# Research Statement: Data Driven Market Design

Michael Albert (malbert@cs.duke.edu)

Much of the modern world is systems of interacting, self-interested, economic agents, and the design of these systems face two separate, but interrelated, challenges. First, in any given system, an optimal solution, whether that be revenue maximizing, social welfare maximizing, production maximizing, or some other criteria, *must be computed*, itself a potentially difficult problem. Second, and often more difficult, the economic agents must be subject to *correctly designed incentives* to encourage interaction that achieves system optimum. Research in *market design* has already developed transformative applications, such as the modern system of online advertising, residency matching, and kidney exchanges. However, the field has identified both computationally intractable problems and problems where the requirement that incentives be correctly structured makes the problem generally impossible. Fortunately, combining the increasing wealth of available data with algorithmic approaches can make the impossible possible and the intractable tractable. **My long term research goal is to apply a data driven approach to optimize the design of markets to solve pressing, otherwise intractable, problems.**

To tackle this research agenda, a truly interdisciplinary approach is necessary. Consider the example of kidney exchanges. Kidney exchanges provide a mechanism to allow a pair of people, one in need of a kidney and one willing to donate a kidney, but unable due to medical constraints, to "swap" kidneys with another pair. However, there may not be a pair that matches with each other, but there might be a triplet (or an even longer *cycle*). This is a *formally*, or NP-hard, problem even without considering that pairs may behave strategically. Moreover, we cannot fix incentives by paying donors, due to legal restrictions. Therefore, a computational, optimization focused approach must be combined with a strong understanding of both economic incentives and mechanisms for aligning those incentives. This line of research has led to practical success; UNOS currently uses algorithms developed by this research community to allocate kidneys nationwide.

Increasingly, the agents of interest are not even human but are instead algorithms that can interact at the frequency of milliseconds. These algorithms possess their own objectives and limitations that must be incorporated into the design of markets. Most of the modern internet is funded through display advertising, and that advertising is targeted. In order to optimally match advertisers and users, online advertising platforms (e.g. Google, Facebook, and Microsoft) conduct an auction, including soliciting bids, for every click of every individual on every ad supported websites, many of which have multiple ad slots. This must all be done between the time when a user clicks a link and the webpage loads. Therefore, the advertising platform must a) be able to very quickly compute a winner for the auction and b) be such that the bidder can very quickly compute their bid. This problem has generated novel solutions that have been widely implemented as a key component of the information economy.

As the world moves towards increasing automation, there will be an opportunity to develop and guide systems that influence purchasing decisions, resource allocation, traffic flows, energy usage, health care outcomes, and supply chains. Moreover, since many systems operate electronically, every interaction can be recorded, analyzed, and incorporated into future iterations of the market. In future work, I intend to focus on three potential markets that are poised to benefit from fundamental shifts in technology in the near term. First, resource allocation systems, such as federated server farms, may be fundamentally transformed due to block-chain based distributed computation. Second, traffic systems and road infrastructure utilization will see the widespread adoption of self-driving vehicles and, with them, a once in a generation opportunity to rethink road systems. Third, the move to a "smart" grid, for the distribution of electricity, and intelligent appliances, that can respond to incentives, will allow for opportunities to significantly affect energy usage through pricing and scheduling.

## Mechanism Design with Unknown Correlated Distributions

The typical challenge in market design (or *mechanism design*) is that the participants in a market (or *mechanism*) have some *private valuation* over the possible outcomes of the mechanism, e.g. participants in an auction know their valuation for the object but the auctioneer does not. It has long been known that optimal mechanism design, particularly when the goal is to either maximize revenue or implement budget balanced, socially efficient outcomes, requires that the mechanism be designed with some understanding of the distribution of agents that will participate. Within this class, the mechanisms with the strongest guarantees are only possible when agents valuations are *correlated*.

