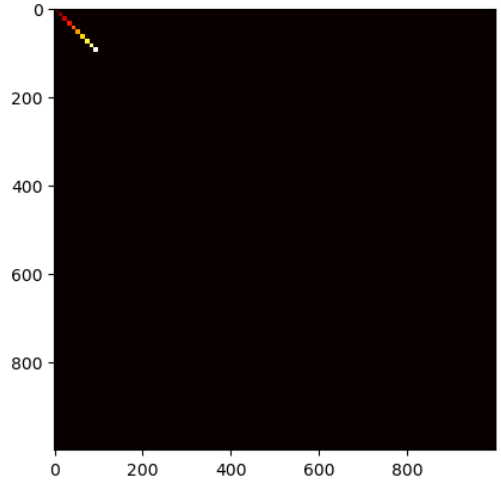
Weekly Report

WU Zihan

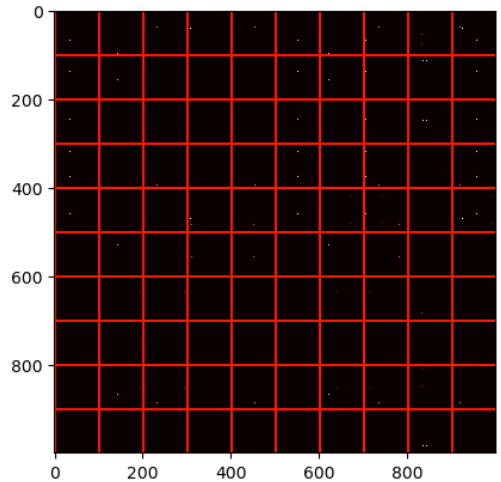
September 14, 2023

Probability Model

- Target: Reconstruct biclusters in big matrix;
- Method: Partition the matrix into small blocks and reconstruct biclusters in each block larger than $T_m \times T_n$;
- Probability Model: We construct the probability model to decide T_p re-partition times to ensure the probability of missing biclusters is less than ϵ.



(a) Original matrix A



(b) Partition matrix A into blocks

Results for simulated data matrix

Experiment settings

- $A \in \mathbb{R}^{M \times N}$,
  $M = N = 1000$
- Biclusters number $K = 10$
- Bicluster height
  $\phi_1 = \phi_2 = \cdots = \phi_K = \phi$
- Bicluster width
  $\psi_1 = \psi_2 = \cdots = \psi_K = \psi$
- Threshold $T_m = T_n = 4$

Experiment results

$$T_p \geq \frac{\ln(1 - p_0)}{-2[m\phi(s^{(k)})^2 + n\psi(t^{(k)})^2]}$$

- According to computation, $T_p \geq 303.024$ to ensure the probability of missing biclusters is less than 1%.
- We set $T_p = 304$ and do 1000 experiments, in which 992 successful detections are observed.

Results

T_p numbers for various bicluster sizes

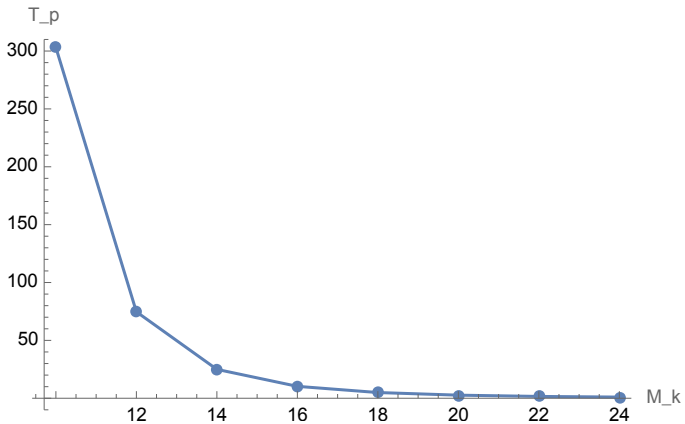


Figure: T_p numbers for various bicluster sizes

Rate of successful detections

Each M_k is tested 100 times.

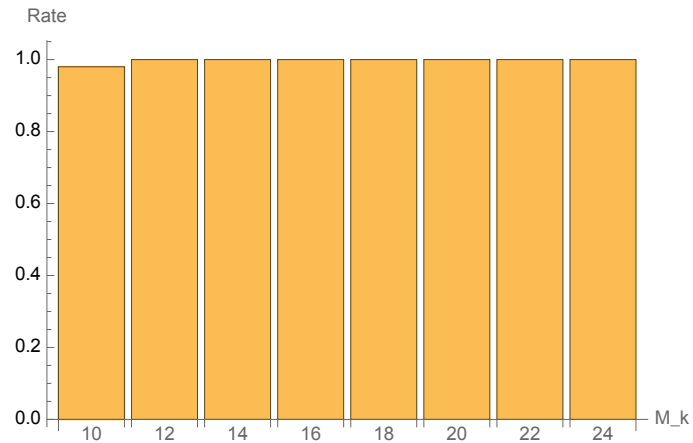


Figure: Rate of successful detections