Learning from Crowds in the Presence of Schools of Thought

Yuandong Tian[1] and Jun Zhu[2]

[1]Carnegie Mellon University

[2]Tsinghua University

Crowd-sourcing

	Worker 1	Worker 2	Worker 3	Worker 4
Task 1	X	X		X
Task 2		X		X
Task 3	X		X	X

Crowd-sourcing

Objective Tasks

E.g. *Labeling dataset*
Knowledge Test

Subjective Tasks

E.g. *Demographical Survey*
Personal Opinions
Creative thoughts
Ill-designed ambiguous tasks.

Crowd-sourcing

Objective Tasks

Subjective Tasks

Noise

Crowd-sourcing

Objective Tasks

Subjective Tasks

Noise

Worker reliability

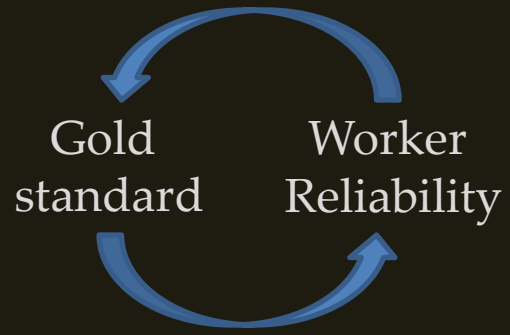
Task clarity

Previous works

Objective Tasks

Subjective Tasks

Majority Voting
[J. Whitehill et al., NIPS'09]
[V.C. Raykar et al., JMLR'10]
[P. Welinder et al., NIPS'10]
…..



Our Contribution

Objective Tasks | Subjective Tasks

Contributions:

1. Applicable to both objective and subjective tasks.
2. Simple , no iterative procedure, no initial guess.

Two Principles

A worker is reliable

if he agrees with other workers in many tasks.

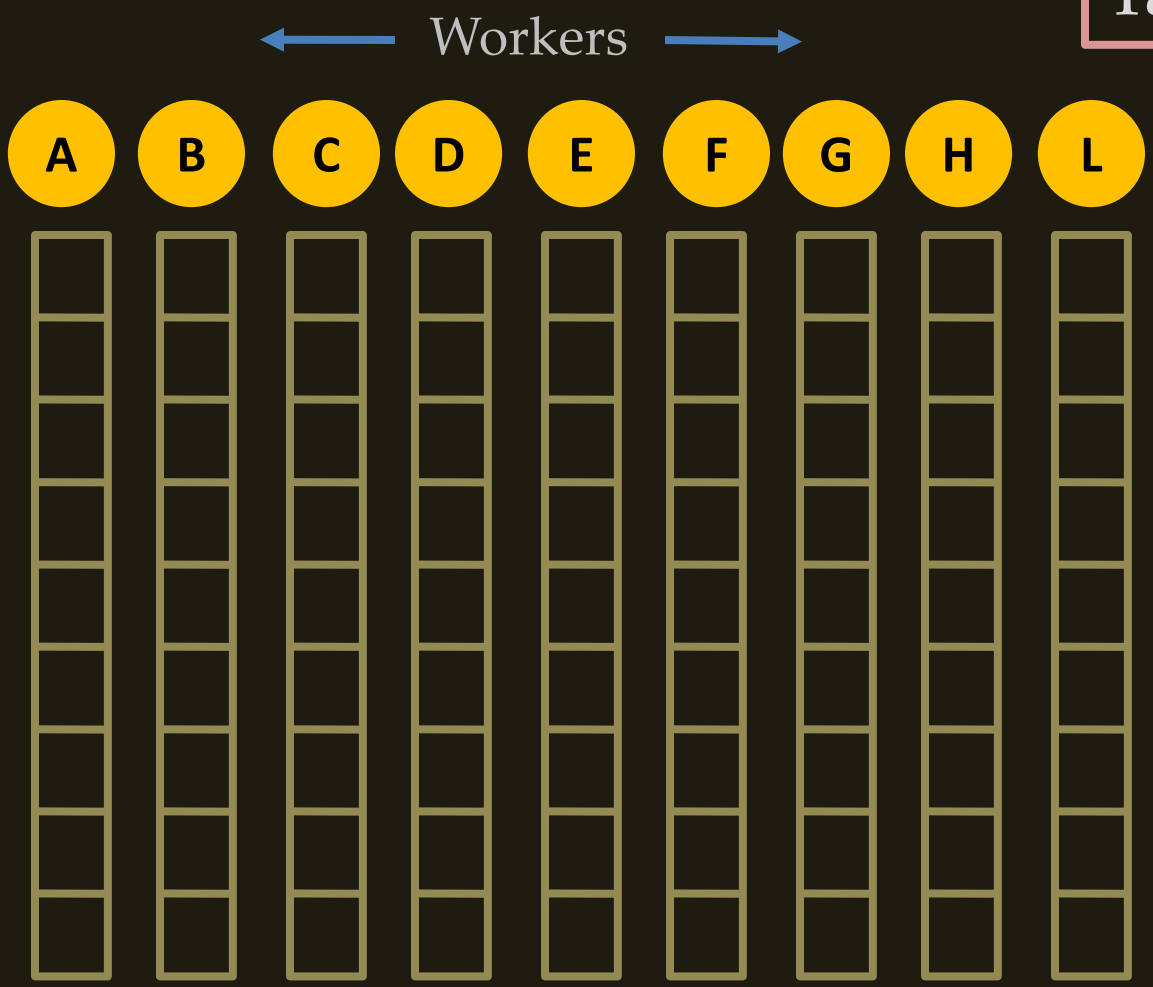
A task is clear

if it has only a few answers.

Clustering Analysis

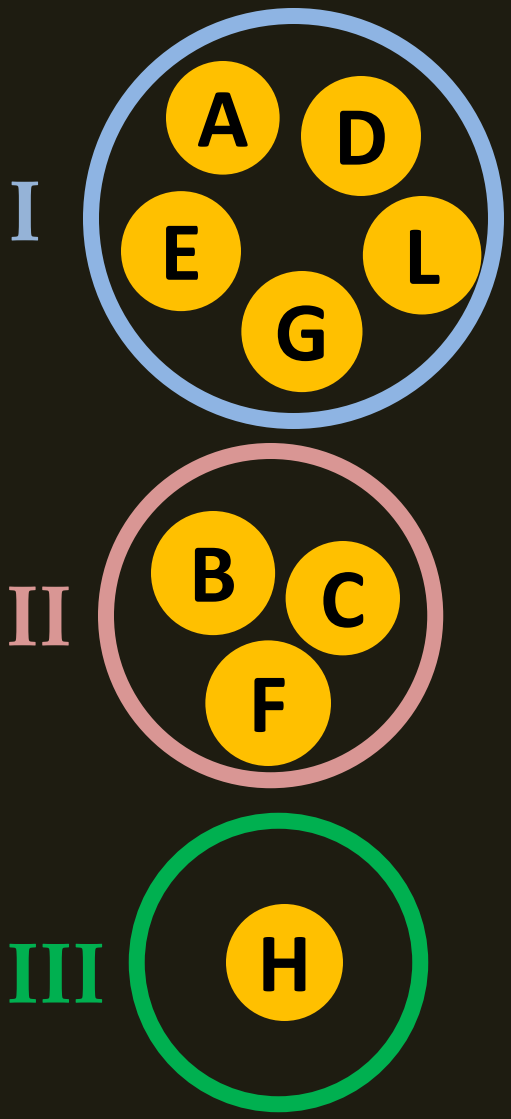
Task k

$\mathbf{x}_{ik}$



Group-size Matrix #Z

Task k



Worker	Assign.	Cluster size
A	I	5
B	II	3
C	II	3
D	I	5
E	I	5
F	II	3
G	I	5
H	III	1
L	I	5

Group-size Matrix #Z

#Z

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
Worker A	5	3	2	3	4	2	6
Worker B	3	3	4	5	4	3	6
Worker C	3	2	2	5	2	4	6
Worker D	5	3	4	5	4	4	6
Worker E	5	2	2	5	2	3	2
Worker F	3	2	2	5	2	4	2
Worker G	5	2	4	3	1	3	6
Worker H	1	1	1	1	2	2	1
Worker L	5	1	4	3	4	4	6

Worker Reliability

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
Worker A	5	3	2	3	4	2	6
Worker B	3	3	4	5	4	3	6
Worker C	3	2	2	5	2	4	6
Worker D	**5**	**3**	**4**	**5**	**4**	**4**	**6**
Worker E	5	2	2	5	2	3	2
Worker F	3	2	2	5	2	4	2
Worker G	5	2	4	3	1	3	6
Worker H	**1**	**1**	**1**	**1**	**2**	**2**	**1**
Worker L	5	1	4	3	4	4	6

