



Blockmodeling et équivalences

Laurent Beauguitte

► **To cite this version:**

Laurent Beauguitte. Blockmodeling et équivalences. Cinquième document produit par le groupe fmr (flux, matrices, réseaux). 2011. <halshs-00566474>

HAL Id: halshs-00566474

<https://halshs.archives-ouvertes.fr/halshs-00566474>

Submitted on 16 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Blockmodeling et équivalences

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

Version 1 - Février 2011



Ce document du groupe fmr est la suite logique du document fmr4[1] consacré aux cliques et aux communautés.

Introduction

Si la recherche des cliques ou des communautés permet de regrouper les sommets les plus centraux d'un graphe, la centralité étant essentiellement fonction du degré, *i*) cela ne permet pas une partition de l'ensemble des sommets et *ii*) la structure des relations entre sommets n'est pas prise en compte.

Les méthodes de *blockmodeling*¹ à l'inverse créent des partitions disjointes de l'ensemble des sommets d'un graphe en se basant uniquement sur les propriétés relationnelles des différents acteurs. Ce souci était très clairement à l'origine un moyen pour refuser toute idée de classe sociale². Dès le départ cependant, certains auteurs utilisent cette technique tout en gardant une approche « structuraliste »³.

Le plan de la présentation est le suivant : la première partie décrit les étapes et principes du *blockmodeling*, les deuxième et troisième parties sont consacrées à deux méthodes classiques de recherche d'équivalence. Une quatrième partie expose les pistes nouvelles proposées par différents auteurs, tant du côté de la sociologie que de la physique et de l'informatique. Enfin, une dernière partie présente les fonctions disponibles dans les logiciels les plus couramment utilisés.

1. Les difficultés terminologiques sont suffisamment nombreuses dans le domaine de l'analyse des réseaux pour nous faire renoncer à proposer un équivalent francophone.

2. Ainsi on lit dans White *et al.*[22] : “ the resulting blockmodel is a view of social structure obtained directly from aggregation of the relational data without imposing any a priori categories or attributes for actors.”

3. Voir par exemple le couple d'articles de Burt en 1977[6][7].

FIGURE 1 – Les étapes du *blockmodeling*

Adjacency matrix							
	A	B	C	D	E	F	G
A	0	0	0	0	0	0	0
B	0	0	1	0	0	1	0
C	0	0	0	0	0	0	0
D	1	0	0	0	0	0	0
E	0	0	0	0	0	0	0
F	0	1	0	1	0	0	0
G	0	0	0	0	0	0	0

Blocked matrix							
	B	F	C	D	E	G	A
B	0	1	1	0	0	0	0
F	1	0	0	1	0	0	0
C	0	0	0	0	1	1	0
D	0	0	0	0	0	0	1
E	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0

Image matrix			
	1	2	3
1	1	1	0
2	0	0	1
3	0	0	0

L'*image matrix* étant constituée uniquement de *zeroblocks* et de *oneblocks*, représenter la matrice de densité est ici inutile.

1 Du graphe à l'*image matrix*

Si le vocabulaire n'est pas totalement harmonisé, la plupart des auteurs en *Social Network Analysis* (SNA) différencient les notions de position et de rôle. La position agrège un ensemble de sommets impliqués dans le même type de relations au sein d'un réseau social, le rôle se réfère aux systèmes de relations entre les sommets ou entre les positions⁴.

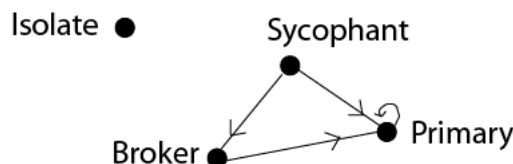
Le principe du *blockmodeling* est relativement intuitif : la première étape consiste à permuter lignes et colonnes d'une matrice d'adjacence afin de positionner les uns à côté des autres les sommets qui ont le même type de relations avec le même type d'acteurs. La matrice d'adjacence obtenue après permutation est nommée une matrice bloquée (*blocked matrix*). La dernière étape consiste à agréger les sommets appartenant aux mêmes blocs pour créer une matrice de densité (*density matrix*). Si le graphe de départ est simple (absence de boucle), les diagonales ne sont pas prises en compte pour le calcul des densités. La matrice de densité est transformée en une *image matrix* correspondant au *blockmodel*. Cette *image matrix* peut être représentée par un graphe réduit (*reduced graph*). Au final, le *blockmodel* consiste en une *image matrix* et en une partition des sommets entre les différents blocs.

