Mesures d'influence via les indicateurs de centralité dans les réseaux sociaux

Oualid Benyahia [1], **Christine Largeron** [2]

1. *Universite Jean Monnet*
Laboratoire Hubert Curien UMR CNRS 5516
18 rue Lauras 42000 Saint-Etienne, France
oualid.benyahia@univ-st-etienne.fr

2. *Universite Jean Monnet*
Laboratoire Hubert Curien UMR CNRS 5516
18 rue Lauras 42000 Saint-Etienne, France
Christine.Largeron@univ-st-etienne.fr

ABSTRACT. *Identifier les acteurs importants dans un réseau est une tâche difficile qui est cependant très utile dans diverses applications comme par exemple dans les systèmes de recommandation, le marketing ou encore l'analyse épidémiologique où il est essentiel de cibler des acteurs susceptibles de favoriser ou limiter un phénomène de propagation. Dans cette optique, plusieurs mesures ont été proposées dans le cadre de l'analyse des réseaux sociaux. C'est le cas notamment des indicateurs de centralité qui permettent d'évaluer, selon différents point de vue, l'importance d'un acteur en fonction de ses relations aux autres. Dans cet article, nous proposons des variantes de ces mesures de centralité qui exploitent non seulement la structure du réseau mais aussi des attributs décrivant les acteurs. Des expérimentations, réalisées à partir de données extraites du réseau ArnetMiner, ont permis de valider l'apport de ces nouvelles mesures.*

KEYWORDS: *analyse des réseaux sociaux, mesure de centralité, mesure d'influence, graphes attribués*

1. Introduction

Dans la littérature sur les réseaux sociaux, des indicateurs permettent d'identifier des acteurs importants dans le réseau comme les mesures de centralité (Freeman, 1979 ; Wasserman, Faust, 1994). Un des avantages de ces mesures est qu'elles sont faciles à mettre en œuvre. De plus, elles peuvent être mises à jour en tenant compte seulement d'une information locale correspondant à une fraction du réseau. Ces mesures permettent en fait de comparer la position d'un acteur relativement à celles des autres selon différents points de vue. Ainsi, la centralité de degré permet de repérer les acteurs qui ont le plus de voisins directs et la centralité de proximité ceux qui peuvent joindre le plus facilement tous les autres membres du réseau. De même la centralité d'intermédiarité permet d'identifier les acteurs qui ont le plus de chance d'être entre deux autres acteurs du réseau alors que la centralité des valeurs propres ou sa variante, le PageRank, ceux qui sont le plus en relation avec des membres eux même fortement reliés à d'autres acteurs du réseau. Ces mesures peuvent être calculées pour des réseaux représentés par un graphe orienté ou non. Ainsi, par exemple, la centralité de degré, peut être définie en considérant tous les voisins d'un nœud ou bien ses successeurs ou encore ses prédécesseurs. De même certaines mesures sont applicables à des graphes valués ou non valués. Ainsi, par exemple, la centralité des valeurs propres peut être déterminée en tenant compte de la valuation du lien existant entre deux acteurs traduisant l'intensité de leur relation, de même que les centralités de degré et de proximité (Barrat *et al.*, 2004 ; Opsahl *et al.*, 2010).

Cependant, proposées dès les débuts de l'analyse des réseaux sociaux, ces mesures ont été conçues pour analyser des réseaux qui peuvent être décrits par un graphe simple dont les sommets correspondent aux acteurs et les liens aux relations qui existent entre eux. Or, les réseaux actuels, en particulier les réseaux d'information, sont représentés sous la forme de graphes plus complexes comme les graphes à attributs dans lesquels des caractéristiques, tel que le sexe ou l'âge, sont associées aux sommets pour les décrire. La prise en compte de ces informations complémentaires s'étant révélée pertinente notamment dans le cadre des modèles de propagation, nous avons donc estimé qu'elle méritait de l'être dans la définition des mesures de centralité.

