M1 Info - Systemes Complexes Avances

Cours 4 - Centralite des noeuds et des liens dans un reseau complexe

Semestre Automne 2022-2023 - Université Côte D'azur

Christophe Crespelle

christophe.crespelle@univ-cotedazur.fr

* Merci a Rémy Cazabet pour ses slides, sur lesquels est base ce cours.

Centralite des noeuds

- Sert a mesurer l'importance des noeuds dans un reseau

- Terme un peu trompeur
  - Pas forcement lie a une notion de centralite geometrique

Centralite des noeuds

- Sert a mesurer l'importance des noeuds dans un reseau
- Terme un peu trompeur
  - Pas forcement lie a une notion de centralite geometrique
- Utilisation des mesures de centralite
  - Ont souvent une interpretation claire dans le contexte

Centralite des noeuds

- Sert a mesurer l'importance des noeuds dans un reseau
- Terme un peu trompeur
  - Pas forcement lie a une notion de centralite geometrique
- Utilisation des mesures de centralite
  - Ont souvent une interpretation claire dans le contexte
  - Peuvent servir de proprietes des noeuds pour l'apprentissage automatique
    - Classification des noeuds
    - Prediction de liens

Centralite de degre

- Le degre du noeud u : $d^\circ(u)$
  - C'est a dire son nombre de voisins
  - Dans le cas oriente, on peut se restreindre au degre sortant $d^+(u)$ ou entrant $d^-(u)$

Centralite de degre

- Le degre du noeud u : $d^{\circ}(u)$
  - C'est a dire son nombre de voisins
  - Dans le cas oriente, on peut se restreindre au degre sortant $d^{+}(u)$ ou entrant $d^{-}(u)$
- Souvent suffisant a trouver les noeuds importants :

Centralite de degre

- Le degre du noeud u : $d^{\circ}(u)$
  - C'est a dire son nombre de voisins
  - Dans le cas oriente, on peut se restreindre au degre sortant $d^{+}(u)$ ou entrant $d^{-}(u)$
- Souvent suffisant a trouver les noeuds importants :
  - Dans une serie, les personnages principaux sont ceux qui parlent au plus d'autres personnages

Centralite de degre

- Le degre du noeud u : $d^\circ(u)$
  - C'est a dire son nombre de voisins
  - Dans le cas oriente, on peut se restreindre au degre sortant $d^+(u)$ ou entrant $d^-(u)$
- Souvent suffisant a trouver les noeuds importants :
  - Dans une serie, les personnages principaux sont ceux qui parlent au plus d'autres personnages
  - Les aeroports les plus importants sont ceux qui ont le plus de connections avec les autres

Centralite de degre

- Le degre du noeud u : $d^{\circ}(u)$
  - C'est a dire son nombre de voisins
  - Dans le cas oriente, on peut se restreindre au degre sortant $d^{+}(u)$ ou entrant $d^{-}(u)$
- Souvent suffisant a trouver les noeuds importants :
  - Dans une serie, les personnages principaux sont ceux qui parlent au plus d'autres personnages
  - Les aeroports les plus importants sont ceux qui ont le plus de connections avec les autres
- Mais pas toujours :

Centralite de degre

- Le degre du noeud u : $d^\circ(u)$
  - C'est a dire son nombre de voisins
  - Dans le cas oriente, on peut se restreindre au degre sortant $d^+(u)$ ou entrant $d^-(u)$
- Souvent suffisant a trouver les noeuds importants :
  - Dans une serie, les personnages principaux sont ceux qui parlent au plus d'autres personnages
  - Les aeroports les plus importants sont ceux qui ont le plus de connections avec les autres
- Mais pas toujours :
  - Les utilisateurs facebook avec le plus d'amis sont des faux

Centralite de degre

- Le degre du noeud u : $d^{\circ}(u)$
  - C'est a dire son nombre de voisins
  - Dans le cas oriente, on peut se restreindre au degre sortant $d^{+}(u)$ ou entrant $d^{-}(u)$
- Souvent suffisant a trouver les noeuds importants :
  - Dans une serie, les personnages principaux sont ceux qui parlent au plus d'autres personnages
  - Les aeroports les plus importants sont ceux qui ont le plus de connections avec les autres
- Mais pas toujours :
  - Les utilisateurs facebook avec le plus d'amis sont des faux
  - Les pages web ou wikipedia avec le plus de liens sont simplements des listes de reference

