



**HAL**  
open science

## Analyse du comportement des contributeurs dans l'Information Géographique Volontaire via la construction de réseaux sociaux

Quy Thy Truong, Cyril de Runz, Guillaume Touya

### ► To cite this version:

Quy Thy Truong, Cyril de Runz, Guillaume Touya. Analyse du comportement des contributeurs dans l'Information Géographique Volontaire via la construction de réseaux sociaux. Atelier Gestion et Analyse de données Spatio-Temporelles (GAST@EGC), 2018, Paris, France. pp.45-54. hal-01924268

**HAL Id: hal-01924268**

**<https://hal.science/hal-01924268>**

Submitted on 15 Nov 2018

**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.

# Analyse du comportement des contributeurs dans l'Information Géographique Volontaire via la construction de réseaux sociaux

Quy Thy Truong<sup>\*,\*\*</sup>, Cyril De Runz<sup>\*</sup>,  
Guillaume Touya<sup>\*\*</sup>

<sup>\*</sup>Modeco, CReSTIC, University of Reims Champagne-Ardenne,  
CS 30012, f-51687, Reims cedex 2, France  
quy-thy.truong@ign.fr

<sup>\*\*</sup>Univ. Paris-Est, LASTIG COGIT, IGN, ENSG,  
F-94160 Saint-Mande, France

**Résumé.** Modéliser les interactions sociales au sein de projets de cartographie volontaire et citoyenne nécessite de définir ce qui relie les contributeurs entre eux dans le temps et l'espace. Dans un souci de réalisme, plutôt que d'étudier un seul type de relation, nous choisissons de construire un réseau social multi-couche contenant différents types d'interactions. L'analyse d'un tel multigraphe devrait permettre de détecter des communautés entre les collaborateurs spatio-temporels et de définir des profils de contributeurs typiques. Un cas d'étude sur OpenStreetMap illustre les déductions pouvant être faites sur les contributions à partir de leurs auteurs.

## 1 Introduction

Les contributions issues de la collecte collaborative d'information géographique sont des données spatio-temporelles. En effet, l'Information Géographique Volontaire (VGI) est une représentation spatiale du monde réel faite par un contributeur à un instant donné. Ce mode d'acquisition collaboratif offre la possibilité aux contributeurs de se compléter et se corriger mutuellement, permettant ainsi l'enrichissement et la mise à jour rapide de la base de données. Bien que cet aspect du fonctionnement collaboratif semble favoriser la qualité des données, la liberté accordée à tout contributeur, du plus expert au plus malintentionné, nuance cette affirmation. La qualité des données VGI reste donc une problématique encore étudiée dans la recherche actuelle.

Pour assurer la qualité des données VGI, Goodchild et Li (2012) ont défini une approche qui "repose sur une hiérarchie d'individus de confiance agissant en tant que modérateurs". Cette approche suppose d'analyser les actions de contribution permettant d'identifier les contributeurs fiables. De plus, elle exige de qualifier les réactions du contributeur face aux contributions d'autrui. Ce papier aborde la problématique suivante : comment modéliser les interactions entre contributeurs à partir de leurs activités spatio-temporelles dans le contexte de qualification des données VGI ? Le sujet de

cet article se positionne comme un cas d'utilisation des données spatio-temporelles. Les interactions entre contributeurs sont obtenues à partir des contributions appartenant à une même fenêtre spatio-temporelle et la collaboration est modélisée sous forme d'un réseau social multiplexe.

De premiers résultats sont présentés à partir d'un cas d'étude sur OpenStreetMap (OSM), montrant que, malgré l'hétérogénéité caractéristique des contributeurs VGI et de leurs contributions, la communauté accorde sa confiance à certains contributeurs et par conséquent à leurs contributions.

## 2 Travaux connexes

Considérer les communautés de VGI comme des réseaux sociaux n'est pas une nouveauté. Mooney et Corcoran (2014) et Stein et al. (2015) ont proposé des modèles de graphes sociaux pour représenter les collaborations entre les contributeurs d'OSM. En général, les communautés de crowdsourcing fonctionnent grâce aux réactions des contributeurs sur les contributions des uns et des autres. Cependant, un seul aspect à la fois était considéré pour modéliser la collaboration.

