

SOD322

RECHERCHE OPÉRATIONNELLE ET DONNÉES MASSIVES

Projet : Modèles pour des réseaux mobiles réels

Alexia CALANDRAS

Imad EL HANAFI

7 Mars 2018



Table des matières

Table des matières	1
Présentation du problème	2
1 Premiers traitements	2
2 Analyse des réseaux	3
2.1 Caractéristiques des réseaux	3
2.2 Créations et suppressions	5
3 Modélisation markovienne	6
3.1 Génération des graphes markoviens	6
3.2 Comparaison avec les graphes réels	6
3.3 Pertinence du modèle	7
4 Pour aller plus loin	8
4.1 Distributions des fractions de création et suppression pour RollerNet	8
4.1.1 Estimation séparée pour p et d	8
4.1.2 Estimation jointe	10
4.2 Suivi saisonnier des fractions de création et suppression pour Infocom06	11
Conclusion	12

Présentation du problème

Dans ce projet, on se propose d'analyser les propriétés de différents réseaux mobiles et comparer les résultats à ceux obtenus sur des réseaux mobiles générés par le *edge-Markovian evolving graph model*.

Nous utiliserons pour cela deux réseaux différents : **RollerNet** et **Infocom06**. Dans une première partie nous effectuerons quelques premiers traitements sur nos réseaux. Dans une seconde partie nous présenterons plusieurs outils d'analyse et leurs résultats. Enfin, nous générerons des réseaux grâce au modèle markovien et comparerons les simulations aux réseaux réels.

1 Premiers traitements

Les réseaux sont tous deux représentés par la données de leurs arêtes à l'instant t . Ces arêtes sont fournies dans un tableau comprenant les entrées suivantes :

- Premier nœud de l'arête ;
- Second nœud ;
- Date de création de l'arête ;
- Date de suppression de l'arête.

RN data - head :

	0	1	2	3
0	1	2	991	991
1	1	2	1192	1192
2	1	2	1210	1210
3	1	2	1290	1290
4	1	2	1319	1319

FIGURE 1 – Aperçu de la table **RollerNet**

IC data - head:

	0	1	2	3
0	1	3	45796	45796
1	1	3	45963	45963
2	1	3	46326	46326
3	1	3	55106	55106
4	1	3	56866	56866

FIGURE 2 – Aperçu de la table **Infocom06**

Nous avons commencé par calculer le nombre de nœuds, le nombre de contacts et la durée totale de l'expérience pour chacun des réseaux :

Caractéristique	RollerNet	Infocom06
Nombre de nœuds	62	98
Nombre de contacts	59 410	148 784
Durée totale de l'expérience (en s.)	9 976	337 418

TABLE 1 – Premiers calculs sur les réseaux

Étant donné que les sets de données ne commencent pas à la date $t = 0$, nous uniformisons les dates de début et fin d'expérience afin que ces dernières débutent à 0. Pour cela, nous retranchons la date minimale aux deux colonnes de temps.

Il peut être utile d'adopter un point de vue temporel, plutôt qu'un point de vue arêtes, et ainsi utiliser un tableau qui rend compte des créations et suppressions d'arêtes dans l'ordre chronologique. Nous retravaillons donc les deux tableaux, afin qu'une ligne soit le reflet d'un événement dans le réseaux :

- Instant t ;

- Premier nœud concerné ;
- Second nœud ;
- Type d'événement : « C » pour une création d'arête ou « S » pour une suppression.

RN data - head :

	0	1	2	3
0	0	22	31	C
1	1	22	31	S
2	12	22	31	C
3	13	22	31	S
4	37	31	44	C

IC data - head :

	0	1	2	3
0	0	13	17	C
1	375	3	14	C
2	463	16	53	C
3	526	4	5	C
4	710	22	25	C

FIGURE 3 – Transformation de la table **RollerNet** FIGURE 4 – Transformation de la table **Infocom06**

2 Analyse des réseaux

2.1 Caractéristiques des réseaux

Pour analyser les réseaux, nous définissons plusieurs mesures :

- La durée inter-contact, c'est-à-dire la durée séparant deux contacts (i.e. existence d'arête) pour une paire de nœuds donnée.
- Le degré moyen des nœuds dans le réseau à l'instant t :

$$\text{Degré moyen} = 2 \times \frac{\text{Nombre d'arêtes à l'instant } t}{\text{Nombre de noeuds}}.$$

Remarque : Il n'est pas nécessaire de calculer le degré de chaque nœud pour obtenir le degré moyen, il suffit de compter le nombre de liens actifs à la date t , les multiplier par 2 car chaque arête implique 2 nœuds, et de diviser cette valeur par le nombre de nœuds dans le graphe.

