



## Graphe de contacts et ondelettes

Benjamin Girault, Paulo Gonçalves, Eric Fleury

► **To cite this version:**

Benjamin Girault, Paulo Gonçalves, Eric Fleury. Graphe de contacts et ondelettes. École d'Été ResCom 2013, May 2013, île de Porquerolles, France. <hal-00907018>

**HAL Id: hal-00907018**

**<https://hal.inria.fr/hal-00907018>**

Submitted on 21 Nov 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.

# Graphe de contacts et ondelettes

Benjamin Girault\*, Paulo Gonçalves<sup>†\*</sup> et Éric Fleury\*

\*École Normale Supérieure de Lyon – LIP – INRIA, DANTE – CNRS : UMR5668 – UCBL – Lyon I – Université de Lyon

<sup>†</sup>Inria Rhône-Alpes, France

## I. MOTIVATIONS

Les paramètres de la diffusion et des mutations des souches bactériennes nosocomiales sont aujourd’hui encore mal comprises. Les mécanismes macroscopiques en jeu lors de la diffusion sont à opposer à des mécanismes microscopiques qui sont eux bien connus et compris. Le passage à l’échelle d’un hôpital conduit alors à l’étude d’un système complexe qui doit être simplifié, modélisé, avant d’en faire une étude épidémiologique.

Ce travail entre dans le cadre d’un travail plus large qui entend donner un début de réponse en étudiant une corrélation entre le réseau (dynamique) de contacts et la diffusion microbiologique. Pour cela, dans le cadre du projet MOSAR (*Mastering hOSpital Antimicrobial Resistance*) au sein du groupe i-Bird (*Individual Based Investigation of Resistance Dissemination*), un réseau de capteurs sans fil fut déployé pendant plusieurs mois sur le site de l’Hôpital Maritime de Berck-sur-Mer conjointement à des analyses microbiologiques.

Les études statistiques en cours montrent que le réseau des contacts comporte plusieurs échelles spatio-temporelles. Par exemple, le réseau présente des pseudo-périodicités journalière et hebdomadaire. Spatialement, les contacts sont organisés par service, avec une affinité plus forte pour les contacts intra-service, intra-bâtiment, ou intra-étage, mais également par catégorie socio-professionnelle pour les personnels.

Ce graphe des contacts est un graphe de terrain présentant de multiples dynamiques à des échelles différentes. Par exemple, l’hôpital support de l’étude est un hôpital qui peut être divisé en deux ailes (bâtiments), eux-mêmes divisés en deux et trois étages. Chaque étage possède alors ses propres spécificités quant aux types de pathologies traitées. Parallèlement à cela, alors que la majeure partie des personnels est affectée à un service donné, certains personnels sont beaucoup plus mobiles et entrent en contacts avec d’autres personnes de manière plus homogène (par exemple, les personnels de nuit ou les ergothérapeutes). On voit donc que si le graphe des contacts expose les mêmes caractéristiques que l’organisation de l’hôpital, alors nous avons affaire à un problème ayant une dynamique structurelle multi-échelle.

En outre, les analyses microbiologiques présentent également plusieurs dynamiques. En effet, les souches bactériennes étudiées peuvent se propager entre individus, mais elles peuvent également muter et développer des résistances aux traitements. Ceci nous force donc à considérer les interactions inter-souches pour un modèle épidémique précis.

Ce jeu de données est donc remarquable par ses multiples dynamiques structurelles, temporelles et de transformation, et donne donc un cadre idéal pour mettre en application les méthodes d’ondelettes sur graphe récemment introduites. Il est également idéal pour mieux maîtriser et comprendre cet outil d’ondelettes sur graphe. On pourra voir par exemple [1] pour une méthode d’ondelette sur graphe, et [2] pour une application de cette méthode à la détection de communautés.

