Trends in Computational Social Choice

15

Cite as: Nicholas Mattei and Toby Walsh. A PrefLib.org Retrospective: Lessons Learned and New Directions. In Ulle Endriss (editor), *Trends in Computational Social Choice*, chapter 15, pages 289–305. AI Access, 2017.

http://www.illc.uva.nl/COST-IC1205/Book/

CHAPTER 15

A PrefLib.org Retrospective: Lessons Learned and New Directions

Nicholas Mattei and Toby Walsh

15.1 Introduction

The Internet enables computers and (by proxy) humans to communicate at distances and speeds previously unimaginable. Many of the benefits from this technology are derived from the ability to connect more decision makers (in computer science (CS) we call these *agents*) into groups, composed of human agents, computer agents, or a mix of the two. These groups of agents must make collective decisions subject to external and internal constraints and preferences in many important real-world settings including: selecting leaders by voting (Faliszewski and Procaccia, 2010), kidney exchanges (Dickerson et al., 2012), matching students to seats in schools (Abdulkadiroğlu et al., 2005), allocating work or resources (Budish and Cantillon, 2012; Aziz et al., 2016), and distributing food to charities (Aleksandrov et al., 2015). In all of these settings, self-interested agents formalize and submit their preferences to a centralized or de-centralized authority and outcomes (kidney matchings, leaders, etc.) are decided by a mechanism. Each mechanism for group decision making may (or may not) satisfy various criteria, e.g., fairness and/or efficiency, that a system designer deems important. Within CS, the study of mechanisms including algorithmic, axiomatic, and practical issues, broadly fall in into the artificial intelligence (AI) related subfields of algorithmic game theory (Nisan et al., 2007), preference reasoning (Domshlak et al., 2011), and computational social choice (ComSoc) (Brandt et al., 2016; Rothe, 2015). Results from these research areas have impact within CS as well as across the sciences and daily life, with applications in recommender systems, data mining, and machine learning (Chevaleyre et al., 2008; Domshlak et al., 2011).

Game theory is an important mathematical framework used to analyze strategic behavior of self-interested agents with applications across a number of domains including economics, biology, and computer science (Maschler et al., 2013). A game-theoretic analysis typically provides an idea of how agents *may* act, within a given context, under assumptions about the rationality of and information available to them. However, as researchers have found, there are many instances in economics and biology (Goeree and Holt, 2001) where the predictions of game theory are contradicted by data or experiment, giving rise to the school of behavioral and experimental economics (Kagel and Roth, 1995; Camerer, 2011). Indeed, many important research results for mechanisms and social choice in economics has come from the development of theory that is specifically informed by real-world data and/or practical application that is then rigorously tested (e.g., Budish and Cantillon, 2012; Dickerson et al., 2012).

Much of the work in ComSoc centers on collective decision making; with a special emphasis on understanding manipulative or strategic behavior by the participating agents. This line of inquiry answers questions about incentives and security: participants in an mechanism should be incentivized to report the truth and/or be unwilling (computationally) or unable (axiomatically) to find a misreporting of their information that is beneficial. For voting and aggregation schemes this most often means studying how agents can strategically misreport their preferences given worst-case assumptions about the knowledge of the manipulators, e.g., complete information, nicely structured preferences of the other agents, or limiting assumptions on the responses of other agents (Brandt et al., 2016). Consequently, these studies provide only limited information about the reasoning complexity in many real-world settings; manipulation is often trivially easy given complete information and strict preferences; or NP-hardness proofs may rely on huge instances or vary particular structures which do not frequently occur in real life (Davies et al., 2011; Mattei and Walsh, 2016).

Indeed, in the paper that laid the intellectual foundation for complexity theoretic analysis of voting and aggregation procedures, Bartholdi, III et al. (1989) warned against this direction: *"The existence of effective heuristics would weaken any practical import of our idea. It would be very interesting to find such heuristics."* While there has been robust work on moving beyond the worst-case in theory, leveraging tools such as fixed parameter tractability (Conitzer, 2010; Faliszewski and Procaccia, 2010; Betzler et al., 2012); average case analysis (Erdélyi et al., 2007; Rothe, 2015; Xia and Conitzer, 2008); and other approximation and heuristic techniques (Skowron et al., 2013); until recently there has not been a similar emergence of *data-driven* research programs that directly questions these worst-case assumptions.

The first goal we had in mind when founding PREFLIB was to address what we see as two fundamental questions in ComSoc that can be addressed with data:

1. How wide is the gap between theoretical intractability results and practical, real-world instances? If the constructions required to prove theoretical intractability are rare, what does this tell us about the practical applicability of these results?

