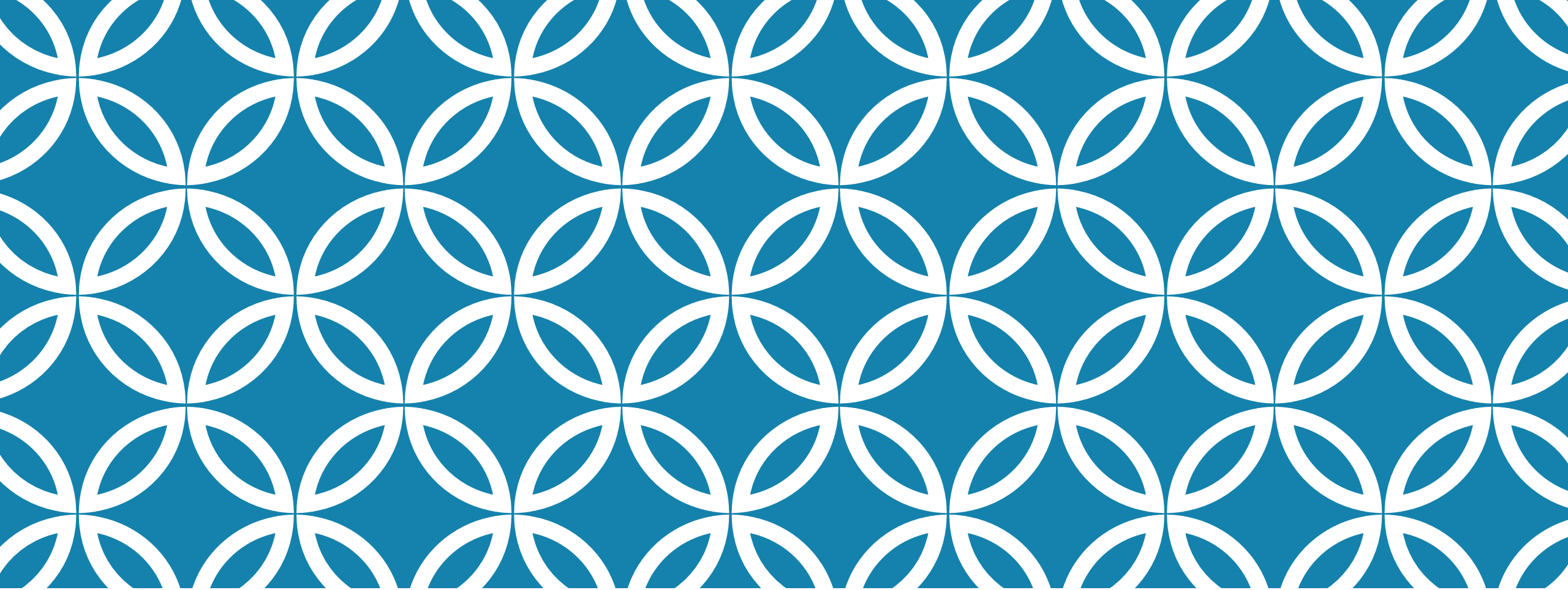
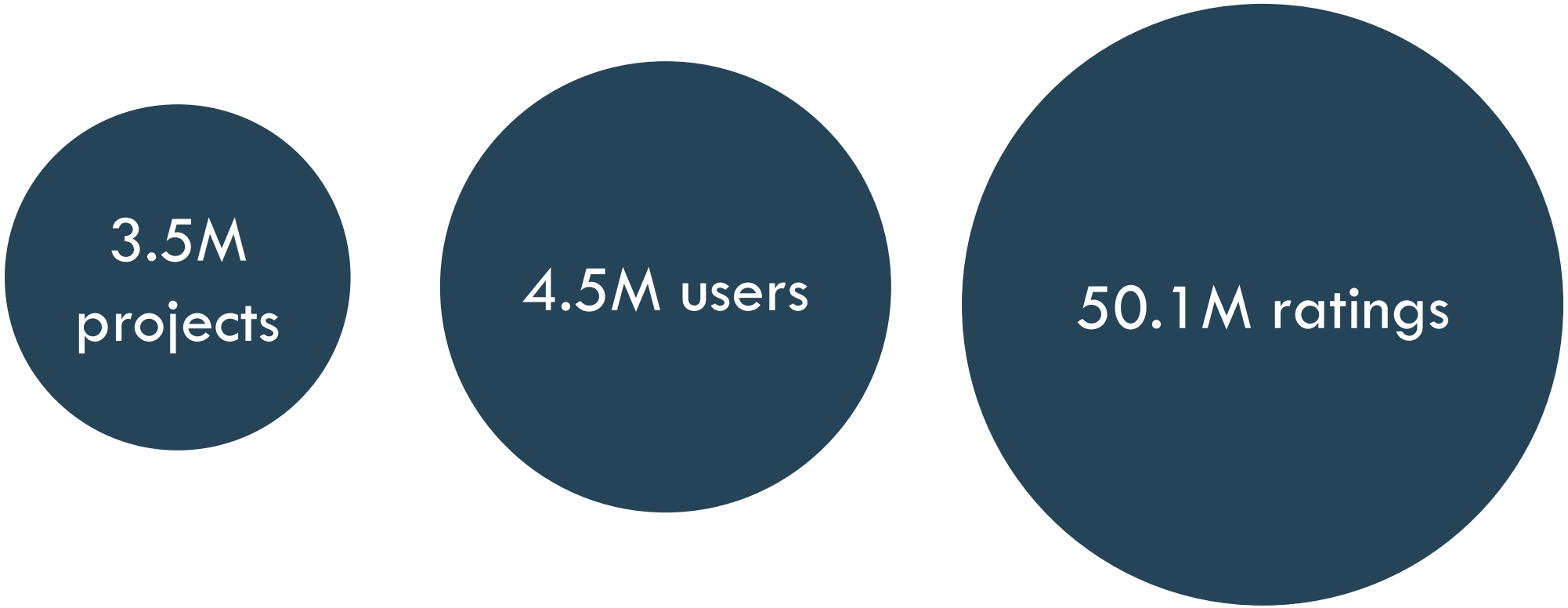


