Published in "Physica A: Statistical Mechanics and its Applications
444: 397–402, 2016"
which should be cited to refer to this work.

Analysis of ground state in random bipartite matching

Gui-Yuan Shi [a], Yi-Xiu Kong [a], Hao Liao [b,*], Yi-Cheng Zhang [a]

[a] *Physics Department, University of Fribourg, Chemin du Musée 3, CH-1700 Fribourg, Switzerland*

[b] *Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Nanhai Road 3688, 518060, Shenzhen, China*

HIGHLIGHTS

- We applied the Kuhn–Munkres Algorithm to analyze ground state of bipartite matching problem.
- We settle down the quantity and distribution of blocking pairs in the ground state.
- The stability of ground state decreases exponentially with better connectivity in the network.
- The scope of application of the ground state is extended to a broader initial conditions.

Bipartite matching problems emerge in many human social phenomena. In this paper, we study the ground state of the Gale–Shapley model, which is the most popular bipartite matching model. We apply the Kuhn–Munkres algorithm to compute the numerical ground state of the model. For the first time, we obtain the number of blocking pairs which is a measure of the system instability. We also show that the number of blocking pairs formed by each person follows a geometric distribution. Furthermore, we study how the connectivity in the bipartite matching problems influences the instability of the ground state.

1. Introduction

Bipartite matching problems, appear in many social processes like the marriage problem between men and women, college admission problem between students and universities, assignment between workers and jobs, and also the choice making between buyers and sellers. Due to its various applications in the real world, not only economists but also the statistic physicists are attracted by the bipartite matching problems.

Gale and Shapley first introduced the stable marriage problem, a one-to-one two side matching [1], which is the most important bipartite matching problem. Bipartite matching problems was rephrased to an optimization problem by assigning agent i an energy term ε_i to represent his/her satisfaction. The so-called optimal matching has the minimum energy among all the possible matchings under certain assumptions. And a matching is called stable only if there are no two agents man i and woman α, each of whom prefer the other to their spouse in x. Such a pair is said to be a blocking pair [2] with respect to x, abbreviated as *BP* hereafter. Gale–Shapley algorithm proves the existence of a stable solution in the matching problem under any circumstance. By neglecting the stability of the state, Mézard and Parisi applied the replica method of spin glass theory to study the global optimal solution of one-to-one two side matching problem [3,4]. Afterward, Zhang et al. studied the scaling behavior and partial information matching of the marriage problem [5–7]. Furthermore, Dzierzawa and Oméro introduced

* Corresponding author.
E-mail addresses: jamesliao520@gmail.com (H. Liao), yi-cheng.zhang@unifr.ch (Y.-C. Zhang).

Table 1
Using of the variables in this paper.

Variable	Description
n	The number of male agents or female agents
$\mathscr{M}$	The set of male agents
$\mathscr{W}$	The set of female agents
x	A matching $x : \mathscr{M} \to \mathscr{W}$
$M_{i,\alpha}$	The energy of man m_i if woman w_α was matched to him
$W_{\beta,j}$	The energy of woman w_β if man m_j was matched to her
$H_{i,\alpha}$	The mean energy of man m_i and woman w_α if man m_i and woman w_α were matched
$E_H(x)$	Average energy per person under a given matching x
$E_H(GS)$	Average energy per person for the ground state
$\langle E_H(GS)\rangle$	Mean value of $E_H(GS)$
$f(\epsilon)$	Probability distribution function of individual energy in the ground state
P_k	The probability that a man forms k BPs
N_{BP}	The number of BPs
N_{bp}	The number of men/women who form BPs (note that a person may have more than one BP)
σ	The fraction that the number of persons one can know in all the agents of the other sex
S_σ	The probability of a person could not find a BP under the constraint of a person only knows $n\sigma$ persons of the other sex.

the acceptance threshold and thus improved the matching result [8]. Recently, Zhou et al. studied the bidirectional selection problems from a new perspective of social networks [9,10]. In this paper, we aim to compute the number of blocking pairs in the global optimal solution and then analyze the properties of this solution.

