

Performance des systèmes d'information

Tien-Nam Le

`tien-nam.le@ens-lyon.fr`
`perso.ens-lyon.fr/tien-nam.le/su`

Sciences U Lyon

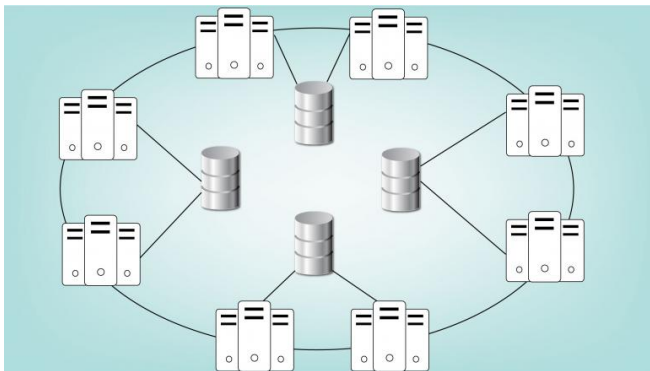
Outline

Partie 1. Big Data



Outline

Partie 2. Performance des systèmes distribués



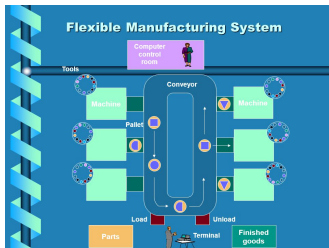
Outline

Partie 3. Performances des architectures de cloud computing



Outline

Partie 4. Performances des systèmes de fabrication et des chaînes d'approvisionnement



Langues

Python avec **Hadoop** et **Spark**



Lecture 1. Big Data

Part 1

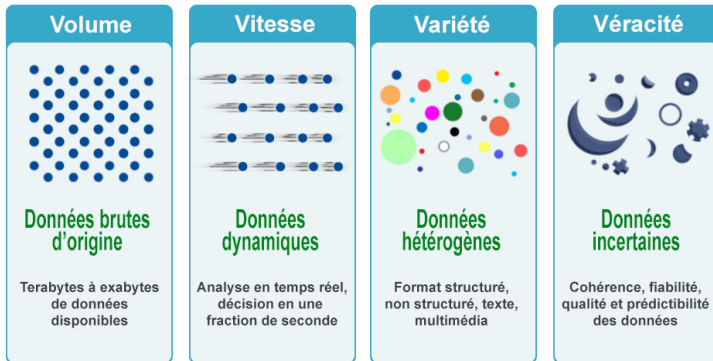
Outline

1. Réviser Big Data
2. Modélisation de Big Data
3. Analyse des modèles: Google PageRank
4. TP: Python pour Data Science

1.1. Réviser Big Data

Quatre "V" de Big Data

Dimensions du Big Data

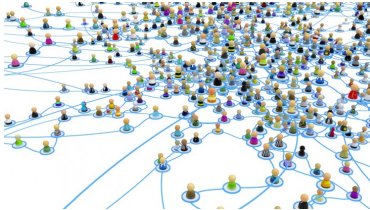


Réseaux (Networks)

Les données peuvent être représentées sous la forme d'un **réseau** avec **connexions** entre des points de données.

Exemples:

- ▶ Réseau utilisateur / client
- ▶ Réseau de collaboration
- ▶ Réseau de médias sociaux
- ▶ répertoire Web
- ▶ Réseau de distribution de produits

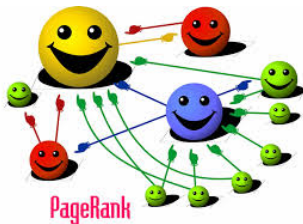


Modélisation des réseaux

Pour exploiter les informations des données, ← on doit comprendre le réseau.

Exemple: Google PageRank

- ▶ **Données:** un ensemble de sites Web
- ▶ **Réseau:** Chaque site est un nœud, Web a \rightarrow Web b s'il existe un lien de site a vers site b
- ▶ **Objectif:** Renvoyer le sites le plus pertinents pour chaque recherche.



Pour comprendre le réseau:

- ▶ D'abord besoin de trouver une bonne façon de modéliser le réseau
- ▶ Puis analyser les caractéristiques du modèle

1.2 Modélisation des réseaux

Graphes

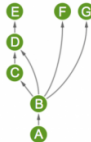
Définition des **graphes**:

- ▶ Nœuds (sommets)
- ▶ Arêtes (liens)
- ▶ Graphes dirigés / non dirigés
- ▶ Graphes pondérés

Undirected



Directed



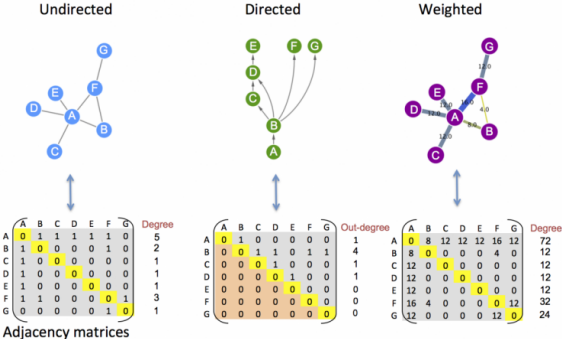
Weighted



Graphes

Autres définitions:

- ▶ Degré des noeuds, distributions de degré
- ▶ Composants (fortement) connexes; composants d'entrée / sortie
- ▶ Représentation matricielle des graphes



Exercises:...

