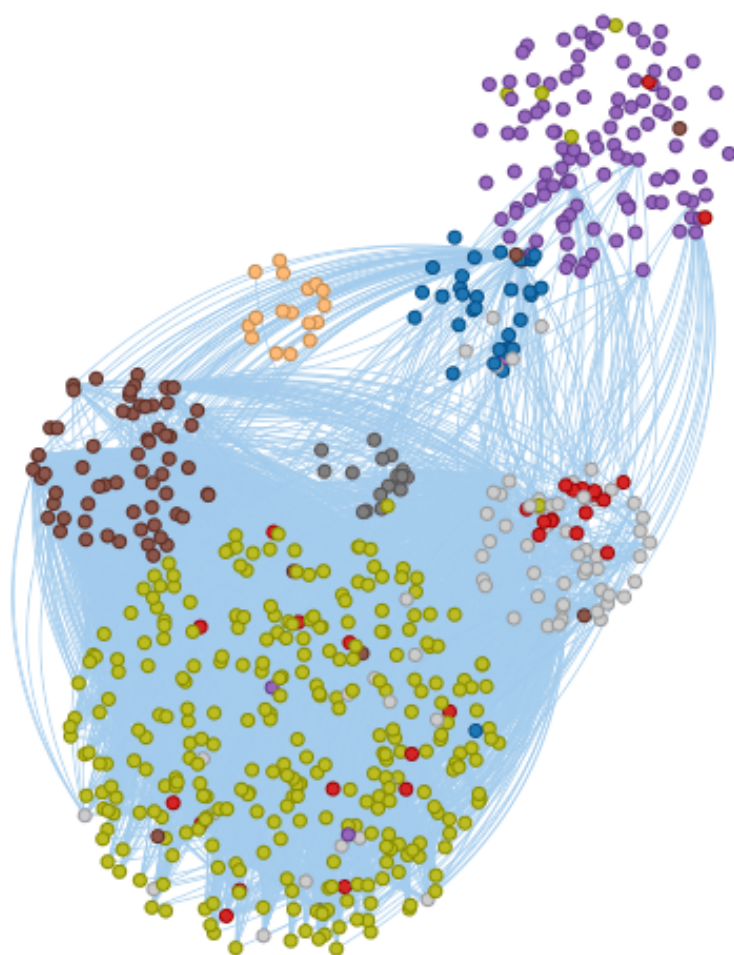


Les enjeux environnementaux à l'Assemblée nationale, entre parole et action

Calliste Ladouas et Louis-Alexis Peneloux

Avril et mai 2025



Sommaire

1	Objectifs du projet	3
2	Construction du graphe "député-amendements"	4
3	Détection de communautés	5
4	Détection des propos liés à l'environnement dans les débats	8
5	Extraction des lois environnementales et des votes des députés	9
6	Détermination des députés les plus influents et actifs	10
7	Analyse sur les votes et les discours environnementaux selon le cluster	12
7.1	Hypothèses à tester et cadre de validité du test Anova	12
7.2	Analyse des votes portant sur des lois environnementales selon le cluster .	13
7.3	Analyse des occurrences liées à des sujets environnementaux selon le cluster	15
7.4	Analyse du rapport vote/discours pour les enjeux environnementaux	17
8	Analyses statistiques et interprétations	18
8.1	Corrélation avec la participation totale aux votes	19
8.2	Corrélation avec la centralité	20
8.3	Corrélation entre parole environnementale et vote environnemental	21
9	Conclusion	22

1 Objectifs du projet

Ce projet part d'une question simple : les discours politiques sur l'environnement sont-ils toujours suivis d'actions concrètes ? À l'Assemblée nationale, certains députés prennent souvent la parole sur les enjeux climatiques. Mais cela se traduit-il dans leur participation aux votes sur les textes de loi liés à l'environnement ? Notre objectif a été d'identifier d'éventuels **décalages entre la parole et l'action**, autrement dit de repérer des formes de greenwashing parlementaire.

Pour cela, nous avons croisé plusieurs sources de données issues de la **XVe législature** : les interventions orales lors des débats, les amendements déposés et co-signés, et les scrutins sur les textes à caractère environnemental. En combinant **analyse de graphes**, **détection de communautés**, **mesures de centralité** et **tests statistiques**, nous avons cherché à caractériser les comportements individuel et collectif des députés selon leur **appartenance politique**. Ce rapport retrace l'ensemble de notre démarche, les outils utilisés, les choix effectués ainsi que les premiers enseignements tirés de cette exploration.

Il y a **trois grandes dimensions à notre projet** :

- **exploration de données**
 - construction du graphe "député-amendements"
 - extraction des lois environnementales et des votes des députés
- **extraction d'informations**
 - détection de communautés
 - détection des propos liés à l'environnement dans les débats
 - détermination des députés les plus influents et actifs
- **analyses statistiques et interprétations**

Afin de clarifier notre démarche, nous déterminons 5 paramètres cruciaux à calculer. Nous pouvons ensuite raisonner à partir de ces **5 caractéristiques par député pour tester différentes hypothèses**. Notons aussi que nous nous attachons à **analyser les spécificités de chaque cluster à l'Assemblée en termes d'implication générale puis environnementale, dans les votes comme dans les débats**.

On calcule donc pour chaque député :

- sa **centralité** C à partir du graphe obtenu ;
- son **nombre de participations** P_{tot} **aux votes** lors de la XVe législature ;
- son **nombre de participations** P_e **aux votes pour les lois environnementales** une fois celles-ci extraites ;
- la **longueur** M_{tot} **des propos** prononcés lors des débats (en nombre de caractères) ;
- le **nombre de mots** M_e **liés aux enjeux environnementaux** prononcés au cours des débats.

Voici deux exemples d’hypothèses que nous allons tester. Nous allons étudier la corrélation entre $\frac{M_e}{M_{tot}}$ et $\frac{P_e}{P_{tot}}$ afin de savoir si la parole (implication dans les débats) et l’action (participation aux votes) vis-à-vis des enjeux environnementaux sont proportionnelles. Autre exemple : on calcule la corrélation entre la centralité C et M_e afin de déterminer si les députés les plus influents sont ceux qui parlent le plus de sujets environnementaux.

2 Construction du graphe ”député-amendements”

Comme première extraction de données, nous avons construit un premier graphe à partir des données officielles de l’Assemblée nationale se basant sur les **amendements** déposés pendant la XVe législature qui a eu lieu de 2017 à 2022. Nous nous concentrons sur cette législature car il s’agit de la législature complète la plus récente (l’Assemblée nationale de la XVIe législature ayant été dissoute au bout de deux ans). L’objectif est de **relier deux députés dès lors qu’ils ont co-signé au moins un amendement**, en pondérant les arêtes du graphe par le nombre total de co-signatures. Ainsi, on espère pouvoir cerner avec ce graphe les **proximités** entre députés ainsi qu’extraire des informations sur la **centralité** de tel ou tel député.

Les données nécessaires sont accessibles sur le site **Open Data de l’Assemblée nationale, sous forme de fichiers XML**. Chaque fichier d’amendement contient notamment les identifiants des députés signataires, au format PAXXXXXX. Nous avons traité environ **200 000 amendements**, ce qui a nécessité un temps de calcul non négligeable.

Ces identifiants n’étant pas directement interprétables, nous utilisons **une seconde base de données** pour associer à chaque identifiant les informations correspondantes : nom, prénom, circonscription, etc. L’appartenance politique n’étant pas systématiquement fournie dans ces fichiers, nous avons complété cette information à partir du site nosdeputes.fr qui met à disposition des informations sur chacun des élus.

Nous exploitons donc cette [base de données](#) pour en extraire la liste des députés et le parti auquel ils sont affiliés. On obtient **13 partis** (dont les non inscrits). Voici la liste ci-dessous :

- LREM - La République en Marche - 299 députés
- NG - Nouvelle Gauche - 2 députés
- N/A Non Affilié - 4 députés
- UDI - Union des Démocrates et Indépendants - 22 députés
- UAI - UDI, Agir et Indépendants - 4 députés
- AE - Agir Ensemble - 23 députés
- MODEM - Mouvement Démocrate et Démocrates apparentés - 66 députés
- LT - Libertés et Territoires - 18 députés
- NI - Non Inscrits - 34 députés

- GDR - Gauche Démocrate et Républicaine - 18 députés
- SOC - Socialistes et apparentés - 35 députés
- LFI - La France Insoumise - 17 députés
- LR - Les Républicains - 117 députés

Au total, il y a **659 personnes qui ont occupé le poste de député pendant la XV^e législature, pour 577 postes fixes**. On fait le choix de **garder** les personnes qui n'ont pas fait un mandat complet.