Notions de entralites basees sur les distances

- Farness

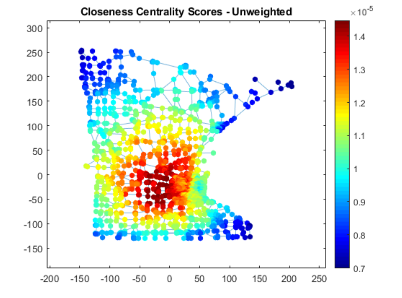
- Closeness

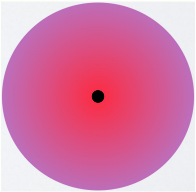
- Centralite harmonique

Farness et closeness

- Idee : quantifier a quel point le noeud u est proche/loin de tous les autres noeuds du reseau

- Similaire au centre d'une figure geometrique
  - Le centre d'un disque est le point ayant la plus petite distance moyenne (et distance max) aux autres points du disque





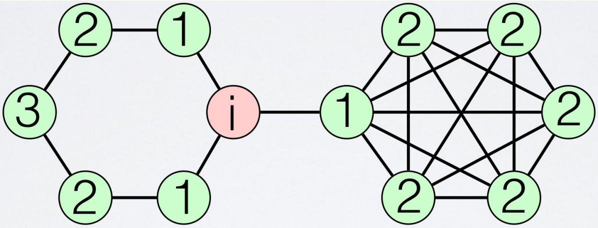
Farness

Definition (Farness)

Distance moyenne du noeud u a tous les autres noeuds :

$$Farness(u) = \frac{\sum_{v \in V \setminus \{u\}} dist(u, v)}{n - 1}$$

- Exemple :



Closeness

- Probleme de la farness
  - Au plus le score est grand, au moins le noeud est "central"

Closeness

- Probleme de la farness
  - Au plus le score est grand, au moins le noeud est "central"
- Correction :

Definition (Closeness)

Inverse de la farness :

$$Closeness(u) = \frac{1}{Farness(u)} = \frac{n-1}{\sum_{v \in V \setminus \{u\}} dist(u,v)}$$

Closeness

- Probleme de la farness
  - Au plus le score est grand, au moins le noeud est "central"
- Correction :

Definition (Closeness)

Inverse de la farness :

$$Closeness(u) = \frac{1}{Farness(u)} = \frac{n-1}{\sum_{v \in V \setminus \{u\}} dist(u,v)}$$

- Remarque : $Closeness(u) \in [0,1]$

Closeness

- Probleme de la farness
  - Au plus le score est grand, au moins le noeud est "central"
- Correction :

Definition (Closeness)

Inverse de la farness :

$$Closeness(u) = \frac{1}{Farness(u)} = \frac{n-1}{\sum_{v \in V \setminus \{u\}} dist(u,v)}$$

- Remarque : $Closeness(u) \in [0, 1]$
  - $Closeness(u) = 1$ ssi tous les neouds sont a distance 1 de u

Closeness

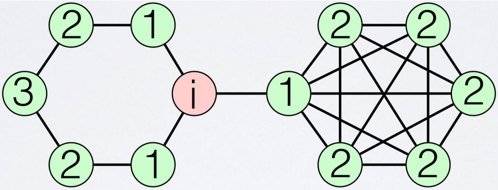
- Probleme de la farness
  - Au plus le score est grand, au moins le noeud est "central"
- Correction :

Definition (Closeness)

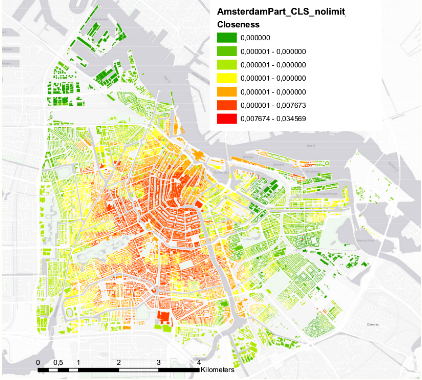
Inverse de la farness :

$$Closeness(u) = \frac{1}{Farness(u)} = \frac{n-1}{\sum_{v \in V \setminus \{u\}} dist(u,v)}$$

- Remarque : $Closeness(u) \in [0,1]$
  - $Closeness(u) = 1$ ssi tous les neouds sont a distance 1 de u
- Exemple :