Pour ces deux caractéristiques, nous obtenons les graphes suivants :

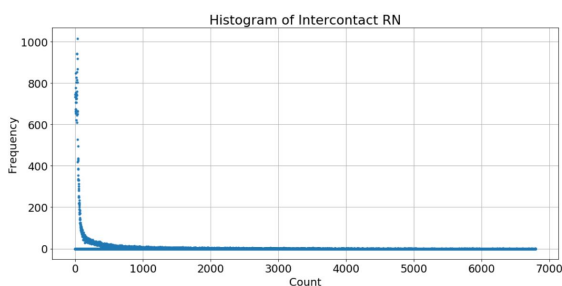


FIGURE 5 – Distribution de la durée inter-contact pour **RollerNet**

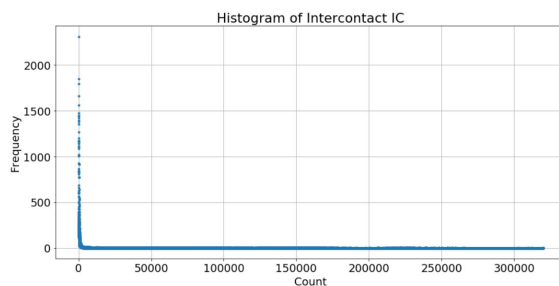


FIGURE 6 – Distribution de la durée inter-contact pour **Infocom06**

Ces graphes ne sont pas évidents à comparer, nous les traçons donc en échelle log-log :

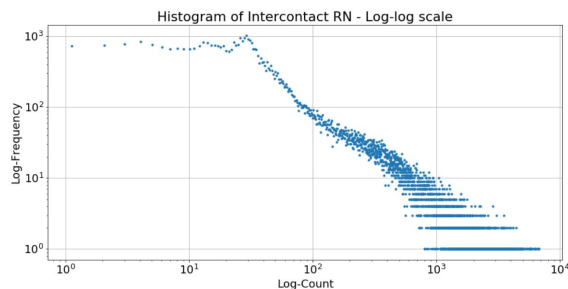


FIGURE 7 – Distribution de la durée inter-contact pour **RollerNet** en échelle log-log

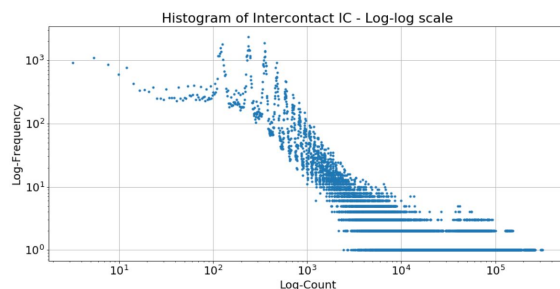


FIGURE 8 – Distribution de la durée inter-contact pour **Infocom06** en échelle log-log

Commentaires

La distribution des durées inter-contacts est très hétérogène pour les deux réseaux puisque celle-ci s'étend sur plusieurs ordres de grandeur. Par ailleurs, elle n'est pas monotone : comme nous l'avons vu en classe, les hausses et les baisses peuvent s'interpréter, notamment dans le cas d'**Infocom06**, par le fait que de nombreux contacts disparaissent en fin de journée et se recréent le lendemain matin, causant ainsi des hausses de la durée inter-contact sur une nuit.

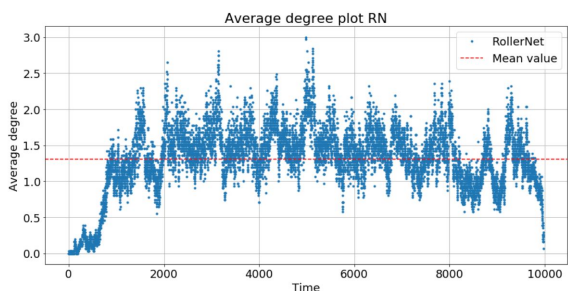


FIGURE 9 – Évolution du degré moyen de **RollerNet** pour cours du temps

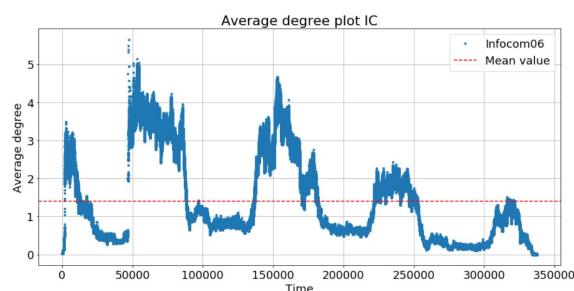


FIGURE 10 – Évolution du degré moyen d' **Infocom06** pour cours du temps

Commentaires

Les deux réseaux ne présentent pas du tout le même comportement en ce qui concerne l'évolution du degré moyen dans le réseau au cours du temps. En effet, l'expérience **RollerNet** se déroulant sur un intervalle assez court, le degré moyen présente trois phases assez marquées. Au début de l'expérience (jusqu'à $t = 1500$), énormément de liens se créent, faisant gonfler le degré moyen pour atteindre un certain seuil (approximativement, la valeur moyenne, représentée en pointillés). Ensuite, le degré moyen oscille fortement autour de ce seuil et enfin le degré chute subitement lorsque l'expérience s'achève.

