



Visualisation de données relationnelles

Thomas Guyet

► **To cite this version:**

Thomas Guyet. Visualisation de données relationnelles. Conférence Internationale de Géomatique et d'Analyse Spatiale (SAGEO), Sep 2013, Brest, France. hal-00916923

HAL Id: hal-00916923

<https://hal.inria.fr/hal-00916923>

Submitted on 10 Dec 2013

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.

Visualisation de données relationnelles

Intégration des graphes spatiaux dans QGis

Thomas Guyet

Laboratoire d'Informatique, AGROCAMPUS-OUEST/IRISA UMR 6074
65 rue de Saint-Brieuc, F-35042 Rennes, France
thomas.guyet@agrocampus-ouest.fr

RESUME.

Cet article présente l'implémentation d'une extension QGis pour la visualisation et la construction interactive de graphes spatiaux. Les graphes spatiaux sont des représentations de l'information spatiale sous la forme d'objets spatiaux reliés entre eux par des relations (spatiales ou non). Cette représentation est adaptée à la modélisation et à l'analyse d'information spatiale par des traitements informatiques (fouille de données, recherche de plus court chemins, etc). L'utilisation des graphes spatiaux pâtit de l'absence d'outils facilitant la construction et la visualisation intégrée. Dans cet article, on présente un extension QGis introduisant un nouveau type de couche, les GraphLayer. Ces nouvelles couches peuvent être intégrées dans les projets SIG. Elles offrent des fonctionnalités riches de visualisation et d'édition interactive.

ABSTRACT. This article presents the implementation of a QGis plugin for the visualization and the interactive construction of spacial graphs. Spatial graphs are accurate representations of spatial information through spatial objects linked by relationships (spatial or not). This representation is suited to the modeling and analysis of spatial information by computer processing (data mining, search for shortest paths, etc.). The use of spatial graph suffers from the lack of tools to facilitate the construction and integrated visualization. In this paper, we present a QGis plugin introducing a new type of layer: GraphLayer. These new layers can be integrated into any GIS projects. They offer rich functionality for visualization and interactive editing.

MOTS-CLES : Graphes spatiaux, information relationnelle, analyse visuelle , couche de graphe

KEYWORDS: Spatial graphs, relational information, visual analysis, graph layer

1. Introduction

La représentation de données est un élément important pour leur analyse (Shneiderman, 2001). Là où une mauvaise représentation noie l'information dans la masse des données, une représentation adéquate facilite l'identification d'éléments intéressants et rend possible leur interprétation. Ce problème de la représentation des données se pose aussi bien pour leur analyse automatique (par des algorithmes d'extraction de connaissance) que pour une analyse par un expert qui visualise les données (visual data mining). Dans le premier cas, il s'agit d'une question de représentation informatique de données, dans le second cas, il s'agit d'une question de représentation visuelle. L'analyse visuelle des données est facilitée par la capacité des outils de visualisation à exprimer le maximum de contenu informationnel des données.

Les données spatiales se prêtent facilement à la visualisation et les systèmes d'information géographiques (SIG) offrent de grandes capacités de visualisation de l'information spatiale (Dodge *et al.*, 2008). Les questions relatives à la représentation des données font ainsi partie intégrante des problématiques de la géomatique (Miller, 2003). Le problème le plus largement discuté pour la représentation visuelle d'une carte est celui de la symbologie, c'est-à-dire le choix du mode de représentation visuelle des entités de la carte. Les évolutions technologiques ont conduit récemment à poser de nouvelles questions de représentation d'information dans une carte : la représentation de l'information dynamique (évoluant dans le temps) (Harrower et Fabrikant, 2008) ou encore la représentation de l'information en 3D (Zlatanova *et al.*, 2004).

La représentation de l'information relationnelle dans les données spatialisées reste, quant à elle, assez peu étudiée dans le domaine des SIG. On entend ici par information relationnelle, des informations qui mettent en relation des entités (spatialisées ou non). Ces relations peuvent être de différentes natures : relations topologiques, fonctionnelles ou logiques (Guyet, 2010). La représentation pose à la fois la question du codage informatique d'une information relationnelle entre entités spatialisées et sa visualisation (représentation visuelle) dans un SIG. Actuellement, les seules possibilités de visualisation de l'information relationnelle consistent à détourner l'usage normal des outils de visualisation de SIG (représentation de données vectorielles). L'analyse visuelle de ce type de données est limitée par l'absence d'outil dédié à leur visualisation. Il nous apparaît alors essentiel d'intégrer cette nouvelle possibilité dans les SIG actuels.

