Algorithms and Games on Social Networks

1 Overview

The Web is one of the most complex human creations, and among its myriad characteristics, the *socio-economic* aspect of the Web stands out as particularly intricate: the amazing growth of electronic commerce through sponsored search, Amazon and eBay, the unexpected success of Wikipedia, and the ever-increasing popularity of social networks such as Facebook represent socio-economic interactions at a scale much larger than could even be imagined up until a couple of decades ago. Further, the content in many of these entities is created primarily by independent users with diverse economic interests, who work in collaboration, and often competition, with each other. Hence, the mathematics required for understanding these aspects of the Web must lie on the intersection of Game Theory, Economics, and especially theoretical Computer Science.

My research agenda is to advance the understanding of the Web by contributing mathematical insights, drawing tools from the above-mentioned areas. An increased understanding of these aspects would help make the present Web economically efficient, and in the long term this would undoubtedly give rise to new socio-economic entities. This line of study also brings the promise of rich theoretical insights, as classical theories need to be re-interpreted in terms of the scale, noise and dynamic nature inherent to the Web as well as its networked nature.

2 Social Network Theory[1]

The recent emergence of social networks such as Facebook, Orkut and MySpace and their amazing popularity has resulted in a new and wonderfully rich dataset for science. Users put detailed profiles containing personal information on these networks, and identify other network users as “friends”, resulting in a richly annotated and intricate graph of social connections. Not surprisingly, this is a gold mine of information for businesses such as online advertisers and accordingly, social networks with large number of users are highly valued. Although targeted advertising is the main source of revenue for these networks, the current advertising technologies used by the networks are modeled on the sponsored search paradigm of contextual advertising and do not leverage the networked nature of the data effectively.[2] In particular, network topology information could potentially be used for targeting advertisements in a more effective manner, which is a situation that would benefit all the three participants in the network: users, advertisers and the social network. To fulfil this need for alternative monetizing schemes for social networks, I (along with my co-authors) have proposed such a scheme in a recent work [1]. Network topology information also implies a “power” structure among the users in the network, and the theoretical study of economic interactions on the network can be very helpful in understanding these properties.

2.1 Marketing on Social Networks

2.1.1 Current Research

There have been several attempts at devising monetizing techniques for social networks [5,7,14], but most of these proposals are based on direct user advertising. The impetus for our investigation came from searching for a monetizing scheme which would exploit the network structure to derive revenue from implementing viral marketing schemes over social networks A unique and important aspect of the marketing scheme we propose is that it distributes a portion of the revenue generated through the sale back to the users of the network, incentivizing them to participate in the process.

The proposed monetization scheme works by selling products on social networks and proposes marketing strategies which can exploit the structure of the underlying social network to optimize revenue generated through the sale. Specifically, consider a seller interested in selling a good or a service on the social network.

[1] Social network theory is a relatively new area that refers to the mathematical study of interactions between strategic agents on a social network. The name for the area is not widely used yet since it is a new field [15].

[2] We substantiate these claims by citing the example of Facebook which has presently over 70 million users [17]. Facebook was valued at $15 billion by Microsoft in 2007 [12], but its estimated revenue for 2008 was $300 million [17]. The Beacon advertising system from Facebook does uses the network to target advertisements, but it has been the success Facebook had hoped for and has been beset with privacy concerns [16].

Once a sale has been made, the seller seeks to capitalize on the *influence* the buyer has on her friends, and asks the buyer to recommend[3] the product to her friends. In return for the recommendation, the seller promises the buyer some cash (called a *cash-back*) if her friends end up buying the product. We assume that each buyer sends a recommendation to all of her friends in the hope that it would receive some money for her efforts. Note that the payment is not certain as a person may choose not to buy a product even on receiving a recommendation. This is formalized using a probabilistic model that encodes a user's probability of buying as a function of the recommendations it receives and the price offered for the product. A natural way to think about the probability is to assume that the probability of buying increases with a decrease in price (the actual model actually does not use this assumption) and increase in recommendations. Thus, the random process of the sale of the product cascades from the initial set of buyers on the network to their friends, the friends of their friends and so on.