The rest of this paper is organized as follows. In Section 2, we introduce the Gale–Shapley model. In Section 3, we use Kuhn–Munkres algorithm to find the global optimal solution of the bipartite matching problem. In Section 4, we analyze the blocking pairs in the ground state, with both numerical and analytical approaches. Then, we analyze the stability of the ground state based on the number of blocking pairs.

2. Model

The Gale–Shapley model is a matching model in which there are two sets of participants, men and women. We denote by $\mathscr{M} = \{m_1, m_2, \ldots, m_n\}$ and $\mathscr{W} = \{w_1, w_2, \ldots, w_n\}$, the set of men and women, respectively. A matching is a one-to-one mapping between the two disjoint sets, *i.e.*, an invertible bijection $x : \mathscr{M} \to \mathscr{W}$. A matching x can be denoted as:

$$x = [(m_1, x(m_1)), (m_2, x(m_2)), \ldots, (m_n, x(m_n))],$$

where $x(m_i) = w_\alpha$ means the woman who matched with man m_i, and $x^{-1}(w_\alpha) = m_i$ means the man who matched with w_α [11].

Previous works [3,7,8] assumed a discrete uniform distribution for the energy term $\varepsilon = 1, 2, \ldots, n$, but we hold that the energy should be independent of the model size n. Therefore we suggest that the energy ε_i follows a continuous uniform distribution on $[0, 1]$, and result will be consistent with previous work [3,7] since the size of the model is large enough to eliminate the difference between discrete and continuous distribution.

For each matched pair (m_i, w_α), we denote by $M_{i,\alpha}$ and $W_{\alpha,i}$ the energy of man i and woman α respectively. We denote by M and W, the matrices composed of the energies $M_{i,\alpha}$ and $W_{\alpha,i}$, respectively. For a given matching $x = [(m_1, x(m_1)), (m_2, x(m_2)), \ldots, (m_n, x(m_n))]$, the average energy per person $E_H(x)$ is

$$E_H(x) = \frac{1}{2n}\left[\sum_{i=1}^{n} M_{i,x(m_i)} + \sum_{i=1}^{n} W_{x(m_i),i}\right] = \frac{1}{n}\sum_{i=1}^{n} H_{i,x(m_i)}, \tag{1}$$

where we defined the mean energy of man i and woman α $H_{i,\alpha} = \frac{1}{2}(M_{i,\alpha} + W_{\alpha,i})$. We denote by $E_H(GS)$ the energy of the ground state x_{GS} of the model, which is the matching x_{GS} that minimizes the energy $E_H(x)$.

Table 1 summarizes the notations adopted in this paper for the variables of the model.

3. Energy analysis

We apply the Kuhn–Munkres algorithm [12,13] to find the ground state of the model. The Kuhn–Munkres algorithm is described in Table 2. We study the properties of the ground state when changing the size n of the model.

Fig. 1 shows how the average energy $\langle E_H(GS)\rangle$ of 100 realizations of ground state depends on the size n of the model. Numerical simulation results were fitted with a power-law $\langle E_H(GS)\rangle = An^{-\beta}$ with the least square method. The estimated exponent is $\beta = 0.50$, the expected average energy per person is $\langle E_H(GS)\rangle = \frac{0.808}{\sqrt{n}}$, which is consistent with the result in Ref. [7] as we expected.

Table 2
Description of the Kuhn–Munkres algorithm.

Step 1.	The elements in M and W are generated randomly. Compute the corresponding elements of the matrix H_1.
Step 2.	Subtract the smallest elements in each row from all the elements of its row, and subtract the smallest element in each column from all the elements of its column. We denote the new matrix by H_1.
Step 3.	Draw lines through appropriate rows and columns so that all the zero elements of the energy matrix H_1 are covered and the minimum number of such lines is denoted by n_k.
Step 4.	If $n_k = n$, then we will have n independent zeros, the corresponding n positions represent the optimal assignment.
Step 5.	If $n_k < n$, determine the smallest element not covered by any line. Subtract this element from each uncovered row, and then add it to each covered column. Return to Step 3.

