massivement, nécessitant souvent des explorations fines avant de pouvoir utiliser les variables. En tout état de cause, le renouveau récent de l'intérêt pour les méthodes d'apprentissage statistique non supervisé constitue un terrain fécond que les chercheuses en sciences (sociales) gagneraient à investir.

³⁹Voir Étienne Ollion et Julien Boelaert, « Au-delà des big data. Les sciences sociales face à la multiplication des données numériques », in *Sociologie*, 2015.

Annexe 1 : Les données de l'enquête

Les données quantitatives mobilisées dans l'enquête ont l'origine suivante :

- ☞ Activités législatives: collectées via le site de l'Assemblée nationale ou l'API du site www.nosdeputés.fr.
- ☞ La catégorie « responsabilité institutionnelle » a été construite en sélectionnant les personnes qui avaient exercé des fonctions à l'Assemblée offrant des ressources ou de la visibilité, pouvant être chronophages, mais ne se traduisant pas forcément dans un surcroît d'activité en ce qui concerne les autres mesures. L'exemple-type est la figure des questeurs, qui disposent de moyens accrus et sont connus de tous, mais dont le travail (important) ne se traduit pas par une augmentation des autres variables. Pour cette catégorie, on a donc pris en compte les activités centrales au fonctionnement de l'institution, à savoir la présidence (1 personne), vice-présidence (6 personnes), questure (3 personnes), présidence de groupe (7 personnes, ainsi que la présidente du Rassemblement national qui fonctionne comme un groupe parlementaire). On y a ajouté les 14 élus ayant un rôle (présidence, secrétariat) dans l'un des sept groupes de travail sur la réforme des institutions car ces positions étaient à la fois chronophages et vecteurs de visibilité relative lors de cette première année. Les présidences de commissions et de délégations n'ont pas été ajoutées à cette catégorie car, comme on le voit sur la carte, elles entraînent une activité particulière et déjà très remarquable.
- ☞ Invitations dans les médias. Audiovisuels : collecte automatisées de données de programme télévision et via des sites spécialisés (<https://www.politiquemaniam.com/>); Mentions dans la presse nationale et locale : collecte automatisée via des bases recensant les articles de presse; Nombre de followers sur Twitter : collecte automatique au 1er septembre 2018.
- ☞ Responsabilités institutionnelles : collecte manuelle, questions au gouvernement, questions écrites, nombre de salles réservées: open data de l'Assemblée nationale.
- ☞ Trajectoire professionnelle et politique des élus : voir (Boelaert, Michon et Ollion, 2018).

Annexe 2 : Statistiques descriptives, par cluster

	Âge	Femmes (%)	Nombre de mandats à l'Assemblée	Primo-députés (%)	Anciens auxiliaires politiques (%)	Vie en politique (années)	Effectifs
Élites	53,0	16,7	1,2	33,3	83,3	24,7	6
1er cercle	46,6	31,4	0,8	61,0	38,9	11,7	118
Rang (investissement externe)	50,8	32,2	0,8	67,3	21,0	13,1	214
Rang (investissement interne)	48,7	50,6	0,2	86,1	14,7	7,2	231

Tableau 4 : Propriétés socio-démographiques et politiques des députés de 2017 selon leur profil d'activités.

	Rang (investissement interne)	Rang (investissement externe)	1 ^{er} cercle	Élites	Ensemble
Agriculteur	2,6	2,9	2,6	0,0	2,7
Artisan commerçant chef d'entreprise	15,3	9,7	6,8	0,0	11,3
Cadre supérieur du privé	26,6	20,3	21,4	16,7	23,1
Cadre supérieur du public	18,3	21,7	25,6	0,0	20,9
Collaborateur	7,9	15,5	24,8	83,3	15,0

politique					
Employé	0,4	1,9	0,9	0,0	1,1
Profession intermédiaire	15,7	16,9	6,8	0,0	14,1
Profession libérale	13,1	11,1	11,1	0,0	11,8
Total	100,0	100,0	100,0	100,0	100,0

Tableau 5 : Profession la plus significative des députés selon leur profil d'activités (%).
 La profession la plus significative est définie comme la profession dans laquelle les députés ont passé le plus de temps avant d'être élus à plein temps. Pour plus d'informations, voir Boelaert, Michon et Ollion, 2017, *op. cit.*, p. 110.

Annexe 3 : Graphique « radar »

