Effect of Limited Number of Interviews on Matching Markets

Hedyeh Beyhaghi * and Éva Tardos **

Cornell University

Abstract. We study outcome of two-sided matching between prospective medical residents who can only apply to a limited number of positions and hospitals who can interview only a limited number of applicants and show non-intuitive effects in the matching outcomes. We study matching size as our notion of efficiency, and show when the number of interviews is limited, a market with limited number of applications achieves a higher efficiency compared to a market with no limit. Also we find that a system of treating all applicants equally (setting the same limit for their number of applications), is more efficient rather than allowing a small set to apply to one more/less position. This comparison results in a scallop-shape figure 2 that shows expected size of matching with respect to expected number of applications. Finally we show that limiting number of interviews does not always hurt efficiency of matching markets and can improve social welfare in certain cases.

1 Introduction

Matching is a fundamental paradigm in a variety of real-world situations and arises in various domains. In some of these domains, matchings are found through a centralized algorithm such as the deferred acceptance algorithm by Gale and Shapley [6]. Prominent examples include: national residency matching program (NRMP) for assigning medical students to hospital residency programs; schools assignments in some cities the US; and college admissions in some countries [11]. The deferred acceptance algorithm is typically studied assuming that participants express their full list of ordered preferences. Under this assumption, the algorithm finds a stable matching, meaning that no pair of participants would rather deviate from their assigned matches and be matched with one another. However, typically both sides of the match are limited in the number of options they can consider. In some systems, such as the national residency matching program, there are application limits imposed by the system, but even without such explicit limits, preparing applications, and interviewing applicants is costly and time consuming, and hence typically extremely limited. Applications, e.g.

* `hedyeh@cs.cornell.edu`, Dept of Computer Science, Cornell University.

** `eva.tardos@cornell.edu`, Dept of Computer Science, Cornell University. Work supported in part by NSF grant CCF-1563714.

for college admission, often require writing specialized essays. Residency programs, as well as some of the colleges, interview their applicants, which requires significant time and effort. In large matching markets it is not feasible for either side to consider a large set. For instance, in NRMP, there are nearly 5000 residency programs and applicants have to limit their choices. Similarly, interviewing consumes significant time of the hospitals, and so hospitals can only grant interviews to a limited set of doctors.

The goal of this paper is to study the effect on the resulting matching of the number of applications and applicants considered on both sides of the matching. We think of our model as a simplified version of the matching process of the national residency matching program. In this program, the matching mechanism has two stages. In the first stage, doctors express their interest in a set of hospitals, and hospitals choose to interview a subset of the applicants. In the second stage, both doctors and hospitals submit their ordered preference list over the set they interviewed with. The centralized matching market uses these preference lists to perform deferred acceptance algorithm and outputs the final matching.

As a primary metric of the social welfare of the system, we will mainly consider the size of the matching found. If there were no constraints on the number of interviews for applicants or programs, and both parties would be willing to list everybody on the other side, the outcome of deferred acceptance algorithm would be a matching of maximum size (fully matching the smaller side of the market). With limited applicants and limited interviews, the resulting matching can be much smaller. There are a number of different sources of this loss in the size of the matching. When doctors apply to a limited set of hospitals, there will be hospitals that have received less applications than their available positions, or even no applications at all. Even if hospitals receive many more applications than their positions, they might end up with free slots in case the doctors they interview choose other hospitals. Limiting the number of interviews by hospitals further reduces number of potential matches, e.g., doctors might end up with too few of no interviews, and even with interviews, they many not get an offer.An alternative view of the limitations of the hospital side is a limit on offers for one position: based interviewing all or none of the applicants, the may rank only a limited number of their favorite applicants. Such selective offer making is observed some time in academic job markets, when some high ranked departments prefer to wait till next year, rather than make offers to applicants they liked less.

The Model We study the matching resulting from a two-stage procedure inspired by NRMP. In the first stage, applicants request interviews from a limited number of programs. We assume that applicants have a preliminary ranking of the hospitals (based on reputation), and will base their selection on this reputation. Next, the hospitals need to grant interviews to a select set of applicants and reject others. This again is based on a preliminary ranking of applicants based on their CV's. After conducting interviews, both applicants and hospitals can sort their interview set based as their refined preference list, and submit this list to the system running the deferred acceptance algorithm.

For the most part of this paper, we consider a model with the same number of doctors and hospital positions. In section 3 we study a uniform and symmetric preference model where both doctors and hospitals have preferences drawn from uniform distributions, and after the interviews, they again have a preference list that is uniform. As we will show, with this assumption, each applicant applying to their top choices for hospitals, and both hospitals and applicants truthfully reporting their preferences forms a Nash equilibrium of this game. In this model, reordering preferences after interviews does not change the distribution of preferences, so it is equivalent to think of the limitation of hospitals as a limit on the number of top applicants they are willing to rank, i.e., accept for a position. Such selective offer making is observed in academic job market, when some high ranked departments prefer to wait till next year, rather than make offers to applicants they liked less.

