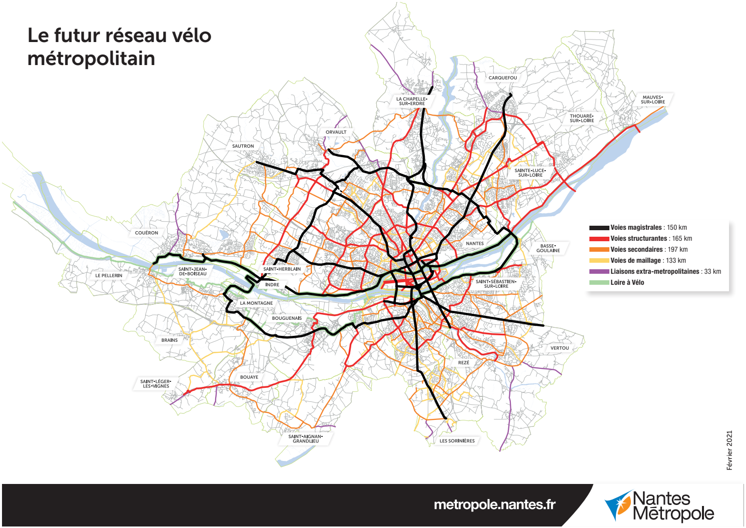
R2.07 Graphes

Corentin Dufourg, Régis Fleurquin et Thibault Godin
IUT de Vannes Informatique

Motivation : SAÉ Vélo



Source : Nantes Métropole

Dans $G = (S, A)$, les voisins d'un sommet s sont les sommets t tels que sAt. On note $V(i)$ cette ensemble.

Dans $G = (S, A)$, les voisins d'un sommet s sont les sommets t tels que sAt.

On note $V(i)$ cette ensemble.

Le degré $d(s)$ d'un sommet s est le de voisins de s dans G.

Dans $G = (S, A)$, les voisins d'un sommet s sont les sommets t tels que sAt. On note $V(i)$ cette ensemble.

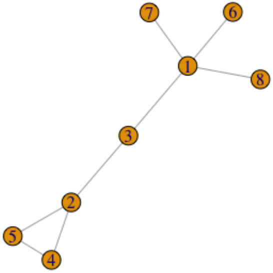
Le degré $d(s)$ d'un sommet s est le de voisins de s dans G.

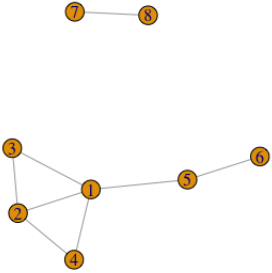
$$d(s) = |V(s)|$$

Dans $G = (S, A)$, les voisins d'un sommet s sont les sommets t tels que sAt. On note $V(i)$ cette ensemble.

Le degré $d(s)$ d'un sommet s est le de voisins de s dans G.

$$d(s) = |V(s)|$$





Centralité de degré

L'indicateur de centralité le plus simple est la centralité de degré ou degree centrality, aussi appelée centralité de prestige :

$$C_D(i) = d(i)$$

Centralité de degré

L'indicateur de centralité le plus simple est la centralité de degré ou degree centrality, aussi appelée centralité de prestige :

$$C_D(i) = d(i)$$

Intuitivement, un sommet très connecté aux autres est très important dans un graphe et on l'appelle un hub.

Centralité de degré

L'indicateur de centralité le plus simple est la centralité de degré ou degree centrality, aussi appelée centralité de prestige :

$$C_D(i) = d(i)$$

Intuitivement, un sommet très connecté aux autres est très important dans un graphe et on l'appelle un hub.

Application : personnalités influentes dans un réseau (ex. Instagram)

L'**excentricité** d'un sommet est la distance maximale de ce sommet aux autres sommets du graphes.

$\xi(i) = \max_j d_G(i, j)$

L'**excentricité** d'un sommet est la distance maximale de ce sommet aux autres sommets du graphes.

$\xi(i) = \max_j d_G(i, j)$

Le **diamètre** d'un graphe est le maximum des excentricité

$\delta = \max_i \max_j d_G(i, j) = \max_i \xi(i)$

L'**excentricité** d'un sommet est la distance maximale de ce sommet aux autres sommets du graphes.

$\xi(i) = \max_j d_G(i,j)$

Le **diamètre** d'un graphe est le maximum des excentricité

$\delta = \max_i \max_j d_G(i,j) = \max_i \xi(i)$

Le **rayon** d'un graphe est le minimum des excentricité

$\rho = \min_i \max_j d_G(i,j) = \min_i \xi(i)$

L'**excentricité** d'un sommet est la distance maximale de ce sommet aux autres sommets du graphes.

