



HAL
open science

Evaluer la qualité d'une fragmentation de graphe multi-niveaux

Carole Blanc, Maylis Delest, Jean-Marc Fédou, Guy Melançon, François
Queyroi

► **To cite this version:**

Carole Blanc, Maylis Delest, Jean-Marc Fédou, Guy Melançon, François Queyroi. Evaluer la qualité d'une fragmentation de graphe multi-niveaux. MARAMI 2010, Jun 2010, Toulouse, France. hal-00542484

HAL Id: hal-00542484

<https://hal.science/hal-00542484v1>

Submitted on 2 Dec 2010

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.

Évaluer la qualité d'une fragmentation de graphe multi-niveaux

Carole Blanc* — Maylis Delest*,** Jean-Marc Fedou*** — Guy Melançon*,**
— François Queyroi*

* Université de Bordeaux, CNRS, LaBRI

** INRIA Bordeaux – Sud-Ouest, France

*** Université de Nice, CNRS UMR 6070 I3S, France

{blanc,maylis,melancon}@labri.fr fedou@unice.fr francois.queyroi@gmail.com

RÉSUMÉ. Nous présentons ici une mesure de qualité multi-niveaux pour les graphes fragmentés prenant explicitement en compte l'aspect hiérarchique de la fragmentation. Cet aspect est très présent dans les sciences sociales et la biologie. Notre mesure de qualité multi-niveaux généralise une mesure introduite par Mancoridis proposée dans un contexte de rétro-ingénierie en informatique. Le critère que nous définissons traverse récursivement la hiérarchie des fragments et aboutit à un polynôme à une variable prenant en compte la connectivité interne et externe des fragments apparaissant à chaque niveau de la hiérarchie. Ce polynôme reflète la qualité de l'ajustement entre le graphe et la hiérarchie proposée. Nous présentons plusieurs exemples de manière à justifier notre concept.

ABSTRACT. We design and study a multilevel modularity quality for clustered graphs, explicitly taking the nesting structure of clusters into account. Multilevel models appear crucial in the natural and social sciences. The multilevel modularity quality measure generalizes a modularity quality measure introduced by Mancoridis in the context of reverse software engineering. The measure we designed recursively traverses the hierarchy of clusters and computes a one-variable polynomial encoding the intra and inter-cluster connectivity ratios appearing at all levels in a hierarchical clustering. The resulting polynomial reflects how the graph combines with the hierarchy of clusters and can be used to assess the quality of a hierarchical clustering. We discuss examples as proof-of-concept.

MOTS-CLÉS : Multi-niveaux, Fragmentation de graphes hiérarchiques, Mesure de qualité, Modularité, q-analog, Combinatoire

KEYWORDS: Multilevel, Hierarchical Graph Clustering, Quality measures, Modularity, q-analog, Combinatorics

1. Introduction

Les graphes de terrain sont aujourd'hui l'objet de nombreux travaux cherchant à comprendre leur structure ou leur évolution. Citons les travaux sur la métrologie ou sur les différentes mesures qu'on peut conduire sur ceux-ci [LAT 08, MAG 10]. Par ailleurs, les graphes de terrain sont devenus des objets familiers des applications grand public du web ou d'autres domaines. Une approche possible pour appréhender ces graphes de grande taille est d'y rechercher une structure de communautés : des sous-groupes denses connectés les uns aux autres de manière plus lâches. Cette notion popularisée par Newman [NEW 06] est héritière de la sociologie et plus particulièrement des travaux de Moreno (voir [FRE 04]) en sociométrie. Les groupes sociaux sont habituellement formés de gens qui, pris deux à deux, se connaissent quasiment tous. Certains membres de ces groupes, ouverts sur l'extérieur agissent en qualité de pivots et sont parfois à l'origine de contacts entre des personnes qui appartenaient au départ à des communautés différentes. Cette organisation des réseaux et leur corrélation à la dynamique sociale a été abondamment étudiée par Burt [BUR 05].

Ainsi, on trouve dans la littérature bon nombre d'approches cherchant à identifier ces communautés (on parle aussi de modules, ou de clusters) [CLA 04, BLO 08] mettant en avant la notion de densité des sous-graphes en nombre d'arêtes comme élément déterminant de la modularité des graphes. L'organisation des groupes en sous-groupes, et sur plusieurs niveaux, est obtenue en réitérant la procédure pour chacune des communautés [BRA 08]. Outre le fait de décrire la structure des graphes, cette organisation par niveaux présente l'avantage de fragmenter le graphe en plus petites entités facilitant de fait leur visualisation [AUB 03, HAM 04, AUB 07].