In section 4, we study effect of limiting interviews in a tiered structure, where high tiers are preferred to low tiers but inside tiers the preferences are drawn from uniform distributions, as before. We will assume that the effect of the tiers is dominant, in that no party ever prefers a low tier option to a high tier option. In this model, both the applicants as well as the hospitals have an important strategic choice to make: how to divide the limited number of applications and interviews between the tiers. To study the effect of this limit on social welfare in a fair way, we have to model the values of low and high tier matches. We'll elaborate on this model further in Section 4.

Our Results We study effect of limitation on the number of applications and interviews in matching markets and show non-intuitive effects on the matching outcomes.

It seems intuitive that allowing applicants to apply to more positions will cause the size of the resulting matching to increase. In Section 3.1 we show that this is not true. We compute the expected size of the matching formally when each hospital select only one applicant. In this special case, it is best to view this as immediately making an offer to the top applicant. When ranking of applications is correlated, such selective offer making is well-known to lead to many departments unable to fill their positions, as top applicants get many offers, and others don't get any. Notice that we observe the same phenomenon in a model where rankings of different hospitals are independent, and drawn from the uniform distribution. Using simulation, we also show that this counterintuitive phenomenon remains true also when hospitals interview, and then rank more applicants, or make offers less selectively. By computing the expected size of matching, we find the optimal number of applications with which market achieves the largest matching. This gives answer to the market design question for finding optimal number of applications.

A different way to study the effect of additional application is to start with a system where all applicants are allowed to apply for the same number of positions, and all hospitals interview the same number of candidates. Surprisingly, we find that this system of treating all applicants equally is best for maximizing the size of the matching! To show this, we allow a small subset of applicants to apply to one extra position. Surprisingly, we find that this increase in applications in this way, decreases the size of the matching. On the other hand, if starting from the case when all doctors apply to the same number of hospitals, we restrict a small subset of doctors to one fewer application, this too decreases the size of the matching. In the case when all hospitals make an offer to their favorite applicant only, we show the negative effect of adding or deleting an application analytically, for larger number of interviews, we establish the same phenomenon by simulation. The resulting effect on the size of the matching is illustrated by Figure 2 resulting in a scallop-shape function that representing size of matching as a function of the expected number of applications.

In section B we show the same phenomenon even in unbalanced networks with different number of applicants and positions. Although allowing a small set to apply to one more/less position does not have as severe effect, we observe that effect of not treating everybody equally can hurt efficiency in more general settings also.

In section 4 we study the same phenomena in a two-tier model: both doctors and hospitals are divided into high-tier and low-tier, and all parties prefer a match to a high-tier partner to a low-tier one. To evaluate social welfare in this model, we need to model this difference in value. We use a parameter v to model the rate at which both sides value high-tier matches compared to low-tier. In this model, both applicants and hospitals need to make a strategic choice about how to divide their applications and interviews between the two tiers. Intuitively one would think that limiting the number of interviews can only decrease the social welfare. Surprisingly we find that social welfare of this setting can be improved with limiting the number of interviews in some cases.

Related Literature While there has been great interest in finding stable matchings for various markets, there is less known about how limitation on number on preference lists or interviews affect the market. Short preference lists were first considered by Immorlica and Mahdian [7], however, in their model participants preferred to stay unmatched to being matched outside their preference lists. In reality, the limited preference lists are in a large part a product of the limitation explained above: limited time for information gathering and preparing applications, and interviewing applicants, etc. Limited applications by one side of the matching market were considered in various models, including [1], [10].

Arnosti [1] studies matching with short preference lists for applicants using a *large market approximation*, which we will also use in our paper. The large market approximation considers the limiting case where number of applicants and positions grow while other parameters of the market, like the reported preference lists, remain fixed.

Small number of applications due to limited budget or time, have been considered by [2] and [3]. In these models, applicants need to be strategic about selecting a small set of positions to apply for. Specially, when the market is correlated, applicants have a higher chance of admission in the less popular positions and they have to strategically come up with a short list balancing between maximizing their chance of admission in a high-ranked position and minimizing their chance of unemployment. Avery and Levin[2] study early admission for college application. In early admission, that has become popular in recent years, students can select one college in early stage. They consider a model with two schools where students can choose between the more popular or the less popular school in the early stage. In [3], the authors of this paper used a tiered model to study the case where deferred acceptance is implemented with applicants submitting a limited list of applications and hospitals granting interviews to all applicants. Their findings show how applicants decide where to apply and what would the quality of outcome be in the resulting equilibrium. In this paper, we focus more on the socially efficiency of outcomes, rather than outcome of strategic behavior. Since we study independent preferences, compared to correlated models in these two papers, participants are not motivated to strategize their actions, and they only report their true preference lists. Also in our model both sides, and not only the applicant side, have limits on their length of their preference lists.