Task Clarity

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
Worker A	5	3	2	3	4	2	6
Worker B	3	3	4	5	4	3	6
Worker C	3	2	2	5	2	4	6
Worker D	5	3	4	5	4	4	6
Worker E	5	2	2	5	2	3	2
Worker F	3	2	2	5	2	4	2
Worker G	5	2	4	3	1	3	6
Worker H	1	1	1	1	2	2	1
Worker L	5	1	4	3	4	4	6

Factorization

$\#Z$

	T 1	T 2	T 3	T4	T5	T6	T 7
WA	5	3	2	3	4	2	6
WB	3	3	4	5	4	3	6
WC	3	2	2	5	2	4	6
WD	5	3	4	5	4	4	6
WE	5	2	2	5	2	3	2
WF	3	2	2	5	2	4	2
WG	5	2	4	3	1	3	6
WH	1	1	1	1	2	2	1
WL	5	1	4	3	4	4	6

$=$

λ

Worker Reliability

μ^T

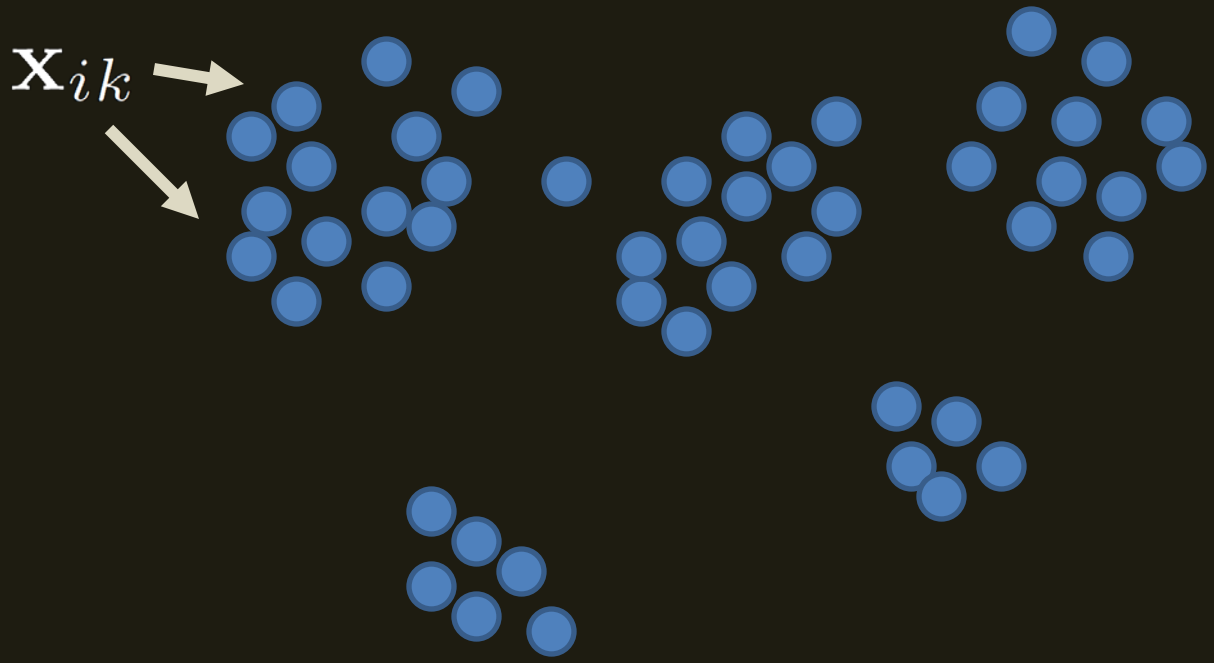
Task clarity

Perron-Frobenius theorem:

$\#Z > 0 \rightarrow \lambda > 0$ and $\mu > 0$

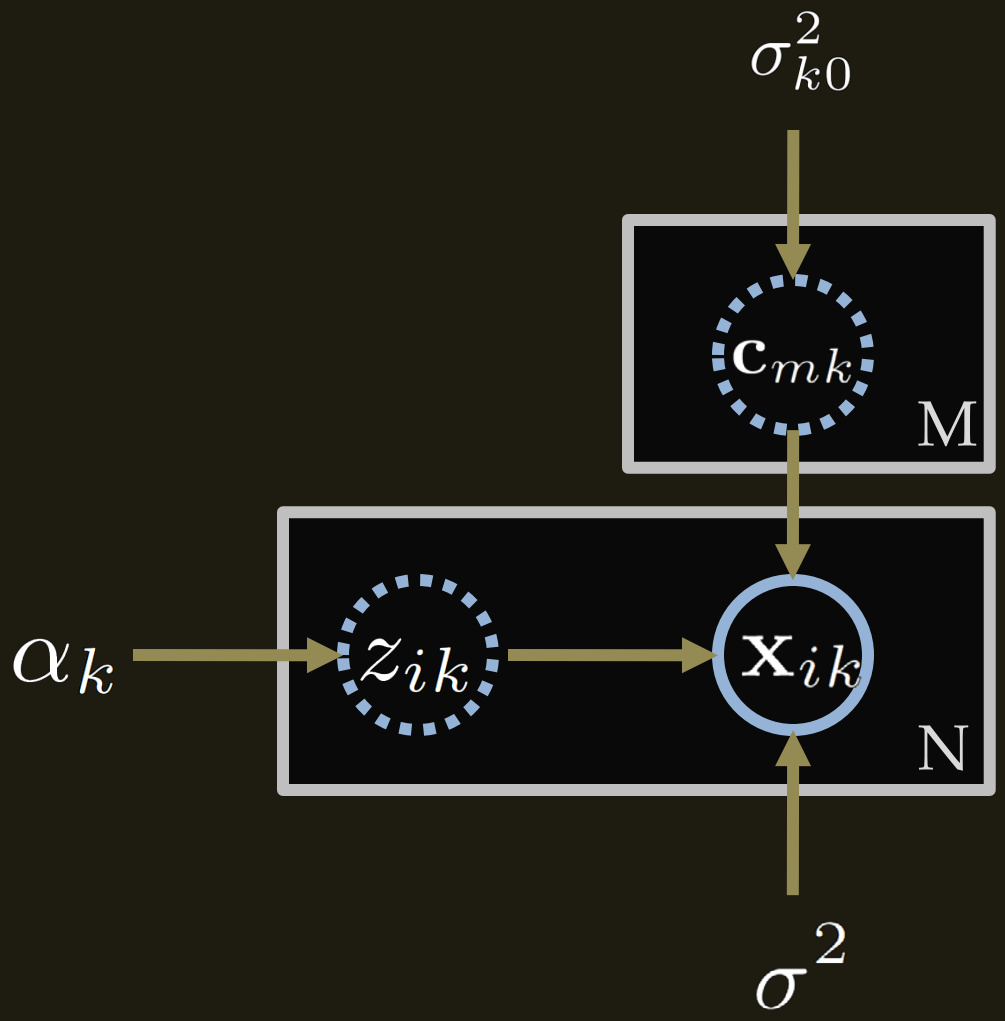
Clustering Model

Task k



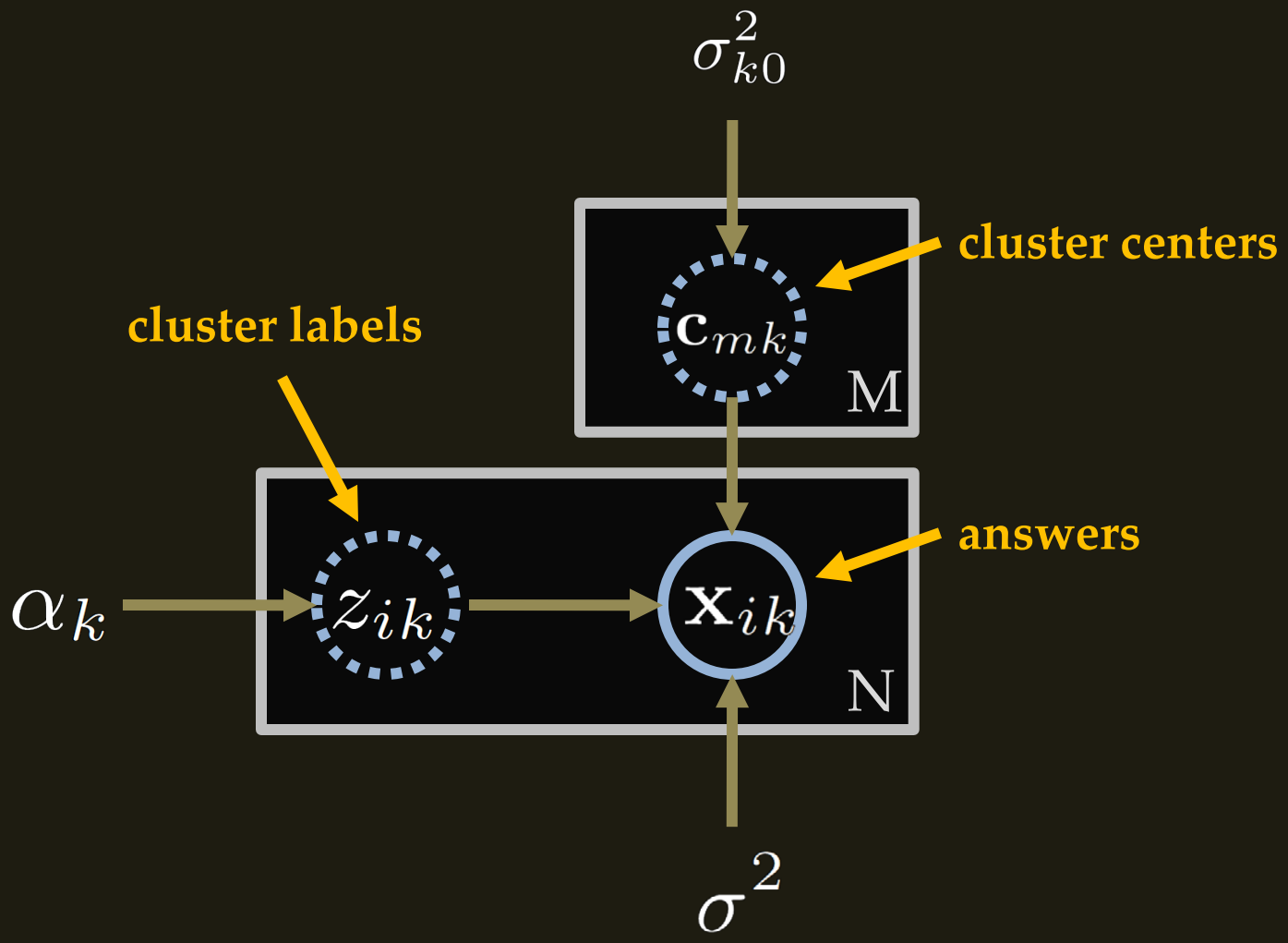
Clustering Model

Task k

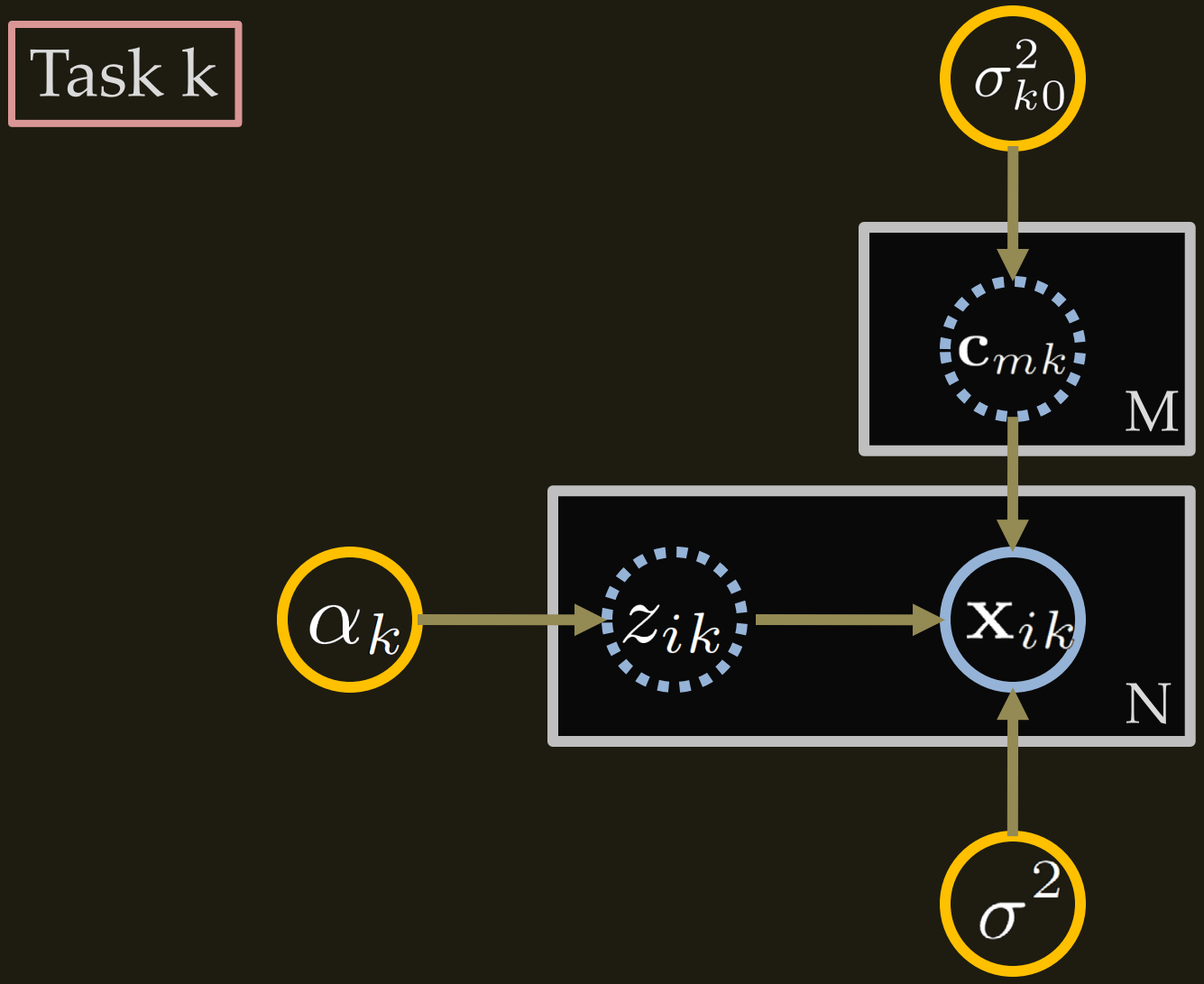


Clustering Model

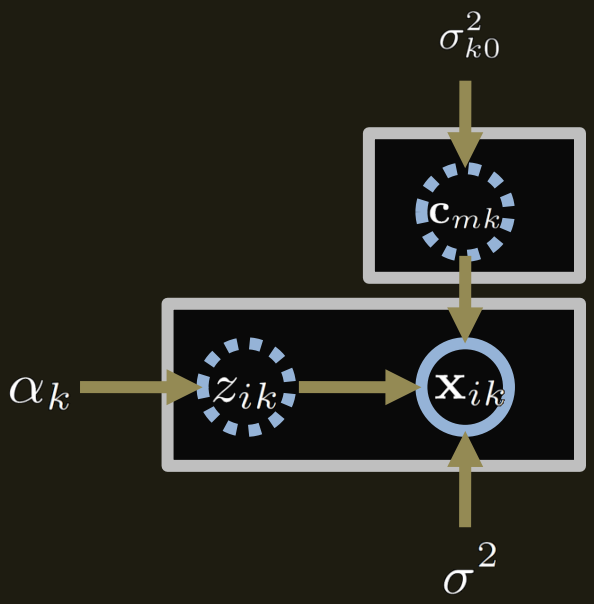
Task k

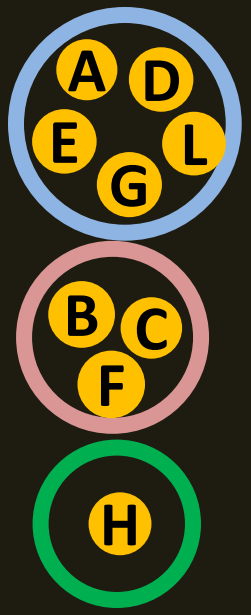


Clustering Model



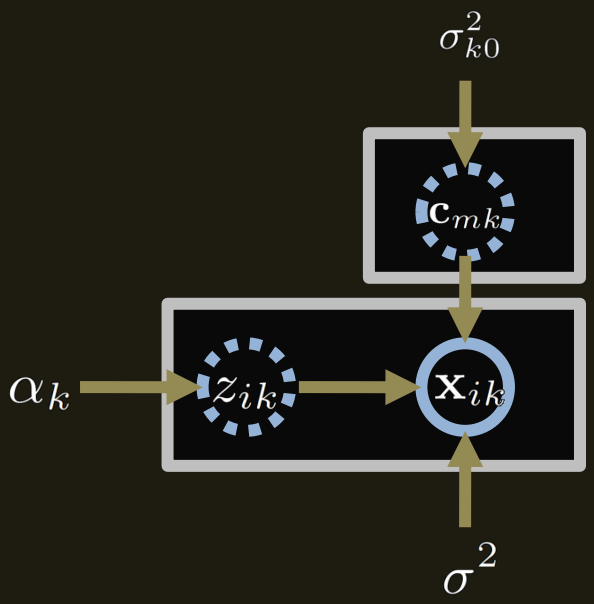
Clustering Model



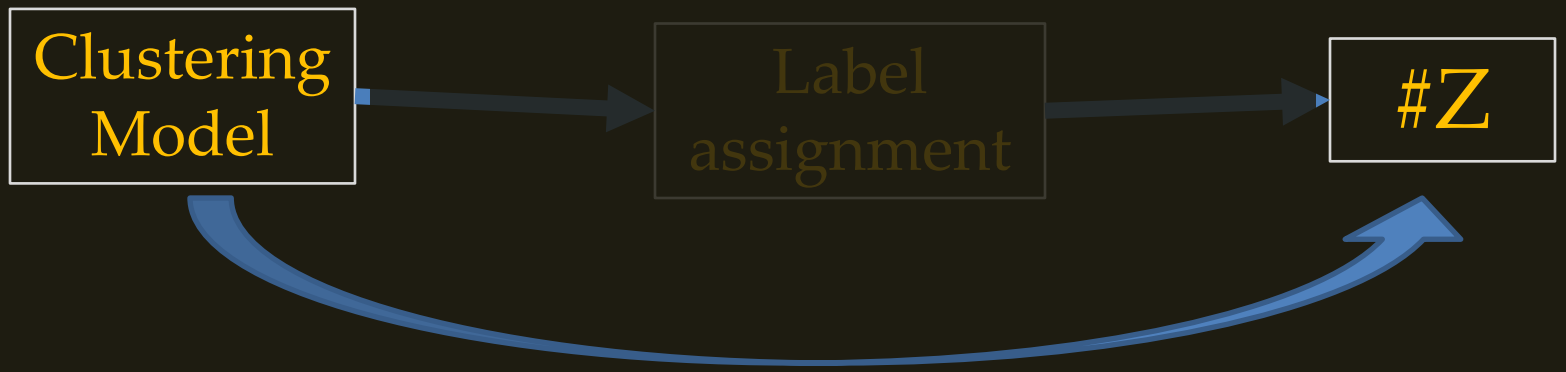


	T 1	T 2	T 3	T4	T5	T6	T 7
W1	5	3	2	3	4	2	6
W2	3	3	4	5	4	3	6
W3	3	2	2	5	2	4	6
W4	5	3	4	5	4	4	6
W5	5	2	2	5	2	3	2
W6	3	2	2	5	2	4	2
W7	5	2	4	3	1	3	6
W8	1	1	1	1	2	2	1
W9	5	1	4	3	4	4	6

Clustering Model



	T 1	T 2	T 3	T4	T5	T6	T 7
W1	5	3	2	3	4	2	6
W2	3	3	4	5	4	3	6
W3	3	2	2	5	2	4	6
W4	5	3	4	5	4	4	6
W5	5	2	2	5	2	3	2
W6	3	2	2	5	2	4	2
W7	5	2	4	3	1	3	6
W8	1	1	1	1	2	2	1
W9	5	1	4	3	4	4	6



Close form solution to #Z

$$\#\tilde{z}_{ik} \approx \sum_{j=1}^{N} \frac{1}{1+\exp\left(\beta_k D_{ij}^k + \beta_{k0}\right)}$$