Selon Kivelä et al. (2014), "réduire un système social à un réseau dans lequel les acteurs sont connectés (...) par un seul type de relation est souvent une approximation extrêmement grossière de la réalité. C'est pourquoi, depuis des décennies, les sociologues reconnaissent qu'il est crucial d'étudier les systèmes sociaux par la construction de plusieurs réseaux sociaux en utilisant différents types de liens entre le même ensemble d'individus". Il s'agit donc de relier deux contributeurs par plusieurs relations s'il partagent des liens sociaux d'ordres différents, par exemple un lien familial, amical, professionnel, etc.

Jankowski-Lorek et al. (2016) ont repris cette idée pour étudier l'impact du travail d'équipe dans la qualité des données collaboratives à partir d'un réseau social multidimensionnel de la communauté de Wikipédia. Stein et al. (2015) ont aussi confronté deux types de collaboration (largeur et profondeur de la collaboration) dans le but d'identifier des profils de contributeurs. En suivant la même idée, nous proposons de modéliser les interactions entre contributeurs par un réseau social multiplexe. Un tel réseau permet de relier les même individus par plusieurs relations de types différents. Ainsi, ce graphe donne une vision plus complète de la collaboration entre les contributeurs d'une communauté.

La modélisation proposée dans ce papier n'apporte rien de nouveau par rapport au concept même des graphes multiplexes. Toutefois, nous proposons de tirer profit des ces travaux existants sur les réseaux sociaux pour les mettre au service de l'étude des contributeurs et de la qualité des contributions VGI. En effet, la modélisation multiplexe de graphes sociaux est, à notre connaissance, une solution innovante dans le domaine de l'Information Géographique Volontaire. L'intérêt de cet article est de montrer les différentes collaborations qui ont lieu au sein d'une même fenêtre spatio-temporelle de contributions, et en quoi ces interactions permettent de qualifier le rôle des contributeurs dans leur communauté, au sens de collaborateurs spatio-temporels.

### 3 Modélisation des contributeurs VGI dans un réseau social multiplexe

Nous construisons un type de réseau multi-couche particulier, appelé réseau multiplexe, qui est une séquence de graphes  $\{G_\alpha\}_{\alpha=1}^b = \{(V_\alpha, E_\alpha)\}$  où  $E_\alpha$  est l'ensemble des arcs,  $V_\alpha$  est l'ensemble des sommets tel que  $V_\alpha = V_\beta = V$  pour tout  $\alpha, \beta$ , et  $b$  est le nombre de graphes (Kivelä et al., 2014). Dans notre cas,  $V$  constitue l'ensemble des contributeurs et représente la seule connexion entre tous ces graphes initialement indépendants. Selon les graphes considérés, certains sommets de  $V$  peuvent être isolés. L'enjeu ici est de définir l'ensemble des arcs  $E_\alpha$ . Les propositions qui suivent reprennent ou s'inspirent des modèles d'interactions sociales de la littérature.

Dans la littérature portant sur l'Information Géographique Volontaire, les collaborations ont surtout été considérées selon le point de vue des éditions d'objets. On peut définir  $E_{co\text{-édition}} = \{(A, B, e)\}$  où le contributeur  $A$  a créé la version  $v+1$  de la version  $v$  du même objet produite par le contributeur  $B$ , et  $e$  l'intensité de la co-édition *i.e.* le nombre de fois où  $A$  a directement édité une nouvelle version d'une contribution précédemment éditée par  $B$  (Mooney et Corcoran, 2014). Par ailleurs, l'étude de la co-édition peut s'étendre à celle d'un entrelacement de collaboration (interlocking collaboration), permettant alors de caractériser l'intensité de la collaboration (Stein et al., 2015). Dans ce cas, l'entrelacement de collaboration n'est pas restreinte aux éditions successives entre deux contributeurs. Ainsi, la largeur de collaboration est définie par  $E_{largeur} = \{(A, B, w)\}$  où  $w$  est le nombre d'objets que les contributeurs  $A$  et  $B$  ont édité en commun ; la profondeur de collaboration est définie par l'ensemble  $E_{profondeur} = \{(A, B, d)\}$  où  $d$  est le nombre maximum de fois où le contributeur  $A$  a répondu (directement ou non) au contributeur  $B$  sur le même objet. Au-delà des différences subtiles qui existent entre ces graphes, ces derniers permettent d'estimer si un contributeur tombe d'accord avec la communauté.