Dans cette optique, Evrett et Borgatti (Everett, Borgatti, 2012) ont considéré le cas des graphes attribués, où les nœuds sont caractérisés par des attributs catégoriels, pour déterminer les groupes de nœuds. Ils intègrent les attributs des sommets dans le calcul des mesures de centralité en utilisant deux métriques connues: le *E-I homophily index* (Krackhardt, Stern, 1988) et le *Gould and Fernandez brokerage metrics* (Gould, Fernandez, 1989). L'index E-I consiste a partitionner la centralité de degré selon les groupes d'appartenance des nœuds. Il est ainsi généralisé pour être intégré dans les autres types de centralité tels que la centralité de proximité et la centralité des valeurs propres. De même, pour la centralité d'intermédiarité elle est généralisée via un partitionnement des nœuds (Gould and Fernandez brokerage metrics) pour diviser le voisinage d'un nœud en différents groupes.

Le but de notre travail a aussi été de revisiter des indicateurs de centralité, en tenant compte de la structure du réseau notamment du sens des relations et des valuations qui peuvent les quantifier mais aussi d'informations complémentaires décrivant les acteurs et qui sont susceptibles d'indiquer la notoriété d'un acteur indépendamment de la structure du réseau. Par contre, à la différence de Everett et Borgatti (2012), nous considérons le cas de graphes attribués où les nœuds sont caractérisés par des attributs numériques.

L'article est organisé comme suit. La section 2 décrit les différentes mesures que nous avons introduites et la section 3 présente les résultats d'une expérimentation menée sur des données réelles pour évaluer l'intérêt des mesures que nous avons proposées.

2. Mesures d'influence par les indicateurs de centralité

Un réseau d'information peut être représenté par un graphe avec attributs $G = (V, E)$ où V est l'ensemble des sommets, $E \subset V \times V$ est l'ensemble des arêtes et où chaque sommet $v_i \in V$ est associé à un vecteur numérique $Y^i = (y_1^i, y_2^i, \ldots, y_L^i)^T$ le décrivant (Zhou *et al.*, 2009). Dans la suite de cet article, y_l^i désigne la modalité ou la valeur de l'attribut $l \in \{1, \cdots, L\}$ observée pour le nœud v_i, où L est le nombre d'attributs et $w_i = \|Y^i\|$ désignera la norme de ce vecteur représentant le poids associé au nœud. Ce poids peut décrire par exemple sa notoriété. Afin de comparer les positions occupées par les sommets dans le graphe, plusieurs mesures de centralité ont été définies : la centralité de degré, de proximité, d'intermédiarité, ou encore de vecteurs propres dont une variante est le PageRank. Elles peuvent être employées pour évaluer l'importance d'un nœud en fonction de son nombre de liens sortants dans un graphe ou parfois même son influence dans le réseau, même si cette dernière notion est trop complexe pour pouvoir être correctement appréhendée par les simples mesures de centralité. Dans la suite, nous proposons d'étendre ces mesures, initialement définies sur des graphes sans attribut, de façon à tenir compte non seulement des relations qui existent entre les sommets du graphe mais aussi d'informations les caractérisant.

2.1. La centralité de Degré (degree centrality)

La centralité de degré mesure l'importance relative d'un nœud dans un graphe en fonction du nombre de ses liens. Elle peut être définie en considérant le nombre de liens sortants (*out-degree*) ou entrants (*in-degree*) dans un graphe orienté, ou simplement le nombre de liens $deg(v_i)$, normalisé par le nombre maximal de liens, dans un graphe non orienté (Freeman, 1979) :

$$Degree(v_i) = \frac{deg(v_i)}{|V| - 1} \tag{1}$$

Dans le cas d'un graphe orienté et valué, elle est calculée en utilisant les poids des liens comme décrit dans (Barrat *et al.*, 2004) :

$$WEDegree(v_i) = \sum_{v_j \in out(v_i)} w_{ij} \tag{2}$$

où w_{ij} est le poids du lien entre le sommet v_i et le sommet v_j, $out(v_i)$ représente l'ensemble des successeurs v_j de v_i. Dans (Opsahl *et al.*, 2010) une autre mesure de centralité de degré est proposée :

$$WEOpsahlDegree(v_i) = (deg_{out}(v_i))^{(1-\alpha)} \cdot \left(\sum_{v_j \in out(v_i)} w_{ij} \right)^{\alpha} \tag{3}$$