Le critère de modularité Q introduit par Newman semble aujourd'hui le plus utilisé [NEW 06]. Un autre critère, MQ introduit par Mancoridis [MAN 98], a retenu notre attention. Ce critère et sa généralisation que nous présentons ici posent les bases conceptuelles d'une théorie du découpage en communautés multi-niveaux des graphes de terrain.

Du point de vue opérationnel, l'idée d'itérer un algorithme d'identification de communautés sur chaque module permet effectivement d'obtenir une structure de communautés emboîtées sur plusieurs niveaux. Mais on peut opposer à cette heuristique que le calcul des sous-communautés se fait alors sans tenir compte du découpage qui est fait des autres sous-groupes. En d'autres termes, le découpage des sous-communautés se fait systématiquement sur la base d'un critère local à la sous-communauté qui est découpée, sans vision globale de l'effet de ce découpage sur l'ensemble de la hiérarchie qu'il s'agit de produire.

Cet article propose donc un critère opérant à la fois au niveau global et local, cherchant à influencer sur la décision du découpage local en fonction de son effet sur les autres composantes de la hiérarchie. Nous motivons la définition du critère et l'illustrons. La première section revient sur la définition de la mesure MQ et une reformulation de celle-ci. L'approche multi-niveaux est décrite dans la deuxième section tandis que dans la troisième section présentent quelques exemples.

2. La mesure de qualité de Mancoridis

2.1. Définition originale de la mesure

Mancoridis *et al.* [MAN 98] ont proposé une mesure de qualité de fragmentation MQ (pour *Modularity Quality*) reposant sur la différence entre les ratios de connectivité interne et externe des fragments. Elle fut employée par ces mêmes auteurs dans le cadre d'un algorithme de partitionnement du code source d'un programme. Ils se placent dans un cadre orienté pour rendre compte des relations d'inclusion entre les modules.

Soit $G = (V, E)$ un graphe, avec un ensemble de sommets V et d'arêtes E . Soit $C = (C_1, \dots, C_k)$ une partition de cardinalité k , on appellera fragment un élément de la partition. Soient deux fragments C_i, C_j , nous notons e_{ij} le nombre d'arêtes liant les sommets de C_i à ceux de C_j ; nous noterons également t_i le cardinal de C_i . La mesure de qualité, notée \widetilde{MQ} , représente la moyenne des différences entre les ratios de connectivité intra et inter fragments. Formellement,

$$\widetilde{MQ}(G; C) = \frac{1}{k} \left(\sum_{i=1}^k \frac{e_{ii}}{t_i^2} - \frac{1}{k-1} \sum_{j \neq i} \frac{e_{ij}}{t_i t_j} \right). \quad (1)$$

2.2. Généralisation de la mesure de qualité de Mancoridis

Nous souhaitons généraliser la mesure de Mancoridis définie en (1) de manière à pouvoir l'appliquer à des types de structures variés. On pourra ainsi exprimer MQ pour toute les configurations possibles de G .

L'influence du type de graphe intervient lorsque l'on cherche à quantifier le maximum d'arêtes entre et à l'intérieur des fragments. On pose donc $\delta(t_i, t_j)$ (noté $\delta(t_i)$ si $i = j$) une fonction dépendante de ce que nous appellerons le *graphe de référence* et qui associe à la taille de deux fragments C_i, C_j le nombre maximal d'arêtes pouvant exister entre eux. Si, par exemple, on prend le graphe non-orienté complet sans boucles comme graphe de référence, on aura $\delta(t_i) = \frac{t_i(t_i-1)}{2}$ et $\delta(t_i, t_j) = t_i \cdot t_j$. Notons que cette définition permet également de tenir compte d'une éventuelle valuation des arêtes. On souhaite également pouvoir modifier la pondération liée aux fragments. On définit ainsi x_i un poids associé au fragment C_i et X_C sa somme sur les k éléments de C . On définit une formule généralisée de la mesure de Mancoridis :

$$MQ(G; C) = \frac{1}{X_C} \sum_{i=1}^k x_i (\alpha_i - \beta_i) \quad (2)$$

Dans cette définition, α_i et β_i sont respectivement les ratios de connectivité interne et externe du fragment C_i . Plus précisément,

$$\alpha_i = \frac{e_{i,i}}{\delta(t_i)}, \quad (3)$$

qui correspond à la densité d'arêtes dans le fragment C_i et

$$\beta_i = \frac{1}{X_C - x_i} \sum_{j \neq i} \frac{x_j e_{i,j}}{\delta(t_i, t_j)}, \quad (4)$$

qui correspond à la moyenne pondérée des ratios de connectivités externes entre C_i et ses frères.