There are also a few papers considering limited lists on both sided of the market. Drummond et al. [4] and Kadam [8] study the effect of limited application and interview lists on the matching. In their models, both doctors and hospitals are (almost perfectly) vertically differentiated, so applicants and hospitals have almost perfect information on who they would match to, given full access to all parties preferences. While we agree that there are better and worse applicants, and better and worse residency programs, we believe that a lot of the choices are much more personal and idiosyncratic. To model this, we assume that applicant's, as well as hospital's, preferences have a significant component that is random.

Drummond et al. [4] investigate an interviewing game, where medical students and hospitals interview each other as the first stage of matching. For their theoretical analysis, they make the following assumptions: hospitals offering the residency programs have a common preference list; preference of medical students is distributed as a slightly different version of a common ordering; residents choose two hospitals to interview with; and hospitals grant interviews to all of their applicants. They show that under this assumption, applicants select two consecutive hospitals in the common ordering that closely correspond to their own rank. Using simulation, they also show that with high dispersion, this is not an equilibrium anymore and doctors select hospitals with less competition in this case. Similar to our model, they consider a two-stage game that involves an interview stage. Compared to their results in the independent model that we study, participants are not strategic and only report their favorite choices. In light of this we study effect of limitation in number of applications without considering effect of strategicness.

Similarly, Kadam [8] studies a two-stage model in which both doctors and hospitals have limitation on number of interviews. Similar to our paper, he shows positive effects of limiting interviews. In that model limiting interviews might increase the size of the matching while decreasing overall social welfare. Also it can result in improved outcomes for weaker applicants while having a negative impact on strong applicants. In both of these papers, they consider markets where both doctors and hospitals are (almost perfectly) vertically differentiated. Since in our model both sides are truthful, unlike these papers our focus is on the effect of number of applications and interviews on size of the matching.

Kleinberg at al.[9] focus on interview as an information gathering step. They design an order for interviews and offers with the goal of minimizing the number of interviews needed in order to find the stable matching. Although the positions are vertically differentiated in their model, they assume that applicants do not know the perfect order prior to the process and the interviews help them to figure it out. They continue the interviewing process until they find an efficient matching, and study the overall number of interviews needed. In our model the number interviews are fixed ahead of time, and we study the effect of this limitation on the matching.

2 Model

We consider a two-sided market consisting of doctors and open positions in hospitals (hospitals for short). We assume there are n doctors and n positions, and that each hospital has only one open position, so we focus on one-to-one matching.

For most of the paper we assume there is a universal limit on number of applications by doctors and number of interviews by hospitals. The fixed limit on the number of applications by a doctor is inspired by NRMP that grants all doctors 10 initial applications. We relax this assumption later and compare effect of having a varied allocation of applications. We also have a fixed limit on the number of interview conducted by hospitals, modeling that hospitals spend more or less the same time and budget on interviews.

We consider a two-stage matching mechanism. In the first stage doctors and hospitals gather information and learn their preferences. In this stage, doctors submit a list of hospitals as their interview requests. Hospitals then choose a subset from their applications to conduct interviews. In the second stage, at the end of interviews, both doctors and hospitals order the list that they interviewed with, based on their preferences. Once doctors and hospitals submit their ordered preferences, a doctor-proposing deferred acceptance algorithm is performed to determine the final assignment of doctors to hospitals. The algorithm starts with all doctors unmatched. In each step, all unmatched doctors, who have not yet exhausted their options, apply to their most preferred hospital among those to which he/she has not yet applied. Now each hospital tentatively accepts their most preferred doctor from the doctors now applying and the one who has been tentatively assigned to the selected hospital, and rejects all other applicants. The procedure is repeated until all unmatched doctors have been rejected from every hospital in their lists.

In the beginning of the procedure, everybody has prior preference over the other side. For sections 3, we assume that everybody's preference comes independently from uniform distribution over the other side. Both doctors and hospitals refine their preference order for the list they interviewed. However, we assume that the overall distribution of doctors preferences and hospitals preferences will be an outcome of new independent draws from uniform distribution. Refinement in the way mentioned, does not have an effect on our analysis. For analytic purposes we can assume that both the initial distribution, as well as the ranking after the interviews, are uniformly random.

We claim that in this model, neither doctors or hospitals will be strategic: When doctors apply to hospitals for interview, they pick their favorite hospitals, with respect to their prior preference. Also hospitals select their favorite doctors to interview based on their prior. After interviews doctors and hospitals refine their preference order based on what they learned in the interviews stage. They are truthful with the orders they use for doctors/hospitals they interviewed. To see why, note that due to the symmetry and independence in the market, no doctor or hospital is considered more popular than others. Therefore neither doctors or hospitals can benefit from applying or making offer to somebody other than their favorite selection. After choosing their lists, again neither hospitals or doctors can benefit from reordering the lists. It is not beneficial for the doctors (the proposing side) as shown in [5]. Also as shown in [7], with high probability it is not beneficial for hospitals either.

