



HAL
open science

Dis moi où tu es, je te dirai ce que tu vaux

Clément Bertier, Farid Benbadis, Marcelo Dias de Amorim, Vania Conan

► To cite this version:

Clément Bertier, Farid Benbadis, Marcelo Dias de Amorim, Vania Conan. Dis moi où tu es, je te dirai ce que tu vaux. ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 2018, Roscoff, France. hal-01781707

HAL Id: hal-01781707

<https://hal.science/hal-01781707v1>

Submitted on 30 Apr 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.

Dis moi où tu es, je te dirai ce que tu vaux

C. Bertier^{1,2}, F. Benbadis², M. Dias de Amorim¹ et V. Conan²

¹ Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, LIP6, F-75005 Paris, France

²Thales SIX GTS, Gennevilliers, France

Peut-on avoir une idée de la centralité d'un nœud uniquement à partir de sa localisation et indépendamment de son identité? Nous étudions la centralité d'intermédiarité (*betweenness-centrality*) dans un contexte véhiculaire, en liant la valeur de la centralité à sa localisation via un découpage de l'espace. Nos résultats montrent que même si la valeur d'intermédiarité des nœuds est compliquée à estimer, notamment à cause de la mobilité des nœuds et de l'influence de son voisinage, grâce au découpage spatial il est en effet possible d'anticiper le rang futur d'un nœud grâce à sa localisation. Nous montrons que cette solution permet de trouver les endroits spécifiques possédant un rang d'intermédiarité prévisible, et donc de prévoir le rang des nœuds s'y trouvant même quand ceux-ci ont tendance à transiter fréquemment dans l'espace.

Mots-clés : Intermédiarité géolocalisée, graphes de contact, réseaux véhiculaires.

1 Introduction

Quand les nœuds (utilisateurs) forment un graphe entre eux selon un rayon de communication (lien), l'étude de ces communications directes passe par l'analyse des métriques de centralité affiliées aux nœuds [Fre77]. Celles-ci ont pour but, entre autres, de mesurer la connectivité (ou utilité) d'un nœud au sein d'un réseau. Dans le cadre des communications directes et des réseaux opportunistes, cela revient généralement à étudier la capacité d'un nœud à diffuser efficacement une donnée.

Parmi ces centralités, la centralité d'intermédiarité (*betweenness-centrality*), est très utilisée dans la littérature à cause de sa capacité naturelle à capturer le potentiel de diffusion d'une information via un nœud [SCJR12]. Mais celle-ci est compliquée à anticiper, notamment parce que les nœuds bougent et changent fréquemment de voisinage. Dans ce papier, on se pose la question de savoir *si la localisation d'un nœud a une influence quelconque sur son intermédiarité*, en particulier dans des réseaux fortement influencés par les contraintes imposées par l'espace géographique sous-jacent. Même si une réponse positive à cette question semble intuitive, peu de travaux se sont intéressés à quantifier cette influence; un exemple notable est celui du placement de services dans une ville en fonction de la centralité du réseau routier [SCI⁺07].

Nous proposons cette analyse dans le cadre des réseaux véhiculaires, notamment à cause de leur corrélation spatiale intuitive et leur mobilité aux propriétés particulières. Le fait que certains lieux fortement empruntés à la congestion présentent des caractéristiques plus favorables à la connectivité semble évident, par exemple un carrefour reliant plusieurs avenues, ou encore l'entrée d'un pont où le nombre de voies est réduit. De plus, les véhicules sont forcés de suivre une route, et ne peuvent pas se déplacer aléatoirement sur la route. De ces propriétés émergent des caractéristiques substantiellement différentes des réseaux mobiles habituels. Par exemple, sur la figure 1, nous voyons un instantané dans la ville de Luxembourg durant lequel

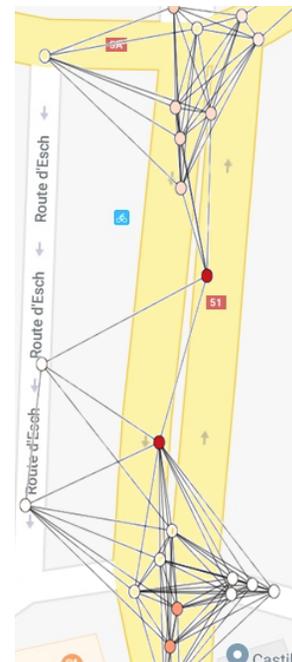


FIGURE 1: Illustration d'une haute valeur de centralité (nœuds foncés) dans la ville de Luxembourg.

deux nœuds possèdent une haute valeur d’intermédiarité (ici colorés en rouge foncé). Mais comme l’un se dirige vers le nord (voie de droite), et l’autre vers le sud (voie de gauche), le lien entre ces deux nœuds est destiné à être rompu sous peu. Suite à ce changement d’emplacement, leur intermédiarité va donc chuter. Leur valeur d’intermédiarité semble donc intuitivement être liée à leur emplacement.

