

STRUCTURE ET DYNAMIQUE DES RÉSEAUX

Master RES – M2 – Université Pierre et Marie Curie

TP 5

Clémence Magnien, Lionel Tabourier, Fabien Tarissan

`clemence.magnien@lip6.fr`

`lionel.tabourier@lip6.fr`

`fabien.tarissan@lip6.fr`

Le but de cette séance de TP est d'analyser les propriétés de plusieurs réseaux mobiles et de comparer celles-ci à celles des réseaux générés par des modèles markoviens. Pour cela, on s'appuiera sur deux jeux de données : ROLLERNET, et INFOCOM06.

1 Rendu du TP

Ce TP est noté. La date limite de rendu est le **4 janvier 2015** à minuit.

Vous devrez rendre **les documents suivants et uniquement ceux-ci** :

- un rapport de 4 pages maximum (courbes incluses) **au format pdf** présentant les courbes obtenues et vos commentaires
- une archive contenant votre code, au format `tar.gz`

Ces deux fichiers sont à envoyer dans un seul mail adressé aux trois enseignants.

Vous pouvez travailler en binôme si vous le souhaitez.

2 Préliminaires

Voici quelques informations sur les deux expériences ayant abouti aux jeux de données sur lesquelles vous allez travailler :

- **RollerNet** : Ce jeu de données a été récolté lors d'une randonnée roller à Paris en août 2006. La randonnée dure environ 3h avec une pause de 30 min et couvre environ 30km. La randonnée se déroule dans les rue de Paris ce qui implique des ralentissements et accélérations notamment lors des passages aux feux. Des *iMotes* ont été distribuées à plusieurs dizaines de participants répartis dans tout le cortège sachant qu'il y avait environ 2500 personnes en roller. Les *iMotes* distribués utilisent la technologie du bluetooth et observent les alentours toutes les 15s.
- **Infocom06** : Cette expérience se déroule lors de la conférence Infocom qui a eu lieu à Barcelone en avril 2006. Comme ROLLERNET, elle utilise la technologie *iMote* avec un scan toutes les 120s. L'expérience compte près d'une centaine de

capteurs, parmi lesquels une majorité sont des participants à la conférence, mais il y a également 17 capteurs statiques et 3 ont été placés dans les ascenseurs.

Ces jeux de données sont constitués de lignes de 4-uplets décrivant les contacts entre participants sous le format :

n_1 n_2 t_d t_f

où n_1 et n_2 sont les identifiants des 2 nœuds impliqués dans le contact, t_d indique le temps de début du contact et t_f le temps de fin du contact, c'est à dire le dernier instant pour lequel le contact entre n_1 et n_2 a été enregistré. Notons que les contacts sont ici bidirectionnels et que, par convention, $n_1 < n_2$.

À titre d'exemple, voici un petit jeu de données impliquant 4 nœuds pendant 10 unités de temps :

```
1 2 126 128
1 3 125 128
1 4 123 128
2 3 121 121
2 4 120 121
2 4 124 126
3 4 122 128
```

Exercice 1 Récupérez les fichiers `ROLLERNET` et `INFOCOM06` depuis le répertoire `/Enseignants-Public/tarissan/sdr/2014/tp5/`

Pour chacun de ces fichiers, calculez le nombre de nœuds, le nombre total de contacts et la durée totale de l'expérience.

3 Formatage

Ces jeux de données ne commencent pas tous au temps 0, ce qui rend difficile leur comparaison. Nous allons donc unifier les temps de début en les faisant tous commencer à 0. Ainsi, sur notre exemple, cela donnerait :

```
1 2 6 8
1 3 5 8
1 4 3 8
2 3 1 1
2 4 0 1
2 4 4 6
3 4 2 8
```

Exercice 2 Écrivez un programme qui normalise les fichiers de façon à ce qu'ils commencent tous au temps 0.