Au contraire, l'expérience **Infocom06** se déroulant sur plusieurs jours, on observe une certaine périodicité dans l'allure du degré moyen. Au début de la journée, le nombre de liens dans le réseau augmente fortement puis il varie tout au long de la journée. En fin de journée, ce degré moyen diminue sensiblement et se maintient ensuite à une valeur très basse (mais non nulle), la variance est également bien plus faible qu'en journée.

On sent déjà que la notion de valeur moyenne pour le réseau **RollerNet** semble avoir un sens puisque les oscillations sont concentrées autour de cette valeur, ce qui n'est pas le cas pour **Infocom06**.

2.2 Créations et suppressions

Nous avons ensuite étudié la dynamique liée à la création et la suppression d'arêtes au cours du temps. Nous rappelons la définition de la part de liens créés à un instant t fixé :

$$\text{Ratio créés} = \frac{\text{Nombre de liens créés}}{\text{Nombre de liens non existants à l'instant précédent}}.$$

La part de liens supprimés est :

$$\text{Ratio supprimés} = \frac{\text{Nombre de liens supprimés}}{\text{Nombre de liens existants à l'instant précédent}}.$$

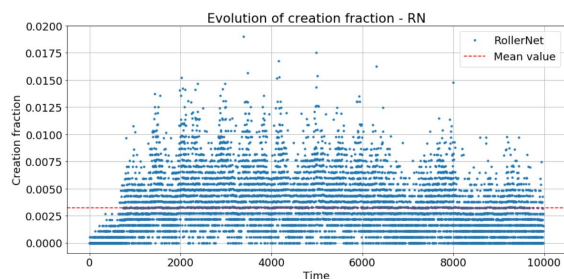


FIGURE 11 – Évolution de la fraction de création de RollerNet pour cours du temps

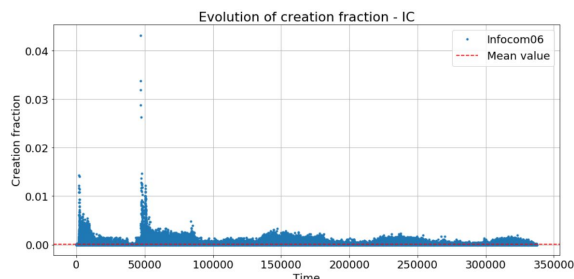


FIGURE 12 – Évolution de la fraction de création d'Infocom06 pour cours du temps

Commentaires

Au vu de la fraction de création des deux graphes, il semble que la notion de moyenne pour RollerNet ait bien du sens. En revanche, Infocom06 prend des valeurs très étendues mais dont la plupart sont concentrées vers 0. Cela a donc pour effet que la valeur moyenne est très écrasée vers l'origine et ne rend donc pas compte de la variabilité de la fraction de création des liens.

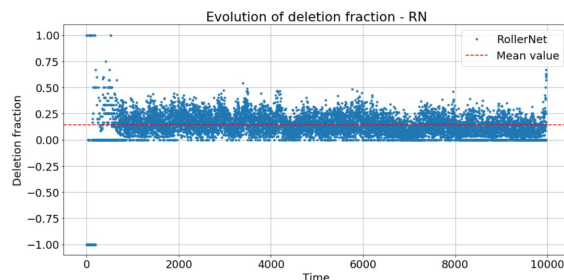


FIGURE 13 – Évolution de la fraction de suppression d' RollerNet pour cours du temps

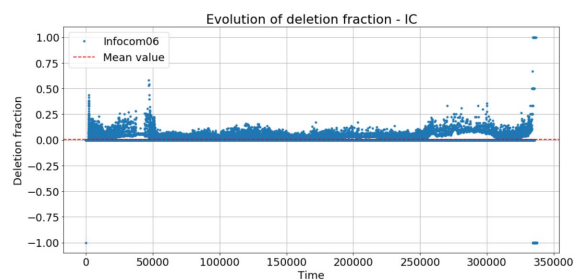


FIGURE 14 – Évolution de la fraction de suppression d' Infocom06 pour cours du temps

Commentaires

Pour ce qui est de la fraction de suppression, mettons de côté les fractions valant -1 pour les instants où il n'existe aucun lien dans le graphe. On peut alors calculer la valeur moyenne de la fraction de suppression pour les deux réseaux et faire le même constat que pour la fraction de création : la notion de moyenne semble avoir du sens pour **RollerNet**, ce n'est pas le cas pour **Infocom06**.