Close form solution to #Z

$$\#\tilde{z}_{ik} \approx \sum_{j=1}^{N} \frac{1}{1+\exp\left(\beta_k D_{ij}^k + \beta_{k0}\right)}$$

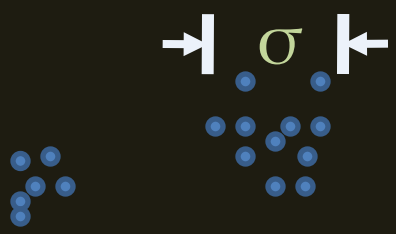
Squared Euclidean Distance
between worker i and worker j in task k

Hyper-Parameters Estimation

Hyper-parameters: α_k, σ_{k0} and σ

$$\mathbb{E}\left[D_{ij}^k\right] = 2d\left(\sigma^2 + \sigma_{k0}^2 \frac{\alpha_k}{1+\alpha_k}\right)$$

$$\mathbb{E}\left[\|\mathbf{x}_{ik}\|^2\right] = d(\sigma^2 + \sigma_{k0}^2).$$



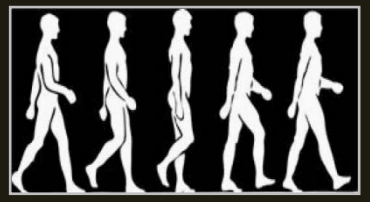
$\sigma = 0.2$

Experiments Setting

Mission I: Image Classification (Sky/Building/Computer)


Do these images contain sky?

Mission II: Counting Objects

Mission III: Images Aesthetics


Do these images look pretty?

Statistics

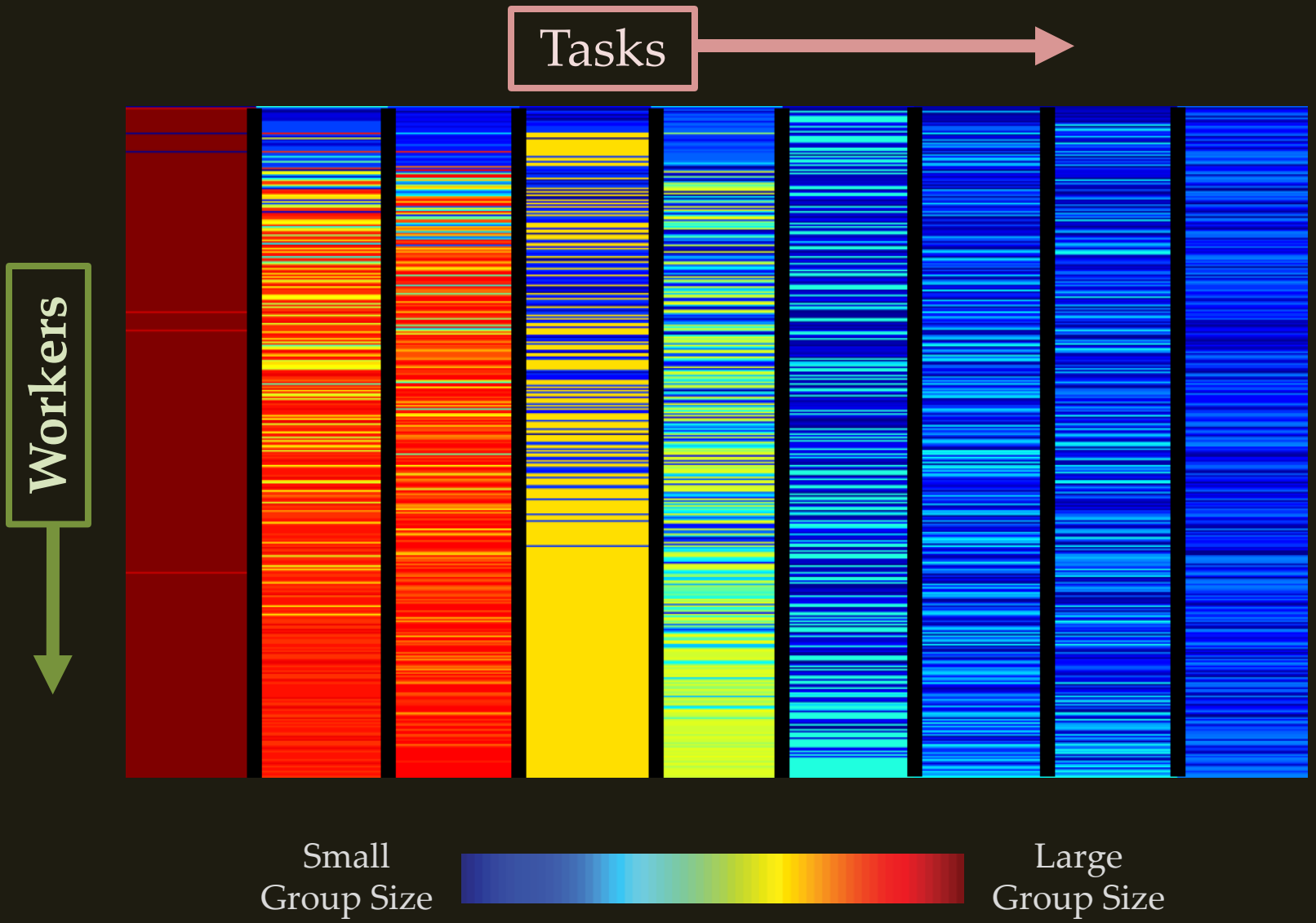
Mission I			Mission II	Mission III
Sky (12)	Building (12)	Computer (12)	Counting (4)	Images Aesthetics (12 + 12)

