Controlling and Assessing Correlations of Cost Matrices in Heterogeneous Scheduling

Louis-Claude Canon, Laurent Philippe and Pierre-Cyrille Héam

Université de Franche-Comté / UBFC – DISC / FEMTO-ST / CNRS
New Challenges in Scheduling Theory

April 2, 2016

Context

Study focused on $R||C_{max}$[1].

A common simple problem

- Hundreds of studies consider it or variations of it.
- An optimization problem with a concise formulation.

Yet, it presents challenges

- NP-Hard: most solutions are heuristics.
- Non-trivial random generation of instances.

[1] Using the $\alpha|\beta|\gamma$ notation

Context

Study focused on $R||C_{max}$[1].

A common simple problem

- Hundreds of studies consider it or variations of it.
- An optimization problem with a concise formulation.

Yet, it presents challenges

- NP-Hard: most solutions are heuristics.
- Non-trivial random generation of instances.

[1] Using the $\alpha|\beta|\gamma$ notation

Context

Study focused on $R||C_{max}$[1].

A common simple problem

- Hundreds of studies consider it or variations of it.
- An optimization problem with a concise formulation.

Yet, it presents challenges

- NP-Hard: most solutions are heuristics.
- Non-trivial random generation of instances.

[1] Using the $\alpha|\beta|\gamma$ notation

Approach

Focus on the evaluation of scheduling heuristics (no design of new heuristics).

Problem

- Several existing heuristics: MET, MinMin, MaxMin, ...
- What are the relative performance of these heuristics?

Solution

- Design of a test bed consisting of a set of instances.
- Instances are obtained by generating random cost matrices.

Approach

Focus on the evaluation of scheduling heuristics (no design of new heuristics).

Problem

- Several existing heuristics: MET, MinMin, MaxMin, ...
- What are the relative performance of these heuristics?

Solution

- Design of a test bed consisting of a set of instances.
- Instances are obtained by generating random cost matrices.

Approach

Focus on the evaluation of scheduling heuristics (no design of new heuristics).

Problem

- Several existing heuristics: MET, MinMin, MaxMin, . . .
- What are the relative performance of these heuristics?

Solution

- Design of a test bed consisting of a set of instances.
- Instances are obtained by generating random cost matrices.

Reproducibility and Open Science

- Tools for consolidating and improving the validity of results.
- Improve research methodology with reproducible practices.
- Rely on the openness to increase studies validity.

But there are other tools.

Reproducibility and Open Science

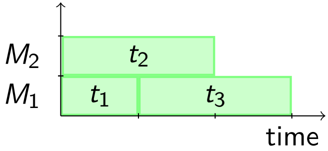
- Tools for consolidating and improving the validity of results.
- Improve research methodology with reproducible practices.
- Rely on the openness to increase studies validity.

But there are other tools.

Scheduling Problem: $R||C_{max}$

- Independent tasks all available simultaneously without preemption.
- Unrelated platform: each task takes a specific duration on any machine.
- Objective: minimizing the maximum completion time.

$$\begin{pmatrix} e_{1,1} = 1 & e_{1,2} = 10 \\ e_{2,1} = 3 & e_{2,2} = 2 \\ e_{3,1} = 2 & e_{3,2} = 1.5 \end{pmatrix}$$



Range-based method [Ali et al., 2000]

Require: n, m, R_{task}, R_{mach}
Ensure: a $n \times m$ cost matrix e

for all $1 \leq i \leq n$ **do**
  $\tau[i] \leftarrow U(1, R_{task})$
  for all $1 \leq j \leq m$ **do**
    $e[i,j] \leftarrow \tau[i] \times U(1, R_{mach})$
return e

Range-based method [Ali et al., 2000]

$$e = \begin{pmatrix} \tau[1]U(1, R_{mach}) & \cdots & \tau[1]U(1, R_{mach}) \\ \vdots & \ddots & \vdots \\ \tau[n]U(1, R_{mach}) & \cdots & \tau[n]U(1, R_{mach}) \end{pmatrix}$$

Summary

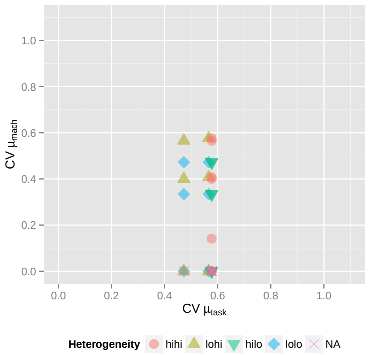
- one vector of n values that follow $U(1, R_{task})$
- n vectors of m values that follow $U(1, R_{mach})$

