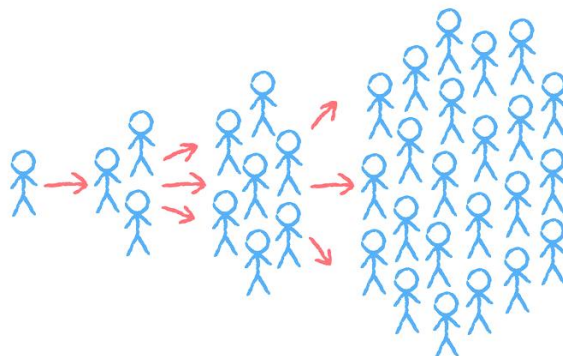


Objectif :

Dans le TP précédent, nous avons utilisé des modèles pour étudier les phénomènes de diffusion. Dans ce TP, nous tenterons de comprendre les phénomènes de diffusion en nous basant sur des données réelles, en particulier sur la publication de messages sur le média social Twitter.



Exercice 1 : Préliminaire

- Comment générer un réseau à partir de Twitter ?
- Quelles sont les particularités de ce réseau ?
- En utilisant les commandes **more** ou **head** sur linux, prendre connaissance des jeux de données :
 - Fichier représentant le réseau des followers (Twitter_Network.net)
 - Fichiers représentant les Tweets sur l'année 2009 de Juin à Décembre (tweets2009-XX.txt)
- Comment sont formatées ces données ?

Exercice 2 : Analyse du réseau

- Le réseau étant extrêmement volumineux, commencez par extraire un sous-graphe de ce réseau à l'aide de la commande **head** sur linux.
- En utilisant l'outil d'analyse **Gephi**, commenter ce réseau
- Ce sous-réseau a-t-il une structure particulière ?
- Refaire les mesures en augmentant la taille du sous-réseau
- Que peut-on conclure en matière de diffusion d'informations ?

Exercice 3 : Etude d'un phénomène de diffusion

- Comment se propage une information sur le réseau Twitter ?
- En utilisant l'outil **ESTweetsAnalyze**ⁱ, comparez l'impact qu'ont eu 2 sujets, de votre choix, sur le réseau Twitter en montrant sous forme de graphique:
 - L'intérêt pour le sujet choisi (i.e. le nombre de Tweets par jour)
 - Le nombre de diffuseurs distincts par jour (en analysant le fichier TWEETS.txt)
- Commenter les tendances observées

ⁱ *L'outil **ESTA** s'utilise de la façon suivante :

```
java -jar ESTweetsAnalyze.jar <dossier_dataset> <nb_de_mois_a_analyser> <mot-clef1> <mot-clef2> <...>
Ex. java -jar ESTweetsAnalyze.jar /home/toto/ESTweetsAnalyze 2 "#MickaelJackson" "#MJ"
```

2 fichiers sont ainsi générés par l'application dans un dossier **resultats** :

- RESULTATS.txt : Calcule, pour chaque jour, le nombre de tweets contenant les mots-clefs recherchés
- TWEETS.txt : Regroupe tous les Tweets qui contiennent les mots-clefs recherchés (date, user, message)