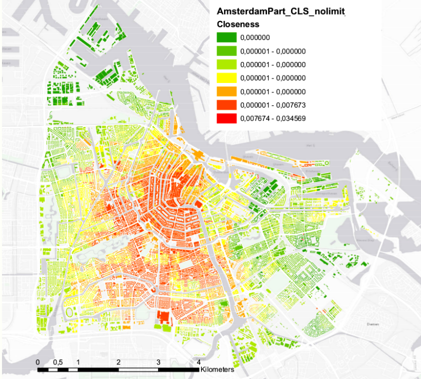


Closeness d'un reseau urbain



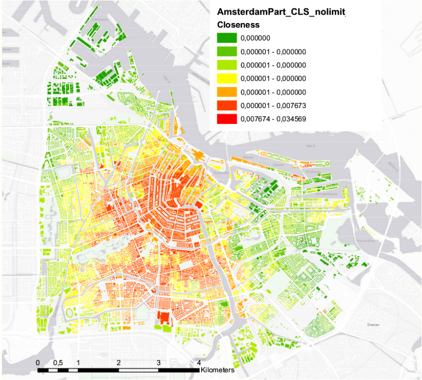
- La closeness a du sens pour les reseaux spatialises

Closeness d'un reseau urbain



- La closeness a du sens pour les reseaux spatialises
- Et pour les autres ? ? ?

Closeness d'un reseau urbain



- La closeness a du sens pour les reseaux spatialises
- Et pour les autres ? ? ?
  - C'est une bonne question pour votre projet ;)

Centralite Harmonique

Definition (Harmonic centrality)

Variante de la closeness definie comme la moyenne de l'inverse des distances a tous les autres noeuds :

$$Harmonic(u) = \frac{1}{n-1} \sum_{v \in V \setminus \{u\}} \frac{1}{dist(u,v)}$$

Centralite Harmonique

Definition (Harmonic centrality)

Variante de la closeness definie comme la moyenne de l'inverse des distances a tous les autres noeuds :

$$Harmonic(u) = \frac{1}{n-1} \sum_{v \in V \setminus \{u\}} \frac{1}{dist(u,v)}$$

- Remarque : bien definie meme pour les reseaux non connexes, en posant $\frac{1}{+\infty} = 0$.

Centralite Harmonique

Definition (Harmonic centrality)

Variante de la closeness definie comme la moyenne de l'inverse des distances a tous les autres noeuds :

$$Harmonic(u) = \frac{1}{n-1} \sum_{v \in V \setminus \{u\}} \frac{1}{dist(u,v)}$$

- Remarque : bien definie meme pour les reseaux non connexes, en posant $\frac{1}{+\infty} = 0$.
- Comme pour la closeness, on a :
  - $Harmonic(u) \in [0,1]$
  - $Harmonic(u) = 1$ ssi tous les neouds sont a distance 1 de u

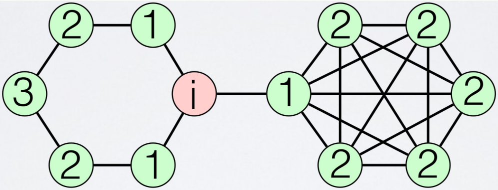
Centralite Harmonique

Definition (Harmonic centrality)

Variante de la closeness definie comme la moyenne de l'inverse des distances a tous les autres noeuds :

$$Harmonic(u) = \frac{1}{n-1} \sum_{v \in V \setminus \{u\}} \frac{1}{dist(u,v)}$$

- Remarque : bien definie meme pour les reseaux non connexes, en posant $\frac{1}{+\infty} = 0$.
- Comme pour la closeness, on a :
  - $Harmonic(u) \in [0,1]$
  - $Harmonic(u) = 1$ ssi tous les neouds sont a distance 1 de u
- Exemple :



Centralite de betweenness

- Mesure a quel point un noeud est present sur les plus courts chemins dans le reseau

Definition (Centralite de betweenness)

Somme des poucentages de plus courts chemins sur lesquels se trouve u, pour tous les couples s, t de sommets :

$$C_B(u) = \sum_{s \neq u \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}}$$

avec σ_{st} le nombre de plus courts chemins de s a t, et $\sigma_{st}(u)$ le nombre de tels chemins qui passent par u.

Question ???

Quel sont les valeurs maximum et minimum de la betweenness d'un sommet ?

Centralite de betweenness

Centralite de betweenness

$$C_B(u) = \sum_{s \neq u \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}}$$

Question ???

