

Modélisation des graphes de terrain

Fabien Tarissan, Matthieu Latapy

`stages@complexnetworks.fr`

`http://complexnetworks.fr`

LIP6 – CNRS et UPMC – Paris

Dans de très nombreux contextes applicatifs, on rencontre de grands graphes n'ayant aucune structure simple apparente, que nous appellerons *graphes de terrain* (par opposition aux graphes explicitement construits par un modèle ou une théorie). Citons par exemple la topologie de l'internet (routeurs et câbles entre eux), les graphes du web (pages web et liens hypertextes entre elles), les échanges divers (pair-à-pair, e-mail, etc), mais aussi les réseaux sociaux, biologiques ou linguistiques.

Il est apparu récemment que la plupart de ces grands graphes ont des propriétés statistiques en commun, et que ces propriétés les différencient fortement des graphes aléatoires¹ utilisés jusqu'alors pour les modéliser. Notamment, ils ont une densité² très faible, une distance moyenne faible, une distribution de degrés³ hétérogène, et une densité locale⁴ forte.

Depuis lors, de nombreux travaux ont été menés visant à capturer ces propriétés dans des modèles, nécessaires tant pour effectuer des simulations que pour étudier formellement ces objets, et bien sûr pour en *comprendre* la nature.

Les graphes aléatoires classiques capturent la densité faible (qui est en fait un paramètre du modèle) et la distance moyenne faible. Par contre, ils ont une distribution des degrés homogène et une densité locale faible.

Nous sommes également en mesure de générer un graphe aléatoire à distribution de degrés donnée. On capture ainsi toutes les propriétés citées ci-dessus sauf la densité locale forte.

Malgré de nombreuses tentatives, générer des graphes ayant également une densité locale forte *tout en gardant leur caractère aléatoire* reste un problème ouvert. Les attentes sont pourtant extrêmement fortes.

Nous proposons d'attaquer ce problème sous deux angles nouveaux.

Dans la première approche, il s'agit de *transformer la contrainte sur la densité locale forte en une contrainte sur les degrés*. En effet, on sait faire beaucoup de choses sur les degrés, et notamment des tirages aléatoires satisfaisant diverses contraintes. Or la densité locale forte peut être vue comme conséquence de la présence de nombreuses cliques, qui peuvent être codées par des distributions de degrés dans un graphe biparti représentant le graphe original. En itérant cette remarque, on espère pouvoir capturer non seulement la densité locale mais des propriétés plus subtiles des graphes considérés.

L'autre approche repose sur la théorie des processus de branchements, qui ont prouvé leur efficacité dans le contexte des graphes aléatoires mais n'arrivent pas à capturer la densité locale. Nous proposons de nous affranchir de certaines contraintes de modélisation afin d'être capables de capturer cette propriété dans des processus de branchements. Ceux-ci ne pourront plus être directement interprétés comme des graphes aléatoires, mais ils pourront capturer des comportements observés empiriquement sur les graphes de terrain.

Les travaux préliminaires que nous avons menés ont montré que ces deux approches sont pertinentes et extrêmement prometteuses. L'objectif de ce stage est de les explorer plus avant.

1. Nous entendons par *aléatoire* un tirage aléatoire uniforme dans un ensemble donné, ici l'ensemble des graphes ayant une taille donnée.

2. *i.e.* probabilité que deux nœuds choisis au hasard soient reliés.