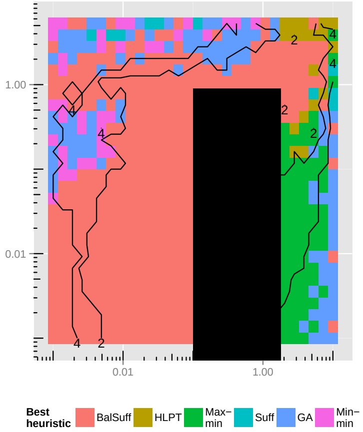
Usage in the literature

- Cited more than 300 times (Google Scholar).
- We covered 160 relevant references (papers using the method, written in English, freely available).
- We extracted the parameters of each of the $37 + 173 = 210$ distinct instances.

Articles reproducibility	Number
Most parameters are given	78
Other instances are referenced	40
No clear description	42

Heterogeneity bias[2]





[2] L.-C. Canon and L. Philippe, **On the Heterogeneity Bias of Cost Matrices when Assessing Scheduling Algorithms**, Euro-Par 2015.

LPT: a 4/3-Approximation for $Q||C_{max}$

Instance for $Q||C_{max}$

- A vector of task weights (w_i for task i).
- A vector of processor cycle times (b_j for machine j).
- A $R||C_{max}$ instance in which $e_{ij} = w_i b_j$.

HLPT: an adaptation of LPT for $R||C_{max}$

- Should work well when $e_{ij} = w_i b_j$.
- How does it work when e_{ij} is arbitrary?

$$E_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} \quad E_2 = \begin{pmatrix} 1 & 6 & 9 \\ 2 & 2 & 3 \\ 6 & 3 & 4 \end{pmatrix}$$

LPT: a 4/3-Approximation for $Q||C_{max}$

Instance for $Q||C_{max}$

- A vector of task weights (w_i for task i).
- A vector of processor cycle times (b_j for machine j).
- A $R||C_{max}$ instance in which $e_{ij} = w_i b_j$.

HLPT: an adaptation of LPT for $R||C_{max}$

- Should work well when $e_{ij} = w_i b_j$.
- How does it work when e_{ij} is arbitrary?

$$E_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} \quad E_2 = \begin{pmatrix} 1 & 6 & 9 \\ 2 & 2 & 3 \\ 6 & 3 & 4 \end{pmatrix}$$

Proximity Measures

- Objective: quantifying the similarity between a $R||C_{max}$ instance and the closest $Q||C_{max}$ instance.
- Task correlation ($\rho^r_{i,i'}$ is the correlation between row i and row i'):

$$\rho_{\text{task}} \triangleq \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i'=1, i' \neq i}^{n} \rho^r_{i,i'} \tag{1}$$

- Machine correlation ($\rho^c_{j,j'}$ is the correlation between column j and column j'):

$$\rho_{\text{mach}} \triangleq \frac{1}{m(m-1)} \sum_{j=1}^{m} \sum_{j'=1, j' \neq j}^{m} \rho^c_{j,j'} \tag{2}$$

Computing the Correlation

Compute each linear coefficient of correlation $\rho^r_{i,i'}$:

$$\rho^r_{i,i'} \quad \begin{pmatrix} \vdots & \ddots & \vdots \\ e_{i,1} & \cdots & e_{i,m} \\ \vdots & \ddots & \vdots \\ e_{i',1} & \cdots & e_{i',m} \\ \vdots & \ddots & \vdots \end{pmatrix}$$

The correlation quantifies the strength of the linear relation between two vectors (close to 1 if $e_{i,j} = \alpha e_{i',j} + \beta$ for $1 \leq j \leq m$).

Computing the Correlation

Compute each linear coefficient of correlation $\rho^r_{i,i'}$:

$$\rho^r_{i,i'} \quad \begin{pmatrix} \vdots & \ddots & \vdots \\ \boxed{e_{i,1} \quad \cdots \quad e_{i,m}} \\ \vdots & \ddots & \vdots \\ \boxed{e_{i',1} \quad \cdots \quad e_{i',m}} \\ \vdots & \ddots & \vdots \end{pmatrix}$$

The correlation quantifies the strength of the linear relation between two vectors (close to 1 if $e_{i,j} = \alpha e_{i',j} + \beta$ for $1 \leq j \leq m$).

