

Graphes, réseaux, réseaux sociaux : vocabulaire et notation

Laurent Beauguitte - UMR Géographie-cités
 beauguittelaurent<at>parisgeo.cnrs.fr

Version 1 - Septembre 2010

groupe f.m.r.

Toute personne désireuse de s'attaquer à l'analyse de réseaux se heurte à deux problèmes liés au vocabulaire utilisé. Le premier problème concerne les préférences disciplinaires : selon que l'on lit un physicien, un sociologue, un mathématicien ou un géographe, soit les mêmes termes n'ont pas le même sens, soit, ce qui est plus courant, plusieurs termes différents en désignent un même objet. Le deuxième problème est lié à la langue. Certains termes anglais ont été adaptés en français de façon parfois maladroite (*demi degré intérieur* pour *in-degree* par exemple). Pour corser le tout, un graphe est par défaut orienté pour les mathématiciens francophones (voir Bergé, 1973[1]) et non orienté pour les mathématiciens anglophones (West, 2001[6])... Le choix fait ici est de systématiquement indiquer entre parenthèses et en italiques le ou les équivalent(s) anglophones aux termes définis, sauf lorsqu'il est quasi équivalent au terme français¹.

Les notations proposées ici ne sont pas universelles. Elles sont basées essentiellement pour l'aspect mathématique sur Bergé (1973), pour l'aspect réseaux sociaux sur Wasserman et Faust (1994[5]) et sur Boccaletti *et al.*(2006[2])

1 Ce qu'est un graphe

Des points et des lignes

Un graphe $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ se définit mathématiquement comme un ensemble fini de sommets $\mathcal{N} \neq \emptyset$ et un ensemble fini (mais éventuellement vide) de liens \mathcal{L} . Le nombre d'éléments dans \mathcal{N} et \mathcal{L} est respectivement noté N et K . Un graphe se définit par son ordre (*order*) soit le nombre de

1. Degré et degree, graphe et graph etc.

sommets, et par sa taille (*size*) définie par le nombre de liens. Il est courant cependant de parler de grands graphes quand ils comportent de nombreux sommets.

Les synonymes les plus courants de sommets sont points, nœuds, acteurs. Les termes anglais sont *nodes*, *vertex* (pluriel *vertices*), *actors* ou *points*. Pour les liens, on rencontre les termes *edges*, *ties*, *lines* ou encore *relations*.

Dans un graphe non planaire \mathcal{G} de taille N , le nombre de liens K est au minimum de 0 et au maximum $N(N-1)/2$. Un graphe $\mathcal{G}_{N,K}$ est dit complet (*complete*) si $K = \binom{N}{2} = N(N-1)/2$, il est noté K_N .

Un sous graphe (*subgraph*) $\mathcal{G}' = (\mathcal{N}', \mathcal{L}')$ de \mathcal{G} est un graphe tel que $\mathcal{N}' \subseteq \mathcal{N}$ et $\mathcal{L}' \subseteq \mathcal{L}$.

Un sommet qui n'est adjacent à aucun lien est dit isolé (*isolate*). Lorsqu'un graphe est constitué de plusieurs sous-graphes non connectés les uns aux autres, ces sous-graphes sont nommés composants (*component*). Un composant formé d'un seul sommet est dit trivial. Si enlever un sommet s d'un graphe \mathcal{G} augmente le nombre de composants du graphe, s est appelé point d'articulation (*cutpoint* ou, plus rarement, *cut-vertex*). Si enlever un lien l d'un graphe \mathcal{G} augmente le nombre de composants du graphe, l est appelé isthme (*bridge*).

Des distances et des itinéraires

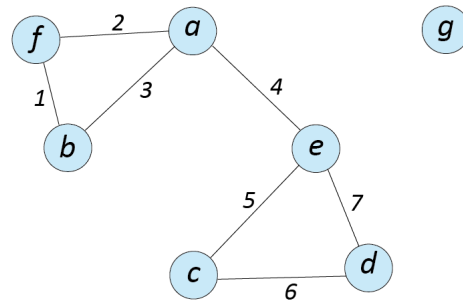
Un sommet a est dit incident à un lien l lorsqu'une des extrémités de l est a . Si l'autre extrémité de l est un sommet b , on dit que a et b sont adjacents ou voisins. La distance la plus courte mesurée en nombre de liens entre deux sommets est nommée distance géodésique (*geodesic*).

On appelle chaîne (*walk*) un parcours sur un graphe non orienté allant d'un sommet à un autre en empruntant des arêtes², lorsque le graphe est orienté, on utilise le terme chemin (*directed walk*). Une chaîne (ou un chemin) est dite élémentaire si chaque sommet y apparaît au plus une fois, elle/il est dit simple si chaque lien apparaît au plus une fois. Lorsque le départ et l'arrivée de cette chaîne (ou de ce chemin) élémentaire est le même sommet, on parle de cycle (graphe non orienté) ou de circuit (graphe orienté) Un graphe est dit connexe (*connected*) lorsqu'il existe un chemin ou une chaîne entre toute paire de sommets.