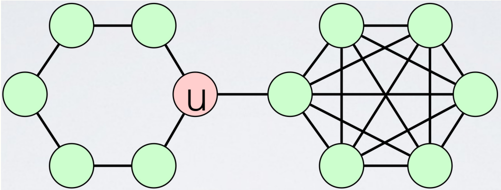
Est-ce grave que ce la betweenness soit sur une echelle ouverte ? (non normalisee entre 0 et 1)

Version normalisee de la betweenness

$$C_B^{norm}(u) = \frac{2C_B(u)}{(n-1)(n-2)}$$

Centralite de betweenness

Exemple :

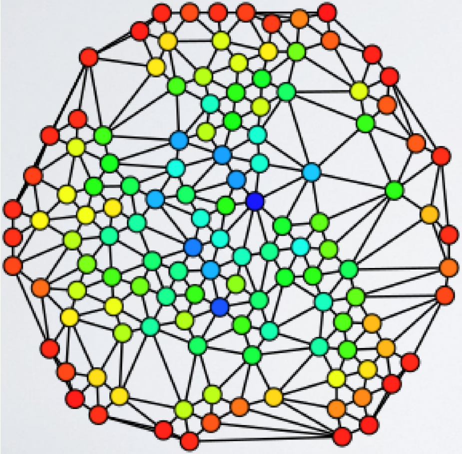


Complexite de calcul :

- Avec Ford-Fulkerson
  - Temps : $O(n^3)$
  - Espace : $O(n^2)$
- Avec BFS (astucieux)
  - Temps : $O(nm)$
  - Espace : $O(m)$

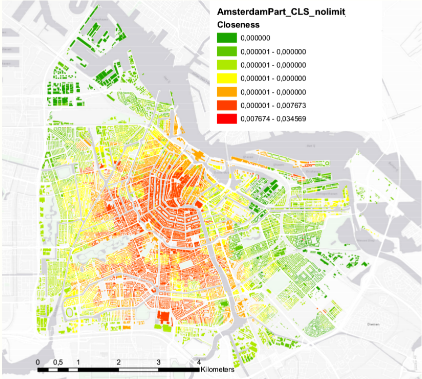
Algorithme pour le calcul de la betweenness

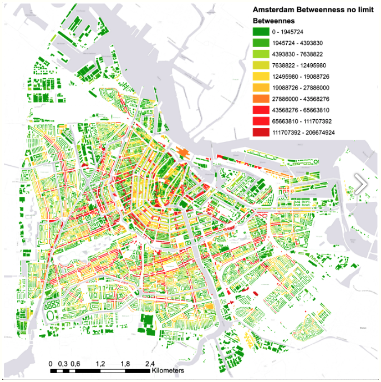
Centralite de betweenness



- Ressemblance avec closeness ?

Comparaison betweenness/closeness





Comparaison betweenness/closeness

Autre exemple de comparaison :

- Biparti complet equilibre + un sommet universel

Betweenness des liens

Une definition recursive de la centralite

- Les noeuds importants sont ceux connectes aux noeuds importants

Une definition recursive de la centralite

- Les noeuds importants sont ceux connectes aux noeuds importants
- plusieurs notions de centralite sont basees sur cette idee :
  - La centralite de vecteur propre (eigenvector centrality)
  - Le PageRank de Google

Une definition recursive de la centralite

Ce qu'on voudrait :

- Chaque noeud a un score (sa centralite)
- Chaque noeud envoie son score a ses voisins
- Le score d'un noeud est la somme des scores de ses voisins

Une definition recursive de la centralite

Ce qu'on voudrait :

- Chaque noeud a un score (sa centralite)
- Chaque noeud envoie son score a ses voisins
- Le score d'un noeud est la somme des scores de ses voisins

$$C_u = \sum_{v \in N(u)} C_v$$

Une definition recursive de la centralite

Ce qu'on voudrait :

- Chaque noeud a un score (sa centralite)
- Chaque noeud envoie son score a ses voisins
- Le score d'un noeud est la somme des scores de ses voisins

$$C_u = \sum_{v \in N(u)} C_v$$

- Un petit probleme :
  - Le score du noeud u est envoye $d^\circ(u)$ fois
  - La somme des scores dans le reseau augmente apres l'envoi
  - On ne peut donc pas avoir egalite des scores avant et apres dans tout le reseau !

