Research Contributions and Its Significance

Swaprava Nath

Email: swaprava@gmail.com Web: http://swaprava.byethost7.com

My research has contributed to the game theoretic modeling of Internet related microeconomic applications and presented mechanism design solutions. In particular, I have focused on the *crowdsourcing* problem. The art of aggregating information and expertise from a diverse population has been in practice since a long time. The Internet and the revolution in communication and computational technologies have made this task easier and given birth to a new era of online resource aggregation, which is now popularly referred to as crowdsourcing. Even though crowdsourcing had been in practice implicitly in many classic applications including the Oxford English Dictionary, it has gained much attention in recent times after the success of Wikipedia, oDesk, Amazon Mechanical Turk, InnoCentive, and several other such applications. Today, crowdsourcing is not only limited to small online games, rather it is commercially used as active means to generate revenue for organizations. With the proliferation of this unique tool, one has to understand two important features of this aggregation technique: (a) crowdsourcing is always human driven, hence the participants are rational and intelligent, and they have a payoff function that they aim to maximize through their choice of actions, and (b) the participants are connected over a social network which helps both the designer and the participants to reach out to a large set of individuals. To understand the behavior and the outcome of such a strategic crowd, we need to understand the economics of a crowdsourcing network using game theoretic modeling. This is where my research contributes significantly to the existing literature. In my thesis, I have considered the following three major aspects of the strategic crowdsourcing problem as shown in Figure 1.

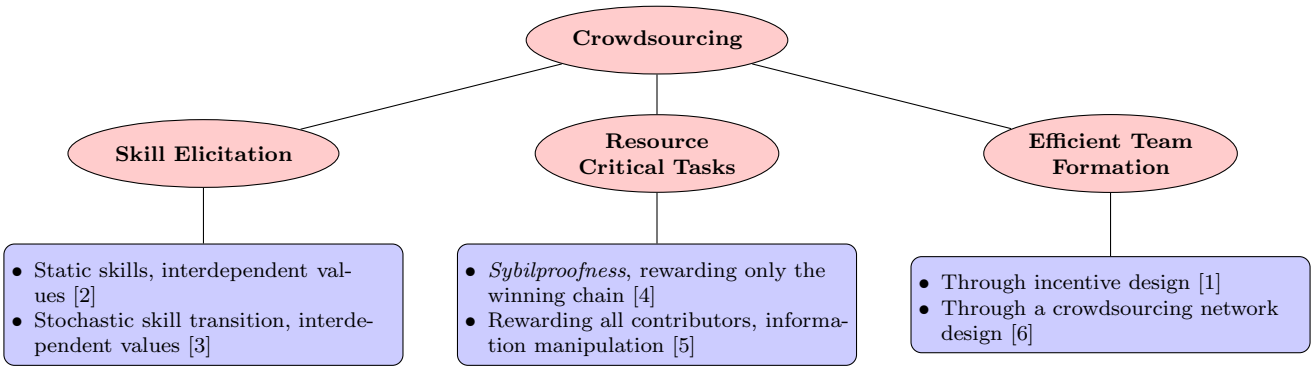


Figure 1: Research components of the crowdsourcing problem.

Eliciting Skills of the Strategic Workers

Since the crowd is heterogeneous, the skill levels of the participants is often unknown to the crowdsourcer, who wants to select an optimal subset of the participants based on that skill. However, the participants may perfectly know their individual skills (hidden from the crowdsourcer) and this induces a game between the crowdsourcer and the participants. We designed mechanisms that help elicit the skills truthfully from the participants ensuring their propensity to play this game.

- First, we addressed the mechanism design problem in a static setting, that is, the private skills are invariant over

time. We provided a mechanism that makes truthful reporting a strictly better strategy for all agents and also guarantees individual rationality under a special setting [2].

- Next, we extended the solution to a setting where the participants' qualities vary stochastically over time according to a Markov process [3].

Resource Critical Task Execution via Crowdsourcing

This applies to the crowdsourcing contests like finding an object or an answer. The objective is not only to find the correct answer in a time optimal or cost minimal fashion, but also to find the right structure of the crowdsourcing network that found the solution. We designed mechanisms that incentivizes the individuals on the network to meet the right objectives.