Dans ce papier, nous proposons d’ignorer l’identité des nœuds pour nous concentrer sur leur localisation. Pour ce faire, nous découpons l’espace pour ensuite analyser les valeurs des nœuds dans ces sous-ensembles de l’espace. Nous observons que le classement des nœuds ayant les valeurs d’intermédiarité les plus élevées du réseau ne sont pas réparties équitablement dans l’espace. Nous montrons que cela permet, dans certains lieux, d’anticiper la valeur future du rang d’un nœud, révélant que les positions peuvent prévaloir sur l’identité de la personne. Dans nos expérimentations, nous sommes capables d’estimer correctement entre 80 et 90% des nœuds qui vont avoir une valeur d’intermédiarité parmi les valeurs les 20% plus hautes, grâce à leur emplacement.

2 Contexte

2.1 Jeu de données

Nous utilisons deux jeux de données véhiculaires : LuST Scenario [CFE15], une simulation de 24 heures contenant approximativement 300 000 véhicules sur 172 km² dans les entourages de la ville de Luxembourg et TAPASCologne project [UTCFO14], une autre simulation de 24 heures contenant approximativement 700 000 véhicules sur 400 km² dans les entourages de la ville de Cologne. Ces simulations ont été créées via SUMO, un simulateur de micro-mobilité (au sens où la totalité des véhicules sont mis à jour sur chaque instantané produit). L’utilisation des simulations permet une plus grande précision dans nos calculs, étant donné que les traces réelles sont généralement très irrégulières dans leur granularité.

Nous avons rétréci la surface des simulations à 20 km², correspondant à la taille des centres-villes, notamment à cause du fait qu’en dehors de ces derniers les contacts entre véhicules sont très rares. Dans cet article, nous considérons que les nœuds (véhicules) forment des liens avec leur voisinage selon un rayon de communication autour d’eux fixé à $\Delta = 50$ m.

2.2 Centralité d’intermédiarité

Les métriques de centralité sont nombreuses, mais dans le domaine des réseaux opportunistes la centralité d’intermédiarité prédomine pour sa capacité à sélectionner des nœuds importants [ZXS13]. Elle est définie tel que [Fre77] :

$$g(v) = \sum_{s \neq v \neq d} \frac{\sigma_{sd}(v)}{\sigma_{sd}}, \quad (1)$$

avec s la source, d la destination, v le nœud sur lequel la valeur d’intermédiarité est calculée, σ_{sd} comme le nombre total de plus courts chemins entre s et d et $\sigma_{sd}(v)$ comme le nombre de ces plus courts chemins qui passent par v . Comme $s \neq v \neq d$, alors les chemins directs (c.-à-d. s est à un saut de d), ne sont pas comptabilisés. Les paires sont uniques, c’est à dire que $\sigma_{sd} = \sigma_{ds}$, et ne sont comptabilisés qu’une seule fois dans le calcul.

Dans notre situation, cette métrique a pour but de révéler les emplacements importants et par conséquent les nœuds fréquentant ces emplacements ; autrement dit, la disparition de ces nœuds a de fortes chances d’affecter la structure du réseau dans son ensemble, soit par l’accroissement des chemins, soit par la disparition complète de chemins entre certains nœuds.

3 Anticipation des positions à forte intermédiarité

Notre intuition est que les positions des nœuds et leurs intermédiarités sont liées. Autrement dit, en regardant la suite de valeurs obtenues dans un lieu précis, on devrait pouvoir observer une certaine redondance dans celles-ci, pour éventuellement anticiper les valeurs à venir.

