

De quoi est faite Wikipédia ?

Quelques pistes pour une analyse de sa régulation au prisme des effacements
réciproques.

Léo Joubert

MCF sociologie à l'Université de Rouen

DySoLab

Laboratoire des dynamiques sociales

Plan

1. Problématique et source de données

- a. La notion de régime de couplage flou
- b. WPTokens

2. Analyse des réseaux

- a. Des régimes de couplage aux sous-corpus
- b. La construction des « graphes de suppression »
- c. La notion de forme

3. Typologie des régimes de couplage flou

- a. Deux analyses en composantes principales (ACP)
- b. Sur le biais de genre
- c. Sur l'exposition médiatique

4. Conclusion



WIKIPÉDIA
L'encyclopédie libre

[Accueil](#)
[Portails thématiques](#)
[Index alphabétique](#)
[Article au hasard](#)
[Contacter Wikipédia](#)

- ▶ [Contribuer](#)
- ▶ [Imprimer / exporter](#)
- ▶ [Boîte à outils](#)

[Article](#) [Discussion](#) [Lire](#) [Modifier](#) [Afficher l'historique](#)

Pomme



Problématique et sources de données

La notion de régime de couplage flou

« Nos sociétés se caractérisent généralement par un 'couplage flou' entre des pratiques interactionnelles et les structures sociales, [...] un ensemble de règles de transformation, une membrane sélectionnant la manière dont diverses distinctions sociales, extérieurement pertinentes, seront prises en compte au cours de l'interaction »

(Goffman 1988)

Hypothèses de travail :

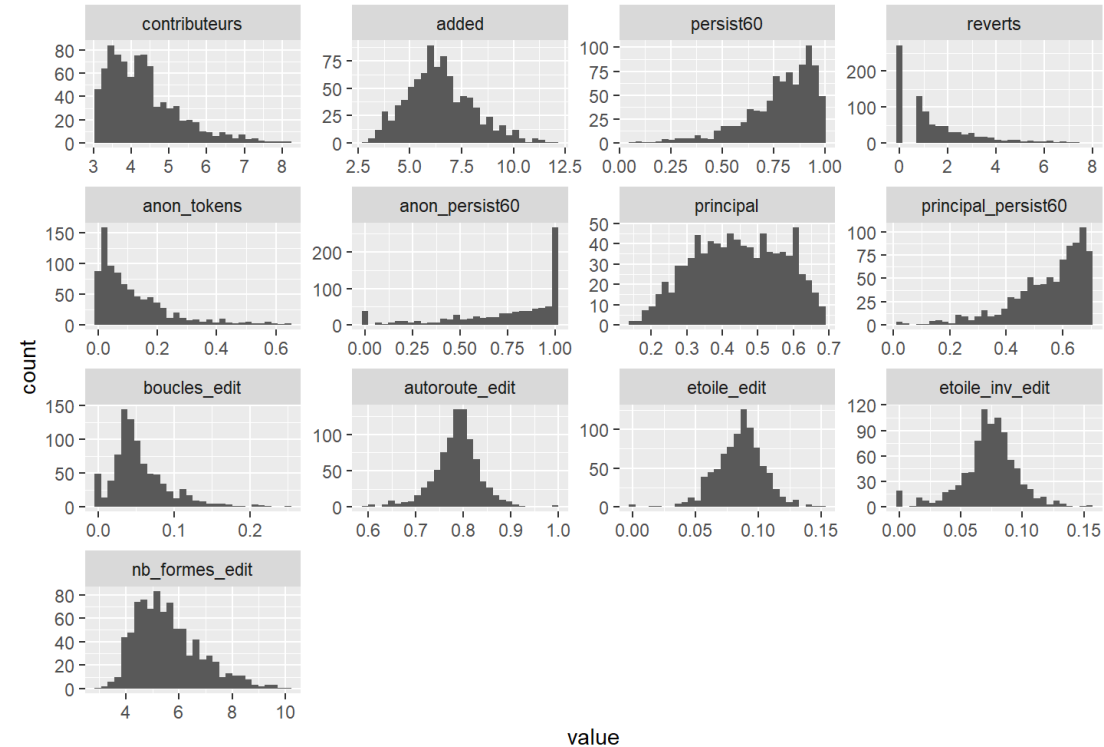
1. C'est en détaillant les flux de coécriture que l'on peut dégager une série de manière dont la « membrane » se constitue.
2. Chaque sous-espace de Wikipédia possède son propre « régime » de couplage

Problématique et sources de données

WPtokens

```
<page>
<title>AccessibleComputing</title>
<ns>0</ns>
<id>10</id>
<redirect title="Computer accessibility" />
<revision>
...
<contributor>
...
</contributor>
</revision>
</page>
```

