Approximate Strategy-Proofness in Large Two-Sided Markets

Lars Ankile, Kjartan Krange, Yuto Yagi

December 5, 2019

Abstract

An approximation of strategy-proofness in large markets is highly evident. Through simulations one can observe that the percentage amount of agents with useful deviations decreases as market size grows. Furthermore, in a matching market, there seems to be a strong connection between the length of preference order lists, the correlation of preferences and the approximation of strategy proofness. Interestingly, approximate strategy proofness is reached easier with shorter length of preference orders and higher preference correlation. These findings justify the use of deferred acceptance algorithm in large two sided matching markets.

1 Introduction

Approximate strategy-proofness is observed in large matching markets where the percentage of agents who can usefully deviate is small. A research paper exploring this concept is Roth and Elliot, 1990[1]. It presents how the labor-market matching between newly graduated physicians and hospitals benefit from the discovery that *"Opportunities for strategic manipulations, are surprisingly small"* i.e. they show how approximate strategy-proofness makes their matching algorithm more robust in a large market. Clearly the implications of the property could in many cases be important when deciding how to implement large matching markets. For example by giving market designers the opportunity to trust the agents to report truthfully even though the DA mechanism is not technically strategy-proof.

In this paper we are running multiple simulations to analyze the approximate strategy proofness. Furthermore, we are aiming at simulating different amounts of correlation between the agents' preferences in the market. In researching this property we hope to find new insights into when strategy-proofness can be a reliable source of security and for what market size and preference correlation it appears effective to a lesser extent.

Our hypothesis is that we will witness a near-strategy-proofness for sufficiently large markets and that the length of the reported preference ordering and the amount of correlation will have an effect on the approximation to strategy-proofness.

2 Theory

The following is a list of theorems with a description of how they apply to our research:

- *Theorem 12.5 Truthful reporting is a dominant strategy for students (teachers) in the student-proposing DA mechanism.* (Parkes and Seuken, 2019[4])
- *Theorem 12.7 No mechanism for two-sided matching is both stable and strategy-proof* (Parkes and Seuken, 2019[4])

  Knowing that deferred acceptance returns a stable matching, and that the proposing side has truthful reporting as its dominant strategy. We can deduce that in general there exists possible useful deviations, and that these will only exits for the agents that are being proposed to.

  Thus, we only have to look for deviations among agents being proposed to.
- *In simple markets, (...) all successful manipulations can (also) be accomplished by truncations (Roth and Peranson, 1999[1])*

  This result is highly useful in the implementation of our simulation. Instead of looking for $k!$ possible permutations of preferences, for preference orders of length k, for the each participant on the receiving side it is sufficient to truncate the preference order once at every position. This makes an intractable problem tractable. **Thus, we only need to check usefulness of truncation deviations**.

  The thorems mentioned allows us to look for deviations in the form of truncation on one side of the matching market without loss of generality, while greatly reducing the computational complexity.

3 Method

In short, our simulation generates matching markets for a given amount of preference correlation, gradually increasing the market size, and for each market size, counts the number of agents who can usefully deviate.

Our algorithm can be divided into three parts: *(i) Preference Generation*, *(ii) Deferred Acceptance*, *(iii) Counting Deviations.*

3.1 Preference Generation

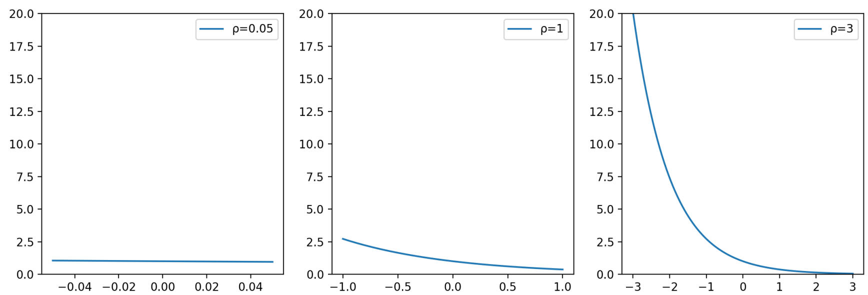

Figure 1: Plots of the three probability distributions we used for modeling different levels of correlation between agents' preferences. Left is the uncorrelated distribution, middle is the moderate, and right is the high correlation.

For our results to be valid we need to realistically model agent preferences in a way that would resemble the distribution over preferences in a real market. An important aspect we needed to capture is the fact that some market participants tend to rise be popular among all agents on the other side, e.g. some men are usually attractive in most women's eyes, which makes them more likely to appear high on women's hypothetical preference orderings. We found that a distribution defined by the inverse exponential function would give us the desired effect. We model three specific market scenarios with differing level of correlation between agents' preferences: (1) almost no correlation, (2) moderate correlation, and (3) high correlation. In figure 1, one can see a plot of the three different distributions used, respectively.

