Mesures locales et globales des réseaux

Ahmed Mohameden
amed.mohameden@gmail.com

Université Cheikh Anta Diop
Faculté des Sciences et Techniques
Département de Mathématiques et Informatique
Section Informatique

1^er^ avril 2018

Plan

A quoi servent les mesures ?

Classification de mesures

Mesures locales de voisinage

Mesures locales d'ensemble

Mesures globales

Limites et faiblesses

A quoi servent les mesures ?

Les mesures locales/globales constituent un outil indispensable dans l'analyse des réseaux sociaux. Ils permettent, entre autre, de :

1. Identifier
    - les nœuds intermédiaires
    - les nœuds populaires ou encore les nœuds isolés
    - les nœuds qui communiquent le plus (nœuds actifs)
    - les nœuds les plus centraux, c'est à dire, les plus proches de l'ensemble des nœuds du réseau
2. Exprimer la probabilité que les voisins d'un nœud soient eux-mêmes connectés par un lien
3. Avoir une idée sur le niveau de propagation d'information au sein du réseau
4. Mesurer le degré d'interconnexion de nœuds du réseau

Classification de mesures

Une classification possible est celle par notion de voisinage, où les mesures se classent en 3 catégories :

1. Mesures locales de voisinage
2. Mesures locales d'ensemble
3. Mesures globales

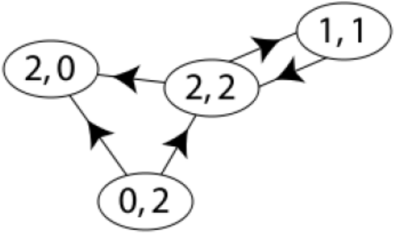
Mesures locales de voisinage

Les mesures locales de voisinage décrivent la situation d'un élément par rapport à ses voisins immédiats, c'est à dire, directement connectés ou adjacents. Exemples :

- Centralité de degré, degré pondéré, centralité combinée, centralité spectrale, ...

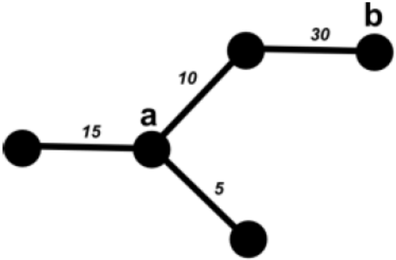
Centralité de degré

La centralité de degré constitue la forme la plus intuitive de centralité. Elle indique le nombre de voisins immédiats d'un nœud.



Degré pondéré

Le degré pondéré permet de prendre en compte l'intensité de flux (nombre de voyageurs, nombre d'appels téléphoniques, nombre de publications scientifiques, ...). Il est défini par la somme de poids des liens adjacents au nœud voulu.



Centralité combinée

Comme son nom l'indique, cette mesure combine la centralité de degré et le degré pondéré dans le but de pallier leurs faiblesses. La centralité combinée constitue le rapport entre le degré pondéré et la centralité de degré à l'exposant α, multiplié par la centralité de degré. Le choix de la valeur de α dépend de la mesure que nous voulons favoriser : si $0 \leqslant \alpha \leqslant 1$ c'est à dire que nous donnons plus d'importance à la centralité de degré, tandis que le choix d'une valeur supérieure à 1 favorise le rôle des poids de liens. Formellement :

$$C(v) = d(v) \times \left(\frac{d_w(v)}{d(v)}\right)^{\alpha} \tag{1}$$

Centralité spectrale

La centralité spectrale (ou centralité des vecteurs propres) mesure à quel degré un nœud est connecté aux nœuds fortement connectés dans le réseau. La connectivité d'un nœud est indiquée par le nombre de ses voisins. Cette mesure est définie par la combinaison linéaire des centralités des noeuds voisins du nœud en question. Formellement :

$$C^{spec}(v_i) = \frac{1}{\lambda} \times \sum_{j=1}^{n} a_{ij} C^{spec}(v_j) \tag{2}$$