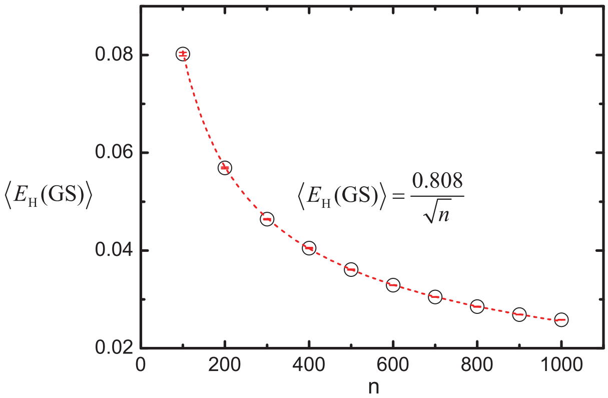

Fig. 1. The relationship between $\langle E_H(GS)\rangle$ and the model size n. The dashed line indicates the fitted curve $\langle E_H(GS)\rangle = \frac{0.808}{\sqrt{n}}$. The results are averaged over 100 independent realizations.

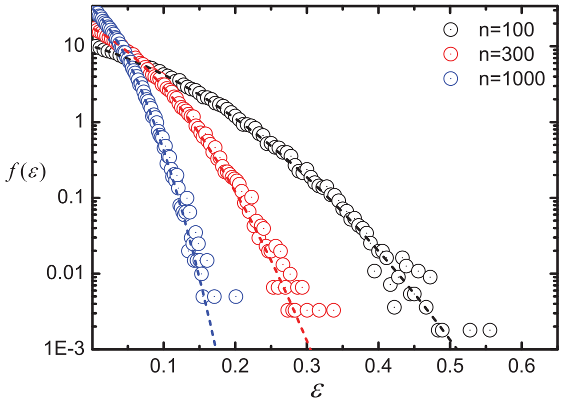

Fig. 2. The PDF of ε with $n = 100$ (black circles), 300 (red circles), 1000 (blue circles). The dashed lines are the corresponding fitting curves by Eq. (2). The results are averaged over 100 independent realizations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2 shows the distribution of individual energy ε. We find that the Probability Density Function (abbreviated as PDF) $f(\varepsilon)$ can be fitted by an exponential function that has a quadratic polynomial in ε as exponent:

$$\frac{f(\varepsilon)}{\sqrt{n}} = C(a,b)\mathrm{e}^{-a(\sqrt{n}\varepsilon)^2 - b(\sqrt{n}\varepsilon)}. \tag{2}$$

The PDF $f(\varepsilon)$ should be normalized to 1,

$$C(a,b) = \left[\frac{1}{2}\sqrt{\frac{\pi}{a}}\mathrm{e}^{\frac{b^2}{4a}} erfc\left(\frac{b}{2\sqrt{a}}\right)\right]^{-1}. \tag{3}$$

Here *erfc* is the complementary error function, it is defined as

$$erfc(z) = \frac{2}{\sqrt{\pi}}\int_z^{\infty} \mathrm{e}^{-t^2}\,\mathrm{d}t. \tag{4}$$