$\xi(i) = \max_j d_G(i, j)$

Le **diamètre** d'un graphe est le maximum des excentricité

$\delta = \max_i \max_j d_G(i, j) = \max_i \xi(i)$

Le **rayon** d'un graphe est le minimum des excentricité

$\rho = \min_i \max_j d_G(i, j) = \min_i \xi(i)$

L'excentricité peut être utilisée comme un indicateur de centralité

$C_{exc}(i) = \xi(i)$

Le diamètre $\rightsquigarrow$ pire des cas diffusion information ou une autre ressource dans le réseau

Le diamètre $\rightsquigarrow$ pire des cas diffusion information ou une autre ressource dans le réseau

WWW $\rightsquigarrow$ distance moyenne entre 2 pages ≈ 19 (Albert, Barabasi, 1999)

Le diamètre $\rightsquigarrow$ pire des cas diffusion information ou une autre ressource dans le réseau

WWW $\rightsquigarrow$ distance moyenne entre 2 pages ≈ 19 (Albert, Barabasi, 1999)

Un(e famille de) réseau *petit monde* si la distances moyenne $\bar{\ell} \propto \log(|V|)$

Le diamètre $\rightsquigarrow$ pire des cas diffusion information ou une autre ressource dans le réseau

WWW $\rightsquigarrow$ distance moyenne entre 2 pages ≈ 19 (Albert, Barabasi, 1999)

Un(e famille de) réseau *petit monde* si la distances moyenne $\bar{\ell} \propto \log(|V|)$

Milgram (1967) six degrees of separation $\bar{\ell} \approx 6$

Le diamètre $\rightsquigarrow$ pire des cas diffusion information ou une autre ressource dans le réseau

WWW $\rightsquigarrow$ distance moyenne entre 2 pages ≈ 19 (Albert, Barabasi, 1999)

Un(e famille de) réseau *petit monde* si la distances moyenne $\bar{\ell} \propto \log(|V|)$

Milgram (1967) six degrees of separation $\bar{\ell} \approx 6$

Bhagat et al. (2016) Facebook $\bar{\ell} \approx 3,5$ (three and a half degrees of separation).

Le diamètre $\rightsquigarrow$ pire des cas diffusion information ou une autre ressource dans le réseau

WWW $\rightsquigarrow$ distance moyenne entre 2 pages ≈ 19 (Albert, Barabasi, 1999)

Un(e famille de) réseau *petit monde* si la distances moyenne $\bar{\ell} \propto \log(|V|)$

Milgram (1967) six degrees of separation $\bar{\ell} \approx 6$

Bhagat et al. (2016) Facebook $\bar{\ell} \approx 3,5$ (three and a half degrees of separation).

Bacon number (films), Erdös number (Maths)

La centralité de proximité (closeness centrality) du sommet s est définie par

$$C_P(s) = \left(\sum_{t \in S} d(s,t) \right)^{-1}$$

La centralité de proximité (closeness centrality) du sommet s est définie par

$$C_P(s) = \left(\sum_{t \in S} d(s,t) \right)^{-1}$$

C'est une mesure proche de l'excentricité

La centralité de proximité (closeness centrality) du sommet s est définie par

$$C_P(s) = \left(\sum_{t \in S} d(s,t) \right)^{-1}$$

C'est une mesure proche de l'excentricité

La centralité d'intermédiarité (betweenness centrality) du sommet s est définie par

$$C_B(s) = \sum_{t,t' \in S} \frac{P_{t,t'}(s)}{P_{t,t'}}$$

avec $P_{t,t'}$ le nombre de plus courts chemins de t à t', et $P_{t,t'}(s)$ le nombre de plus courts chemins de t à t' passant par s

La centralité de proximité (closeness centrality) du sommet s est définie par

$$C_P(s) = \left(\sum_{t \in S} d(s,t) \right)^{-1}$$

C'est une mesure proche de l'excentricité

La centralité d'intermédiarité (betweenness centrality) du sommet s est définie par

$$C_B(s) = \sum_{t,t' \in S} \frac{P_{t,t'}(s)}{P_{t,t'}}$$

avec $P_{t,t'}$ le nombre de plus courts chemins de t à t', et $P_{t,t'}(s)$ le nombre de plus courts chemins de t à t' passant par s

$C_B(s)$ élevé $\rightsquigarrow$ passage sur un grand nombre de chemins les plus courts $\rightsquigarrow$ part importante de la communication passe par s