Coefficient de clustering

Le coefficient de clustering est égal au nombre de triades fermées (triangles) divisé par le nombre de triades possibles. Il représente la probabilité que quand deux sommets ont un voisin en commun, le troisième lien existe. Plus ce coefficient est grand, plus le voisinage est proche d'une clique. Formellement :

$$C_i = \frac{\sum_{j \in V_i} \sum_{k \in V_i} a_{jk}}{d_i(d_i - 1)} \tag{3}$$

Mesures locales d'ensemble

Les mesures locales d'ensemble prennent en compte la situation d'un élément par rapport à tous les autres éléments de même nature présents dans le réseau. Exemples :

- Nombre de Koenig, centralité d'intermédiarité, centralité de proximité, ...

Nombre de Koenig

Le nombre de Koenig, appelé aussi excentricité, désigne le nombre de liens servant à connecter le sommet le plus distant. Formellement :

$$e(v) = \max_{u \in V} d(u, v) \tag{4}$$

Centralité d'intermédiarité (1/2)

La centralité d'intermédiarité mesure le nombre de fois où un nœud est l'intermédiaire sur le plus court chemin entre deux autres. Formellement :

$$C^{int}(v_i) = \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{g_{jk}(v_i)}{g_{jk}} \tag{5}$$

Où $g_{jk}(v_i)$ est le nombre total de chemins géodésiques entre les noeuds v_j et v_k qui passent par le nœud v_i, et g_{jk} est le nombre total des plus courts chemins entre les noeuds v_j et v_k.

Centralité d'intermédiarité (2/2)

Un niveau élevé de centralité d'intermédiarité n'est pas forcement corrélé avec un score élevé de centralité de degré. Les noeuds ayant une forte centralité d'intermédiarité constituent des « points de passages importants » pour relier rapidement deux sommets du graphe.

Centralité de proximité

La centralité de proximité dépend inversement de l'indice de Shimbel, défini par la somme des longueurs des plus courts chemins permettant de relier tous les autres sommets. Elle permet de déterminer les sommet les plus centraux en identifiant ceux qui sont les plus proches de l'ensemble des sommets du graphe. Formellement :

$$C^{prox}(v_i) = \frac{1}{\sum_{j=1}^{n} d(v_i, vj)} \tag{6}$$

Un score de proximité élevé indique que le nœud est un élément efficace dans la transmission de l'information au sein du réseau.

Mesures globales

Les mesures globales décrivent la situation globale du réseau. Exemples :

1. Densité
2. Rayon et diamètre
3. Distribution des degrés

Densité

La densité d'un réseau désigne le rapport entre le nombre de liens présents et le nombre de liens possibles. Elle varie entre 0 (graphe sans aucun lien) et 1 (graphe complet).

Diamètre et rayon

Le diamètre d'un réseau signifie la plus grande distance possible qui puisse se trouver entre deux de ses nœuds. La distance entre deux sommets étant définie par la longueur d'un plus court chemin entre ces deux sommets.
En revanche, le rayon désigne la plus petite distance possible qui puisse exister entre deux de ses nœuds.

Distribution des degrés

La distribution des degrés représente la distribution des probabilité de degré de noeuds. Formellement, elle est définie par la fraction de nœuds ayant un degré égal à k. Par conséquent, si le réseau est composé de n nœud dont n_k d'eux sont de degré k, alors $P(k) = \frac{n_k}{n}$.

Limites et faiblesses des mesures précédentes

Les principales faiblesses des mesures précédentes sont :

1. La même mesure peut qualifier des graphes de structures très différentes. Exemple : la densité
2. L'utilisation de la somme conduit à un biais si la répartition des valeurs sommées (nombre de voisins, poids, ...) est déséquilibrée
3. L'aspect manquant aux mesures précédentes et la combinaison entre l'aspect topologique et l'intensité de communication