Dans nos expérimentations, nous avons choisi de fixer le paramètre α à 0.5 pour équilibrer l'apport des liens par rapport aux poids de ces derniers.

Dans le cas d'un graphe avec des attributs décrivant les nœuds, nous introduisons des variantes de ces mesures notées respectivement *WNDegree* et *WNEDegree* ou *WNEOpsahlDegree* selon que le graphe est valué ou non :

$$WNDegree(v_i) = w_i \cdot Degree(v_i) \tag{4}$$

$$WNEDegree(v_i) = w_i \cdot WEDegree(v_i) \tag{5}$$

$$WNEOpsahlDegree(v_i) = w_i \cdot WEOpsahlDegree(v_i) \tag{6}$$

2.2. La centralité de proximité (closeness centrality)

D'après la centralité de proximité, un nœud est considéré comme important s'il peut rapidement atteindre les autres sommets ((Wasserman, Faust, 1994 ; Hakimi, 1964)). La mesure usuelle est égale à l'inverse de la somme des distances du nœud v_i à tous les autres nœuds du graphe. La distance entre deux nœuds correspond à la longueur du plus court chemin (distance géodésique) les reliant :

$$CCentr(v_i) = \frac{1}{\sum_{\substack{v_j \in V \\ j \neq i}} |ShortPath(v_i, v_j)|} \tag{7}$$

où $ShortPath(v_i, v_j)$ représente un des chemins les plus courts entre deux sommets v_i et v_j et $|ShortPath(v_i, v_j)|$ la longueur, en nombre de liens, du chemin. Par convention, $|ShortPath(v_i, v_j)|$ est égal à $|V|$ ou à ∞ s'il n'existe pas un tel chemin.

Pour les graphes valués, (Opsahl *et al.*, 2010) ont défini une variante tenant compte du poids des liens qui fait appel à l'algorithme de Dijskstra. Nous proposons de mesurer la centralité de proximité d'un nœud par la somme des poids moyens des chemins les plus courts le reliant aux autres nœuds du graphe :

$$CWECentr(v_i) = \sum_{\substack{v_j \in V \\ j \neq i}} \frac{\sum\limits_{e \in ShortPath(v_i, v_j)} w(e)}{|ShortPath(v_i, v_j)|} \tag{8}$$

où $w(e)$ est le poids, mesurant l'intensité de la relation entre deux nœuds, d'un lien $e \in E$ appartenant à un chemin. Dans le cas ou aucun chemin n'existe entre les deux nœuds v_i et v_j, alors le terme au numérateur est nul. L'idée ici est que l'influence se fait essentiellement via les plus courts chemins. Cette influence est d'autant plus importante que le poids moyen du chemin est plus grand. S'il existe plusieurs plus courts chemins entre deux nœuds, c'est celui avec le plus grand poids moyen qui est retenu.

