

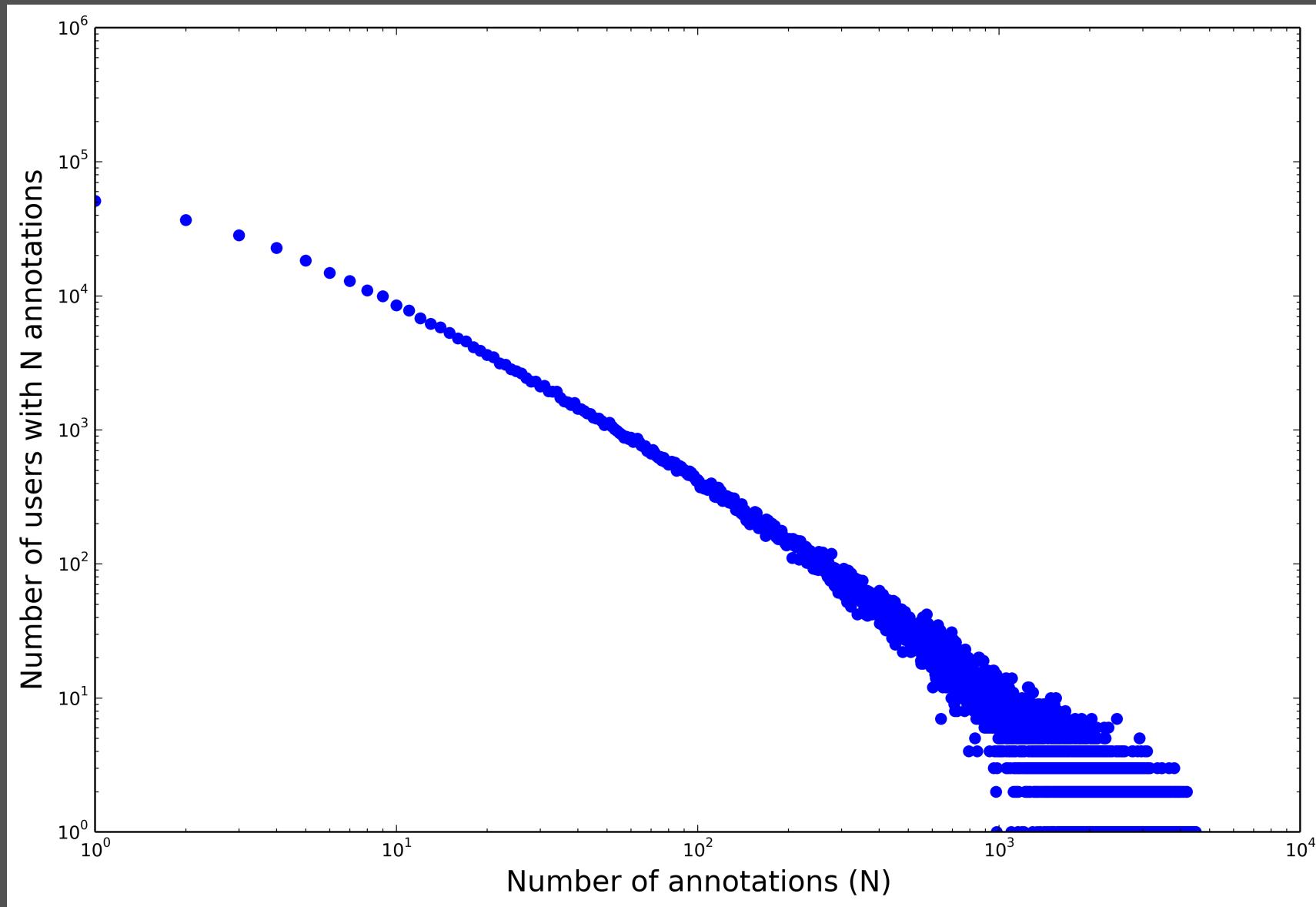


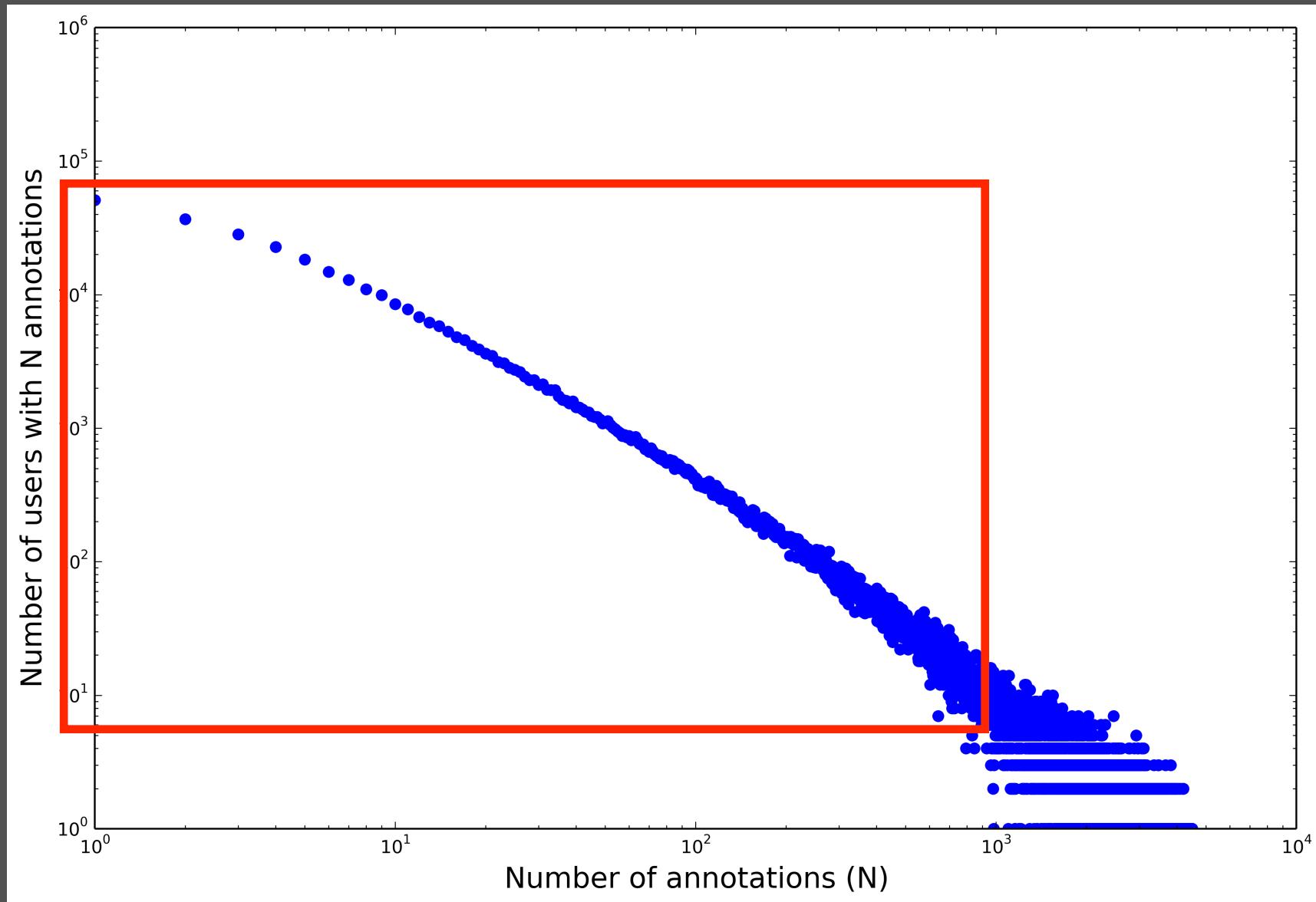
“Supertagger” Behavior in Building Folksonomies