We assume that throughout the whole process, there is no coordination between doctors or hospitals. If hospitals knew what applicants had more/less interviews/offers, they may prefer making offers to doctors with less interviews, as these are more likely to accept. Our model assumes no such strategic behaviour.

For most of the paper we focus on size of the matching as our notion of efficiency. We define efficiency as ratio of size of matching outcome to size of maximum matching. Since we assume all doctors and hospitals prefer to be matched rather than unmatched, they find every applicant/position acceptable, therefore in a maximum-size matching everybody on the small side of the market is matched.

Definition 1 (Social Welfare). *We define social welfare to be the ratio of matching outcome compared to maximum-size matching. In a two sided market, maximum-size matching will be the size of the smaller side of the market.*

Throughout this paper we study a "large market approximation" of the setting as introduced by [1]. This assumes that properties of the market, such as number of applications, number of interviews and the ratio between sizes of the two sides of the market, is fixed while the size of market grows to ∞ and studies properties of the matching in the limiting case. The example below is a simple example of this approximation that will be used in next sections.

Example 1. When doctors are allowed to apply to one positions, in case they apply to their favorite hospital, social welfare of matching is $(1 - 1/e) \approx 0.63$. This is because the size of matching is equal to the number of matched hospitals and any hospital that has an application can be matched by offering a position to one of its applicants. Since applicants have only this one application, they accept the offer. The probability of a hospital having any application is $1-(1-1/n)^n$ which tends to $1-1/e$ with n approaching ∞.

In section 3, we evaluate the size of matching as function of number of applications with different number of offers or interviews by hospitals. We assume that the market is symmetric in the sense that doctors apply to the same (closely bounded) number of applications . In section B, we study unbalanced networks where the set of doctors and hospitals are not the same size. In section 4, we study effect of interviews in a more general model, explained later. Although it is intuitive to believe that limiting number of interviews only hurts social welfare, we show an example with improved social welfare when limiting the number of interviews.

3 Effect of Number of Interviews

In this section we study effect of number of interviews/offers. In 3.1 to 3.2 we find the efficiency of matching as a function of the expected number of applications by doctors, when where hospitals only select one applicant. We begin with the case where all doctors apply to the same number of hospitals k and then we move to doctors selecting varying number of applications. We find allocating the number of allocations equally results in the most efficient matching and if we allow a small subset to apply to an extra position or restrict a small set to apply to one less position, this results in a smaller matching. In section 3.3 we study the matching outcome when hospitals grant multiple interviews/offers. We find that treating applicants equally is most efficient even in this more general model.

3.1 Integer Number of Applications and Making One Offer

We study social welfare of the matching with respect to number of applications by doctors. If all applicants are allowed to apply to the same number of positions, it is intuitive that size of matching is increasing in the number of applications. In this part we show that it is not true. Having more applications can have both positive and negative effects. Although the number of hospitals who receive applications is increasing as the number of application grows, the number of doctors who receive multiple applications grows also. These two opposite effects create a trade-off between higher and lower limits on interviews number. In this section we study the trade-off between two effects and find the optimal number of applications that results in largest matching size.

We start with the simple case where doctors are allowed to send one application. In this case each doctor receives at most one offer. Therefore doctors accept all offers they receive and hospitals know that if they make offer to any of their applicants they will be matched. So in this case the size of matching is the number of hospitals who receive any application. As shown in example 3 the size of matching tends to $1-1/e$ in the limit as n increases.

To observe the effect of having more applications, consider the extreme case where doctors do not have a limit and are allowed apply to all hospitals. If the size of matching is monotone increasing in the number of applications, this case creates the largest matching. Since hospital preferences are uniformly random, each hospital selects an applicant to make an offer to uniformly random, and so the number of doctors who receive an offer is:

$$lim_{n\to\infty}(1-\frac{1}{n})^n = 1-\frac{1}{e}$$

which shows that although the number of applications increased from one to N there is no increase in the size of matching. Therefore either $1-\frac{1}{e}$ is the largest matching possible or applying to $1 < k < n$ positions, achieves a larger matching which, if true means that matching size is not increasing in the number of applications.

Definition 2 (covered hospital). *A hospital is covered if it receives at least one application.*

Suppose that all doctors are allowed to apply to k positions. When everybody has exactly k applications, the number of total application is nk. With a similar analysis to example 3, the number of covered hospitals is $n(1-e^{-k})$.

Proposition 1. *When all applicants apply to exactly k uniformly random positions, with large market approximation the expected fraction of covered hospitals is* $(1-e^{-k})$.

Proof. Similar to example 3, the probability of a hospital receiving any application is $1-(1-1/n)^{nk}$, which tends to $(1-e^{-k})$ in the limit.

Next we find the probability of acceptance of a random application. In the previous section we showed that doctors apply to their favorite positions and since doctors preferences are independent, their applications are also independent. Also since with large n and small number of applications, the probability that two random application of doctor go to the same hospital approaches to 0, with large market assumptions, we assume that applications of a doctor are independent. And we can assume that the destination hospitals of all applications are selected uniformly and independently at random. With this symmetry in the market, the probability of acceptance of any two random application is the same.

