

Etude du réseau ERUA de l'Université Paris 8 & Erasmus

Férielle KHELIFA et Mouhamed Bachir DIOP - M2 GSI

02

Qu'est-ce que le réseau ERUA ?

Le réseau ERUA, c'est quoi ?

- Une alliance européenne de 5 universités réformatrices issues de Bulgarie, du Danemark, de France et de Grèce avec des valeurs et des objectifs communs en matière de recherche, d'enseignement et de développement institutionnel.

Quel est son objectif ?

- Promouvoir la mobilité et réinventer l'enseignement supérieur et la recherche.





03

Pourquoi ERUA & Erasmus ?

Le choix de ce sujet se justifie notamment par l'importance du programme ERASMUS dans le projet ERUA.

Erasmus est un acteur décisif :

- Initiateur de projets
- Nombreux partenariats internationaux / Dimension internationale
- Mobilité des jeunes et du personnel

04

Les sites d'entrée utilisés

Dans le cadre de notre étude, notre première mission a été de :

- Rechercher et lister quelques URL de sources des laboratoires de l'Université Paris 8 appartenant au réseau ERUA et des URL de sources Erasmus.
- Importer dans l'outil de visualisation de données Hyphe.

Quelques exemples de liens recensés :

- <https://info.erasmusplus.fr/>
- <https://erasmus-plus.ec.europa.eu/fr>
- <https://www.education.gouv.fr/le-programme-erasmus-9890>
- <https://circeft.fr/>
- <https://cemti.univ-paris8.fr/>
- <http://www.arscan.fr/arscan-presentation/>
- <http://www.cognition-usages.org/chart2/>
- <https://aiac.univ-paris8.fr/>

05

Hyphe : l'outil de visualisation de données



06

Le déroulé du Crawl

DEFINE WEB ENTITIES
Check the boundaries of each web entity before creating it

Move all sliders TO THE LEFT TO THE RIGHT

1	Arscan.fr /arscan-presentation	http .fr arscan www. /arscan-pres...	Already exists
2	Circeft.fr	https .fr circeft	Already exists
3	Cognition-Usages.org /chart2	http .org cognition-us... www. /chart2	Already exists
4	Education.gouv.fr	https .gouv.fr education www. /le-programm...	
5	Erasmusplus.fr	https .fr erasmusplus info.	
6	Europa.eu	https .eu europa ec. erasmus-plus. /fr	
7	Univ-Paris8.fr	https .fr univ-paris8 aiac.	Already exists Same web entity defined row 8
8	Univ-Paris8.fr	https .fr univ-paris8 cemti.	Already exists Same web entity defined row 7

CREATE WEB ENTITIES

1-ère étape :

- Vérifier si les entités web sont bien définies
Si ce n'est pas, étirer manuellement le curseur de sorte à sélectionner l'entièreté du lien hypertexte.

2-ème étape :

- Définir la profondeur des clics

3-ème étape :

- Lancer le crawler

musidanse.Univ-Paris8.fr

Prévu hier

Finished in 1 minute

Q Profondeur 2

INFRACTUEUX

0 page explorée (+1 erreur)

Univ-Paris8.fr /-La-MSH-Paris-Nord-

Prévu hier

Finished in 1 hour

Q Profondeur 2

ATTEINT

1 page explorée

lasectionclinique.Un

Prévu hier

Finished in 1 hour

Q Profondeur 2

ATTEINT

78 pages explorées

transcrit.Univ-Paris8

Prévu hier

Finished in 1 hour

Q Profondeur 2

ATTEINT

1 page expl

Le déroulé du Crawl

Après avoir attendu près de 24 heures pour crawler l'intégralité des sources d'entrées, Hyphe est parvenu à analyser 37 sources sur 41.

Les sources peuvent être infructueuse, si :

- Entité web pas bien défini
- Problème d'URL

Solution :

- Chercher s'il y a d'autres URL sources pour les entités web infructueuses.

Le processus de nettoyage

Le processus de nettoyage consiste à éliminer les données créant du bruit.