- We have shown that in crowdsourcing contests like the DARPA red balloon challenge, certain desirable properties are impossible to satisfy together. We introduced approximate versions of these properties and provided a first attempt in the context of crowdsourcing to design *approximately sybilproof mechanisms* [4].
- In these kinds of contests, a great deal of human effort can get wasted, as people can potentially go explore the same incorrect solution already explored by someone else. We discussed about how a *synergistic* mechanism can be designed using the tools from information theory and prediction market to mitigate this problem [5].

Efficient Team Formation from a Crowd

The third direction considers the whole crowdsourcing network as a consolidated organization, where the designer aims to maximize the net productive output of the system. We have taken two complementary approaches here.

- First, we analyzed how individuals connected in a network trade-off between their production and communication effort given the *network positions* and the *reward sharing scheme*, and show how the reward sharing scheme can be designed optimally so as to maximize the total output of the network [1].
- On the other hand, we showed how to *design the network* that leads to an optimal performance [6].

Putting everything together, my research addressed several interesting problems on a very recent research area named crowdsourcing and provided novel solutions to its economic aspect, which makes my research clearly distinguishable in the literature that lies in the intersection of computer science and microeconomics.

References

[1] S. Nath and B. Narayanaswamy. Improving Productive Output in Influencer-Influencee Networks. *arXiv preprint arXiv:1302.3045*, 2013.

[2] S. Nath and O. Zoeter. A Strict Ex-post Incentive Compatible Mechanism for Interdependent Valuations. submitted to Economics Letters, May 2013.

[3] S. Nath, O. Zoeter, Y. Narahari, and C. Dance. Dynamic Mechanism Design for Markets with Strategic Resources. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI 2011)*, pages 539–546, 2011.

[4] S. Nath, P. Dayama, D. Garg, Y. Narahari, and J. Zou. Mechanism Design for Time Critical and Cost Critical Task Execution via Crowdsourcing. In *Proceedings of Internet and Network Economics (WINE 2012)*, pages 212–226, 2012.

[5] S. Nath, P. Dayama, R. D. Vallam, D. Garg, and Y. Narahari. Synergistic Crowdsourcing. Working Paper, 2012.

[6] S. Nath, B. Narayanaswamy, K. Kandhway, B. Kotnis, and D. C. Parkes. On Profit Sharing and Hierarchies in Organizations. Presented in Asian Meeting of the Econometric Society (AMES), December 2012.

Proposed Plan of Research

Swaprava Nath

Email: swaprava@gmail.com Web: http://swaprava.byethost7.com

The focus of my research lies in the broad area of *Internet economics*. In particular, I work on the theory and applications of microeconomics that actively involve computational science and the World Wide Web as a method of implementation. My research explores solutions against the strategic behavior of the individuals in these applications, and uses game theory and mechanism design to yield efficient and robust solutions. My Ph.D. thesis has considered one such application, namely *crowdsourcing*, and provided solutions to several important questions in different aspects of crowdsourcing. This proposal discusses the plan of a larger class of problems that I would like to pursue in future. My analysis techniques involve real analysis, optimization, probability theory, and similar mathematical treatment. In the following, I am going to motivate my research agenda with suitable examples, state the research objectives and the plan of the methods to be used, and finally, present its merit in the current research state-of-the-art.

1 Motivation

Innovations in the computational and communication sciences have revolutionized the way we talk, disseminate information, aggregate our opinions or beliefs, and even do trade with the consumers. A large section of the Web today is involved in generating content by the users, outsourcing tasks to individuals in a crowd, and generating revenue through advertising. Technology has made this job easier by providing an atmosphere where these applications can run efficiently. However, one can observe that even though the applications are run by machines, the brain behind every application or web content is human. Being rational and intelligent, the human users choose their actions to optimize their individual objectives. More generally, all human driven applications on the Internet have a certain pattern and that is why it needs a detailed microeconomic study for a better performance.