Proposition 2. *When applicants apply to k random positions and all hospitals offer a position randomly to one of their applicants, the probability for a doctor to be matched is:*

$$(1-(1-\frac{(1-e^{-k})}{k})^k)$$

Proof. With the above arguments we can assume that probability of acceptance of random applications is the same. Therefore the probability of acceptance of a random application is equal to the average probability of an application. This average probability, depends on two parameters: total number of offers and total number of applications. As discussed previously the total number of offers is equal to number of covered hospital, that is $n(1 - e^{-k})$. The total number of applications is nk in the symmetric case where all doctors apply to the same number of position. Therefore expected probability of acceptance is $(1 - (1 - \frac{n(1-e^{-k})}{nk})^k)$.

Figures 3.1 shows the size of matching with respect to the number of applications.

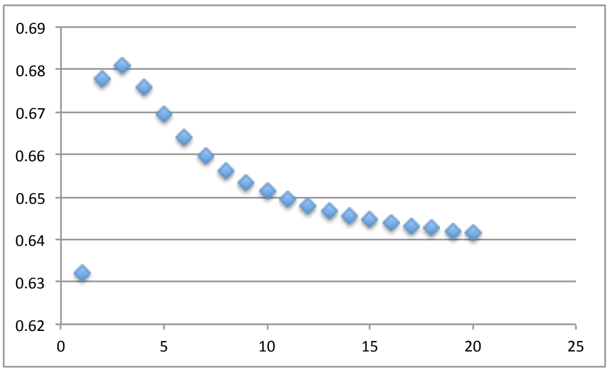

Fig. 1. Size of matching with respect to number of applications, when hospitals make only one offer.

As seen in figure 3.1 the maximum size of matching occurs when all doctors apply to 3 hospitals. Therefore form a market design point of view allowing three applications in this case is optimal.

Proposition 3. *In markets where with the same number of applicants as positions, if all applicants are allowed to apply to k positions and hospitals grant a single offer, allowing applicants to apply to* 3 *hospitals results in the largest matching outcome.*

The biggest jump in the size of matching occurs when moving from one application to two applications. When increasing the number of applications from one to two, the fraction of covered hospitals changes from $1 - 1/e \approx 0.63$ to $1 - 1/e^2 \approx 0.86$, which is the highest increase when adding more applications. This causes the highest increase in the matching size. The increase stops with only three number of applications, with which the fraction of covered hospitals is $1 - 1/e^3 \approx 0.95$. From this point forward the negative effect takes over and because of the random allocation of hospitals some doctors receive multiple offers while others receive none.

3.2 Making One Offer, Fractional Points: scallop-shape

With the above formula we find the optimum number of applications when doctors apply to the same number of hospitals. In this section we study the effect of allowing some doctors to apply to one more/less position on matching size. Surprisingly, we observe that granting extra applications to a small set of doctors and also retracting an application from a small set both hurt the market; suggesting that unfair treatment is not efficient in terms of social welfare.

For this part we find the size of matching as function of expected number of applications. To be compatible with the above section at integer points, we assume that for expected number of applications $k < x < k + 1$, the applicants either apply to k or $k + 1$ positions as was explained in the Section 2, in this case the fraction of covered hospitals is $(1 - e^{-x})$.

Proposition 4. *The fraction of covered hospitals with expected number of applications x is $(1 - e^{-x})$.*

Theorem 1. *The size of matching with expected number of applications $k < x < k + 1$, when applicants either apply to k or $k + 1$ positions is:*

$$(\lceil x \rceil - x)(1 - (1 - \frac{(1 - e^{-k})}{k})^k) + (x - \lfloor x \rfloor)(1 - (1 - \frac{(1 - e^{-k})}{(k + 1)})^{k+1})$$

Proof. Same as previous section, the probability of an acceptance of an application equals $\frac{\text{number of covered hospitals}}{\text{number of applications}}$. Therefore the fraction of matched applicants or the probability of an applicant being matched is:

$$(\lceil x \rceil - x)(1 - (1 - \frac{(1 - e^{-k})}{k})^k) + (x - \lfloor x \rfloor)(1 - (1 - \frac{(1 - e^{-k})}{(k + 1)})^{k+1}).$$

The scallop-shape figure 2 represents the above formula, i.e. the size of matching as function of expected number of applications. The unusual behavior of the function at integer points shows that allowing a small group to apply to one more position, or limiting the number of applications of a small group to one less application, has a negative effect on size of matching. Also, if we compare this diagram with figure 4, we observe that with the same structure of applications; i.e. same number of doctors applying to k or $k + 1$ positions, we observe that this unusual shape belongs to the case with limited number of applications, and if there is no limit on the number of applications the size of the matching is monotone increasing.