2. Models of agent behavior and rationality seem to be largely driven by intuitive feeling (e.g., a left to right political spectrum) or mathematical expediency (preferences are complete, strict linear orders). How realistic are these assumptions? Do we ever see them in real-world data? Can we derive or learn the assumptions we should use from data?

The first push of data to PREFLIB: A Library for Preferences (Mattei and Walsh, 2013) was completed on March 15th, 2013. It contained 40 data files from 5 different sources totaling about 10MB, all from our prior publications. Since that time PREFLIB has grown to encompass three distinct types of data and includes over 100,000 data files from 40+ sources totaling more than 10GB. We have organized four instances of the EXPLORE Workshop, held at the International Conference on Autonomous Agents and Multiagent Systems (AAMAS), which focuses on the use of data in ComSoc. Through this effort we have seen a sharp rise in the number of papers using experiments to validate or inform worst-case assumptions. There has also been an increase in the number of tools being built and deployed within the ComSoc community — a sure sign that the community is looking to translate research into impact.

The second goal behind creating PREFLIB was to help social choice and preference research in computer science walk the same road that Kagel and Roth (1995) describe for experimental economics: evolving from theory, to simulated or re-purposed data, to full fledged laboratory and field experiments. This progression enables a "conversation" between the experimentalists and the theoreticians which allowed the field to expand, evolve, and have the impact that it does today (Camerer, 2011). We want data work to feedback into basic theoretical research in CS creating a virtuous circle: if we can verify preference models and input languages, we can build more computational tools, if we can rule out certain behaviors in practice then we can be more confident when deploying tools and mechanisms. Closing this feedback loop will enable practitioners to rigorously test their theoretical assumptions before deployment, providing concrete guidance and adding methods to the theoretical analysis toolbox that are built on well studied, practical foundations. We have seen this progression within other fields in computer science, including machine learning fueled by the UCI Machine Learning Repository (Bache and Lichman, 2013); constraint programming fueled by CSPLIB (Gent and Walsh, 1999), and most recently the explosion of deep learning fueled by resources like ImageNet (Deng et al., 2009).

In this chapter we look back at the process of designing, building, supporting, and promoting PREFLIB. We discuss the basic ideas used and challenges overcome in creating the website and dataset itself including some (hard) lessons learned for others who wish to create and maintain community resources. We then look at some of the publications that have leveraged PREFLIB as well as new tools and services related to or using PREFLIB, surveying the new impact and research directions. We finally look ahead to the next few years of PREFLIB and detail our (biased) view of important research challenges we see on the horizon including expanding the coverage of library and tool chain; using the library to learn well-founded domain restrictions or trends in preferences; expanding the scope of empirical testing and evaluation in social choice; and encouraging stronger links with other aspects of computer science.

15.2 Looking Back: Motivations and Challenges

In our paper that introduced PREFLIB we outlined a set of motivations for building PREFLIB and a set of challenges that we saw on the horizon. Taking each of these sets in turn we discuss the current priorities of PREFLIB and how we think we did agains the challenges. As discussed in the introduction of this chapter there were a number of motivations behind building PrefLib. While our thinking has changed over the years we remain true to a number of our original motivations.

Challenges and Competitions. When we started out we had intentions of establishing the library itself as a set of data on which to run competitions and challenges, much like the MAX-SAT Competition hosted at `http://www.maxsat.udl.cat/` or the Netflix Prize Challenge (Bennett and Lanning, 2007). This explicit motivation has fallen away given privacy concerns around releasing data, e.g., the lawsuits surrounding the sequel to the Netflix Prize Challenge, and the fact that the research priorities of ComSoc are not explicitly amenable to competitions. While we could imagine competitions around various preference reasoning algorithms, given that the majority of PrefLib contains voting data and multi-attribute preference data, it is not clear what kinds of goals this competition would have.

Benchmarking. We feel that PrefLib has just recently crossed the threshold where we can begin using the library as a benchmark for various algorithms in the ComSoc community. We have started to see some of this work, for instance the work of Skowron et al. (2015) on approximating hard to compute proportional representations. We see a this type of research expanding as benchmarking could be very interesting for looking at average case or approximation ratios for various voting and assignment objectives that are computationally hard to compute, see, e.g., (Aziz et al., 2017) and (Bouveret et al., 2016).

Realism. Perhaps the key motivating factor behind assembling PrefLib was a desire to have realistic data. Many of the models studied in classical social choice seem to be chosen because they *seem* reasonable or were explicitly chosen for mathematical expediency. Perhaps nothing is more of an exemplar here than the fact that out of over 300 profiles containing strict, complete preference relations, absolutely none are single-peaked, a common profile restriction that has been called “natural” or “well-motivated” numerous times since its introduction by Black (1948). Collecting data has helped us to quantify what is reasonable. Now we have to start using the data.