Dans (Heaberlin et DeDeo, 2016), un réseau de normes Wikipédia a été défini comme un graphe dont les sommets sont des pages Wikipédia. Deux pages sont connectées lorsque l'une fait référence à l'autre via un lien hypertexte. Réutiliser la contribution d'autrui pour faire sa propre contribution constitue donc un autre type d'interaction. L'ensemble  $E_{réutilisation} = \{(A, B, d)\}$  estime la relation de confiance entre deux utilisateurs  $A$  et  $B$ , en particulier dans le cas où  $d > 0$ , puisqu'un contributeur utilise la contribution d'un autre contributeur si le premier a confiance en le second (c'est du moins notre hypothèse). Contrairement aux graphes précédents, le graphe de réutilisation est une nouvelle proposition puisqu'il provient d'une adaptation du réseau de normes Wikipédia.

Enfin, connecter les utilisateurs ayant contribué dans la même zone permet de révéler des collaborations existantes dans la communauté, surtout s'ils contribuent en même temps. Ainsi, on peut définir l'ensemble  $E_{co\text{-location}} = \{(A, B, l)\}$  qui connecte les contributeurs selon le chevauchement  $l$  de leurs zones de contribution respectives (Neis et Zipf, 2012; Zielstra et al., 2014), et l'ensemble  $E_{co\text{-temporalité}} = \{(A, B, t)\}$  qui décrit le lien temporel séparant un couple de contributeurs. L'agrégation des graphes de co-location et de co-temporalité correspond au modèle spatio-temporel de co-occurrence introduit dans (Crandall et al., 2010). Ce modèle avait été défini pour identifier des liens

sociaux, par conséquent l'analyse simultanée de ces deux graphes devrait permettre d'étudier ce même problème. Notre proposition consiste à diviser en deux graphes distincts le graphe de co-occurrence.

Par rapport aux différents travaux cités, notre modèle reprend et réadapte les graphes proposés dans un même réseau. En représentant ces différentes relations dans même un système, des contributeurs remarquables pourront être identifiés, en particulier après détection de communauté (clustering).

## 4 Cas d'application : analyse des contributeurs d'OpenStreetMap

Pour démontrer l'utilité d'un tel graphe multiplexe, un graphe de trois couches a été réalisé sur un cas d'étude OpenStreetMap (OSM). Ces couches correspondent aux interactions de largeur de collaboration, de profondeur de collaboration et de réutilisation. Ces graphes ont été construits à partir du chargement des contributions sur une fenêtre spatio-temporelle qui s'étend de 2010 à fin 2014 et sur le centre de Paris (île de la Cité et ses alentours). Les graphes de largeur et de profondeur de collaboration sont produits à partir des contributions OSM de type ponctuels (node) tandis que le graphe de réutilisation a été produit à partir des contributions OSM de type linéaires (way), c'est-à-dire que l'on considèrerait uniquement les objets ways composés de nodes édités par un auteur différent. Pour limiter l'affichage aux interactions les plus fortes, les figures suivantes présentent les sommets dont les arcs (entrant ou sortants) sont de poids supérieur à 2. Un filtrage supplémentaire sur les sommets restreint l'affichage des sommets de degré supérieur à 2. Ces graphes sont affichés sur les FIG. 1, 2, 3, et TAB. 1 explicite les paramètres d'affichage des graphes.

Paramètre	Définition
Couleur du sommet	Proportionnel au degré entrant, en suivant les couleurs suivantes : bleu = faible ; blanc = intermédiaire ; rouge = élevé
Taille du sommet	Proportionnel au degré sortant
Couleur de l'arc	Identique à la couleur du sommet initial
Taille de l'arc	Proportionnel au poids <i>i.e.</i> l'intensité de la collaboration

TAB. 1 – *Paramètres de visualisation des graphes. D'après la définition d'un réseau multiplexe, chaque graphe est composé du même ensemble de sommets, mais pour des raisons de lisibilité seuls les arcs de poids supérieur à 2 et les sommets de degré supérieur à 2 sont affichés.*

Dans les graphes de largeur et de profondeur de contribution, on remarque trois contributeurs très actifs (représentés par des gros sommets, c'est-à-dire caractérisé par un fort degré sortant) : #17397, #18855 et #33745. Or, le contributeur #17397 présente un degré entrant beaucoup plus faible que les deux autres : celui-ci a donc été moins édité par le reste de la communauté. De plus, le graphe de réutilisation montre que ses contributions ont été largement réutilisées par les autres contributeurs.