3.3 Granting multiple interviews

In this section we focus on the more realistic case, where hospitals grant multiple interviews. Unlike the previous subsections, this case actually uses the usual procedure of Gale-Shapley mechanism. In this scenario, hospitals have backup options in case their top applicant turns down the offer.

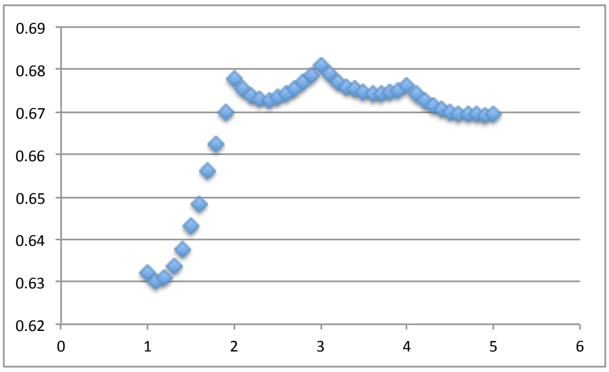

Fig. 2. Size of matching with respect to expected number of applications, when hospitals make only one offer.

After hospitals conduct their interviews, they submit an ordered list of their interview set as input to deferred acceptance algorithm. Therefore, the procedure is more complicated then in the previous subsections: any application by doctors can be rejected before the interview or after the interview in deferred acceptance process. Because of this complication, we used simulation to find the resulting matching size.

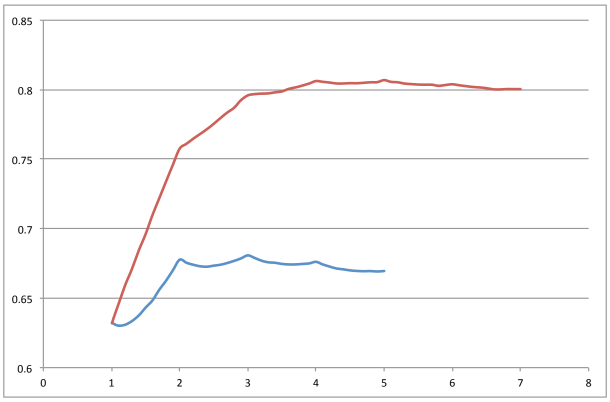

Fig. 3. Size of matching with respect to expected number of application. The top red curve represents the setting where hospitals allow two interviews. The lower blue curve represents the setting with one interview and is re-drawn from Figure 2 for reference.

We considered a market with 10,000 hospitals and doctors and simulated the outcome of the procedure explained in the introduction for different expected number of applications. The red curve in figure 3, shows the size of matching with respect to expected number of applications when hospitals grant two interviews.

As the figure suggests, the structure of the matching size even with multiple interviews, has the same structure as single interviews (the blue curve).

4 Improvement of social welfare with restricting interviews

In this section we study effect of limited number of interviews in a more general model studied in [3]. In the models studied in previous sections, limiting the number of interviews by hospitals only had a negative effect on matching outcome. We show that this property does not extend to all markets, and limiting the number of interviews can have positive effect on social welfare. This finding is similar to [8] who shows limiting the number of interviews can increase the size of matching.

In the models studied so far, limiting the number of interviews by hospitals is hurtful in terms of social welfare. With limiting number of interviews for hospitals some applications are rejected before the interview stage, causing both doctors and hospitals to submit shorter lists to as input to deferred acceptance algorithm. It is not hard to see that this results in smaller matchings.

In a previous paper [3], the authors of this paper studied a market that was a combination of independence and correlation. They assumed that both sides are divided into tiers, where everybody prefers higher tiers to lower tiers (correlation) and inside a tier the preferences are uniformly random (independence). We can view the model studied in previous section as their model with only one tier. In a two-tier market they assume that matching to a high-tier doctor/hospital has value $v > 1$ for participants, while matching to a low-tier doctor/hospital has value 1, and social welfare is the sum of the values of all participants.

Unlike the model studied in previous section, in a two-tier model doctors are strategic when they are limited in number of applications. Since the market is not symmetric and high-tier hospitals are more popular, doctors strategically decide how to select their limited number of hospitals both to increase the likelihood of being matched to a high-tier hospital, and having backup offers from low-tiers in case they do not succeed in receiving an offer from high-tier.

Definition 3 (social welfare in a two-tier market [3]). *Social welfare is the sum of the value of matched doctors and hospitals, such that the value of matching to a high-tier doctor/hospital is v while the value of matching to a low-tier doctors/hospitals is* 1*.*

They show that strategic action of doctors can hurt social welfare. As they show, in equilibrium in a two-tier market, doctors have a tendency to apply more to high-tier hospitals. This strategic behavior sometimes has negative impact on social welfare of the matching. Applying to more high-tier hospitals makes high-tier hospitals congested and also reduces the number of matches of the high-tier doctors (more valuable doctors in terms of social welfare) since they do not use the benefit from potential matches to the low tier.