Insularity. The final motivation was that many groups within ComSoc were rather insular: most groups worked on their own problems and their own datasets. Additionally these resources were dispersed and not well interconnected through common portals. An additional concern was that we were not collecting data and interacting with more data-centric communities such as the Data Mining and Machine Learning communities, where we think much of the work in ComSoc has applications. We have started to bridge these gaps in big ways: we survey the large selection of tools now available in ComSoc which are mostly interlinked on the web. While PrefLib was not the impetus for all of this, we like to think we helped.

In addition to the motivations behind building PrefLib we also foresaw a number of hurdles and challenges that we would face on establishing the library.

Variety and Over-fitting. We painted these two challenges as two sides of a coin. Variety meant there were too many shapes and forms that preferences came in, while over-fitting is a challenge if PrefLib was too small. Rather than try to cover the entire gambit of preference formalisms we focused more tightly on some of the more common formalisms: preference orders and ratings on combinatorial domains. This allowed us to gather a large amount of data across voting, allocation, and matching domains where many groups are doing research. In this way we have (hopefully) addressed both of these concerns.

Elicitation and Modeling. Eliciting and modeling user preferences are both hard problems. Finding the proper formalism and then devising a structure to encode that formalism are both necessary and difficult research problems. We wanted to ensure that while collecting available data into a large database we did not take focus away from these other problems. We may have been overthinking our ambitions at the beginning. There are still rich and ongoing research programs on both of these topics. However, like before, most of this research takes place in other fields like psychology and machine learning (Allen et al., 2015); and we must admit that perhaps some of the formal preference reasoning research in ComSoc has fallen away, evidenced by the lack of such a chapter in the Handbook (Brandt et al., 2016).

Privacy and Data-Silos. We wrote that others may be reluctant to share data for a number of reasons or it may require serious effort to put data in common formats. On the latter point, we even underestimated the challenge; ball-parking the man hours required to put everything in sane and common formats is beyond us at this point. However, we have been encouraged by all the groups, both within ComSoc and beyond, that have approached us to donate their data (even more when they convert it before sending to us). However, we will never overcome the challenge of releasing data and the inherent tension it brings between privacy, exclusivity, and the advancement of science. We have been happy so many have been willing partners.

15.3 Building PrefLib

PrefLib is technically three different systems corresponding to two different GitHub repositories and several thousand individual text files. The first GitHub repository is code and templates for generating the website itself, including the scripts to build the indexing and cross-linking. The second GitHub repository is the tools, useful not only for conducing experiments but also for reading and writing the text files in the various PrefLib formats. The final and largest piece is the several thousand text files which make up the "database" of preferences. We will discuss each of these three core components in turn and discuss the design decisions, technologies, and lessons learned from creating them.

The heart and soul of PrefLib is the data itself. We started off with data files from various projects that we had done in the past, devised a common file format, converted our existing files into those formats, uploaded them to the web, and boom, PrefLib was born... almost.

From the beginning we wanted to design for both extensibility and ease of use across not only researchers in the ComSoc community but also researchers from psychology, sociology, and political science. Most projects that upload data to the web and walk away are doomed to fail; it requires sustained effort and intentional maintenance to translate a pile of data on the web into something that can be used. The UCI Machine Learning Repository (Bache and Lichman, 2013) and high impact toolkits like scikit-learn (Buitinck et al., 2013) have required full-time developers and committed support; we had two people.

15.3.1 The Data and Website

From the beginning we wanted to integrate files donated by a wide variety of researchers in social choice and beyond. This was the driving force behind using simple, comma separated value based file formats. We hoped that this would mean that others could easily translate their files and send them to us when they heard about the project. Hence, the construction of the database is about as old school as it gets. When we see cool experiments or datasets, we ask to host them. We think of data we would like to have and either go out and collect it or we look for partners (like IFAAMAS) who are willing to help us collect and then publish the data online. We organize the EXPLORE series of workshops as a way to get the word out and hopefully attract even more submissions.

The website and the data are inextricably linked and we cannot explain one without explaining the other. PrefLib is a series of static pages that are uploaded onto a private server that we maintain along with a large directory structure containing the data. We chose this approach over something more complex, e.g., keeping the text files in a large database and dynamically generating the pages when people loaded the site, because (1) we honestly do not update the text files that often and (2) we are not web designers.

Typically our data is collected when either we get in touch with someone, or they contact us about hosting data on PrefLib. We collect the requisite information about our ability to publish the data, the required citations from the collector/author, and any special notes they would like distributed. We[1] then convert the data into one or more of the various PrefLib formats and add it to our index.