Certaines propriétés des graphes méritent d'être citées car elles connaissent de nombreuses applications, notamment en informatique. Un graphe eulérien (*Eulerian graph*) désigne un graphe pour lequel il existe une chaîne (*walk*) passant une fois et une seule par tous les liens du graphe. Dans un graphe hamiltonien (*Hamiltonian graph*), il existe un cycle passant par tous les sommets du graphe.

2. Les francophones nomment arête (*edge*) un lien dans un graphe non orienté, arc un lien dans un graphe orienté.

FIGURE 1 – Quelques propriétés d'un graphe



groupe fmr, 2010

Ce graphe non orienté est d'ordre 7 et de taille 7. Le sommet g est isolé, le nombre de composants non triviaux est de 1. Deux cycles sont présents : $\{cde\}$ et $\{abf\}$. Les points a et e sont des points d'articulations, le lien 4 est un isthme.

Un arbre (*tree*) est un graphe où tout sommet est un point d'articulation (*cutpoint*) et tout lien est un isthme (*bridge*). Il ne contient donc aucun cycle³.

Enfin, en mathématiques, la représentation graphique qui peut en être faite n'a aucun intérêt - ce qui n'est absolument pas vrai pour les sciences sociales où tout graphe produit une certaine image de la réalité.

La figure 1 tente de résumer la plupart des notions abordées dans cette première partie.

2 Les différents types de graphes

La typologie des graphes peut se faire selon plusieurs logiques qui se complètent plus souvent qu'elles ne s'opposent. Une première distinction, sur laquelle le groupe fmr reviendra souvent, concerne les graphes planaires (*planar*) et non planaires. Un graphe planaire peut être projeté sur un plan sans qu'aucune lien n'en croise un autre. Ce type de graphes a longtemps été privilégié en géographie (voir Chorley et Haggett[3] ou Mathis[4]).

Il est possible de distinguer les graphes pour lesquels seule la présence ou l'absence de liens importe, et ceux pour lesquels la direction du lien est prise en compte. On peut différencier 3 types de graphes selon la nature des liens eux-mêmes. Enfin, le graphe peut être simple ou... non simple.

3. C'est par exemple le type de graphe que l'on obtient par la méthode des flux dominants.

Graphe orienté et non orienté

Le fait de prendre en compte ou non la direction des liens crée la différence entre graphe orienté (*directed graph* ou *digraph*) et graphe non orienté (*graph*). Rappelons-le, par défaut pour les mathématiciens anglophones, un graphe est non orienté ; par défaut pour leurs collègues francophones, il est orienté.

Si le graphe est non orienté, la matrice d'adjacence correspondante est symétrique. En effet, tout lien l_{ij} entraîne la présence du lien l_{ji} . Si le graphe est orienté, la matrice d'adjacence correspondante est (le plus souvent) non symétrique.

La prise en compte de la direction des liens a des conséquences méthodologiques importantes pour mesurer les propriétés du graphe.

La nature des liens

On distingue trois types de graphes en fonction de la nature des liens existants :

- les graphes binaires ou booléens (*binary*) où les liens traduisent l'existence d'une relation entre deux sommets
- les graphes valués (*valued*) où les liens traduisent la présence et l'intensité de la relation
- les graphes signés (*signed*) où les liens expriment soit l'attraction, soit le rejet

Il est tout à fait possible d'imaginer des graphes à la fois valués et signés (soldes commerciaux ou migratoires entre paires d'États), mais la littérature sur le sujet manque.

Les méthodes d'analyse disponibles sont plus nombreuses et mieux connues pour les graphes binaires. Ainsi, il est fréquent que les graphes valués soient transformés en graphes binaires avant d'être étudiés. Tout le problème est alors de justifier le seuil retenu pour effectuer cette opération. . .

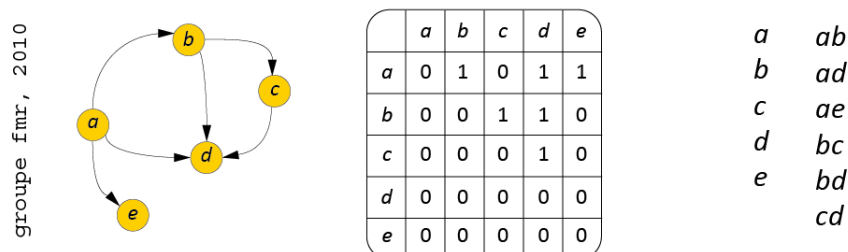
Graphes simples et autres graphes

Pour la théorie des graphes, un graphe simple est constitué d'un seul ensemble de sommets et ne comprend ni boucle - lien d'un sommet vers lui-même (*loop*, *reflexive tie* ou *self-tie*) - ni liens multiples (dans ce dernier cas, on parle de multigraphe).

Si la définition du graphe simple est claire, celle du graphe complexe l'est moins. Dans un bon article de synthèse ([2], les graphes complexes sont définis ainsi « des réseaux dont la structure est irrégulière, complexe et évolutive dans le temps »⁴. Un graphe possède une structure régulière lorsque

4. 'networks whose structure is irregular, complex and dynamically evolving in time', p.177.