On s'intéresse dans cet article à visualiser l'information relationnelle dans un SIG en s'appuyant sur des représentations existantes de l'information relationnelle. On utilise une représentation de l'information relationnelle sous la forme d'un graphe et

on présente une extension de QGis¹ pour la construction et la visualisation de graphes spatiaux. L'objectif de ce travail est d'introduire un nouveau type de couche de données, les couches de graphes spatiaux, spécifiquement dédié à la visualisation d'informations relationnelles dans un projet SIG. Ce type de couche ajoute des capacités de visualisation, d'analyse automatique et d'analyse visuelle de l'information relationnelle.

2. État de l'art de la représentation de l'information relationnelle dans les SIG

Dans le domaine de la géographie et de l'environnement les entités spatiales ne peuvent être vues comme indépendantes, il est primordiale d'appréhender l'espace étudié dans son ensemble et donc de prendre en compte les relations qui lient les entités qui composent l'espace.

Les modèles proposés par (Pech, 2005) et (Guyet, 2010) sont basés sur une représentation de l'information spatiale à l'aide de trois sortes de relations binaires : les relations topologiques, fonctionnelles ou logiques. Ces deux modèles sont orientés vers des tâches de fouille de données. Les relations fonctionnelles incluent des relations spécifiques à un domaine d'application. Il peut s'agir de relations liées à la géographie comme des relations de transfert hydrologiques entre parcelles agricoles, ou bien de relations plus abstraites comme des relations d'appartenance à une exploitation agricole (une entité pouvant représenter une exploitation agricole). (Degenne *et al.*, 2010) introduit Ocelet, un outil pour la définition et la représentation d'informations relationnelles (dynamique) en environnement. Le formalisme de représentation d'Ocelet est plus riche que les approches précédentes. Les relations ne se limitent pas aux relations binaires. De part cette expressivité, Ocelet est orienté par la représentation des connaissance plutôt que par l'analyse des données représentée.

Les relations topologiques ont particulièrement été étudiées. D'un point de vue algébrique, les formalisme de représentation des relations spatiales qualitatives tels que le RCC8 (Region Connection Calculus) propose des outils de représentation et de raisonnement sur les relations topologiques (Renz, 2002). Le domaine des ontologies s'est également particulièrement intéressé à l'organisation des relations topologiques. (Katz et Grau, 2005) ont ainsi traduit la RCC8 dans le formalisme OWL en proposant OWL-DL. Le système ONTOAST (Miron *et al.*, 2007) a étendu le système AROM-ST pour représenter les relations topologiques mais également des informations d'orientation et de distance. L'Open Geospatial Consortium (OGC) a défini GeoSPARQL, une ontologie des relations topologiques des relations.

¹ QGis est un Système d'Information Géographique *opensource* largement utilisé, voir <http://www.qgis.org/>

Dans notre approche, on utilise les graphes comme support de la représentation de l'information relationnelle. Les formalismes qui utilisent des graphes sont contraints par la représentation exclusive de relations binaires. Toutes les relations du RCC8 sont des relations binaires et peuvent donc être représentées par un graphe. Les relations d'ordre supérieurs ne peuvent être représentées, mais sont peu utilisées. L'utilisation de graphes spatiaux est principalement orientée vers la représentation de réseaux (transports, électriques, etc.), de relations d'interactions (pour modéliser des informations en économie) ou de flux (flux migratoires, flux hydrologiques, etc.).