Une fois ces correspondances établies, nous pouvons contruire le graphe. Chaque nœud représente un député, et une arête relie deux nœuds si les députés concernés ont co-signé au moins un amendement comme expliqué précédemment. Le poids de l'arête correspond donc au nombre total d'amendements co-signés. Le résultat est un graphe dense qui constitue **notre base pour les analyses ultérieures** : détection de communautés, mesures de centralité ou encore étude des dynamiques de collaboration à l'Assemblée nationale.

3 Détection de communautés

Nous testons **plusieurs méthodes de détection de communautés** sur le graphe député-amendements. L'objectif est d'identifier des groupes de députés qui collaborent fréquemment entre eux, à partir des co-signatures d'amendements. Au-delà de la simple vérification que ces groupes recoupent les appartenances politiques officielles, nous voulons aussi voir si certains députés non inscrits, ou appartenant à des formations très peu représentées, peuvent être rattachés à des dynamiques collectives plus larges.

Dans un premier temps, nous utilisons l'**algorithme de propagation de labels** (label propagation) **dans sa version pondérée**. Bien qu'il soit simple et rapide, son résultat se révèle peu pertinent dans notre cas : les clusters manquent de cohérence au niveau de l'orientation politique. De surcroît, LREM, parti largement majoritaire, se retrouve sur-représenté dans quasiment toutes les communautés. Cela rend les résultats difficilement interprétables.

Nous optons alors pour l'**algorithme de Louvain**, basé sur la maximisation de la modularité. Cette méthode donne des résultats nettement plus satisfaisants : les clusters obtenus sont beaucoup plus **homogènes politiquement**. Certains regroupent exclusivement des députés LFI, MODEM ou LR, tandis que d'autres, mixtes, restent interprétables. Le grand cluster LREM apparaît naturellement mais, cette fois, les autres grands groupes sont bien séparés, et les non-inscrits ou membres de petits partis sont en grande partie rattachés à des clusters dominants, ce qui permet de les positionner politiquement à partir de leurs collaborations. On **raisonne alors maintenant en terme de clusters plutôt que de partis politiques** pour étudier les dynamiques collectives.

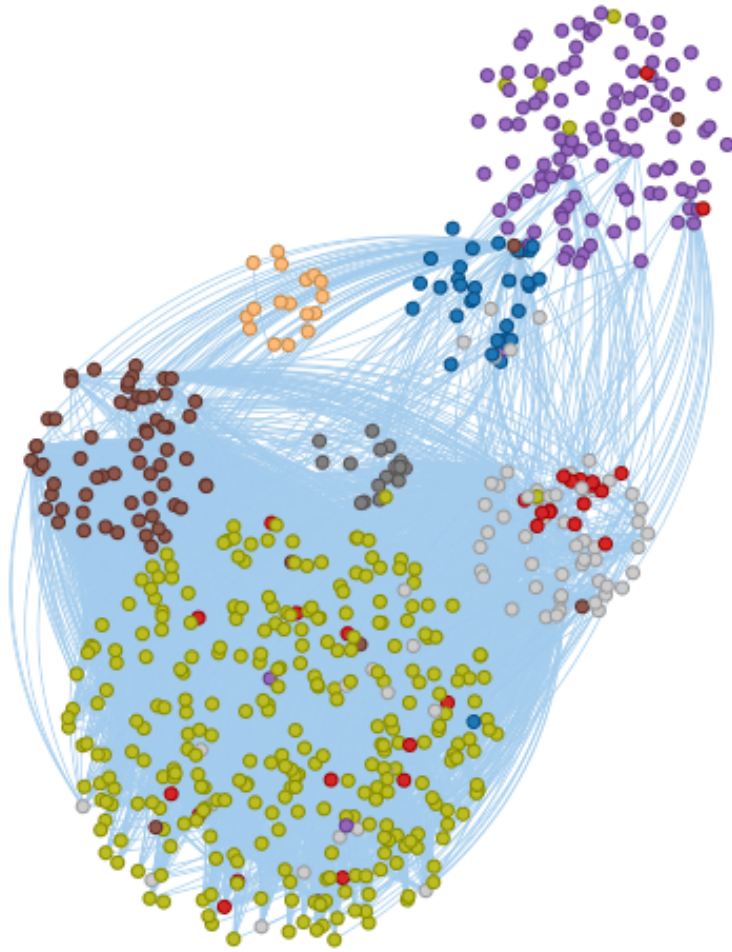


Figure 1: Représentation du graphe député-amendements. Les clusters sont mis en évidence. Seules les arêtes inter-cluster avec les plus grands poids sont représentées.

Gardons cet ordre et nommons chaque cluster selon son parti dominant :

Cluster 1 - **maj SOC** : Socialistes à 80% - 39 députés

Cluster 2 - **maj GDR** : GDR à 95% - 18 députés

Cluster 3 - **maj LR** : LR à 94% - 119 députés

Cluster 4 - **NI** : Non inscrits à 100% - 11 députés

Cluster 5 - **LFI** : LFI à 100% - 17 députés (la totalité de LFI)

Cluster 6 - **MODEM** : MODEM à 100% - 63 députés

Cluster 7 - **maj LREM** : LREM à 89% - 328 députés

Cluster 8 - **mixte** : Mixte avec 38% de UDI suivi par LT et AE - 55 députés

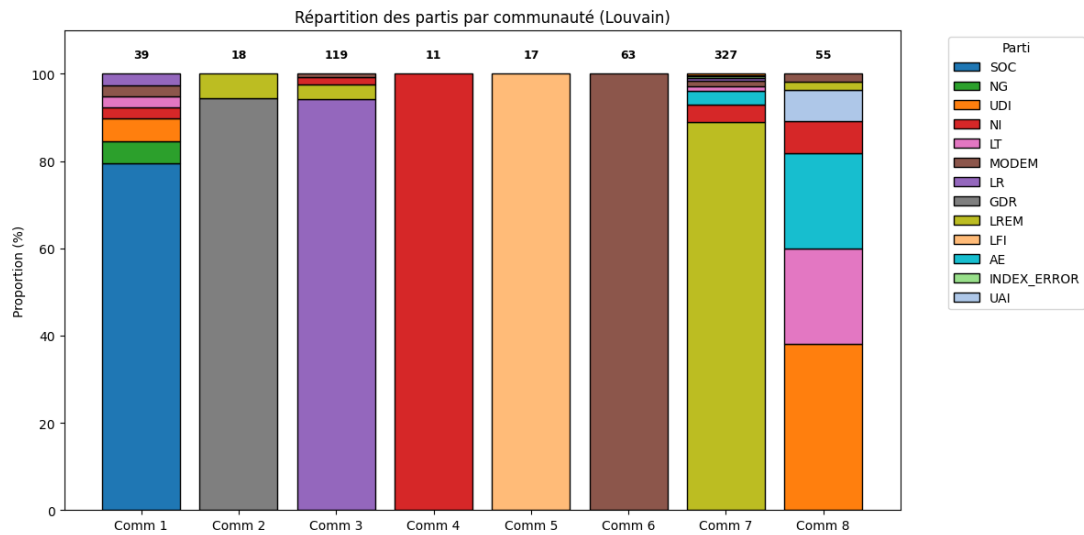


Figure 2: Répartition des partis au sein des clusters obtenus avec l'algorithme de Louvain