La figure 1 montre les étapes à partir d'un cas idéal.

Avec des données réelles, il est souvent nécessaire de fixer un seuil de densité pour déterminer quels liens on garde entre les différentes positions obtenues. Burt[5] a proposé une typologie des positions souvent reprise par la suite où deux critères permettent de créer quatre catégories : le fait de

4. 'In social network analysis position refers to a collection of individuals who are similarly embedded in networks of relations, while role refers to the patterns of relations which obtain between actors or between positions' ([21], p. 348). Pour une discussion approfondie du concept, voir [13].

FIGURE 2 – Typologie des positions selon R. Burt



Ces quatre labels peuvent être adaptés, ainsi dans une étude sur des flux commerciaux ou financiers, on pourrait utiliser les termes d'isolat, de périphérie (intégrée ou non) et de centre.

recevoir ou non un lien d'une autre position ; le fait d'avoir une majorité de liens à l'intérieur de la position ou vers une autre. Ce seuil, tout comme les étiquettes proposées, peut être modifié (voir figure 2).

Les sommets regroupés en blocs le sont non parce qu'ils sont liés les uns aux autres, mais parce qu'ils émettent et/ou reçoivent des liens vers le même type d'acteurs. Ils sont dits équivalents.

Trois formes d'équivalence sont classiquement distinguées : l'équivalence structurale (*structural equivalence*), automorphique (*automorphic equivalence*) et régulière (*regular equivalence*).

2 L'équivalence structurale

Principes généraux

Selon la définition devenue classique de Lorrain et White[16], deux sommets sont dits structurellement équivalents s'ils ont les mêmes liens avec les mêmes sommets au sein d'un graphe⁵. Le cas est relativement rare comme on peut le voir sur le graphe orienté de la figure 3⁶.

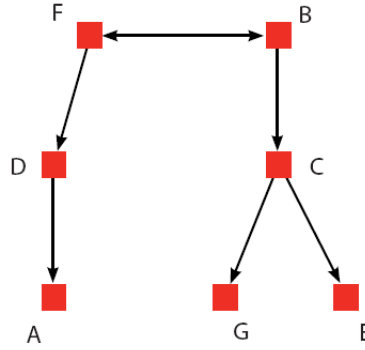
Burt[5] a proposé l'utilisation de la distance euclidienne pour mesurer l'équivalence structurale. Soit deux sommets i et j et d_{ij} la distance entre i , j et les autres sommets d'un graphe d'ordre g :

$$d_{ij} = \sqrt{\sum_{k=1}^g [(x_{ik} - x_{jk})^2 + (x_{ki} - x_{kj})^2]}$$

5. "a is structurally equivalent to b if a relates to every object x of C in exactly the same ways as b does", C désignant le graphe.

6. Les deux figures sont tirées de L. Beauguitte, 2010, "Basic notions on SNA - Equivalence(s)", disponible sur http://cnrs4.nfrance.com/euro_drupal/node/26

FIGURE 3 – Équivalence structurale, équivalence régulière



Les sommets E et G sont les seuls à être structurellement équivalents (ils reçoivent tous deux un lien du même acteur). Les sommets concernés par l'équivalence régulière sont les suivants AEG , CD et BF .

pour $i \neq k$ et $j \neq k$. Les deux sommets seront structurellement équivalents si $d_{ij} = 0$.

L'utilisation de la corrélation est également fréquemment utilisée pour la recherche de l'équivalence structurale. La moyenne des relations en ligne de l'acteur i est notée \bar{x}_i , et la moyenne des relations en colonne du même acteur est notée $\bar{x}_{.i}$. On obtient :

$$r_{ij} = \frac{\sum(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) + \sum(x_{ik} - \bar{x}_{.i})(x_{jk} - \bar{x}_{.j})}{\sqrt{\sum(x_{ki} - \bar{x}_i)^2 + \sum(x_{ik} - \bar{x}_{.i})^2} \sqrt{\sum(x_{kj} - \bar{x}_j)^2 + \sum(x_{jk} - \bar{x}_{.j})^2}}$$