Peu de travaux s'intéressent à la visualisation de graphes spatiaux. (Pech, 2005) visualise ses graphes au moyen de l'outil GraphViz (Ellson *et al.*, 2003) spécialisé dans la représentation de graphes, mais ne permettant pas de représenter l'information spatiale tel que les SIG. Ocelet offre la possibilité de visualiser ses informations au sein de Google Earth à l'aide du format KML non-spécifique aux graphes. Des outils de visualisation tels que MBone, Constellation ou FlowMap sont utilisés pour représenter des graphes de flux (Munzner, 2000). À notre connaissance, il n'existe pas d'outils intégré à des SIG pour l'affichage générique de graphes. ArcGis comprend une extension pour l'analyse de réseaux de transports avec des fonctionnalités de recherche de plus court chemin. L'extension Roadgraph de QGis est également disponible pour le même type de fonctionnalité. Ce type de représentation ne nous semble pas approprié pour toutes les informations relationnelles. Notre objectif a donc été de proposer l'intégration de la visualisation de l'information relationnelle par des graphes spatiaux dans QGis.

3. Extension QGis pour la visualisation de graphes

3.1. Présentation générale

L'extension QGis pour la visualisation introduit un nouveau type de couche spécifique aux graphes. Chaque couche correspond à un fichier décrivant un graphe et à une modalité d'affichage (*cf.* section 3.2). Une couche ne contient qu'un seul graphe, mais ce graphe n'est pas nécessairement entièrement connecté.

Une couche de graphe peut être sauvegardée et chargée à l'intérieur d'un projet. On fournit également des outils pour l'édition de graphes (*cf.* section 3.3). Ces fonctions d'éditions permettent de créer un graphe directement dans QGis et de le modifier.

Le modèle d'un graphe est proche de celui proposé par Pech (Pech, 2005) et est illustré dans la Figure 1, les éléments principaux caractérisant un graphe sont les suivants :

- un graphe est défini par un ensemble de nœuds, d'arcs et il définit les attributs qui vont décrire les nœuds et les arcs, ainsi que le système de coordonnées dans

lequel sont exprimés les coordonnées du graphe. Le graphe dispose également d'une liste d'attributs/valeurs qui donnent des informations à l'échelle du graphe ;

- un nœud (node) est décrit par une liste de valeurs qui correspondent aux attributs qui le caractérisent. Un nœud est localisé dans l'espace, il précise les coordonnées auxquelles il doit être positionné ;

- un arc (edge) est décrit par ses deux nœuds d'extrémités (source et destination). Un arc est donc implicitement dirigé, mais il peut être considéré comme non-dirigé. Comme les nœuds, un arc porte également des valeurs pour les attributs qui le décrivent ;

- un attribut qui caractérise un nœud ou un arc peut être d'un des trois types suivants : des nombres entiers, des nombres réels et des chaînes de caractères. L'information relationnelle modélisée peut donc être qualitative ou quantitative.

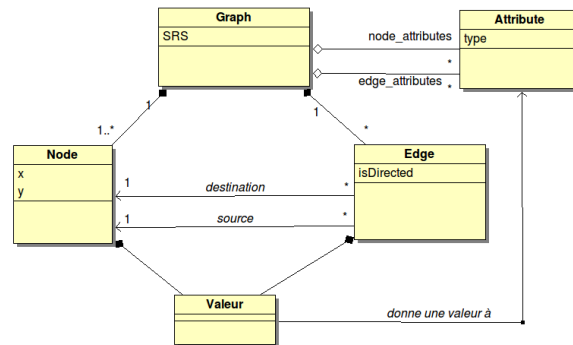


Figure 1. Schéma UML illustrant le modèle d'un graphe

3.2. Fonctionnalités de visualisation

Un graphe est visualisé spatialement par l'affichage des nœuds et des arcs. La visualisation des nœuds et des arcs peut être configurée pour répondre aux besoins de la représentation de l'information portée sur le graphe. Les caractéristiques géométriques suivantes peuvent être adaptés :

- caractéristiques des nœuds : taille d'affichage, largeur de la bordure, couleur de bord et couleur de fond

- caractéristiques des arcs : largeur de trait, couleur de trait, style de ligne.

Les fonctionnalités de visualisation permettent de définir chacune de ces propriétés pour l'ensemble des nœuds/arcs ou pour chaque élément individuel en fonction d'une valeur portée par celui-ci. La Figure 2 illustre l'édition de la propriété de couleur de fond d'un nœud. La couleur dépend ici de la valeur de l'attribut *cl*.

Cette répartition de couleur a été obtenue automatiquement par classification des valeurs.