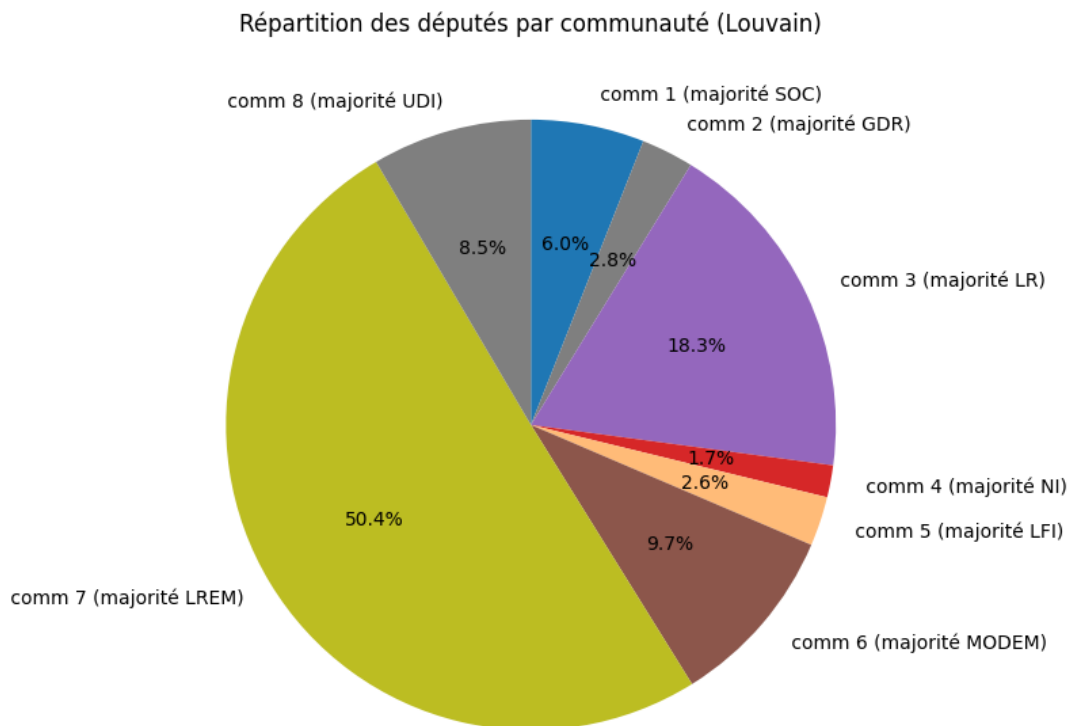


Figure 3: Répartitions des députés selon les différents clusters

4 Détection des propos liés à l’environnement dans les débats

On analyse les compte-rendus de l’Assemblée nationale de la XVe législature en filtrant avec des mots-clés reliés aux enjeux environnementaux. Pour une url d’une séance donnée, on extrait le nom des orateurs qui parlent d’enjeux environnementaux et on compte le nombre d’occurrences de mots-clés. On compte les répétitions car, par exemple, si un député utilise plusieurs fois le mot ”réchauffement climatique” dans son intervention, alors c’est un signe d’importance de ces enjeux dans sa prise de parole.

On veut maintenant extraire les url des 1562 séances du [site de l’Assemblée](#) pour ensuite y appliquer l’algorithme qui détecte dans le code HTML le nombre d’occurrences de mots-clés liés aux enjeux environnementaux pour chaque intervenant. Étrangement, quelques pages telles que la page 24 (36 séances apparaissent sur chaque page) ne passent bien dans l’algorithme. On a essayé de résoudre ce problème mais en vain, même si la structure HTML nous semblait être la même.

Nous avons donc basculé sur la [base de données open source de l’Assemblée nationale](#). Nous comptons dans un dictionnaire le nombre d’occurrences de mots-clés environnementaux. Cette sélection de mots nous donne un bon premier aperçu de l’intérêt des députés pour les sujets environnementaux.

Barbara Pompili arrive en tête avec 1419 occurrences sur la XVe législature, puis Mathilde Panot avec 955, François de Rugy avec 816, Dominique Potier avec 809 et François-Michel Lambert avec 780.

Voilà la répartition générale :

102 députés ont une occurrence ou moins.
82 députés ont entre 2 et 5 occurrences.
175 députés ont entre 5 et 10 occurrences.
136 députés ont entre 11 et 20 occurrences.
139 députés ont entre 21 et 40 occurrences.
132 députés ont entre 41 et 100 occurrences.
116 députés ont utilisé plus de 100 occurrences.

Voyons maintenant si un ou plusieurs clusters se distinguent en termes d’engagement sur les sujets environnementaux en recoupant les résultats obtenus ci-dessus et le clustering effectué précédemment.

On obtient le résultat suivant en rapportant le nombre d’occurrences au nombre de députés de chaque cluster.

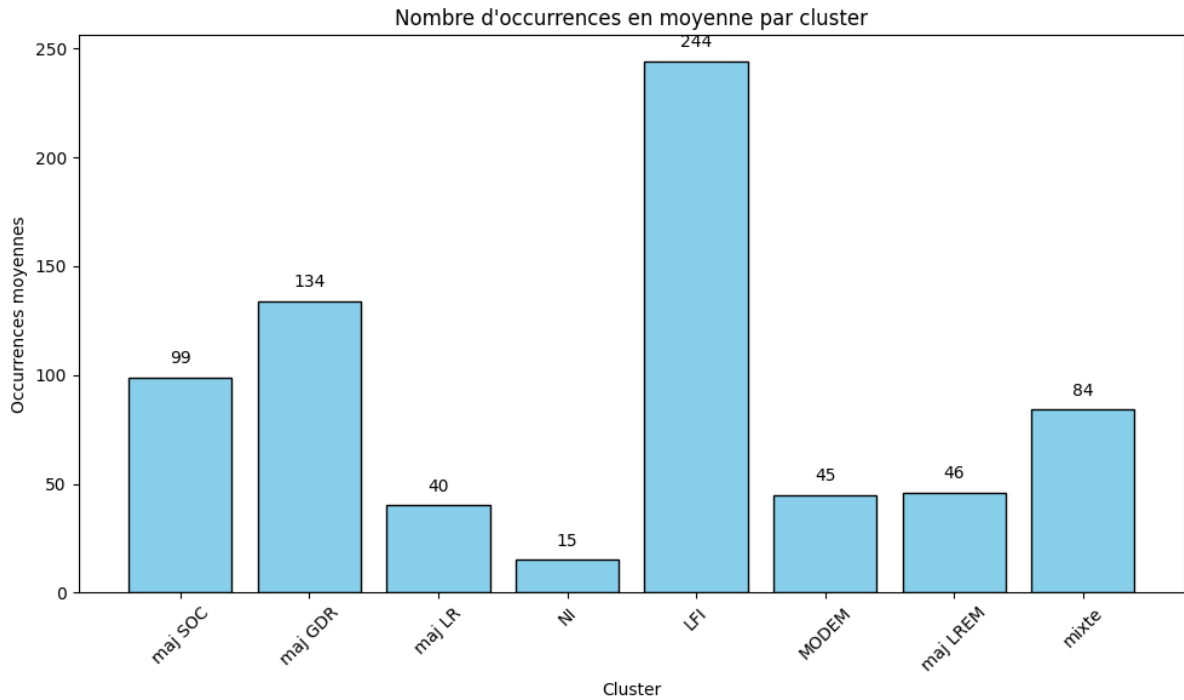


Figure 4: Nombre moyen d’occurrences liées aux sujets environnementaux selon le cluster

On voit que LFI s’empare davantage des sujets environnementaux proportionnellement au nombre de députés. Le nombre d’occurrences de mots autour des enjeux environnementaux est plus de **5 fois plus élevé** pour LFI par rapport au cluster 7 à majorité LREM. Les 10 députés les plus engagés sur ces sujets sont rattachés aux partis suivants dans l’ordre : LREM, LFI, LREM, SOC, LT, LR, LFI, LFI et LREM. Finalement **deux partis minoritaires** (LFI avec 17 députés et LT avec 18 députés) sont **particulièrement actifs dans les débats autour des sujets environnementaux**.

On note que les deux députés les plus actifs sur ces sujets sont **deux femmes**, et de loin. Rappelons que sur les 649 députés, 263 sont des femmes, soit une proportion de 41%. Notons qu’il y a un biais ici car Barbara Pompili a eu l’occasion de s’exprimer à l’Assemblée nationale de juillet 2020 à mai 2022 en tant que Ministre de la transition écologique, et pas en tant que simple députée.

5 Extraction des lois environnementales et des votes des députés

Après avoir analysé la prise de parole des députés, nous avons cherché à **comparer cet engagement discursif avec leur comportement lors des votes**. Pour cela, nous avons extrait et traité les données de scrutin disponibles sur le portail Open Data de l’Assemblée nationale. L’objectif est d’**identifier les textes de loi en lien avec l’environnement** puis d’analyser pour chacun la position adoptée par les différents députés.

Les données de scrutin sont relativement bien structurées, mais elles posent **plusieurs problèmes pratiques**. D’abord, elles ne contiennent pas l’identifiant du document

soumis au vote. Il a donc fallu passer par les dossiers législatifs, à partir d'une **autre base de données** de l'Assemblée nationale, qui recense d'un côté les métadonnées des documents (en l'occurrence, ceux portant sur les textes) et de l'autre les dossiers législatifs, lesquels contiennent les identifiants des scrutins qui leur sont associés. Pour chaque document législatif, nous avons récupéré l'identifiant du dossier, puis les scrutins associés, en choisissant celui qui avait mobilisé le plus grand nombre de votants en cas de doublon. Enfin, nous avons tenté de récupérer le contenu du texte via l'URL construite à partir de son identifiant. Certaines requêtes ont échoué, probablement en raison de versions inaccessibles ou de textes abandonnés en cours de procédure.