Computing the Correlation

Compute each linear coefficient of correlation $\rho^r_{i,i'}$:

$$\rho^r_{i,i'} \quad \begin{pmatrix} \vdots & \ddots & \vdots \\ \boxed{e_{i,1} \quad \cdots \quad e_{i,m}} \\ \vdots & \ddots & \vdots \\ \boxed{e_{i',1} \quad \cdots \quad e_{i',m}} \\ \vdots & \ddots & \vdots \end{pmatrix}$$

The correlation quantifies the strength of the linear relation between two vectors (close to 1 if $e_{i,j} = \alpha e_{i',j} + \beta$ for $1 \leq j \leq m$).

Computing the Correlation

Compute each linear coefficient of correlation $\rho^r_{i,i'}$:

$$\rho^r_{i,i'} \quad \begin{pmatrix} \vdots & \ddots & \vdots \\ \boxed{e_{i,1} \quad \cdots \quad e_{i,m}} \\ \vdots & \ddots & \vdots \\ \boxed{e_{i',1} \quad \cdots \quad e_{i',m}} \\ \vdots & \ddots & \vdots \end{pmatrix}$$

The correlation quantifies the strength of the linear relation between two vectors (close to 1 if $e_{i,j} = \alpha e_{i',j} + \beta$ for $1 \leq j \leq m$).

Examples

$$E_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 10 \end{pmatrix} \quad E_2 = \begin{pmatrix} 1 & 6 & 10 \\ 2 & 2 & 3 \\ 6 & 3 & 4 \end{pmatrix}$$

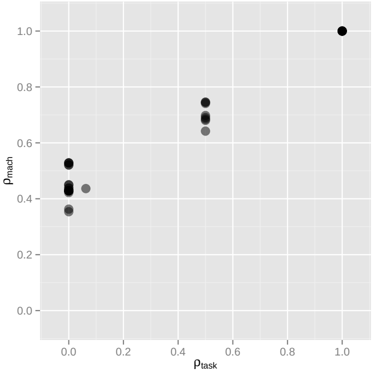
- E_1: correlations close to one.
- E_2: correlations close to zero.

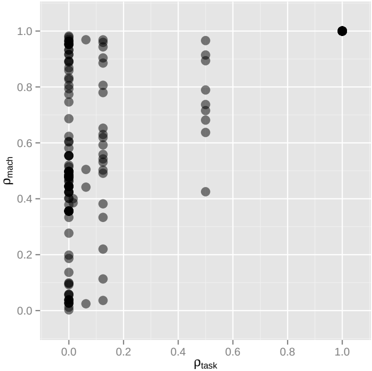
Examples

$$E_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 10 \end{pmatrix} \quad E_2 = \begin{pmatrix} 1 & 6 & 10 \\ 2 & 2 & 3 \\ 6 & 3 & 4 \end{pmatrix}$$

- E_1: correlations close to one.
- E_2: correlations close to zero.

Correlation Properties used in the Literature





Controlling the Correlation

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \begin{pmatrix} U(0,1) & \cdots & U(0,1) \\ \vdots & \ddots & \vdots \\ U(0,1) & \cdots & U(0,1) \end{pmatrix}$$

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \alpha \begin{pmatrix} c_1 & \cdots & c_1 \\ \vdots & \ddots & \vdots \\ c_n & \cdots & c_n \end{pmatrix} + \beta \begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix}$$

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \gamma \begin{pmatrix} r_1 & \cdots & r_m \\ \vdots & \ddots & \vdots \\ r_1 & \cdots & r_m \end{pmatrix} + \delta \begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix}$$

Controlling the Correlation

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \begin{pmatrix} U(0,1) & \cdots & U(0,1) \\ \vdots & \ddots & \vdots \\ U(0,1) & \cdots & U(0,1) \end{pmatrix}$$

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \alpha \begin{pmatrix} c_1 & \cdots & c_1 \\ \vdots & \ddots & \vdots \\ c_n & \cdots & c_n \end{pmatrix} + \beta \begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix}$$

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \gamma \begin{pmatrix} r_1 & \cdots & r_m \\ \vdots & \ddots & \vdots \\ r_1 & \cdots & r_m \end{pmatrix} + \delta \begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix}$$

Controlling the Correlation

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \begin{pmatrix} U(0,1) & \cdots & U(0,1) \\ \vdots & \ddots & \vdots \\ U(0,1) & \cdots & U(0,1) \end{pmatrix}$$