However, most of the literature assumes that valuations are independent, even though only minuscule amounts of correlation are necessary for the strongest results, i.e. the correlation condition is *generic*. Moreover, in any setting where the value for an item has a common value component (e.g. future resale) valuations will be mechanically correlated. Unfortunately, in Albert, Conitzer, and Lopomo (2015), we demonstrate that the literature makes an independence assumption, at least partially, because the strongest mechanisms are very sensitive to the estimated distribution. However, many settings of interest to mechanism design are rapidly repeated, online advertising auctions being a prime example, and as more settings are automated this is becoming more common. A main insight of this line of work is that these repeated settings provide an opportunity to *learn* the information necessary to construct the optimal mechanism. A crucial component is that this learning process must and can be done *robustly* by algorithmically constructing mechanisms tailored to the specific setting.

## Robust Revenue Maximizing Mechanisms

The difficulty with designing the optimal mechanism in a correlated valuation setting is the extreme sensitivity to the distribution. However, if the correlation is not taken into account, the performance of the mechanism remains bounded arbitrarily far away from optimal. Therefore, traditional mechanism design procedures either *overfit* the estimated distribution or *effectively disregard* the estimate altogether (see the Bayesian and Ex-Post performance in Figure 1, respectively).

This research agenda combines techniques from robust optimization with techniques from automated mechanism design to incorporate estimates of the true distribution into the mechanism design process, while avoiding overfitting. To achieve this, we fully characterized, for the first time, the optimal mechanisms under a generic correlation condition (Albert, Conitzer, and Lopomo, 2016). Through exploring the space of mechanisms that could feasibly be learned, in Albert, Conitzer, and Stone (2017b), we prove that the ability to learn the optimal mechanism is highly dependent on the amount of correlation present in the original distribution, in contrast to previous work on learning in correlated mechanism design.

This line of research has culminated in the development of $\epsilon$-robust mechanism design, a computationally efficient and sample efficient methodology to incorporate estimates of the true distribution. We demonstrate in Albert, Conitzer, and Stone (2017a) that the revenue optimal mechanism in this class could be computed in polynomial time, and in an article currently under submission at Operations Research (Albert et al., 2017), we demonstrate that an arbitrarily close approximation to the revenue optimal mechanism, under complete certainty, can be achieved with a polynomial number of samples from the true distribution. Figure 1 demonstrates the performance of our $\epsilon$-robust mechanism design procedure.

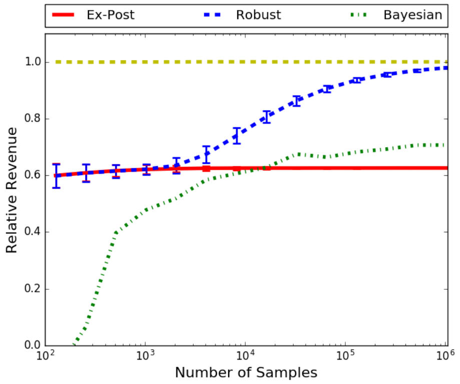

Figure 1: The performance of $\epsilon$-robust mechanism design relative to traditional approaches when the distribution is estimated (full revenue is 1).

## Budget Balanced and Socially Efficient Resource Sharing

While our early research has focused on the revenue maximizing setting, another setting of great interest to me and an area of ongoing work is that of budget balanced, socially efficient mechanisms. Due to the well known Myerson-Satterthwaite Impossibility Theorem, it is generally *impossible* to achieve budget balance, i.e. there is no net monetary loss or gain, and socially efficiency, i.e. the socially optimal outcome is chosen.

This is unfortunate because it seems to rule out many important applications. Specifically, any setting where a group would like to pool resources, such as computing resources within or across firms, must either a) have some outside party that puts money into or takes money out of the system or b) the group must settle for subpar utilization of resources. However, under correlated distributions, this result does not hold. Moreover, in the particular example of shared computing resources, using data from a Google cluster, my colleagues, B. Lee, V. Conitzer, S. Zahedi, P. Zheng, and I show that correlation (between user demand and total demand excluding the user) is very high, exceeding 40% for the majority of users. In on-going work, we are extending the $\epsilon$-robust mechanism design framework in order to achieve budget-balance and social efficiency in a federated server farm setting, i.e. all participants share computing power.