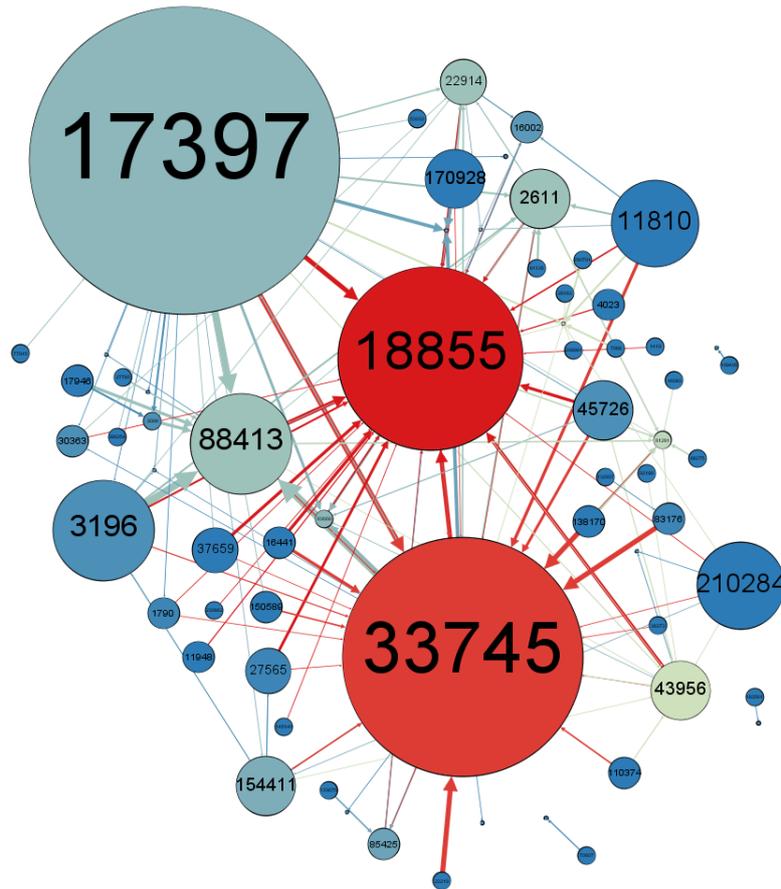


FIG. 1 – *Graphe de largeur de collaboration des contributeurs d'OSM sur Paris (les identifiants sont anonymisés). Un arc dirigé d'un contributeur A vers un contributeur B signifie que A a modifié un objet qui a été précédemment édité par B. L'épaisseur de l'arc correspond au nombre d'objets uniques que le contributeur A a édité après l'utilisateur B.*