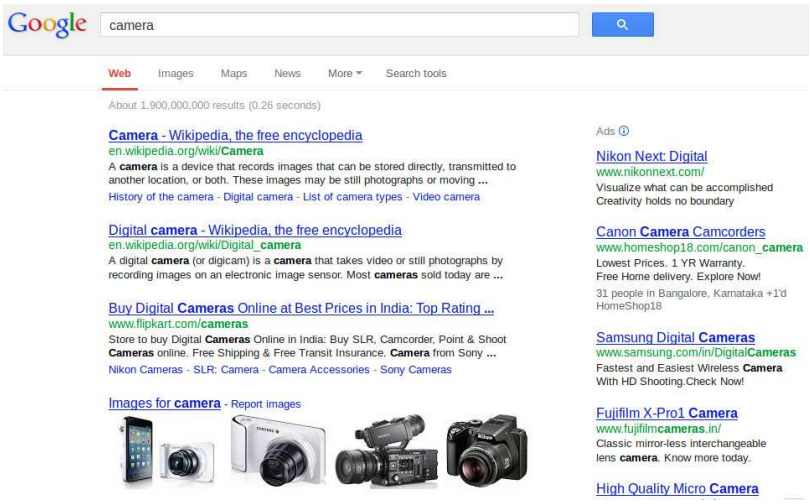
Game theory and mechanism design are the tools from microeconomics tailored to address the questions of this kind. My research develops theories that help explain certain phenomena on the Internet-related applications in the light of game theory and mechanism design. Before going forward, let us look at a few examples that illustrate the need for an economic insight into those applications.

1.1 Crowdsourcing: Wikipedia

Wikipedia is a classic example of harnessing the knowledge of a crowd. Conceptualized by opensource stalwart Richard Stallman in December 2000, Wikipedia was officially launched on January 15, 2001 by Jimmy Wales and Larry Sanger, using the concept and technology of a 'wiki' pioneered in 1995 by Ward Cunningham. As opposed to the expert based system, the idea of letting the crowd edit and control the quality of the content revolutionized the content generation process for Wikipedia. As of May 2013, Wikipedia includes over 26 million freely usable articles in 285 languages, written by over 39 million

registered users and numerous anonymous contributors worldwide [3]. According to Alexa-Internet [1], Wikipedia is the world's sixth-most-popular website, visited monthly by around 11% of all Internet users. To understand why people contribute to Wikipedia, we need to understand the behavioral economics of this domain. [1]

1.2 Online Advertising: Sponsored Search Auctions



A significant portion of today's web traffic is directed to the search operation. Google is a pioneer of the web search applications. However, a major effort goes in earning revenue from this application as the sponsored advertisements which show up during the search operation yield money to Google whenever any user clicks on those ads (see the figure). The ads are placed according to the outcome of an auction that Google runs with those advertisers. These auctions are called the *sponsored search auctions*. The placement of the ads and the proper payment schemes are questions that are efficiently answered by mechanism design. [2]

There are several other application domains, e.g., product recommendation systems, user reviews, resource allocation, voting, matching etc. that are part of the bigger class of problems called *Internet economics*. Figure 1 shows an important subset of problems in this class that my research addresses. The following section provides a detailed description of the sub-domains.

2 Objectives

The following sub-domains of Internet economics, as shown in Figure 1, offer a bunch of important questions that are the need for the hour of the Web applications. This proposal aims to sequentially address all of these problems.

- **Crowdsourcing**: This domain addresses the problem of efficiently harnessing information, expertise, and labor from a diverse, unstructured, and heterogeneous *crowd*, often with a suitable monetary compensation. The basic research questions that lie in this space are: (a) how to pick the right set of experts as their skill set is unknown to the designer, (b) what are the limits of achievability when the task is resource critical, (c) how to align the incentives so that some achievable global objectives are satisfied.
- **Online advertising**: A major source of generating revenue in the Web is via advertising. Ranging from (a) sponsored search auctions to (b) recommendation systems, the basic philosophy is to understand the behavior of the users and to show appropriate ads and properly charging the advertisers. A related domain is to advertise in the (c) user review forums depending on the topics of discussion.

[1] Image courtesy: http://www.wikipedia.org/
[2] Image courtesy: http://www.google.com/