Déwikification
Suppression des
mots vides



Code pour l'application du *diff* aux *dumps* et la construction de la base de données

<https://gitlab.huma-num.fr/ljoubert/wptokens>

Analyse des réseaux

Les sous-corpus

Objet de la comparaison	Sous-corpus		Sous-corpus complémentaire	
<i>Biais de genre</i>	Femmes docteurs		Hommes docteurs	
<i>Exposition médiatique</i>	Candidat.es aux élections présidentielles de 2007 à 2022		Présidents du Conseil	

Chaque sous-corpus a un « complémentaire » avec lequel on le compare pour créer un contraste

Analyse des réseaux

Des régimes de couplage flou aux sous-corpus

Objet de la comparaison	Sous-corpus	N	Sous-corpus complémentaire	N
<i>Biais de genre</i>	Femmes docteurs	94	Hommes docteurs	679
<i>Exposition médiatique</i>	Candidat.es aux élections présidentielles de 2007 à 2022	26	Présidents du Conseil	65

Chaque sous-corpus a un « complémentaire » avec lequel on le compare pour créer un contraste

Analyse des réseaux

Construction des « graphes de suppression »

Chaque contributeur est un nœud

- Dont la taille est proportionnelle à la quantité de contenu ajouté

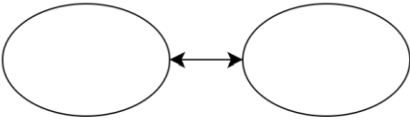
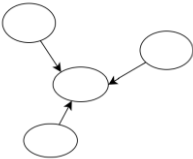
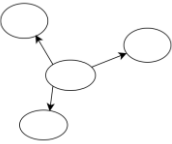

Une arête existe entre A et B si A a effacé du contenu écrit par B (minimum de 50 mots)

- Dont la taille est proportionnelle à la quantité de contenu supprimé

Graphe orienté valué

Analyse des réseaux

Construction des « graphes de suppression »

<i>Forme</i>	<i>Nom</i>	<i>Interprétation</i>
	<i>Autoroute</i>	Deux contributeurs effacent mutuellement leurs ajouts
	<i>Etoile inversée</i>	Plusieurs contributeurs effacent les contributions d'un seul
	<i>Etoile</i>	Un contributeur efface les contributions d'un autre
	<i>Boucle</i>	Un contributeur efface ses propres contributions

Typologie des couplages flous

Analyses en composantes principales

Variables de description du flux de coécriture :

- nombre de mots ajoutés (**added**)
- nombre de contributeurs (**contributeurs**)
- % de mots persistants après 60 jours (**persist60**)
- nombre de reverts (**reverts**)
- nombre de mots ajoutés par des contributeurs « anonymes » (**anons_tokens**)
- % des mots ajoutés par des contributeurs anonymes persistant après 60 jours (**anons_persist60**)
- % de mots ajoutés par le premier contributeur (**principal**)
- % de mots ajoutés par le premier contributeur persistants après 60 jours (**principal_persist**)

Variables de comptage des formes (rapportées au nombre total de formes) :

- autoroutes (**autoroute_edit**)
- étoiles (**etoile_edit**)
- étoiles inversées (**etoile_inv_edit**)
- boucles (**boucle_edit**)