Although restricting the number of interviews can be hurtful in most cases, we show an example where it can improve social welfare.

Example 2. Consider a case where the number of top-tier doctors, low-tier doctors, top-tier hospitals and low-tier hospitals is the same. Suppose all doctors have 2 applications and hospitals grant at most 6 interviews. When v, the value for high-tier, is ≈ 1.91, with the setting where all applications are accepted for interview, all high doctors send both of their applications to the high tier. While with limited number of interviews the probability of their applications turning into offers decreases. So high doctors have incentives to apply to the low tier and they randomize between sending one or two applications to the top tier. This new flow of applications from top tier doctors to low tier hospital increases the size of the matching and improves the overall social welfare. Therefore in this case a limit on the number of interviews by hospitals (6) results in a more efficient matching than unlimited number of interviews.

References

1. N. Arnosti. Centralized clearinghouse design: A quantity-quality tradeoff. *Working Paper*, 2016.
2. C. Avery and J. Levin. Early admissions at selective colleges. *The American Economic Review*, 100(5):2125–2156, 2010.
3. H. Beyhaghi, D. Saban, and E. Tardos. Effect of selfish choices in deferred acceptance with short lists. 2017. Match-Up.
4. J. Drummond, A. Borodin, and K. Larson. Natural interviewing equilibria for stable matching.
5. D. Gale and L. S. Shapley. College Admissions and the Stability of Marriage. *American Mathematical Monthly*, 69(1):9–15, 1962.
6. D. Gale and M. Sotomayor. Ms. machiavelli and the stable matching problem. *The American Mathematical Monthly*, 92(4):261–268, 1985.
7. N. Immorlica and M. Mahdian. Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 53–62 (electronic). ACM, New York, 2005.
8. S. V. Kadam. Interviewing in matching markets. *Working Paper*, 2015.
9. R. Kleinberg, B. Waggoner, and E. G. Weyl. Descending price optimally coordinates search. In *Proceedings of the ACM Symposium of Ecomomic and Computation.* ACM, New York, 2016.
10. F. Kojima and P. A. Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–27, 2009.
11. A. E. Roth and M. Sotomayor. Two-sided matching. *Handbook of game theory with economic applications*, 1:485–541, 1992.

Appendix

In section A we study the matching when there is no limit on number of interviews by hospitals. In section sec:unbalanced we study unbalanced markets with unequal number of doctors and positions.

A Results without Limit for Granting Interviews

In this section we recall the analysis from [3] where there is no limit on number of interviews. As seen in figure 4 the size of matching in this case is increasing with respect to expected number of application.

We recall the analysis for the baseline case, studied in our previous paper [3]. In this case there is no limit on the interviews by hospitals. We considered the game when doctors are allowed to list only up to k hospitals, for parameters $k << n$.

Example 3. When $k = 1$, and all doctors apply to their favorite hospital, social welfare of matching is $(1 - 1/e) \approx 0.63$. This is because the size of matching is equal to the number of matched hospitals and any hospital that has an application can be matched by offering a position to one of its applicants. Since applicants have only this one application, they accept the offer. The probability of a hospital having any application is $1 - (1 - 1/n)^n$ which tends to $1 - 1/e$ with n approaching ∞.

When doctors apply to more than one hospital ($k > 1$), Arnosti [1] proposes the following formula for the probability p of getting an offer from a single application as n grows to ∞. The main idea is that in the limit as n goes to infinity, the different applications of an applicant have independent probability leading to an offer.

Proposition 5. *[1] Using the above large market approximation with each doctor applying to k hospitals, the probability of a single application getting accepted is defined by equation:*

$$(1 - p)^k = e^{-(1-(1-p)^k)/p} \tag{1}$$

Proof. Since each application is accepted independently with probability p, the probability that a doctor with k applications eventually gets matched is $(1-(1-p)^k)$, and so the expected expected number of hospitals a single doctor applies to throughout the Gale-Shapley process to is

$$1 + (1 - p) + (1 - p)^2 + ... + (1 - p)^{k-1} = \frac{1}{p}(1 - (1 - p)^k)$$

Each hospital remains matched if it receives any applications. Using the approximation that all these offers are independent, the probability that a hospital didn't get an application is then

$$(1 - 1/n)^{\frac{n}{p}(1-(1-p)^k)} \approx e^{-(1-(1-p)^k)/p}$$

and so the expected number of matched hospitals is then $n(1-e^{-(1-(1-p)^k)/p})$. The equation claimed by the lemma then follows, as the number of matched hospitals is the same as the number of matched doctors.

Figure 4 shows the expected size of matching as a function of expected number of applications by doctors. We extended the figure for a fractional number of applications. When expected number of applications is some value x, such that $k < x < k+1$, doctors have either k or $k+1$ applications. As seen in the figure, the matching size is increasing in the number of applications.