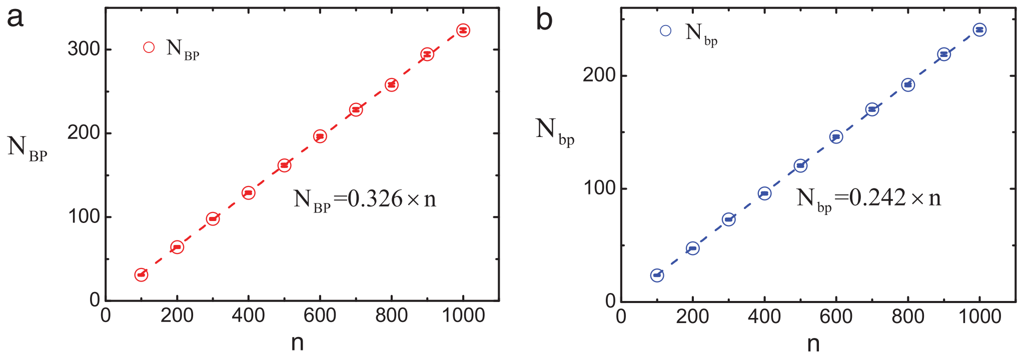

Fig. 3. The result of *BPs* in the numerical simulation. (a) The number of *BPs* N_{BP} (red circles) and (b) the number of men who form *BPs* N_{bp} (blue circles) versus the model size n. The corresponding dashed lines are the linear fits of the data, slop shown in the figure. The results are averaged over 100 independent realizations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We fitted the numerical simulation result with $\frac{f(\varepsilon)}{\sqrt{n}}$ using $\sqrt{n}\varepsilon$ as independent variable, obtaining $a = 0.2287$ and $b = 0.6411$.

4. Blocking pairs

In this section, we start to consider the probability that a man cannot form a *BP*. For example, a man i has energy ε_i in a given matching x, which means that there are $n\varepsilon_i$ women who are better than the spouse of man i. If the man i cannot form any *BP*, all of those $n\varepsilon_i$ women have better spouses than man i, which implies the inequality:

$$W_{\alpha_1,x^{-1}(\alpha_1)} < W_{\alpha_1,i}, W_{\alpha_2,x^{-1}(\alpha_2)} < W_{\alpha_2,i}, \ldots, W_{\alpha_{n\varepsilon},x^{-1}(\alpha_{n\varepsilon})} < W_{\alpha_{n\varepsilon},i}. \tag{5}$$

Note that all the energy terms $M_{i,\alpha}$ and $W_{\alpha,i}$ are uniformly distributed on $[0, 1]$, which implies that the probability that man i with energy ε_i does not form a *BP* with woman α is $P_0(i, \alpha) = (1 - W_{\alpha,x^{-1}(\alpha)})$, thus, for man i with energy ε_i, the probability that he does not form any *BPs* satisfy the following equation:

$$\begin{aligned} P_0(\varepsilon_i) &= \prod_{i=1}^{n\varepsilon_i}(1 - W_{\alpha_i,x^{-1}(\alpha_i)}) \\ &\approx \mathrm{e}^{-\sum_{i=1}^{n\varepsilon_i} W_{\alpha_i,x^{-1}(\alpha_i)}} \\ &\approx \mathrm{e}^{-0.808\sqrt{n}\varepsilon_i}. \end{aligned} \tag{6}$$

Here the first approximately equal sign is used due to $1 - W_{\alpha_i,x^{-1}(\alpha_i)} \ll 1$. Hence, the probability P_0 that a man cannot form a *BP* satisfies:

$$\begin{aligned} P_0 &= \int_0^1 f(\varepsilon)P_0(\varepsilon)\,\mathrm{d}\varepsilon \\ &\approx \int_0^1 f(\varepsilon)\mathrm{e}^{-0.808\sqrt{n}\varepsilon}\,\mathrm{d}\varepsilon \\ &> \int_0^1 f(\varepsilon)(1 - 0.808\sqrt{n}\varepsilon)\,\mathrm{d}\varepsilon \\ &= 1 - 0.808^2 = 0.35. \end{aligned} \tag{7}$$

As aforementioned in Ref. [7], the number of *BPs* is estimated to be $(n\varepsilon)^2 \sim 0.65n$, which happens to be consistent with Eq. (7). Furthermore, if we substitute Eq. (2) into Eq. (7),

$$\begin{aligned} P_0 &\approx \int_0^1 1.0034\mathrm{e}^{-0.2287(\sqrt{n}\varepsilon)^2-0.6411(\sqrt{n}\varepsilon)}\mathrm{e}^{-0.808\sqrt{n}}\,\mathrm{d}\varepsilon \\ &\approx \int_0^\infty 1.0034\mathrm{e}^{-0.2287\varepsilon^2-1.4491\varepsilon}\,\mathrm{d}\varepsilon \\ &= 0.59. \end{aligned} \tag{8}$$

The numerical simulation result in Fig. 3 shows that $P_0 = 0.758$, which shows our estimation is closer to fact than the estimation in Ref. [7]. The difference between our estimation and the numerical result is because the assumption that all the events related to probabilities $P_0(i, \alpha)$ are independent on each other is not valid.

