Collaborative Filtering in Redshift

How to do recommendations?

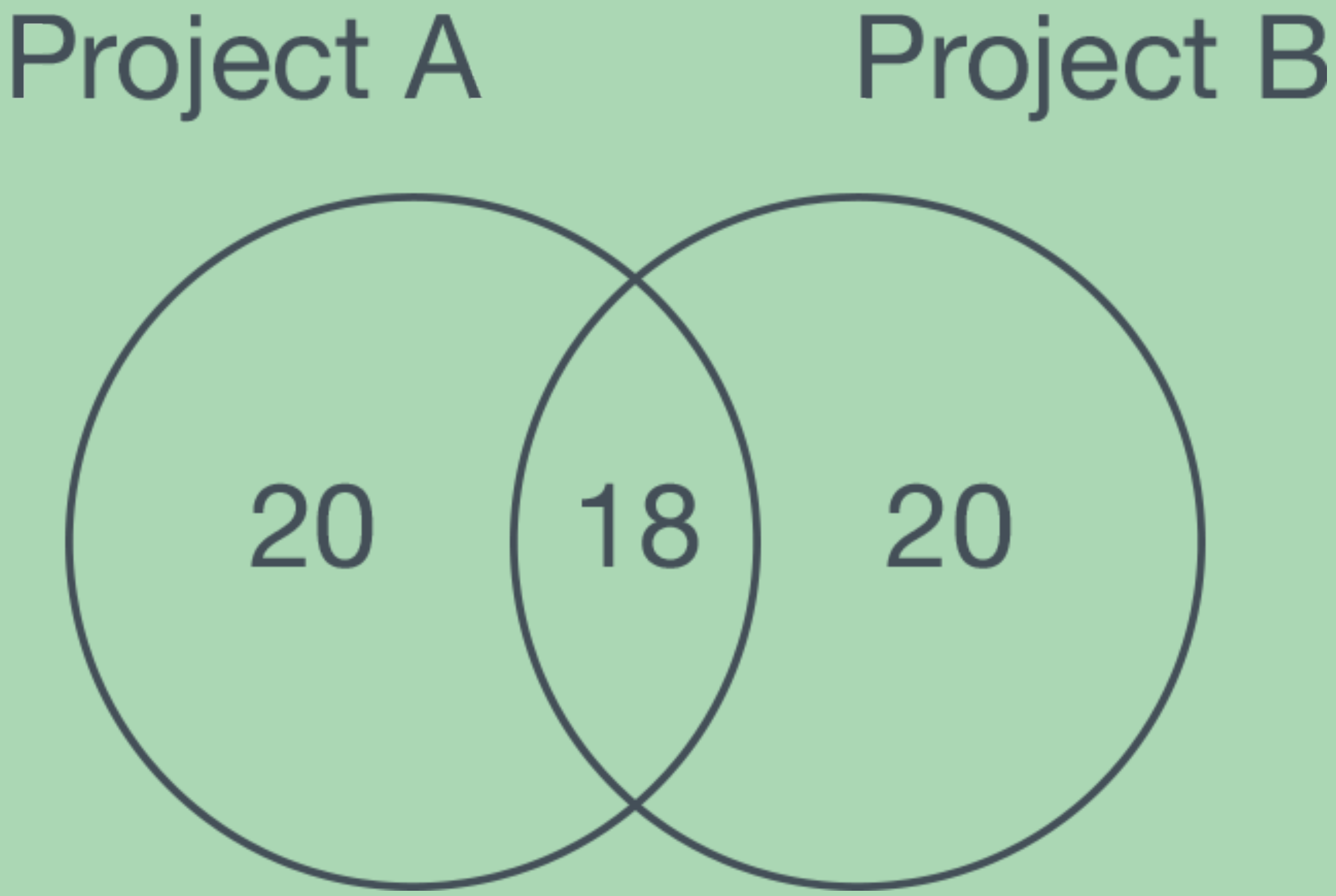
Two projects are "similar" if...