La centralité de proximité (closeness centrality) du sommet s est définie par

$$C_P(s) = \left(\sum_{t \in S} d(s,t)\right)^{-1}$$

C'est une mesure proche de l'excentricité

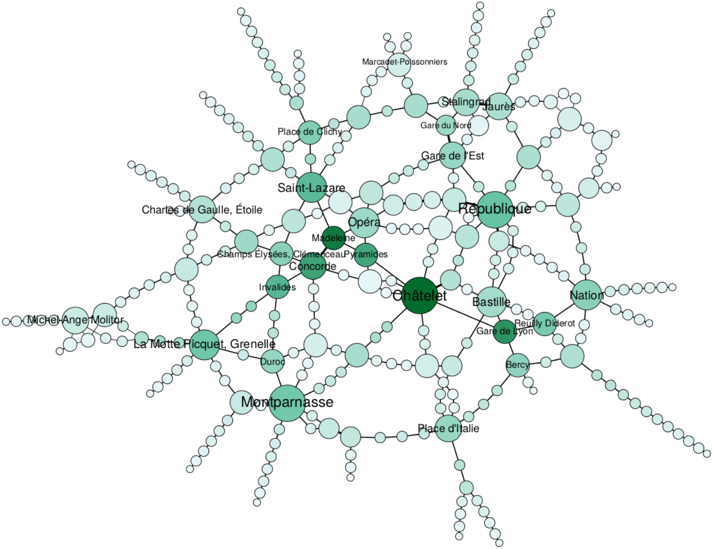
La centralité d'intermédiarité (betweenness centrality) du sommet s est définie par

$$C_B(s) = \sum_{t,t' \in S} \frac{P_{t,t'}(s)}{P_{t,t'}}$$

avec $P_{t,t'}$ le nombre de plus courts chemins de t à t', et $P_{t,t'}(s)$ le nombre de plus courts chemins de t à t' passant par s

$C_B(s)$ élevé $\rightsquigarrow$ passage sur un grand nombre de chemins les plus courts $\rightsquigarrow$ part importante de la communication passe par s

même définition pour les arêtes



Représentation du métro parisien avec degré (taille) et centralité d'intermédiarité (couleur). Source : R. Charbey

https://networkx.org/documentation/stable/reference/algorithms/centrality.html

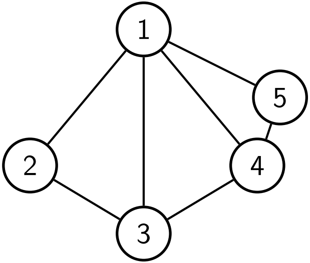
Définition

Une composante connexe $\mathcal{C}$ d'un graphe $\mathcal{G} = (S, A)$ est un sous-ensemble *maximal* de sommets tq toute paire de sommets est relié par une chaîne. Un graphe $\mathcal{G}$ est dit **connexe** ssi il possède une unique composante connexe.

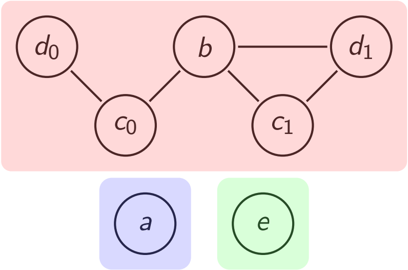
Définition

Une composante connexe $\mathcal{C}$ d'un graphe $\mathcal{G} = (S, A)$ est un sous-ensemble *maximal* de sommets tq toute paire de sommets est relié par une chaîne. Un graphe $\mathcal{G}$ est dit **connexe** ssi il possède une unique composante connexe.

Un graphe connexe :



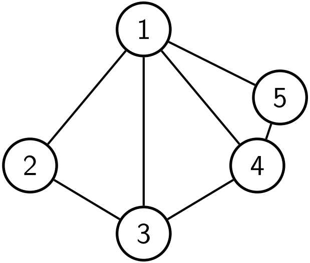
Un graphe décomposé en composantes connexes :



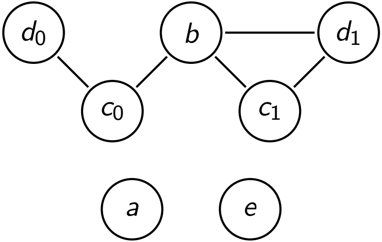
Définition

Une composante connexe $\mathcal{C}$ d'un graphe $\mathcal{G} = (S, A)$ est un sous-ensemble *maximal* de sommets tq toute paire de sommets est relié par une chaîne. Un graphe $\mathcal{G}$ est dit **connexe** ssi il possède une unique composante connexe.