La différence de ces ratios sera pondérée par x_i pour chaque fragment C_i . L'expression originale de la mesure donnée par Mancoridis (1) pour un graphe orienté s'obtient en prenant $\delta(t_i) = t_i(t_i - 1)$, $\delta(t_i, t_j) = t_i \cdot t_j$ et $x_i = 1$ pour chaque fragment C_i . On pourrait choisir $x_i = t_i$ de manière à accorder moins de poids aux petits fragments. MQ serait alors le barycentre de la différence entre connectivités interne et externe sur la fragmentation.

La mesure varie dans l'intervalle borné $[-1, 1]$. Elle attribue un score maximal aux cliques (sous-graphes ayant un nombre maximal d'arêtes internes dans le cadre du graphe de référence) et favorise les sous-ensembles compacts plutôt que les sous-graphes plus larges mais moins denses.

3. L'approche multi-niveaux de la modularité

Notre généralisation passe par une définition récursive de MQ prenant en compte une variable q . En effet, la mesure définie dans (2) peut être calculée en considérant chaque arête de E une par une, et en leur assignant un poids positif (respectivement négatif) si celles-ci connectent deux sommets d'un même fragment (resp. de deux différents). En laissant de côté les pondérations et les constantes, ces poids correspondent à 1 (resp. -1).

Dans le cadre d'une fragmentation multi-niveaux, on veut tenir compte des profondeurs auxquelles une arête est prise en compte. L'intuition est qu'une arête doit se voir assigner un poids positif proportionnel à $1 + q + q^2 + \dots + q^{p-1}$ tant que l'arête reste interne lors d'un parcours des p premiers niveaux de l'arbre de fragmentation. De la même manière, l'arête se verra assignée un poids négatif proportionnel à $q^p + q^{p+1} + \dots$ si elle lie deux sommets issus de fragments différents à partir du niveau $p + 1$.

Les motivations pour l'emploi de cette formulation de MQ par des q -analogue proviennent du domaine de la combinatoire énumérative pour l'étude récursive d'objets discrets. La variable q permet de garder une trace de la profondeur intrinsèque de l'objet. Dans la plupart des cas, les objets peuvent être décrits par des langages formels générés par des grammaires algébriques, appelées *grammaires attribuées* avec l'introduction d'une variable de comptage q (voir [DEL 92] et [MIS 03]). On peut ainsi prendre en compte dans une même série énumératrice à la fois un paramètre algébrique tel que la longueur du mot et un paramètre non-algébrique tel que la position d'une lettre dans un mot. ceci permet d'obtenir par exemple sur des objets de type "polyominos" des séries énumératrices suivant l'aire et le périmètre.

3.1. Définition récursive

Une fragmentation hiérarchique est décrite par un arbre T tel que chaque fragment C_i issu de la racine peut constituer une sous-hiérarchie $(C_{i_1}, \dots, C_{i_{k_i}})$ et ainsi de suite pour chaque $(C_{i_j})_{1 \leq j \leq k_i}$. On code chaque nœud de l'arbre T par un mot w formé d'entiers $i_1 \dots i_{p-1} i_p$ représentant le chemin $i_1 \dots i_{p-1}$ depuis la racine et le nombre i_p de frères à gauche du nœud (plus un) au niveau p . Le nœud racine est codé avec le mot vide. Dans l'exemple donné en figure [1], le nœud noté "222" au niveau $p = 3$ a pour ancêtre le nœud "22" et un frère à gauche.