3 Modélisation markovienne

3.1 Génération des graphes markoviens

Nous rappelons rapidement le modèle *edge-Markovian* vu en classe. Étant donné un graphe $G_t = (V, E_t)$ à l'instant t , la structure du graphe à l'instant suivant $G_{t+1} = (V, E_{t+1})$ est donnée par :

Pour toutes les paires $(n_1, n_2) \in V \times V$ avec $n_1 < n_2$,

- 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 .

Pour effectuer la simulation, nous calculons donc pour chaque réseau les probabilités p et d i.e. la fraction de liens créés et la fraction de liens supprimés moyennées sur le temps :

Réseau	p	d
RollerNet	0.0032	0.139
Infocom06	9.54e-05	0.0045

TABLE 2 – Probabilités utilisées par le modèle markovien

Nous ferons les deux dernières hypothèses suivantes : le graphe G_0 ne comporte aucune arête au départ (comme dans les deux expériences) et le nombre d'instants T est pris identique à celui des expériences.

3.2 Comparaison avec les graphes réels

Afin de comparer les simulations aux réseaux réels, nous commençons par tracer l'évolution du degré moyen des réseaux simulés au cours du temps. Pour faciliter la comparaison, nous superposons les tracés correspondants aux réseaux réels (en bleu) et simulés (en orange). Les pointillés rouges représentent la valeur moyenne du degré moyen pour le réseau réel, et les pointillés noirs celle du réseau simulé.

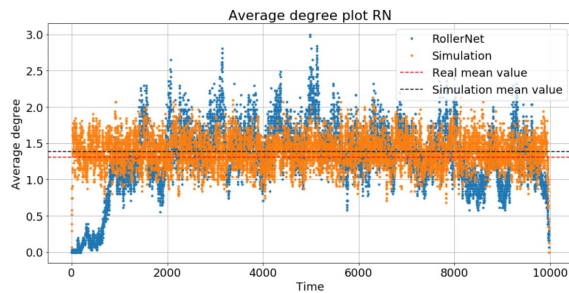


FIGURE 15 – Comparaison du degré moyen au cours du temps pour **RollerNet** réel et généré

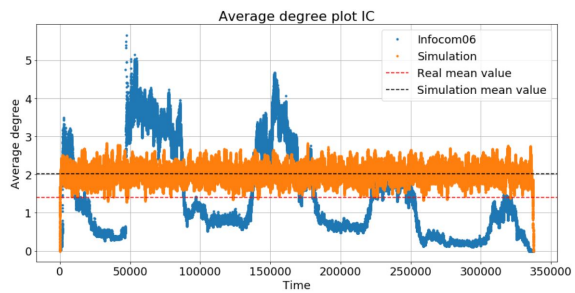


FIGURE 16 – Comparaison du degré moyen au cours du temps pour **Infocom06** réel et généré

Commentaires

Pour **RollerNet**, le réseau simulé semble plutôt bien approcher le réseau réel puisque les deux tracés semblent osciller autour de valeurs moyennes comparables. Cependant, le tracé du degré moyen pour le réseau simulé est légèrement plus écrasé (sa variance est donc plus faible) et ne reprend pas les effets de bords en début et fin d'expérience. La latence à l'établissement des liens dans le réseau réel n'est pas capturée par le réseau simulé.

Au contraire, pour **Infocom06**, le réseau simulé ne parvient pas du tout à capturer les oscillations du degré moyen dans le réseau réel. Le modèle présente un degré moyen particulièrement écrasé autour de la valeur 2, tandis que le réseau réel présente une certaine périodicité liée à la tenue de l'expérience sur plusieurs jours, mais aussi une grande variabilité lorsque l'on se restreint à des intervalles de temps plus courts. C'est cette variabilité au sein du réseau réel qui explique que celui-ci ne se maintient quasiment jamais au niveau de sa moyenne sur l'expérience et donc est mal simulé par le réseau markovien qui est basé sur cette hypothèse.

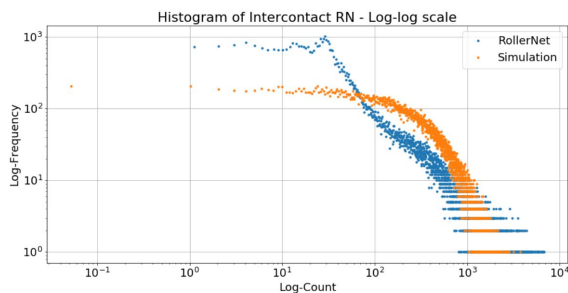


FIGURE 17 – Comparaison de la durée inter-contact pour **RollerNet** réel et généré