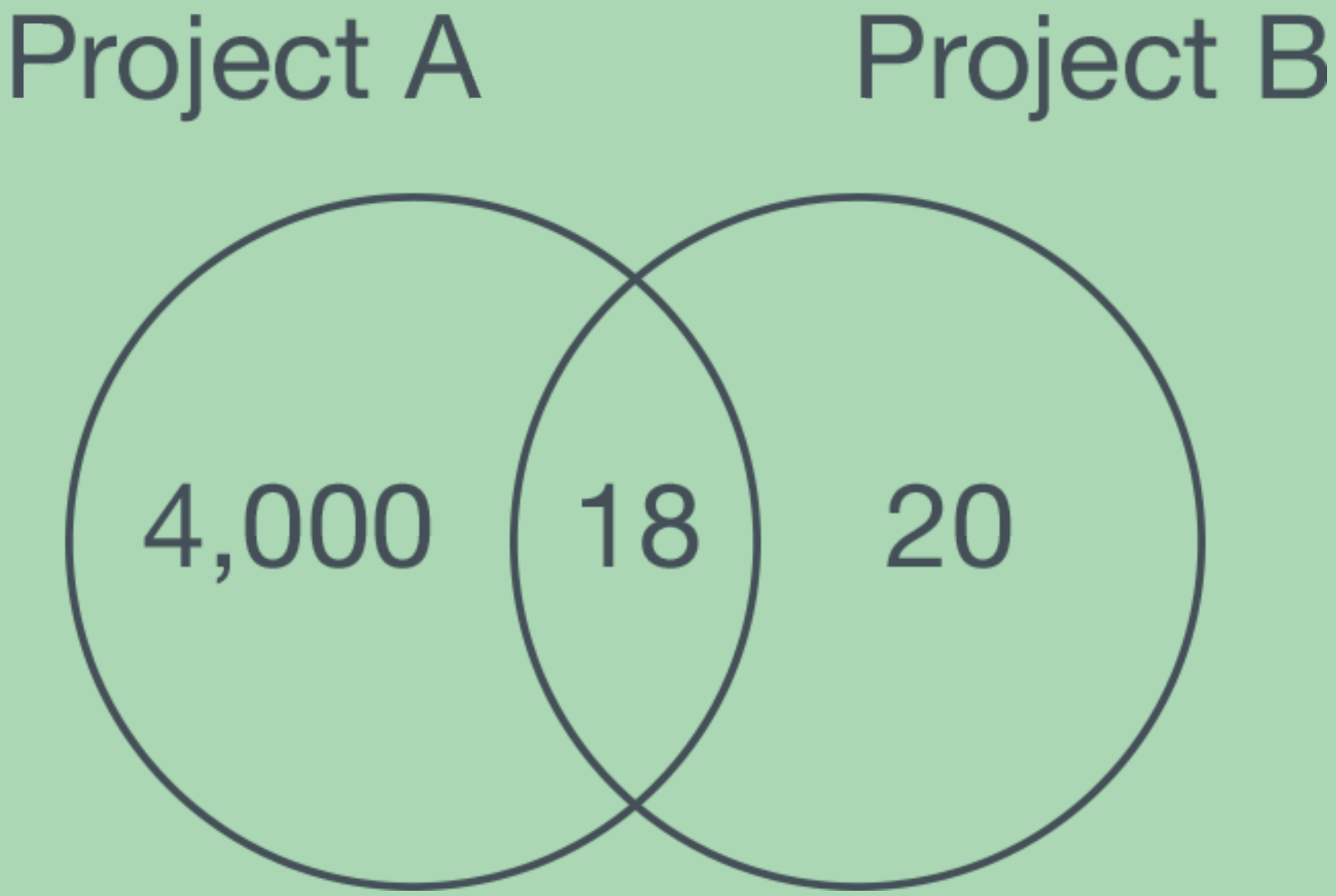
☞ ...their content is "similar" (e.g. description, rewards) using Bayesian Classifiers or something.