The screenshot shows a software interface for managing web entities. At the top, it displays "ENTITÉS WEB 1,958 / 2,144". Below this are four filter buttons: "DANS 42" (checked), "INDÉCIS 6" (checked), "SUR 192" (unchecked), and "DÉCOUVERT 1 910" (checked). A "Rechercher" input field contains the word "Facebook". On the right, there are "APPLIQUER LES MODIFICATIONS" and "ANNULER" buttons. The main area is a table with columns: "NOM", "ST", "C", "Le", "CITÉ ↑", and "DERNIER MOD". The table lists several entities:

NOM	ST	C	Le	CITÉ ↑	DERNIER MOD
Univ-Paris8.fr	DANS	265	é	29	Hier
Archives-Ouvertes.fr	DÉC...	62,431	é	16	Hier
Hypotheses.org	DÉC...	80	é	16	Hier
Cnrs.fr	DÉC...	334	é	15	Hier

1-ère étape :

- Explorer ce que Hyphe a crawlé et examiner la liste des sources liées.

2-ème étape :

- Classer les URL sources en fonction des plus citées.

3-ème étape :

- Rechercher à l'aide des filtres les sources provenant des GAFAM et réseaux sociaux afin de limiter le bruit.

08

Hyphe : la consultation du graphe et les paramétrages



La consultation du graphe et les paramétrages

Après toutes ces étapes, nous accédons à la visualisation du corpus web généré par Hyphe sous forme de graphe.

Les filtres permettent d'affiner la visualisation et la construction des données en fonction du nombre d'entités web qui la citent ou des algorithmes de spatialisation.

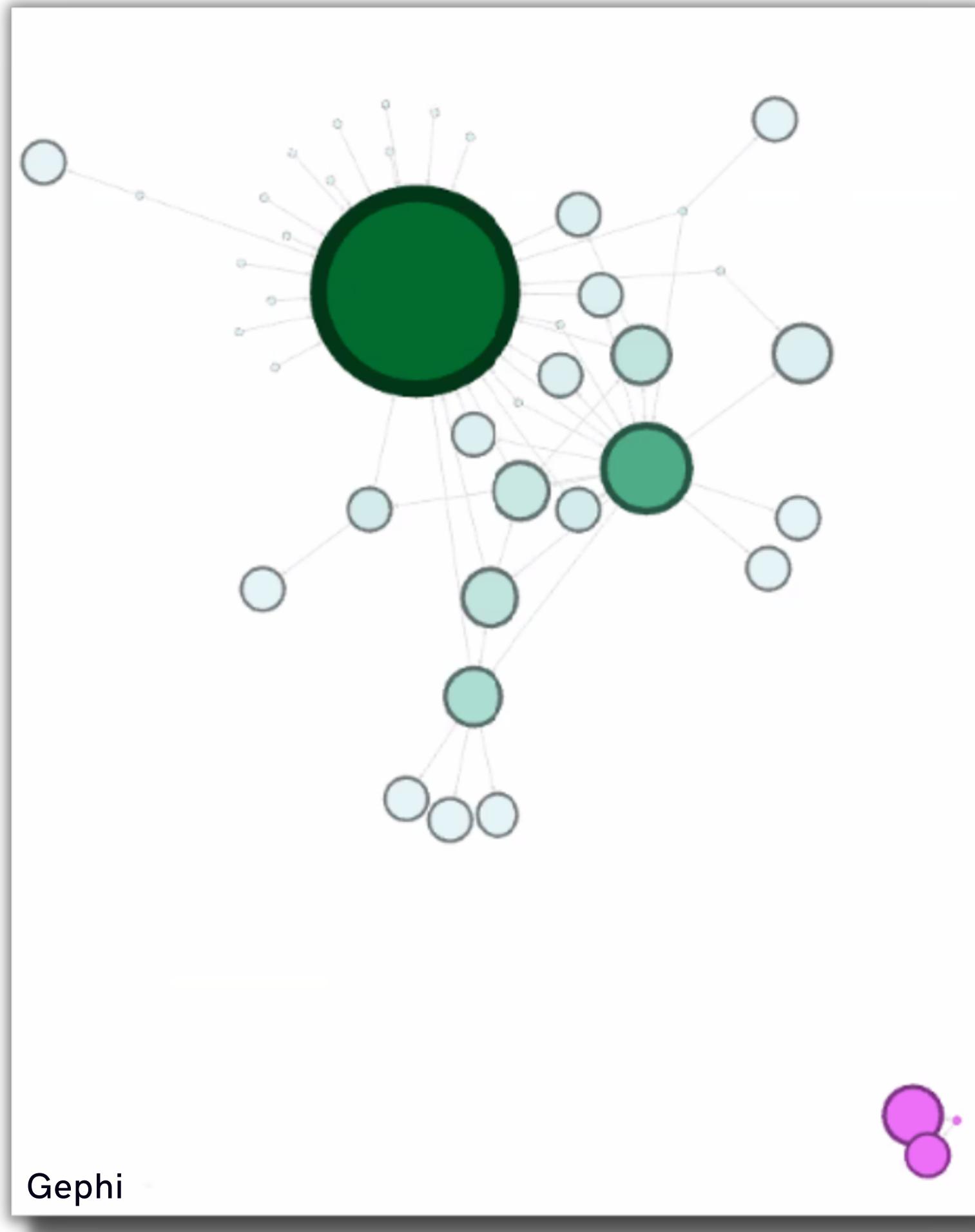
C'est en fonction des différents paramétrages que le graphe et les ressources évoluent.