Since the seller can choose the price at which a product is offered to the recipient of the recommendation, a far-sighted seller can trade-off influence and price in his marketing strategy. For instance, it might offer the product at a low price early on, thereby increasing the probability of purchase and later leverage the influence created by the initial sales to charge higher prices. An obvious question arises for the seller who thinks of implementing such a strategy: *How does one choose prices for the users of the network?* Note that it is not clear if a user will be offered different prices at different times in the sale (an *adaptive* strategy), or fixed in advance before the marketing process starts (called a *non-adaptive* strategy).

In our paper [1], we show that computing the optimal (one that maximizes expected revenue) non-adaptive strategy for a seller in this setting is NP-Hard. In a positive result, we show that there exists a non-adaptive strategy for the seller such that the expected revenue for the seller that is within a constant factor of the expected revenue of by the optimal *adaptive* policy (showing that the adaptivity gap is constant). This strategy is based on an *influence-and-exploit* policy which computes a *max-leaf spanning tree* of the graph, and offers the product to the interior nodes of the spanning tree for free, later on exploiting this influence by extracting profit from the leaf nodes of the tree. We also show that the approximation guarantee of the strategy holds for fairly weak conditions on the function which specifies the dependence of probability of acceptance of an offer on the price.

2.1.2 Directions for Future Research

The model described in the previous section is limited because it does not provide a lot of strategic flexibility to people in terms of *when* they choose to buy a product. For instance, consider a networked game being sold on a network-enabled device, such as the iPhone. Then, at any given time, a person can either buy the product then or choose to *wait-and-watch*. The trade-off that governs the person's decision is between garnering social status and deriving maximum utility from the product itself: buying early would enable the person to have a higher score (or status) in the network than people who join late. On the other hand, buying at a later stage would have more utility for the person as the value of the game would be higher to a person if more of the person's friends also have the game.

Even this extremely simple model can result in counter-intuitive game dynamics due to specific network topology. For instance, consider the following situation in a cyclic network with three users: one of the users has the product and recommends it to both the other users. But the parameter values are such that the two users need each other to buy the product before the usefulness of the product justifies its cost. On the other hand, if both the nodes were to buy the product simultaneously, they would be happier than not buying as then their utility will be higher. Note that this situation is describing a Nash equilibrium of the game, albeit it is an equilibrium the seller would want to avoid. In this setting, the main question we want to understand is: *Can one characterize the Nash equilibrium of the game w.r.t. the price specified by the seller?* If one can get such a characterization, the next logical step would be to ask: *What price of the product maximizes the seller's revenue, and can this price be computed efficiently?*

These are important questions to understand by themselves, but are by no means isolated problems. Indeed, considering nodes in the network to be strategic opens up a host of other natural questions. For instance, note that an important assumption in the model described in the earlier section is that the nodes do not have any *cost* of sending a message (cost here is a proxy of being perceived as sending spam messages

[3] A recommendation can be implemented easily on the online social networks as an "email to friends" option after a buyer completes a purchase.

to your friends). If we consider such a cost, observe that the nodes will self-select recommendation receivers. A promising research direction is to understand how this would affect the seller's revenue, and how will the seller's pricing strategy change in this case. I am currently thinking about the above two problems, and believe that the combination of game theory with networks is a nascent area that offers many unexplored and promising directions for research.

2.2 Bargaining on Social Networks

One of the questions I believe to be central in the domain of social network theory is to understand the effect of network topology on various network processes, such as the marketing process described in the previous section. Economic interactions on the network are crucial in furthering this understanding, and this section is about a fundamental type of interaction: bargaining between players. A bargaining problem is one in which two (or more) players decide whether to deal with each other (e.g., split a sum of money), where each of the players can choose to *walk away* from the deal in which case they have an alternate deal (e.g., a fall back amount).

Bargaining is one of the oldest and most fundamental problems in game theory that has wide applicability in real-life situations such as retail markets and business contracts [13]. Bargaining games have been extensively studied by economists and sociologists since a solution for the two person setting was described in a seminal paper by Nash [11]. A special class of bargaining games is where players are connected through an underlying (social) network and can only interact with their neighbors in the network. These have been particularly well-studied in sociological literature, and there is now a large body of experimental work in this area of Network Exchange Theory [18]. Recently, there has also been increasing interest in mathematical models of network exchange games, especially from a computational point of view [3, 4, 9]. This study of bargaining on social networks is useful in providing insights into the "power" of a node's position in the network, and provides a beautiful instance of local interactions among nodes converging to global economic state of the network. The computational point of view is quite pertinent here, as powerful techniques from graph theory (such as theory of matchings) can help characterize the outcome of the interactions on the network [9].