## II. DONNÉES I-BIRD

Le présent travail a pour but d’analyser le jeu de données issu du projet i-Bird, projet mené de Mai 2009 à Octobre 2009 à l’Hôpital Maritime de Berck-sur-Mer. Cet hôpital de moyenne à longue durée (de séjour) a été choisi en raison du caractère de « réservoir » à infection de ce type d’établissement de soins. Lors de la période de l’étude, c’est l’ensemble des personnels et des patients ayant donné leur accord qui a été mis à contribution, à savoir plus de 450 patients et près de 350 personnels. La question posée était alors de quantifier l’influence des contacts sur une diffusion microbienne.

Dans toute l’étude i-Bird, il est supposé qu’un contact physique entre deux personnes peut être approximé par une courte distance entre ces personnes (distance inférieure à un mètre et demi). Partant de cette supposition, les participants à l’étude ont été équipés de capteurs sans fil enregistrant toutes les trente secondes les capteurs détectés, afin d’obtenir une image du réseau de contacts au cours du temps.

L’objectif de cette étude de terrain est en outre d’étudier la propagation et la résistance de souches bactériennes responsables d’infections nosocomiales. Deux grandes familles de souches ont fait l’objet de prélèvements (écouvillons) hebdomadaires : les staphylocoques dorés (*Staphylococcus aureus*, responsables de 13% des infections nosocomiales) et les entérobactéries (responsables de 5% de ces infections). Ces prélèvements ont ensuite été mis en culture et analysés à la recherche des souches présentes et de leurs résistances aux différents traitements antibiotiques. Nous nous intéressons dans ce travail à la propagation et à l’évolution des souches de staphylocoques dorés au sein de l’hôpital, ces bactéries se propageant par contact physique.

À l’issue de l’étude, l’ensemble des données disponibles comprend près de un milliard et demi de contacts bidirectionnels horodatés, et près de 7000 résultats de prélèvements. Cet ensemble de données peut alors être vu comme un grand graphe dynamique (de contact) pour lequel les nœuds sont le support de plusieurs (une par souche microbienne) fonctions

de portage (de la souche). Notre but est alors d'étudier ces fonctions.

### III. ANALYSE EN ONDELETTES DE L'AGRÉGATION TEMPORELLE DU GRAPHE DES CONTACTS

Ce travail préliminaire a pour but de présenter un cadre d'étude expérimental particulièrement bien adapté à l'application de certaines approches du traitement du signal orienté graphes. Si le but à terme est l'étude simultanée de toutes les dynamiques du jeu de données, nous nous intéressons dans un premier temps à une structure de données réduite. Pour cela, nous faisons l'hypothèse qu'un contact entre deux individus est une variable aléatoire et que cette variable aléatoire vue comme une série temporelle est stationnaire. Sous cette hypothèse il est alors raisonnable de représenter la variable contact par une grandeur statistique (moyenne d'ensemble estimée empiriquement), comme par exemple la durée moyenne de contact, la durée moyenne d'inter-contact, le nombre de contacts (par jour, par semaine), ou plus simplement la durée cumulée des contacts. Le choix de cette grandeur est déterminant en termes de signification épidémiologique laquelle peut permettre de mieux comprendre la diffusion macroscopique d'une souche bactérienne.

Cette grandeur statistique doit en outre être calculée sur une échelle de temps suffisamment grande pour ne pas être trop dépendante de phénomènes individuels ou des périodes liées à l'activité des services. Étant donné que les prélèvements n'ont eu lieu qu'au plus une fois par semaine et qu'on observe une périodicité des contacts à l'échelle de la semaine, nous calculons la valeur moyenne choisie par agrégation des données de contacts sur plusieurs semaines consécutives. La question est alors de savoir comment le comportement à grande échelle (*i.e.* sur ce graphe agrégé) influence la diffusion, et en particulier quelle grandeur explique le mieux la diffusion observée.

Comme on l'a vu précédemment, l'organisation de l'hôpital est hiérarchisée. Dès lors, on s'attend à ce que le graphe des contacts révèle les mêmes caractéristiques que cette organisation, et en particulier, ait une dynamique structurelle multi-échelle. Ainsi, les outils d'analyse multirésolutions tels que les ondelettes semble un choix pertinent pour caractériser la dynamique (spatiale) de ces objets. C'est ainsi que les approches récentes d'ondelettes sur graphes telles que définies dans [1] sont au centre de l'étude de notre système.