The screenshot shows the 'Paramètres de visualis...' (Visualization parameters) dialog box. At the top, there's a search bar labeled 'Rechercher une WebEntity'. Below it is a 'Filtrage' (Filtering) section with three checkboxes: 'DANS' (checked), 'INDÉCIS' (checked), and 'DEHORS' (unchecked). A note below says 'Pas de filtrage' (No filter). To the right, there are several filtering options: 'Affichage uniquement connecté DÉCOUVERT' (Display only connected DISCOVERED), 'Afficher uniquement DÉCOUVERT avec 2+ liens' (Display only DISCOVERED with 2+ links), 'Afficher uniquement DÉCOUVERT avec 3+ liens' (Display only DISCOVERED with 3+ links), and 'Afficher uniquement DÉCOUVERT avec 4+ liens' (Display only DISCOVERED with 4+ links). On the right side of the dialog, there's a 'Clé' (Key) section with a note: 'Chaque point ou nœud est un entité Web' (Each point or node is a Web entity). Below this, there's a 'COULEUR DU NŒUD' (Node color) section titled 'Web entity status' with four categories: 'DANS' (black dot), 'INDÉCIS' (grey dot), 'DEHORS' (red dot), and 'DÉCOUVERT' (blue dot). Further down, there's a 'TAILLE DU NŒUD' (Node size) section titled 'Degré' with a plus/minus button. A note explains: 'Le DEGRÉ d'une entité Web est le nombre d'autres entités Web qu'elle cite plus le nombre d'autres entités Web qui la citent.' (The degree of a Web entity is the number of other Web entities it cites plus the number of other Web entities that cite it). At the bottom, there are two notes: 'Plus petit nœud DEGRÉ de 0' (Smallest node DEGREE of 0) and 'Le plus grand nœud DEGRÉ de 29' (Largest node DEGREE of 29).