The present literature on computational aspects of bargaining problems is extremely thin, especially when compared to more than half century of work done in economics on bargaining. The first challenge for computational scientists is to understand the computational requirements of the various bargaining solutions that have been proposed in the economics literature. For instance, Nash Bargaining and Proportional Bargaining are two main classes of solutions that have been studied extensively by economists. My current line of research examines the Nash bargaining solution proposed in [9] closely, and derives more insights into the solution using tools from matching theory. The research direction I plan to pursue in the future is to examine the implications of bargaining for social networks. In particular, I believe that applying the insights learnt from bargaining to formulate and revisit areas of social network theory such as network formation would be valuable research.

3 Algorithms for the Web

My stated research agenda of contributing to the understanding of the Web through algorithmic ideas has led me to investigate a variety of problems. In all the problems I have worked on, I have found tools from algorithms to be extremely useful in characterizing and finding efficient solutions to problems. In particular, approximation and random sampling are indispensable for working with web scale data that is noisy and constantly changing. The following sections describe two such problems, both of which are about seemingly different domains, but with the common theme of having efficient algorithms for the Web.

3.1 Diversification of Search Results

Search is one of the primary user activities on the Web, and despite the unequivocal success of search engines, search is still believed by many to be an open problem. Part of the reason behind this belief is that search engines do not always understand *user intent* behind a query, and this is key to designing an effective ranking system in a search engine. Without any explicit knowledge of user intent, search engines want to diversify results to improve user satisfaction. In such a setting, the probability ranking principle-based approach of presenting the most relevant results on top can be sub-optimal, and hence the search engine would like to trade-off relevance for dissimilar results.

In joint work with Sreenivas during the summer of 2008 at Microsoft Research [6], we examined a general framework for diversification systems. In the context of web search, given a query, our diversification system for web pages consists of three separate parts: a relevance function for the web pages given the query, a pairwise similarity function between web pages and a diversification function which produces a diverse ranking given the relevance and distance function. The main contribution of the work is two criteria for selecting among the various diversification objectives that can be thought of for such a framework: a mathematical characterization of diversification systems in terms of axioms, and an evaluation methodology based on a data set derived from the disambiguation pages listed in Wikipedia. The proposed mathematical characterization consists of a set of natural axioms that a diversification system is expected to satisfy, and we proved that no diversification function can simultaneously satisfy all the axioms. We also proposed a novel experimental evaluation methodology for the a diversification function using a data set derived from the disambiguation pages listed in Wikipedia. Further as a proof-of-concept, we provided three examples of diversification objectives that satisfy a different subset of the axioms. We showed that these objectives can be reformulated as different flavors of the well-studied facility dispersion problem, which characterizes their computational complexity and also provides approximation algorithms. We also implemented these algorithms for our objectives and evaluated them using the proposed methodology. The results from the investigation have been very positive and demonstrate the effectiveness of the framework in distinguishing between various criteria.

The primary advantage of this framework is its modular nature, which coupled with the formal guarantees can encompass a wide range of situations. In particular, this work will serve as a basis for implementing a new product diversification system that would be used in Live Search.

3.2 Reverse Nearest Neighbors

Proximity problems have found wide applicability in information retrieval, including answering queries, duplicate detection and finding pages similar to a given page. In particular, the setting of many of these problems is as follows: there is a large dataset that one can pre-process. Then during the "online" phase, one would like to answer a large number of proximity queries with some speed-up gained due to the pre-processing. The setting can be abstracted as follows: the dataset (set of documents) and the query are points in some metric space, i.e. there is a notion of pairwise distance between the documents that constitutes a metric. A typical example of a proximity query here is the Nearest Neighbor ($\mathcal{NN}$) query, which seeks the find the *nearest* (minimum distance) point in the dataset to a given query point.