402 workers

Dataset link:

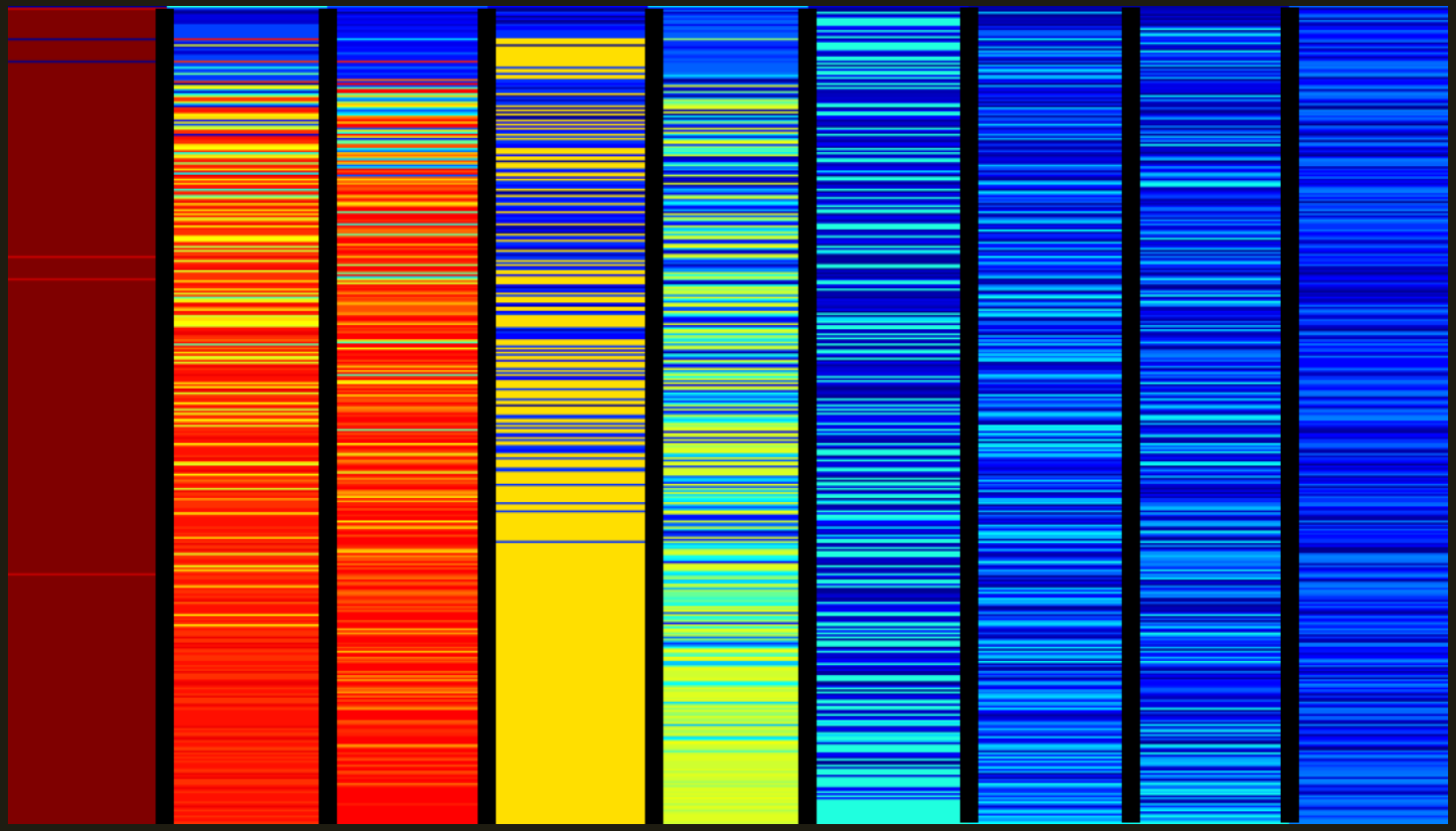
http://www.cs.cmu.edu/~yuandong/kdd2012-dataset.zip

The Groupsize Matrix

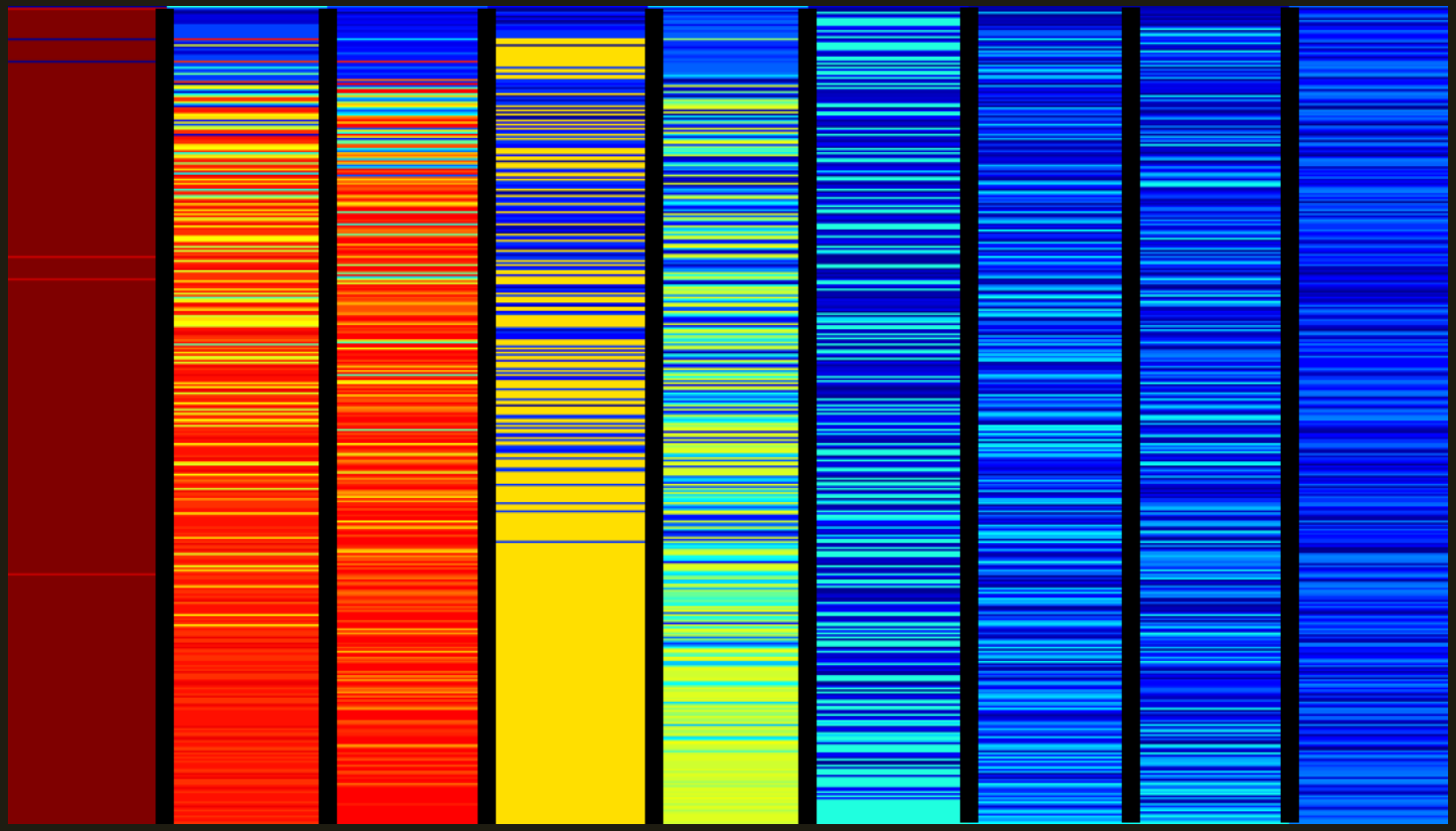


Rank-1 Factorization

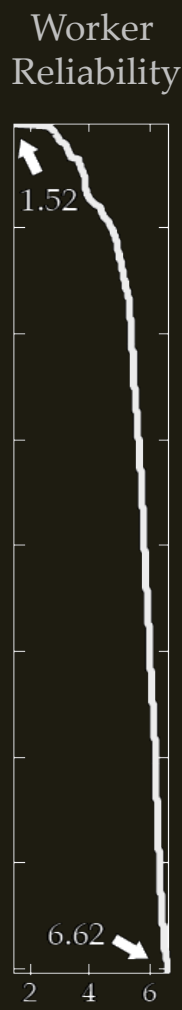
$$\|\#\tilde{\mathbf{Z}} - \boldsymbol{\lambda}\boldsymbol{\mu}^T\|_F / \|\#\tilde{\mathbf{Z}}\|_F = 0.27$$