In the first iteration of the site, which was online between 2013 and 2015, a Python script read one giant .csv file which contained meta-data and the path for every data file within PrefLib. In retrospect this was not the best design choice as the file quickly became large and unmaintainable. The only practical upside was that the entire database index was in a single file that we could put under version control easily.

We moved to the current design in 2015, motivated by a number of factors, but mostly due to the size and time it took to maintain the index file. We essentially redesigned the entire process with the design goal that researchers could download a single archive file which would be entirely self-indexing, with no special soft-

[1] Not always and thanks to everyone who sends in correctly formatted files!!

ware required. To this end we decided to use the directory structure itself as the index method, with each folder representing a complete dataset and meta-data. The new indexing script simply walks the directory structure of the /data/ folder, builds index.html pages within each folder given the info.txt file in that folder, and builds a top level index page to interlink with the main static portion of the site. This entire structure is then rsynced to the web server. The entire set of scripts and static webpage files is available at `https://github.com/nmattei/PrefLib-www`.

A current major design challenge we are facing is how to revision the data files themselves. There over 5000 uncompressed data files in the index that range from a few KB to several GB in size. We currently have some "manual" versioning that happens in the form of pushing a dated archive of the entire /data/ directory to the /archive page of PREFLIB. However, this solution is not optimal and we hope to move to something more inline with modern development practice in the near term like git-lfs or some another system for versioning both extremely large and extremely numerous files.

15.3.2 The Tools

The PREFLIB Tools project, available at `https://github.com/nmattei/PrefLib-Tools`, was not originally planned as part of PREFLIB. However, after looking at the framework that we needed to build just to merge our two datasets, not to mention the amount of code that we needed to write in order to translate the various formats that we had coming in as donations, we decided that maybe a Python module that could read and write the file formats was in order.

The initial launch of the code was just functions in Python that could read and write the file formats listed on PREFLIB. We also included the functions necessary to convert between some of the different formats on the site, e.g., turn a strict order into a set of pairwise comparisons. We packaged this up as a single file and posted it on the site.

Over the first years we kept getting requests for more data that was generated according to a particular distributions (also known as cultures in the wider social choice arena, see e.g. Mattei (2011) for more discussion) or had different numbers of candidates. Adding to the pressure to publish more code was reading about experiments which claimed (incorrectly) to generate profiles or structures at random (for a longer example see Allen et al. (2016)). Finally, since our goal was to expand and facilitate a culture of empirical experimentation (Cohen, 1995) in ComSoc we felt the community needed at least some tools to support those just starting out. So we published generators and a command line script to generate unlimited data according to many of the statistical distributions that have been used in social choice research in the past.

After finishing most of the generators we moved the code to GitHub in 2015 in order to make it more accessible and allow others to contribute to the code base. Along with this move we added functions to check for domain restrictions such as single-peakedness (Black, 1948), functions to compute various randomized allocations (Aziz et al., 2015), and have uploaded examples and tutorials that we have given at various conferences over the past several years.

After five years we still have a long way to go to make the tools more generally useable. While they are reasonably well documented, they were never "designed." We have begun a process of refactoring the code to bring some consistency to the objects and call structures we use. We hope that this process will make the code more useable and more extensible for others to use in the future.

15.4 More Tools in ComSoc

While we see PrefLib as a library and platform to enable research there have been a number of tools developed and deployed online by members of the ComSoc community. We see the broader movement towards implementation and providing useful apps as a sign of budding maturity within ComSoc. We highlight some of the most interesting and useful tools in this section. For a more comprehensive list of other tools in ComSoc as well as other public datasets please visit `http://www.preflib.org` where we maintain a comprehensive list.

- Whale3, which stands for WHich ALternative is Elected, is an open source web application created by Sylvain Bouveret and is available at `http://strokes.imag.fr/whale3/`. Whale3 is one of the first online polling systems developed by members of the ComSoc community and put online. The app allows for a number of input preference types including approval voting and rank order ballots, and a number of voting rules including Plurality, Borda, and STV. There are also a number of visualizations to analyze the output of a particular poll. You can read more about Whale3 in Chapter 20 of this book.

- The Spliddit project run by Goldman and Procaccia (2014), available at `http://www.spliddit.org/` is a web-based tool to facilitate the splitting of a variety of divisible and indivisible goods from rent to cab fares. It is a front end to a variety of game theoretic and social choice algorithms developed over the years including the Shapley Value (Shapley, 1953), for splitting cabs, and the Dollar Share (de Clippel et al., 2008), for dividing credit.