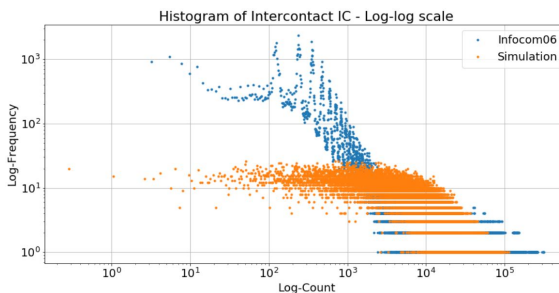


FIGURE 18 – Comparaison de la durée inter-contact pour **Infocom06** réel et généré

Commentaires

Pour **RollerNet**, la distribution de la durée inter-contact du réseau simulé est moins hétérogène que celle du réseau réel, les valeurs sont bien plus resserrées. Cela semble cohérent avec le fait que p et d aient été choisis constants lors de la simulation. L'allure générale de la courbe orange (réseau simulé) ne suit pas vraiment celle du réseau réel.

Ces remarques sont encore plus vraies pour **Infocom06** : la distribution du réseau simulé diffère vraiment du réseau réel. Les variations observées sur le réseau réel ne sont pas du tout rendues par la simulation.

Pour les deux expériences, le modèle markovien peine donc à capturer la durée inter-contact.

3.3 Pertinence du modèle

Finalement, le réseau le mieux représenté par ce modèle markovien est le réseau **RollerNet** mais cette satisfaction est à nuancer. En effet, si la notion de degré moyen semble assez bien capturée par la simulation, ce n'est pas le cas de la durée inter-contact dont la variabilité se trouve écrasée par le recours aux moyennes.

Pour l'expérience **Infocom06**, ni le degré moyen au cours du temps, ni la durée inter-contact n'est rendu correctement. Cela est dû au fait que cette expérience a lieu sur plusieurs jours et présente donc une périodicité qui n'est forcément plus capturée lorsque l'on utilise un modèle moyenné.

On aurait bien sûr pu prévoir de tels résultats puisque, lors de l'étude des fractions de créations et suppressions au paragraphe 2.2, on avait vu que la notion de moyenne semblait compromise pour **Infocom06**. Elle semblait avoir davantage de sens pour **RollerNet**, ce qui se confirme avec les résultats du réseau simulé.

Notons tout de même que la pertinence de la notion de moyenne dans un réseau n'implique pas que le modèle markovien parviendra bien à reproduire toutes les propriétés globales du réseau.

4 Pour aller plus loin

Dans cette partie, nous allons tenter d'améliorer le modèle markovien précédemment évoqué en lui apportant des modifications. Comme nous l'avons vu, utiliser uniquement les valeurs moyennes des fractions de création et suppression (p_{moyen} et d_{moyen}) ne nous permet pas de retranscrire les spécificités des expériences (notamment la périodicité liée au jour et à la nuit).

Nous envisageons donc de ne plus moyenner les valeurs de p et d mais de déterminer la distribution suivies par ces dernières, et de générer un modèle markovien tirant p et d à chaque instant selon les distributions établies.

4.1 Distributions des fractions de création et suppression pour RollerNet

4.1.1 Estimation séparée pour p et d

Dans cette partie, nous tentons d'estimer les densités de probabilité de p et d afin de simuler un modèle markovien avec p et d tirés selon cette distribution. Notons tout de même que le code que nous avons mis en place requiert un certain temps de calcul, et nous ne l'avons donc pas implémenté sur le réseau **Infocom06**. Nous proposons une autre méthode pour ce réseau dans la section 4.2.