Moreover, this approach provides even stronger guarantees than traditional budget balance. Specifically, in most budget balanced mechanisms, there is a problem where many "low value" agents participate in order to get payments from "high value" bidders. With correlation, we can ensure that not only are payments budget balanced, but they are also zero in expectation *for all participants*, low or high. Additionally, the $\epsilon$-robust mechanism design approach generally scales polynomially with the support of the distribution of bidders, implying an exponential scaling in the number of bidders. However, for the specific case of socially efficient mechanisms, the scaling can be done linear in the number of bidders by decoupling the optimization problem.

A particularly interesting interface of this work with recent technological developments come from block chain based distributed computation, such as Ethereum. By combining this $\epsilon$-robust mechanism design approach for resource allocation with a system for distributed computation, fully decentralized, optimal resource allocation should be possible. However, many open questions remain: In Ethereum, computation is costly, so how do we balance the optimality of the mechanism with the cost of computation? How do we structure incentives in early rounds of the mechanism in order to avoid agents strategically mis-reporting in order to influence later rounds?

## Data Driven Optimization of Traffic Systems

I am also very interested in market based mechanisms to optimize the use of existing road infrastructure by influencing traffic. It has long been known that optimal tolling can lead to both better travel times and increased social welfare on existing infrastructure. However, traditional tolling systems have been restricted to a single major roadway due to the cost of implementation and inconvenience for the driver on collecting the toll. With the approaching adoption of autonomous, connected vehicles, there is the opportunity to explore entirely novel tolling systems. However, privacy concerns are likely to limit the amount of information that can be collected in these tolling systems. As part of a mega-project funded by the Texas Department of Transportation, my colleagues and I introduced a novel tolling system, $\Delta$-Tolling (Sharon et al., 2017b). This tolling scheme uses purely local, link based information, specifically the difference between free flow travel time and congested travel time, to efficiently compute tolls for an *entire city in real time.* This approach is novel for both being model free and having extremely low data requirements. In fact, the system could be implemented with a smart phone app. We demonstrated that this tolling scheme is both theoretically optimal, under certain assumptions, and in micro, meso, and macro simulations, using empirically collected travel demand data, the improvements in travel time was 33% for the road network of Austin, TX.

However, even with dramatic improvements in travel times and social welfare, tolling systems are politically difficult to implement. Therefore, a more practical approach may be an opt-in tolling schemes, where a participant receives a positive incentive, initially, to enroll and then faces tolls that affect their driving behavior. The question is, how many participants are necessary? We developed a novel approach to computing, in polynomial time in the size of the network, the necessary amount of traffic that must opt-in in order to achieve the system optimal result. We demonstrate (Sharon et al., 2017a), for six networks, that the average percentage of compliant agents necessary over our test networks is 25%. This line of work can influence the design and targeting of any such opt-in system.

## Scheduling Charging in a Smart Grid

Given the move to renewable energy and the adoption of electric vehicles, there is both a need and an ability to time-shift loads on the electric grid. Specifically, the grid can delay charging of an autonomous vehicle, or even use the vehicle as a temporary battery, in order to smooth demand. Therefore, my colleagues and I (de Weerdt, Albert, and Conitzer, 2017) exam the problem of charge scheduling in the smart grid. We investigate more than 20 variants of the charge scheduling problem, and we prove the computational complexity of each variant.

Unfortunately, many of the interesting variants are computationally intractable. Moreover, there is an incentive issue in charge scheduling, in that the users of the grid need to report their time constraints and desired quantity of charge. While approximate solutions to the scheduling problem would seem to be the answer, given the computational intractability, traditional mechanism design fails if the outcome that is chosen is approximately optimal. Therefore, we need solutions that are *maximal in range.* However, determining the correct range to maximize over is a problem that can only be answered by examining realistic distributions of charging habits and supply fluctuations. I intend to combine sophisticated analysis of energy usage data with targeted mechanism design to incentivize accurate reporting of preferences over charging.