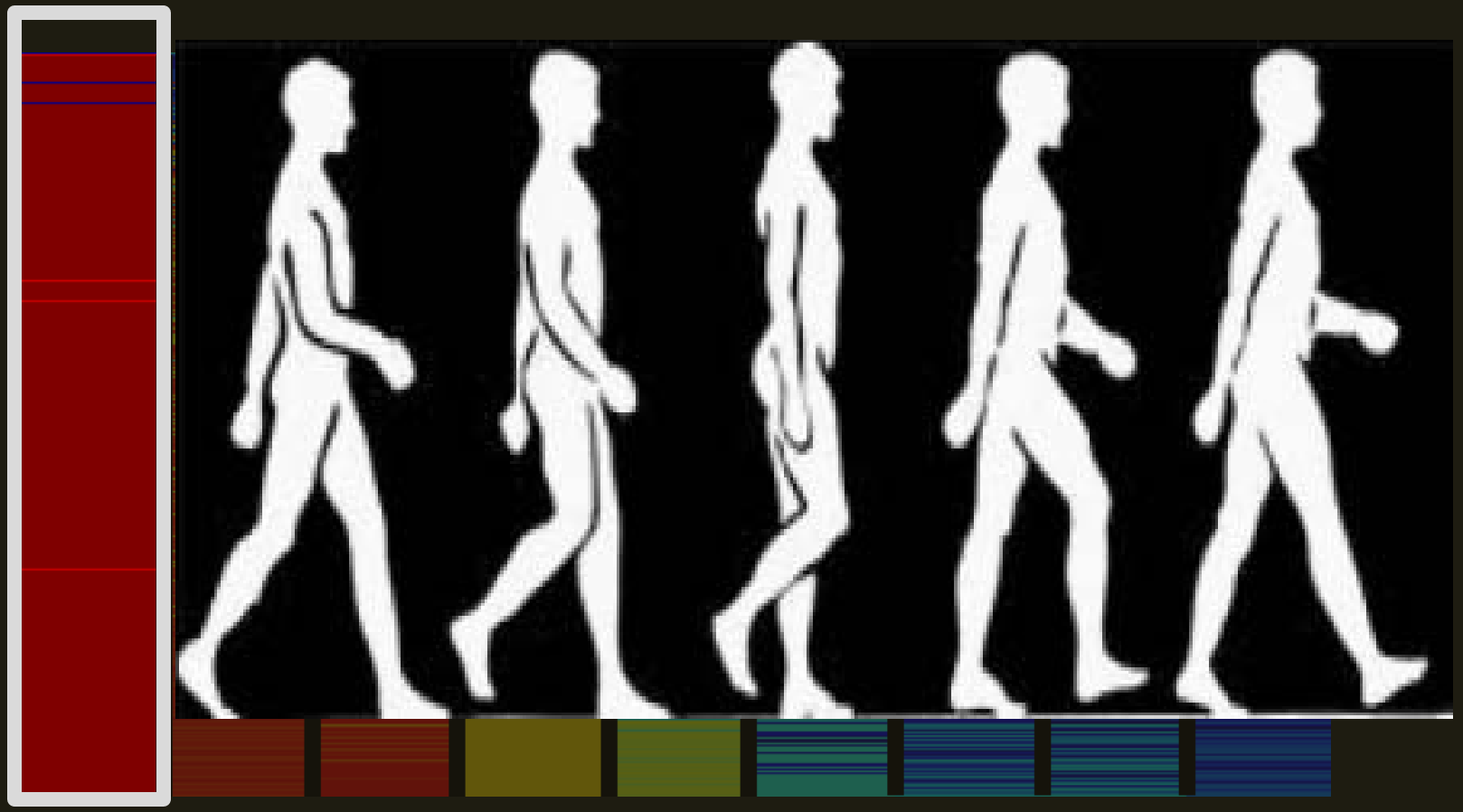
Rank-1 Factorization



Task:	Count2	Sky	Building	Count3	Computer	Count1	Beauty1	Beauty2	Count4
Clarity μ_k:	69.4	52.3	52.0	34.8	28.1	13.9	12.4	11.8	10.2

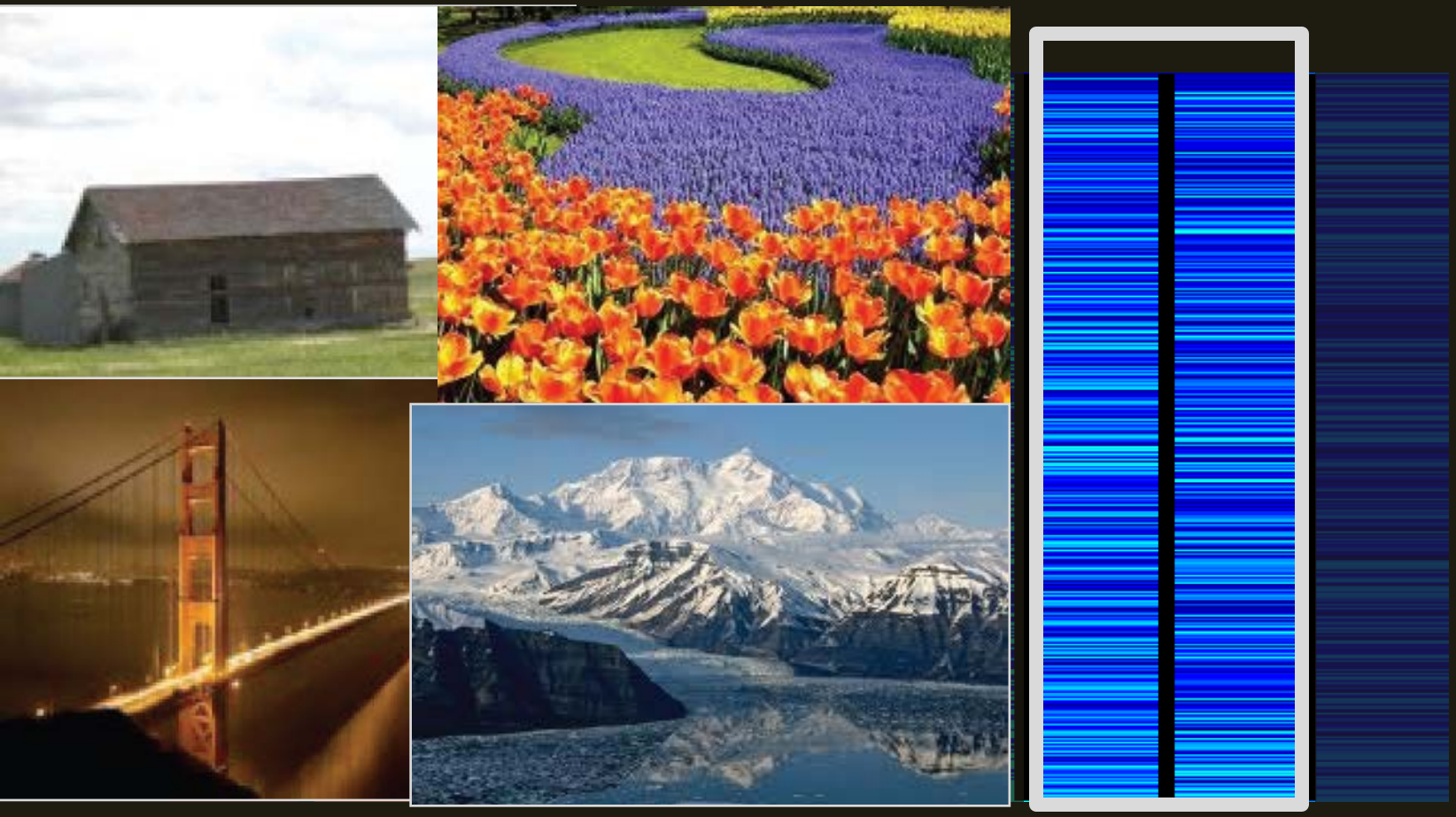


Tasks' clarity



Count 2: Clarity = 69.4

Task's clarity



Task:	Count2	Sky	Building	Count3	Computer	Count1	Beauty1	Beauty2	Count4
Clarity μ_k:	69.4	52.3	52.0	34.8	28.1	13.9	12.4	11.8	10.2