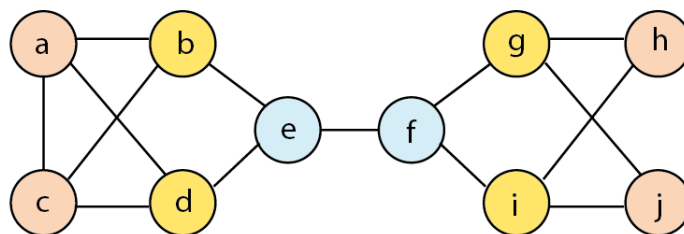
où $i \neq k$ et $j \neq k$. La diagonale n'est pas prise en compte dans les calculs. Si deux acteurs sont structurellement équivalents, la corrélation entre leurs lignes et colonnes respectives sera de 1.

CONCOR et CAH

CONCOR (acronyme de CONvergence of iterated CORrelations) est une méthode basée sur des corrélations itératives mise au point dans les années 70[4]. Si on corrèle une matrice de distance avec elle-même x fois, on obtient une matrice contenant uniquement des 1 et des 0 ce qui crée une première partition en deux sous groupes. Les deux matrices issues de cette partition sont à nouveau soumises à une corrélation par itération pour obtenir une partition en quatre groupes. Cette classification descendante ne peut donc que produire des nombres de classe pairs.

L'inconvénient principal de cette méthode, par ailleurs efficace, est que les fondements mathématiques de la méthode restent obscurs et que des tests

FIGURE 4 – Sommets automorphiquement équivalents



effectués sur des matrices dont les blocs sont connus ont souvent donné des résultats décevants. Elle reste cependant utilisée en raison de sa robustesse.

Les deux sociologues Snyder et Kick[20] ont proposé une partition du monde en centre, semi-périphérie et périphérie en utilisant CONCOR sur quatre matrices représentant quatre types de liens (économiques, diplomatiques, militaires et traités). Cet article a donné lieu à une descendance fournie (dont l'article de Smith et White sur les seuls flux commerciaux[19]) et l'intérêt de la méthode pour tout géographe étudiant les dynamiques du Système-Monde est réel.

Les autres méthodes permettant de mettre en évidence une équivalence structurale sont bien connues des géographes dans la mesure où il s'agit des classifications ascendantes hiérarchiques appliquées sur une matrice de distance (similarité, dissimilarité, corrélation, distance euclidienne, etc.) entre les sommets. Les différentes options disponibles (lien complet, lien moyen, lien simple, etc.), leurs avantages et inconvénients sont décrits dans tout bon manuel de statistiques multivariées. Il peut être intéressant de tester plusieurs méthodes pour valider la robustesse des partitions produites.

3 L'équivalence automorphique

Les termes d'équivalence automorphique (*automorphique equivalence*) et d'isomorphisme structural (*structural isomorphism*) sont synonymes. Cette forme d'équivalence est plus souple que la précédente, il n'est en effet pas nécessaire pour que deux sommets puissent être équivalents qu'ils partagent les mêmes relations avec les mêmes acteurs mais qu'ils aient la même structure de relations. Ou, de façon plus intuitive, si les noms des sommets sont effacés, on peut intervertir deux sommets isomorphes et la structure du graphe est inchangée. Deux sommets appartenant à des sous-graphes non connexes (composants) peuvent donc être automorphiquement équivalents (mais non structurellement équivalents).

Dans la figure 4 tirée de Borgatti et Everett[3], les différents couleurs

montrent les blocs des sommets isomorphes. Ces deux auteurs ont proposé d'appeler *orbits* les blocs créés, ce terme n'a pas été repris ensuite.

4 L'équivalence régulière

L'équivalence régulière regroupe les sommets qui ont le même type de liens avec le même type d'acteurs. Quatre positions de base existent dans un graphe orienté : les isolés, les émetteurs (*source, transmitter*), les récepteurs (*sink, receiver*) et les intermédiaires (*carrier*)⁷.