The Reverse Nearest Neighbor ($\mathcal{RNN}$) is the inverse problem of $\mathcal{NN}$, and seeks to retrieve the points in a given point set that have the query point as their nearest neighbor. The $\mathcal{RNN}$ problem was introduced in [10], and finds applications in database systems, marketing and maintaining nearest neighbor structures. Although many heuristic approaches have been proposed for the $\mathcal{RNN}$ problem, the algorithmic complexity of this problem remained an open question. I resolved this question in recent joint work with David and Steve [2]. We show that during the pre-processing stage, the $\mathcal{RNN}$ problem can be reduced to the $\mathcal{NN}$ problem in any metric space, which is a surprising observation as no formal relationship had hitherto been known between these problems. We further proved that this reduction, along with a simple extension of the locality-sensitive hashing technique [8] yields a sub-linear time algorithm for online $\mathcal{RNN}$ queries. Note that the output of the $\mathcal{RNN}$ query is a *set of points*, as against a single point in the case of the $\mathcal{NN}$ problem. A nice (and practically useful!) outcome of our algorithm is that the query time is sensitive to the size of the output set in the best possible manner, i.e. in terms of an additive factor.

The $\mathcal{RNN}$ problem is a generic problem that appears in a surprisingly wide variety of areas, and this work would enable researchers to apply our algorithmic ideas to their specific problem domain.

4 Final Words

Social network theory is an interdisciplinary area that brings together a diverse set of ideas towards understanding important problems on the Web. My perspective as an applied mathematician complements the traditional development of network exchange theory as an experimental science, and enables me to apply powerful mathematical techniques from economics, game theory and theoretical computer science towards solving these problems. I am excited to be working on this area for my thesis research, and intend to contribute to its growth by developing elegant models and creatively developing new algorithms and theory for these important problems.

References

[1] David Arthur, Rajeev Motwani, Aneesh Sharma, and Ying Xu. Pricing strategies for viral marketing on Social Networks, under submission.

[2] David Arthur, Steve Oudot, and Aneesh Sharma. Approximating Reverse Nearest Neighbors using Locality Sensitive Hashing, under submission.

[3] N. Braun and T. Gautschi. A Nash bargaining model for simple exchange networks. *Social Networks*, 28(1):1–23, 2006.

[4] Tanmoy Chakraborty and Michael Kearns. Bargaining Solutions in a Social Network. *Workshop on Internet and Network Economics*, page To appear, 2008.

[5] Stephen Foley. Google searches and finds myspace deal. `http://www.independent.co.uk/news/business/news/google-searches-and-finds-myspace-deal-411027.html`, 2006.

[6] Sreenivas Gollapudi and Aneesh Sharma. An axiomatic approach for result diversification. *To appear in the Proceedings of the 18th international conference on World Wide Web*, 2009.

[7] J. Hartline, V. Mirrokni, and M. Sundararajan. Optimal Marketing Strategies over Social Networks. *Proceedings of the 17th international conference on World Wide Web*, 2008.

[8] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 604–613, 1998.

[9] J. Kleinberg and É. Tardos. Balanced outcomes in social exchange networks. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 295–304. ACM New York, NY, USA, 2008.

[10] F. Korn and S. Muthukrishnan. Influence sets based on reverse nearest neighbor queries. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 201–212. ACM New York, NY, USA, 2000.

[11] John Forbes Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

[12] BBC News. Facebook valued at $15 billion. `http://news.bbc.co.uk/2/hi/business/7061042.stm`, 2007.

[13] M.J. Osborne. *An introduction to game theory.* Oxford University Press New York, 2004.

[14] Katharine Q. Seelye. Microsoft to provide and sell ads on facebook. `http://www.nytimes.com/2006/08/23/technology/23soft.html`, 2006.

[15] Siddharth Suri. Research on social network theory. `http://www.cs.cornell.edu/~ssuri/About_Me_files/SidSuriRS.pdf`, 2008.

[16] Wikipedia. The beacon advertisement system. `http://en.wikipedia.org/wiki/Facebook_Beacon`, 2008.

[17] Wikipedia. Facebook revenue in 2008. `http://en.wikipedia.org/wiki/Facebook`, 2008.

[18] D. Willer, editor. *Network Exchange Theory.* Praeger/Greenwood, 1999.