Modélisation de réseaux par graphes

- ▶ **Question:** Comment modéliser un réseau quand on n'a pas les données complètes.
- ▶ **Idée:** Générer un graphe reproduisant la complexité du réseau réel.
- ▶ **Problème:** décider où mettre des arêtes entre les nœuds dans le graphe généré
- ▶ **Méthodes communes:** générer
 - ▶ Réseaux aléatoires
 - ▶ Réseaux sans échelle (scale-free networks)
 - ▶ Réseaux petit monde (small-world networks)

Réseaux aléatoires

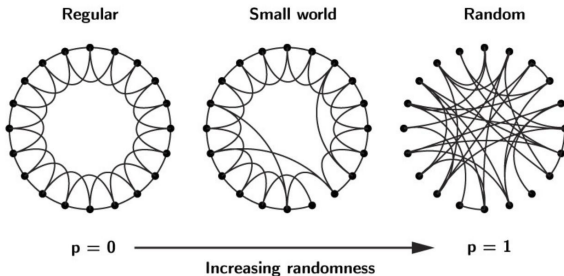
- ▶ Supposons que le réseau a n nœuds
- ▶ Chaque deux nœuds a et b ont un arête avec une probabilité fixe p
- ▶ c'est-à-dire les liens d'un réseau aléatoire se produit indépendamment et au hasard.
- ▶ Les réseaux aléatoires générés avec le même ensemble des paramètres n et p semblent légèrement différents les uns des autres.
- ▶ Le degré de nœuds dans un réseau aléatoire suit une distribution de Poisson.
- ▶ Les réseaux aléatoires nous fournissent un modèle utile faire une analyse initiale.

Réseaux sans échelle

- ▶ Réseau sans échelle: il en existe un grand nombre des nœuds hautement connectés.
- ▶ La distribution des degrés des nœuds suivre une loi de puissance.
- ▶ Les noeuds de haut degré s'appellent des "hubs"
- ▶ Exemple: réseaux sociaux ou Internet
- ▶ Un réseau sans échelle résiste aux pannes de réseau. Probabilité qu'un nœud avec un degré élevé échouer est faible → le réseau est presque toujours connecté.
- ▶ Générer un réseau sans échelle: Ajouter un nouveau lien au noeud avec une probabilité liée au nombre de liens c'est déjà fait.

Réseaux petit monde

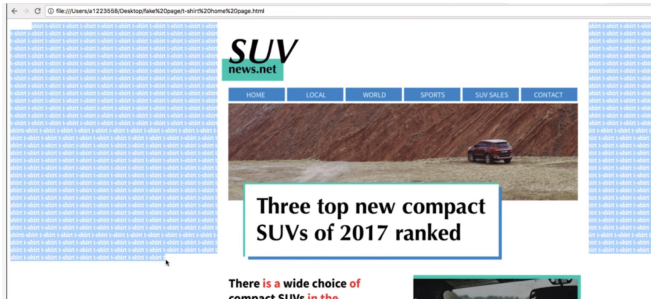
- ▶ La plupart des nœuds ne sont pas voisins mais la distance entre deux nœuds est très petite
- ▶ Il y a de très gros 'hubs'
- ▶ Propriété "Six degrés de séparation"
- ▶ Exemples: sites Web avec menus de navigation, réseaux électriques, réseaux de neurones du cerveau, réseaux de co-occurrence de mots
- ▶ Génération de réseaux petit monde:



1.3 Analyse des modèles: Google PageRank

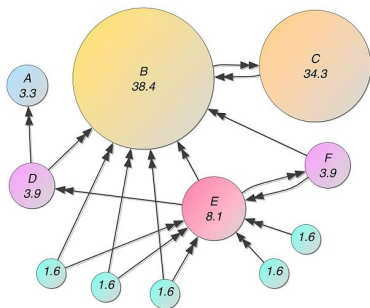
Moteurs de recherche

- ▶ **Objectif:** renvoyer les pages les plus pertinentes pour chaque mot clé
- ▶ **Techniques avant Google:** basé sur nombre d'apparences des mots clés.
- ▶ **Problème:** spam de mot clé



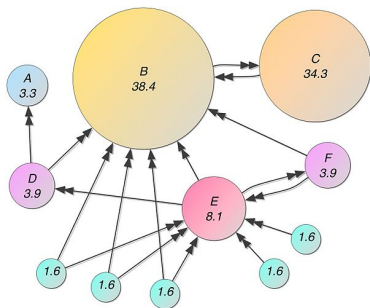
Google PageRank

- ▶ Méthode PageRank: renvoie les pages les plus importantes contenant les mots-clés (ex. Wikipedia, sites Web gouvernementaux, grands journaux, etc.).
- ▶ Quelles pages sont importantes?
- ▶ Chaque page se voit attribuer une valeur qui résume l'importance de cette page.
- ▶ Une valeur élevée signifie que la page Web est importante.



Google PageRank

- ▶ Essentiellement, la valeur attribuée à une page Web est la probabilité d'atterrir sur cette page après une série de clics aléatoires.
- ▶ Calculer PageRank: (tableau noir)



Exercises...

TP: Python pour Data Science

Libraries

- ▶ Numpy
- ▶ Pandas
- ▶ Matplotlib