Une definition recursive de la centralite

Ce qu'on voudrait :

- Chaque noeud a un score (sa centralite)
- Chaque noeud envoie son score a ses voisins
- Le score d'un noeud est la somme des scores de ses voisins

$$C_u = \sum_{v \in N(u)} C_v$$

- Un petit probleme :
  - Le score du noeud u est envoye $d^\circ(u)$ fois
  - La somme des scores dans le reseau augmente apres l'envoi
  - On ne peut donc pas avoir egalite des scores avant et apres dans tout le reseau !
- $\Longrightarrow$ il faut normaliser (facteur λ ci-dessous)

$$C_u = \frac{1}{\lambda} \sum_{v \in N(u)} C_v$$

La methode des puissances

$$C_u = \frac{1}{\lambda} \sum_{v \in N(u)} C_v$$

Deux questions :

- Comment trouver de tels scores ?
- Comment trouver λ ?

La methode des puissances

$$C_u = \frac{1}{\lambda} \sum_{v \in N(u)} C_v$$

Deux questions :

- Comment trouver de tels scores ?
- Comment trouver λ ?

La methode des puissances (power method) :

- On initialise tous les scores a des valeurs quelconques, exemples :

La methode des puissances

$$C_u = \frac{1}{\lambda} \sum_{v \in N(u)} C_v$$

Deux questions :

- Comment trouver de tels scores ?
- Comment trouver λ ?

La methode des puissances (power method) :

- On initialise tous les scores a des valeurs quelconques, exemples :
  - Le meme pour tous les noeuds

La methode des puissances

$$C_u = \frac{1}{\lambda} \sum_{v \in N(u)} C_v$$

Deux questions :

- Comment trouver de tels scores ?
- Comment trouver λ ?

La methode des puissances (power method) :

- On initialise tous les scores a des valeurs quelconques, exemples :
  - Le meme pour tous les noeuds
  - Un score aleatoire pour chaque noeud

La methode des puissances

$$C_u = \frac{1}{\lambda} \sum_{v \in N(u)} C_v$$

Deux questions :

- Comment trouver de tels scores ?
- Comment trouver λ ?

La methode des puissances (power method) :

- On initialise tous les scores a des valeurs quelconques, exemples :
  - Le meme pour tous les noeuds
  - Un score aleatoire pour chaque noeud
  - N'importe quoi d'autre (suppose plus proche du resultat final)

La methode des puissances

$$C_u = \frac{1}{\lambda} \sum_{v \in N(u)} C_v$$

Deux questions :

- Comment trouver de tels scores ?
- Comment trouver λ ?

La methode des puissances (power method) :

- On initialise tous les scores a des valeurs quelconques, exemples :
  - Le meme pour tous les noeuds
  - Un score aleatoire pour chaque noeud
  - N'importe quoi d'autre (suppose plus proche du resultat final)
- Chaque score est mis a jour par la regle ci-dessus, en notation matricielle : $X^{t+1} = \frac{1}{\lambda} A X^t$

La methode des puissances

$$C_u = \frac{1}{\lambda} \sum_{v \in N(u)} C_v$$

Deux questions :

- Comment trouver de tels scores ?
- Comment trouver λ ?

La methode des puissances (power method) :

- On initialise tous les scores a des valeurs quelconques, exemples :
  - Le meme pour tous les noeuds
  - Un score aleatoire pour chaque noeud
  - N'importe quoi d'autre (suppose plus proche du resultat final)
- Chaque score est mis a jour par la regle ci-dessus, en notation matricielle : $X^{t+1} = \frac{1}{\lambda} A X^t$
- On prend λ de sorte a garder la norme du vecteur constante, c'est a dire $\lambda = |AX^t|$.

La methode des puissances

Question ???

Est-ce que chacun des scores convergent vers une valeur stable ?

La methode des puissances

Question ???

Est-ce que chacun des scores convergent vers une valeur stable ?

Théorème (Perron-Frobenius)

Si la matrice A est la matrice d'adjacence d'un graphe non oriente connexe, alors :

La methode des puissances

Question ???

Est-ce que chacun des scores convergent vers une valeur stable ?

Théorème (Perron-Frobenius)

Si la matrice A est la matrice d'adjacence d'un graphe non oriente connexe, alors :

1. *Les valeurs propres de norme maximum en comprennent une (seule) qui est reelle et positive, notee λ.*

La methode des puissances

Question ???