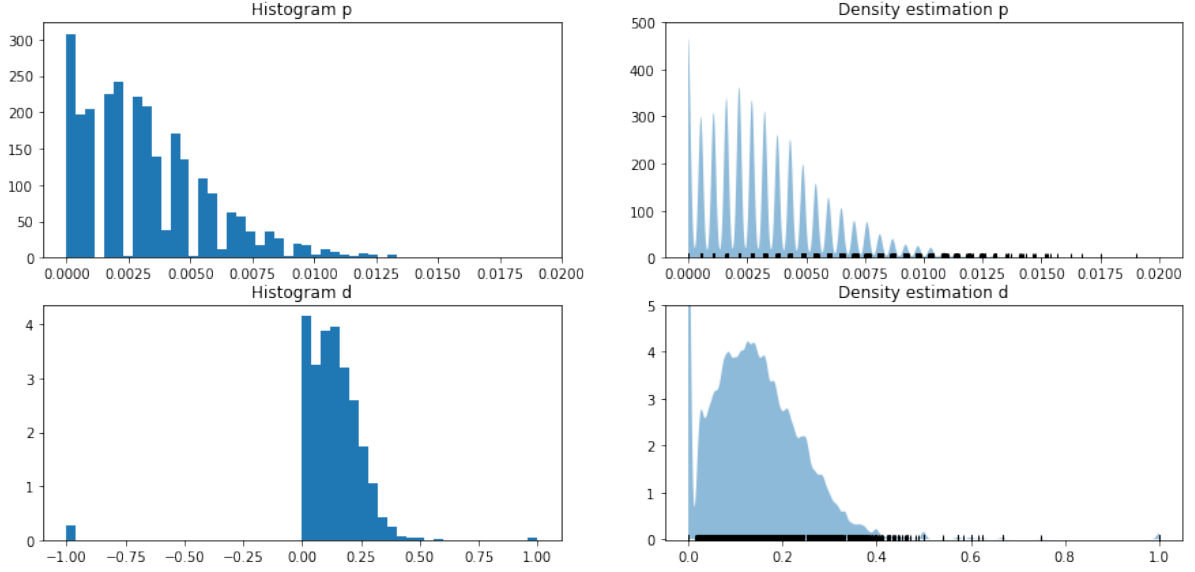
Pour déterminer ces densités, nous choisissons d'effectuer une estimation par noyau. Soit x_1, x_2, \dots, x_N un échantillon i.i.d. d'une variable aléatoire (de densité f). L'estimateur non-paramétrique de la densité par la méthode des noyaux est alors :

$$\hat{f}_h(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right)$$

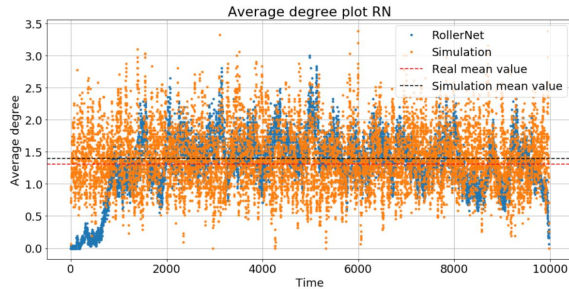
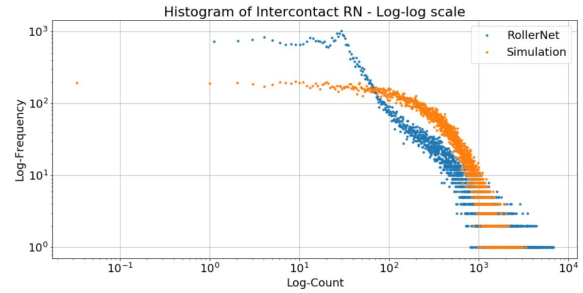
où :

- K est un noyau (« kernel ») ;
- h est la fenêtre qui contrôle le degré de lissage de l'estimation.

Sur le graphe ci-dessous, nous avons tracé les histogrammes de p et d pour le réseau **RollerNet** (graphes de gauche). Ensuite, nous avons obtenu un estimateur de la densité de chacune des variables en utilisant un noyau gaussien (graphes de droite).

FIGURE 19 – Histogrammes et estimations des densités pour p et d

Toujours pour **RollerNet**, nous avons simulé un réseau qui tire à chaque instant t les fractions p_t et d_t selon les distributions déterminées précédemment.

FIGURE 20 – Comparaison du degré moyen au cours du temps pour **RollerNet** réel et généréFIGURE 21 – Comparaison de la durée inter-contact pour **RollerNet** réel et généré

Commentaires

L'utilisation d'une distribution hétérogène nous permet d'obtenir un meilleur réseau simulé qui s'approche davantage du réseau réel que le modèle markovien, du moins au niveau des degrés moyens. En effet, les variations obtenues avec ce nouveau modèle sont plus représentatives du réseau réel (variance plus élevée que le premier modèle). En ce qui concerne la durée inter-contact, celle-ci n'est toujours pas capturée.

Cependant, cette méthode ne tient pas compte des corrélations qui pourraient exister entre p et d . Nous allons donc proposer une seconde méthode d'estimation, qui cette fois-ci tenterait d'estimer la densité jointe (p, d) en utilisant toujours la méthode des noyau mais en dimension 2 cette fois.

4.1.2 Estimation jointe

Considérons toujours le réseau **RollerNet**, le graphe ci-dessous représente le nuage de points (« scatter plot ») des variables p et d ainsi que la distribution de (p, q) estimée (projeté en dimension 2 dans le graphe de type « contour », i.e. en courbes de niveaux).