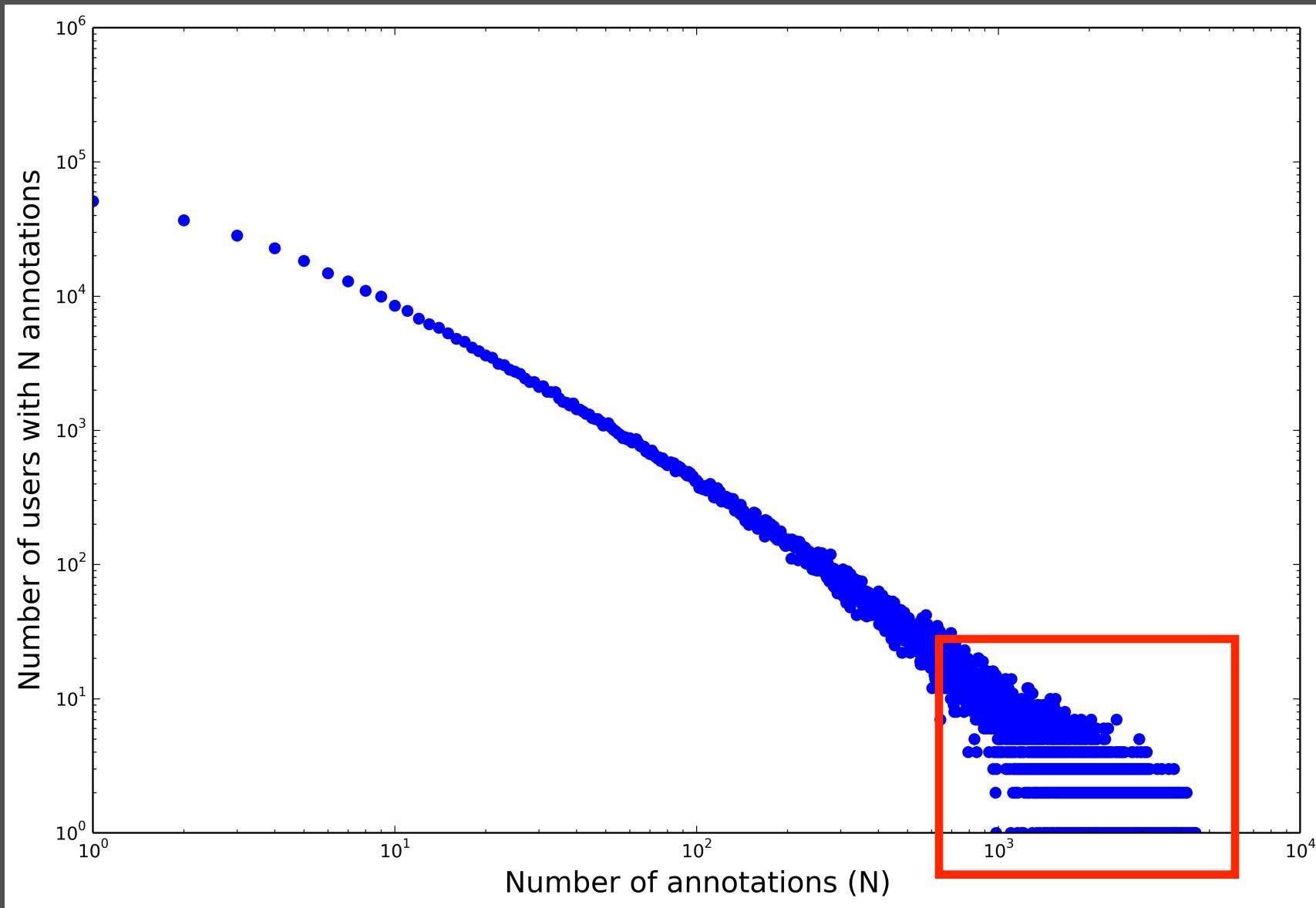
Jared Lorince¹, Sam Zorowitz², Jaimie Murdock¹, and Peter M. Todd¹

