Les mesures locales d'un réseau

César Ducruet

Les mesures locales d'un réseau

César Ducruet, CNRS, UMR Géographie-cités
ducruet<at>parisgeo.cnrs.fr

Version 1 - Décembre 2010

groupe f.m.r.

Introduction

Une mesure locale s'attache à décrire la situation d'un élément au sein du réseau par rapport aux autres éléments, qu'il s'agisse de sommets ou bien de liens. Les mesures locales peuvent être interprétées, selon l'approche et la thématique en jeu, comme des mesures de centralité et d'accessibilité ; elles expriment avant tout un potentiel. L'analyse statistique de ces mesures, prises dans leur individualité ou bien deux à deux, peut permettre de vérifier certaines lois d'organisation des réseaux en construisant des mesures globales ([8]).

Les mesures locales sont bien plus nombreuses et diverses que les mesures globales. Nous proposons de les classer en deux grands types : les mesures locales de voisinage et les mesures locales d'ensemble. Les premières décrivent la situation d'un élément par rapport à ses voisins immédiats (directement connectés ou adjacents), tandis que les secondes rendent compte de la situation d'un élément par rapport à tous les autres éléments de même nature présents dans le réseau. Quand cela est possible, des sous-catégories permettent de mieux différencier les mesures de nature différente, c'est-à-dire qui sont habituellement appliquées dans des champs scientifiques séparés (ex : centralité dans un réseau social, accessibilité dans un réseau de transport). Or il n'y a aucune raison valable de conserver ces coupures thématiques et disciplinaires puisque ces mesures s'appliquent à tout réseau.

1 Les mesures locales de voisinage

Certaines mesures peuvent connaître de nombreuses variantes en fonction du critère retenu. Par exemple, la définition du degré ou degré d'incidence (*degree centrality*) peut s'appliquer à une profondeur (*depth*) d'une valeur de 1 (sommets voisins immédiats, adjacents, directement connectés), de 2 (voisins proches, reliés par l'intermédiaire des voisins immédiats), de 3, etc. En d'autres termes, il s'agit de prendre ou non en compte le degré des voisins. Le degré peut également être pondéré en fonction du poids des liens (*weighted degree*), ce qui équivaut dans le cas d'un réseau de transport à calculer la somme, par exemple, du trafic des liens d'un sommet (voyageurs, tonnes) même si cette somme peut varier en fonction de la profondeur retenue (Barrat *et al.*, 2004[1]). Dans un graphe orienté, on distingue le degré « entrant »(*in-degree*) du degré « sortant »(*out-degree*), les deux pouvant être identiques ou non.

Calculer la part du flux majeur dans le total du trafic revient à déterminer un niveau de vulnérabilité d'un sommet (*hub dependence*, voir Ducruet, 2008[7]). En effet, cette mesure illustre à quel point le lien principal compte dans l'activité générale du sommet, et quelle part de son activité serait perdue en cas de disparition de ce lien (ou du voisin relié *via* le flux majeur). Dans le cas où un sommet serait relié par plusieurs liens de poids identique ou comparable, la prise en compte du seul flux majeur peut s'avérer trop simpliste (Puebla, 1987[17]). On peut dans ce cas appliquer un seuil : liens ayant un poids d'au moins 10% dans le degré pondéré ; liens dont le poids total constitue au plus 50% du degré pondéré, etc. afin de prendre en compte les flux dominants (*dominant flows*) dans l'analyse. Enfin, le degré peut devenir une mesure locale d'ensemble si l'on rapporte le degré au nombre de sommets du graphe (pourcentage de sommets reliés dans le total) ou bien si l'on calcule le poids du degré pondéré dans le poids total du graphe $Q(G)$.

La « centralité combinée » est d'après Opsahl *et al.* (2010[16]) plus efficace que les mesures précédentes puisqu'elle combine le degré d'incidence et le degré pondéré : topologie et poids. En effet, le degré considère tous les liens de la même façon (0, absence de lien et 1, présence de lien), tandis que le degré pondéré ne fait pas de distinction entre un seul lien de valeur 2 et deux liens de valeur 1. Tout est dans le paramètre alpha proposé par les auteurs, dont le réglage (entre 0 et 1) dépend de la thématique abordée : un coefficient faible donnera plus d'importance au degré, tandis qu'un coefficient fort favorisera le rôle du poids des liens.