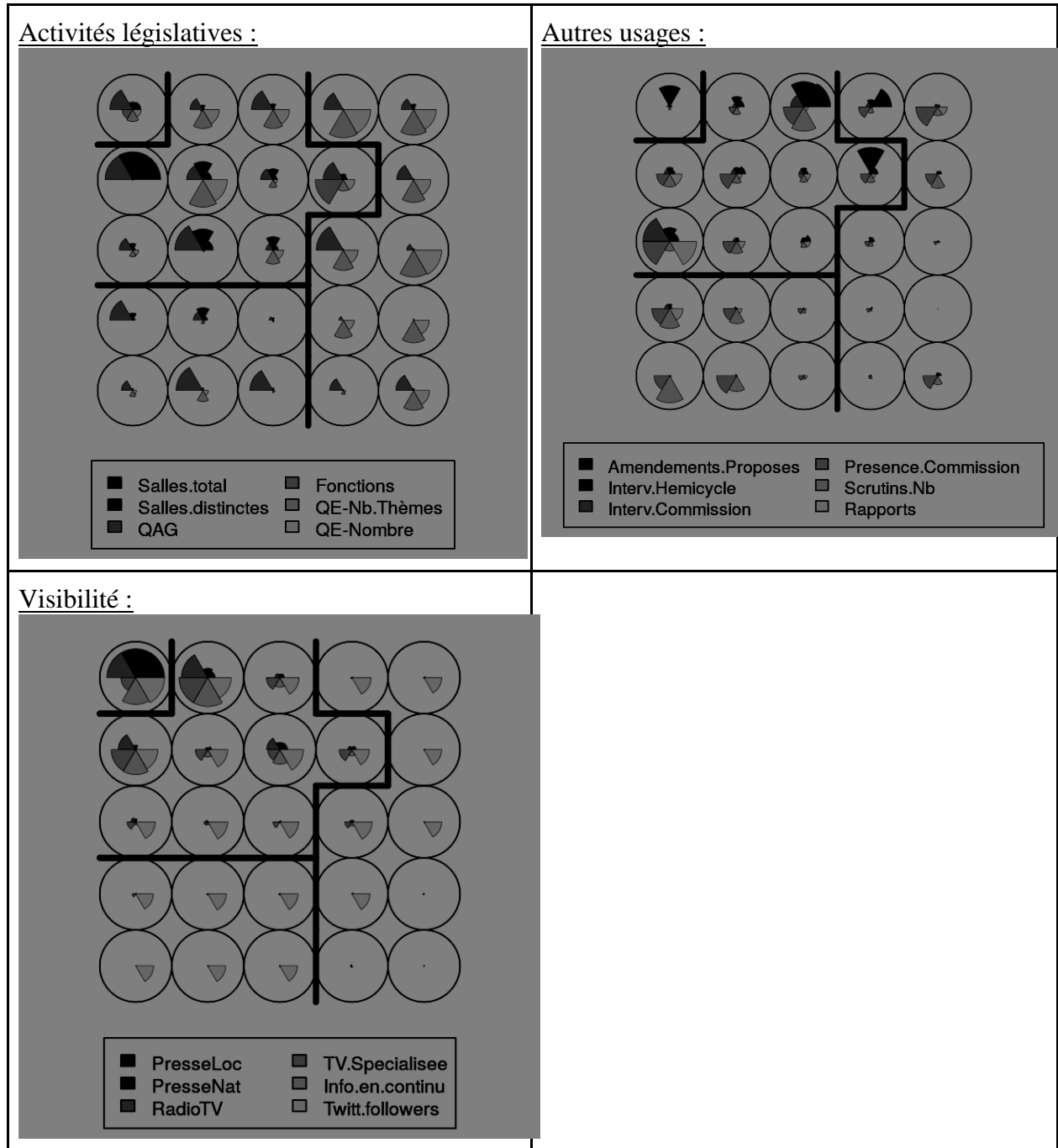


Figure 7 : Graphiques en radar des variables actives de la SOM apprise (activités des députés).

L'information représentée est strictement la même que celle de la figure 2, à savoir les valeurs des prototypes de la carte. L'intensité de chaque activité dans une cellule est représentée par la taille du segment de disque correspondant.

Annexe 4 : ACP

	Axe 1 (25% inertie)		Axe 2 (13% inertie)		Axe 3 (12% inertie)	
	Corrélation	Contribution	Corrélation	Contribution	Corrélation	Contribution
PresseNat	0.77	13.27	-0.27	3.13	-0.22	2.11
PresseLoc	0.60	8.14	-0.36	5.48	-0.11	0.54
RadioTV	0.82	15.04	-0.19	1.52	-0.28	3.45
TV.Specialisee	0.75	12.82	-0.01	0.00	-0.06	0.15
Info.en.continu	0.82	15.24	-0.15	0.90	-0.27	3.37
Twitt.followers	0.44	4.40	0.13	0.72	-0.04	0.09
Amendements.Pr oposes	0.42	3.90	-0.04	0.06	0.59	15.82
Interv.Hemicycl e	0.63	8.84	0.01	0.00	0.37	6.16
Interv.Commissi on	0.37	3.15	0.34	5.02	0.54	13.31
Presence.Commi ssion	0.08	0.14	0.48	9.71	0.53	12.67
Scrutins.Nb	0.07	0.10	0.50	10.56	0.25	2.91
Rapports	0.21	1.00	0.49	9.94	0.24	2.59
Salles.total	0.50	5.63	0.41	7.15	-0.21	1.98
Salles.distinctes	0.44	4.39	0.52	11.55	-0.13	0.71
QAG	0.23	1.15	0.19	1.54	0.19	1.72

Fonctions	0.34	2.64	-0.15	0.97	0.01	0.00
QE-Nb.Thèmes	0.04	0.04	-0.62	16.15	0.58	15.50
QE-Nombre	0.07	0.12	-0.61	15.58	0.61	16.93

Tableau 6 : Caractérisation des trois premiers axes de l'ACP