Le mécanisme de gestion de l'affichage permet de configurer plusieurs propriétés de visualisation en même temps. On peut, par exemple, faire modifier la taille d'un nœud en fonction d'un premier attribut et sa couleur en fonction d'un autre.

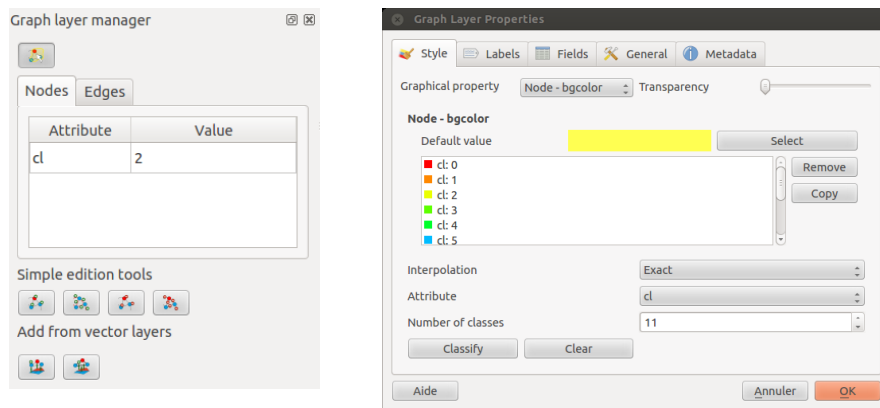


Figure 2. Sur la gauche, boîte à outils pour la consultation et l'édition d'une couche de graphe. Sur la droite, fenêtre de configuration de l'affichage d'un graphe. La fenêtre illustre la configuration de la couleur de fond d'un nœud à partir de l'attribut cl.

3.3. Fonctionnalités d'édition

L'édition des couches de graphes comporte deux types d'opérations :

- les opérations de modification manuelle des éléments du graphe : ajout/suppression d'un nœud, ajout/suppression d'un arc entre nœuds, modification des valeurs des attributs.

- les opérations de construction d'un graphe à partir de données provenant de couches vectorielles. Ces opérations permettent, d'une part, de construire des nœuds à partir d'une couche vectorielle en utilisant les centroïdes des géométries qui la composent et en récupérant les attributs. D'autre part, une opération similaire permet de récupérer les attributs d'une couche vecteur lors de la création d'arcs entre deux nœuds.

4. Exemples d'utilisation

On illustre maintenant la construction et la visualisation des graphes obtenus dans QGis sur deux exemples de cartes.

La carte de la Figure 3 représente un graphe de proximité entre les parcelles cadastrales dans un environnement urbain. Un arc indique une distance entre deux centroïdes de parcelles cadastrales inférieures à 100m. En milieu urbain, le nombre de parcelles répondant à ce critère de proximité peut être très important. Pour rendre le graphe plus lisible, nous avons construit uniquement les 5 arcs les plus proches. Cette couche de graphe a été obtenue intégralement par l'utilisation des fonctionnalités automatiques de notre extension QGis.