## Financial Economics Research

Though my recent work has been data intensive and algorithmic, my early research used more standard techniques from the financial economics literature. For completeness, I will briefly discuss this older work, and explain how the tools and methodology continue to inform my current work. As an undergraduate and early PhD student, I, along with my co-authors, developed and published a computationally efficient approach to pricing a specific kind of financial option, a barrier option (Albert, Fink, and Fink, 2008). Our approach used increasing refinements of our model around a strong non-linearity that affected the pricing. This insight allowed us to efficiently and accurately take into account the non-linearity without paying a global computational cost.

During my PhD, I focused on two areas: the interaction between intellectual property rights and economic growth (Albert, 2011) and the optimal CEO compensation contract in order to provide incentives for optimal firm debt choice (Albert, 2014). While these two areas may seem disparate, they are both, at the core, optimal contracting problems, i.e. how do we set up the system of incentives for economic agents to behave optimally. I argue in Albert (2011) that a key component of the wedge between developed and developing economies is intellectual property rights. Specifically, developing economies generally have very weak intellectual property rights. Weak intellectual property rights allow firms to acquire, extra-legally, intellectual property from other nations. The farther behind the *intellectual frontier* the economy is the more valuable the ability to steal. However, as an economy develops, eventually the inability to use intellectual property to secure capital, due to the inability to legally protect intellectual property, outweighs the positive benefits of being able to ride off the coat tails of more advanced economies. Therefore, the typical arc of intellectual property rights, starting weak and growing stronger as an economy develops, can be viewed as an optimal economic response.

For my dissertation, I looked at constructing incentives for CEOs in order to encourage optimal debt decisions. It is a widely recognized puzzle in the literature that firms seem to use too little debt relative to the tax benefits that debt provides. However, CEO compensation contracts are structured in a way that interacts strongly with firm debt decisions. Specifically, a CEO is often given large equity stakes, stocks and options, which become effectively worthless in a bankruptcy event. Moreover, the CEO is nearly always fired when a firm declares bankruptcy, so a CEO is hit twice in bankruptcy, making a *risk-averse* CEO more likely to overly weight bankruptcy in her decision making process. I detailed this effect and demonstrated that it is present in the actions of CEOs using firm level data. Therefore, the optimal contracting response is to provide positive incentives during bankruptcy, often referred to as "golden parachutes." Though highly derided in the media, I demonstrate that they are an optimal response to an incentive misalignment between a firm and a CEO.

# References

Albert, M.; Conitzer, V.; Lopomo, G.; and Stone, P. 2017. Mechanism design with correlated valuations: Efficient methods for revenue maximization. Under Submission at Operations Research.

Albert, M.; Conitzer, V.; and Lopomo, G. 2015. Assessing the robustness of Cremer-McLean with automated mechanism design. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI '15, 763–769. Austin, TX, USA: AAAI Press.

Albert, M.; Conitzer, V.; and Lopomo, G. 2016. Maximizing revenue with limited correlation: The cost of ex-post incentive compatibility. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*.

Albert, M.; Conitzer, V.; and Stone, P. 2017a. Automated design of robust mechanisms. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*.

Albert, M.; Conitzer, V.; and Stone, P. 2017b. Mechanism design with unknown correlated distributions: Can we learn optimal mechanisms? In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '17, 69–77. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Albert, M.; Fink, J.; and Fink, K. E. 2008. Adaptive mesh modeling and barrier option pricing under a jump-diffusion process. *Journal of Financial Research* 31(4):381–408.

Albert, M. 2011. Financial development, intellectual property rights, and growth. Working Paper.

Albert, M. 2014. Executive compensation and firm leverage. Working Paper.

de Weerdt, M.; Albert, M.; and Conitzer, V. 2017. Complexity of Scheduling Charging in the Smart Grid. *ArXiv e-prints*.

Sharon, G.; Albert, M.; Rambha, T.; Boyles, S.; and Stone, P. 2017a. Traffic Optimization For a Mixture of Self-interested and Compliant Agents. *ArXiv e-prints*.

Sharon, G.; Hanna, J. P.; Rambha, T.; Levin, M. W.; Albert, M.; Boyles, S. D.; and Stone, P. 2017b. Real-time adaptive tolling scheme for optimized social welfare in traffic networks. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2017)*.