Une fois ce lien établi entre textes et scrutins, nous avons compté le nombre d'occurrences de mots liés à des enjeux environnementaux dans chaque texte. Pour ne pas inclure des lois très générales (comme les lois de finances) contenant simplement quelques mentions ponctuelles, nous avons défini deux seuils : **un seuil absolu d'occurrences (au moins 30) et un seuil relatif**, basé sur la densité des mots-clés environnementaux par rapport à la taille du texte. Cette sélection a permis d'**isoler une centaine de textes que nous avons jugés pertinents comme relatifs à l'environnement**.

Pour chaque scrutin associé à ces textes, nous avons ensuite extrait la liste des députés ayant participé au vote. La participation était codée dans les fichiers selon différents statuts (pour, contre, abstention, non votant). Dans un premier temps, nous avons retenu uniquement le fait de participer ou non au vote, **quel que soit le sens du vote**, en considérant que cela constitue déjà une **forme d'engagement**.

Ces données ont été compilées dans un tableau récapitulatif indiquant, pour chaque député, le **nombre total de votes relatifs à des lois environnementales** auxquels il ou elle a participé. Ce tableau, croisé avec les clusters définis précédemment, nous a permis de **tester statistiquement s'il existe des différences significatives de participation entre clusters**, et de mettre en regard ces résultats avec l'analyse de leurs prises de parole.

6 Détermination des députés les plus influents et actifs

Afin de déterminer les députés les plus influents dans leur travail parlementaire, on adopte un critère de **centralité** dans le graphe précédemment construit. Après quelques essais, nous écartons certaines métriques. La *degree centrality*, qui mesure simplement le nombre de voisins, ne tenait pas compte de la pondération des arêtes (le nombre d'amendements cosignés). L'*eigenvector centrality* n'a pas convergé, probablement en raison de la structure trop dense du graphe.

Nous avons finalement retenu la ***betweenness centrality***, qui mesure à quel point un député se trouve sur les plus courts chemins reliant d'autres députés. Pour que les poids des arêtes soient correctement pris en compte, il était nécessaire d'**inverser les pondérations des arêtes**.

Le tableau ci-dessous présente les dix députés avec la centralité la plus élevée, accompagnés de leur nombre de participations aux votes (tous textes confondus).

Député	Centralité ($\times 10^2$)	Participations	Classement participation	Cluster
Emmanuelle Anthoine	3.31	653	240	2
Valérie Bazin-Malgras	3.28	583	370	2
Josiane Corneloup	0.82	544	408	2
Véronique Louwagie	0.58	545	407	2
Patrick Vignal	0.57	601	354	6
Bruno Bilde	0.54	444	515	3
Patrick Hetzel	0.50	564	390	2
Christine Pires Beaune	0.31	625	307	0
Emmanuelle Ménard	0.31	745	71	7
Edith Audibert	0.29	186	610	2

Table 1: Classement des députés selon leur centralité

En parallèle, nous avons extrait les dix députés ayant participé au plus grand nombre de scrutins sur l'ensemble des textes de la législature. Nous indiquons pour chacun sa centralité dans le graphe et son classement selon cette mesure.

Député	Participations	Centralité ($\times 10^4$)	Classement centralité	Cluster
Michèle Peyron	1025	0.0004	437	6
Jean-Luc Fugit	928	0.3816	277	6
Catherine Osson	909	0.1125	337	6
Michèle de Vaucouleurs	907	1.5677	181	5
Guillaume Gouffier-Cha	894	0.0477	404	6
Jean-Jacques Bridey	883	0.0004	442	6
Marie-Christine Verdier-Jouclas	863	0.8911	227	6
Marion Lenne	861	1.0018	218	6
Aude Amadou	848	0.0	548	6
Stéphane Testé	843	7.2986	64	6

Table 2: Députés ayant le plus fort taux de participation aux scrutins, avec leur centralité

Cette double lecture permet de comparer deux formes d'activité parlementaire : d'une part, la centralité dans les réseaux de rédaction d'amendements ; d'autre part, l'assiduité aux scrutins. Les députés les plus influents dans le graphe ne sont pas nécessairement ceux qui votent le plus, et inversement. Par exemple, Michèle Peyron, en tête en nombre de participations, a une centralité quasi nulle. Pour confirmer cette hypothèse, on peut utiliser un test de **corrélation de Pearson** (où l'hypothèse nulle est celle d'une dé-corrélation des données).

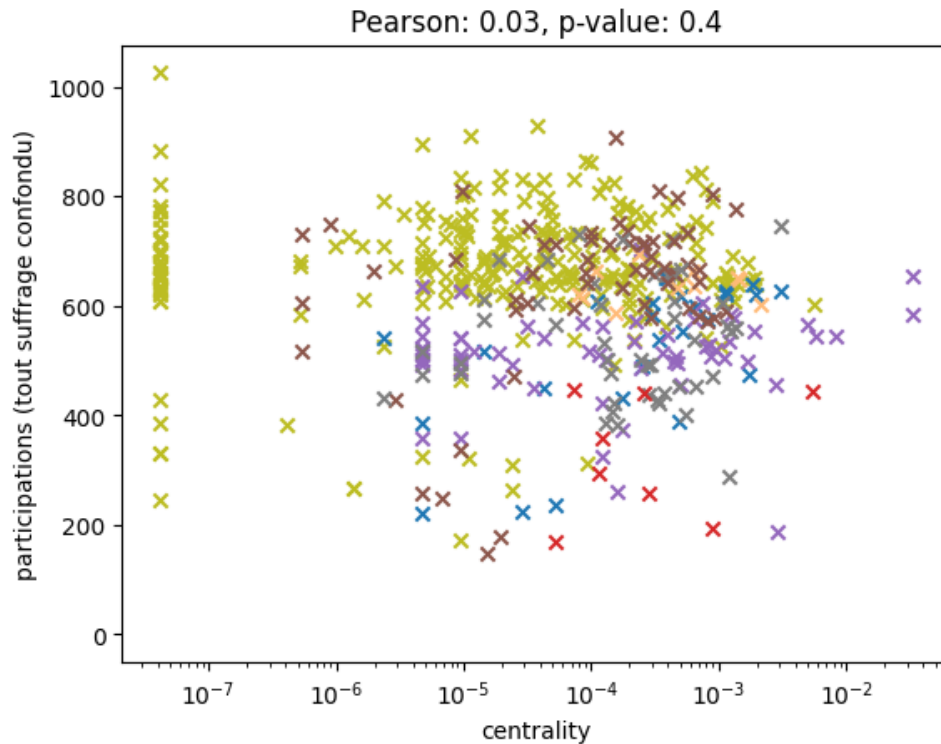


Figure 5: Représentation de chaque député sur le plan (centralité, participations), tous votes confondus. L'échelle en abscisse est logarithmique.

La **p-value** du test étant bien supérieure à 0.05 (0.4 exactement), on maintient l'hypothèse nulle : on considérera ces deux critères comme décorrélés. **L'engagement dans les urnes ne se recoupe pas avec la centralité des députés.**

7 Analyse sur les votes et les discours environnementaux selon le cluster

7.1 Hypothèses à tester et cadre de validité du test Anova

On cherche dans cette partie à tester les **trois hypothèses** suivantes.

- Hypothèse 1 : le nombre de votes relatifs à des lois environnementales ne dépend pas du cluster.
- Hypothèse 2 : la place de l'environnement dans les discours ne dépend pas du cluster.
- Hypothèse 3 : le rapport discours/vote ne diffère pas selon les clusters.

Pour appliquer le **test ANOVA (Analyse de la variance)**, les échantillons doivent satisfaire les conditions suivantes :

- **Indépendance** des échantillons.
- **Normalité des populations** : Les données de chaque échantillon doivent provenir de populations qui suivent une distribution normale.

- **Homoscédasticité (égalité des variances)** : Les variances des populations doivent être approximativement égales. Cette condition garantit que la dispersion des données est comparable d'un groupe à un autre, permettant une comparaison équitable entre eux.

Les **deux derniers points** nécessitent d'être vérifiés. On utilise **Shapiro-Wilk** pour tester la normalité des échantillons ainsi que **Levene** pour tester l'égalité des variances.