- The Pynx project run by Brandt et al. (2015), available at `https://pnyx.dss.in.tum.de` is an easy to use web based tool for preference aggregation. It is designed to run decentralized surveys or polls and automatically selects from a variety of rules including Kemeny's Rule (Kemeny, 1959) and Fishburn's Rule (Fishburn, 1984), also known as maximal lotteries. The inclusion of Fishburn's rule makes Pynx the only online tool to offer randomized rules (Brandl et al., 2016).

- The UNOS Kidney Paired Donation Pilot Program creates a matching market where a donor/receiver pair that are incompatible are matched with different donor/receiver pair that are incompatible such that the cycle (or longer chain) is compatible. Hence, if a husband cannot donate to his wife due to incompatibly, he may be able to donate to another woman whose husband can donate their kidney back. Finding cycles of these possible donations in large groups of people is a computationally difficult problem.

The Kidney Exchange research program run by Dickerson et al. (2012) provide deep technical expertise and custom tools to support the UNOS in this effort. This research program has led to a number of fundamental advances in matching theory (Dickerson et al., 2014) and the group has released a number of tools (and provided datasets to PREFLIB); an overview of these tools is available on John P. Dickerson's GitHub page at `https://github.com/JohnDickerson/KidneyExchange`.

- The Votelib project run by Tal et al. (2015), available at `http://votelib-hdm.ise.bgu.ac.il/` is a collection of data about strategic voting behavior. The group conducted a number of studies with properly incentivized participants in their lab. These participants attempted to vote strategically on a number of tasks. The group then attempted to evaluate the types of strategies used by these voters to solve manipulative voting problems.

- The CRISNER project, which stands for Conditional & Relative Importance Statement Network PrEference Reasoner, was developed by Santhanam et al. (2010) and available at `http://www.ece.iastate.edu/~gsanthan/crisner.html`. The goal of the project was to provide fast software to solve dominance queries for CP-nets using advances from the model checking community. Since then CRISNER has expanded to other preference formalisms and provides fast solutions to many problems proven to be NP-hard or harder in the preference reasoning literature (Domshlak et al., 2011).

- The Democratix project is run by Charwat and Pfandler (2015) and is available at `http://democratix.dbai.tuwien.ac.at/`. This project consists of ASP implementations of many voting rules, including some that are computationally hard such as Kemeny's and Dodgson's voting rules. The ASP implementations are very fast and capable of computing solutions for fairly large instance sizes. The website itself is a nice interface to the system and even takes PREFLIB formats as input! The code is open source and allows others to create new voting rules using ASP statements.

- The RoboVote project is run by Ariel Procaccia and his team at Carnegie Mellon University and is available at `http://robovote.org/`. The site is an elegant and easy to use interface for a number of voting and selection rules divided into the two traditional views of voting: aggregating subjective preferences or aggregating objective preferences subject to noise. To this end the site implements a number of new voting rules that are optimal for these two views of voting given certain noise functions and/or assumptions about the views of the voters (Caragiannis et al., 2017; Boutilier et al., 2015; Procaccia et al., 2016).

15.5 Leveraging PREFLIB

In this section we survey some of the papers that have used data to explore topics in ComSoc. In each of these papers, empirical experiments were run that compliment comprehensive theoretical results. We feel that each of these papers is made stronger, and the results more impactful, by the inclusion of experiments run on real-world data.

- In "Achieving Fully Proportional Representation: Approximability Results", by Skowron et al. (2015), the authors study the complexity of approximate winner determination under the Monroe and Chamberlin-Courant multi-winner voting rules. Though the outcomes of these rules are hard to compute in theory, the approximation algorithms presented in the paper are often tractable and give good results in theory. The empirical experiments use data collected by the authors, and donated to PrefLib, to show that in practice, the approximation ratios are often significantly better than those guaranteed by the theoretical results. This should give implementers confidence in using approximation algorithms to achieve good results in practice.
- In "Voting with Rank Dependent Scoring Rules", by Goldsmith et al. (2014), the authors detail a new class of voting rules which combine Order Weighted Averages with traditional scoring rules. The main thrust of the theoretical work in the paper is the axiomatic characterization of these rules, which show that they have a mix of properties, some better, some worse, than existing rules. To compliment these axiomatic results, there are also empirical experiments on data from PrefLib, showing that in practice, the rules perform better than traditional scoring rules at being robust to noise, a stated design goal of rank-dependent scoring rules.
- In "Optimal Aggregation of Uncertain Preferences", by Procaccia and Shah (2016), the authors provide polynomial time algorithms to aggregate complete rankings of agents when their preferences are expressed as distributions over rankings. This is an important step in relaxing the common strict assumptions over the preference orders of agents. The algorithms presented are complex but yield polynomial time results for minimizing the expected sum of Kendall tau distances between the set of input rankings and the final output ranking. The experiments in this paper are designed to show that *ignoring* this uncertainty can lead to very sub-optimal results. Here we see experiment bolstering the impact of theoretical work by showing how bad things can get when one ignores uncertainty.
- In "Elections with Few Candidates: Prices, Weights, and Covering Problems", by Bredereck et al. (2015), the authors detail algorithms and empirical experiments for problems that occur when voters have prices associated with changing their votes, known as the bribery problem in ComSoc. The authors close a number of open problems in the literature and provide a high level algorithm that encompasses many of the known results. The algorithmic results are mostly in FPT, which provides one measure of computational hardness. Nicely complementing these results is a set of empirical experiments using custom algorithms and MILP formulations, on data from PrefLib, that shows tractability on real-world instances.
- In "Empirical Analysis of Plurality Election Equilibria," by Thompson et al. (2013), the authors design and run a series of comprehensive experiments to