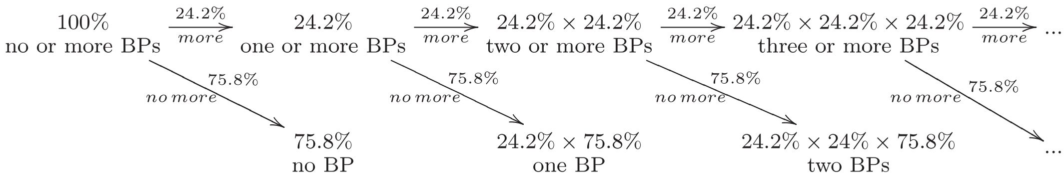

Fig. 4. Illustration of the probability that one person forms *k BPs*.

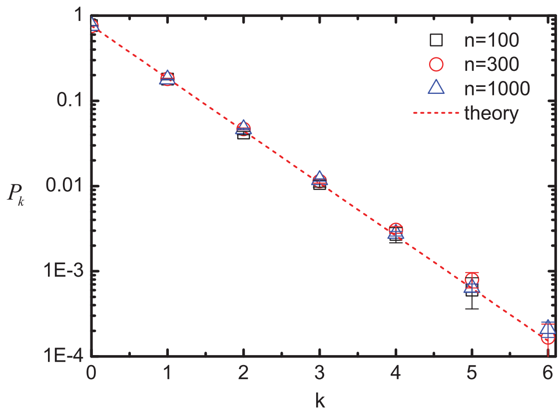

Fig. 5. The probability that a man/woman forms *k BPs* in models consist of 100 (black squares), 300 (red circles), and 1000 (blue triangles) men or women. The red dashed line is the theoretical prediction by Eq. (9). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Next, we analyze the probability that a single person forms *k BPs*. From the numerical simulation, a man/woman has a chance of 24.2% to form a *BP* in average. For those people who have already formed a *BP*, they still approximately have a probability of 24.2% to form another *BP* if we neglect the slight change of the samples space as each of them has already formed a *BP*, otherwise they will finally form only one *BP*. And of course this process can continue to the situation that one forms *k BPs*. The process is illustrated in Fig. 4.

Thus, we conclude that the probability of the single person forms *k BPs* is given:

$$P_k = 0.758 \times (1 - 0.758)^k, \quad k = 0, 1, 2, \dots. \tag{9}$$

The result of Eq. (9) and the numerical simulation result of the probability of a man/woman forms *k BPs* are shown in Fig. 5. The total number of *BPs* is then calculated with Eq. (9):

$$N_{BP} = \sum_{k=1}^{\infty} 0.758 \times 0.242^k \times k \times n = 0.319n, \tag{10}$$

Fig. 3 shows that our result is in good agreement with the outcome of the numerical simulation.

The results and discussions above are based on the perfect information assumption which means all the agents know everybody of the other sex. In the following, we study the situation that a man i who only knows a fraction σ $(0 < \sigma \ll 1)$ of the n women in the model. A man/woman is satisfied only if he/she cannot form a *BP*. Under the assumption that everyone has an equal possibility to form a *BP*, the probability that he/she is satisfied can be estimated as

$$S_\sigma = \left(1 - \frac{0.326}{n}\right)^{n\sigma} \approx 1 - 0.326\sigma. \tag{11}$$

So probability that the ground state is stable is

$$S_\sigma^n \approx (1 - 0.326\sigma)^n \approx \frac{1}{e}^{0.326n\sigma} \approx 0.72^{n\sigma}, \tag{12}$$

where $n\sigma$ is the average number of women/men a man/woman knows.

As we can see from Eq. (12), the ground state stability decays exponentially when increasing σ. That is reasonable because in our model which does not contain constrains like moral standard and religion in real society, the *BPs* will match with their *BP* partner than the assigned one in the ground state if the *BPs* know each other. Obviously, if one knows more people, he/she has a larger possibility to know his/her *BP* partner, so the ground state of the model is more unstable.