Analyse des contributeurs VGI via la construction de réseaux sociaux

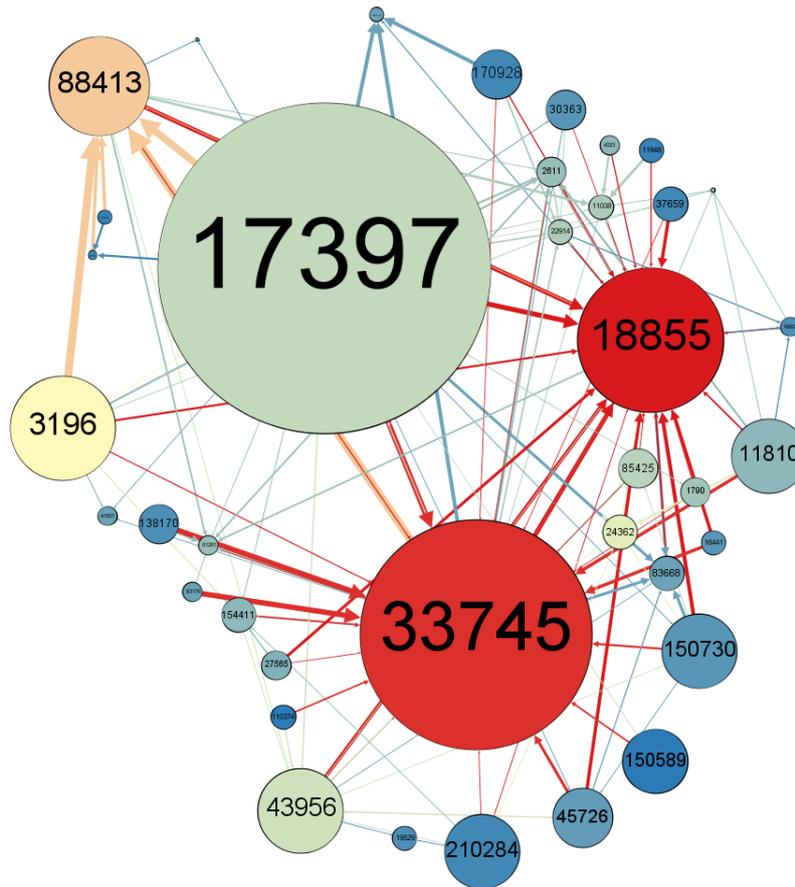


FIG. 2 – Graphe de profondeur de collaboration des contributeurs OSM sur Paris (les identifiants sont anonymisés). Comme le graphe de largeur de collaboration, un arc dirigé d'un contributeur A vers un contributeur B signifie que A a édité un objet précédemment édité par B. L'épaisseur de l'arc correspond au nombre maximal de fois où A a édité B sur un même objet (mais pas nécessairement de manière directe, d'autres contributeurs ont pu éditer l'objet entre A et B).

