Research Statement

Pushkar Tripathi

My research is primarily focused on the study of allocation problems. A common characteristic of the real world instantiations of these problems is the lack of complete and accurate information about the input. For example, in many practical settings the input may be revealed over time (streaming), or may be drawn from a distribution that may not even be known in advance. Alternately, the cost of acquiring the entire input may be prohibitively large making it infeasible to use conventional methods that require access to the entire input at once. Modeling and addressing these concerns has been the central theme of my work.

Towards this end, I have designed approximation algorithms for several online assignment problems. I have also studied allocation problems in the stochastic paradigm. Another line of my work relates to testing the limits of approximability of various classical allocation problems under general objective functions.

Streaming Input Model

Online Bipartite Matching: This problem involves a bipartite graph $G(L, R, E)$, with one side L (ads, jobs, or items to sell, in different motivating examples) known beforehand to the algorithm, and vertices from the other side R (ad-slots, job-candidates, or buyers) arriving one by one online. When a vertex $r \in R$ arrives, its incident edges are revealed, and the algorithm can match it to some currently unmatched neighbor in L. The objective is to maximize the size of the matching obtained at the end.

This problem was first studied by Karp et al. in [17] where they gave an optimal $1-1/e$ competitive algorithm. The stochastic variant of this problem where the arriving vertices are drawn from a known distribution has received considerable attention [3, 8, 11, 18] of late and the $1-1/e$ barrier has been breached for this model. However all prior work on this problem required access to the underlying distribution from which incoming vertices are drawn. Furthermore, this distribution is assumed to have a small support.

In [16], along with Mehta and Karande, I studied this problem when the arriving vertices are drawn from a fixed but unknown distribution that may have an exponentially large support. We designed a randomized algorithm that attains a competitive ratio of 0.656. In fact our results hold in the more general *random order arrival* model (incoming vertices arrive in random order), and are based on exploiting certain symmetry properties of the worst input. We also showed an example to establish that our analysis was off by a factor of at most 0.03.

Online Budgeted Allocation (OBAP): In this problem we are given a set of agents each of which specifies a budget denoting her maximum spending capacity. Items arrive online and each agent places a bid on them. Each arriving item can be allocated to at most one agent and the agent is charged her bid for the item from her budget. The objective is to maximize the amount of money spent by the agents.

Apart from the online bipartite matching problem mentioned above this problem also generalizes two well studied problems -

- Online Vertex Weighted Matching (refer [1]) : This is a generalization of the online bipartite matching problem where the vertices on the left have weights and the objective is to maximize the weighted sum of the matched vertices.
- Adwords with small bids (refer [5, 21]) : The setup is identical to the OBAP with the added constraint that each of the bids is much smaller than the budget.

However unlike the aforementioned problems, OBAP is known to be NP-hard even for the offline setting (refer [6]). This excludes traditional approaches for analyzing online algorithms for this problem that rely on linear programming and the primal dual schema.

One can show that the greedy algorithm that assigns each item to the agent who can pay the most for it attains a factor of 0.5. Beating this bound has been a long standing open problem [19, 20, 21]. In [24] I present an algorithm that attains a competitive ratio of $0.5 + \epsilon$ for some fixed constant ϵ that is bounded away from 0. The analysis relies on the randomized primal dual technique that was recently introduced by Jain and Devanur [9].

Ongoing Work: I am currently working with László Végh on improving my result in [24] for the OBAP. We strongly believe that there exists an algorithm that attains a factor of $1 - 1/e$ for the OBAP which would present a unified framework for understanding online allocation problems and yield valuable insights in to their structure.

Future Work: An interesting generalization of OBAP is the online submodular welfare maximization problem. In this problem we are given a universe of items and a set of agents. Each agent has a monotone submodular utility function over the set of items. The items arrive online and we are required to assign them to the agents as they arrive to maximize the total utility derived by the agents. Clearly the problem generalizes the OBAP. The problem is hard to approximate to any factor better than $1 - 1/e$ even in the information theoretic sense (not contingent on $P \neq NP$) [22], and unlike the OBAP there is no natural LP relaxation for this problem. Despite these challenges it is interesting to note that the greedy algorithm (assign each item to the agent who gets maximum incremental benefit) still attains a factor of 0.5. Beating this factor is a major open problem in this area.

Query Commit Model

Even though efficient combinatorial algorithms for the maximum matching problem have been known [10] for several years, in certain situations we are required to design algorithms that do not utilize global structures such as *blossoms* and are constrained to make decisions based solely on locally available information. For example consider the kidney exchange problem - Often patients with a kidney disease have a family member who is willing to donate his/her kidney. Unfortunately, these donors are sometimes blood-type incompatible. To solve this problem, a kidney exchange is performed in which patients swap their incompatible donors to get a compatible donor. Owing to the cost involved in medical tests, incentive issues, and due to ethical concerns, it is desired that an exchange is performed whenever the test indicates that the exchange is possible. This situation can be visualized as a maximum matching problem in a general graph in the query-commit model described below.