# DISTRIBUTED MATRIX FACTORIZATION

Jenny Lu

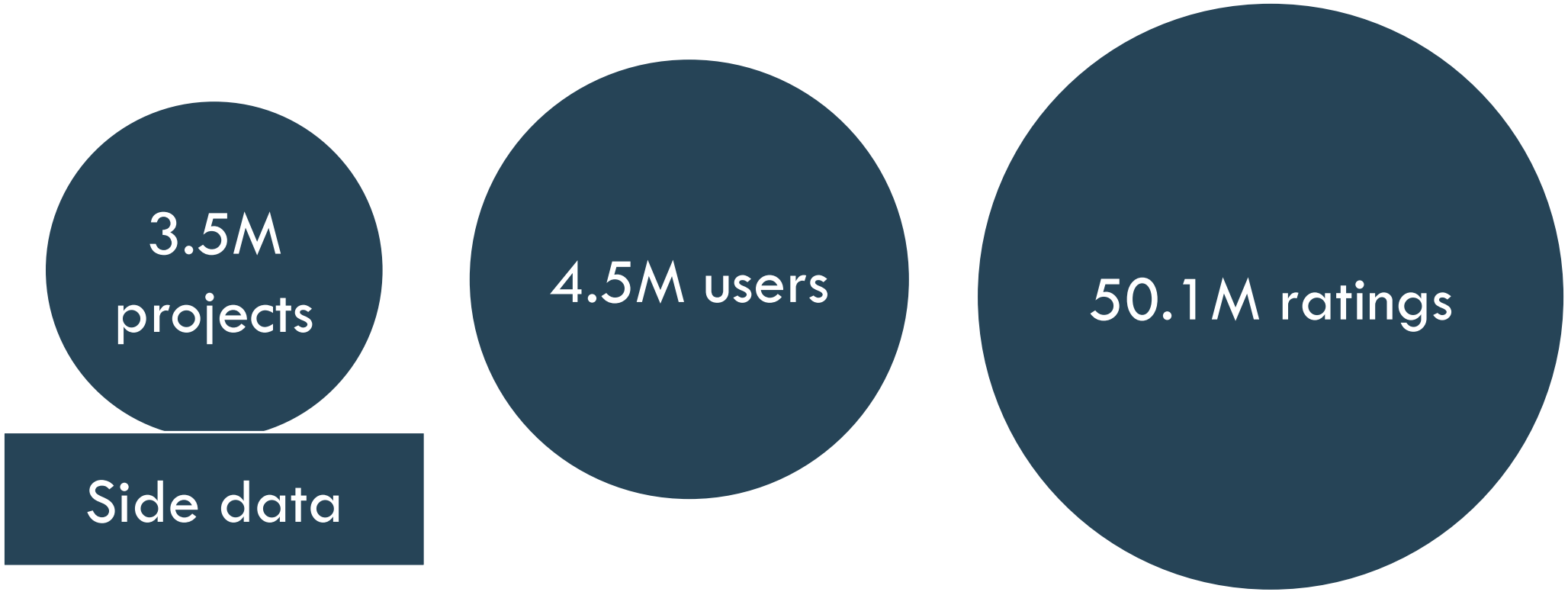
# GOALS

Create a recommender for GitHub's open source projects



# GOALS

Create a recommender for GitHub's open source projects



# MATRIX FACTORIZATION

## ALS

Parallelizable

## SGD

Easier and faster

Suited for sparsity

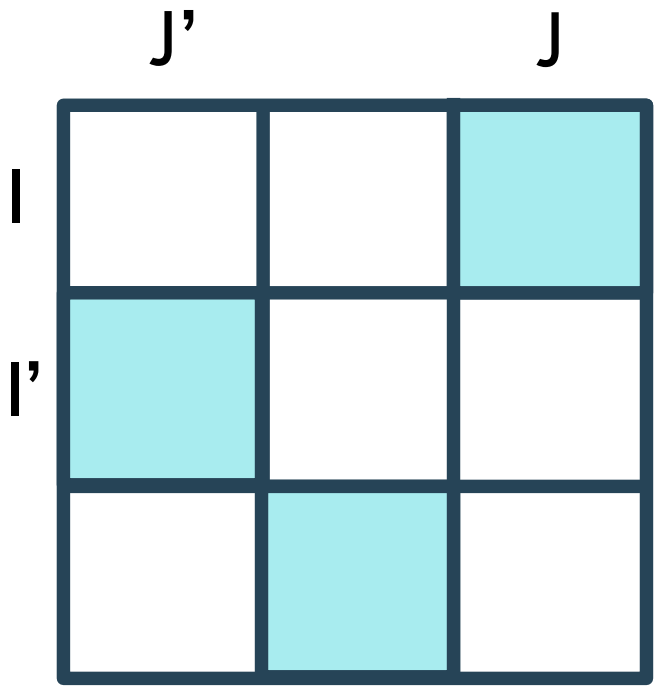
# IMPLEMENTATION

## 1. Partition data

- Find interchangeable matrix blocks such that:

$$I \cap I' = \emptyset$$

$$J \cap J' = \emptyset$$



# IMPLEMENTATION

## 1. Partition data

- Find interchangeable matrix blocks such that:

$$I \cap I' = \emptyset$$

$$J \cap J' = \emptyset$$

## 2. Perform SGD

- Have each worker do SGD updates in parallel

# IMPLEMENTATION

## 1. Partition data

- Find interchangeable matrix blocks such that:

$$I \cap I' = \emptyset$$

$$J \cap J' = \emptyset$$

## 2. Perform SGD

- Have each worker do SGD updates in parallel

## 3. Combine+repeat

- Combine results from each worker
- Repeat until convergence

# DESIGN CHOICES

1 iteration ≠ 1 epoch

Half precision float

# CHALLENGES

## Problem

Results diverged rapidly

## Solution

Scaled ‘rating’ with k

Decreased learning rate

Added learning schedule

# RESULTS

Faster for same number of epochs and k

Lower MSE than ALS

Improved MSE with side data

# POTENTIAL APPLICATIONS

Large matrices (dense and sparse)

Side data

More termination options

# FURTHER WORK

Better side data evaluation

Better heuristics for parameter tuning

# THANK YOU!

Questions?

Jenny Lu

ulynnej@gmail.com | github.com/ulynnej | pltalot.com