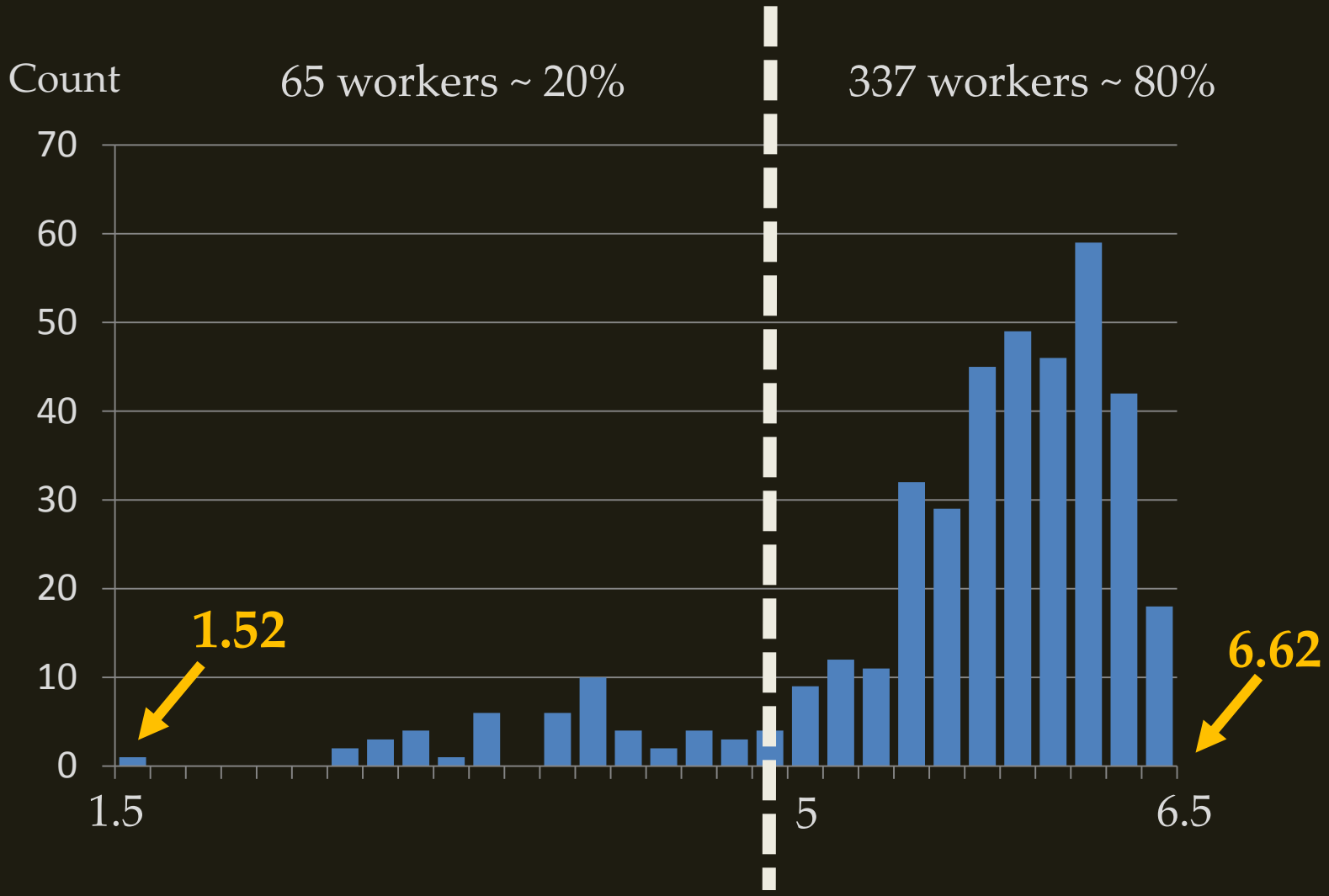
Beauty1 and Beauty2: Clarity = 12.4/11.8

Task's clarity

Task:	Count2	Sky	Building	Count3	Computer	Count1	Beauty1	Beauty2	Count4
Clarity μ_k:	69.4	52.3	52.0	34.8	28.1	13.9	12.4	11.8	10.2

Count 4: Clarity = 10.2

Workers' Reliability



Ranking Workers

Mission I

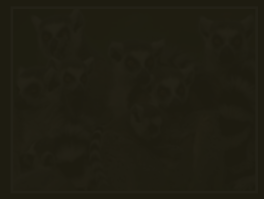
Sky (12)

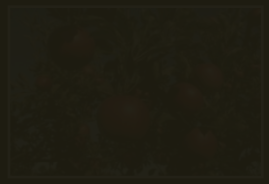
Building (12)

Computer (12)

Mission II

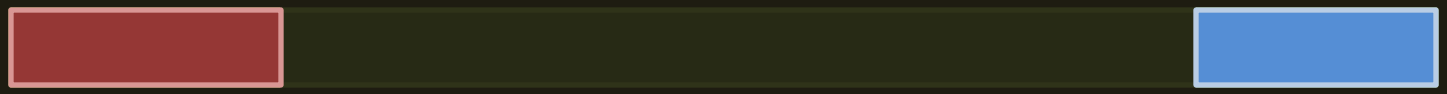
Counting (4)





Mission III

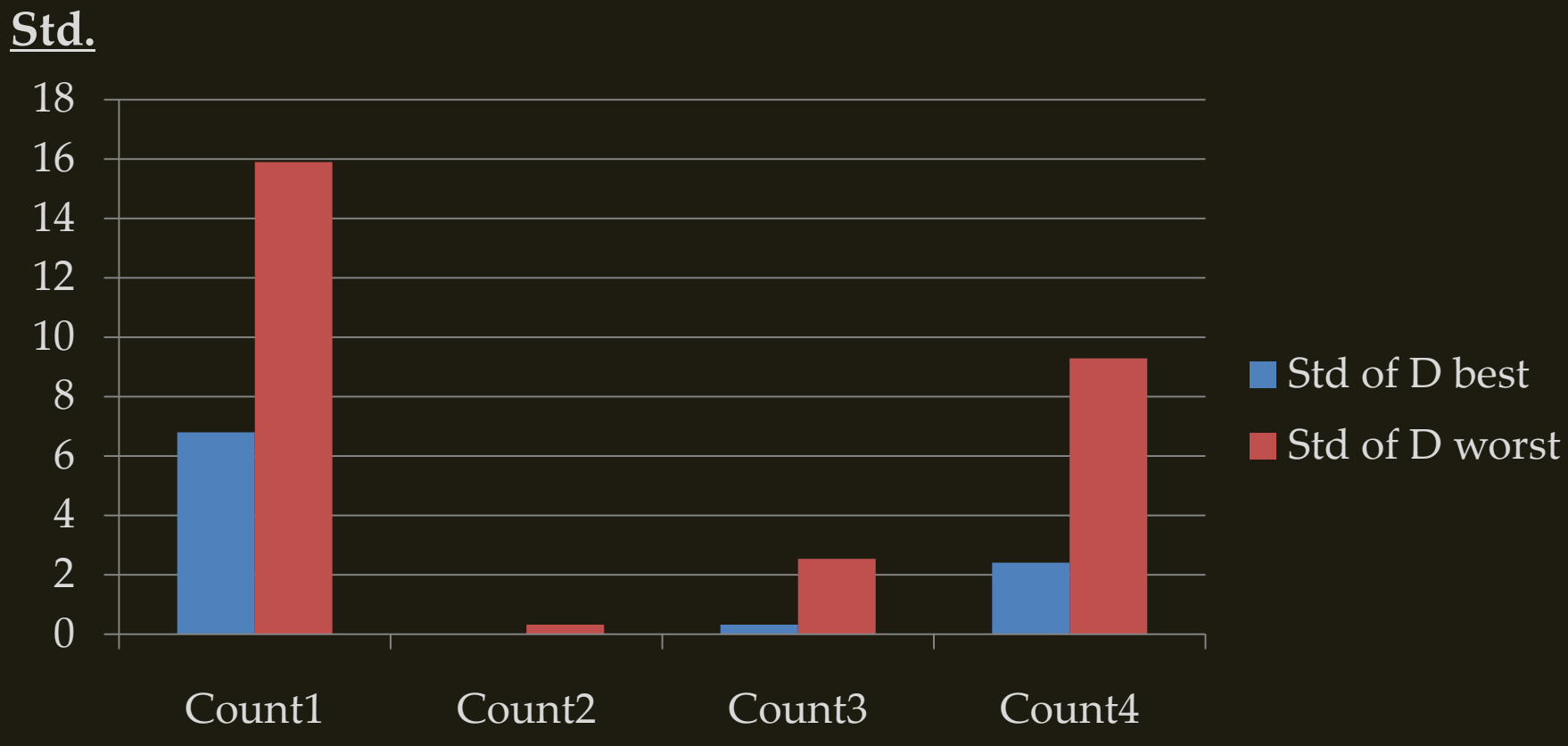
Images Aesthetics (12 + 12)