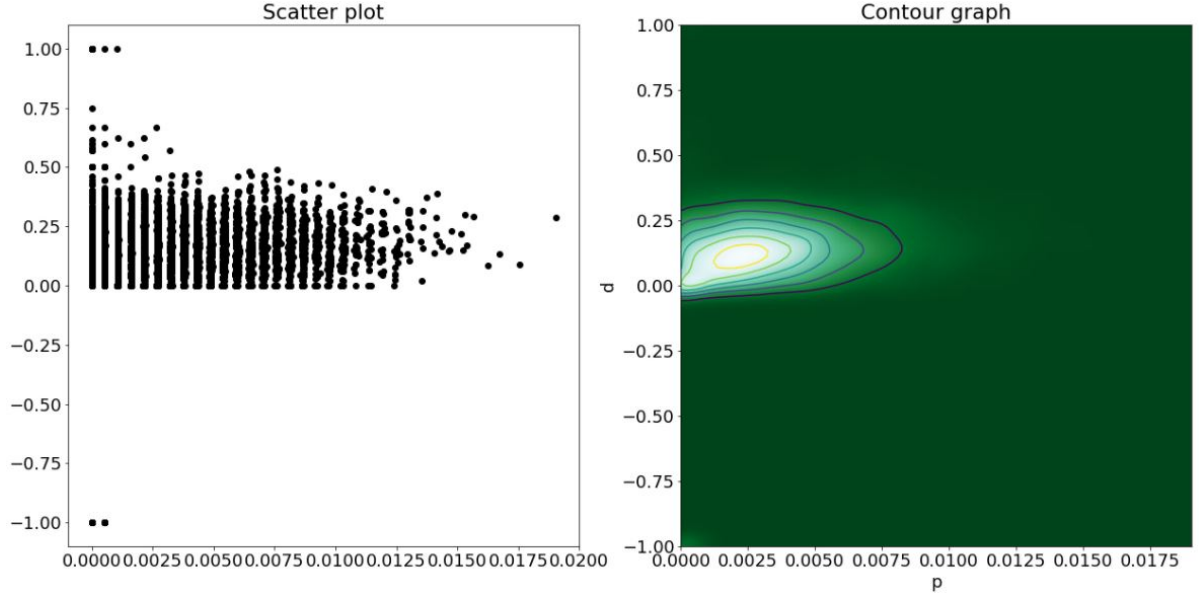


FIGURE 22 – Scatter plot et distributions jointe des fractions de création et suppression

Cette méthode présente a priori l'avantage de pouvoir capturer les relations/corrélation qui existent entre les fractions de création et suppression, ce qui pourrait nous permettre de produire un réseau simulé plus proche du réel.

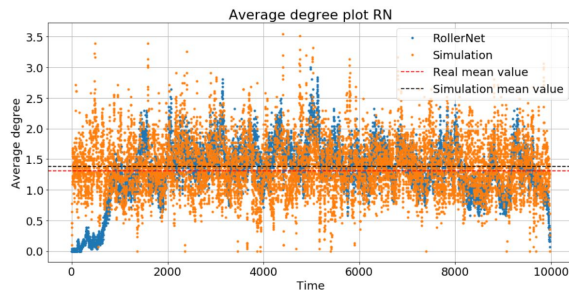


FIGURE 23 – Comparaison du degré moyen au cours du temps pour RollerNet réel et généré

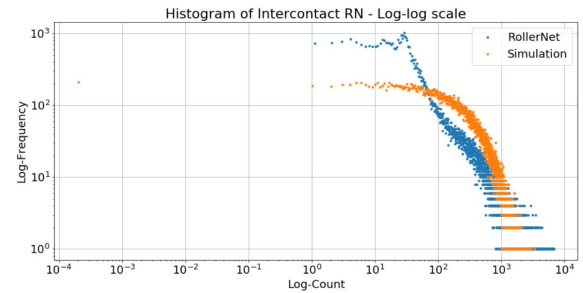


FIGURE 24 – Comparaison de la durée inter-contact pour RollerNet réel et généré

Commentaires

Cette seconde simulation basée sur l'estimation jointe de (p, d) ne semble finalement pas spécialement plus convaincante que la simulation basée sur l'estimation séparée de p et d .

Pour les deux dernières simulations de **RollerNet**, les valeurs moyennes simulées sont très proches de la vraie valeur moyenne. La variance du réseau est bien imitée, mais elle ne colle évidemment pas parfaitement au réseau réel. Ces deux modèles constituent tout de même une bonne base pour étudier le réseau **RollerNet**.

Autres méthodes possibles

Malgré les améliorations obtenues en utilisant des distributions hétérogènes, nous pensons que d'autres méthodes pourraient peut-être conduire à de meilleures simulations. Nous pensons en particulier aux *chaînes de Markov* et les *séries chronologiques*. Ces outils permettraient de capturer l'évolution temporelle des variables p et d , et ainsi trouver un modèle plus réaliste.