¹Indiana University

²John Hopkins University







What's the difference between prolific
and non-prolific tagging behavior?

- 1: How do prolific and non-prolific *users* differ?
- 2: How do the *folksonomies* generated by these users differ?
- 3: What causal factors might be driving these differences?

Dataset

last.fm

Dataset

last.fm

Total users	1,884,597
Friendship relations	24,320,919
Total annotations	50,372,895
Users with ≥ 1 annotation	521,780
Total unique tags	1,029,091
Unique items tagged	4,477,593
Total Scrobbles	1,181,674,857
Users with scrobbles recorded	73,251
Unique items scrobbled	32,864,795
Total loved tracks	162,788,213
Users with ≥ 1 loved track	1,355,859
Total banned tracks	23,321,347
Users with ≥ 1 banned track	502,758
Unique Groups	117,663
Users with ≥ 1 group membership	827,232

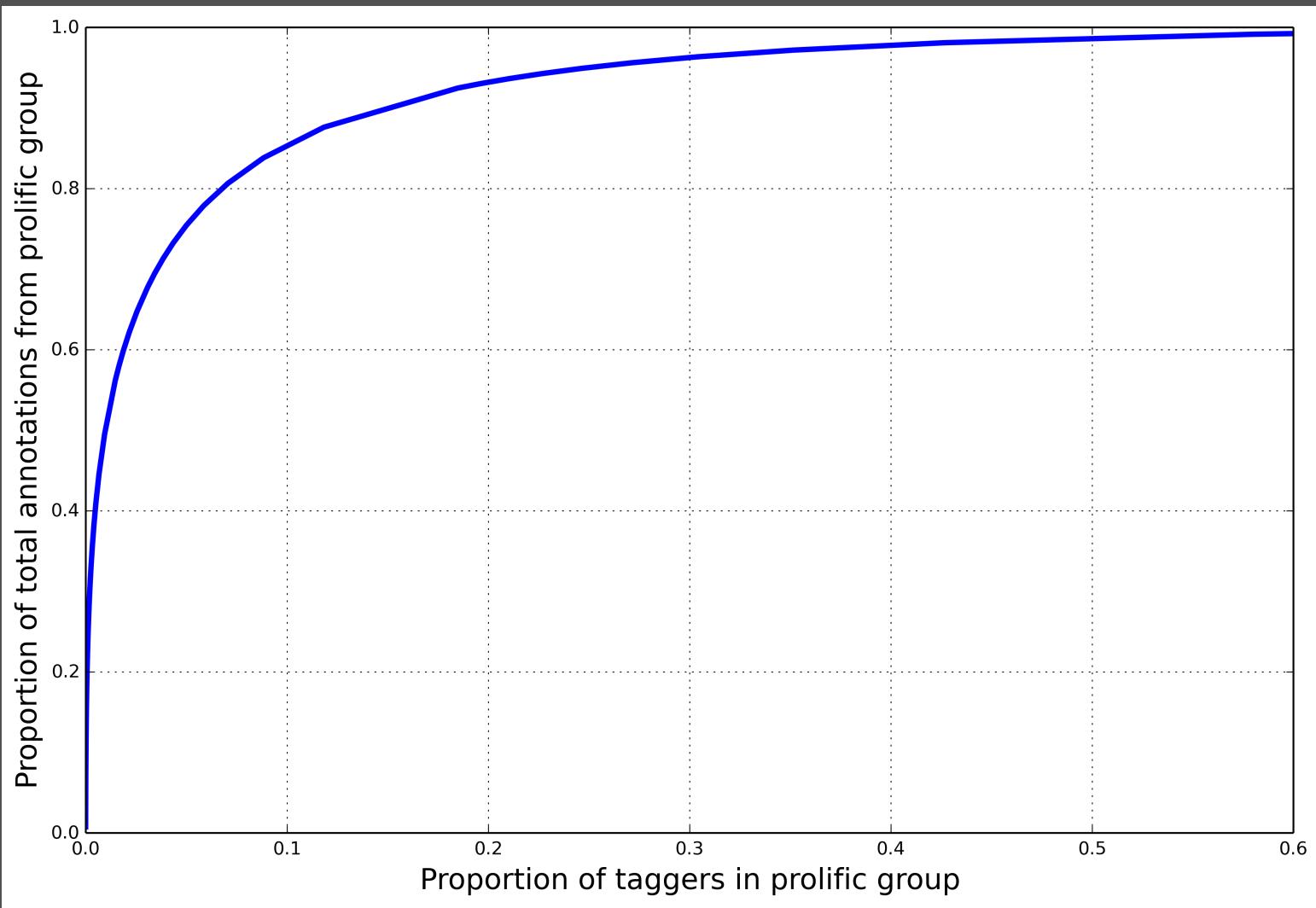
Dataset

last.fm

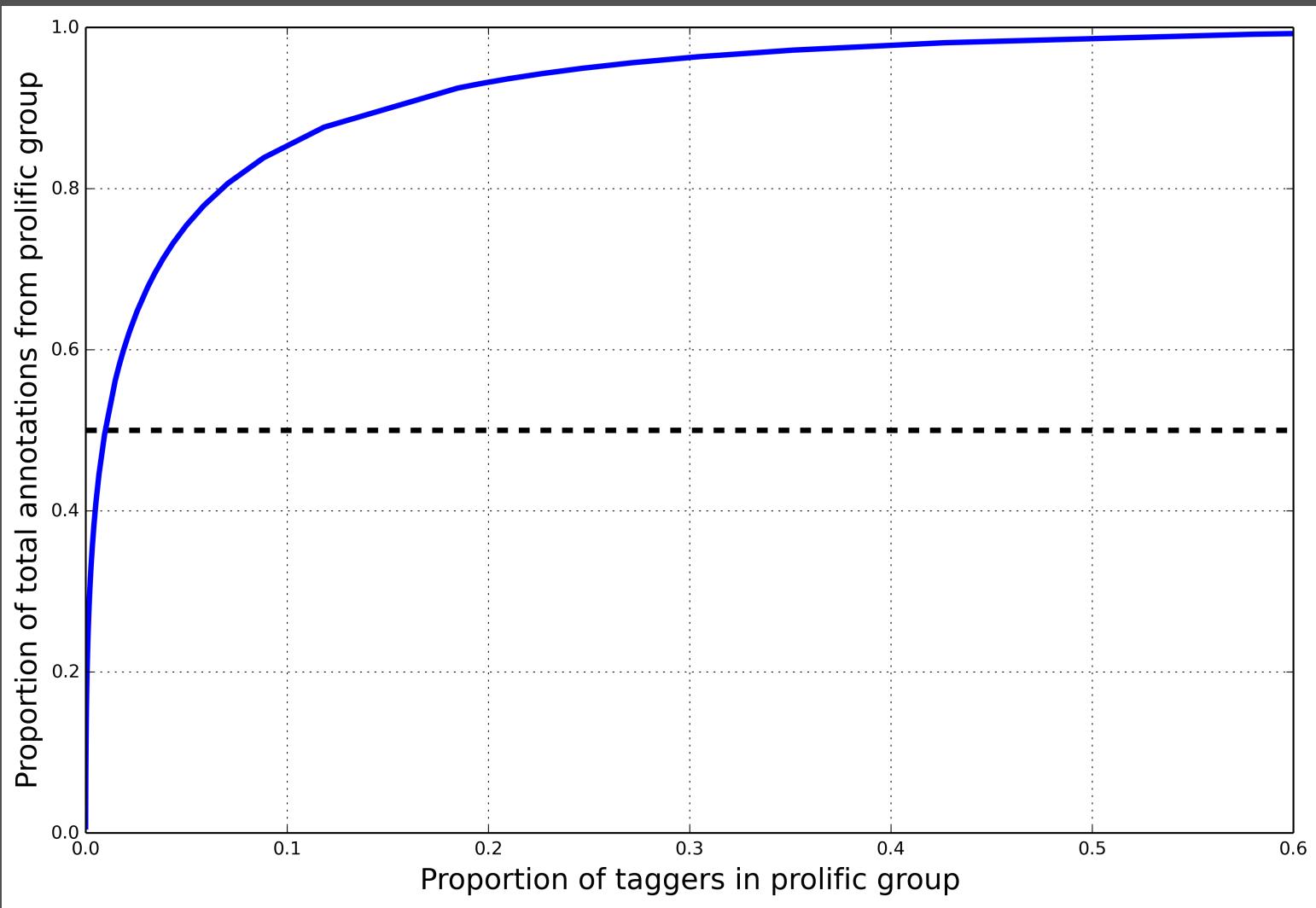
Total users	1,884,597
Friendship relations	24,320,919
Total annotations	50,372,895
Users with ≥ 1 annotation	521,780
Total unique tags	1,029,091
Unique items tagged	4,477,593
Total Scrobbles	1,181,674,857
Users with scrobbles recorded	73,251
Unique items scrobbled	32,864,795
Total loved tracks	162,788,213
Users with ≥ 1 loved track	1,355,859
Total banned tracks	23,321,347
Users with ≥ 1 banned track	502,758
Unique Groups	117,663
Users with ≥ 1 group membership	827,232