investigate the equilibrium states that occur under a variety of information assumptions on the parts of the voters. This work nicely encapsulates the idea that though voters may be strategic, they may not be able to correctly guess at what equilibria other voters are playing. A variety of test settings are considered and they show that, despite the worst-case assumptions, plurality often still leads to reasonable equilibria. The comprehensive set of tools developed for this paper are available under the PrefLib site in the /tools/ section.

15.6 Looking Ahead: The Next Five Years

We are entering a period of research in ComSoc where one can grab data from a variety of sites and analyze it using a number of online and offline systems. This new ecosystem of data and tools is opening up new avenues of research and exciting new questions. We broadly consider this ecosystem and suggest new and exciting research directions that can be tackled.

Learning and Using Domain Restrictions: As we have seen, some of the assumptions in ComSoc are made more for mathematical expediency rather than motivated by data or experiment. While traditional game theory tells us what may happen if agents are perfectly rational, lessons from behavioral game theory (Camerer, 2011) into how *humans* typically act has not been leveraged in ComSoc. In mechanism design we are starting to see work along this line (Wright and Leyton-Brown, 2012) and it is helping to deliver better impact in areas including auctions (Hartford et al., 2016). We should work with researchers in preference learning, deep learning, and other fields to mine our available preference data, including VoteLib (Tal et al., 2015) and PrefLib, for models of how agents are likely to act in the settings under study in ComSoc.

Expanding PrefLib and Empirical Testing: There is even more room to see the empirical work in ComSoc increase. There does not exist a good culture around experimentation and comparative work such as the research programs for AI outlined by Cohen (1995). While we have attempted to address this gap through the EXPLORE workshop we can expand more. In the coming years it would be good to establish benchmark sets of preferences for the voting and allocation domains and devise a competition around solvers for problems that are known to be NP-hard; e.g., computing Kemeny Winners (Kemeny, 1959) or approximating various hard to compute fairness properties (Bouveret and Lemaître, 2016).

New Communities and Tools: We should continue to expand the publication and use of tools that allow us to translate theoretical results in ComSoc into practice. We have surveyed a number of these tools and research programs around these tools like the work of Qing et al. (2014) are moving these tools into other research areas. New tools are coming online such as the OPRA system from RPI, (`https://opra.cs.rpi.edu/polls/main`) and we expect this trend to continue. The next step is delivering these tools into research with even more communities outside ComSoc.

Preference Drift: Till now, PREFLIB has largely treated preferences as static, as that is the type of data we have received. There are now datasets like the ANES Vote Survey (ED-00013) and the data from VoteLib (Tal et al., 2015) that have a temporal basis. Using these datasets would allow our users to study preference drift, e.g., models of preference change over time, which we see as an exciting avenue for new research.

Hidden Preferences: Since we started PREFLIB, it has largely dealt in explicit preferences. However, there are many settings when preferences are implicit or must be teased out of other signals. For instance, systems collect the books you buy and the songs you listen to and want to learn from this an overall "preference model" for your tastes. While PREFLIB currently doesn't contain this data, it is an exciting avenue for future research.

15.7 Conclusion

In this chapter we have looked back at the first five years of designing, building, supporting, and promoting PREFLIB. Evaluating ourselves on how we measured up to our original intent goals, we notice that we fulfilled most of our goals, shifting a few so that the project stays more focused. We hope that our discussion of the technology used to deploy the site (and the amount of elbow work required) will help others who are considering undertaking the task of building resources for the research community. We are excited by the new avenues of research and new tools that have come online since we established PREFLIB and we look forward to the next five years of research in computational social choice.

