

Can we predict the popularity  
of an article from its title?

Jacob Andreas  
HackNY FA'11

6,764 titles

from 4 tech websites

(data courtesy of Parse.ly)

## 820 features

“My Brief OKCupid Affair with a World Champion Magic:  
The Gathering Player” (1348 hits)

(popular,

“my brief okcupid affair with a world champion magic : the  
gathering player”,

“PRP\$ NN VBD NN IN DT NN NN NN : DT NN NN”)

(popular,

[“PRP\$”, “NN VBD”, “DT NN NN”, “brief affair”, . . . ])

# Some basic results

(MNB classifier with 10-fold cross-validation vs. majority baseline)

	Acc.	$f_{pop}$	$f_{avg}$
baseline	.837	0	.763
us	.818	.491	.824

# Some useful features

(Top features by info. gain ratio)

1. iphone 5
2. IN PRP (e.g. "to him")
3. kindle fire
4. fire
5. DT NN NN (e.g. "the blue cat")
6. kindle
7. WRB PRP (e.g. "why she")
8. is
9. PRP (e.g. "him")

Let's play a game.