Ces fichiers sont triés par ordre croissant d'identifiants. Par la suite, il sera plus simple d'utiliser une version du jeu de données trié par temps et dans laquelle le début et la fin d'un contact sont dissociés. Ainsi, une ligne

`n1 n2 td tf`

donnera lieu à deux lignes

`td n1 n2 C`

et

`tf+1 n1 n2 S`

Ici, le quatrième champs est un caractère (C ou S) indique si l'événement concernant le début d'un contact ou la fin d'un contact (c'est à dire la création d'un lien ou la suppression d'un lien).

Note : La suppression du lien a lieu au temps $t_f + 1$, t_f étant le dernier instant où le contact existe.

Sur notre exemple, cela donne :

```
0 2 4 C
1 2 3 C
2 2 3 S
2 2 4 S
2 3 4 C
3 1 4 C
4 2 4 C
5 1 3 C
6 1 2 C
7 2 4 S
9 1 2 S
9 1 3 S
9 1 4 S
9 3 4 S
```

Exercice 3 *Écrivez un programme permettant de générer une telle version*

On rappelle que la commande `sort -n -kx,x fic` ordonne le fichier `fic` suivant l'ordre numérique croissant des valeurs la colonne x .

4 Analyse

Cette partie est dédiée à l'analyse de différentes propriétés des réseaux de contacts. En fonction des exercices, il sera judicieux d'utiliser l'un ou l'autre des formats (le format initial trié par nœud ou le format indiquant les créations et suppressions de liens, trié par temps).

4.1 Inter-contact

On appelle *temps d'inter-contact* le temps écoulé entre deux contacts impliquant les mêmes nœuds. Pour le deuxième contact entre 2 et 4 de l'exemple, le temps d'inter-contact est 2.

Exercice 4 *Écrivez un programme qui calcule le temps inter-contact associé à chaque contact observé.*

Exercice 5 *Calculez et affichez la distribution des temps d'inter-contact. Commentez.*

4.2 Degré moyen

On s'intéresse à maintenant l'évolution du degré des nœuds dans le réseau.

Exercice 6 *Écrivez un programme qui calcule le degré moyen à chaque instant. Est-il nécessaire de calculer le degré de chaque nœud pour faire ce calcul ?*

Exercice 7 *À l'aide de gnuplot, affichez l'évolution du degré moyen au cours du temps pour les jeux de données ROLLERNET et INFOCOM06. Commentez.*

4.3 Création et suppression de liens

On s'intéresse maintenant à la dynamique liée à la création et suppression des liens au cours du temps. On rappelle que la fraction de liens créés à un instant donné correspond au nombre de liens créés par rapport au nombre de liens qui *auraient pu* être créés, c'est à dire par rapport au nombre de liens inexistantes à l'instant précédent. De même, la fraction de liens supprimés correspond au nombre de liens supprimés par rapport au nombre de liens existants à l'instant précédent.

Le calcul de ces deux quantités demande donc de mémoriser à la fois le nombre de liens qui sont créés (resp. supprimés) à un instant donné mais aussi le nombre de liens inexistantes (resp. existants) à l'instant précédent.

Exercice 8 *Écrivez un programme qui calcule la fraction de liens créés et supprimés à chaque instant de temps et le mémorise dans un fichier. Par convention, quand la fraction ne peut être calculée¹ la valeur par défaut sera -1 .*

5 Modélisation et simulation

Les deux quantités étudiées précédemment vont servir de base pour le modèle markovien vu en cours. En effet dans ce modèle, étant donné un graphe G_t à un instant t donné, la structure du graphe à l'instant suivant G_{t+1} dépend directement de celle de G_t et est gouvernée par deux paramètres indépendants : la probabilité d'apparition p d'un nouveau lien et la probabilité d de disparition d'un lien existant. Ainsi si $G_t = (V, E_t)$, alors $G_{t+1} = (V, E_{t+1})$ avec E_{t+1} déterminé comme suit. Pour tout couple $(n_1, n_2) \in V \times V$ avec $n_1 < n_2$:

1. Par exemple lorsque le graphe est vide, on ne peut pas calculer la fraction de liens supprimés.

- si $(n_1, n_2) \notin E_t$, alors $(n_1, n_2) \in E_{t+1}$ avec probabilité p .
- si $(n_1, n_2) \in E_t$, alors $(n_1, n_2) \notin E_{t+1}$ avec probabilité d .

