

Activités de recherche actuelles

Mots clefs : algorithme, graphe, approximation, online, optimisation, analyse en moyenne, groupes dynamiques, structure de connexion, Vertex Cover, Independent Set.

Contexte

J'ai effectué ma thèse au sein de l'équipe OPAL (OPTimisation et ALgorithme), axe *Méthodes et Outils pour les Systèmes Complexes Artificiels (MOSCA)*, du laboratoire IBISC (Informatique, Biologie Intégrative et Systèmes Complexes, EA 4526) de l'Université d'Évry-Val d'Essonne. J'ai été encadré par M. Christian LAFOREST, professeur à l'université Blaise Pascal (Clermont Ferrand) sur le sujet « *Au delà de l'évaluation en pire cas : comparaison et évaluation en moyenne de processus d'optimisation pour le problème du vertex cover et des arbres de connexion de groupes dynamiques* ».

Présentation synthétique de mes activités de recherche

La théorie de l'approximation polynomiale est un axe très important dans le domaine de l'algorithmique. Ceci est principalement dû au fait que la plupart des problèmes d'optimisation combinatoire que l'on rencontre dans la pratique sont des problèmes NP-difficiles. Ainsi, on s'intéresse aux algorithmes d'approximation avec garanties de performance. Il est d'usage, lors de l'analyse des performances d'un algorithme d'approximation, de se référer au pire cas et au rapport d'approximation en pire cas. Cependant, ce pire cas est souvent le fruit d'une instance pathologique rare. L'analyse en moyenne permet d'obtenir des résultats sur les performances concrètes d'un algorithme. Mes travaux ont pour objet de mieux comprendre la dynamique de certains algorithmes d'approximation ou online et se découpent en deux parties autonomes :

Comparaison d'algorithmes d'approximation pour le problème du vertex cover. Le problème du vertex cover est le suivant : on souhaite couvrir toutes les arêtes du graphe avec un nombre minimum de sommets. Il s'agit d'un problème NP-complet classique en optimisation combinatoire et il existe de nombreux algorithmes d'approximation dont les rapports d'approximation en pire cas vont de 2 à n , la taille du graphe.

Nous avons montré que bien qu'ayant un mauvais rapport d'approximation (non-borné par une constante et pouvant être arbitrairement grand), certains algorithmes obtiennent de meilleurs résultats que ceux ayant un très bon rapport d'approximation sur une grande majorité des classes de graphes couramment utilisées, soulignant ainsi que la notion d'approximation en pire cas n'est pas suffisamment fine pour classer des algorithmes d'approximation en fonction de leurs performances. Nous avons donc décidé d'étudier le comportement en moyenne de ces différents algorithmes, ce qui nous a permis d'obtenir des résultats parfois surprenant. Par exemple, l'algorithme *MDG* (on sélectionne le sommet de degré maximum, on le retire du graphe avec ses arêtes incidentes et on recommence tant qu'il existe une arête), qui est $\Theta(\log(n))$ -approché, possède un rapport d'approximation moyen qui tend vers 1 (c'est-à-dire qu'en moyenne, il retourne une solution optimale) sur la classe des graphes spécialement conçue pour montrer que son comportement en pire cas peut être désastreux.

L'étude du comportement en moyenne de ces algorithmes nous a aussi permis de concevoir de nouveaux algorithmes qui possèdent de meilleures performances. Par exemple, nous avons modifié un algorithme de liste (un tel algorithme considère les sommets un par un dans un ordre déterminé à l'avance. Lorsque l'algorithme considère un sommet, il doit prendre une décision irrévocable à son sujet, c'est-à-dire l'ajouter ou non à la solution), List-Left, présenté par Avis et Imamura, afin que ses performances soient toujours meilleures que l'algorithme original. De plus, l'évaluation en moyenne de ce nouvel algorithme (List-Right) a permis de montrer que, même si il possède un mauvais rapport d'approximation, l'espérance de la taille de la solution retournée

est, pour tout graphe, strictement inférieure à deux fois la taille de l'optimal, montrant ainsi que ce nouvel algorithme est capable de rivaliser avec les meilleurs. Ces résultats sur les algorithmes de liste ont donné lieu à deux publications dans le journal international « Information Processing Letters », l'une en 2008, l'autre en 2009.

Nous avons poursuivi ces travaux en montrant analytiquement que, sur des classes de graphes précises telles que les graphes aléatoires d'Erdős-Rényi, l'espérance de la taille de la solution retournée List-Right est plus petite que celle de l'algorithme le plus couramment utilisé (l'algorithme Edge Deletion, qui retourne les sommets d'un couplage maximal et qui possède un rapport d'approximation de 2). Nous avons par ailleurs effectué une analyse exacte de l'espérance de la taille des solutions retournées par tous ces algorithmes sur des chemins de taille n .