Pour être plus précis, cette étude se focalise sur la diffusion microbienne, et à ce titre nous voulons étudier les fonctions de portage mesurées. Les questions posées sont donc comment se diffusent les souches microbiennes et quels nœuds du graphe des contacts sont des vecteurs de diffusion ou au contraire des freins à l'infection. Les ondelettes sur graphe semblent alors être un outil particulièrement adapté pour conduire l'analyse des variations spatiales d'une fonction portée par un graphe tout en intégrant dans ce schéma la structure du graphe.

Dans un premier temps, nous nous intéressons à la réponse impulsionnelle de la transformée en ondelettes sur le graphe. La conjecture que nous nous proposons de vérifier est que cette réponse impulsionnelle peut se comparer à la diffusion

bactérienne d'une souche partant du sommet portant l'impulsion. Un exemple de décomposition en ondelettes (ici fonction d'échelle) est présenté Figure 1.

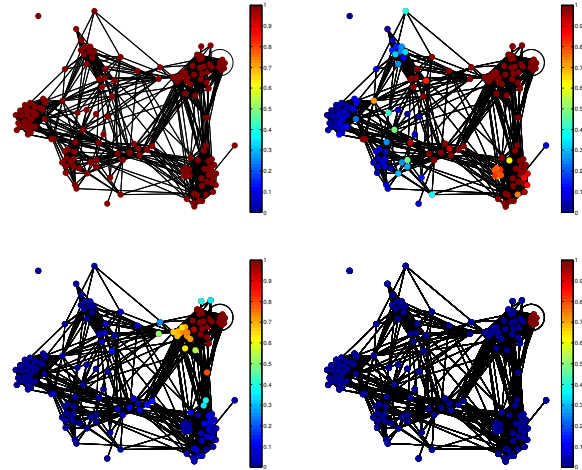


FIGURE 1. Coefficients de la fonction d'échelle de la réponse impulsionnelle en un point (ici entouré) sur le graphe des contacts pondéré par la durée cumulée des contacts, normalisés par l'énergie totale de l'échelle. Trois échelles sont ici représentées, de la plus grossière (a) à la plus fine (d). Seules les arêtes de poids suffisamment grand sont représentées.

### IV. PERSPECTIVES

Comme évoqué précédemment, ce travail est un préliminaire nécessaire à une étude correcte des données i-Bird par les outils de traitement du signal. En particulier, la question de la bonne grandeur représentative d'un contact qui expliquerait au mieux les différentes diffusions bactériennes reste ouverte et implique dans notre cas une comparaison systématique des ondelettes sur graphe avec les fonctions de diffusion.

D'un point de vue purement épidémiologique, comprendre quelle grandeur est la plus explicative aura des conséquences sur l'organisation structurelle des futurs hôpitaux. En effet, dégager de l'étude un facteur défavorisant une diffusion bactérienne est crucial pour lutter contre la prolifération des bactéries au sein d'un hôpital en ajustant par exemple sa géométrie ou les soins effectués par les personnels.

Enfin, les différentes souches bactériennes peuvent s'échanger de l'information, et à ce titre développer des résistances au différents traitements. Le projet i-Bird a ainsi parmi ses buts de comprendre comment les différents individus au sein d'un hôpital favorisent un échange d'information génétique des bactéries tel que ces bactéries deviennent multi-résistantes. Dans ce cadre, identifier les mécanismes macroscopiques en jeu dans le développement des multi-résistances est un pré-requis pour comprendre comment les freiner.

### RÉFÉRENCES

- [1] D.K. HAMMOND, P. VANDERGHEYNST et R. GRIBONVAL : Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.
- [2] Nicolas TREMBLAY et Pierre BORGAT : Multiscale Community Mining in Networks Using Spectral Graph Wavelets. Preprint, décembre 2012.