Le modèle nécessite donc 3 informations à chaque instant : l'ensemble E_t des liens existant à l'instant t et les probabilités p et d . Ces dernières correspondent dans notre étude à la valeur moyenne des fractions calculées au cours du temps.

Il faut deux autres informations pour pouvoir simuler le modèle :

- **Le graphe initial** G_0 : afin de reproduire le contexte des différentes expériences, nous considérerons uniquement le graphe vide dans lequel $E_0 = \emptyset$
- **Le nombre d'étape** T : afin de reproduire le contexte des différentes expériences, nous prendrons pour valeur la durée de l'expérience considérée.

Exercice 9 *Écrivez un programme qui génère une trace aléatoire suivant le modèle markovien. Ce programme prendra en paramètre le nombre de nœuds, la probabilité de création des liens, la probabilité de suppression des liens et le nombre d'étapes (instants de temps) voulues.*

Exercice 10 *À partir des programmes précédents, comparez l'évolution du degré moyen au cours du temps sur les traces générées aléatoirement par le modèle markovien et celles issues des jeux de données. Vous utiliserez pour cela le programme précédent en prenant des valeurs correspondantes au jeux de données considérés. Commentez.*

Exercice 11 *Écrivez une synthèse qui discute la pertinence du modèle markovien vis-à-vis des deux jeux de données étudiés. Vous aborderez notamment les questions suivantes (ce n'est pas une liste exhaustive) :*

- *Est-ce que le modèle se comporte de manière identique sur les deux jeux de données considérés ?*
- *Est-ce que l'analyse faite en amont (section 4.3) permettait de prédire le comportement du modèle ?*
- *Quelles propriétés sont bien reproduites par le modèles ? Sur quels jeux de données ?*
- *Quelle conclusion tirez-vous de ce travail ?*
- *Quelles pistes proposez-vous pour améliorer le modèle ?*

6 Approfondissement de l'analyse

Cette partie est destinée à l'approfondissement de l'analyse conduite jusqu'ici. Les questions sont volontairement succinctes et il vous est demandé de proposer vous-même une méthodologie pertinente pour répondre à ces questions.

Si vous avez terminé les exercices précédents, vous pouvez traiter les exercices ci-dessous, dans l'ordre que vous préférez.

6.1 Comparaison plus fine

Exercice 12 *Complétez la comparaison entre les jeux de données réels et le modèle markovien en étudiant les distributions cumulatives inverses des degrés moyens. Commentez.*

Exercice 13 *Complétez la comparaison entre les jeux de données réels et le modèle markovien en étudiant les distributions cumulatives inverses des fractions de liens créés. Commentez. Faites de même pour les liens supprimés.*

6.2 Autres propriétés comme point de comparaison

On s'est servi jusqu'ici de la notion de degré pour comparer les jeux de données et le modèle. On aurait pu choisir d'autres propriétés, par exemple la distribution des temps inter-contact ou encore le coefficient de clustering,

Exercice 14 *Complétez la comparaison des jeux de données et du modèle en vous appuyant sur d'autres propriétés structurelles des réseaux.*

6.3 Exploration d'autres modèles

À la question 8, on a calculé la fraction de liens créés et supprimés au cours du temps. Puis on s'est servi de la moyenne de ces valeurs pour simuler le modèle markovien. Au lieu de se servir de la moyenne, on aurait pu considérer un modèle qui reproduise la distribution des fractions observées au cours du temps, c'est à dire qui reproduise *globalement* la séquence des fractions observées mais pas l'ordre dans lequel elles ont lieu. Ce modèle, plus contraint que le précédent, est intuitivement plus proche des jeux de données car il reproduit exactement les apparitions/suppressions observées globalement mais change le moment où celles-ci ont lieu.

Exercice 15 *Proposez et implémentez un modèle basé sur une séquence de probabilités d'apparition/disparition des liens. Commenter l'efficacité de ce modèle au regard des propriétés des jeux de données.*

Vous pourrez bien évidemment aussi explorer vos propres modèles, en fonction des extensions que vous avez proposées pour la question 11.