On applique ensuite le test Anova si les deux dernières hypothèses sont bien vérifiées pour les votes portant sur des lois environnementales ainsi que sur les occurrences liées à des sujets environnementaux.

7.2 Analyse des votes portant sur des lois environnementales selon le cluster

On construit les **8 listes contenant le nombre de votes relatifs aux lois environnementales pour chaque député** de tel ou tel cluster.

✓Hypothèse de normalité vérifiée pour les 7 premiers clusters mais pas le huitième. Pour les 7 premiers clusters, la p-value est inférieure à 0,05 mais, pour le cluster mixte, la p-value vaut 0,08, ce qui reste acceptable.

✓Test de Levene concluant. Pour le détail, on obtient : `LeveneResult (statistic = np.float64(3.2987), pvalue = np.float64(0.0019))`. La p-value est bien inférieure à 0,05.

On applique donc Anova et on obtient une **p-value de 10^{-78} et une statistique F de 74**.

La p-value minuscule signifie qu'il y a une différence extrêmement significative entre les groupes. De plus, la F statistique mesure le ratio entre la variance entre les groupes et la variance au sein des groupes. On obtient une F statistique grande devant 1, ce qui corrobore la **différence statistiquement significative entre les clusters quant au nombre de votes portant sur des lois environnementales**.

Par exemple, notons qu'un simple calcul des moyennes nous donne respectivement pour les 8 clusters : 45, 36, 30, 8, 45, 56, 58 et 34. Une participation qui peut donc varier d'un facteur 7 ! Anova nous permet d'étudier plus en détail la répartition entre clusters.

Pour aller plus loin dans l'analyse et cerner les clusters qui se démarquent, on opte pour le **test de Tukey**. On choisit de représenter **graphiquement** plutôt que dans un tableau les résultats du test de Tukey.

La lecture du graphique se fait comme suit. **Si un intervalle en chevauche un autre, cela signifie qu'il n'y a pas de différence significative entre les groupes**. Par exemple, le cluster 1 chevauche les clusters 2 et 5 uniquement. L'intervalle autour de la ligne est l'intervalle de confiance à 95%. Par exemple, le cluster 1 a une moyenne de votes plus élevée que le cluster 3 avec un niveau de confiance de 95%.

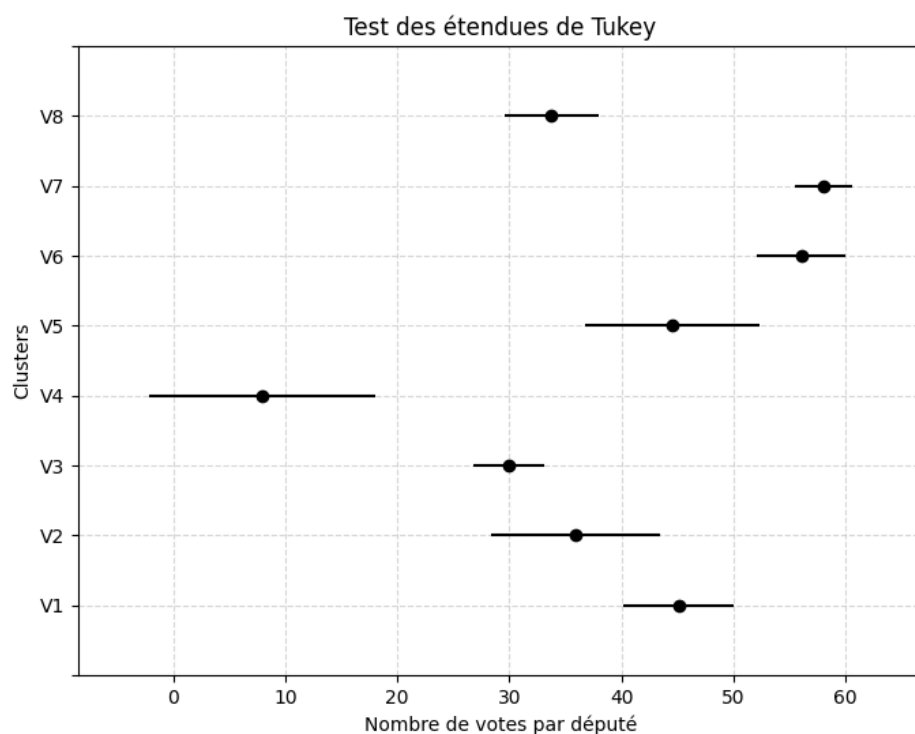


Figure 6: Test des étendues de Tukey pour le nombre de votes pour les lois liées à l'environnement

Pour affiner cette représentation, on utilise la bibliothèque *seaborn* pour une représentation en forme de violon.

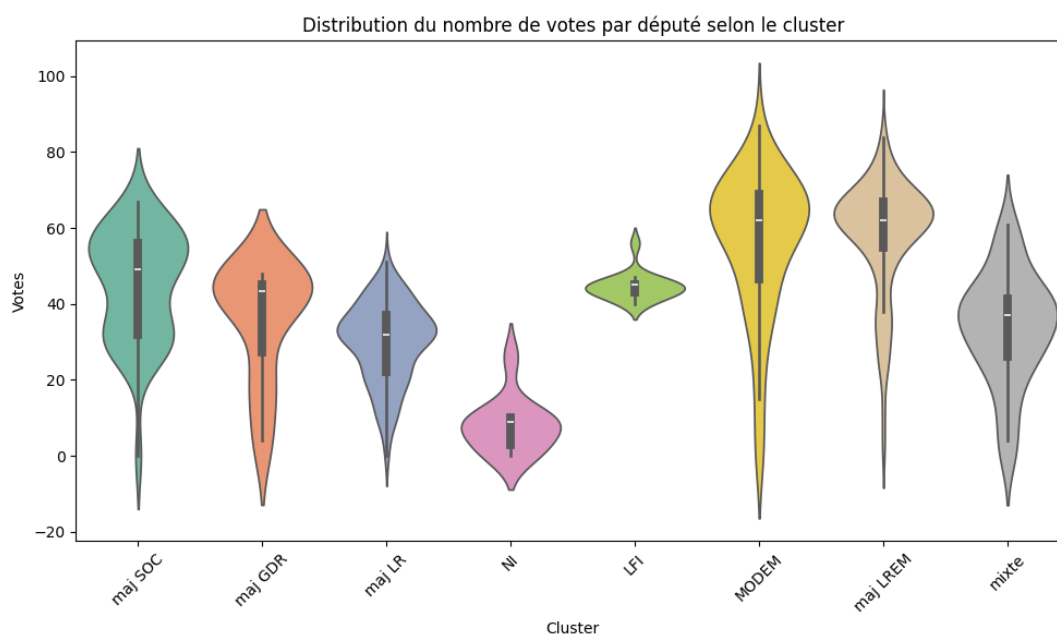


Figure 7: Diagramme en violon du nombre de votes pour les lois liées à l'environnement

Les différents éléments du *violin plot* peuvent s'interpréter de la façon suivante :

- Médiane haute \Rightarrow Cluster globalement engagé
- Médiane basse \Rightarrow Faible mobilisation environnementale du cluster
- Distribution fine \Rightarrow Cohérence interne forte
- Distribution large \Rightarrow Variabilité importante des comportements
- Valeurs extrêmes \Rightarrow Présence de députés atypiques dans leurs votes

On peut l'analyser comme suit :

Cluster	Taille	Médiane votes env.	Cohérence interne
maj SOC	39	Moyenne-élevée	Faible
maj GDR	18	Moyenne-élevée	Moyenne
maj LR	119	Moyenne	Moyenne
NI	11	Très faible	Forte
LFI	17	Moyenne-élevée	Très forte
MODEM	63	Élevée	Moyenne
maj LREM	328	Élevée	Forte
Mixte	55	Moyenne	Faible

Table 3: Résumé comparatif des clusters selon les votes de lois environnementales

On remarque notamment que LFI et LREM ont une cohérence interne notable, ils **votent en bloc sur les sujets environnementaux**. A l'inverse, le MODEM a une distribution **beaucoup plus large, presque bimodale**, signe d'une faible cohérence dans l'engagement des députés quand il s'agit de voter. Il apparaît donc que l'hypothèse 1 est rejetée : **le nombre de votes relatifs à des lois environnementales dépend du cluster**, tant dans sa médiane que dans sa distribution.

7.3 Analyse des occurrences liées à des sujets environnementaux selon le cluster

On mène le même travail pour les occurrences liées à des sujets environnementaux. On construit les **8 listes contenant le nombre d'occurrences autour de ces enjeux** pour tel ou tel cluster.