Est-ce que chacun des scores convergent vers une valeur stable ?

Théorème (Perron-Frobenius)

Si la matrice A est la matrice d'adjacence d'un graphe non oriente connexe, alors :

1. *Les valeurs propres de norme maximum en comprennent une (seule) qui est reelle et positive, notee λ.*
2. *L'espace propre associe a λ est de dimension 1.*

La methode des puissances

Question ???

Est-ce que chacun des scores convergent vers une valeur stable ?

Théorème (Perron-Frobenius)

Si la matrice A est la matrice d'adjacence d'un graphe non oriente connexe, alors :

1. *Les valeurs propres de norme maximum en comprennent une (seule) qui est reelle et positive, notee λ.*
2. *L'espace propre associe a λ est de dimension 1.*
3. *Il contient un vecteur propre v dont toutes les composantes sont strictement positives (unique sous $|v| = 1$).*

La methode des puissances

Question ???

Est-ce que chacun des scores convergent vers une valeur stable ?

Théorème (Perron-Frobenius)

Si la matrice A est la matrice d'adjacence d'un graphe non oriente connexe, alors :

1. *Les valeurs propres de norme maximum en comprennent une (seule) qui est reelle et positive, notee λ.*
2. *L'espace propre associe a λ est de dimension 1.*
3. *Il contient un vecteur propre v dont toutes les composantes sont strictement positives (unique sous $|v| = 1$).*

Et enfin, **la methode des puissances converge vers** v.

Centralite de vecteur propre

Definition :

La centralite de vecteur propre est celle qui a chaque noeud associe sa composante dans le vecteur propre de Perron-Frobenius.

Centralite de vecteur propre

Definition :
La centralite de vecteur propre est celle qui a chaque noeud associe sa composante dans le vecteur propre de Perron-Frobenius.

- Cette centralite est celle qui correspond a l'idee que l'importance d'un noeud est une sorte de moyenne de l'importance de ses voisins

Centralite de vecteur propre

Definition :
La centralite de vecteur propre est celle qui a chaque noeud associe sa composante dans le vecteur propre de Perron-Frobenius.

- Cette centralite est celle qui correspond a l'idee que l'importance d'un noeud est une sorte de moyenne de l'importance de ses voisins
- La centralite de vecteur propre peut se calculer avec la methode des puissances

Centralite de vecteur propre

Definition :
La centralite de vecteur propre est celle qui a chaque noeud associe sa composante dans le vecteur propre de Perron-Frobenius.

- Cette centralite est celle qui correspond a l'idee que l'importance d'un noeud est une sorte de moyenne de l'importance de ses voisins
- La centralite de vecteur propre peut se calculer avec la methode des puissances
- Elle marche aussi dans le cas des graphes orientes (avec condition supplementaire)

Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

- La matrice d'adjacence est assymetrique

Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)

Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)

Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)
    - on utilise les vecteurs propres a droite

Centralite de vecteur propre pour graphes orientes

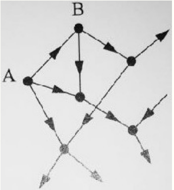
Quelques difficultes specifiques aux graphes orientes :

- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)
    - on utilise les vecteurs propres a droite
    - Ce sont les voisins entrants qui contribuent

Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

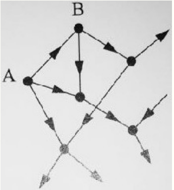
- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)
    - on utilise les vecteurs propres a droite
    - Ce sont les voisins entrants qui contribuent
- Les noeuds sources n'ont pas de voisins entrants !



Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

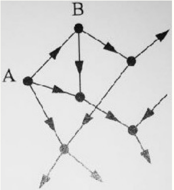
- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)
    - on utilise les vecteurs propres a droite
    - Ce sont les voisins entrants qui contribuent
- Les noeuds sources n'ont pas de voisins entrants !
  - A a seulement des voisins sortants : $C_A = 0$



Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

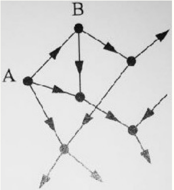
- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)
    - on utilise les vecteurs propres a droite
    - Ce sont les voisins entrants qui contribuent
- Les noeuds sources n'ont pas de voisins entrants !
  - A a seulement des voisins sortants : $C_A = 0$
  - Le seul voisin entrant de B est A (et $C_A = 0$) : $C_B = 0$



Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

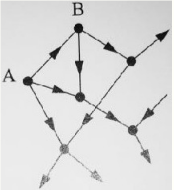
- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)
    - on utilise les vecteurs propres a droite
    - Ce sont les voisins entrants qui contribuent
- Les noeuds sources n'ont pas de voisins entrants !
  - A a seulement des voisins sortants : $C_A = 0$
  - Le seul voisin entrant de B est A (et $C_A = 0$) : $C_B = 0$
  - etc.



Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)
    - on utilise les vecteurs propres a droite
    - Ce sont les voisins entrants qui contribuent
- Les noeuds sources n'ont pas de voisins entrants !
  - A a seulement des voisins sortants : $C_A = 0$
  - Le seul voisin entrant de B est A (et $C_A = 0$) : $C_B = 0$
  - etc.
    - Seules les composantes fortement connexes ont centralite non nulle



Centralite de vecteur propre pour graphes orientes

Quelques difficultes specifiques aux graphes orientes :

- La matrice d'adjacence est assymetrique
  - 2 ensembles de vecteurs propres (a gauche et a droite)
  - 2 vecteurs propres principaux (ceux de Perron-Frobenius)
    - on utilise les vecteurs propres a droite
    - Ce sont les voisins entrants qui contribuent
- Les noeuds sources n'ont pas de voisins entrants !
  - A a seulement des voisins sortants : $C_A = 0$
  - Le seul voisin entrant de B est A (et $C_A = 0$) : $C_B = 0$
  - etc.
    - Seules les composantes fortement connexes ont centralite non nulle

Question ???

Dans quel type de reseau n'y a-t-il jamais de composantes fortement connexes ?

PageRank (Google)

Centralite de vecteur propre generalisee

PageRank

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Brin, S. and Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: Seventh International World-Wide Web Conference (WWW 1998), April 14-18, 1998, Brisbane, Australia.

Sergey Brin and Lawrence Page

Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/

PageRank (Google)

Sergey Brin received his B.S. degree in mathematics and computer science from the University of Maryland at College Park in 1993. Currently, he is a Ph.D. candidate in computer science at Stanford University where he received his M.S. in 1995. He is a recipient of a National Science Foundation Graduate Fellowship. His research interests include search engines, information extraction from unstructured sources, and data mining of large text collections and scientific data.

Lawrence Page was born in East Lansing, Michigan, and received a B.S.E. in Computer Engineering at the University of Michigan Ann Arbor in 1995. He is currently a Ph.D. candidate in Computer Science at Stanford University. Some of his research interests include the link structure of the web, human computer interaction, search engines, scalability of information access interfaces, and personal data mining.

PageRank

Cadre :

graphes orientes (web, liens hypertextes)

Deux ameliorations/changements par rapport a eigenvector

- Probleme des noeuds sources
  - Ajout d'un gain de centralite constant a tout les sommets
- Les noeuds de forte centralite donnent une forte centralite a tous leurs voisins
  - La centralite d'un noeud est plutot divisee entre ses voisins (conservativite)

$$C_u^{t+1} = \frac{1}{\lambda} \sum_{v \in N^-(u)} C_v^t \quad \Longrightarrow \quad C_u^{t+1} = \alpha \sum_{v \in N^-(u)} \frac{C_v^t}{d^+(v)} + \beta$$

Avec, par convention, $\beta = 1$ et α un parametre (souvent 0.85) qui controle l'importance de β

PageRank - interpretation matricielle

- Le pagerank est juste le vecteur propre principal (Perron-Frobenius) de la matrice modifiee de Google
- La matrice de Google G est definie ainsi
  - Normalisation S des colonnes de A
  - On ajoute $\frac{1-\alpha}{n}$ a toutes les entrees
- On obtient $G = \alpha S + \frac{1-\alpha}{n} E$
  - Avec E la matrice ou toutes les entrees valent 1.