5. Conclusion

In this paper we find the ground state in bipartite matching problem by using Kuhn–Munkres algorithm. By combining numerical and analytical method, we compute the number of blocking pairs in the ground state, and also study the probability distribution of a person forms one or more blocking pairs. Furthermore, we discover that the ground state stability decays exponentially with the growth of connectivity in the model. Our approach to analyze blocking pairs can be also extend to study stability of other matching problems in the big data era [14].

Some problems still remain unsolved. For example, one of the problems will be the analytical result of the individual energy distribution in the ground state. When the analytical result is available, it is possible for find the relationship between individual energy distribution and the number of *BPs*. Furthermore, if we compare the different states of the model, for example Gale–Shapley stable solution, optimal stable solution and the ground state, how many people will be harmed or benefit for this change? How much they sacrifice or improve? We believe our result may provide some inspirations to find the solution to the above questions.

Acknowledgments

The authors would like to thank Manuel Sebastian Mariani and Zhuo-Ming Ren for their valuable comments on this work. This work was partially supported by the EU FP7 Grant 611272 (project GROWTHCOM) and by the Swiss National Science Foundation (grant no. 200020-156188). We also thank the support from China Scholarship Council. H.L acknowledges the support from the Science and Technolgy Innovation Commission of Shenzhen (grant no. JCYJ20150625101524056), Guangdong Key Laboratory Projects (grant no. 2012A061400024 and 2014A030313553), China 863 project (grant no. 2015AA015305), NSF China projects (grant no. U1301252, 61170076, 61471243, 2014ZM0079, 61503140), National Education Information and Technology Research Projects (grant no. 136241506, 136221507), and Tencent Open Research Fund.

References

[1] D. Gale, L.S. Shapley, College admissions and the stability of marriage, Amer. Math. Monthly 69 (1) (1962) 9–15.
[2] A.E. Roth, J.H.V. Vate, Random paths to stability in two-sided matching, Econometrica 58 (6) (1990) 1475–1480.
[3] M. Mézard, G. Parisi, Replicas and optimization, J. Phys. Lett. 46 (17) (1985) 771–778.
[4] M. Mézard, G. Parisi, On the solution of the random link matching problems, J. Physique 48 (9) (1987) 1451–1459.
[5] Y.-C. Zhang, Happier world with more information, Physica A 299 (1) (2001) 104–120.
[6] P. Laureti, Y.-C. Zhang, Matching games with partial information, Physica A 324 (1) (2003) 49–65.
[7] M.J. Oméro, M. Dzierzawa, M. Marsili, Y.-C. Zhang, Scaling behavior in the stable marriage problem, J. Physique I 7 (12) (1997) 1723–1732.
[8] M. Dzierzawa, M.J. Oméro, Statistics of stable marriages, Physica A 287 (1) (2000) 321–333.
[9] B. Zhou, S. Qin, X.P. Han, Z. He, J.R. Xie, B.H. Wang, A model of two-way selection system for human behavior, PLoS One 9 (1) (2014) e81424.
[10] B. Zhou, Z. He, L.L. Jiang, N.X. Wang, B.H. Wang, Bidirectional selection between two classes in complex social networks, Sci. Rep. 4 (2014) 7577.
[11] A.E. Roth, The economics of matching: Stability and incentives, Math. Oper. Res. 7 (4) (1982) 617–628.
[12] H.W. Kuhn, The Hungarian method for the assignment problem, Naval Res. Logist. Q. 2 (1–2) (1955) 83–97.
[13] J. Munkres, Algorithms for the assignment and transportation problems, J. Soc. Ind. Appl. Math. 5 (1) (1957) 32–38.
[14] R. Mao, H. Xu, W. Wu, J. Li, Y. Li, M.H. Lu, Overcoming the challenge of variety: Big data abstraction, the next evolution of data management for AAL communication systems, IEEE Commun. Mag. 53 (1) (2015) 42–47.