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \alpha \begin{pmatrix} c_1 & \cdots & c_1 \\ \vdots & \ddots & \vdots \\ c_n & \cdots & c_n \end{pmatrix} + \beta \begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix}$$

$$\begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix} \leftarrow \gamma \begin{pmatrix} r_1 & \cdots & r_m \\ \vdots & \ddots & \vdots \\ r_1 & \cdots & r_m \end{pmatrix} + \delta \begin{pmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,m} \end{pmatrix}$$

Formal Derivations

Let r_{task} and r_{mach} be the desired correlations between the rows and columns.

$$\alpha = \sqrt{r_{\text{mach}}}$$
$$\beta = \sqrt{1 - r_{\text{mach}}}$$
$$\gamma = \sqrt{r_{\text{task}}}$$
$$\delta = \sqrt{1 - r_{\text{task}}}$$

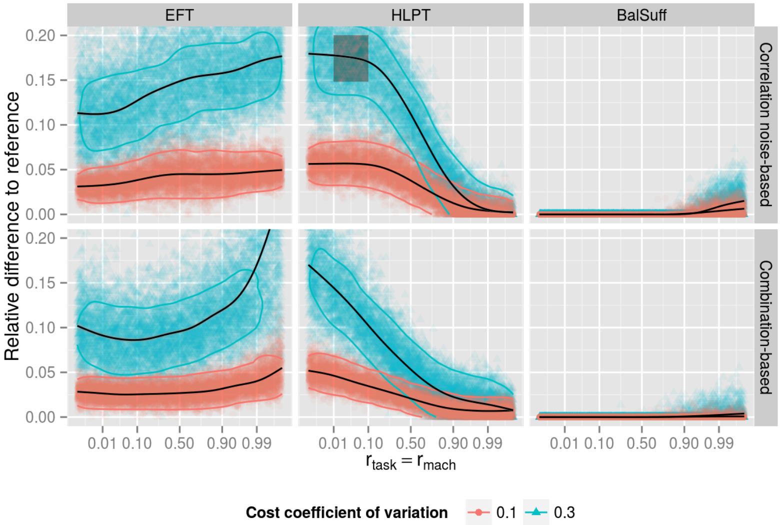
Coefficient of variation of c_i:

$$\frac{\sqrt{r_{\text{task}}} + \sqrt{1 - r_{\text{task}}}\left(\sqrt{r_{\text{mach}}} + \sqrt{1 - r_{\text{mach}}}\right)}{\sqrt{r_{\text{task}}}\sqrt{1 - r_{\text{mach}}} + \sqrt{1 - r_{\text{task}}}\left(\sqrt{r_{\text{mach}}} + \sqrt{1 - r_{\text{mach}}}\right)}$$

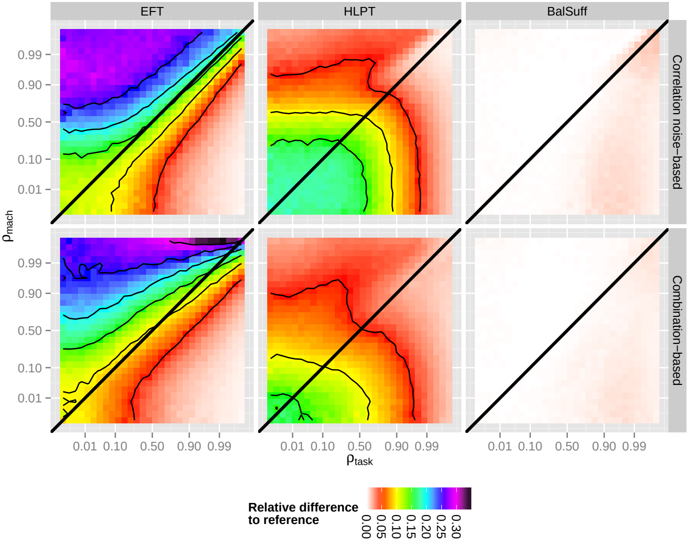
Coefficient of variation of r_i:

$$\sqrt{1 - r_{\text{mach}}}\frac{\sqrt{r_{\text{task}}} + \sqrt{1 - r_{\text{task}}}\left(\sqrt{r_{\text{mach}}} + \sqrt{1 - r_{\text{mach}}}\right)}{\sqrt{r_{\text{task}}}\sqrt{1 - r_{\text{mach}}} + \sqrt{1 - r_{\text{task}}}\left(\sqrt{r_{\text{mach}}} + \sqrt{1 - r_{\text{mach}}}\right)}$$