Acknowledgments

We would like to thank everyone who has donated their time, energy, data, or support to the PREFLIB project: Haris Aziz, Robert Bredereck, Rafael Bordini, Carleton Coffrin, Sanmay Das, John P. Dickerson, Edith Elkind, Ulle Endriss, Piotr Faliszewski, Toshihiro Kamishima, Omer Lev, David Manlove, Andrew Mao, Jeffrey O'Neill, Dominik Peters, Florenz Plassmann, Nicolaus Tideman, and Hongning Wang. There are others we have certainly left off this list. We attempt to keep an accurate record at `http://www.preflib.org/about.php`.

Our research and the development of PREFLIB has been supported by a number of organizations over the years including NICTA, Data61, CSIRO, IBM, UNSW, the Australian Research Council, the European Research Council, and AOARD.

Bibliography

A. Abdulkadiroğlu, P. Pathak, and A. E. Roth. The New York City high school match. *American Economic Review*, 95(2):364–367, 2005.

M. Aleksandrov, H. Aziz, S. Gaspers, and T. Walsh. Online fair division: Analysing a food bank problem. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

T. Allen, M. Chen, J. Goldsmith, N. Mattei, A. Popova, M. Regenwetter, F. Rossi, and C. Zwilling. Beyond theory and data in preference modeling: Bringing humans into the loop. In *Proceedings of the 4th International Conference on Algorithmic Decision Theory (ADT)*, 2015.

T. Allen, J. Goldsmith, H. Justice, N. Mattei, and K. Raines. Generating CP-nets uniformly at random. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2016.

H. Aziz, S. Gaspers, S. Mackenzie, N. Mattei, N. Narodytska, and T. Walsh. Equilibria under the probabilistic serial rule. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

H. Aziz, O. Lev, N. Mattei, J. S. Rosenschein, and T. Walsh. Strategyproof peer selection: Mechanisms, analyses, and experiments. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2016.

H. Aziz, G. Rauchecker, G. Schryen, and T. Walsh. Algorithms for max-min share fair allocation of indivisible chores. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

K. Bache and M. Lichman. UCI Machine Learning Repository, 2013. URL `http://archive.ics.uci.edu/ml`. University of California, Irvine, School of Information and Computer Sciences.

J. Bartholdi, III, C. Tovey, and M. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

J. Bennett and S. Lanning. The Netflix Prize. In *Proceedings of the KDD Cup and Workshop at ACM Knowledge Discovery and Data Mining (KDD)*, 2007.

N. Betzler, R. Bredereck, J. Chen, and R. Niedermeier. Studies in computational aspects of voting - A parameterized complexity perspective. In *The Multivariate Algorithmic Revolution and Beyond — Essays Dedicated to Michael R. Fellows on the Occasion of His 60th Birthday*, pages 318–363, 2012.

D. Black. On the rationale of group decision-making. *Journal of Political Economy*, 56(1):23–34, 1948.

C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, and O. Sheffet. Optimal social choice functions: A utilitarian view. *Artificial Intelligence*, 227: 190–213, 2015.

S. Bouveret and M. Lemaître. Characterizing conflicts in fair division of indivisible goods using a scale of criteria. *Autonomous Agents and Multiagent Systems*, 30 (2):259–290, 2016.

S. Bouveret, Y. Chevaleyre, and J. Lang. Fair allocation of indivisible goods. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 12, pages 284–311. Cambridge University Press, 2016.

F. Brandl, F. Brandt, and H. G. Seedig. Consistent probabilistic social choice. *Econometrica*, 84(5):1839–1880, 2016.

F. Brandt, G. Chabin, and C. Geist. Pnyx: A powerful and user-friendly tool for preference aggregation. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1915–1916, 2015.

F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

R. Bredereck, P. Faliszewski, R. Niedermeier, P. Skowron, and N. Talmon. Elections with few candidates: Prices, weights, and covering problems. In *Proceedings of the 4th International Conference on Algorithmic Decision Theory (ADT)*, pages 414–431, 2015.

E. Budish and E. Cantillon. The multi-unit assignment problem: Theory and evidence from course allocation at Harvard. *The American Economic Review*, 102(5):2237–2271, 2012.

L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux. API design for machine learning software: Experiences from the scikit-learn project. In *Proceedings of the ECML PKDD Workshop on Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

C. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, 2011.

I. Caragiannis, S. Nath, A. D. Procaccia, and N. Shah. Subset selection via implicit utilitarian voting. *Journal of Artificial Intelligence Research*, 58:123–152, 2017.

G. Charwat and A. Pfandler. Democratix: A declarative approach to winner determination. In *Proceedings of the 4th International Conference on Algorithmic Decision Theory (ADT)*, pages 253–269, 2015.

Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. Preference handling in combinatorial domains: From AI to social choice. *AI Magazine*, 29(4):37–46, 2008.