Based on these distributions we first generated preference orderings for one side of the market, where they were given a k length preference list drawn at random according to the distributions described above. Since it e.g. does not makes sense for a hospital to have preferences for students they have not interviewed with or for Norwegian universities to have preferences over students that did not apply to them, we chose to base the other side's preferences on the preferences of the first. This will also make sure we have as high match ratio as possible. We allowed the proposal receiving side to have a variable length preference ordering. E.g. if every male was allowed to have preference orderings of length 10, and there is 5 males who have one specific female among their preferences, that female will have a preference ordering of length 5. With everyone's preferences in order we are ready to proceed to the actual matching.

3.2 Deferred Acceptance

In order to find stable matches, we used deferred acceptance algorithm which is known to produce stable matches. This algorithm takes incomplete preference orders and interpret it as the agent preferring being unmatched to being matched to unlisted agents. i.e. Suppose m's preference over f_1, f_2, f_3 given as $f_1 \succ_m f_3$. Then this will be interpreted as $f_1 \succ_m f_3 \succ_m \emptyset \succ_m f_2$. The following is the pseudo code of our deferred acceptance algorithm.

Algorithm 1 DeferredAcceptance

```
Initialize p in proposers and r in recipients unmatched
while exists p that is unmatched and has someone yet to propose to do
  for p in unmatched proposers do
    r = most preferred r which p has not proposed yet
    if p is not in r's preference list then
      break
    else if r is unmatched then
      (p, r) become matched
    else
      if r prefers p to current match then
        (p, r) become matched
        r's previous match p' becomes unmatched
      else
        no change
      end if
    end if
  end for
end while
```

3.3 Counting Deviations

The paper is trying to find a connection between $D(n) : D =$ number of agents with useful deviations, and $n =$, the agent count on one side of the market. Subsection 3.3 describes the calculation of $D(n)$.
As presented in the 2 Theory , for a given mechanisms resulting stable matching, misreports in the form of truncation deviations on the proposal recipient side of the market is the only possibility for useful deviation. For simplicity we call a recipient *female* in this subsection. When our simulation returns a stable market based on a truthful reporting (given a preference generation dependent on ρ and n) we analyze this matching in the following way:

1. While there are females not analyzed, select a female.

2. Given a female, truncate her last preference order entry.

3. Based on this new preference or-

der, run deferred acceptance and check if this female benefits.

4. Repeat step 2 and 3 until deviation is found ($D(n) = D(n) + 1$) or preference list is of length one in this case mark female as analysed and go to step 1.

We are counting how many females have at least one useful misreport. The $D(n)$ returned will be logged with the corresponding n = market size, k = preference length and ρ = preference correlation. Giving us our final results.

4 Results

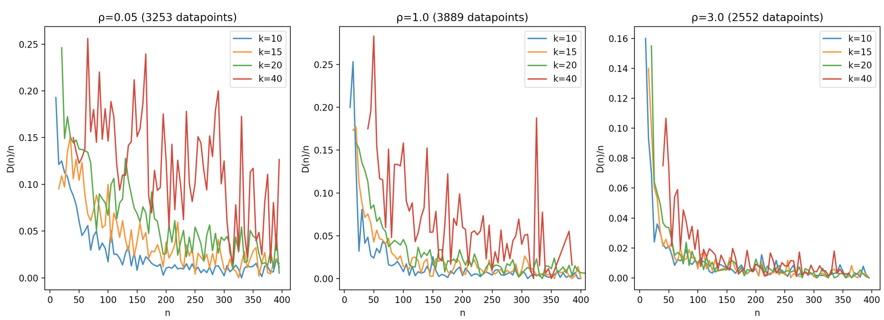

Figure 2: Simulations for $\rho = 0.05, 1.0, 3.0$ each with $k = 10, 15, 20, 40$

5 Discussion

5.1 The General Findings

There are two main types of simulations we have done. The first is for three different preference length ($k = 10, 15, 20$), and three amounts of preference correlations ($\rho = 0.05, 1.0, 3.0$), with a market size running from 5 to 400 (n). In general this shows us the effect of different k and ρ on the approximate strategy-proofness. The second simulation shows how as the market grows even larger (up to $n = 1000$), $D(n)/n$ converges to 0. We show this while keeping k constant at 20.

As k and n grows large, the amount of computation needed grows large. Indeed, our deferred acceptance implementation has a running time of $\mathcal{O}(n^2)$ we run the DA for n recipients with an average preference length k, leaving our total simulation at a average running time of $\mathcal{O}(n^3 k)$. An unrealistic worst case of $\mathcal{O}(n^4)$.

As a general finding, the results show that the market quickly approaches strategy-proofness when the markets grows from a size of $n = 10$ up to around $n = 100$. In fact, in this segments of n, given a high ρ the percentages of agents with useful deviations goes from around 16% to a mere 0.01%. As the market grows even larger, the percentages of agents with useful deviations approaches zero but the speed of convergence slows down.

Our last figure shows how with an increasing number of simulation runs, the data plot gradually becomes smoother with less random spikes in $D(n)/n$. Since the generation of the preference orders are partially based on randomness, having the data based on multiple simulations give us model results gradually approaching an average value. The random based algorithm design, makes our results more easily applicable to the real world as they are based on a huge set of different kind of matching market situations. The plot gradually converges towards a monotone decreasing function. Leading us to conclude that approximate strategy proofness exists in large markets.