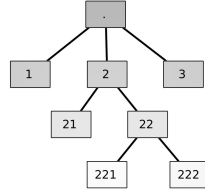


Figure 1. Un arbre de fragmentation annoté

Nous définissons maintenant la mesure de qualité multi-niveaux :

Definition 1 Pour tout nœud w de T . Soit C_w le fragment correspondant, G_w le sous-graphe induit par C_w et T_w le sous-arbre de T issu de w . Le nombre de descendants directs de w est noté k_w et l'indice w_i désigne le i -ème descendant de w . On pose également $\gamma_w = (\alpha_w - \beta_w)$, la différence entre les ratios des connectivités interne(3) et externe(4) sur C_w . Pour $q \in [0, 1]$, on a :

$$MQ(G_w; T_w; q) = \begin{cases} 0 & \text{si } k_w = 0 \\ \frac{1}{X_{C_w}} \sum_{i=1}^{k_w} x_{w_i} \gamma_{w_i} \cdot (1 + q \cdot MQ(G_{w_i}; T_{w_i}; q)) & \text{sinon} \end{cases} \quad (5)$$

Par abus de langage, nous écrirons $MQ(q)$ ou plus simplement MQ pour se référer à la fonction (5) calculée sur l'ensemble du graphe en partant de la racine de T lorsqu'il n'y aura pas d'ambiguïtés. Si on étudie un arbre de fragmentation plat on retrouve bien la formule donnée en (2).

Comme on peut le voir dans l'équation (5), MQ est pondérée par la différence des ratios de connectivités interne et externe de C_w , fragment racine de T_w . On répercute ainsi la bonne ou mauvaise qualité d'une fragmentation aux niveaux suivants.

3.2. Formulation sur la base de chemins dans l'arbre de fragmentation

Le coefficient q^p dans MQ noté $[MQ; q^p]$ peut être exprimé individuellement en utilisant la notion de chemins dans l'arbre de fragmentation. L'idée est d'obtenir une formulation de MQ plus compréhensible et facile à implémenter.

On note \mathcal{D} l'ensemble des mots, \mathcal{L} l'ensemble des mots correspondant aux feuilles de l'arbre, la notation $u \prec w$ signifie que u est un préfixe non-vide de w , classiquement $|w|$ désigne la longueur du chemin.

Théorème 3.1 *Si la mesure x_w est additive pour tout nœud de T codé par w , on a*

$$MQ(G; T; q) = \frac{1}{X_C} \sum_{w \in \mathcal{L}} x_w \left(\sum_{v \prec w} q^{|v|-1} \left(\prod_{u \prec v} \gamma_u \right) \right) \quad (6)$$

Une autre formulation est possible si on parcourt l'arbre de manière itérative en prenant les coefficients un par un.

Théorème 3.2 *Soit l'ensemble $\mathcal{D}_p = \{w \in \mathcal{D}, |w| = p + 1\}$, si la mesure x_w est additive pour tout nœud de T codé par w ,*

$$[MQ; q^p] = \frac{1}{X_C} \sum_{v \in \mathcal{D}_p} x_v \prod_{u \prec v} \gamma_u \quad (7)$$

Cette formule donne des valeurs pour chaque coefficient égales à celles obtenues à partir de la formule (6).

La mesure MQ s'obtient en sommant $[MQ; q^p]$ sur toutes ses valeurs non-nulles *i. e.* telles que $0 \leq p \leq h(T) - 1$ où $h(T)$ désigne la hauteur de l'arbre T .

$$MQ(G; T; q) = \frac{1}{X_C} \sum_{p=0}^{h(T)-1} q^p \sum_{v \in \mathcal{D}_p} x_v \prod_{u \prec v} \gamma_u \quad (8)$$

Le calcul de (8) est de complexité $O(n^2)$.

3.3. Interprétation de la mesure

Remarquons que les objectifs cités au début de cette section sont atteints car une arête sera prise en compte plusieurs fois, une première en tant qu'arête de G , ensuite en tant qu'arête de $(G_i)_{1 \leq i \leq k}$ et ainsi de suite, contribuant à MQ à travers différentes puissances de q , le poids attribué dépendant de la profondeur.

Intéressons nous maintenant à l'interprétation de la variable q . Le cas où $q = 1$ correspond à la situation extrême où le poids positif d'une arête dans MQ est proportionnel à la profondeur à laquelle cette arête est encore interne et où son poids négatif est proportionnel à la différence entre la profondeur de l'arbre et la profondeur à laquelle l'arête est devenue externe. Comme nous le verrons dans la section 4, la valeur de q joue un rôle important dans le choix entre une fragmentation plus ou moins profonde. De manière générale, selon la valeur de q , un fragment dense pourra avoir

une contribution plus petite qu'un fragment moins dense mais se situant à un niveau hiérarchique plus élevé.