4.2 Suivi saisonnier des fractions de création et suppression pour Infocom06

Comme nous l'avons dit en début de partie 4., les calculs sur le réseau **Infocom06** peuvent prendre un temps relativement considérable. Nous avons donc choisi une autre approche pour ce réseau. La périodicité observée dans l'allure du degré moyen du réseau **Infocom06** suggère un découpage du réseau en intervalles $[t_i, t_{i+1}]$ pour lesquels le degré moyen est relativement homogène, comme suit :

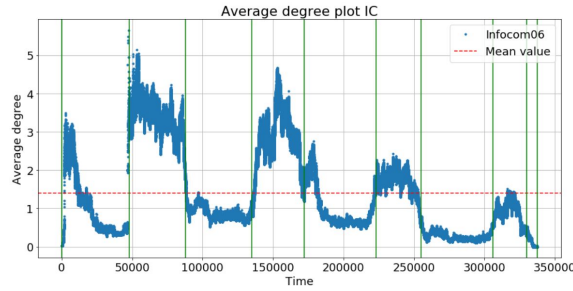


FIGURE 25 – Évolution du degré moyen d'Infocom06

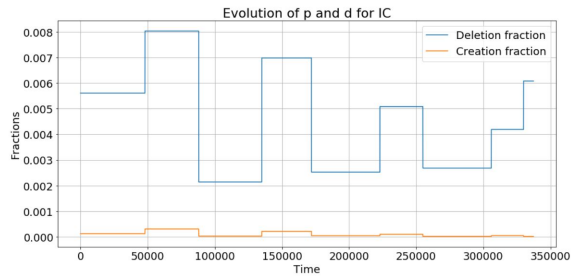


FIGURE 26 – Évolution de p et d utilisés pour la génération de réseau

Pour chaque intervalle (délimité par les lignes vertes du graphe de gauche), nous avons estimé des fractions de création et suppression p_{I_t} et d_{I_t} , où I_t désigne un intervalle du temps. Nous obtenons les valeurs représentées sur le graphe de droite.

Après avoir obtenu les valeurs moyennes de p et d sur chaque intervalle, nous relançons une simulations du réseau comme avant et nous constatons que le réseau obtenu s'approche significativement du réel.

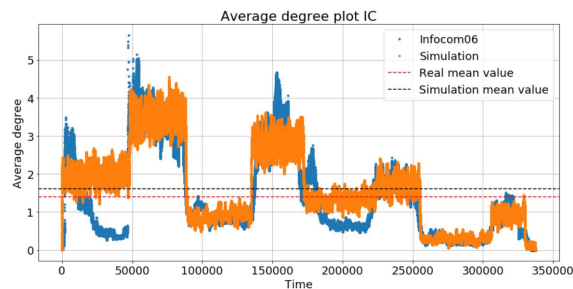


FIGURE 27 – Comparaison du degré moyen au cours du temps pour Infocom06 réel et généré

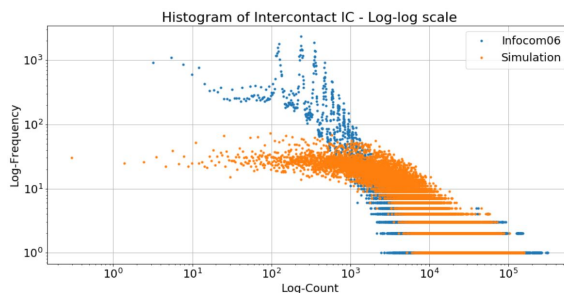


FIGURE 28 – Comparaison de la durée inter-contact pour Infocom06 réel et généré

Commentaires

Ce découpage en intervalles de temps améliore sensiblement le résultat sur le réseau Infocom06. En effet, à l'échelle d'une journée au cours de l'expérience, la notion de moyenne a davantage de sens que la moyenne sur l'ensemble de l'expérience. Ainsi, le degré moyen est approximé par morceaux et colle bien mieux à la courbe du réseau réel. De même, l'ordre de grandeur de la durée inter-contact est également amélioré : les valeurs sur le début de l'expérience sont rehaussées, plus proches de celles du réseau réel. En revanche, cette seconde propriété n'est toujours pas très bien capturée.

Conclusion

Au cours de ce projet, nous avons mis en œuvre plusieurs modèles de génération de réseau à partir de réseaux réels. Le réseau markovien à génération d'arêtes n'ayant pas réussi à capturer les propriétés de degré moyen et de durée inter-contact de manière fidèle, nous avons tenté de l'améliorer en estimant les lois selon lesquelles tirer les probabilités de création et suppression d'arêtes au cours du temps. Nous sommes parvenus à affiner un peu le modèle pour RollerNet mais celui-ci reste encore perfectible. Dans le cas d'Infocom06, en raison de la périodicité, nous avons proposé un modèle approprié pour simuler un tel réseau dynamique. Le résultat est nettement amélioré en ce qui concerne la propriété de degré moyen du graphe au cours du temps.