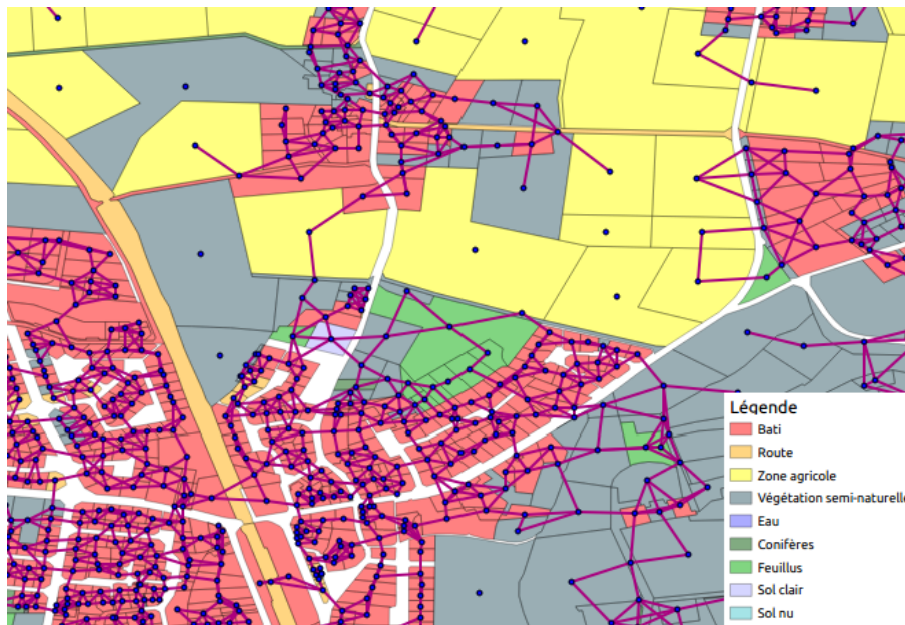


Figure 3. Illustration d'un graphe de proximité entre parcelles cadastrales. Chaque nœud représente une parcelle. Elle est connectée à au plus 5 parcelles distantes de moins de 100m.

Dans la Figure 4, nous illustrons les propriétés d'affichage de notre extension en représentant un graphe de connectivité entre les parcelles agricoles : un arc est indiqué lorsque deux parcelles sont connexes. Le graphe ne concerne que les parcelles agricoles dont l'information est portée par les nœuds. Les arcs représentent

l'interface entre les parcelles connectées : type de haie ou absence de particularité (en bleu).

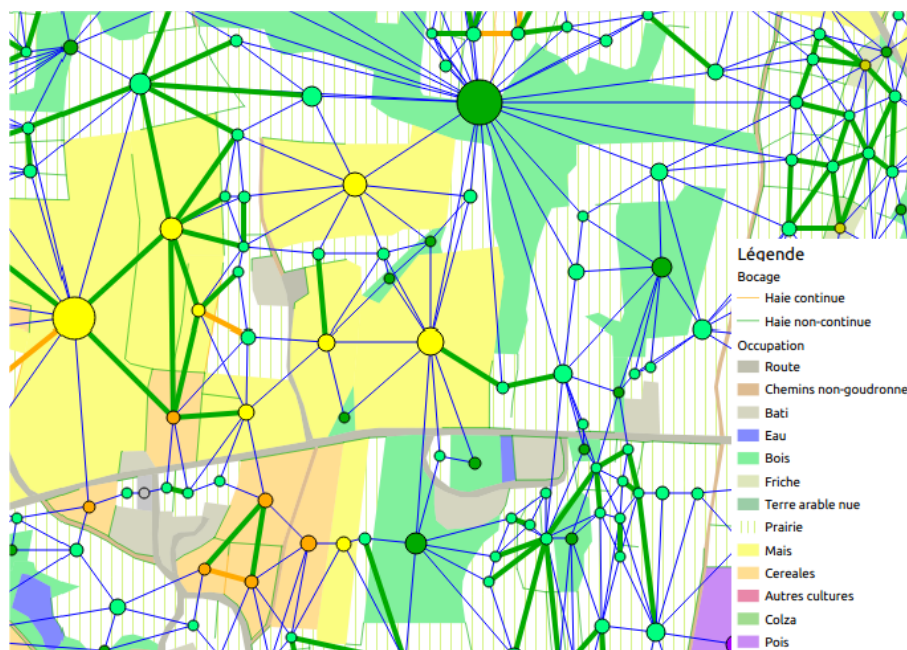


Figure 4. Graphe de connectivité entre parcelles agricoles. La taille des nœuds est proportionnelle aux tailles de parcelle. La couleur des nœuds correspond à l'occupation du sol. La couleur des arcs indique le type d'interface.

5. Perspectives et conclusion

Dans ce papier, nous avons présenté un outil de visualisation de graphes spatiaux intégré dans QGIS sous la forme d'un nouveau type de couche. L'intérêt de ce nouveau type de couche est de représenter une information relationnelle entre les entités spatiales.

La version du logiciel présentée dans cet article, implémentée en C++ sur la version de développement de QGIS (v1.9.0), est une version préliminaire qui pourra être mise à disposition sur demande afin d'en étendre le modèle de données et les outils de génération par rapport aux usages qui peuvent être attendus d'un tel outil. À terme, il offrira également à chacun la possibilité de proposer ses propres traitements de graphes.