Ces différentes analyses théoriques ont été complétées par une étude expérimentale, effectuée sur de nombreuses classes de graphes (chemins, arbres, grilles, hypercubes, graphes de pire cas pour les différents algorithmes, graphes d'Erdős-Rényi, ...). Il ressort de cette étude que les algorithmes ayant un rapport d'approximation mauvais sont pratiquement toujours meilleurs en moyenne que les algorithmes qui possèdent un rapport d'approximation de 2. Ces travaux ont été rédigés et sont en cours de soumission dans la revue « ACM Journal of Experimental Algorithmics ».

Arbres de connexion pour des groupes dynamiques dans un réseau. La construction d'une structure de connexion dans un réseau d'un sous-ensemble statique (qui n'évolue pas dans le temps) de membres est un problème classique. On peut en effet évoquer toute forme de forum ou de réunion via un réseau, où les membres veulent échanger des données via une structure dédiée. Dans ce type de rassemblements virtuels, il n'est cependant pas toujours possible de connaître toutes ces données à l'avance : les participants peuvent arriver et/ou partir, dans un ordre et à des moments inconnus à l'avance (version dynamique du problème). Par exemple, dans les systèmes pair à pair (*peer-to-peer*) il est inconcevable de prédire qui va échanger des données avec qui. Les échanges se font au fur et à mesure, en fonction de paramètres qu'il est difficile de quantifier à l'avance (modèle online).

Cette situation online peut-être modélisée naturellement sous forme de problèmes de graphes. Des travaux précédents ont permis la conception d'algorithmes construisant au fur et à mesure une structure couvrante (un arbre, bâti sur le réseau sous-jacent) dont la qualité est maîtrisée. Ce dernier point est particulièrement important. En effet, lorsque les membres communiquent via leur structure dédiée, la qualité de service offerte (notamment en terme de délais d'acheminement des données point à point) doit être suffisante. Cette exigence de qualité a été traduite sous la forme du problème d'optimisation suivant. Nous devons maintenir tout au long de la réunion un arbre qui minimise la distance maximum entre les membres du groupe (correspondant au temps de communication maximum entre les membres).

Ces travaux ont montré que le maintien, tout au long de la réunion, d'un arbre qui garantit cette qualité de service implique la *reconstruction* totale de l'arbre à certaines étapes. Une telle reconstruction nécessite, d'un point de vue réseau, la mise à jour des tables de routage de chaque membre du groupe, perturbant ainsi les communications en cours et induisant un fort trafic de contrôle. L'objectif est alors de minimiser le nombre d'étapes où ont lieu de telles reconstructions. Il a été montré que tout algorithme va effectuer, en pire cas, un nombre linéaire (en nombre de requêtes d'arrivée et de départ de membres) de reconstructions.

Mes travaux avaient pour objectif de proposer une méthode pour la résolution de ce problème, puis d'évaluer l'espérance du nombre d'étapes où ont lieu des reconstructions. Après avoir fourni une formule permettant de calculer l'espérance exacte du nombre de reconstructions, nous avons montré que pour tout graphe, cette espérance est constante ou bien logarithmique en nombre de requêtes, suivant les hypothèses de départ sur la probabilité d'ajout et de retrait de chaque membre du groupe. Ces résultats sont très encourageants, car ils montrent que même si il existe un pire cas extrêmement défavorable, l'algorithme proposé possède un excellent comportement moyen. Ces travaux ont été rédigés et ont été soumis pour publication dans la revue « Technique et Science Informatiques ». Une version préliminaire a fait l'objet d'une publication dans la conférence nationale (avec comité de lecture) « ALGOTEL 2007 ».

Perspectives de ces travaux

Pour ces travaux, mes perspectives à court terme sont les suivantes :

- Concernant l'algorithme List-Right, nous n'avons trouvé aucun graphe pour lequel l'algorithme Edge Deletion possède un meilleur comportement en moyenne, nous amenant à penser que pour tout graphe, List-Right domine en moyenne l'algorithme Edge Deletion. Il faudrait donc démontrer cette conjecture.
- Porter ces différents résultats sur d'autres problèmes d'optimisation combinatoire comme par exemple le problème de l'ensemble indépendant.
- Concernant le travail sur les groupes dynamiques, il serait intéressant d'explorer de nouveaux modèles probabilistes et effectuer une analyse similaire en considérant cette fois-ci la distance moyenne des membres dans l'arbre plutôt que son diamètre.
- Observer le comportement moyen de nouveaux algorithmes de connexion de groupe pour les comparer à celui que nous avons fourni, de la même façon que ce que nous avons fait pour le vertex cover.

À plus long terme, je souhaite explorer de nouveaux domaines en m'intégrant dans de nouveaux projets. Par exemple, je souhaite étudier les techniques de résolution exacte (dites de « kernelization ») pour le vertex cover (ou tout autre problème NP-complet). Je suis aussi relativement intéressé par les grands graphes, les réseaux dynamiques et les algorithmes distribués.