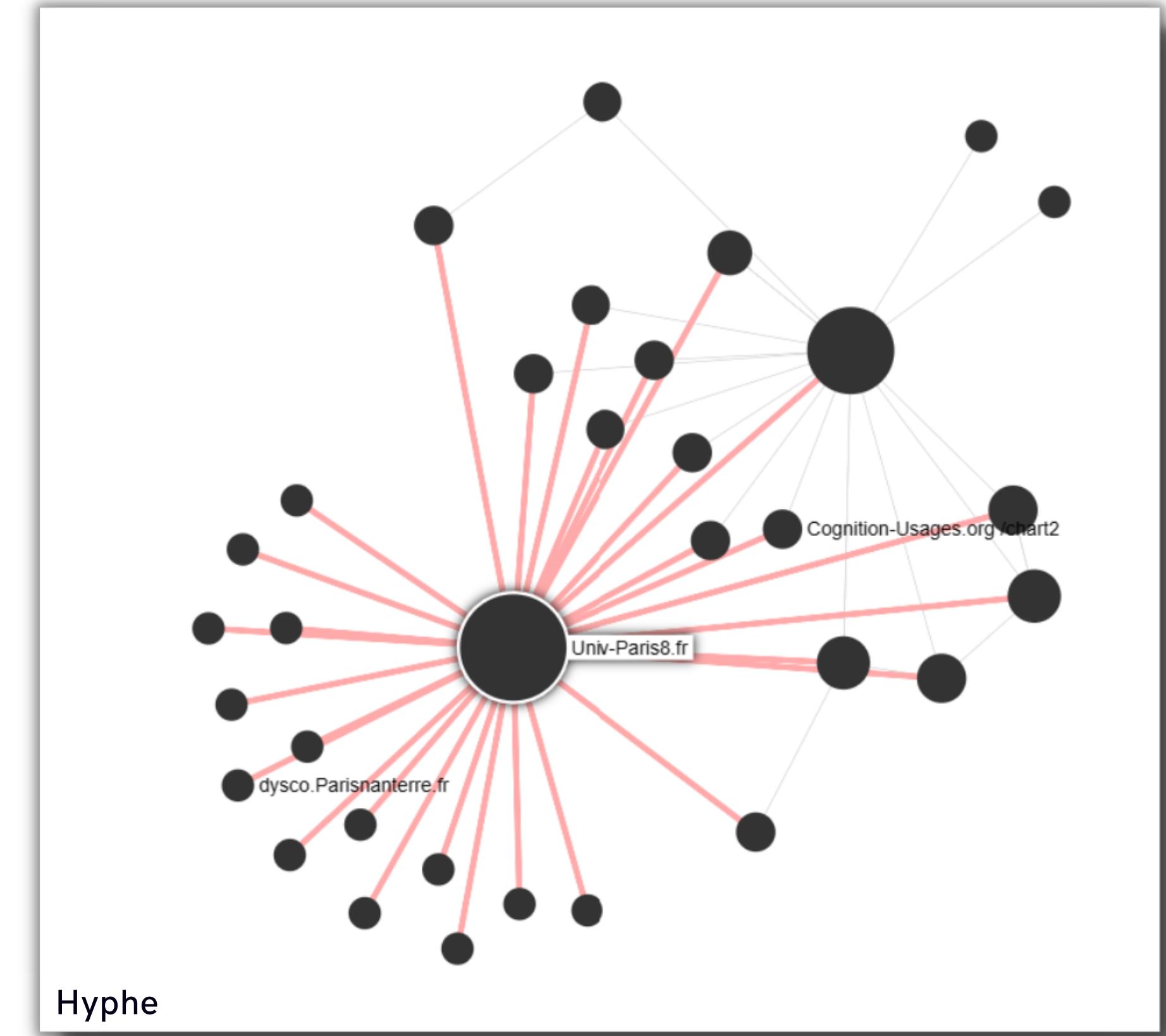


09

Hyphe & Gephi : notre analyse des graphes



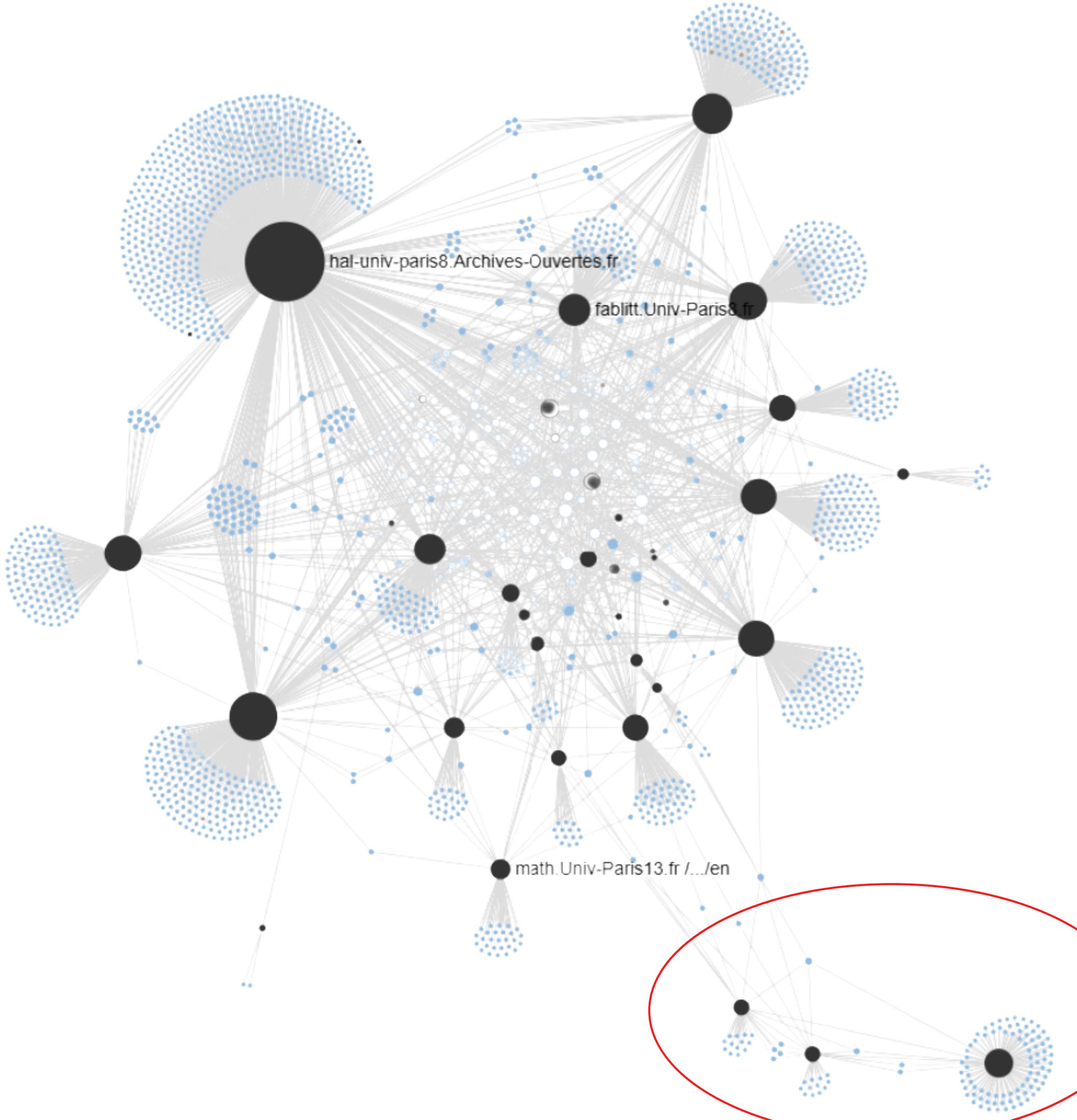
Gephi



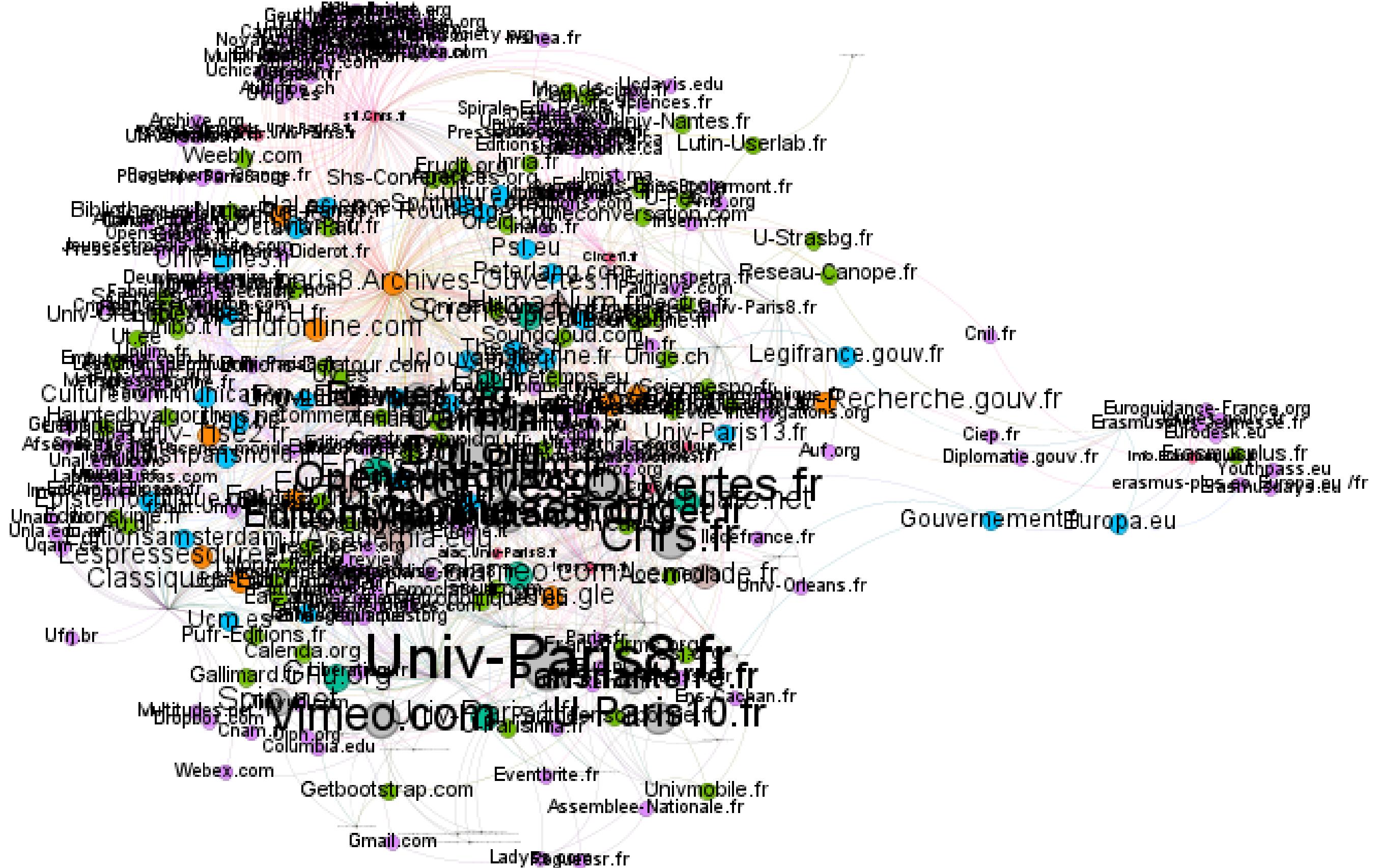
Hyphe