Il est évident que deux arbres de fragmentation T, T' appliqués à un même graphe donneront deux polynômes différents, qui ne différeraient que légèrement si les deux fragmentations étaient "proches". On s'attend également à ce qu'un graphe G' plus grand et disposant d'une structure hiérarchique similaire à celle de G donne un polynôme similaire. Visuellement, si on trace les courbes associées pour $q \in [0, 1]$, elles doivent être proches et avoir un comportement similaire.

4. Exemples et applications

Nous nous intéressons ici à l'utilisation du critère MQ multi-niveaux pour comparer différentes fragmentations hiérarchiques. Cette mesure est une mesure de qualité, elle ne constitue donc pas un critère absolu de sélection mais plutôt une aide à la décision. Une manière directe de présenter les résultats obtenus pour différentes fragmentations est de tracer les courbes associées aux polynômes MQ sur un même graphique, c'est d'ailleurs la méthode qui sera employée dans les exemples ci-dessous. Cette visualisation permet de comparer les mesures en ne perdant pas d'informations.

On pourra cependant reprocher à cette méthode qu'elle ne peut pas être utilisée au sein d'un algorithme de fragmentation à maximisation de critère. Dans ce cadre, nous proposons de comparer les intégrales en q des mesures sur l'intervalle $[0, 1]$.

4.1. Juger la qualité d'un découpage sur des niveaux supplémentaires

Le critère MQ peut être utilisé pour déterminer si le découpage d'un fragment à un niveau donné p est opportun. Notons que les coefficients de q de puissance inférieure ou égale à $p-1$ ne dépendent pas de la hiérarchie existante au niveau p . Deux configurations qui partagent la même hiérarchie jusqu'au niveau p auront leurs $p-1$ premiers coefficients de MQ égaux. On se ramène dans ce cadre à une comparaison locale. L'intérêt de poursuivre la fragmentation au niveau $p+1$ sera représenté par la valeur du coefficient de q de puissance p .

Pour un arbre de fragmentation T , on notera $[C_i \dots C_j]$ un sous-arbre de l'arbre se terminant par les feuilles correspondant aux fragments notés $(C_i \dots C_j)$ dont la racine est l'union de ces fragments, à savoir $C_{i \cup \dots \cup j}$.

Considérons, pour illustrer notre propos, un exemple très simple : la figure [2] présente deux configurations menant à un découpage de trois groupes (C_1, C_2, C_3) partitionnant les sommets d'un graphe G non-orienté. Notons que les arbres $T = [C_{1 \cup 2}, C_3]$ et $T' = [[C_1, C_2], C_3]$ partagent le même premier niveau. Dans ce cadre, comparer la qualité de ces deux configurations revient à évaluer la qualité de la séparation de $C_{1 \cup 2}$ dans T' .

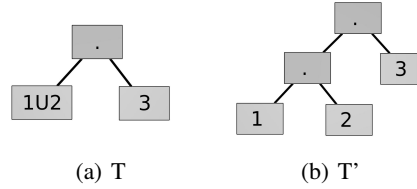


Figure 2. Deux différentes fragmentations d'un graphe formé de trois cliques.

Considérons le cas extrême où (C_1, C_2, C_3) sont des cliques de taille n et qu'il existe bn^2 , avec $0 \leq b \leq 1$, arêtes entre C_1 et C_2 et que C_1 et C_2 sont déconnectés de C_3 . Si on prend comme pondération la taille des groupes, on a alors

$$MQ(T) = \frac{1}{3} \left(1 + \frac{2n(1+b) - 2}{2n-1} \right)$$

$$MQ(T') = \frac{1}{3} \left(1 + \frac{2n(1+b) - 2}{2n-1} [1 + q(1-b)] \right)$$

Pour $q = 0$, on aura $MQ(T) = MQ(T')$. La comparaison repose bien sur la valeur du coefficient $[MQ(T'), q]$. Remarquons que ce dernier est fonction décroissante de b . Cela illustre un phénomène assez évident : plus le groupe $C_{1 \cup 2}$ est dense, moins il est intéressant de le séparer. Pour $b = 1$, la pente de $MQ(T')$ est d'ailleurs nulle.