Un graphe connexe :



Un graphe décomposé en composantes connexes :




```
nx.components(G)
```

Un *point d'articulation* ou cutpoint est un sommet qui, si on le supprime du graphe, augmente le nombre de composantes connexes.

Un *point d'articulation* ou cutpoint est un sommet qui, si on le supprime du graphe, augmente le nombre de composantes connexes.

Aux points d'articulation avec une centralité d'intermédiarité élevée, le graphe est très vulnérable en terme de communication.

Un *point d'articulation* ou cutpoint est un sommet qui, si on le supprime du graphe, augmente le nombre de composantes connexes.

Aux points d'articulation avec une centralité d'intermédiarité élevée, le graphe est très vulnérable en terme de communication.

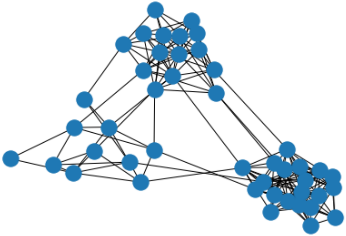
un graphe connexe est dit k-sommet-connexe s'il possède au moins $k + 1$ sommets et s'il reste encore connexe après avoir supprimé $k - 1$ sommets.

Un pont ou isthme ou bridge est une arête qui, si on la retire du graphe, augment le nombre de composantes connexes

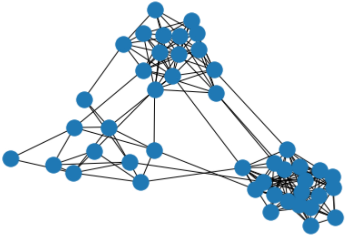
un graphe connexe est dit k-arêtes-connexe s'il possède au moins $k+1$ sommets et s'il reste encore connexe après avoir supprimé $k-1$ arêtes.

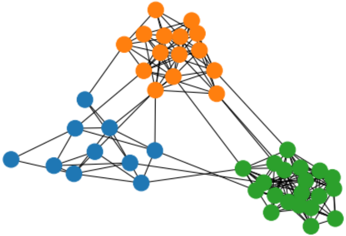
```
nx.is_k_edge_connected(G)
```

Clustering



Clustering





Clustering

coefficient de clustering global :

$$C = \frac{3 \times |\text{triangles}|}{|\text{paires de voisins distincts}|}$$

Clustering

coefficient de clustering global :

$$C = \frac{3 \times |\text{triangles}|}{|\text{paires de voisins distincts}|}$$

coefficient de clustering local d'un sommet :

$$C_i = \frac{|\text{triangles de sommet } i|}{|\text{paires de voisins distincts de } i|}$$

soit

$$C_i = \frac{|\text{triangles de sommet } i|}{\binom{d_i}{2}}$$

C'est la fraction de ses paires de voisins connectés, égale à 0 si $d_i \leq 1$ par convention.

Clustering

On a $C_i \leq 1$, avec égalité si et seulement si le sommet i et son voisinage forment une clique d'au moins 3 nœuds.

Clustering

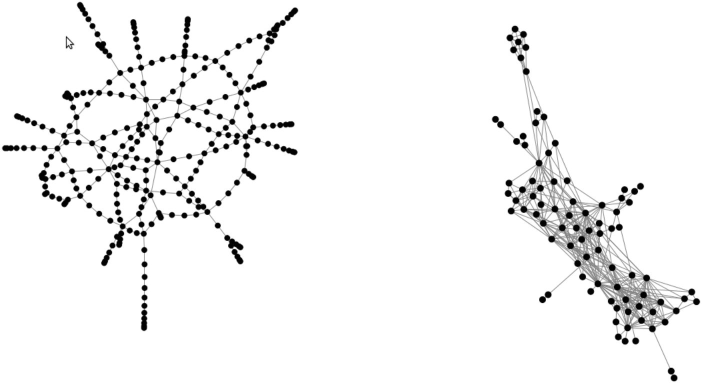
On a $C_i \leq 1$, avec égalité si et seulement si le sommet i et son voisinage forment une clique d'au moins 3 nœuds.

En prenant la moyenne des coefficients locaux, on obtient le coefficient local moyen :

$$\bar{C} = \frac{\sum_{i \in V} C_i}{|V|}$$

On a également $\bar{C} \leq 1$, avec égalité si et seulement si le graphe est un ensemble de cliques de taille au moins 3.

Analyse



métro	1 069	418	56	10	2	0
social	10 660	5 760	8 646	376	2 213	727

Deux réseaux de même nombre d'arêtes, à gauche
le réseau du métro parisien et à droite un réseau égocentré

Comparaison de deux réseaux. Source : R. Charbey