Problem Formalization

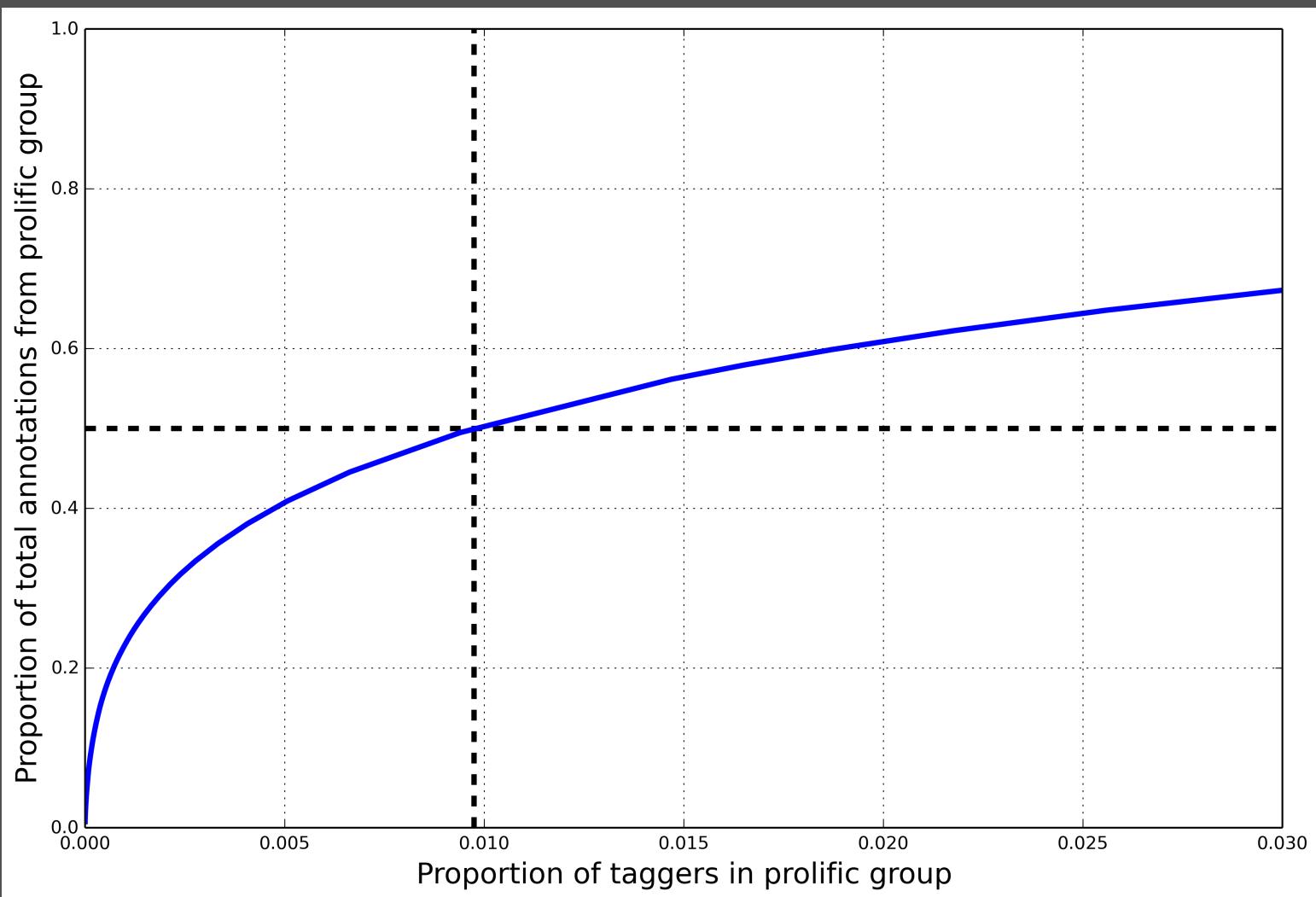
Problem Formalization



Problem Formalization



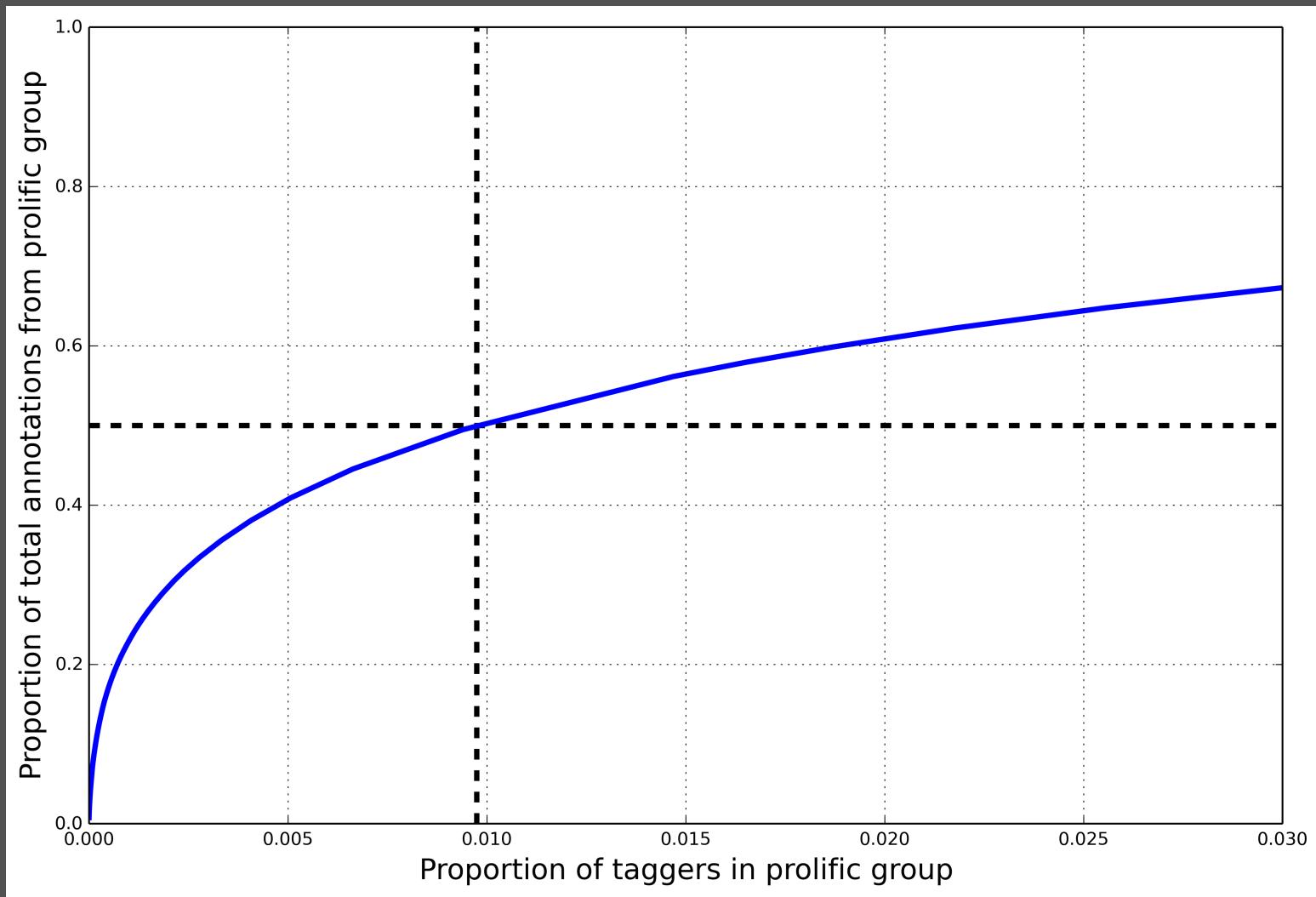
Problem Formalization



Problem Formalization

5,086
prolific taggers

