

How Well Does This Approach Work?

Doc d_i

Cavaliers vs. Golden State Warriors: NBA playoff finals ...
basketball game ... **travel** to Cleveland ... **star** ...

θ_1 "Sports"

$$\pi_{i1} \propto c(\text{"sports"}, d_i) = 0$$

1. Need to count related words also!

θ_2 "Travel"

$$\pi_{i2} \propto c(\text{"travel"}, d_i) = 1 > 0$$

...

2. "Star" can be ambiguous (e.g., star in the sky).

θ_k "Science"

$$\pi_{ik} \propto c(\text{"science"}, d_i) = 0$$

3. Mine complicated topics?