Si les sommets du graphe sont caractérisés par des attributs alors nous proposons de calculer la centralité de proximité, selon que le graphe soit non valué ou valué, respectivement par :

$$CWNCentr(v_i) = w_i \cdot CCentr(v_i) \tag{9}$$

$$CWNECentr(v_i) = w_i \cdot CWECentr(v_i) \tag{10}$$

2.3. *La centralité d'intermédiarité (betweenness centrality)*

D'après la centralité d'intermédiarité, un nœud est considéré comme important s'il est placé sur un nombre important de chemins (géodésiques) reliant les autres nœuds du graphe (Wasserman, Faust, 1994 ; Freeman, 1979). Cet indicateur est défini par :

$$BCentr(v_i) = \sum_{\substack{(v_k, v_j) \in V \times V \\ i \neq k \neq j}} |g_{kj}(v_i)| / |g_{kj}| \tag{11}$$

où $g_{kj}(v_i)$ (resp. $|g_{kj}(v_i)|$) est l'ensemble des plus cours chemins entre les sommets v_k et v_j qui passent par v_i (resp. son cardinal) et g_{kj} (resp. $|g_{kj}|$) l'ensemble des chemins les plus courts de v_k à v_j (resp. son cardinal).

Ainsi, plus il y a de chemins qui passent par un nœud, plus ce dernier est important. Ces nœuds intermédiaires jouent un rôle important dans les processus de diffusion ou de propagation puisqu'ils ont un certain contrôle sur les interactions indirectes qui peuvent exister entre les nœuds non adjacents du réseau.

Dans le cas d'un graphe valué, cette mesure peut facilement être adaptée de façon à tenir compte des poids associés aux liens :

$$BWECentr(v_i) = \sum_{\substack{(v_k, v_j) \in V \times V \\ i \neq j \neq k}} \frac{\sum\limits_{S \in g_{kj}(v_i)} \sum\limits_{e \in S} w(e)}{\sum\limits_{S \in g_{kj}} \sum\limits_{e \in S} w(e)} \tag{12}$$

où $w(e)$ est le poids du lien $e \in E$ appartenant au plus court chemin $S \in g_{kj}$ entre les deux sommets v_k et v_j.

Si les nœuds du graphe sont caractérisés par des attributs alors, l'indicateur est pondéré par le poids du nœud w_i, donnant différents définitions selon que le graph soit valué ou non :

$$BWNCentr(v_i) = w_i \cdot BCentr(v_i) \tag{13}$$

$$BWNECentr(v_i) = w_i \cdot BWECentr(v_i) \tag{14}$$

2.4. La centralité des vecteurs propres et d'autorité (Eigenvector Centrality et PageRank)

La centralité des vecteurs propres est basée sur l'idée que le score d'un nœud est plus élevé s'il est connecté à des nœuds ayant eux mêmes un score élevé que s'il est connecté à des nœuds ayant des scores faibles (Bonacich, Paulette, 2001 ; Ghosh, Lerman, 2010). En général, elle est calculée de façon itérative comme suit :

$$EVCentr(v_i) = \frac{1}{\lambda_1} \cdot \sum_{v_j \in V} a_{ij} \cdot EVCentr(v_j) \tag{15}$$

où $A = \{a_{ij}\}$ est la matrice d'adjacence du graphe et λ_1 correspond à la plus grande valeur propre solution de l'équation $AX = \lambda X$.

Le $PageRank$ est une variante de la centralité des vecteurs propres (Brin, Page, 1998 ; Ghosh, Lerman, 2010). Introduite initialement pour mesurer la popularité d'une page Web. Elle est définie usuellement par :

$$PRankCentr(v_i) = (1 - \beta) \cdot W_0 + \beta \cdot \sum_{v_j \in V} a_{ji} \cdot \frac{PRankCentr(v_j)}{deg(v_j)} \tag{16}$$

où W_0 est habituellement choisi uniformément ($W_0 = 1/|V|$) pour tous les nœuds du graphe.

Dans le cas d'un graphe avec attributs, nous proposons d'adopter une forme personnalisée $PRankWNECentr$ du PageRank (Sergey *et al.*, 1998 ; Jeh, Widom, 2003) avec une distribution non-uniforme des poids initiaux en remplaçant W_0 par les poids w_i définis à partir des attributs pour chaque sommet v_i.

3. Expérimentations

Des expérimentions ont été menées sur des graphes attribués réels pour évaluer la pertinence des mesures proposées.