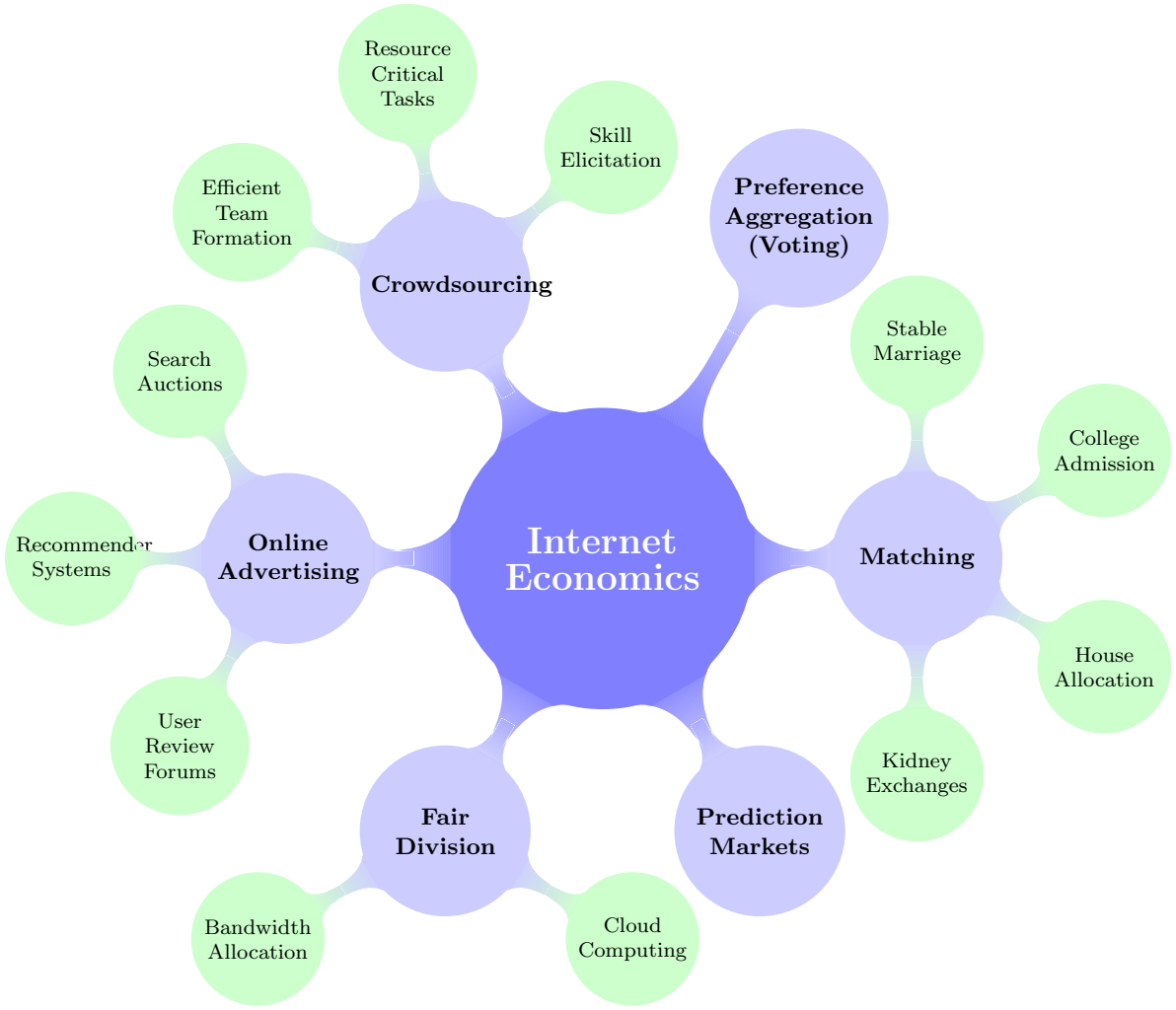


Figure 1: Graphical overview of the major research components of Internet economics.

- **Fair division**: Even though this is a classical economic problem, the bandwidth allocation problem in shared Internet access points or the resource sharing in cloud computing bring us back to the fair division problem.
- **Prediction markets**: These markets allow users to buy and sell securities on potential outcomes of an uncertain event, and helps the designer determine the probability of the events by efficiently aggregating the knowledge and information of the participants. The goal here is to elicit the *true* information from the agents.
- **Matching**: Even though the problems of matching (a) students to colleges, (b) houses to its potential occupants, (c) kidney donors to its acceptors, or (d) men with women for marriage are classical problems of stable matching theory, many applications in the Internet, e.g., matching the Web content providers to the consumers, involve matching a two sided market with one or both sides having preferences over the other side. The ideas from the classical theory needs to be tailored to cater the requirements of the Web domain.
- **Preference aggregation**: Collective decision making in the facility installation problem or in selecting the right candidate for a job is similar to the voting procedures, which is another classical area of microeconomics. The proliferation of the computer technologies helped to aggregate preferences and the subsequent computations

efficiently. However, the computational aspects of voting is still a research area that is under-explored.

Several problems under the areas (a-c) in the crowdsourcing sub-domain (first bullet above) has already been solved in my Ph.D. thesis [2]. Even then, there are a few things that can add to the problems addressed there.