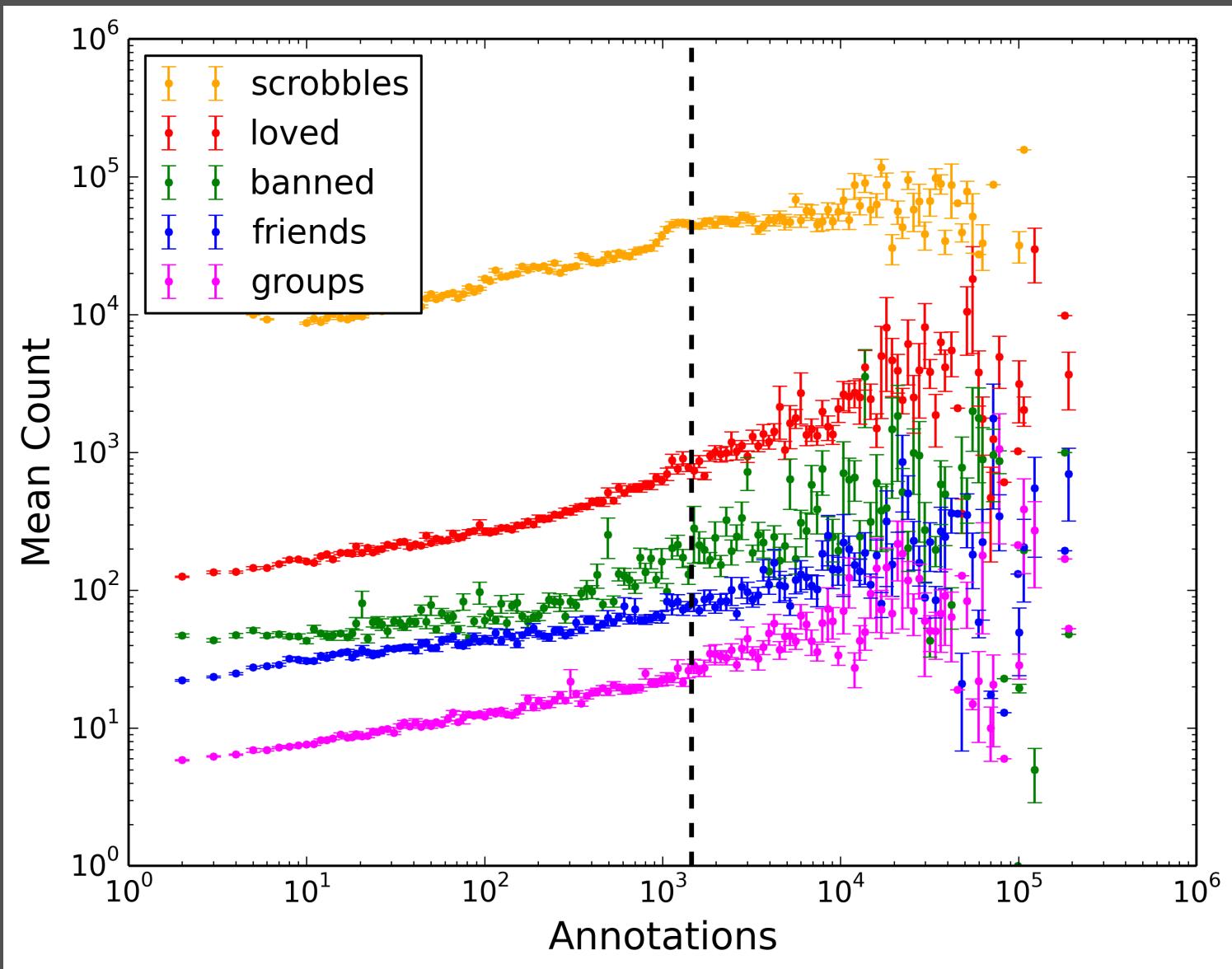
516,694 non-
prolific taggers



Question 1: Are the users different?

Question 1: Are the users different?

Yes, but not in a surprising way.



Question 2: Do the users tag differently?