✓Hypothèse de normalité vérifiée pour les 8 clusters.

✓Test de Levene concluant : LeveneResult (statistic = np.float64(4.9189), pvalue = np.float64(1.9860e-05)). La p-value est largement inférieure à 0,05.

On applique donc Anova et on obtient une **p-value de 3.10^{-13} et une statistique F de 11**. La différence de répartition des propos environnementaux est donc significative entre clusters. Avec Tukey, on cherche à mettre en évidence les différences entre clusters en termes d'occurrences de mots environnementaux.

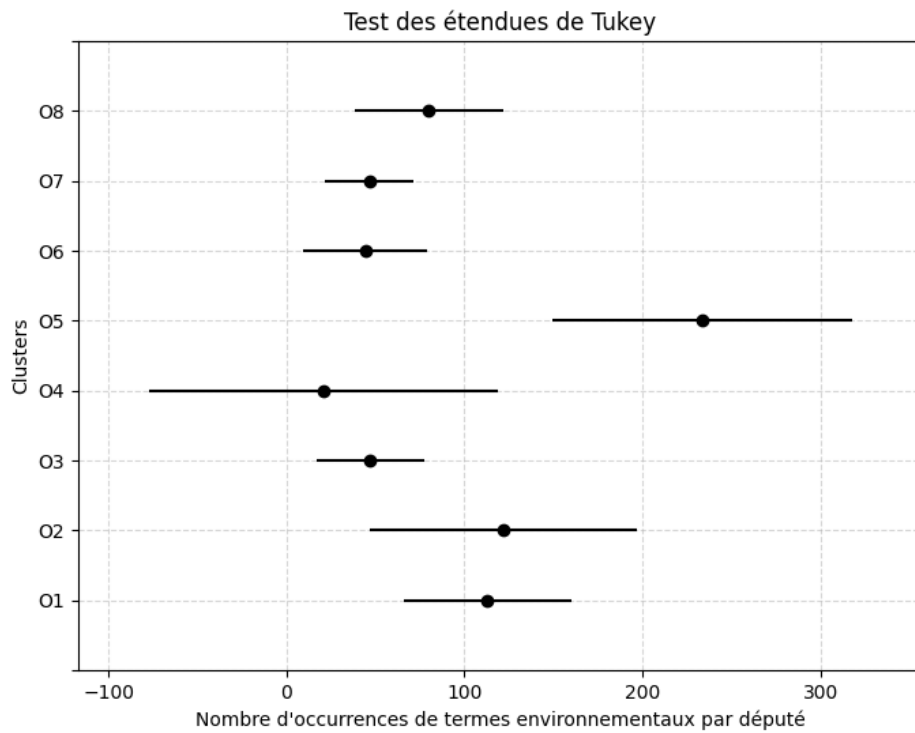


Figure 8: Test des étendues de Tukey pour le nombre d'occurrences liées à l'environnement

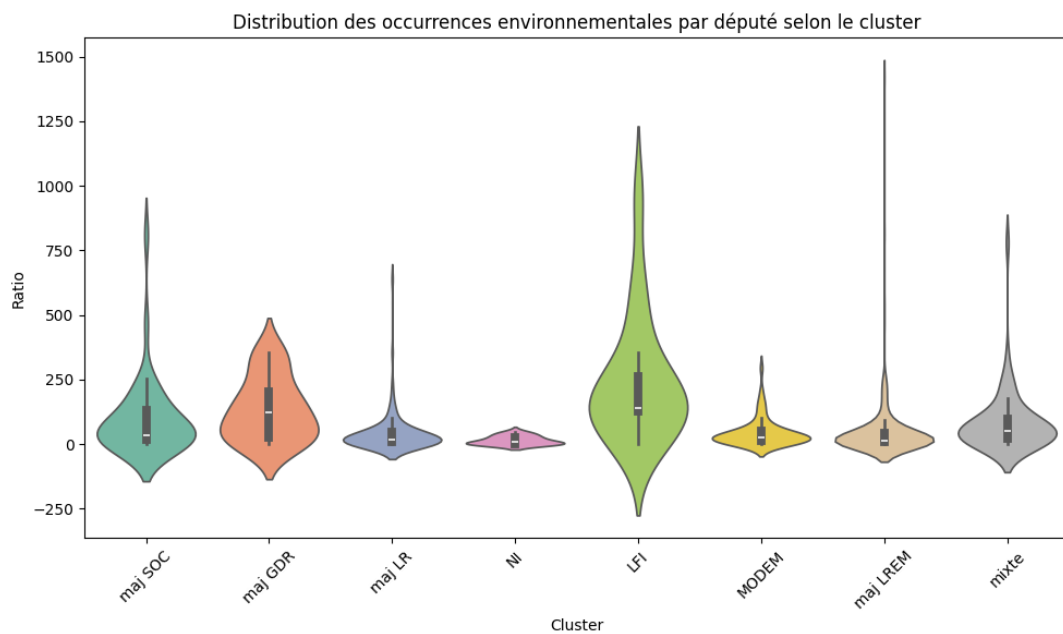


Figure 9: Diagramme en violon pour le nombre d'occurrences liées à l'environnement

Le test des étendues de Tukey met en évidence un engagement environnemental dans les débats très marqué chez **LFI**. Sa médiane nettement supérieure aux autres clusters traduit un engagement environnemental discursif **important**. Trois clusters, à savoir ceux à majorité **LR**, **MODEM** et **LREM**, se distinguent par leur cohérence interne. La

ligne partagée se retrouve par une **participation homogène** dans les débats environnementaux. Mis à part les non inscrits, ces trois clusters ont un **engagement discursif environnemental particulièrement faible**. Le **cluster mixte**, composé de divers petits partis (UDI, LT, AE), montre une cohérence moyenne avec un engagement modeste, certainement lié à une priorité accordée à d'autres thématiques dans leur agenda politique.

Le test Anova et ce diagramme révèlent une **forte hétérogénéité dans l'investissement discursif des députés sur les enjeux environnementaux**. On peut donc rejeter l'hypothèse 2 : **la place de l'environnement dans les discours dépend du cluster**.

7.4 Analyse du rapport vote/discours pour les enjeux environnementaux

On mène le même travail pour les occurrences liées à des sujets environnementaux. On calcule pour chaque député le ratio votes sur des lois environnementales/occurrences liées à des enjeux environnementaux lors des débats.

✓Hypothèse de normalité vérifiée pour les 8 clusters.

✓Test de Levene concluant : LeveneResult (statistic = np.float64(5.4127), pvalue = np.float64(4.7846e-06)). La p-value est largement inférieure à 0,05.

On peut donc appliquer Anova et on obtient une **p-value de 9.10^{-8}** et une **statistique F de 6,8**. La différence de répartition du ratio votes environnementaux/occurrences environnementales est donc significative entre clusters. L'hypothèse 3 est rejetée.

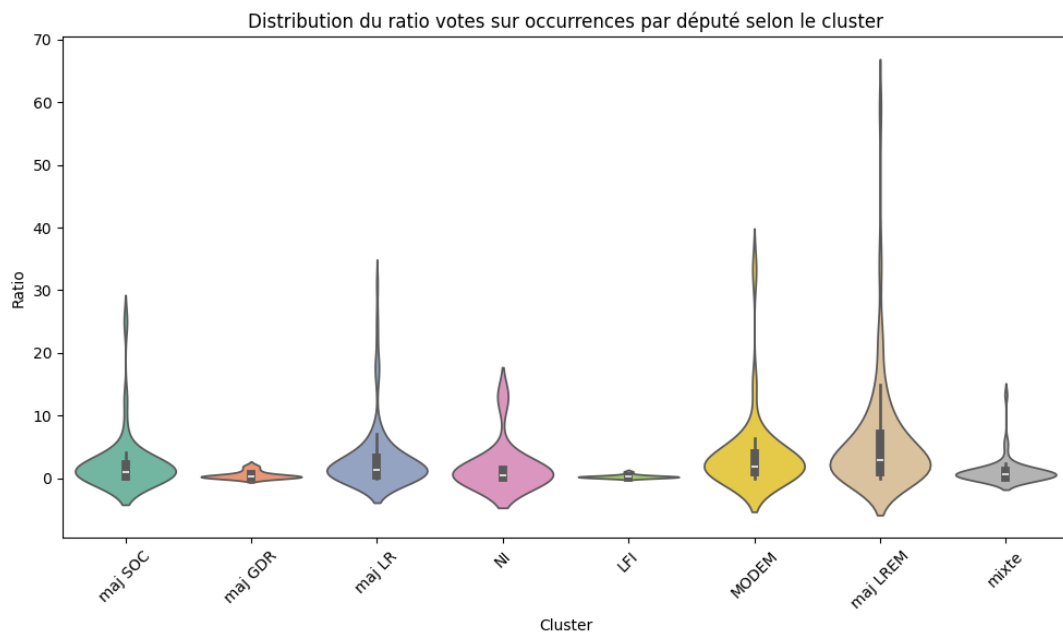


Figure 10: Diagramme en violon pour le rapport votes sur nombre d'occurrences pour les sujets environnementaux

On utilise à nouveau le diagramme en violon mais, cette fois-ci, du fait des maxima beaucoup plus marqués, le diagramme est moins lisible.