L'un des équations les plus utilisées pour la mesure de l'équivalence régulière[12] entre deux sommets i et j se calcule ainsi :

$$M_{ij}^{t+1} = \frac{\sum_{k=1}^N \max_{m=1}^N \sum_{q=1}^Q M_{km}^t (i_{qj} M_{km}^t + j_{qi} M_{km}^t)}{\sum_{k=1}^N \max_m^* \sum_{q=1}^Q (i_{qj} \text{Max}_{km} + j_{qi} \text{Max}_{km})}$$

où N est le nombre de sommets et Q le nombre de liens. M varie de 0 (équivalence nulle) à 1 (équivalence parfaite).

Si la méthode était initialement prévue pour des graphes binaires et orientés, des auteurs ont proposé des adaptations, tant pour les graphes signés[10] que pour les graphes valués[17] ou les graphes ordinaux (voir notamment l'algorithme CATREGGE développé par Borgatti[2] et implémenté dans Ucinet).

5 Autres méthodes de *blockmodeling*

Dans un panorama assez technique des méthodes disponibles, Nunkesser et Sawitzki[18] évoquent également l'utilisation de modèles statistiques (p_1, p^*) comme option possible. En deux mots⁸, il s'agit de tester la robustesse d'une partition *a priori* d'un graphe et d'utiliser cette partition comme variable dans un modèle de régression. Comme le soulignent les auteurs en conclusion, ces méthodes restent utilisées de façon marginale comparées aux méthodes traditionnelles de *blockmodeling*.

Les physiciens, qui ont beaucoup retravaillé le degré et l'intermédiarité[11] ont peu mobilisé l'équivalence. Une des rares pistes présentées se trouve dans un article de Guimerà et Amaral[15] où les auteurs commencent par expliquer que les méthodes traditionnelles ne permettent pas d'obtenir de résultats intéressants. Ils proposent à la place une mesure fondée sur le *within-module degree* et le coefficient de participation (*participation coefficient*), mesures qui ne peuvent cependant être réalisées qu'une fois les partitions pertinentes

7. Voir la définition simple donnée par Borgatti et Everett[3] : "A source is a node with outdegree, but not indegree. A sink is a node with indegree, but no outdegree."

8. Ces modèles statistiques seront traités dans un papier fmr futur.

(appelées *modules* par les auteurs) connues⁹. Il semble qu'ils n'aient pas développé cette piste ultérieurement.

6 Quelques logiciels adaptés

STRUCTURE est l'un des premiers programmes, développé par Burt en 1981, permettant de calculer l'équivalence régulière basée sur la distance euclidienne. L'algorithme est implémenté Ucinet. Ce dernier est sans doute le logiciel qui propose le plus de méthodes de *blockmodeling* dans la mesure où la recherche des trois types d'équivalence est disponible. Des options permettent également de choisir l'algorithme adapté au type de réseau. Les sorties sont particulièrement claires (composition des positions, matrice bloquée, matrice de densité).

Si le nombre de sommets dépasse 256, Pajek propose la recherche de l'équivalence régulière et structurale, et les différentes distances de CAH sont présentes (lien faible, moyen, fort, Ward). Les deux commandes sont marquées d'une étoile * ce qui signifie que le graphe doit être de taille raisonnable (et votre ordinateur de puissance correcte).

Les différents modules d'analyse de graphe du logiciel R sont les seuls (à ma connaissance) à permettre le *blockmodeling* de graphes de grande taille, de type différent, et sans faire mouliner durant des heures un ordinateur aux performances modestes. Si le package 'statnet' ne permet que le *blockmodeling* de graphes binaires[14], le package 'blockmodeling' est lui spécifiquement dédié aux graphes valués[23].

Il semble enfin que les logiciels développés par des informaticiens (Gephi, Tulip) ne proposent pas ce type d'outils.

Conclusion

À ma connaissance, les géographes n'ont quasiment jamais utilisé ces méthodes. Or elles peuvent s'avérer tout à fait adaptées pour produire une partition à partir d'une matrice de flux entre unités spatiales ou pour une étude sur des publications scientifiques[8][9]. Claude Grasland a ainsi développé récemment un programme R utilisant le principe de CONCOR pour produire des partitions à l'échelle mondiale à partir de flux commerciaux (papier en cours d'écriture).

Leur implantation dans la plupart des logiciels disponibles ainsi que la littérature théorique, méthodologique et empirique existant sur le sujet rendent ces méthodes aisément accessibles et l'un des objectifs du groupe fmr est justement de les utiliser sur des matrices afin d'étudier leur pertinence et leur utilité en géographie. Affaire à suivre donc...