Le degré moyen des voisins les plus proches est utile puisqu'il permet de révéler la structure d'ensemble du réseau. En anglais *assortative network* signifie que les sommets voisins ont un degré comparable, *disassortative network* signifie le contraire (Newman, 2002 et 2003[13][14]). Deux mesures globales permettent d'en rendre compte (cf. [8]). L'*assortativity coefficient* est donc une mesure globale qui correspond à la corrélation de Pearson au sein de chaque couple de sommets entre leurs degrés respectifs. Le *neighbor connectivity* est la corrélation entre le degré d'un sommet et la moyenne des degrés des sommets voisins. On peut aussi raisonner en termes de probabilités en fonction de la distribution statistique des valeurs. L'eigenvector centrality se différencie du degré par le fait qu'elle tient compte du degré des

Tableau 1 – Les mesures locales de voisinage

Degré, degré d'incidence (*degree centrality*) : nombre de sommets adjacents

$$k_i = C_D(i) = \sum_j^N x_{ij}$$

Degré pondéré (*weighted degree*) : somme des poids des liens adjacents

$$s_i = C_D^w(i) = \sum_i^N w_{ij}$$

Centralité combinée (*combined centrality*) : degré pondéré sur degré d'incidence à l'exposant α

$$C_D^{w\alpha}(i) = k_i \times \left(\frac{s_i}{k_i}\right)^\alpha = k_i^{(1-\alpha)} \times s_i^\alpha$$

Degré moyen des voisins les plus proches (*average nearest neighbours degree*)

$$k_{nn,i} = \frac{1}{k_i}\sum_j a_{ij}k_j$$

Centralité des vecteurs propres (*eigenvector centrality*)

$$x_i = \frac{1}{\lambda}\sum_{j=1}^n A_{ij}x_j$$

Vulnérabilité (*hub dependence*) : part du lien le plus fort dans le degré pondéré (%)

Transitivité (*transitivity, clustering coefficient*) : nombre de triades fermées sur nombre de triades possibles

$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)}$$

Indice de force (*strength index*) : probabilité que les liens adjacents appartiennent à des cycles 3 et 4

$$w_s(u) = \frac{\sum_{e \in adj(u)} w_s(e)}{deg(u)}$$

Indice de Strahler (*Strahler index, Horton-Strahler number*) : nombre d'embranchements en aval du sommet sommets adjacents. Cette centralité tient donc compte à la fois du nombre et de la qualité des liens ([3]).

Certaines mesures locales de voisinage se prêtent à un type particulier de réseau. L'indice de Strahler, créé en hydrologie, s'applique uniquement aux graphes en arbre (*tree graph*) et mesure le niveau de ramification des sommets. C'est le nombre d'embranchements (confluences) en aval de chaque sommet. Inversement, la transitivité ne peut s'appliquer à un graphe en arbre car elle s'intéresse avant tout à l'existence de cycles (cf.[2]) dans le graphe. Plus précisément, la transitivité correspond au nombre de triades (*triads*) fermées sur le nombre de triades possibles (trois sommets voisins interconnectés). C'est donc la probabilité pour un sommet donné que ses deux voisins immédiats soient également connectés entre eux. En théorie, l'existence d'un réseau petit-monde (*small-world network*) repose sur une forte transitivité moyenne, ainsi que sur un diamètre faible et une distance moyenne faible des plus courts chemins (cf. [8]). L'indice de force (*strength index*) est identique à cela près qu'il ne considère pas les triades mais les cycles de niveau 3 et 4 [1](Bourqui *et al.*, 2009[4]). L'indice de force est en fait la densité de la cohésion des liens autour d'un sommet donné. Une force élevée correspondra à des nœuds situés dans un environnement dense, comme un graphe ou sous-graphe complets, une force faible correspondra à des *hubs*, dont les sommets adjacents ne sont que faiblement reliés entre eux.

2 Les mesures locales d'ensemble

La plupart des mesures d'ensemble sont des mesures d'accessibilité prises sous différents angles. Elles prennent en compte soit la distance géodésique, soit la distance euclidienne (ex : kilomètres). La majorité des mesures d'ensemble a été appliquée aux réseaux de transport (le plus souvent sur des graphes planaires), visant à rendre compte de l'inégale situation des sommets (villes, gares, etc.) dans le réseau routier ou ferroviaire. Leur point commun est de décrire la situation d'un sommet par rapport à tous les autres, par rapport aux chemins possibles, à leur longueur (totale, la plus courte, moyenne), et à leurs caractéristiques (poids, qualité, etc.).

