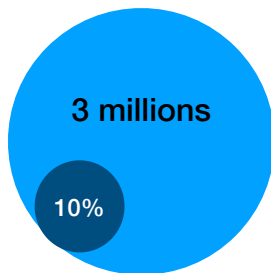


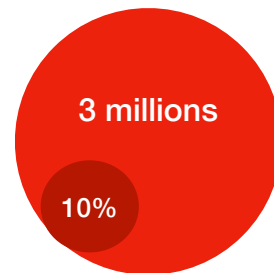


# Schéma de notre démarche en data mining

1. Sélection arbitraire de média de départ rassemblant un nombre important d'individus, légèrement orientés politiquement.



2. Récupération de 10% des individus s'abonnant à leur contenu

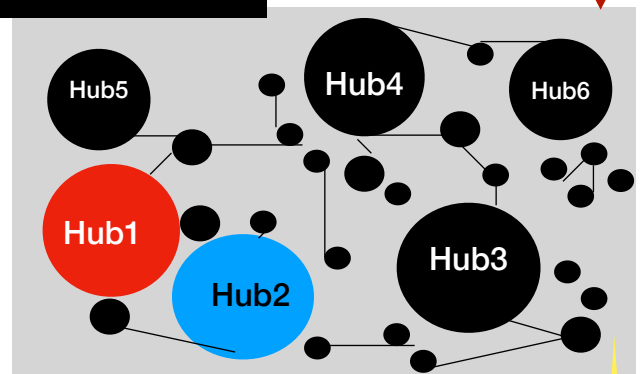
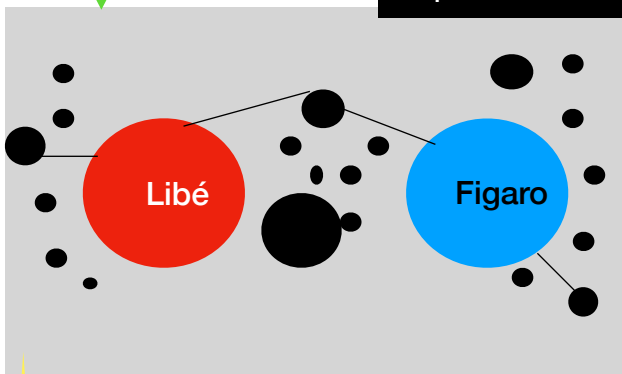


Peu de contraintes  
(600 000/2h)

3. Elaboration de deux graphes

Beaucoup de contraintes  
(130 utilisateurs/2h)

Un point est un individu, un lien est un abonnement.



Le graphe « par hubs de départ », permet d'avoir une visualisation des 600 000 individus. Un individu est soit lié à Figaro, soit lié à Libération, soit les deux. Un individu est plus ou moins imposant selon son nombre d'abonnés (popularité sur Twitter).

Permet d'observer si les individus des deux hubs de départ s'opposent ou non. Si oui, lesquels ? (Parmi les populaires)

Le graphe « par recherche d'intérêts », permet d'avoir une visualisation en profondeur des communautés de départ en recherchant pour chaque individu tous ses abonnements. Un individu est alors lié sur le graphe à tous ses abonnements. Un individu est plus ou moins imposant selon le nombre de liens qu'il reçoit (popularité dans la communauté).

Permet d'identifier des nouveaux hubs et avoir une répartition par communautés.