

Estimation d'un modèle à blocs stochastiques dense échantillonné par la méthode Respondent-Driven Sampling (RDS).

Thi Phuong Thuy VO, LAMA - Université Gustave Eiffel - Champs-sur-Marne
Viet Chi TRAN, LAMA - Université Gustave Eiffel - Champs-sur-Marne

L'échantillonnage en fonction des répondants ("Respondent - Driven Sampling", RDS) peut être utilisé pour découvrir des réseaux sociaux dans des population cachées. Les personnes interrogées sont invitées à indiquer leurs partenaires et un certain nombre de coupons sont remis à certaines des ces personnes. Ceci peut conduire à l'étude d'une chaîne de Markov sur un graphe aléatoire.

Nous considérons la chaîne de référence sur le modèle à blocs stochastiques (Stochastic Block Model, SBM). Dans notre étude, le graphe et la chaîne de référence sont construits simultanément. Nous nous intéressons au problème de récupérer des informations statistiques sur un modèle à bloc stochastique à partir du sous-graphe découvert par une marche aléatoire (correspondant à un RDS à un coupon). Nous considérons ici le cas dense où le réseau aléatoire peut être approché par un graphon. Tout d'abord, nous écrivons la vraisemblance du sous-graphe découvert par la marche aléatoire : des biais émergent car les "hubs" et les types majoritaires sont plus susceptibles d'être échantillonnés. Même dans le cas où les types sont observés, l'estimateur du maximum de vraisemblance n'est plus explicite. Lorsque les types de sommets ne sont pas observés, nous utilisons un algorithme SAEM ("Stochastic Approximation version of Expectation-Maximization algorithm") pour maximiser la vraisemblance. Deuxièmement, nous proposons une stratégie d'estimation différente en utilisant les nouveaux résultats d'Athreya et Röllin. Elle consiste à dé-biaisier l'estimateur EM variationnel proposé par Daudin et al. et qui ignore les biais.