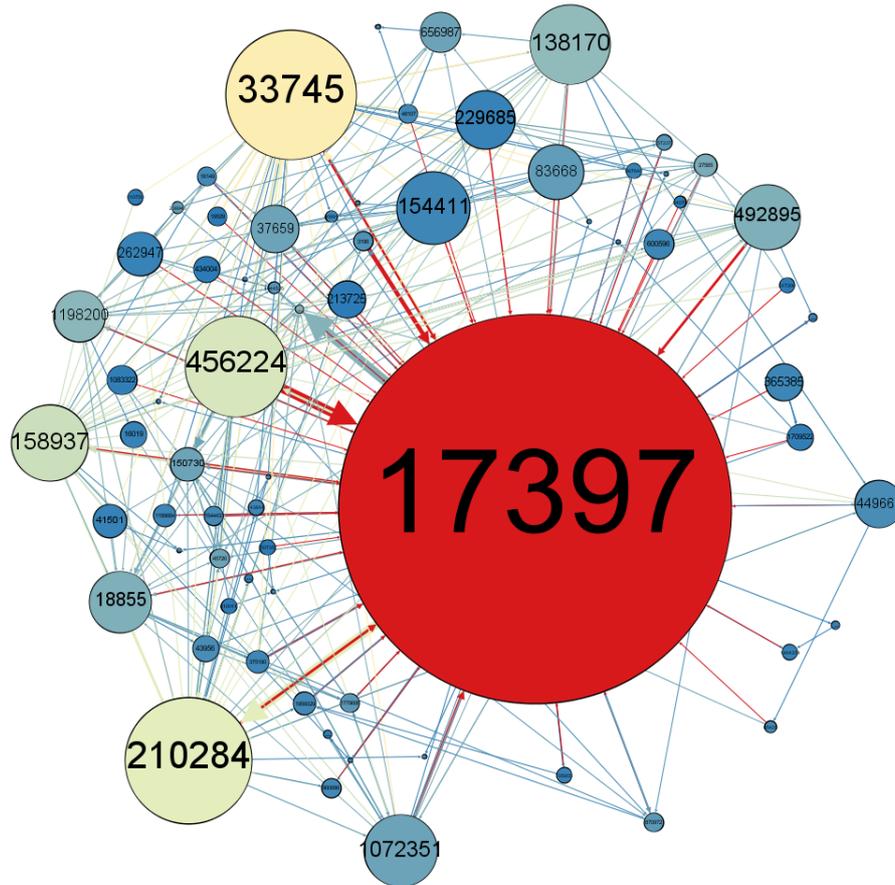


FIG. 3 – Graphe de réutilisation OSM sur Paris (les identifiants des contributeurs sont anonymisés). Un arc dirigé d'un contributeur A vers un contributeur B signifie que A a créé un nouvel objet composé d'une contribution de B. En particulier, dans ce cas d'étude on se restreint à l'utilisation d'objets ponctuels composant des objets linéaires.

La réutilisation de ces données donne du poids à la fiabilité de ces données : en effet, on peut supposer qu'un contributeur A va réutiliser les données d'un autre contributeur B si A considère que les contributions de B sont correctes. Par conséquent, #17397 a un profil proche de celui d'un modérateur, puisqu'il édite beaucoup en étant lui-même peu édité et ses propres contributions sont considérées comme fiables par les autres contributeurs.

Les contributeurs #18855 et #33745 sont édités par de nombreux contributeurs. Les arcs rouges dirigés vers ces deux sommets ont un poids important dans les graphes de largeur et de profondeur de collaboration. Concernant le graphe de largeur de collaboration, cela signifie que ces deux contributeurs ont été édités sur plusieurs objets différents ; par conséquent la largeur de collaboration entre ces deux contributeurs avec le reste de la communauté correspond à une édition qui s'étend spatialement. Quant au graphe de profondeur de collaboration, l'épaisseur d'un arc traduisant le nombre d'interactions maximal avec un contributeur sur un objet donné, on pourra considérer que la profondeur de la collaboration correspond à une interaction qui s'étend dans le temps. Sans même parler de guerre d'éditions, on constate ici que certains contributeurs vont revenir éditer plusieurs fois un même objet après les passages d'un même contributeur. C'est également le cas du contributeur #88413 dont plusieurs arcs entrants de poids relativement important caractérisent une interaction temporellement étendue avec d'autres utilisateurs. On peut donc se poser la question de la fiabilité de ces contributeurs et par conséquent, de leurs contributions. De plus, bien que des graphes de co-localisation et de co-temporalité auraient été pertinents ici pour décrire la proximité spatio-temporelle entre les contributeurs, les graphes de collaboration (en largeur et en profondeur) permettent déjà de mettre en avant l'intensité spatio-temporelle de la collaboration.

Dans le graphe de réutilisation, le contributeur #33745 apparaît en jaune, ce qui montre qu'un certain nombre de ses contributions ont été réutilisées par d'autres utilisateurs. Cela nuance notre analyse et démontre que les contributions de cet utilisateur ne sont pas forcément de mauvaise qualité. Cela démontre aussi que les graphes de largeur et de profondeur de collaboration ne suffisent pas pour qualifier les contributeurs, ce qui justifie ainsi la nécessité d'une analyse conjointe des différents graphes du réseau multiplexe. Le contributeur #18855 apparaît également sur le graphe de réutilisation, mais de manière moins visible. Précisons toutefois que cela ne signifie pas forcément qu'il est moins fiable que le contributeur #33745 : peut-être que ses contributions ne sont simplement pas faites pour être réutilisées dans des objets linéaires. En effet, tous les objets ponctuels n'ont pas vocation à composer des objets linéaires. Par exemple, les routes ne peuvent pas être construites à partir de points d'intérêt.

## 5 Conclusion et perspectives

De la même façon que certains travaux menés dans le domaine de l'analyse des réseaux sociaux, ce papier propose de modéliser le comportement des contributeurs VGI au moyen d'un graphe social multiplexe. Les interactions entre contributeurs sont alors récupérées à partir des différents modes de contributions, permettant de reconstituer de manière plus détaillée les liens sociaux dans le temps et dans l'espace. Il est vrai

qu'un tel réseau social met en lumière les perceptions des contributeurs sur les uns et les autres, mais l'analyse va plus loin en abordant la problématique de la qualité des données VGI. En effet, l'application de ce modèle sur OpenStreetMap a pu mettre en évidence des pistes permettant de qualifier les contributions OSM.