3.1. Dataset

Pour l'expérimentation, nous avons exploité le jeu de données utilisé par (Tang *et al.*, 2009). Il permet de construire des graphes réels de terrain avec des liens valués et des attributs décrivant les nœuds. Le graphe est issu des publications dans différents domaines académiques [1] et représente les relations de co-publications entre les auteurs (graphe de co-auteurs). Il a été construit à partir du système de recherche académique Arnetminer [2]. Il comprend au total 640134 auteurs et 1554643 relations de co-publications dans différents sujets.

Plus formellement, pour chaque sujet de recherche, on dispose d'un graphe $G = (V, E)$ de co-publications dont l'ensemble V des nœuds correspond aux auteurs et où chaque auteur v_i est caractérisé par un attribut w_i correspondant à son nombre de publications, recensé dans Arnetminer, tous sujets confondus (indépendamment du graphe considéré). L'ensemble E des liens entre les nœuds représente des liens de co-publications qui sont valués par le nombre total d'articles w_{ij} co-écrits par les deux auteurs v_i et v_j.

Pour les expérimentations nous avons retenu trois graphes. Le premier graphe est consacré au sujet *Data-mining* et comprend 679 nœuds et 1687 liens non orientés. Le deuxième graphe, portant sur le thème *Information Retrieval*, comprend 657 nœuds et 1907 liens non orientés. Le troisième graphe, ayant *Bayesian Network* comme sujet, dispose de 554 nœuds et de 1238 liens de co-publications.

3.2. Évaluation

Pour évaluer les mesures de centralité que nous avons proposées (cf. section 2), nous considérons deux autres indicateurs comme référence de l'importance réelle d'un auteur : le *H-Index* et le nombre de *Citations*. Ces indicateurs ont été récupérés à partir du site Arnetminer [3]. Le *H-index* (ou indice de HIRSCH) (Hirsch, 2005) est un indice permettant d'évaluer le rang d'un chercheur : un chercheur a un indice h si h de ses articles ont chacun au moins h citations, et les autres articles ont au plus h citations.

Ainsi, après avoir calculé la centralité des acteurs dans le graphe à l'aide d'une des mesures que nous avons définie, nous pouvons comparer le résultat par rapport au *H-Index* ou au nombre de *Citations*. Nous avons choisi de prendre pour chaque indicateur, les tops 20 meilleurs auteurs classés par leur rang par rapport au *H-Index* d'une part et par rapport au nombre de *Citations* d'autre part, l'objectif étant ici d'évaluer la pertinence des mesures de centralité pour identifier quelques acteurs très importants. Afin de déterminer la performance d'une mesure de centralité, on calcule l'indice de Jaccard entre l'ensemble des auteurs retournés par cette mesure et l'ensemble des auteurs de référence (*i.e.* suivant le *H-Index* ou le nombre de *Citations*).

Par ailleurs, chaque mesure de centralité est aussi évaluée en fonction de sa *Précision* et de son *Rappel*. Il faut noter que le nombre total d'auteurs retournés par une

1. http://arnetminer.org/lab-datasets/soinf/.
2. http://www.arnetminer.org.
3. http://arnetminer.org/person-ranklist/hindex/89.

mesure étant égal au nombre total d'auteurs influents par rapport au *H-index* ou par rapport au nombre de *Citations*, les taux de précision et de rappel sont donc identiques.

3.3. Résultats et analyse

Les tableaux 1, 2 et 3 présentent respectivement les résultats (indice de Jaccard, Précision/Rappel) des expérimentations pour les cinq familles de mesures (18 mesures au total) de centralité : Degree, Closeness, Betweenness, PageRank et Eigenvector, pour les trois thèmes retenus pour l'expérimentation : Data-Mining, Information Retrieval (IR) et Bayesian Networks. La centralité d'un auteur est calculée en fonction des mesures classiques et des variantes que nous avons proposées, notées (*) dans les tableaux, et qui prennent en compte les valuations des liens ou les attributs des nœuds. Les scores de la plupart des mesures de centralité sont meilleurs quand elles sont comparées au *H-Index* que par rapport au nombre de *Citations* notamment pour le taux de Précision/Rappel dans les trois graphes.