Question 2: Do the users tag differently?

A: Do they tag with similar vocabularies?

B: Do they tag similar content?

	<i>Prolific</i>	<i>Non-Prolific</i>
Total users	5,086	516,694
Total Annotations	25,185,083	25,187,812
Total tags	399,552 (78.56)	797,784 (1.54)
Unique tags	231,307	629,539
Shared tags		168,245
Total Items	2,992,046 (588.29)	2,515,070 (4.87)
Unique items	1,962,523	1,485,547
Shared Items		1,029,523

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Rank	Tag (P)	Tag (NP)
1	“rock”	“seen live”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Rank	Tag (P)	Tag (NP)
1	“rock”	“seen live”
2	“electronic”	“rock”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Rank	Tag (<i>P</i>)	Tag (<i>NP</i>)
1	“rock”	“seen live”
2	“electronic”	“rock”
3	“pop”	“indie”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Rank	Tag (<i>P</i>)	Tag (<i>NP</i>)
1	“rock”	“seen live”
2	“electronic”	“rock”
3	“pop”	“indie”
4	“alternative”	“alternative”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Rank	Tag (<i>P</i>)	Tag (<i>NP</i>)
1	“rock”	“seen live”
2	“electronic”	“rock”
3	“pop”	“indie”
4	“alternative”	“alternative”
5	“indie”	“electronic”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Rank	Tag (<i>P</i>)	Tag (<i>NP</i>)
1	“rock”	“seen live”
2	“electronic”	“rock”
3	“pop”	“indie”
4	“alternative”	“alternative”
5	“indie”	“electronic”
6	“female vocalists”	“pop”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Rank	Tag (<i>P</i>)	Tag (<i>NP</i>)
1	“rock”	“seen live”
2	“electronic”	“rock”
3	“pop”	“indie”
4	“alternative”	“alternative”
5	“indie”	“electronic”
6	“female vocalists”	“pop”
7	“metal”	“female vocalists”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

Rank	Tag (<i>P</i>)	Tag (<i>NP</i>)
1	“rock”	“seen live”
2	“electronic”	“rock”
3	“pop”	“indie”
4	“alternative”	“alternative”
5	“indie”	“electronic”
6	“female vocalists”	“pop”
7	“metal”	“female vocalists”
8	“jazz”	“metal”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

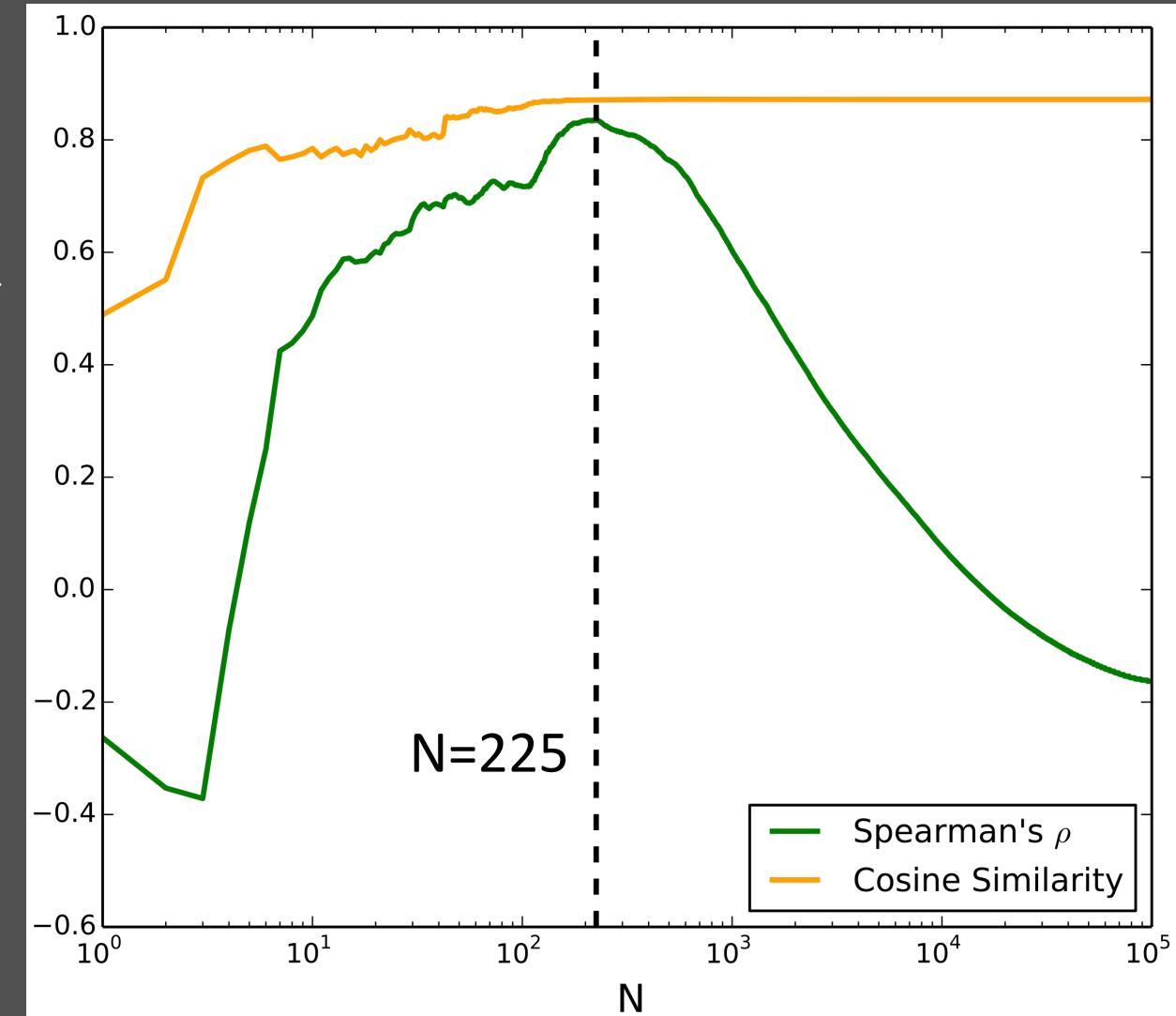
Rank	Tag (<i>P</i>)	Tag (<i>NP</i>)
1	“rock”	“seen live”
2	“electronic”	“rock”
3	“pop”	“indie”
4	“alternative”	“alternative”
5	“indie”	“electronic”
6	“female vocalists”	“pop”
7	“metal”	“female vocalists”
8	“jazz”	“metal”
9	“experimental”	“alternative rock”