Pour plus de lisibilité, on met ici les valeurs des moyennes : **2.31, 0.54, 3.03, 1.85, 0.29, 3.74, 6.1 et 1.12** respectivement. Les députés du **cluster à majorité LREM** ont tendance à voter beaucoup de lois environnementales relativement à la place qu'ils accordent aux sujets environnementaux dans leurs prises de parole. Le ratio est en effet **21 fois plus élevé que le cluster 100% LFI**. On peut le comprendre de la façon suivante : les députés LFI, étant **minoritaires, occupent beaucoup l'espace de parole** en s'exprimant fréquemment sur des sujets environnementaux. Le cluster LREM **vote en bloc** pour les lois environnementales mais sans qu'il ait besoin d'occuper autant l'espace de débat, leur nombre leur garantissant que les lois pour lesquelles ils votent pour soient validées. Les **non inscrits** ont tendance à parler très peu de ces sujets. On note que la dispersion au sein des non inscrits, malgré le fait qu'ils ne soient que 11, est importante. A l'inverse, des clusters comme le MODEM et ceux à majorité LR et LREM sont **cohérents** pour la majorité des députés.

Finalement, on a mis en évidence que **le rapport entre vote environnemental et discours environnemental dépend du cluster**. Une analyse plus en détail nous montre les **différentes stratégies politiques adoptées à l'Assemblée vis-à-vis des enjeux environnementaux**.

8 Analyses statistiques et interprétations

On se propose ici de poursuivre l'analyse en étudiant certaines corrélations éventuelles via des tests statistiques. Avec la précédente extraction de données, nous avons à notre disposition un **jeu de données tabulaire** (cf. figure 10), sur lequel nous pouvons **tester des hypothèses supplémentaires**.

	IDENT	nom	prenom	cluster	participations	centrality	participations_climat	occurrences_env	taille_parole
0	PA721710	Peyron	Michèle	6	1025	4.148135e-08	83	5	120166
1	PA722366	Fugit	Jean-Luc	6	928	3.816284e-05	87	249	367572
2	PA720492	Osson	Catherine	6	909	1.124961e-05	76	21	193257
3	PA721924	de Vaucouleurs	Michèle	5	907	1.567744e-04	87	50	575291
4	PA721296	Gouffier-Cha	Guillaume	6	894	4.770355e-06	78	14	441374
...
644	PA783463	Métadier	Sophie	7	34	0.000000e+00	0	17	40146
645	PA720668	Blairy	Emmanuel	3	28	0.000000e+00	0	0	6630
646	PA719914	Darrieussecq	Geneviève	5	0	0.000000e+00	0	44	474740
647	PA563	Bompard	Jacques	7	0	0.000000e+00	0	0	5496
648	PA720182	Cesar	Pascale	5	0	0.000000e+00	0	0	1988

649 rows × 9 columns

Figure 11: Données tabulaires extraites

Rappel charte de couleurs



8.1 Corrélation avec la participation totale aux votes

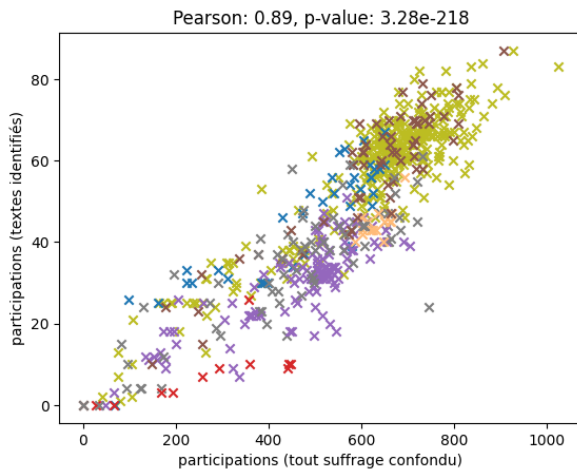


Figure 12: Plan (participation totale, participation aux scrutins cibles).

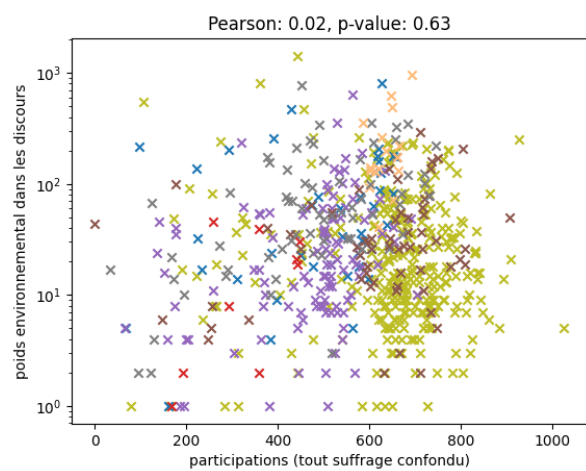


Figure 13: Plan (participation totale, occurrence dans les discours). L'échelle des ordonnées est logarithmique.

On effectue ici deux tests de corrélation : **entre la participation (tous textes confondus) et la participations aux scrutins ciblés (càd portant sur des lois environnementales)** ainsi qu'**entre la participation et les occurrences de termes environnementaux dans les débats**. On garde la même couleur afin de repérer les clusters (cf. charte de couleurs).

Les résultats mettent en évidence des dynamiques très différentes. On observe d'abord

une forte **corrélation linéaire** ($r = 0.89$, $p < 10^{-200}$) entre la participation totale aux scrutins et la participation aux votes identifiés comme environnementaux (figure 11). Cela montre que les députés assidus de manière générale le sont aussi sur les textes liés à l’environnement ; cette tendance reflète avant tout une **logique d’assiduité et de concordance avec l’engagement dans les scrutins du cluster auquel le député appartient**, sans que cela n’indique nécessairement un intérêt particulier pour les enjeux écologiques.

À l’inverse, la **corrélation entre participation générale et fréquence d’utilisation de termes environnementaux dans les discours est quasi nulle** ($r = 0.02$, $p = 0.63$) (figure 12). Cela suggère que la prise de parole sur ces sujets ne dépend pas du niveau d’activité parlementaire global. Certains députés très présents en séance n’évoquent presque jamais l’environnement, tandis que d’autres, moins actifs sur les scrutins, y font souvent référence dans leurs interventions. **La parole sur les enjeux écologiques semble donc relever d’un positionnement politique, plutôt que d’une simple présence ou régularité dans le travail parlementaire.**

8.2 Corrélation avec la centralité

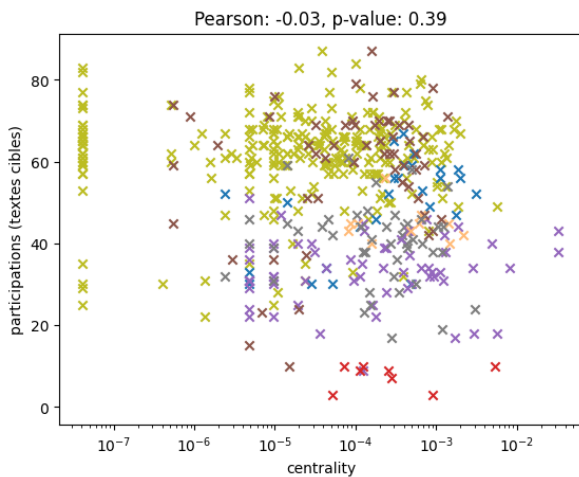


Figure 14: Plan (centralité, participation aux scrutins cibles).