4.2. Comparaison de fragmentations sur un exemple générique

Prenons des graphes qui impliquent une fragmentation évidente, nous montrons comment la mesure MQ peut indiquer que cette fragmentation est optimale. On dispose de quatre cliques (C_1, C_2, C_3, C_4) de même taille n . On considère que les fragments C_1 et C_2 sont complètement déconnectés de C_4 . Nous allons comparer les quatre configurations suivantes en modifiant les différents ratios de connectivité (notés $(b_{i,j})_{1 \leq i < j \leq 4}$) entre les cliques :

Arbre plat : $[C_1, C_2, C_3, C_4]$ **Arbre parfait** : $[[C_1, C_2], [C_3, C_4]]$
Arbre 3-2 : $[[C_1, C_2], C_3, C_4]$ **Arbre linéaire** : $[[[C_1, C_2], C_3], C_4]$

Les résultats sont présentés dans la figure [3], voici les conclusions que l'on peut tirer de cette analyse :

– Comme on pouvait s'y attendre, des valeurs égales des $b_{i,j}$ font que l'arbre plat semble être le meilleur choix. Cette configuration n'est pas optimale quel que soit $q \in [0, 1]$ mais elle est clairement préférable aux autres configurations si on considère l'intégrale des mesures.

– Si $b_{1,2}$ est bien plus grand que les autres $b_{i,j}$, l'arbre 3-2 est la meilleure fragmentation.

- Si on prend $b_{1,2}$ et $b_{3,4}$ plus grands que les autres $b_{i,j}$ alors le maximum de MQ est donné par l'arbre parfait.
- Si $b_{1,2} \gg b_{1,3} \simeq b_{2,3} \gg$ autres $b_{i,j}$, la configuration linéaire est la meilleure.

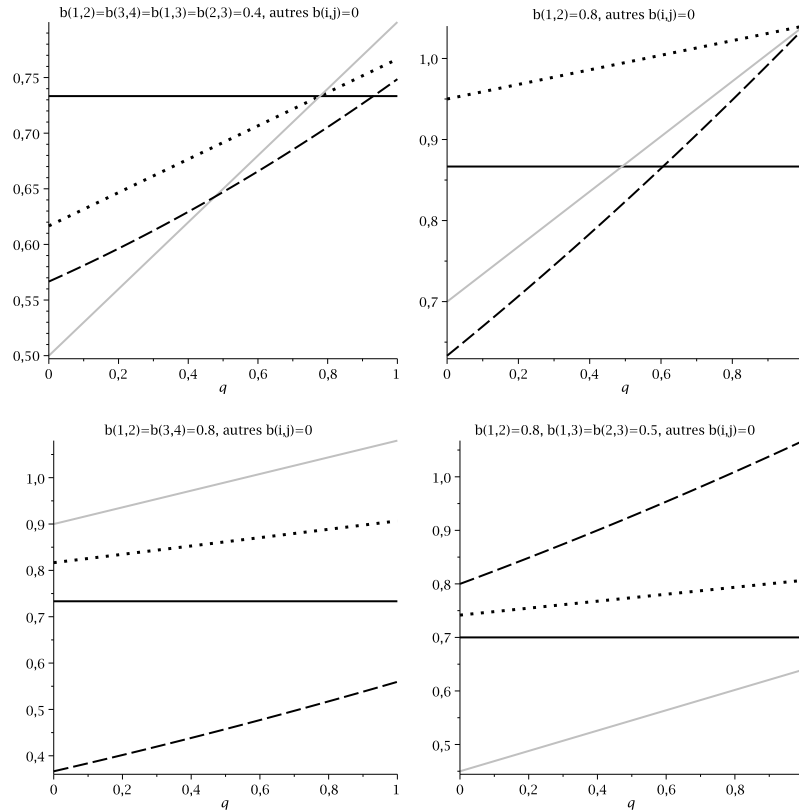


Figure 3. Courbes de MQ pour les arbres Plat (Noir), 3-2 (Pointillés), Parfait (Gris) et Linéaire (Tirets)

4.3. Application : Étude de différentes fragmentations du graphe NCAA

Nous allons ici nous intéresser à l'utilisation de la mesure MQ pour comparer différentes fragmentations appliquées à un cas réel.

Le réseau dont on dispose est une représentation de la division 1A de la ligue universitaire de football américain (NCAA Football) lors de la saison 2000 [GIR 02]. Chaque sommet est associé à une équipe universitaire, une arête est présente entre deux sommets si les deux équipes aux extrémités ont joué ensemble lors de la saison régulière. Cette division rassemble 115 équipes avec un total de 613 rencontres. Les équipes

se répartissent en 11 conférences comprenant chacune entre 5 et 13 équipes (on ne considère pas ici les équipes de la conférence "indépendants").