9. "Note that these two properties are easily computed once the modules of a network are known."

Références

- [1] L. BEAUGUITTE : Cliques, communautés et dérivés. *Groupe fmr*, 8p., 2010 (<http://halshs.archives-ouvertes.fr/FMR/fr/>).
- [2] S.P. BORGATTI : Two algorithms for computing regular equivalence. *Social Networks*, 15(4):361–376, 1993.
- [3] S.P. BORGATTI et M.G. EVERETT : Notions of position in social network analysis. *Sociological Methodology*, 22:1–35, 1992.
- [4] R.L. BREIGER, S.A. BOORMAN et P. ARABIE : An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12(3):328–383, 1975.
- [5] R. BURT : Positions in Networks. *Social Forces*, 55(1):93–122, 1976.
- [6] R.S. BURT : Positions in multiple network systems, part one : A general conception of stratification and prestige in a system of actors cast as a social topology. *Social Forces*, 56(1):106–131, 1977.
- [7] R.S. BURT : Positions in multiple network systems, part two : Stratification and prestige among elite decision-makers in the community of Altneustadt. *Social Forces*, 56(2):551–575, 1977.
- [8] P. DOREIAN : Structural equivalence in a psychology journal network. *Journal of the American Society for Information Science*, 36(6):411–417, 1985.
- [9] P. DOREIAN : Testing structural-equivalence hypotheses in a network of geographical journals. *Journal of the American Society for Information Science*, 39(2):79–85, 1988.
- [10] P. DOREIAN : A multiple indicator approach to blockmodeling signed network. *Social Networks*, 30(3):247–258, 2008.
- [11] C. DUCRUET : Les mesures globales d'un réseau. *Groupe fmr*, 8p., 2010 (<http://halshs.archives-ouvertes.fr/halshs-00541902>).
- [12] K. FAUST : Comparison of methods for positional analysis : Structural and general equivalences. *Social Networks*, 10(4):313–341, 1988.
- [13] K. FAUST et S. WASSERMAN : Blockmodels : Interpretation and evaluation. *Social Networks*, 14(1-2):5–61, 1992.
- [14] S. GOODREAU, M. HANDCOCK, D. HUNTER, C. BUTTS et M. MORRIS : A statnet tutorial. *Journal of Statistical Software*, 24(9):1–26, 2007.
- [15] R. GUIMERÀ et L.A.N. AMARAL : Cartography of complex networks : modules and universal roles. *Journal of Statistical Mechanics : Theory and Experiment*, 2005(2), 2005.
- [16] F. LORRAIN et H.C. WHITE : Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology*, 1(1):49–80, 1971.

- [17] C. NORDLUND : Identifying regular blocks in valued networks : A heuristic applied tot the St. Marks carbon flow data, and international trade in cereal products. *Social Networks*, 29(1):59–69, 2007.
- [18] M. NUNKESSER et D. SAWITZKI : *Network Analysis*, chapitre Blockmodels, pages 253–292. Springer, 2005.
- [19] D.A. SMITH et D.R. WHITE : Structure and dynamics of the global economy : Network analysis of international trade 1965-1980. *Social Forces*, 70(4):857–893, 1992.
- [20] D. SNYDER et E.L. KICK : Structural position in the world system and economic growth, 1955-1970 : A multiple-network analysis of transnational interactions. *American Journal of Sociology*, 84(5):1096–1126, 1979.
- [21] S. WASSERMAN et K. FAUST : *Social Network Analysis. Methods and Applications*. Structural analysis in the social sciences. Cambridge University Press, 1994.
- [22] H. WHITE, S. BOORMAN et R. BREIGER : Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions. *The American Journal of Sociology*, 81(4):730–780, 1976.
- [23] A. ZIBERNA : Generalized blockmodeling of valued networks. *Social Networks*, 29(1):105–126, 2007.

Table des matières

1	Du graphe à l’<i>image matrix</i>	2
2	L’équivalence structurale	3
3	L’équivalence automorphique	5
4	L’équivalence régulière	6
5	Autres méthodes de <i>blockmodeling</i>	6
6	Quelques logiciels adaptés	7