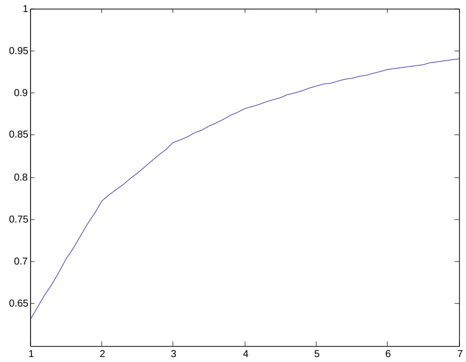

Fig. 4. Size of matching, with no limit on interviews. The horizontal axis is the expected number applications and the vertical axis is the size of matching. The size of the matching is increasing in the expected number of applications.

B Unbalanced Networks

In this section we show that the phenomenon of the positive effect of symmetric number of applications is not limited to the setting with same number of applicants and positions. By generalizing the model from previous sections to unbalanced networks, where the number of applicants and positions is not equal, we still observe the scallop-shape structure. We study how the difference between the size of the sides changes the scallop-shape structure and optimum number of applications.

We study the case where hospitals make one offer and the ratio of number of hospitals to number of doctors is r. Similar to the previous section the probability of acceptance of a random application in this case is the ratio of number of total interviews to number of total applications. The number of total applications, when all doctors apply to k positions is nk, therefore the expected number of application a hospitals receives is k/r. Similar to example 3, the number of covered hospitals (number of total interviews) in this case comes out to be

$rn(1-e^{\frac{k}{r}})$ with large market approximation. Therefore the probability of a random application turning into an offer is:

$$p = \frac{rn(1-e^{k/r})}{nk} = \frac{r(1-e^{k/r})}{k}$$

The figure B below shows the size of matching for $r = \frac{1}{2}, 1, 2$. Note that by size of matching, we mean ratio of matching to maximum size matching. In the below picture red function refers to $r = 2$, blue to $r = 1/2$ and the green to $r = 1$.

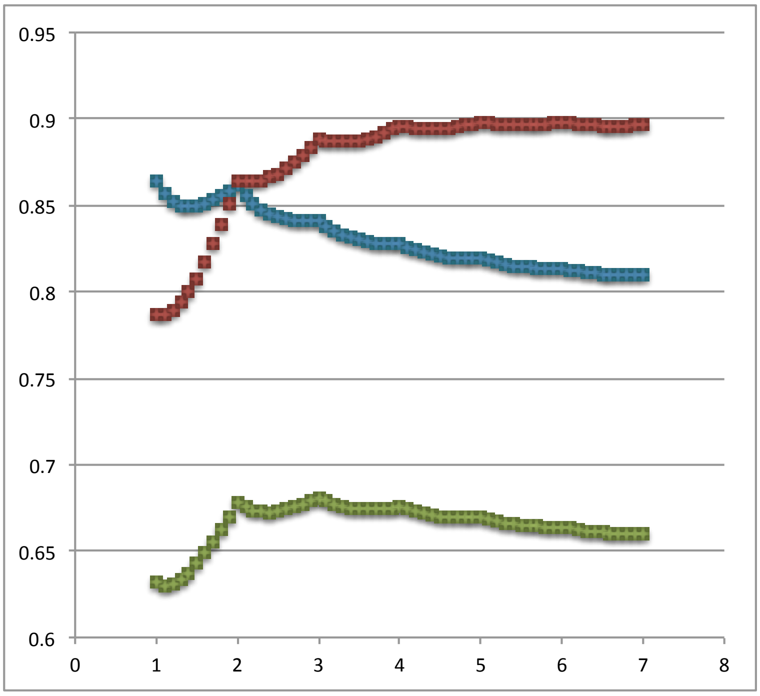

Fig. 5. Size of matching with respect to expected number of applications. The red curve refers to $r = 2$, blue curve to $r = 1/2$ and the curve to $r = 1$, where r is the ratio of number of hospitals to doctors.

As seen in the figure the same structure happens for unbalanced network and the optimal number of applications depend on factor of balancedness. When the number of hospitals is half the number of doctors, the market achieves its maximum size when doctors apply to just one position each. With more applications, hospitals become congested and the probability that they grant interviews to doctors who receive multiple interviews increases. In contrast, when the number of hospitals is more than number of doctors, higher number of applications has a positive effect. With more than one application, more hospitals receive at least one application and while when k is still small, the hospitals are not too congested. Therefore, probability of a doctor receiving many interviews is small.

Note also that unbalanced networks achieve a larger fraction of the maximum matching compared to balanced network. To understand why this is the consider the three cases when all doctors apply to one position. As discussed previously, in this case all hospitals who receive an application will be matched. When $r = 1/2$, i.e., there are two as many doctors as hospitals, therefore compared to the balanced case, a larger fraction of hospitals receive at least one application. Also in the case where $r = 2$, i.e., there are twice as many hospitals as doctors, with one application doctors have twice potential positions compared to the balanced case, therefore there is less congestion on the hospitals side and a large fraction of doctors can be matched.