

## Linkage : analyse conjointe de réseaux et de corpus

### Objectifs et contenu

- Comprendre les bases de l'analyse des réseaux et des corpus de documents
- Apprendre à utiliser le logiciel Linkage
- Apprendre à structurer ses propres données pour les analyser avec Linkage

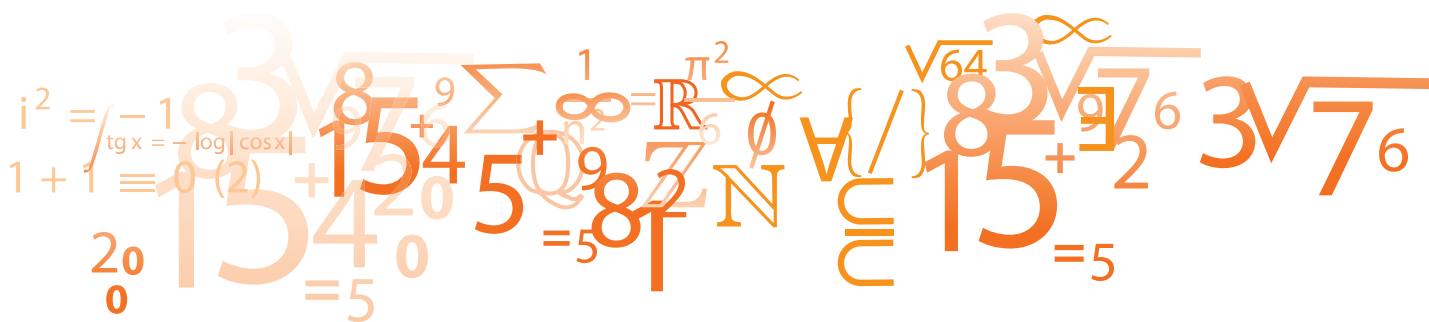
L'intérêt croissant pour l'analyse des réseaux de communication s'explique par la forte présence de ce type de données dans le monde numérique d'aujourd'hui et par les progrès récents dans leur modélisation et leur traitement. Les méthodes de clustering permettent de découvrir une structure en groupes cachés dans le réseau. Dans ce cadre, les méthodes statistiques d'intelligence artificielle présentent l'avantage d'offrir une segmentation fine des données dont l'interprétation est facilitée par le modèle statistique sous-jacent. Aujourd'hui, la plupart des réseaux sociaux sont associés à du texte mais les outils actuels ne permettent pas de réaliser une analyse conjointe.

La méthodologie implémentée dans le logiciel Linkage permet de pallier ce manque et autorise l'analyse conjointe de réseaux et de textes. Linkage requiert uniquement la donnée d'un ensemble d'échanges de textes entre des individus (ou des entités). Par exemple, on peut considérer les échanges de textes ou d'emails entre des individus d'un réseau social ou encore les co-publications scientifiques (ou brevets). Linkage est également capable de déterminer automatiquement le nombre de groupes d'individus et de topics utilisés. Les sorties du logiciel sont les classifications des individus dans les groupes (assortis de probabilités) ainsi qu'une description des topics à l'aide des mots les plus représentatifs.

Cette formation permettra aux stagiaires de s'approprier le modèle statistique de Linkage et d'apprendre à maîtriser le logiciel ainsi que ses différentes sorties graphiques. Elle sera illustrée par des exemples concrets sur données réelles provenant de différents contextes (bibliométrie, réseaux sociaux, finance, marketing, sécurité...)

- Les basiques
- Analyse des réseaux
- Analyse de corpus de documents
- Le modèle statistique dans Linkage / éléments d'inférence
- Présentation générale du logiciel
- Etude bibliométrique
- Deux études de cas : un réseau social - le scandale Enron
- Formatage des données en CSV

Alternance de cours (50 %) et de travaux dirigés (50 %)



## Intervenants

Pierre Latouche (Université Paris 1 Panthéon-Sorbonne)

Page personnelle : <http://samm.univ-paris1.fr/Pierre-Latouche>

Page du laboratoire : <http://samm.univ-paris1.fr/>

Charles Bouveyron (MAP5, Université Paris-Descartes).

Page personnelle : <http://w3.mi.parisdescartes.fr/~cbouveyr/>

Page du laboratoire : <http://www.math-info.univ-paris5.fr/map5/>

## Public visé

Statisticiens, analystes de données, data miners, actuaires, biologistes, etc.

Afin d'adapter le contenu du stage aux attentes des stagiaires, un questionnaire téléchargeable sur le site CNRS Formation ([https://cnrsformation.cnrs.fr/data/STG\\_19316\\_20222.doc](https://cnrsformation.cnrs.fr/data/STG_19316_20222.doc)) devra être complété et renvoyé au moment de l'inscription.

## Prérequis

Connaissances de base des outils informatiques. Des connaissances en analyse de textes ou réseaux sont un plus mais non nécessaires.

## Lieu

Paris

## Coût pédagogique

700 Euros.

## Durée, emploi du temps et organisation

1 jour : le vendredi 06/03/2020

De 4 à 20 stagiaires

TD encadrés par deux intervenants

## Infrastructure requise

Il est demandé aux stagiaires d'apporter leur propre ordinateur portable. Les logiciels nécessaires à la formation seront installés en début de stage.

## A l'issue de la formation

Evaluation de la formation par les stagiaires.

Envoi d'une attestation de formation.