Tableau 1. Résultats des mesures de centralité sur le graphe ayant Data-mining comme sujet.

	Jaccard (H-Index)	Jaccard (C)	P/R (H-Index)	P/R (C)
Degree	0.3333	0.1765	0.5	0.3
WNDegree(*)	0.3333	0.1765	0.5	0.3
WEDegree	0.3333	**0.2121**	0.5	**0.35**
WEOpsahlDegree	0.3333	**0.2121**	0.5	**0.35**
WNEDegree(*)	0.3793	**0.2121**	0.55	**0.35**
WNEOpsahlDegree(*)	0.3793	**0.2121**	0.55	**0.35**
CCentr	0.3333	0.1765	0.5	0.3
CWNCentr(*)	0.3333	**0.2121**	0.5	**0.35**
CWECentr(*)	0.0	0.0256	0.0	0.05
CWNECentr(*)	0.25	0.1765	0.4	0.3
BCentr	0.3333	0.1429	0.5	0.25
BWNCentr(*)	0.3793	0.1765	0.55	0.3
BWECentr(*)	0.2903	0.1765	0.45	0.3
BWNECentr(*)	0.3793	0.1765	0.55	0.3
PRankCentr	**0.4286**	0.1765	**0.6**	0.3
PRankWNECentr(*)	0.2903	0.1765	0.45	0.3
EVCentr	0.2121	0.1111	0.35	0.2
EVWNECentr(*)	**0.4286**	**0.2121**	**0.6**	**0.35**

On peut observer que pour le graphe Data-Mining ce sont les mesures d'autorité à savoir le PageRank classique **PRankCentr** et la nouvelle variante de la centralité des vecteurs propres **EVWNECentr**) qui sont les mieux classées, en terme d'indice de Jaccard et de Précison/Rappel, par rapport au *H-Index*. Par rapport au nombre de *Citation* ce sont plutôt les indicateurs de degré classique du graphe valué **WEDegree** et **WEOpsahlDegree**) et les nouvelles variantes **WNEDegree** et **WNEOpsahlDegree**) qui donnent des scores élevés. Ceci souligne l'apport des poids des liens (nombre

Tableau 2. Résultats des mesures de centralité sur le graphe ayant Information Retrieval (IR) comme sujet.

	Jaccard (H-Index)	Jaccard (C)	P/R (H-Index)	P/R (C)
Degree	0.25	0.2121	0.4	0.35
WNDegree(*)	**0.4286**	**0.3333**	**0.6**	**0.5**
WEDegree	0.25	0.2121	0.4	0.35
WEOpsahlDegree	0.2903	0.25	0.45	0.4
WNEDegree(*)	0.2903	0.2121	0.45	0.35
WNEOpsahlDegree(*)	0.3793	0.2903	0.55	0.45
CCentr	0.25	0.2121	0.4	0.35
CWNCentr(*)	0.3793	0.2903	0.55	0.45
CWECentr(*)	0.0	0.0	0.0	0.0
CWNECentr(*)	0.2121	0.1765	0.35	0.3
BCentr	0.1765	0.1429	0.3	0.25
BWNCentr(*)	0.3793	0.2903	0.55	0.45
BWECentr(*)	0.25	0.2121	0.4	0.35
BWNECentr(*)	0.3793	0.2903	0.55	0.45
PRankCentr	0.25	0.2121	0.4	0.35
PRankWNECentr(*)	0.3793	0.25	0.55	0.4
EVCentr	0.25	0.2121	0.4	0.35
EVWNECentr(*)	0.2903	0.25	0.45	0.4

Tableau 3. Résultats des mesures de centralité sur le graphe ayant Bayesian Networks comme sujet.

