Improved Idea: Topic = Word Distribution

θ_1 "Sports"

$P(w|\theta_1)$

sports	0.02
game	0.01
basketball	0.005
football	0.004
play	0.003
star	0.003
...	
nba	0.001
...	
travel	0.0005
...	

θ_2 "Travel"

$P(w|\theta_2)$

travel	0.05
attraction	0.03
trip	0.01
flight	0.004
hotel	0.003
island	0.003
...	
culture	0.001
...	
play	0.0002
...	

• • •

θ_k "Science"

$P(w|\theta_k)$

science	0.04
scientist	0.03
spaceship	0.006
telescope	0.004
genomics	0.004
star	0.002
...	
genetics	0.001
...	
travel	0.00001
...	

$$\sum_{w \in V} p(w|\theta_i) = 1$$

Vocabulary Set: V={w1, w2,….}