

Instructions générales

Rapport de 15-20 pages sur un des 13 blocs ci-dessous, à effectuer en binôme, à rendre le dimanche 22 mars avant minuit. Soutenance (20 minutes d'exposé, 10 minutes de questions), dates à fixer (entre le lundi 23 mars et le vendredi 3 avril).

Le rapport est à voir comme un petit cours écrit de niveau M2 portant sur le sujet traité dans l'article. Quand on doit présenter un résultat de recherche dans le cadre d'un cours, on ne peut pas se contenter d'énoncer le théorème et d'en donner la preuve. Il faut expliquer le cadre théorique, les motivations, les précédents résultats du domaine, les éventuelles conjectures... avant de se focaliser sur la preuve d'un résultat particulier, pour que celui-ci puisse être situé dans son contexte scientifique. On ne peut donc pas se limiter à la lecture du seul article étudié : il est nécessaire d'aller lire d'autres choses. En fait, l'article étudié peut être vu comme un point de départ pour l'exploration du domaine de recherche et votre rapport peut finalement se concentrer sur la preuve d'un résultat énoncé autre part (mais dans ce cas, il est préférable de m'en tenir au courant).

Il n'y a pas de plan imposé mais disons qu'un bon mémoire pourrait être articulé de la façon suivante :

- une présentation du domaine de recherche, les notations essentielles (mais il est préférable de ne pas introduire trop de notations dès le début, juste celles qui permettent de bien poser la problématique), les motivations pour étudier ce problème ;
- une présentation des résultats principaux du domaine, et plus spécifiquement de celui qui sera prouvé dans le rapport, avec une mise en perspective bibliographique : en quoi ce résultat est nouveau ? Qu'est-ce qui avait été fait avant ? Quelles sont les hypothèses ? Est-ce que, depuis, ce résultat a été amélioré ? Pour écrire cette partie, il est nécessaire d'aller voir d'autres articles connexes, ceux qui sont cités dans l'article étudié, et éventuellement ceux qui le citent ultérieurement. Bref, il s'agit de vraiment ancrer le résultat dans une lignée de recherche.
- un aperçu général des techniques de preuve utilisées ;
- une présentation plus détaillée de la preuve. Ici, il ne s'agit pas de faire du copier-coller, mais de montrer que vous avez su vous approprier la preuve, la refaire par vous-même. Vous avez tout à fait le droit de présenter les choses différemment que dans l'article si cela vous semble plus naturel. Il ne sera pas forcément nécessaire de tout prouver, certains résultats intermédiaires et techniques pourront être énoncés comme des lemmes admis. Mais bien sûr, les étapes cruciales devront, elles, être développées.
- éventuellement une mise en oeuvre pratique du résultat présenté. Dans ce cas, le document sera accompagné d'un notebook généré à partir de RMarkdown ou Python détaillant les simulations effectuées. Les figures issues de ces simulations pourront être intégrées au document. Cette partie est **obligatoire pour les blocs 2 et 3**.

Liste d'articles

1. Clique cachée

N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4) :457–466, 1998.

Y. Dekel, O. Gurel-Gurevich, and Y. Peres. Finding hidden cliques in linear time with high probability. *Combinatorics, Probability and Computing*, 23(1) :29–49, 2014.

(Pour une introduction au problème de la clique cachée, on pourra se référer au Chapitre 1 de G. Lugosi. Lectures on combinatorial statistics. *47th Probability Summer School, Saint-Flour*, pages 1–91, 2017.)

2. Clustering spectral

U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4) :395–416, 2007.

A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering : Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.

3. Modularité

M. E. Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23) :8577–8582, 2006.

M. E. Newman. Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6) :066133, 2004.

A. Clauset, M. E. Newman, and C. Moore. Finding community structure in very large networks. *Physical review E*, 70(6) :066111, 2004.

4. Estimation de paramètres de graphes par marches aléatoires

A. Ben-Hamou, R. I. Oliveira, and Y. Peres. Estimating graph parameters with random walks. *Mathematical Statistics and Learning*, 1(3/4) :375–399, 2018.

5. Reconstruction exacte dans SBM dense

E. Abbe and C. Sandon. Community detection in general stochastic block models : Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE, 2015.

(Pour des une présentation de ces résultats, on pourra lire le Chapitre 1 de S. Bubeck et M. Z. Rácz, Basic models and questions in statistical network analysis. *Statistics Surveys*, 11 : 1–47, 2017.)

6. Reconstruction partielle dans SBM sparse

O. Guédon and R. Vershynin. Community detection in sparse networks via Grothendieck’s inequality. *Probability Theory and Related Fields*, 165(3-4) :1025–1049, 2016.

7. Reconstruction des communautés dans les modèles à blocs réguliers

G. Brito, I. Dumitriu, S. Ganguly, C. Hoffman, and L. V. Tran. Recovery and rigidity in a regular stochastic block model. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, pages 1589–1601. SIAM, 2016.

8. Détection de sous-graphes denses

E. Arias-Castro, N. Verzelen, et al. Community detection in dense random networks. *The Annals of Statistics*, 42(3) :940–969, 2014.

N. Verzelen, E. Arias-Castro, et al. Community detection in sparse random networks. *The Annals of Applied Probability*, 25(6) :3465–3510, 2015.

9. Apprentissage de graphes par requêtes d'arêtes

C. Giraud, Y. Issartel, L. Lehéricy, and M. Lerasle. Pair matching : When bandits meet stochastic block model. *arXiv preprint arXiv :1905.07342*, 2019.

10. Tests et estimation dans les graphes géométriques aléatoires

S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3) :503–532, 2016.

(On pourra lire le Chapitre 2 de S. Bubeck et M. Z. Rácz, Basic models and questions in statistical network analysis. *Statistics Surveys*, 11 :1–47, 2017.)

11. Reconstruction des positions dans les graphes aléatoires géométriques

E. Arias-Castro, A. Channarond, B. Pelletier, and N. Verzelen. On the estimation of latent distances using graph distances. *arXiv preprint arXiv :1804.10611*, 2018.

12. Estimation de la racine dans les arbres croissants aléatoires

S. Bubeck, L. Devroye, and G. Lugosi. Finding adam in random growing trees. *Random Structures & Algorithms*, 50(2) :158–172, 2017.

(On pourra lire le Chapitre 4 de S. Bubeck et M. Z. Rácz, Basic models and questions in statistical network analysis. *Statistics Surveys*, 11 :1–47, 2017.)

13. Reconstruction du graphe sous-jacent à un modèle d'Ising

A. Anandkumar, V. Y. Tan, F. Huang, A. S. Willsky, et al. High-dimensional structure estimation in Ising models : Local separation criterion. *The Annals of Statistics*, 40(3) :1346–1375, 2012.

G. Bresler. Efficiently learning ising models on arbitrary graphs. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 771–782, 2015.