Sans grande surprise, la diversité des comportements permet d'expliquer l'hétérogénéité des données. Pour autant, les contributeurs ayant été considérés comme des personnes de confiance dans la communauté témoignent de la fiabilité de leurs contributions. L'analyse de graphes supplémentaires de manière automatisée permettra, selon nous, de saisir la plupart des modes d'opération des contributeurs. En effet, l'intérêt d'utiliser un réseau multiplexe est la possibilité d'y ajouter d'autres graphes, à condition que l'ensemble des sommets soit égal à  $V$  (rappelons qu'il est possible que des sommets soient isolés dans un graphe donné). Néanmoins, un écueil à éviter serait d'ajouter des graphes sans s'assurer de leur réel apport vis-à-vis des couches présentes dans le réseau multiplexe initial. Par exemple, on peut remarquer l'existence d'un recouvrement entre les graphes de co-édition et les graphes de collaboration (en largeur et en profondeur), puisqu'ils portent tous sur l'activité d'édition des contributions. Il sera judicieux de faire le tri parmi ces graphes.

L'objectif principal de la construction de ces graphes est de faire ressortir une typologie selon laquelle seront classifiés les contributeurs VGI de manière automatique. A cette fin, nos recherches futures porteront d'une part sur la détection de communauté dans les graphes multiplexes y compris des méthodes de partitionnement (*i.e.* méthodes de clustering) (Battiston et al., 2014). D'autre part, la typologie résultante devra être comparée aux enquêtes qualitatives effectuées, comme celles de Coleman et al. (2010) et Duféal et al. (2016).

## Références

- Battiston, F., V. Nicosia, et V. Latora (2014). Structural measures for multiplex networks. *Phys. Rev. E* 89, 032804.
- Coleman, D. J., Y. Geogiadou, et J. Labonte (2010). Volunteered geographic information : The nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research* 4.
- Crandall, D. J., L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, et J. Kleinberg (2010). Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences* 107(52), 22436–22441.
- Duféal, M., C. Jonchères, et M. Noucher (2016). ECCE CARTO - DES ESPACES DE LA CONTRIBUTION A LA CONTRIBUTION SUR L'ESPACE - Profils, pratiques et valeurs d'engagement des contributeurs d'OpenStreetMap (OSM). Research report, UMR 5319.
- Goodchild, M. F. et L. Li (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*.
- Heaberlin, B. et S. DeDeo (2016). The evolution of wikipedia's norm network. *Future Internet* 8(2), 14+.

- Jankowski-Lorek, M., S. Jaroszewicz, Ł. Ostrowski, et A. Wierzbicki (2016). Verifying social network models of wikipedia knowledge community. *Information Sciences* 339, 158–174.
- Kivelä, M., A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, et M. A. Porter (2014). Multilayer networks. *Journal of Complex Networks* 2(3), 203–271.
- Mooney, P. et P. Corcoran (2014). Analysis of interaction and co-editing patterns amongst openstreetmap contributors. *Transactions in GIS* 18(5), 633–659.
- Neis, P. et A. Zipf (2012). Analyzing the contributor activity of a volunteered geographic information project — the case of openstreetmap. *International Journal of Geo-Information*, 146–165.
- Stein, K., D. Kremer, et C. Schlieder (2015). Spatial collaboration networks of OpenStreetMap. In J. Jokar Arsanjani, A. Zipf, P. Mooney, et M. Helbich (Eds.), *OpenStreetMap in GIScience*, Lecture Notes in Geoinformation and Cartography, pp. 167–186. Springer International Publishing.
- Zielstra, D., H. Hochmair, P. Neis, et F. Tonini (2014). Areal delineation of home regions from contribution and editing patterns in openstreetmap. *ISPRS International Journal of Geo-Information* 3(4), 1211–1233.

## Summary

Modelling social interactions in volunteered geographic information projects requires defining what binds contributors together in time and space. In order to be as realistic as possible, instead of choosing one social aspect to study, we choose to build a multi-layered social network that contains several types of interaction between VGI contributors. The analysis of such a multigraph should allow the detection of communities and the definition of typical profiles of contributors. A use case on OpenStreetMap illustrates what inferences can be made about contributions based on their authors.