Oblivious Query Commit Model: Consider a general non-bipartite graph where each patient-donor pair represents a node of the graph and an edge between two nodes indicates if an exchange is possible. For every pair $(u, v) \in V \times V$ we are *not* told a priori whether there is an edge connecting these vertices, until we *probe/scan* this pair. If we scan a pair of vertices and find that there is an edge connecting them we are constrained to *pick* this edge and in this case both u and v are removed from the graph. However, if we find that u and v are not adjacent, they continue to be available to be matched in the future. The goal is to maximize the number of vertices that get matched.

It is easy to see that any algorithm that probes all permissible edges is a $1/2$ approximation since we are guaranteed to pick a maximal matching. One can also construct examples to show that no deterministic algorithm can do strictly better than $1/2$ on all instances. Thus our only hope is to seek a randomized algorithm that beats the barrier of $1/2$ in expectation.

In [13] together with Goel, I studied the following randomized algorithm for finding a large matching in a given graph - Shuffle the vertices according to a uniformly random permutation and iterate through them one at a time. If the current vertex is already matched then ignore it else scan edges incident to it in the order dictated by the permutation until it gets matched or there are no vertices left. Then proceed to the next vertex. Note that this algorithm can be simulated in the oblivious query-commit model by iterating through the vertices based on a randomly chosen ordering and then scanning all the edges incident on a vertex according to this ordering. We showed that this algorithm attains a factor of 0.5983.

Our algorithm is the first significant improvement on a 15 year old result by Aronson, Dyer, Frieze, and Suen in [2] where they analyzed a closely related randomized algorithm and showed that it attained a factor of 0.50000025. In their algorithm we pick a vertex at random and match it to one of its unmatched neighbors uniformly at random.

From the hardness perspective we considered a large family of algorithms (called vertex iterative algorithms), that iterate through the vertices and explore their neighborhood. We showed that no vertex iterative algorithm can attain a factor better than 0.75 for this problem.

Stochastic Query Commit Model: The query commit problem has also been studied in the stochastic setting in [4, 7, 23]. This setting is a relaxation of the oblivious model where we assume that for every pair of vertices we are additionally given the probability that they are adjacent. In joint work with Costello and Tetali, I gave a 0.671 factor algorithm for this problem using a novel sampling technique. Apart from this problem, I believe our sampling technique would find applications in other domains of approximation algorithms. We also showed that no randomized algorithm could achieve a competitive ratio greater 0.896 in this model.

Non-linear Objective Functions

A major focus of computer science has been on the study of linear cost functions owing to their simplicity and malleability. However linear cost functions do not always capture the intricate dependencies that exist in real-world settings. For example the marginal (incremental) cost of producing a good only decreases with increasing scale of production. This property cannot be modeled though linear cost functions. Therefore linear functions only serve as an approximation to the original functions, thus even though we may have a good approximation algorithm for solving the linear optimization problem the output is still suboptimal.

Another feature that arises in practical settings is the presence of multiple agents with possibly different cost functions who wish to collaborate to accomplish a certain task. My endeavor in this line of research has been to model these properties and assess the limits of approximability of classical optimization problems in this paradigm.

Combinatorial Optimization over Submodular Functions: Consider the settings where we are given a network, represented by a graph and multiple agents wish to collaborate to build a combinatorial structure, say spanning tree, over this graph. Each agent specifies a cost function over the edges of the graph. Motivated by properties such as decreasing marginal cost and economies of scale we use submodular functions to model the agents' cost functions. Our task is to minimize the total cost incurred by the agents for building this structure.

Even though this model generalizes the classical paradigm of linear cost functions these problems are not amenable to analysis by traditional techniques of combinatorial optimization. In [12], together with Goel and Wang, I studied the limits of approximability of several classical optimization problems in this setting and presented (tight) polynomial information theoretic lower bounds for them. The information theoretic lower-bounds indicated that it was the lack of information and not the combinatorial structure that was impeding the search for efficient algorithms. To rectify this, in a subsequent paper [14] with Goel and Wang, I studied a class of succinctly representable submodular cost functions called *discounted cost functions*.

In a discounted cost function we are given an additive weight function w that assigns a weight to every edge and an increasing concave function $d : R^+ \to R^+$. The cost of any subset S of edges is defined as $d(w(S))$. In [14], we gave tight logarithmic approximation algorithms for edge cover and spanning tree in this model. On the other hand we also proved that perfect matching and shortest path are hard to approximate to within a polylog factor even in this model.

Minimum Linear Ordering Problems (MLOP): Linear ordering problems can be used to model several allocation problems such as Min Sum Set Cover, Minimum Latency Set Cover, Multiple Intents Ranking, Minimum Linear Arrangement. In its most general form we are given a nonnegative set function f on a finite set V and our task is to find a linear ordering on V such that the sum of the function values for all the suffixes is minimized.