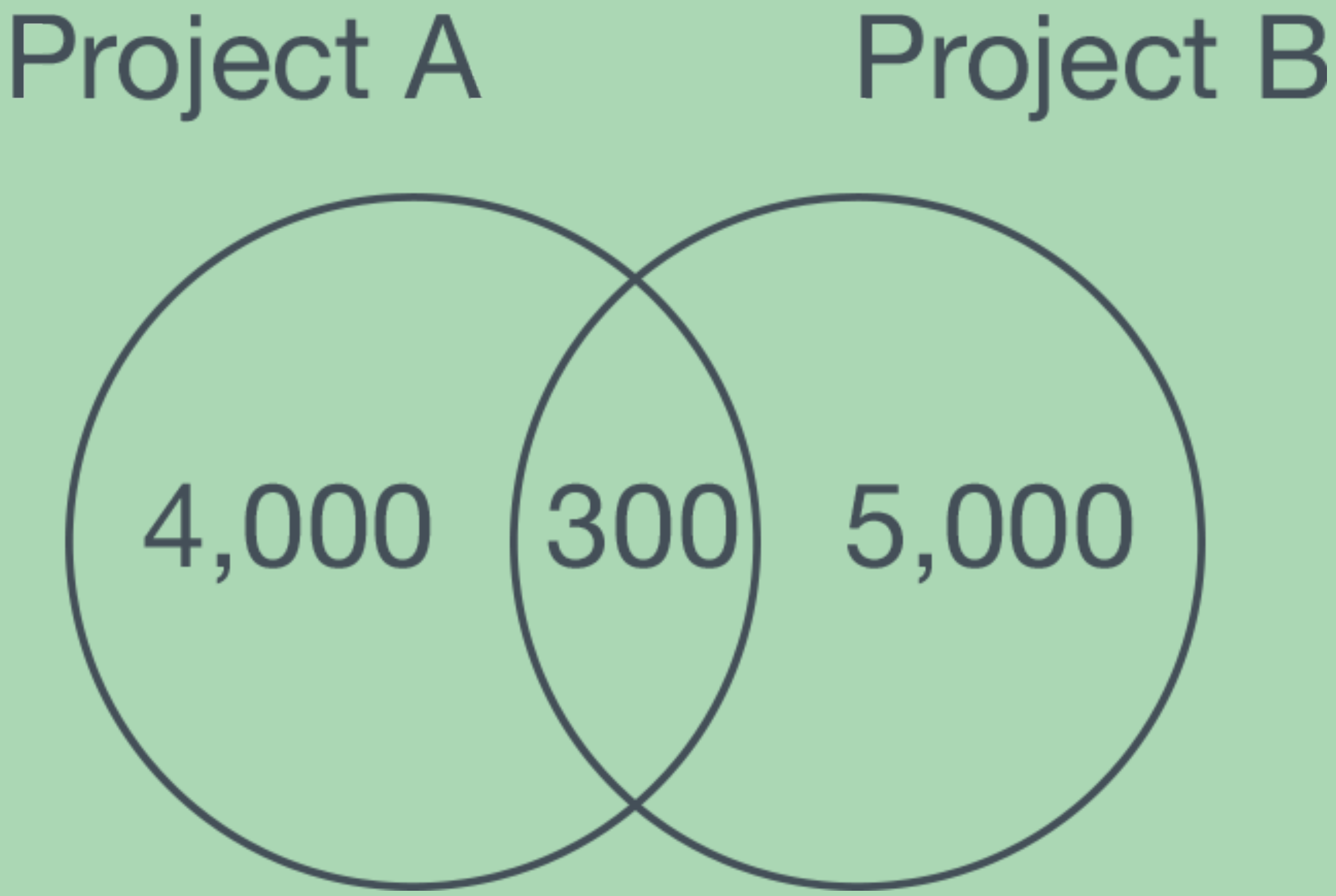
☞ ...they have a high number of common backers.

Shortcomings

- ☞ Content similarity isn't going to surface diverse projects.
- ☞ High # of common backers has bias towards big projects.


Project A
Project B
20
18
20


Project A
Project B
4,000
18
20


Project A
Project B
4,000
300
5,000

Collaborative filtering

- ☞ Represent a project as an "arrow" in the "space" of all Kickstarter users.
- ☞ Two projects are "similar" if their arrows are nearly facing the same direction
  - ☞ i.e. the angle between the arrows is nearly 0.0

Some math

☞ Let `v = (v_1, v_2, ..., v_n)`
and `w = (w_1, w_2, ..., w_n)`

☞ The length is defined as:
`|v| = sqrt(sum(v_i))`

☞ The inner product is defined as:
`v · w = sum(v_i * w_i)`

☞ Then the cosine of the angle between vectors is:
`cos angle = v · w / (|v| |w|)`

The math in our context

- ☞ The vectors consist of just 0's and 1's (0 for not a backer, 1 for a backer).
- ☞ The length of a vector is simply the `sqrt` of the `backers_count`.
- ☞ The inner product is simply the # of common backers.
- ☞ The cosine of the angle is the thing we want to computer and sort by!

Computing the inner product in Redshift

(live code in redshift)

Implementation details

- ☞ Sorted set in redis for every project containing "similar" projects weighted by their "angle."

    - ☞ sets will contain only live projects, and probably only acceptable and better.

- ☞ Recommended projects for a user = union of all sorted sets corresponding to projects they backed.

- ☞ Further refine in SQL by category and making sure we don't recommend something they already backed.

Pitfalls of Collaborative Filtering

- Data
  - new projects
  - cold starts
- Diversity

Future tweaks that can be made

- ☞ Taking into account pledge amount and average pledge per project.

Future applications

☞ Social (recommended user to follow)