Short term goals. The immediate research plan of this proposal is to finish off the interesting research directions associated with the areas addressed in my thesis [2]. The two major problems among them are the following. First, *learning* the skill of the crowd helps design an efficient outsourcing process. In most of the real-world settings, experts may not exactly *know* their qualities rather *experience* by doing it. The mechanism design question would be to make the outsourcing truthful for the experts even when the skills are unknown to both the outsourcer and the experts. Second, in crowdsourcing, we are actually interested in scavenging the *true* information from the strategic crowd in a *time-efficient* manner. The goal is to design (exact or approximate) sybilproof mechanisms using tools like *information theory* and *prediction markets*, so that the two above mentioned objectives can be satisfied.

Long term goals. In the longer run, my plan of research is to sequentially address all other sub-domains mentioned above which serve to perform as solutions to a significant share of the economic problems on the Internet.

3 Outline of the desired results and methodology

A common thread in all the problems mentioned in this proposal is that it involves the Internet as a media of connecting individuals who are strategic. We hope to obtain certain results in this domain, as listed below.

Desirable results:

- In the task crowdsourcing domain, we want to design mechanisms or algorithms that *learn* the skills of the agents and update the list of people for assigning a task.
- In the context of time critical tasks, we want to design a mechanism such that the participants inform partial solutions as quickly as they can, and thereby inducing *synergy* among the people working towards a common goal.
- Another goal is to *design the crowdsourcing network* so that it is stable, and also meets certain performance guarantees.
- In online advertising, our goal is to design schemes so that (a) sponsored search auctions can yield provable performance guarantees for multiple slots, (b) recommendation mechanisms ensure improved revenue guarantees, and (c) incentivizing users to put their honest efforts into reviewing a content (e.g., reviewing the food of a restaurant or technical quality of a journal paper).
- Another interesting area of research is the fair division of common resources. We hope to find a mechanism design scheme that divides shared resources like bandwidth or computing power without any monetary compensation meeting some well-defined design objectives.

We would also like to contribute in the domain of voting, matching, and prediction markets. However, the exact plan of desirable results would become clear after the above mentioned areas are well understood.

Problem handling approach: Our aim is to come up with game theoretic models of these domains and provide mechanisms that satisfy the desirable properties as mentioned earlier. In the process, we might discover certain limits of achievability and design approximately optimal schemes. We would also aim to generalize the results in

order to gain a broader insight on the above mentioned problems in Internet economics. The approach to solve these problems would be mostly theoretical and the treatment would involve real analysis, logic, linear algebra, calculus, probability, and similar mathematical tools.

4 Significance in the context of the current state-of-the-art

Even though Internet was originally conceived as a medium of communication, in the current context, it has grown much beyond the scope of simple communication. Today, for almost all applications including commodity purchase, knowledge dissemination, work outsourcing, predicting events, or aggregating preferences, we use the computational and communicative power of the Internet. With the technological advances of the Internet, it is also extremely important to design the right protocols keeping in mind the human participants who are the end users of these online applications. The current literature that delves into the economic aspects of the Internet is rather thin, and therefore my research agenda is a significant value addition to understand the dynamic human behavior, the limits of achievability, and to design schemes that are strategic manipulation-proof. In particular, the plan outlined in this research proposal would:

1. yield a holistic solution to the strategic crowdsourcing problem.
2. provide a generalized solution for the sponsored search auction.
3. bring more credibility to the scientific and online review process.
4. under suitable settings, provide efficient schemes of shared resource allocation.
5. contribute to online matching, prediction markets, and voting.

The above points distinguish this research from the current state-of-the-art in the respective domains.

5 Prospect as a long-term research agenda

The Internet is getting monetized in a much rapid fashion, and as mentioned earlier, the use of the Internet as a means to generate revenue is on the rise. It is evident that the applications of the problems outlined in this proposal would only multiply in future. Therefore, the understanding of the economics of the network is indispensable to design the next generation Web applications. Soon, these applications would pervade our everyday life through the mobile phones, tablets, and all micro-communication media. Hence, Internet economics stands to perform as a lively area of research for a significant period in the future.

References

[1] Alexa-Internet. Wikipedia.org site info, May 2013. URL `http://www.alexa.com/siteinfo/wikipedia.org`.

[2] S. Nath. *Mechanism Design for Strategic Crowdsourcing.* PhD thesis, Indian Institute of Science, Bangalore, July 2013. submitted.

[3] Wikimedia.org. Wikipedia statistics, article count (official), February 2013. URL `http://stats.wikimedia.org/EN/TablesArticlesTotal.htm`.