P. R. Cohen. *Empirical Methods for Artificial Intelligence*. MIT Press, 1995.

V. Conitzer. Making decisions based on the preferences of multiple agents. *Communications of the ACM*, 53(3):84–94, 2010.

J. Davies, G. Katsirelos, N. Narodytska, and T. Walsh. Complexity of and algorithms for Borda manipulation. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI)*, pages 657–662, 2011.

G. de Clippel, H. Moulin, and N. Tideman. Impartial division of a dollar. *Journal of Economic Theory*, 139(1):176–191, 2008.

J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pages 248–255. IEEE, 2009.

J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Optimizing kidney exchange with transplant chains: Theory and reality. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 711–718, 2012.

J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Price of fairness in kidney exchange. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1013–1020, 2014.

C. Domshlak, E. Hüllermeier, S. Kaci, and H. Prade. Preferences in AI: An overview. *Artificial Intelligence*, 175(7):1037–1052, 2011.

G. Erdélyi, L. A. Hemaspaandra, J. Rothe, and H. Spakowski. On approximating optimal weighted lobbying, and frequency of correctness versus average-case polynomial time. In *International Symposium on Fundamentals of Computation Theory*, pages 300–311, 2007.

P. Faliszewski and A. D. Procaccia. AI's war on manipulation: Are we winning? *AI Magazine*, 31(4):53–64, 2010.

P. C. Fishburn. Probabilistic social choice based on simple voting comparisons. *The Review of Economic Studies*, 51(4):683–692, 1984.

I. P. Gent and T. Walsh. CSPLib: A benchmark library for constraints. In *Proceedings of the 5th International Conference on Principles and Practice of Constraint Programming (CP)*, pages 480–481, 1999.

J. K. Goeree and C. A. Holt. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, pages 1402–1422, 2001.

J. Goldman and A. D. Procaccia. Spliddit: Unleashing fair division algorithms. *ACM SIGecom Exchanges*, 13(2):41–46, 2014.

J. Goldsmith, J. Lang, N. Mattei, and P. Perny. Voting with rank dependent scoring rules. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, 2014.

J. S. Hartford, J. R. Wright, and K. Leyton-Brown. Deep learning for predicting human strategic behavior. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2424–2432, 2016.

J. H. Kagel and A. E. Roth. *The Handbook of Experimental Economics*. Princeton University, 1995.

J. G. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.

M. Maschler, E. Solan, and S. Zamir. *Game Theory*. Cambridge University Press, 2013.

N. Mattei. Empirical evaluation of voting rules with strictly ordered preference data. In *Proceedings of the 2nd International Conference on Algorithmic Decision Theory (ADT)*, 2011.

N. Mattei and T. Walsh. PrefLib: A library for preferences. http://www.preflib.org. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT)*, 2013.

N. Mattei and T. Walsh. Empirical evaluation of real world tournaments. *CoRR*, abs/1608.01039, 2016. URL http://arxiv.org/abs/1608.01039.

N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press Cambridge, 2007.

A. D. Procaccia and N. Shah. Optimal aggregation of uncertain preferences. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 608–614, 2016.

A. D. Procaccia, N. Shah, and Y. Zick. Voting rules as error-correcting codes. *Artificial Intelligence*, 231:1–16, 2016.

C. Qing, U. Endriss, R. Fernández, and J. Kruger. Empirical analysis of aggregation methods for collective annotation. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, 2014.

J. Rothe. *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*. Springer, 2015.

G. R. Santhanam, S. Basu, and V. Honavar. Dominance testing via model checking. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, pages 357–362, 2010.

L. S. Shapley. A value for n-person games. In H. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games*, volume 2 of *Annals of Mathematical Studies*. Princeton University Press, 1953.

P. Skowron, P. Faliszewski, and A. Slinko. Achieving fully proportional representation is easy in practice. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 399–406, 2013.

P. Skowron, P. Faliszewski, and A. Slinko. Achieving fully proportional representation: Approximability results. *Artificial Intelligence*, 222:67–103, 2015.

M. Tal, R. Meir, and Y. Gal. A study of human behavior in online voting. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 665–673, 2015.

D. R. M. Thompson, O. Lev, K. Leyton-Brown, and J. Rosenschein. Empirical analysis of plurality election equilibria. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 391–398, 2013.

J. R. Wright and K. Leyton-Brown. Behavioral game theoretic models: A Bayesian framework for parameter analysis. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 921–930, 2012.

L. Xia and V. Conitzer. Generalized scoring rules and the frequency of coalitional manipulability. In *Proceedings of the 9th ACM Conference on Electronic Commerce (ACM-EC)*, pages 109–118, 2008.