Ce travail s'inscrit dans un projet de réalisation d'un outil d'analyse automatique de graphes spatiaux. Dans un premier temps, des outils d'analyse des propriétés du

graphe (statistiques sur les nœuds et les arcs du graphes : distribution des connectivités des longueurs d'arcs, etc.) permettront aux utilisateurs d'extraire des informations sur les caractéristiques générales du graphe. Dans un second temps, des outils de fouille de données s'intéresseront à identifier des structures intéressantes dans le graphe : structures informationnelles, p. ex. SUBDUE (Holder *et al.*, 2002), ou structures fréquentes (Kuramochi et Karypis, 2002) ou bien de classer les graphes selon leurs structures (Horváth *et al.*, 2004). Finalement, des outils d'interrogation riches permettront aux utilisateurs d'explorer de manière interactive les données. L'intégration de ces fonctionnalités dans QGis doit permettre de les rendre plus accessibles aux gestionnaires des données.

Remerciements

L'auteur tient à remercier l'unité SAD-Paysage de l'INRA et le laboratoire COSTEL du CNRS pour le parcellaire agricole et les données de bocage. Les données de zone urbaine ont été acquises par télédétection par le laboratoire Agrocampus-Ouest/INRA SAS dans le cadre du projet PayTal.

Bibliographie

- Degenne P., Ait Lahcen A., Curé O., Forax R., Parigot D., Lo Seen D., (2010) Modelling the environment using graphs with behaviour: do you speak Ocelet?, International Congress on Environmental Modelling and Software, pp. 8.
- Dodge M., McDerby M., Turner M. (2008) The Power of Geovisualization. In: Dodge, M., McDerby, M., and Turner, M., (eds.), Geographic Visualization: Concepts, Tool, and Applications, Wiley, pp. 1-10.
- Ellson J., Gansner E., Koutsofios E., North E., Woodhull G. (2003) Graphviz and dynagraph – static and dynamic graph drawing tools, Graph Drawing Software, Springer-Verlag, pp. 127-148.
- Guyet T (2010) Fouille de données spatiales pour la caractérisation spatiale de paysages en lien avec des fonctionnalités agro-écologiques, Spatial Analysis and GEOmatics, pp. 3.
- Harrower M., Fabrikant S. I. (2008) The Role of Map Animation in Geographic Visualization. In: Dodge, M., Turner M., and McDerby, M., (eds.), Geographic Visualization: Concepts, Tools and Applications, Wiley, pp. 49-65.
- Holder, L., Cook, D., Gonzalez, J., Jonyer, I. (2002). Structural pattern recognition in graphs. journal of Pattern Recognition and String Matching, vol. 13, pp. 93.
- Horváth, T., Gärtner, T., & Wrobel, S. (2004). Cyclic pattern kernels for predictive graph mining. In proceedings of the conference on Knowledge discovery and data mining, pp. 158-167.
- Katz Y., Grau B. (2005), Representing Qualitative Spatial Information in OWL-DL, Proceedings of OWL: Experiences and Directions.

- Kuramochi, M., Karypis, G. (2002). Discovering frequent geometric subgraphs. In proceedings of International Conference on Data Mining, pp. 258-265.
- Miller H., Wentz E. (2003), Representation and Spatial Analysis in Geographic Information Systems, *Annals of the Association of American Geographers*, 93(3):574–594
- Miron A. D., Gensel J., Villanova-Oliver M., Martin H. (2007) Towards the Geo-spatial Querying of the Semantic Web with ONTOAST, *Actes de W2GIS*, pp. 121-136.
- Munzner T. (2000) *Iterative visualization of large graphes and networks*. Thèse de doctorat. Stanford University.
- Pech Palacio, M. (2005) *Spatial Data Modeling and Mining using a Graph-based Representation*, These de Doctorat, INSA Lyon.
- Renz, J. (2002) *Qualitative Spatial Reasoning with Topological Information*, LNCS, Vol. 2293, Springer-Verlag.
- Shneiderman B. (2001) *Inventing Discovery Tools: Combining Information Visualization with Data Mining*, *Information Visualization(2001)*, In *Discovery Science*. Vol. 2226, pp. 17-28.
- Zlatanova S., Rahman A. , Shi W. (2004) *Topological models and frameworks for 3D spatial objects*, *Journal of Computers & Geosciences*, Vol. 30, No. 4, pp. 419-428