Typologie des couplages flous

Deux analyses en composantes principales

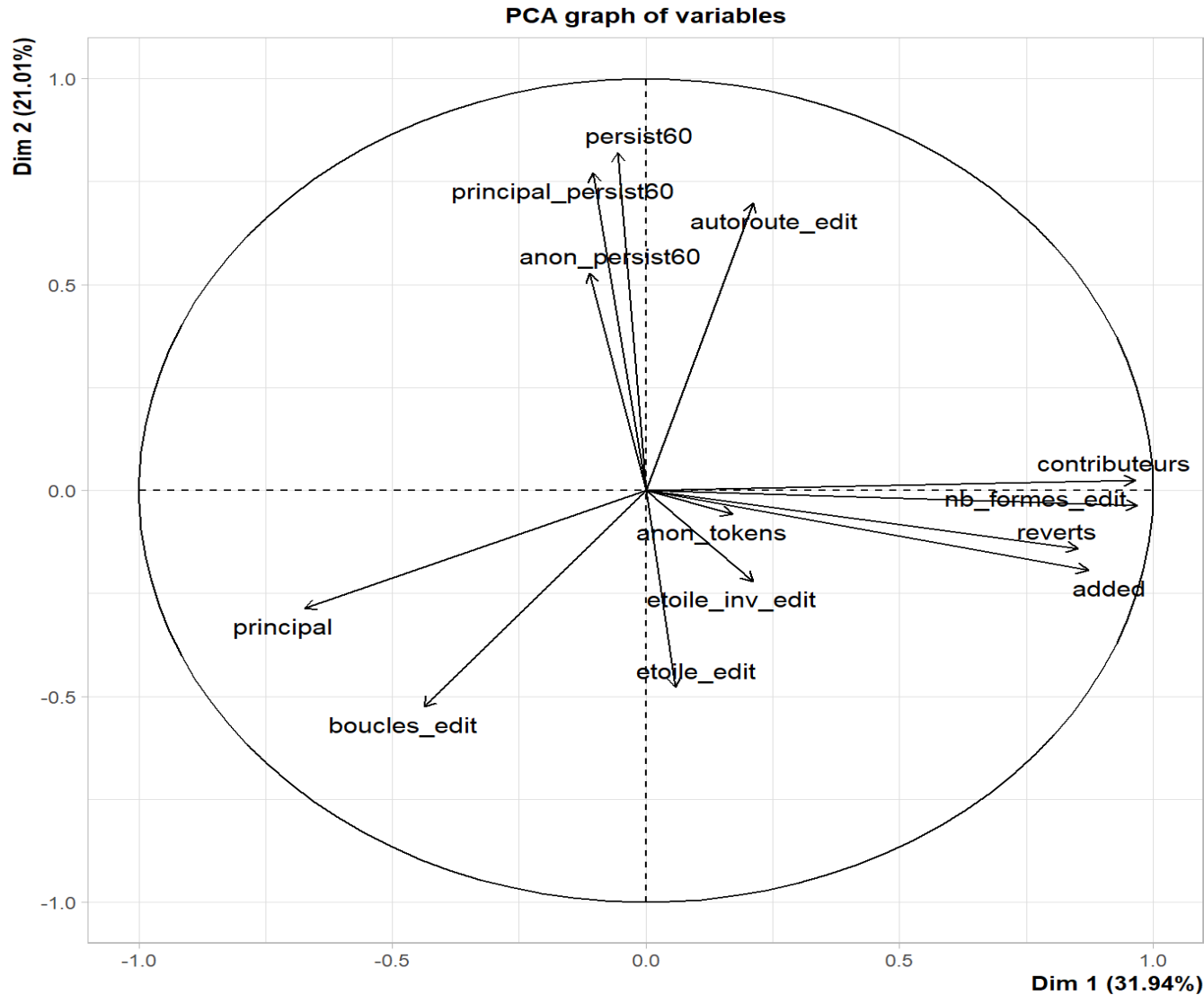
Deux ACP, avec une variable supplémentaire qualitative à deux modalités :

- Appartenance de la page au sous-corpus
- Appartenance de la page au sous-corpus complémentaire

	Sous-corpus	N	Sous-corpus complémentaire	N	Nombre total de pages dans l'analyse
<i>Première ACP</i>	Femmes docteurs	94	Hommes docteurs	679	773
<i>Deuxième ACP</i>	Candidat.es aux élections présidentielles de 2007 à 2022	26	Présidents du Conseil	65	91