5.2 Effect of Varying Preference-Order Length

Figure 3: Longer preference orderings tend to make the opportunities for strategic behavior larger.

In our first figure we see how the three plots: k=10 blue, k=15 orange, k=20 green. The important thing to note is that in general, the plots of $D(n)/n$ for a given n increases with k. Intuitively this makes sense as a larger k quite simply gives more chances of a deviation. For a market designer, the general advice would be that a smaller preference reporting allowed restricts the amount of beneficial misreports available for all agents in total.

5.3 Effect of Varying the Preference Correlation

Figure 4: Lower correlation between agents' preferences tend to increase the amount of opportunities for strategic behavior.

We note that an increasing correlation ρ (ρ=0.05, ρ=1 and ρ=3), leads to smaller ratio of agents who can usefully deviate ($D(n)/n$). This allows market designers with a knowledge of low correlation to expand the market to escure approximate strategy proofness. As for the reason why stronger correlation leads to less possibility for useful deviation, having more agents interested in the same matches makes the rejection chain shorter, thus making it unlikely to return to the devouter.

6 Applications and Improvements

As discussed, a matching market designer has to choose between a stable and strategy proof matching. Parkes and Seuken (2019[4]) argue that unstable markets have a tendency to unravel (possibly with catastrophic consequences) which makes stability an arguably more important property than strategy-proofness. However, as shown in this paper, large market can gain approximate strategy-proofness while maintaining the stability of the outcome.

The proposal receiving side gets their least preferred match and in some cases it might actually be more important for institutions to get a optimal matching. An example could be a market trying to match medical students to residencies at hospitals. It is potentially better if the hospital-optimal matching was achieved instead of the student-optimal one, because the needs for specific talent and knowledge at the hospitals might be more important for a society at large the individual wants of the students.

The results in this paper are likely applicable to the real world. We have seen that in two-sided matching markets it quickly becomes difficult to find useful deviations for the receiving side. This effect is particularly evident when we have relatively short preference orderings and a high correlation between agents in the market. Both of these assumptions we think are common in real applications.

As a familiar instance, Norway has an coordinated admissions system for almost every institution of higher learning, *Samordna opptak*. There, students are ranked according to their high school grades, and students are allowed to create a strict preference order over at most 10 schools. Here we see that every agent on the proposing side has a preference ordering of length at most 10, while every recipient has a preference ordering over all students who have them on their preference ordering. Furthermore, we know from open statistics that a select few schools are extremely popular, while many schools that are not popular and struggle to fill up their class size. This effect is probably common, and is well captured by our medium or high correlation distribution. These observations, together with the fact that the market is very large (tens of thousand of agents), make us believe that this market would potentially reach strong approximate strategy-proofness.

Furthermore, we believe that most large markets will elicit similar characteristics, and will similarly reach approximately strategy-proof.

7 Potential Issues and Next Steps

The following are some points of issues and possible future implementations:

- Our simulation assumes same number of proposer and recipients, and have not modeled the case where there are more proposer than proposal recipients. In the future we should make our simulation compatible with these sorts of markets.

- If we were to use our code to model more varied markets, we would gain a lot from having a highly optimized computational complexity. For instance, adding dictionaries with a $\mathcal{O}(1)$ lookup, could at some lines in the code, drastically speed it up. Greatly reducing the running time for big sets of data.

- We have modelled one-to-one two sided matching market, but there are many other form of matching that require analysis. They include matching with couples, one-to-many matchings, and many-to-many matchings. Next step would be to exoand our model to include these different maching markets.

- This paper has given a empirical proof on the existence of approximate strategy proofness in large two sided matching markets. The next step would be to provide a theoretical proof as to how and why this property arise in large markets.

8 Conclusion

We have found approximate strategy-proofness to be prevalent in large markets. Our models show that percentage amount of deviations decreases most rapidly between a market size of 10 and 100. Approximate strategy proofness increases with the length of preference orders and decreases with preference correlation. Lastly, there are a number of real world applications, like the Norwegian University Application Process, where these findings are of interest.

References

[1] Alvin E. Roth and Elliott Peranson. *The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design.* https://web.stanford.edu/~alroth/papers/rothperansonaer.PDF

[2] Eduardo M. Azevedo and Jacob D. Leshno. *A Supply and Demand Framework for Two-Sided Matching Markets.* https://eduardomazevedo.github.io/papers/Azevedo-Leshno-Supply-and-Demand-Matching.pdf

[3] Benjamin Lubin and David C. Parkes. *Approximate Strategyproofness.* https://dash.harvard.edu/bitstream/handle/1/11879945/Parkes_ApproximateStrategyproofness.pdf

[4] David C. Parkes and Sven Seuken *Algorithmic Economics: A Design Approach.* Harvard University/University of Zurich, 2019.

[5] Alvin E. Roth. *The economics of matching: Stability and incentives.* Mathematics of Operations Research, pages 617–628, 1982.