Dis moi où tu es, je te dirai ce que tu vaux

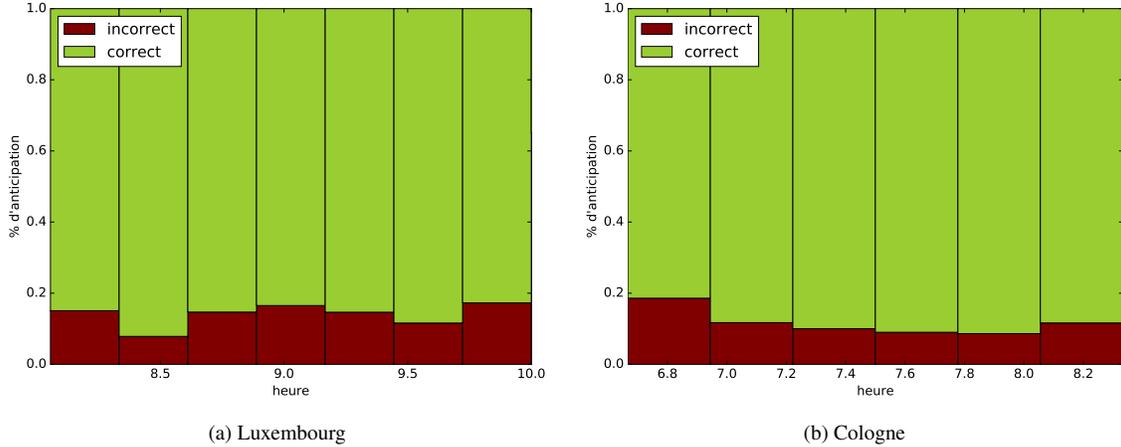


FIGURE 2: Représentation de la probabilité d’anticiper avec succès un nœud appartenant au premier quintile (top 20%). Les horaires pour les deux graphiques ne sont pas les mêmes car l’heure de pointe ne commence pas exactement au même moment.

3.1 Paramètres et possibilité d’anticipation

Dans nos calculs, nous discrétisons le temps en une suite d’instantanés t du réseau, séparés par σ secondes, et découpons l’espace en carrés notés c . Pour le calcul des intermédialités des carrés, nous utilisons une fenêtre glissante de taille W , donc dans la fenêtre $]t - W, t]$ on retrouve $\lfloor \frac{W}{\sigma} \rfloor$ instants au total.

À chaque instantané, nous calculons la valeur d’intermédialité des nœuds existants durant celui-ci. Comme la valeur absolue d’intermédialité est difficile à interpréter, nous faisons un classement des valeurs d’intermédialité. De cette manière, chaque nœud obtient un rang en fonction de sa valeur ; ce rang n’est pas unique, deux nœuds ayant la même intermédialité obtiennent donc le même rang. Ces rangs sont ensuite segmentés en quintiles (c.-à-d., top 20%, 20-40%, etc.), de manière à séparer ces rangs par paliers. Nous notons $\gamma(t, n)$ comme étant le quintile d’un nœud n à un instant donné t ($\gamma: X \mapsto [1, 5]$, avec 1 étant le top 20%).

Nous attribuons ensuite ces valeurs aux carrés respectifs dans lesquels les nœuds résident au moment de l’instantané. Les carrés, qui ont donc en mémoire chaque valeur γ des nœuds ayant résidé dans leur périmètre dans la fenêtre $]t - W, t]$, nous permettent de faire la distribution des quintiles dans la zone qu’ils délimitent, et notons $\delta_q(t, c)$ la probabilité du quintile $q \in [1, 5]$ pour un carré c durant la fenêtre $]t - W, t]$. Nous notons $\hat{q}(t, c)$ comme étant le quintile ayant la plus grande probabilité.

Nous avons pu observer qu’un quintile a une probabilité souvent beaucoup plus importante que les autres, et ce notamment dans les zones à fort trafic. Il devient ainsi possible d’anticiper avec une certaine confiance la valeur du quintile d’un nœud n entrant dans un carré c à l’instant $t + \sigma$.

Nous considérons dans cet article qu’un carré permet une possibilité d’anticipation si $\delta_{\hat{q}}(t, c) > \epsilon$. C’est à dire seulement si le carré c possède un quintile avec une densité de probabilité supérieure à une valeur minimale ϵ . Si un carré ne respecte pas cette condition, alors nous n’essayons pas d’anticiper la valeur des nœuds dans son périmètre ; il est considéré comme étant trop instable. Si un carré ne contient aucune valeur γ , c’est qu’il n’a donc pas été visité et nous ne pouvons également pas anticiper la valeur des nœuds futurs.

Dans cet article, nous fixons ϵ à 0.7 et la taille de fenêtre W à 60 secondes. Les instantanés sont séparés par $\sigma = 10$ secondes. Nous gardons l’étude précise de ces paramètres comme travail futur.