	Jaccard (H-Index)	Jaccard (C)	P/R (H-Index)	P/R (C)
Degree	0.2121	0.1429	0.35	0.25
WNDegree(*)	0.3333	0.25	0.5	0.4
WEDegree	0.25	0.1765	0.4	0.3
WEOpsahlDegree	0.2903	0.2121	0.45	0.35
WNEDegree(*)	**0.3793**	0.2903	**0.55**	0.45
WNEOpsahlDegree(*)	**0.3793**	0.2903	**0.55**	0.45
CCentr	0.2121	0.1765	0.35	0.3
CWNCentr(*)	0.3333	0.25	0.5	0.4
CWECentr(*)	0.0526	0.0256	0.1	0.05
CWNECentr(*)	0.2903	0.2121	0.45	0.35
BCentr	0.2903	0.25	0.45	0.4
BWNCentr(*)	**0.3793**	**0.3333**	**0.55**	**0.5**
BWECentr(*)	0.25	0.25	0.4	0.4
BWNECentr(*)	0.3333	0.25	0.5	0.4
PRankCentr	0.3333	0.25	0.5	0.4
PRankWNECentr(*)	0.25	0.2121	0.4	0.35
EVCentr	0.1111	0.0811	0.2	0.15
EVWNECentr(*)	0.3333	0.25	0.5	0.4

de co-publications) dans la pertinence des résultats et la prépondérance des mesures d'autorité dans le graphe ayant Data-Mining comme sujet.

Pour le graphe IR, c'est la variante **WNDegree** des mesures de degré qui obtient les meilleurs résultats par rapport aux deux références: le *H-Index* et le nombre de *Citation*. On peut aussi distinguer l'apport des valuations, spécialement les attributs des nœuds, dans l'amélioration des scores obtenus par les mesures classiques (WNEOpsahlDegree, CWNCentr, BWNCentr, BWNECentr, PRankWNECentr) même pour la mesure de proximité qui donne en général de très faibles résultats.

Pour le graphe Bayesian Network, c'est les trois variantes **WNEDegree**, **WNEOpsahlDegree** et **BWNCentr** qui obtiennent les meilleurs scores pour les deux références (*H-Index* et nombre de *Citation*). Les valuations apportent des améliorations plus ou moins importantes dans le score des mesures classiques sauf pour le PageRank.

En conclusion, les variantes des mesures de centralités introduites sont bien classées par rapport aux mesures classiques ; ce qui confirme l'apport des attributs dans le calcul de la centralité d'un auteur. Il faut noter que la centralité de proximité, ainsi que ses variantes, donne de piètres résultats car elle n'est pas bien adaptée pour les graphes étudiés. D'une manière générale, même si les résultats varient d'un graphe à l'autre et dépendent forcément du sujet traité, on peut constater une tendance caractérisant les variantes des centralités de degré, intégrant à la fois les attributs et les poids des liens, qui sont toujours classées parmi les meilleures mesures sur les trois graphes. En effet, les nouvelles variantes de la centralité de degré, et spécifiquement les deux mesures **WNEDegree** et **WNEOpsahlDegree**, qui prennent en charge les valuations des liens et les attributs des nœuds, donnent en général de très bons résultats sur les trois graphes et ne sont pas gourmandes en matière de temps de calcul. En effet, elles ont l'avantage d'êtres calculées localement, *i.e.* ne nécessitent pas d'avoir connaissance de la structure globale du graphe, elles sont donc plus rapides à calculer par rapport aux autres mesures.

Pour les centralités classiques du degré, de l'intermédiarité et des vecteurs propres l'apport des valuations des liens a permis d'améliorer la pertinence des résultats, tout particulièrement si on se réfère au *H-index*. Cette amélioration est encore plus forte si on prend en compte, en plus des valuations des liens, les attributs décrivant les nœuds. Ceci n'est pas toujours le cas pour les autres familles de mesures basées sur la proximité ou le PageRank. Dans ces deux familles de mesures, les résultats des variantes basées sur les valuations des liens ne sont pas homogènes et sont parfois encore plus faibles que les mesures classiques.