https://samszo.github.io/M2GSI_22-23/pages/Ferielle_Bachir.html

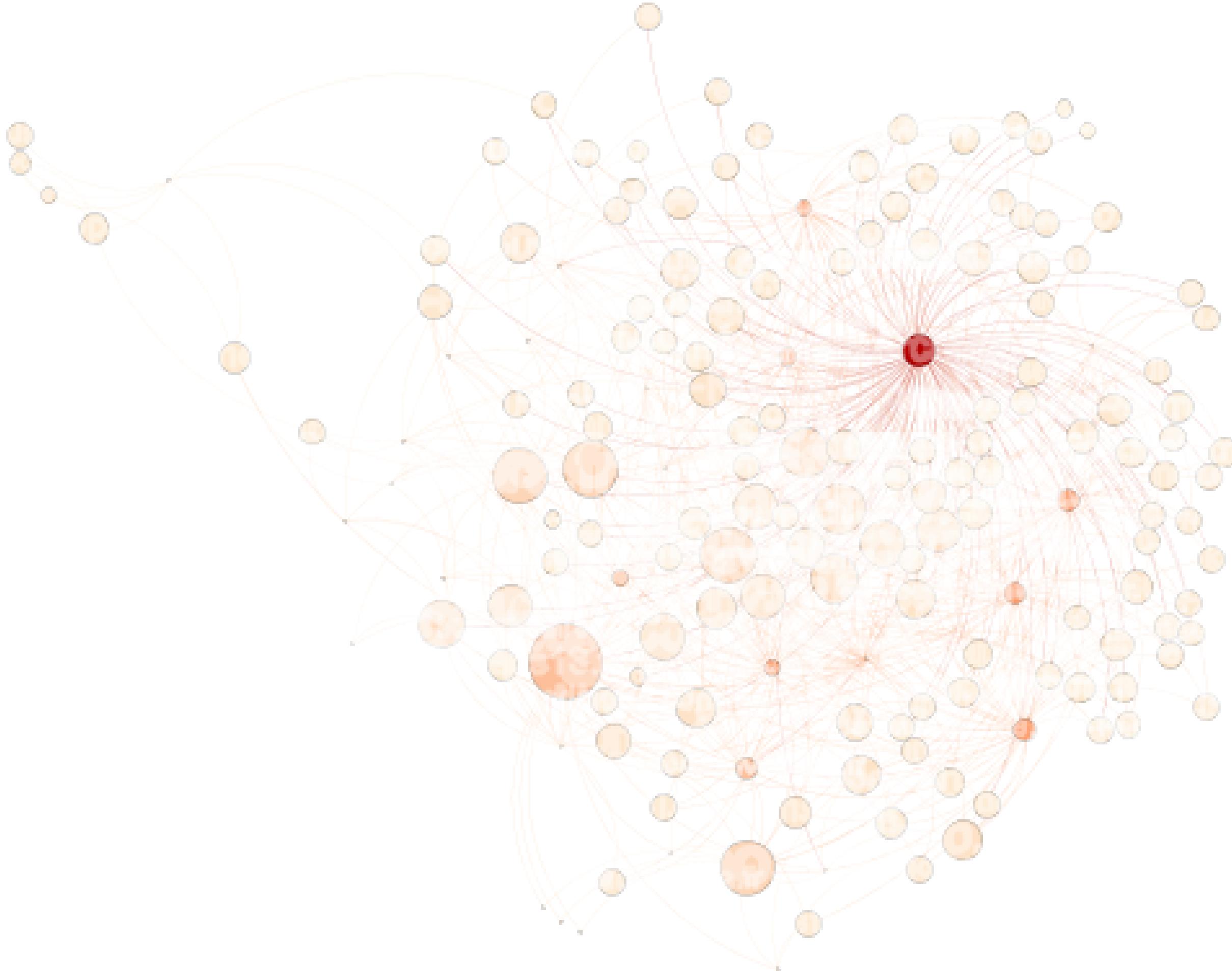
Hyphe



Gephi



Gephi



https://samszo.github.io/M2GSI_22-23/gephi.html?path=Ferielle_Bachir-3.gexf

Université Paris 8 & Erasmus - Férielle Et Bachir

Exported 2023-01-26T08:40:34.682Z

CIRCEFT.FR

ID	3
NAME	Circeft.fr
PREFIXES AS URL	
•	http://circeft.fr
•	https://www.circeft.fr
•	http://www.circeft.fr
•	https://circeft.fr
INDEGREE	1
TOTAL KNOWN PAGES	559
CRAWLED PAGES	130
STATUS	IN
LAST MODIFICATION DATE	2023-01-25T09:51:00.744Z



**Merci à tous pour
votre attention**

**FAITES-NOUS SAVOIR SI VOUS AVEZ DES QUESTIONS
OU BESOIN DE PRÉCISION.**