D most unreliable

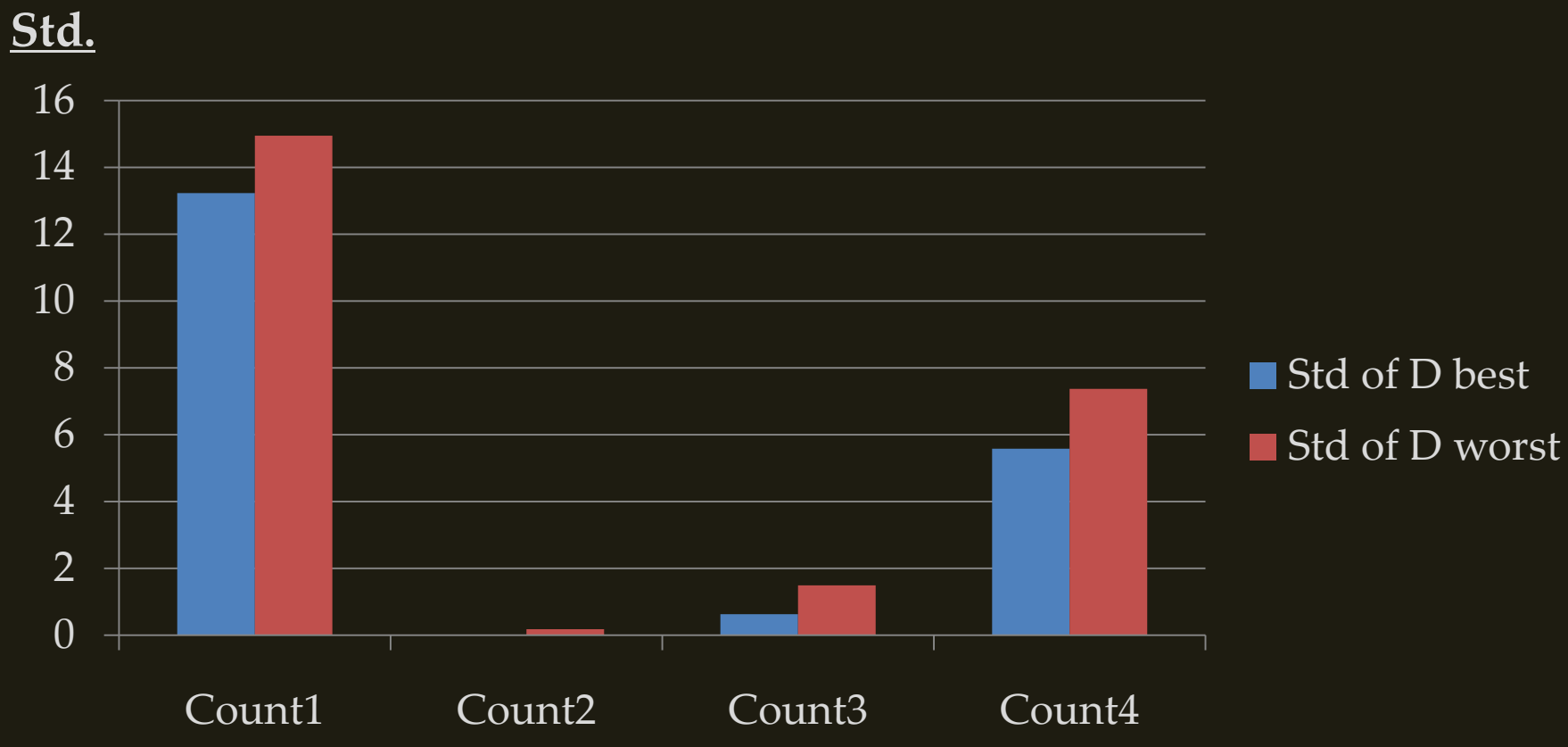
D most reliable

Ranking Workers



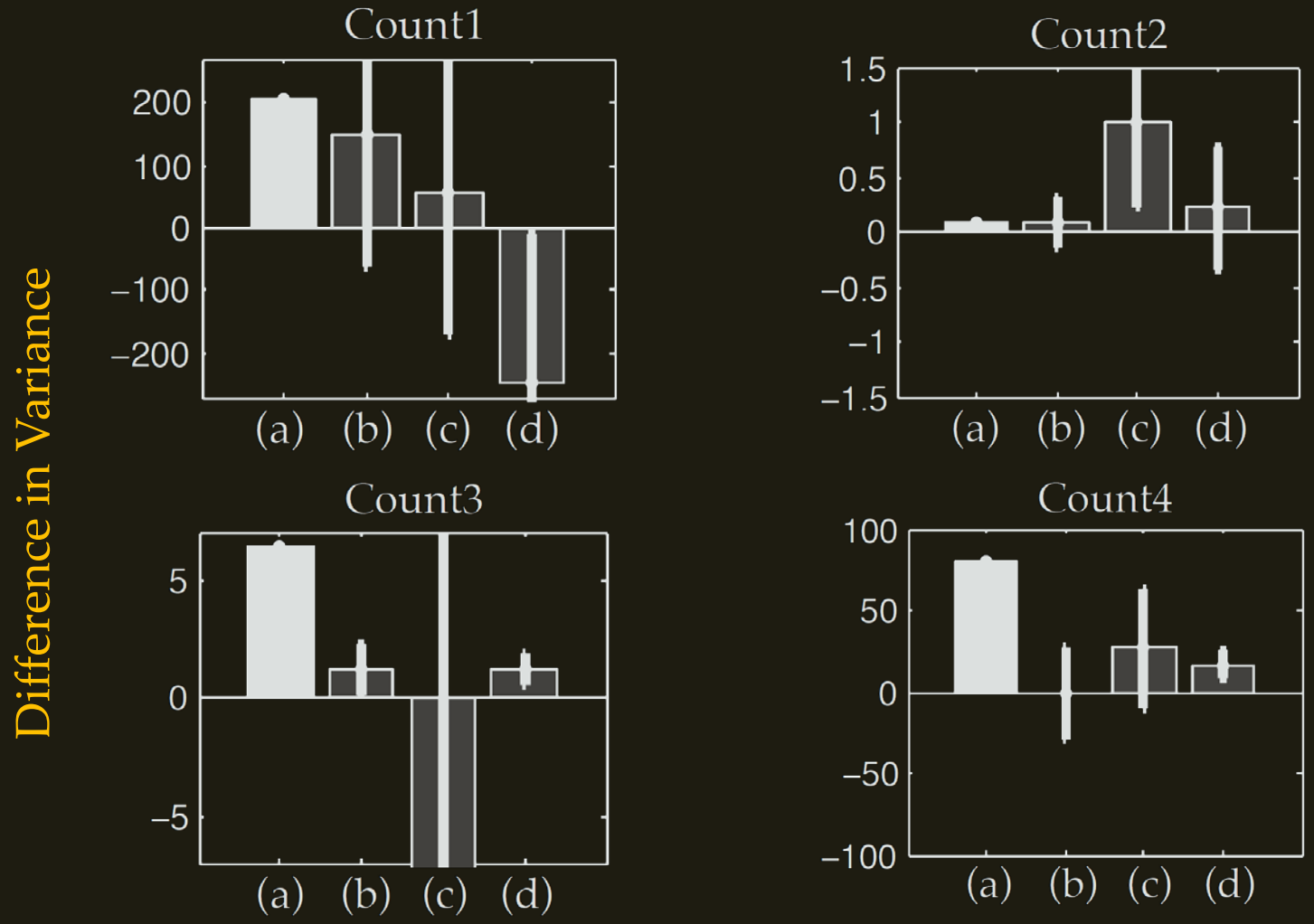
D = 10

Ranking Workers



D = 30

Comparison with Clustering



(a) Our Approach

(b) Spectral Clustering

(c) PCA-Kmeans

(d) Gibbs Sampling

Time Cost

	Methods	Time (sec)
(a)	Our approach	1.41 ± 0.05
(b)	Spectral Clustering	3.90 ± 0.36
(c)	PCA-Kmeans	0.19 ± 0.06
(d)	Gibbs Sampling	53.63 ± 0.19

Predicting Ground truth

	Count1	Count2	Count3	Count4
Ours, D = 5/10	65	5	8	26
Majority Voting	53.7	5.0	9.9	22.9
Majority Voting (Median)	60	5.0	8	24
Learning from Crowd [JMLR'10]	56	5	8	24
Multidimensional Wisdom of Crowds [NIPS'10]	63.7	5	8	26.0
Ground truth	65	5	8	27

Conclusion and Future Work

Conclusion

1. Estimating workers' reliability and tasks' clarity in the presence of **schools of thought**.
2. Applicable to both objective and subjective tasks.
3. Simple solution without iteration, no initial guess.

Future Work

Handling possible missing entries
Improving the scalability.

Thanks!