Durant une saison, les équipes jouent en théorie plus de matchs avec d'autres équipes issues leurs conférences. Toutefois, la proximité géographique influe de façon importante sur les rencontres, les équipes proches ayant plus de chance de s'affronter même si elles n'appartiennent pas à la même conférence. Nous allons comparer trois configurations. Les courbes de MQ associées sont disponibles en figure [4].

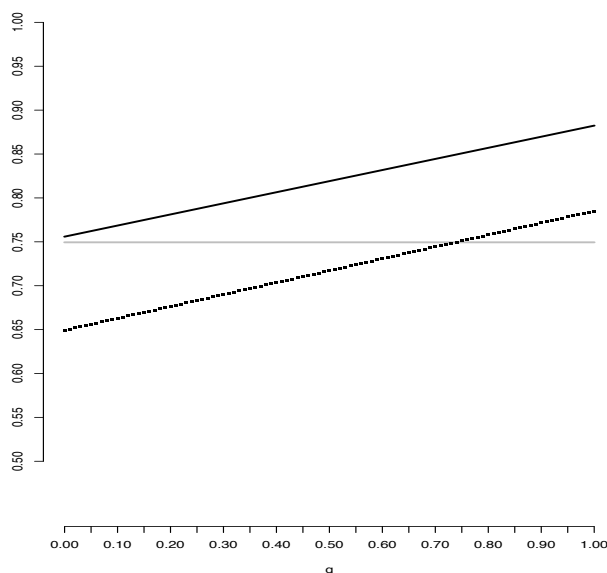


Figure 4. Courbes de MQ associés à trois fragmentations de la Division I-A de la NCAA Football : Découpage en conférence (Gris), Algorithme agglomératif de Newman (Pointillées) et Strength Clustering (Noir).

Les détails des algorithmes ne seront pas rappelés ici. Pour obtenir des fragmentations hiérarchiques, nous avons simplement appliqué récursivement les procédures aux fragments issus de l'itération précédente.

Un partitionnement similaire au découpage en conférences se retrouve dans les configurations résultantes des deux algorithmes utilisés. Ces dernières ajoutent cependant une dimension géographique à ce découpage en regroupant des ensembles d'équipes appartenant au même État ou la même région. Ce comportement se retrouve dans le premier niveau de la fragmentation issue d'un algorithme agglomératif de Newman[NEW 04] et au second niveau de la fragmentation issue d'un algorithme de *Strength Clustering*[AUB 03].

Les pentes des courbes de MQ pour les deux configurations obtenues algorithmiquement indiquent qu'une structure hiérarchique est bien présente au sein de ce réseau. La fragmentation issue du *Strength Clustering* semble être la meilleure. Nous pouvons donc en conclure que la composante géographique doit venir compléter un découpage en conférence et non pas l'inverse.

5. Travail futur

Nous avons donc présenté une mesure de la qualité d'un découpage d'un graphe en modules (fragments) sur plusieurs niveaux, qui prend effectivement en compte l'incidence du découpage d'un module sur le découpage des autres modules et sous-modules du graphe. Sa formulation et son calcul s'appuient sur une combinatoire des chemins de la hiérarchie décrivant l'emboîtement des modules. Des cas d'école simples mais réalistes viennent confirmer la pertinence de l'étude et de l'utilisation d'un tel critère.

Une piste évidente à creuser consiste à proposer une stratégie pour calculer une hiérarchie T optimisant $MQ(G, T, q)$. Ce critère de qualité est cependant déjà utile, par exemple, pour confirmer la pertinence d'un découpage d'un graphe sur plusieurs niveaux. En effet, en comparant les différentes courbes obtenues entre plusieurs fragments hiérarchiques, l'on est à même de choisir le meilleur découpage multi-niveaux (au vu du critère MQ). Cela permet en particulier de déterminer s'il est opportun de fragmenter plus avant certains modules, en mesurant le gain apporté à la valeur MQ . Cette approche objective a été mise en oeuvre de manière probante sur nombre d'exemples (que nous n'avons pas pu présenter ici) qui ont fait l'objet de travaux antérieurs [AUB 03, DEL 04, AMI 05].

Par ailleurs, une étude des propriétés statistiques de la mesure MQ à un niveau a été proposée par [DEL 06] en se basant sur une hypothèse de distribution uniforme des arêtes dans un graphe complet. Une généralisation possible passe par l'utilisation d'autres modèles aléatoires. La détermination de la loi de statistique de $MQ(G, C)$, pour une fragmentation C fixée, permettra l'évaluation quantitative de l'efficacité des algorithmes de fragmentation ou la comparaison de valeurs de MQ normalisées.