L'indice de Shimbel se calcule à partir d'une matrice des distances : celles-ci sont additionnées en ligne ou en colonne. Le sommet dont la somme est la moindre est donc le plus accessible depuis et vers tous les autres sommets. La centralité de proximité (*closeness centrality*) est simplement l'inverse de l'indice de Shimbel, avec l'avantage d'être normalisée de 0 à 1, 1 étant la plus forte accessibilité. L'accessibilité géographique est également un dérivé de l'indice de Shimbel, puisqu'il suffit de diviser ce dernier par le nombre de sommets présents sur les plus courts chemins. De la même façon, l'accessibilité potentielle divise un attribut quelconque d'un sommet donné (ex :

1. Contenant 3 ou 4 sommets.

population, surface commerciale, richesse) par l'indice de Shimbel (Rodrigue *et al., 2009*[18]). L'accessibilité potentielle se basant sur des valeurs associées aux sommets, les résultats varient en ligne ou en colonne : on parle alors d'émissivité (somme en ligne ; capacité à quitter un lieu) ou d'attractivité (somme en colonne ; capacité à atteindre un lieu).

Un bon exemple est l'analyse par Chapelon (2006[5]) de l'accessibilité des ports à la richesse et à la population européennes. La recherche de plus courts chemins dans le graphe valué (algorithme de Floyd) ne se borne pas à la simple topologie du réseau routier ; elle prend également en compte la législation en vigueur, la qualité des liens (hiérarchie, longueur). De nombreux travaux appliquant ces mesures aux réseaux de transport (ex : indice de Shimbel) ont pu souligner les dynamiques territoriales et réticulaires sous-jacentes, comme dans le cas des stratégies de hub en transport aérien ou encore l'influence de l'intégration européenne sur l'accessibilité des régions. Wang *et al.* (2010[20]) ont notamment mis en évidence la forte corrélation entre différents types de centralité (degré, d'intermédiarité, de proximité) et les caractéristiques socio-économiques des villes dans le cas du réseau aérien chinois, en termes de produit régional brut et de population urbaine. Gleyze (2007[10]) propose de distinguer l'effet réseau de l'effet spatial dans le niveau de centralité des sommets, à partir de l'exemple du réseau routier d'Indianapolis.

La centralité d'intermédiarité (*betweenness*), dans sa version topologique ou pondérée, a retenu l'attention de nombreux travaux récents. Les sommets ayant un niveau élevé de centralité d'intermédiarité sont ceux qui permettent de limiter la distance à parcourir dans le réseau. Leur suppression conduirait donc à ralentir le flux puisque ce dernier devrait passer par des chemins plus longs. La centralité d'intermédiarité a notamment été utilisée en géographie par Comin (2009[6]) pour montrer la polarisation du réseau des collaborations scientifiques européennes par les grandes villes, ainsi que par Rozenblat (2010[19]) pour mettre en évidence les logiques de localisation des multinationales dans les villes européennes.Cette centralité se calcule également sur les liens : Girvan et Newman (2002[9]) s'en servent comme base de leur algorithme (divisive hierarchical clustering algorithm) pour trouver des sous-graphes dans le réseau, puisque la suppression des liens les plus centraux met en valeur des cliques. Newman (2005[15]) reste critique à l'égard de la centralité d'intermédiarité puisque d'après lui, les flux dans un réseau ne suivent pas forcément le plus court chemin ou le chemin le plus efficace. Il propose donc de prendre en compte les chemins aléatoires (random-walk betweenness) dans le calcul de cette centralité, et reconnaît l'existence de multiples mesures existantes allant dans ce sens [2] .

2. Newman (2005) mentionne notamment la *power centrality* de Bonacich, la *random-walk centrality* de Noh et Rieger, l'*information centrality* de Stephenson et Zelen. Wasserman et Faust (1994[21]) mentionnent également la *stress centrality* et la *graph centrality*.