Typologie des régimes de couplage

Sur le biais de genre

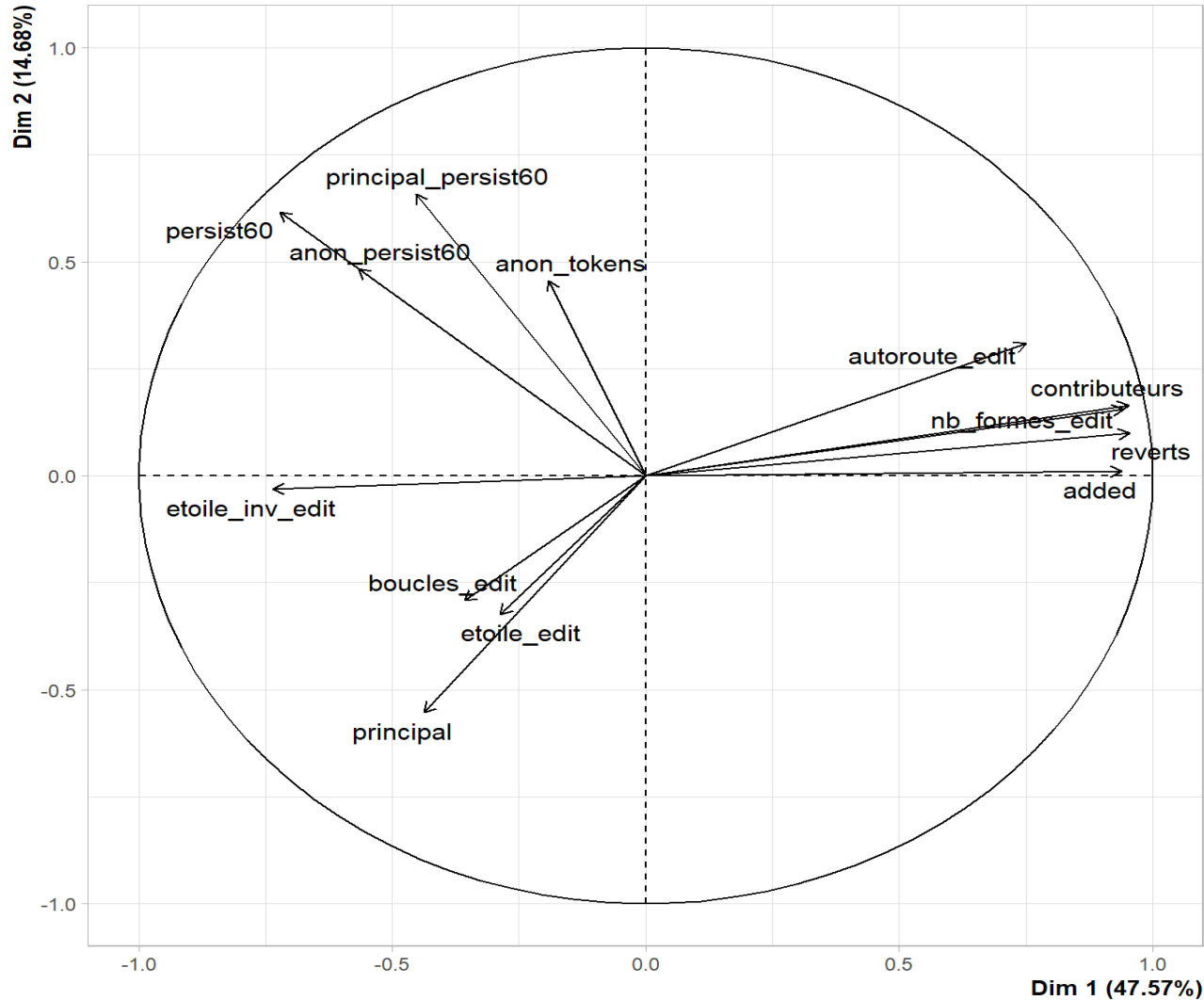


Axe	Modalité	V-test
1	<i>Femmes docteurs</i>	- 0,59
1	<i>Hommes docteurs</i>	+ 0,59
2	<i>Femmes docteurs</i>	- 5,27
2	<i>Hommes docteurs</i>	+ 5,27

Typologie des régimes de couplage

Sur l'exposition médiatique

PCA graph of variables



Axe	Modalité	V-test
1	Candidat.es aux élections présidentielles de 2007 à 2022	+ 6,08
1	Présidents du Conseil	- 6,08
2	Candidat.es aux élections présidentielles de 2007 à 2022	+ 1,14
2	Présidents du Conseil	- 1,14

Pour ne pas conclure...

Partie quantitative : analyse de réseau

- Combiner plusieurs critères dans la construction des sous-corpus et de leur complémentaire
- Intégrer plus directement des outils de la théorie des graphes
- Spécifier les liens à partir de l'analyse du contenu effacé (mots)

Pour ne pas conclure...

Partie qualitative : analyse de contenu (*content analysis*) + analyse de trajectoire (*lifecourse trajectory*)

- Sous-corpus de pages liées à une controverse autour de la naissance du Mouvement de Libération des Femmes (MLF) ; avec Marion Charpenel
 - Histoire du féminisme, Antoinette Fouque, Mouvement de libération des femmes, Le torchon brûle, Editions Des Femmes, Monique Wittig, Différentialisme, Dépôt de gerbe à la femme du Soldat inconnu, Jacqueline Feldman, Anne Zelenskyn, Liliane Kandel, Françoise Picq, Nadja Ringart, Cathy Bernheim, Christine Delphy, Christiane Rochefort, Emmanuèle de Lesseps, Gouines rouges, La Pensée straight, Féminisme matérialiste, Deuxième vague féministe, Sam Bourcier, Caroline Fourest
- Sous-corpus de pages liées à la présentation du personnel politique sur Wikipédia ; avec Jeanne Vermeirsche
 - Emmanuel Macron, Marine Le Pen, Jean-Luc Mélenchon, Dora Moutot, Nationalisme en France, Les Identitaires, Extrême droite, Rassemblement national, Féminisme, Islamisme

Dans ces sous-corpus, trouver des contributeurs à interroger pour les positionner dans le réseau, et analyser ensemble « trajectoire » et « position »