6. Bibliographie

- [AMI 05] AMIEL M., MELANCON G., ROZENBLAT C., Réseaux multi-niveaux : l'exemple des changes aériens mondiaux , *M@ppemonde*, vol. 79, n° 3-2005, 2005.
- [AUB 03] AUBER D., CHIRICOTA Y., MELANCON G., JOURDAN F., Multiscale navigation of Small World Networks , *IEEE Symposium on Information Visualisation*, Seattle, GA, USA, 2003, IEEE Computer Science Press, p. 75-81.
- [AUB 07] AUBER D., CHIRICOTA Y., DELEST M., DOMENGER J.-P., MARY P., MELANCON G., Visualisation de graphes avec Tulip : exploration interactive de grandes masses de

- données en appui à la fouille de données et à l'extraction de connaissances , NOIRHOMME-FRAITURE M., VENTURINI G., Eds., *Revue des Nouvelles Technologies de l'Information (actes EGC 2007)*, vol. 1, p. 147-156, Cépaduès, 2007.
- [BLO 08] BLONDEL V. D., GUILLAUME J.-L., LAMBIOTTE R., LEFEBVRE E., Fast unfolding of communities in large networks , *Journal of Statistical Mechanics : Theory and Experiment*, vol. 2008, 2008, page P10008.
- [BRA 08] BRANDES U., GAERTLER M., WAGNER D., Engineering graph clustering : Models and experimental evaluation , *Journal of Experimental Algorithmics (JEA)*, vol. 12, 2008, p. 1-1, ACM.
- [BUR 05] BURT R. S., *Brokerage and Closure*, Oxford University Press, 2005.
- [CLA 04] CLAUSET A., NEWMAN M. E. J., MOORE C., Finding community structure in very large networks , *Physics Reviews E*, vol. 70, n° 6, 2004, page 66111, APS.
- [DEL 92] DELEST M., FÉDOU J.-M., Attribute grammars are useful for combinatorics , *Theoretical Computer Science*, vol. 98, n° 1, 1992, p. 65-76.
- [DEL 04] DELEST M., MUNZNER T., AUBER D., DOMENGER J.-P., Exploring InfoVis Publication History with Tulip (2nd place - InfoVis Contest) , *IEEE Symposium on Information Visualization*, Venice, Italy, 2004, page 110.
- [DEL 06] DELEST M., FÉDOU J.-M., MELANCON G., A Quality Measure for Multi-Level Community Structure , *SYNASC 2006 8th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, Timisoara, Romania, 2006, p. 63-68.
- [FRE 04] FREEMAN L. C., *The Development of Social Network Analysis : A Study in the Sociology of Science*, Empirical Press, Vancouver, 2004.
- [GIR 02] GIRVAN M., NEWMAN M., Community structure in social and biological networks , *Proceedings of the National Academy of Sciences*, vol. 99, n° 12, 2002, page 7821, National Acad Sciences.
- [HAM 04] VAN HAM F., VAN WIJK J. J., Interactive Visualization of Small World Graphs , *IEEE Symposium on Information Visualisation*, IEEE Computer Science press, 2004, p. 199-206.
- [LAT 08] LATAPY M., MAGNIEN C., Complex Network Measurements : Estimating the Relevance of Observed Properties , SOCIETY I. C., Ed., *IEEE INFOCOM*, 2008, p. 1660-1668.
- [MAG 10] MAGNIEN C., Intégrer mesure, métrologie et analyse pour l'étude des graphes de terrain dynamiques , Mémoire d'habilitation à diriger des recherches, Université Piette et Marie Curie / LIP6, Juillet 2010.
- [MAN 98] MANCORIDIS S., MITCHELL B., RORRES C., CHEN Y., GANSNER E., Using automatic clustering to produce high-level system organizations of source code , *Proceedings of the 6th International Workshop on Program Comprehension*, Citeseer, 1998, page 45.
- [MIS 03] MISHNA M., Attribute grammars and automatic complexity analysis , *Advances in Applied Mathematics*, vol. 30, n° 1-2, 2003, p. 189-207.
- [NEW 04] NEWMAN M., Fast algorithm for detecting community structure in networks , *Physics Reviews E*, vol. 69, 2004, page 066133.
- [NEW 06] NEWMAN M. E. J., Modularity and community structure in networks , *Proceedings of the National Academy of Sciences, USA*, vol. 103, 2006, p. 8577-8582.