Tableau 2 – Les mesures locales d'ensemble

Nombre de Koenig, excentricité (*Koenig number, associated number, eccentricity, dag level*) : nombre de liens servant à connecter le sommet le plus distant $$e(x) = max_{y \in X} d(x, y)$$
Indice de Shimbel (*Shimbel index, Shimbel distance, nodal accessibility, nodality*) : somme des longueurs des plus courts chemins permettant de relier tous les autres sommets $$A_i = \sum_{j=1}^{n} d_{ij}$$
Accessibilité géographique* (*geographic accessibility*) : distance minimale divisée par le nombre de sommets du graphe $$A(G) = \frac{\sum_i^n \left(\sum_j^n d_{ij} \right)}{n}$$
Accessibilité potentielle** (*potential accessibility*) : accessibilité géographique divisée par un attribut du sommet choisi en fonction de la thématique traitée $$A(P) = \frac{\sum_i^n P_i + \sum_j^n P_j}{d_{ij}}$$
Centralité de proximité (*closeness centrality, distance centrality*) : inverse de l'indice de Shimbel $$D_c(v) = \frac{1}{\sum_{t \in G} d_G(v, t)}$$
Centralité d'intermédiarité (*betweenness centrality, shortest-path betweenness*) : nombre de plus courts chemins du graphe passant par chaque sommet $$C_B(i) = \frac{g_{jk}(i)}{g_{jk}}$$
Centralité d'intermédiarité pondérée* (*weighted betweenness centrality, flow betweenness*) : trafic passant par le sommet i entre les sommets j et k par rapport au trafic maximum entre j et k $$C_i^F = \frac{\sum_{j<k \in G} m_{jk}(i)}{\sum_{j<k \in G} m_{jk}}$$
*Dans un graphe valué (ex : distance kilométrique pour les liens) **Dans un graphe valué (ex : distance kilométrique pour les liens ; population ou richesse pour les sommets)

3 Des mesures locales aux mesures globales

L'étude de la distribution statistique des mesures locales est un aspect essentiel de la recherche sur les réseaux, puisque l'on peut en déduire le type de réseau rencontré.

La distribution des degrés (*degree distribution*)

L'exemple le plus communément admis dans les recherches récentes est celui de la distribution des degrés des sommets, qui permet de déterminer l'existence d'une structure invariante d'échelle (*scale-free*, voir séance fmr 2). Le graphique bi logarithmique du degré (abscisses) et de la fréquence des sommets (ordonnées) sert à vérifier l'applicabilité de la loi de puissance. La pente de la courbe (coefficient) est supérieure ou égale à 1 dans le cas des réseaux invariants d'échelle, même si la littérature suggère des valeurs plus élevées (2 à 3) pour valider la structure en question. Cette analyse n'a guère de sens dans le cas des graphes planaires où la dimension hiérarchique est modérée.

Degré et centralité d'intermédiarité

Croiser ces deux centralités dans un graphique permet de repérer la situation exceptionnelle de certains sommets dans le réseau. La centralité désignée comme « anormale »(*anomalous centrality*) par Guimera *et al.* (2005[11]) s'explique par des sommets étant mieux placés au niveau global (forte centralité d'intermédiarité) qu'au niveau local (degré plus faible) : ce sont des ponts, des connecteurs (*bridges*, *connectors*) servant, dans le réseau aérien par exemple, de relais entre régions distantes, comme Anchorage (Alaska) entre l'Asie et l'Amérique du Nord. Guimera *et al.* (2005) vont plus loin en soulignant que ce type de configuration est directement lié à l'existence de sous-groupes ou communautés (voir séance fmr4) fortement interconnectés mais étant reliées par peu de sommets. Le profil opposé existe également, celui d'un sommet ayant beaucoup de liens avec ses voisins (degré élevé) mais n'ayant pas ce rôle de connecteur entre différentes communautés. Or ces écarts n'apparaissent pas dans tous les réseaux car ils dépendent en grande partie de facteurs géographiques et politiques sous-jacents à leur organisation, l'aérien étant plus favorable à la formation de communautés que le maritime (Hu et Zhu, 2009[12]). Le plus souvent, la corrélation entre degré et centralité d'intermédiarité est plutôt forte, les sommets au degré élevé étant également les plus centraux au niveau global ; on parle alors de *hub*[3].

3. Selon les disciplines, le hub signifie un sommet de fort degré, un sommet de fort degré très central, ou un sommet très central de faible degré (pont).

Autres combinaisons possibles

Certaines comparaisons permettent de mettre en évidence d'autres propriétés globales d'un réseau. Par exemple, celle de l'indice de force (*strength*) et du degré dans un graphique bi logarithmique peut suggérer que si la force augmente plus vite que le degré (pente de la droite de régression supérieure à 1), les sommets à fort degré auront tendance à attirer plus de trafic. De la même façon, combiner le degré avec la transitivité confirme que les sommets à fort degré ont, en général, une transitivité faible puisqu'ils connectent par leur position centrale beaucoup de sommets non connectés entre eux (et ne formant donc pas de triades ou de cycles en général). A l'inverse, les sommets de faible degré auront plus de chance de se trouver intégrés dans une triade, donc avec une transitivité maximale.