In [15], together with Iwata and Tetali, I studied this problem for several broad classes of the function f. We showed that the greedy algorithm provides a factor 4 approximate optimal solution when the cost function f is supermodular. We also presented a factor 2 rounding algorithm for MLOP with a monotone submodular cost function, using the convexity of the Lovász extension. These are among very few constant factor approximation algorithms for NP-hard minimization problems formulated in terms of submodular/supermodular functions. In contrast, when f is a symmetric submodular function, the problem has an information theoretic lower bound of 2 on the approximability. In addition, we also gave a randomized rounding algorithm for the Min Sum Vertex Cover problem of factor 1.79, improving over the factor 2 algorithm described by Feige, Lovász, and Tetali (2004).

Final Words

In the future I would like to continue working in the areas of randomized and approximation algorithms. Despite finding applications in several areas of algorithm design, randomization has not been fully exploited in the context of primal dual algorithms. In particular, there are no known primal dual algorithms that use randomization in the dual ascent stage. Motivated by this, and in light of my recent results in [24], I am interested in combining techniques from the analysis of randomized algorithms with those used in designing primal dual algorithms. Moreover, I am determined to explore connections between my areas of expertise and other research areas and look forward to collaborating with researchers working in completely new fields.

References

[1] Gagan Aggarwal, Gagan Goel, Chinmay Karande, and Aranyak Mehta. Online vertex-weighted bipartite matching and single-bid budgeted allocations. *SODA*, 2010.

[2] Jonathan Aronson, Martin Dyer, Alan Frieze, and Stephen Suen. Randomized greedy matching. ii. *Random Struct. Algorithms*, 6:55–73, January 1995.

[3] Bahman Bahmani and Michael Kapralov. Improved bounds for online stochastic matching. *ESA*, 2010.

[4] Nikhil Bansal, Anupam Gupta, Jian Li, Julián Mestre, Viswanath Nagarajan, and Atri Rudra. When lp is the cure for your matching woes: improved bounds for stochastic matchings. In *Proceedings of the 18th annual European conference on Algorithms: Part II*, ESA'10, pages 218–229, 2010.

[5] N. Buchbinder, K. Jain, and J.S. Naor. Online Primal-Dual Algorithms for Maximizing Ad-Auctions Revenue. In *Algorithms–ESA 2007 15th Annual European Symposium, Eilat, Israel, October 8-10, 2007: Proceedings*, page 253. Springer, 2007.

[6] Deeparnab Chakrabarty and Gagan Goel. On the approximability of budgeted allocations and improved lower bounds for submodular welfare maximization and gap. *SIAM J. Comput.*, 39:2189–2211, March 2010.

[7] Ning Chen, Nicole Immorlica, Anna R. Karlin, Mohammad Mahdian, and Atri Rudra. Approximating matches made in heaven. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming: Part I*, ICALP '09, pages 266–278, 2009.

[8] N. Devanur and T Hayes. The adwords problem: Online keyword matching with budgeted bidders under random permutations. In *ACM Conference on Electronic Commerce*, 2009.

[9] Nikhil Devanur and Kamal Jain. Understanding karp-vazirani-vazirani's online matching (1990) via randomized primal-dual. Approximation Algorithms Workshop, Princeton, USA, 2011.

[10] Jack Edmonds. Paths, trees, and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.

[11] Jon Feldman, Aranyak Mehta, Vahab S. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *FOCS*, pages 117–126, 2009.

[12] Gagan Goel, Chinmay Karande, Pushkar Tripathi, and Lei Wang. Approximability of combinatorial problems with multi-agent submodular cost functions. *Foundations of Computer Science, Annual IEEE Symposium on*, pages 755–764, 2009.

[13] Gagan Goel and Pushkar Tripathi. Matching with your eyes closed. Manuscript, 2011.

[14] Gagan Goel, Pushkar Tripathi, and Lei Wang. Combinatorial Problems with Discounted Price Functions in Multi-agent Systems. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2010)*, volume 8 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 436–446, Dagstuhl, Germany, 2010. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[15] Satoru Iwata, Prasad Tetali, and Pushkar Tripathi. Approximating minimum linear ordering problems. Manuscript, 2011.

[16] Chinmay Karande, Aranyak Mehta, and Pushkar Tripathi. Online bipartite matching in the unknown distributional model. In *STOC*, pages 106–117, 2011.

[17] R.M. Karp, U.V. Vazirani, and V.V. Vazirani. An optimal algorithm for online bipartite matching. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing*, 1990.

[18] Vahideh H. Manshadi, Shayan Oveis-Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. In *SODA*, 2011.

[19] Aranyak Mehta. Online matching and adwords. ARC Colloquium, Georgia Institute of Technology, Atlanta, USA, 2011.

[20] Aranyak Mehta. Online matching and adwords. Approximation Algorithms Workshop, Princeton, USA, 2011.

[21] Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. In *FOCS*, 2005.

[22] Vahab Mirrokni, Michael Schapira, and Jan Vondrak. Tight information-theoretic lower bounds for welfare maximization in combinatorial auctions. In *Proceedings of the 9th ACM conference on Electronic commerce*, EC '08, pages 70–77, New York, NY, USA, 2008. ACM.

[23] Marco Molinaro and R. Ravi. The query-commit problem. Manuscript, `arxiv.org/abs/1110.0990`, 2011.

[24] Pushkar Tripathi. Algorithm for the online budgeted assignment problem. Manuscript, 2011.