Impact of the Correlation



Average Impact of the Correlation



TMA (Task-Machine Affinity) [Al-Qawasmeh et al., 2011]

Require: a $n \times m$ cost matrix
Ensure: the TMA of this matrix

$\{E_{i,j}\}_{1\leq i\leq n, 1\leq j\leq m} \leftarrow \left\{\frac{1}{e_{i,j}}\right\}_{1\leq i\leq n, 1\leq j\leq m}$

repeat

  $\{E_{i,j}\}_{1\leq i\leq n, 1\leq j\leq m} \leftarrow \left\{\frac{\sqrt{n}E_{i,j}}{\sqrt{m}\sum_{i'=1}^{n} E_{i',j}}\right\}_{1\leq i\leq n, 1\leq j\leq m}$

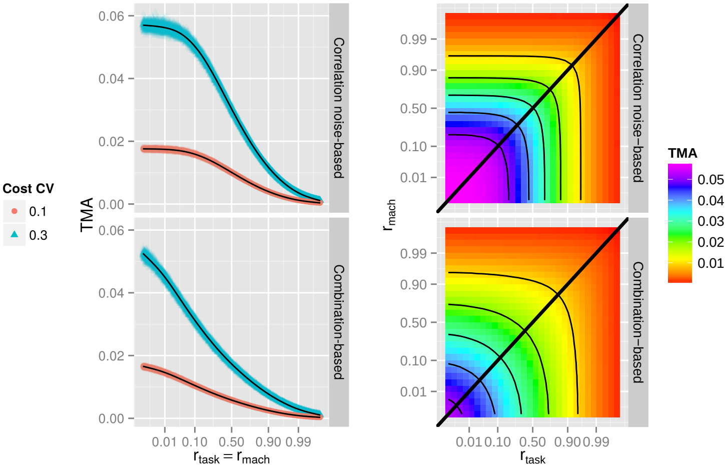
  $\{E_{i,j}\}_{1\leq i\leq n, 1\leq j\leq m} \leftarrow \left\{\frac{\sqrt{n}E_{i,j}}{\sqrt{m}\sum_{j'=1}^{m} E_{i,j'}}\right\}_{1\leq i\leq n, 1\leq j\leq m}$

until the normalization procedure has converged

compute the singular values $\{\sigma_i\}_{1\leq i\leq \min(n,m)}$

return $\frac{1}{\min(n,m)-1}\sum_{i=2}^{\min(n,m)} \sigma_i$

Average Impact of the TMA



Resources

Articles and notes

- Companion research report: http://lifc.univ-fcomte.fr/~publis/papers/pub/2016/RR-FEMTO-ST-1191-2016.pdf.
- Research notes: http://lifc.univ-fcomte.fr/~lccanon/site/blog.html.

Data and code

- Raw data: on-demand (around 500 GB).
- Processed data (21 MB) with the source code of research notes (knitr with Markdown): https://dx.doi.org/10.6084/m9.figshare.2858638.v2

Resources

Articles and notes

- Companion research report: http://lifc.univ-fcomte.fr/~publis/papers/pub/2016/RR-FEMTO-ST-1191-2016.pdf.
- Research notes: http://lifc.univ-fcomte.fr/~lccanon/site/blog.html.

Data and code

- Raw data: on-demand (around 500 GB).
- Processed data (21 MB) with the source code of research notes (knitr with Markdown): https://dx.doi.org/10.6084/m9.figshare.2858638.v2

Conclusion

Presented contributions

- Generate cost matrices for the $R||C_{\max}$ problem.
- Propose a measure for quantifying the proximity with $Q||C_{\max}$.
- Design a method for controlling the correlation.
- Assess the impact of the correlation on the performance of several heuristics.

Perspectives

- Design a generation method fixing both the heterogeneity and the correlation.
- Considering other fields requiring matrix random generation.
- Conduct a full analysis of the performance of existing heuristics based on the heterogeneity, correlation and other measures.

Conclusion

Presented contributions

- Generate cost matrices for the $R||C_{\max}$ problem.
- Propose a measure for quantifying the proximity with $Q||C_{\max}$.
- Design a method for controlling the correlation.
- Assess the impact of the correlation on the performance of several heuristics.

Perspectives

- Design a generation method fixing both the heterogeneity and the correlation.
- Considering other fields requiring matrix random generation.
- Conduct a full analysis of the performance of existing heuristics based on the heterogeneity, correlation and other measures.