(a)
$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

(c)
$$S = \begin{pmatrix} 0 & 1/2 & 1/3 & 0 & 1/5 \\ 1 & 0 & 1/3 & 1/3 & 1/5 \\ 0 & 1/2 & 0 & 1/3 & 1/5 \\ 0 & 0 & 1/3 & 0 & 1/5 \\ 0 & 0 & 0 & 1/3 & 1/5 \end{pmatrix}$$

(e)
$$G = \begin{pmatrix} 0.03 & 0.455 & 0.313 & 0.03 & 0.2 \\ 0.88 & 0.03 & 0.313 & 0.313 & 0.2 \\ 0.03 & 0.455 & 0.03 & 0.313 & 0.2 \\ 0.03 & 0.03 & 0.313 & 0.03 & 0.2 \\ 0.03 & 0.03 & 0.03 & 0.313 & 0.2 \end{pmatrix}$$

PageRank - interpretation comme marche aleatoire

PageRank : une marche aleatoire avec reinitialisation

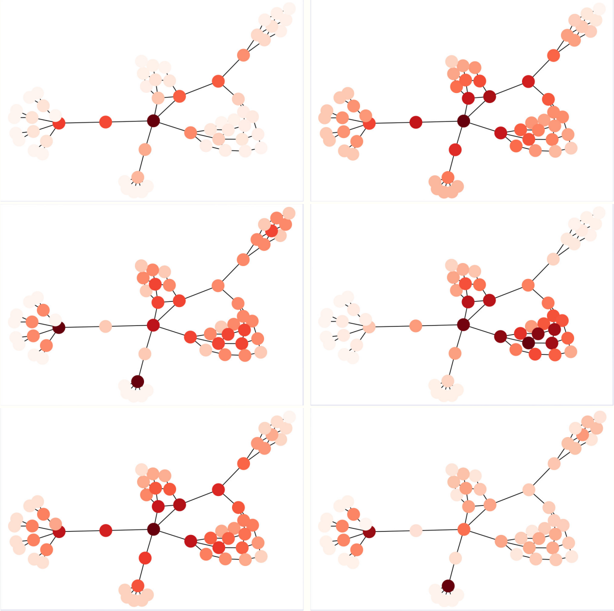
- Une marche aleatoire
  - Quand il est sur un sommet v,
  - le marcheur se deplace sur un voisin sortant u de v
  - choisi unif. alea. parmi les voisins de v
    - D'ou la division par $d^+(v)$
- On ajoute une possibilite de teleportation (reinitialisation)
  - Avec proba α le marcheur suit le comportement de la marche aletoire classique
  - Avec proba $1 - \alpha$, il se teleporte sur n'importe lequel des n noeuds du reseau (y compris lui-meme)
    - D'ou l'ajout du terme $\frac{1-\alpha}{n}$ sur toutes les entrees
    - Sur le neoud d'arrivee, la somme de toutes les possibilites de teleportation donne : $n\frac{1-\alpha}{n} = 1 - \alpha = cte = \beta$ precedant
- Le PageRank du noeud u est la probabilite pour le marcheur d'etre sur u apres un nombre infini d'etapes

PageRank (Google)

Comment fait Google pour traiter nos recherches ?

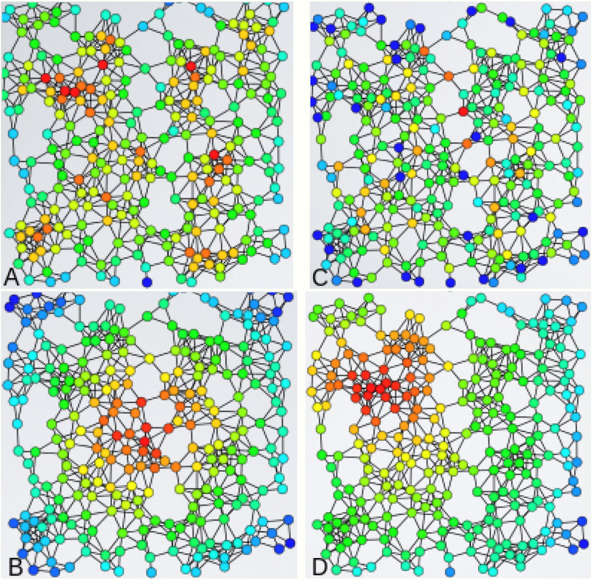
- Calcule le PageRank (avec la methode des puissances, ou autre)
- Sort les documents lies a notre recherche range par PageRank decroissant
- Bien sur, maintenant, Google est certainement beaucoup plus complexe...
  - "Most search engine development has gone on at companies with little publication of technical details. This causes search engine technology to remain largely a black art" **[Page, Brin, 1997]** ;) :D

Qui est qui ?



Degree
Closeness
Harmonic
Betweenness
Eigenvector
PageRank

Qui est qui ? (2)



Degree
Closeness
Betweenness
Eigenvector