Conclusion

Les mesures locales sont très diverses et cet inventaire est loin d'être exhaustif. On peut retenir que le degré est bien souvent la seule mesure locale ne soulevant aucune controverse. Simple à la fois par sa définition et son calcul, sa portée est grande et elle suffit dans bien des cas à résumer la situation d'un sommet dans le graphe, même si cela est surtout valable dans le cas des graphes non planaires où la hiérarchie du degré est bien plus prononcée que dans le cas des graphes planaires. De plus, la plupart des mesures locales dans des graphes non planaires sont fortement corrélées au degré. Conclure sur l'utilité de mesures complémentaires n'aurait donc de sens que par rapport aux questions posées. Ainsi en géographie, la centralité d'un nœud ou d'un territoire doit être interrogée par rapport aux autres composantes de l'espace étudié.

Références

[1] A. Barrat, M. Barthélemy, R. Pastor-Satorras et A. Vespignani : The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752, 2004.

[2] L. Beauguitte : Graphes, réseaux, réseaux sociaux : vocabulaire et notation. *Groupe fmr*, 7p., 2010 (http ://halshs.archives-ouvertes.fr/halshs-00541898/en/).

[3] P. Bonacich : Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, 2007.

[4] R. Bourqui, F. Gilbert, P. Simonetto, F. Zaidi, U. Sharan et F. Jourdan : Detecting structural changes and command hierarchies in dynamic social networks. *In Social Network Analysis and Mining,*

2009. ASONAM'09. International Conference on Advances in, pages 83–88, 2009.

[5] L. Chapelon : L'accessibilité, marqueur des inégalités de rayonnement des villes portuaires en Europe. *Cybergeo*, 2006 (http ://cybergeo.revues.org/index2463.html).

[6] M.N. Comin : *Réseaux de villes et réseaux d'innovation en Europe : Structuration du système des villes européennes par les réseaux de recherches sur les technologies convergentes.* Thèse de doctorat, Université de Paris I Sorbonne, 2009.

[7] C. Ducruet : Hub dependence in constrained economies : the case of North Korea. *Maritime Policy & Management*, 35(4):377–394, 2008.

[8] C. Ducruet : Les mesures globales d'un réseau. *Groupe fmr*, 8p., 2010 (http ://halshs.archives-ouvertes.fr/halshs-00541902).

[9] M. Girvan et M.E.J. Newman : Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821, 2002.

[10] J.F. Gleyze : Effets spatiaux et effets réseau dans l'évaluation d'indicateurs sur les nœuds d'un réseau d'infrastructure. *Cybergeo : European Journal of Geography*, 2007 (http ://cybergeo.revues.org/index5532.html).

[11] R. Guimera, S. Mossa, A. Turtschi et L.A.N. Amaral : The worldwide air transportation network : Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22):7794–7799, 2005.

[12] Y. Hu et D. Zhu : Empirical analysis of the worldwide maritime transportation network. *Physica A*, 388:2061–2071, 2009.

[13] M.E.J. Newman : Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002.

[14] M.E.J. Newman : Mixing patterns in networks. *Physical Review E*, 67(2):26126, 2003.

[15] M.E.J. Newman : A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005.

[16] T. Opsahl, F. Agneessens et J. Skvoretz : Node centrality in weighted networks : Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, 2010.

[17] J.G. Puebla : Spatial structures of network flows : A graph theoretical approach. *Transportation Research Part B*, 21(6):489–502, 1987.

[18] J.P. Rodrigue, C. Comtois et B. Slack : *The geography of transport systems.* Routledge, Taylor and Francis Books, 2009.

[19] C. Rozenblat : Opening the black box of agglomeration economies for measuring cities' competitiveness through international firm networks. *Urban Studies*, 47(13):2841–2865, 2010.

[20] J. Wang, H. Mo, F. Wang et F. Jin : Exploring the network structure and nodal centrality of China's air transport network : A complex network approach. *Journal of Transport Geography*, 2010.

[21] S. Wasserman et K. Faust : *Social Network Analysis. Methods and Applications*. Structural analysis in the social sciences. Cambridge University Press, 1994.

Table des matières