FIGURE 2 – Trois modes de représentation pour un même objet



tous les sommets sont incidents au même nombre k de liens : on parle alors de graphe k -régulier (k -regular). Or, quasiment tous les graphes issus de données empiriques sont irréguliers. Il semble que l'étiquette 'complexe' soit très utilisée chez les physiciens, moins utilisée dans d'autres disciplines⁵.

Si le graphe comprend deux ensembles de sommets \mathcal{A} et \mathcal{B} et que chaque lien a une extrémité dans \mathcal{A} et l'autre dans \mathcal{B} , on parle de graphe biparti (*bipartite graph*). Il est possible d'imaginer des graphes tri ou quadriparti, mais les méthodes d'analyse disponibles restent en grande partie à créer.

3 Modes de représentation

Il existe trois grands modes de représentation des graphes ; les listes de liens, les matrices d'adjacence (*adjacency* ou *connectivity matrix*) et le graphe proprement dit, à savoir sa représentation graphique. Ces trois modes sont strictement équivalents et la figure 2 montre trois fois le même graphe.

La forme 'liste de sommets, liste de liens' peut paraître étrange, elle est pourtant souvent utilisée pour deux raisons complémentaires : certains logiciels (Pajek notamment) demandent ce type de format de données en entrée, et elle est la plus économique en terme de volume. Ajoutons également qu'elle est la plus proche de la définition mathématique donnée plus haut.

La forme matricielle est elle extrêmement utile pour calculer certains indicateurs tant au niveau local (sommets ou liens) que global (graphe dans son ensemble).

Il est possible enfin de représenter un graphe *via* une matrice dite d'incidence de taille $N \times K$: dans ce cas, les sommets sont en lignes, les liens sont en colonnes et les cases contiennent un 1 lorsque le sommet s est incident au lien l , un 0 dans le cas contraire.

5. Le terme n'est par exemple quasiment jamais utilisé dans la revue *Social Networks*.

4 Du graphe au réseau (social)

Utiliser le graphe pour symboliser un réseau (le plus souvent technique) ou un réseau social est fréquent, cela oblige pourtant à choisir les méthodes, outils, et concepts adaptés.

Le passage du niveau graphe au réseau se fait dès le moment où on ajoute des informations portant sur les sommets (un nom, des coordonnées, une masse) et/ou sur les liens.

Lorsque l'on étudie un réseau ayant une existence matérielle tangible (réseau routier, réseau hydrographique), il est fréquent que l'on ait recours aux graphes planaires. Ce choix *a priori* logique est pourtant discutable. Il suffit d'ajouter quelques tunnels et ponts à un réseau routier pour qu'il perde son caractère planaire... Lorsqu'on étudie des flux (de marchandises, de personnes, d'informations), il est fréquent en revanche de les symboliser à l'aide de graphes non planaires.

Le choix entre ces deux types de graphes n'est pas anodin, car les mesures et méthodes mobilisables pour ces deux types de graphes ne sont pas les mêmes.

Il importe également de souligner que si le graphe peut prétendre modéliser⁶ toutes formes de relations, l'image produite n'est toujours qu'une image parmi d'autres possibles et qu'elle est la conséquence de choix (seuillage, algorithme de visualisation) rarement explicités.

Conclusion provisoire

Ce petit balayage terminologique n'a pas la prétention d'être complet. D'autres disciplines mobilisent le vocabulaire et certains outils de la théorie des graphes, comme les réseaux de neurones en intelligence artificielle, les réseaux bayésiens qui associent graphes et probabilités, et il en est sans doute d'autres.

Il est important (et cette phrase reviendra souvent) de se souvenir que la complexité apparente du (des) vocabulaire(s) est uniquement due à l'utilisation par des disciplines différentes de termes identiques pour désigner des outils différents, *et* à l'utilisation de termes différents pour désigner des outils identiques.

Références

- [1] C. Bergé. *Graphes*. Gauthier-Villars, 2 édition, 1973.
- [2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.U. Hwang. Complex networks : Structure and dynamics. *Physics Reports*, 424(4-5) :175–308, 2006.

6. Le terme modèle est utilisé ici comme représentation simplifiée de la réalité.

- [3] P. Haggett and R.J. Chorley. *Network analysis in geography*. Edward Arnold, 1969.
- [4] P. Mathis. *Graphes et réseaux. Modélisation multiviveau*. coll. « Information Géographique et Aménagement du Territoire ». Lavoisier, 2003.
- [5] S. Wasserman and K. Faust. *Social Network Analysis. Methods and Applications*. coll. 'Structural analysis in the social sciences'. Cambridge University Press, 1994.
- [6] D.B. West. *Introduction to graph theory*. Prentice Hall Upper Saddle River, 2001.

Table des matières

1	Ce qu'est un graphe	1
2	Les différents types de graphes	3
3	Modes de représentation	5
4	Du graphe au réseau (social)	6