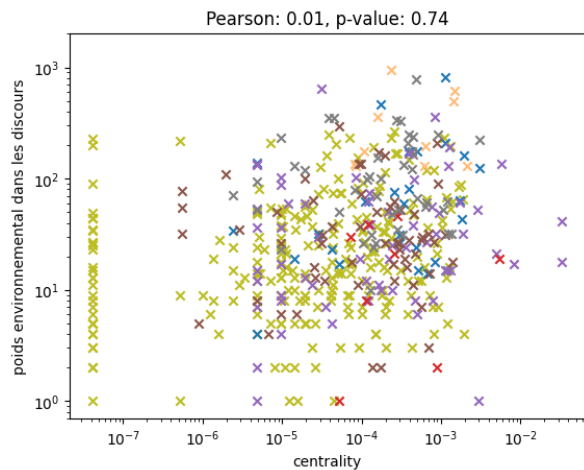


Figure 15: Plan (centralité, occurrence dans les discours).

Les deux figures ci-dessus examinent la corrélation entre la centralité des députés dans le graphe de co-signatures d’amendements (mesurée par la *betweenness centrality*) et deux formes d’engagement sur les enjeux environnementaux. Dans les deux cas, les résultats sont sans appel : aucune corrélation significative n’est observée. **La centralité ne permet ni de prédire la participation aux votes ciblés** ($r = -0.03$, $p = 0.39$, figure 14) **ni le poids environnemental des discours** ($r = 0.01$, $p = 0.74$, figure 15).

Ces résultats invitent à **nuancer la pertinence de la centralité comme indicateur d’influence ou d’engagement politique**. La *betweenness centrality*, en particulier, repose sur l’hypothèse que les nœuds les plus souvent situés sur les plus courts chemins jouent un rôle important dans le réseau. Mais dans le cas du graphe de co-signature, cette interprétation peut être discutable : les liens reflètent des signatures

conjointes d'amendements, **souvent dictées par des stratégies collectives** (notamment par parti politique), et non des relations personnelles fortes. Il est donc possible que d'autres mesures de centralité ou des approches plus qualitatives soient plus adaptées pour caractériser l'influence réelle de chaque député.

8.3 Corrélation entre parole environnementale et vote environnemental

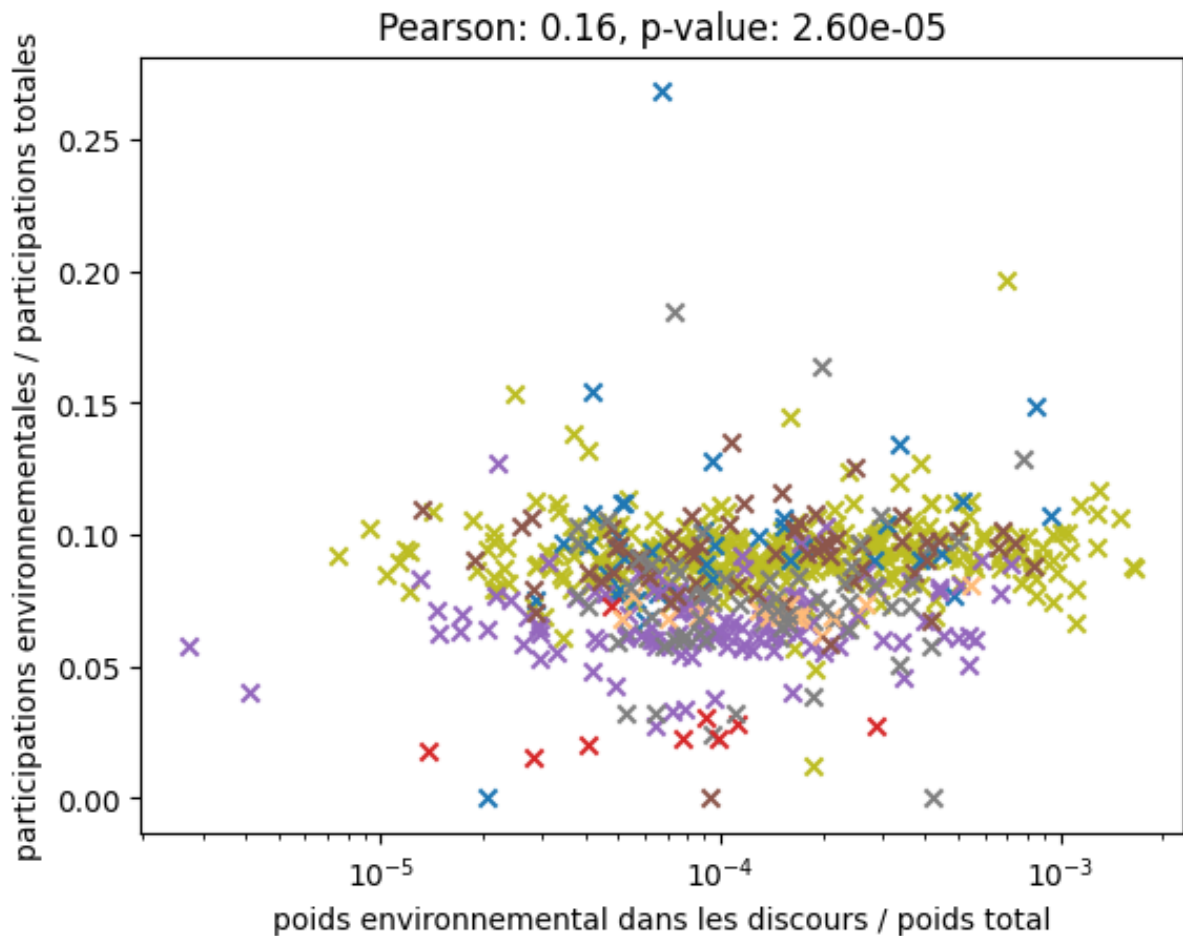


Figure 16:

Enfin, dans un dernier test, nous avons cherché à évaluer une éventuelle corrélation entre deux indicateurs **relatifs** d'engagement. Le premier est le poids environnemental dans les discours, rapporté au temps de parole total de chaque député. Le second est le taux de participation aux scrutins ciblés, rapporté à la participation totale tous scrutins confondus. Ces deux ratios visent à isoler des formes d'engagement associées aux enjeux environnementaux, indépendamment de l'activité parlementaire.

Le test de corrélation de Pearson donne une p-value inférieure à 0,05, ce qui indique une corrélation statistiquement significative. Cependant, le coefficient de corrélation reste

faible (de l'ordre de 0,1), ce qui limite la portée de ce lien. D'un point de vue graphique, la répartition du taux de participation environnementale semble globalement constante selon le ratio de discours environnementaux. En d'autres termes, un député qui consacre une grande partie de son temps de parole à l'environnement ne vote pas nécessairement davantage que les autres sur les textes en lien avec ces sujets.

Cette dissociation entre discours et vote se confirme ici. Elle renforce l'idée que **la parole et l'action parlementaire sur l'environnement obéissent à des logiques assez différentes**.

9 Conclusion

Cette étude, basée sur différentes bases de données publiques, a ainsi permis de mettre en lumière certains résultats sur la place des enjeux environnementaux à l'Assemblée, **souvent compris intuitivement mais quantitativement non étayés**.

Parmi eux, on trouve notamment des **écarts d'engagement significatifs** entre les différents clusters politiques. Certains groupes, comme LFI, prennent largement la parole sur les sujets environnementaux, tandis que d'autres, comme LREM, se distinguent davantage par leur participation aux votes. Cette **dissociation entre discours et action** est confirmée par l'absence de corrélation forte entre les deux indicateurs. Par ailleurs, nous avons montré que la participation aux scrutins environnementaux est très corrélée à la participation générale, ce qui n'est pas le cas de la prise de parole, plus liée à un positionnement politique qu'à l'assiduité. Nous avons également analysé les **spécificités de chaque cluster en termes de prise de position environnementale dans les débats et les scrutins**.

La centralité dans le graphe de co-signature n'est pas non plus liée à un engagement particulier sur les enjeux environnementaux. Ces éléments confirment que la parole et l'action par le vote relèvent de dynamiques distinctes : l'engagement écologique à l'assemblée peut prendre des **formes variées selon les groupes et les individus**. Notre analyse fait ressortir les **stratégies politiques liées aux enjeux environnementaux** au sein des différents groupes de l'Assemblée.

En guise de continuation de notre travail, on pourrait mener cette même étude sur chacune des dernières législatures. Ceci permettrait de discuter de **l'évolution de la place des sujets environnementaux** à l'Assemblée nationale au sein des différents partis ainsi qu'au niveau individuel. On pourrait également envisager une analyse plus fine des termes utilisés pour parler des enjeux environnementaux afin de détecter un éventuel éco-blanchiment.