3.2 Interprétation

Il pourrait sembler correct de simplement comptabiliser combien de nœuds ont vu leur quintile correctement anticipé par les carrés. Néanmoins, cette méthode présente des résultats biaisés, dû au fait qu’une large portion des nœuds a une intermédialité de valeur 0. En effet, comme nous considérons un centre-ville vaste, beaucoup de voitures n’ont aucune communication ou ne sont sur aucun plus court chemin. Ces valeurs de

0 représentent 53% des données sur Luxembourg et 42% sur Cologne ; c'est une conséquence des zones très éparées, dans lesquelles les valeurs d'intermédiarité, si elles existent (car certaines zones ne sont pas accessibles à la circulation), sont généralement très basses. L'anticipation de ces valeurs est intuitive, à cause du peu de personnes résidant dans ces carrés.

Sur la figure 2, pour une question de lisibilité, nous montrons les résultats par tranche de 1 000 secondes. La portion verte représente le pourcentage de nœuds pour lesquels le premier quintile (donc le nombre de nœuds ayant un rang parmi les 20% les plus hauts) a été correctement anticipé, tandis que la couleur rouge représente le taux de mauvaise anticipation.

Sur la figure 2, le début des heures tracées ne sont pas exactement identiques sur les deux jeux de données. Cela est dû au fait que l'heure de pointe matinale, nécessaire à la création de congestion (c.-à-d. de contacts), ne commence pas exactement au même moment dans les deux villes. Malgré cela, nous voyons que durant cette heure de pointe matinale, la capacité à anticiper efficacement les nœuds dont la valeur appartient au premier quintile est notable, jusqu'à se maintenir entre 80% à 90% de valeurs correctement anticipées, malgré le fait que ces deux villes soient très différentes d'un point de vue topographique. À titre d'illustration, Luxembourg a beaucoup de congestion en son centre-ville alors que Cologne voit sa périphérie plus embouteillée. Cette heure de pointe engendre un type de trafic substantiellement différent de celui trouvé durant la nuit (par exemple de minuit à 5 heures du matin), où nos évaluations ont montré que cette anticipation s'avérait beaucoup moins efficace.

Même s'il est nécessaire d'avoir une densité minimale en terme de trafic pour pouvoir anticiper correctement (dans le cas contraire les valeurs trop faibles d'intermédiarité rendent le classement trop variable pour être utilisable), notre intuition initiale qui était que certaines zones permettent de déterminer, et donc d'anticiper, la valeur relative d'un nœud semble ici être avérée.

4 Conclusion et travaux futurs

Dans cet article, nous avons présenté une étude de la valeur de la centralité d'intermédiarité liée à sa localisation, et nous avons montré qu'après un découpage géographique, les valeurs d'intermédiarité les plus hautes étaient prévisibles via la valeur de quintile la plus représentée dans le carré, et ceci sur deux jeux de données véhiculaires à la topographie et trafic substantiellement différents. Dans nos travaux futurs, nous souhaitons évaluer l'influence des multiples paramètres sur ces résultats.

Références

- [CFE15] Lara Codeca, Raphael Frank, and Thomas Engel. Luxembourg SUMO Traffic (LuST) Scenario : 24 hours of mobility for vehicular networking research. In *2015 IEEE Vehicular Networking Conference (VNC)*, volume 2016-Janua, pages 1–8. IEEE, dec 2015.
- [Fre77] Linton C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1) :35, mar 1977.
- [SCI⁺07] Emanuele Strano, Alessio Cardillo, Valentino Iacoviello, Vito Latora, Roberto Messori, Sergio Porta, and Salvatore Scellato. Street centrality vs. commerce and service locations in cities : a Kernel Density Correlation case study in Bologna, Italy. pages 1–14, jan 2007.
- [SCJR12] Mary R. Schurgot, Cristina Comaniciu, and K. Jaffres-Runser. Beyond traditional DTN routing : social networks for opportunistic communication. *IEEE Communications Magazine*, 50(7) :155–162, jul 2012.
- [UTCFO14] Sandesh Uppoor, Oscar Trullols-Cruces, Marco Fiore, and Jose M. Barcelo-Ordinas. Generation and Analysis of a Large-Scale Urban Vehicular Mobility Dataset. *IEEE Transactions on Mobile Computing*, 13(5) :1061–1075, may 2014.
- [ZXS13] Ying Zhu, Bin Xu, Xinghua Shi, and Yu Wang. A survey of social-based routing in delay tolerant networks : Positive and negative social effects. *IEEE Communications Surveys and Tutorials*, 15(1) :387–401, 2013.