La complexité en temps de calcul des nouvelles variantes des mesures de centralité reste dans le même ordre de grandeur que celle des mesures classiques de centralité. Pour la centralité du degré, la somme des poids des liens adjacents à un nœud est calculée sur des valeurs numériques plutôt que sur des valeurs binaires. Pour la centralité des vecteurs propres et le PageRank, le même processus itératif de calcul est utilisé pour atteindre une distribution stationnaire. Pour les deux centralités de proximité et d'intermédiarité, il est nécessaire de calculer les plus courts chemins entre toutes les paires de nœuds dans le graphe. La tâche est encore plus difficile dans le cas des graphes valués, où il faut en plus calculer les poids moyens cumulés des plus courts chemins, mais ceci reste une opération avec une complexité linéaire comparable à celle des mesures usuelles.

4. Conclusion

Nous avons présenté dans ce travail un ensemble de mesures de centralité permettant d'évaluer la position occupée par les acteurs dans un réseau social. Ces variantes de mesures classiques prennent en compte non seulement la structure du réseau mais elles intègrent aussi des valuations qui caractérisent à la fois les liens qui existent entre les acteurs et les attributs spécifiques à ces derniers. Une série d'expérimentations a été réalisée sur trois graphes de co-publications scientifiques.

Ces expérimentations ont permis d'évaluer la performance de ces mesures pour identifier les utilisateurs considérés comme les plus influents dans leur domaine de recherche. Les résultats montrent que notre approche est encore plus pertinente que les indicateurs classique de centralité notamment pour les variantes pondérées de la centralité du degré.

Bibliographie

Barrat A., Barthelemy M., Pastor-Satorras R., Vespignani A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, nº 11, p. 3747–3752.

Bonacich P., Paulette L. (2001, jul,). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, vol. 23, nº 3, p. 191–201.

Brin S., Page L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the seventh international conference on world wide web 7*, p. 107–117.

Everett M. G., Borgatti S. P. (2012). Categorical attribute based centrality: E-I and G-F centrality. *Social Networks*, vol. 34, p. 562 - 569.

Freeman L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, vol. 1, nº 3, p. 215-239.

Ghosh R., Lerman K. (2010, July). Predicting influential users in online social networks. In *Proceedings of kdd workshop on social network analysis (sna-kdd).*

Gould R. V., Fernandez R. M. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology*, vol. 19, nº 1, p. 89.

Hakimi S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, vol. 12, nº 3, p. 450–459.

Hirsch J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, nº 46,, p. 16569-16572.

Jeh G., Widom J. (2003). Scaling personalized web search. In *Www '03: Proceedings of the 12th international conference on world wide web*, p. 271–279. New York, NY, USA, ACM Press.

Krackhardt D., Stern R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*, vol. 51, n° 2, p. 123 - 140.

Opsahl T., Agneessens F., , Skvoretz J. (2010, jul,). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, vol. 32, n° 3, p. 245–251.

Sergey B., Motwani R., , Page L., Winograd T. (1998). What can you do with a web in your pocket? *IEEE Data Engineering Bulletin*, vol. 21, p. 37-47.

Tang J., Sun J., , Wang C., Yang Z. (2009). Social influence analysis in large-scale networks. In *Proceedings of the 15th acm sigkdd international conference on knowledge discovery and data mining*, p. 807–816. New York, NY, USA, ACM.

Wasserman S., Faust K. (1994). *Social network analysis, methods and application* (4th edition: 1998, 1999 éd.; C. U. Press, Ed.). Cambridge University Press, New York, USA.

Zhou Y., Cheng H., Yu J. X. (2009). Graph clustering based on structural/attribute similarities. *Proceedings of the VLDB Endowment*, vol. 2, p. 718–729.