Vocabulary Similarity:

Cosine Similarity: 0.872

Rank Correlation: -0.219

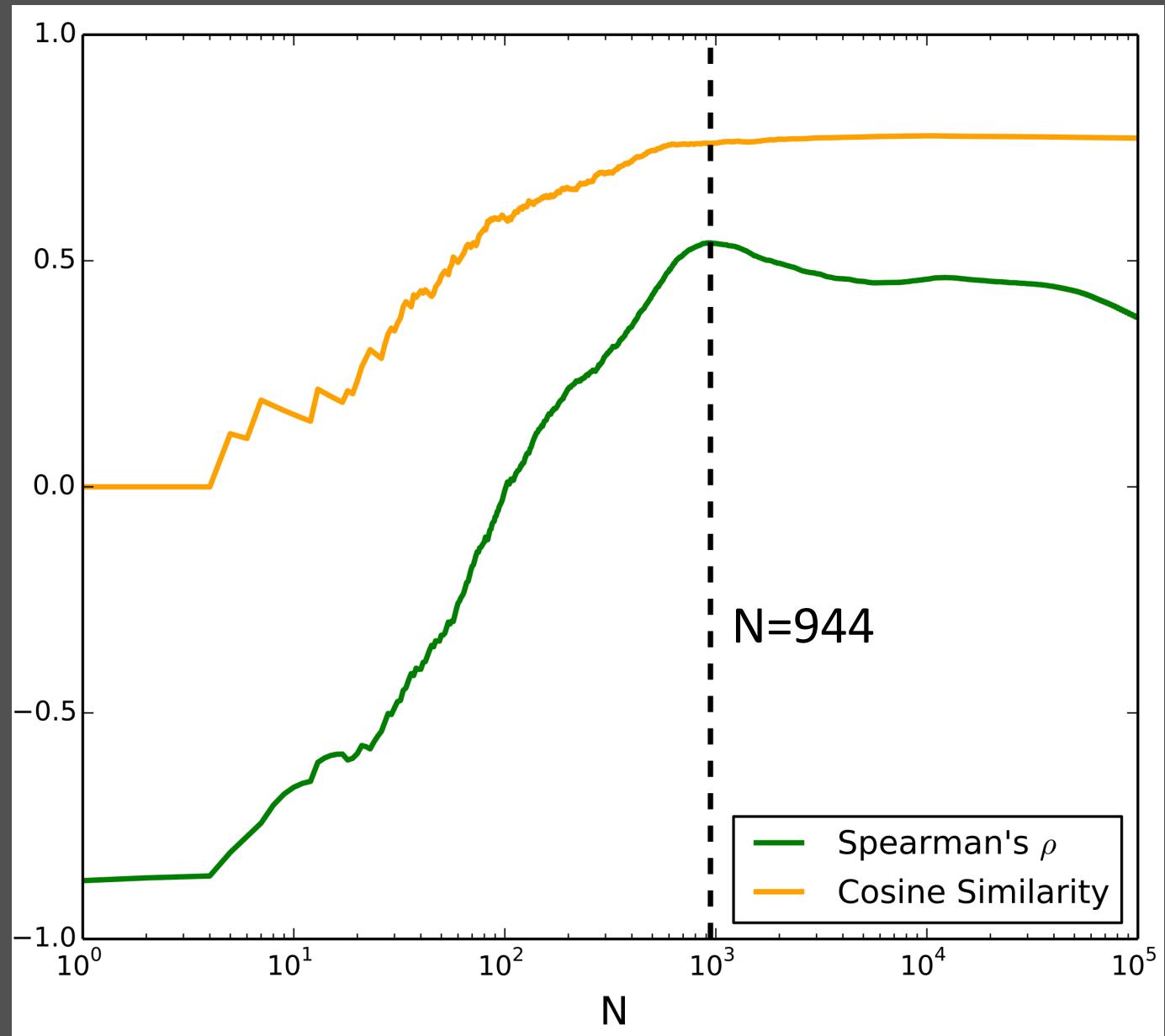
Rank	Tag (P)	Tag (NP)
1	“rock”	“seen live”
2	“electronic”	“rock”
3	“pop”	“indie”
4	“alternative”	“alternative”
5	“indie”	“electronic”
6	“female vocalists”	“pop”
7	“metal”	“female vocalists”
8	“jazz”	“metal”
9	“experimental”	“alternative rock”
10	“classic rock”	“classic rock”
...



Item distribution Similarity:

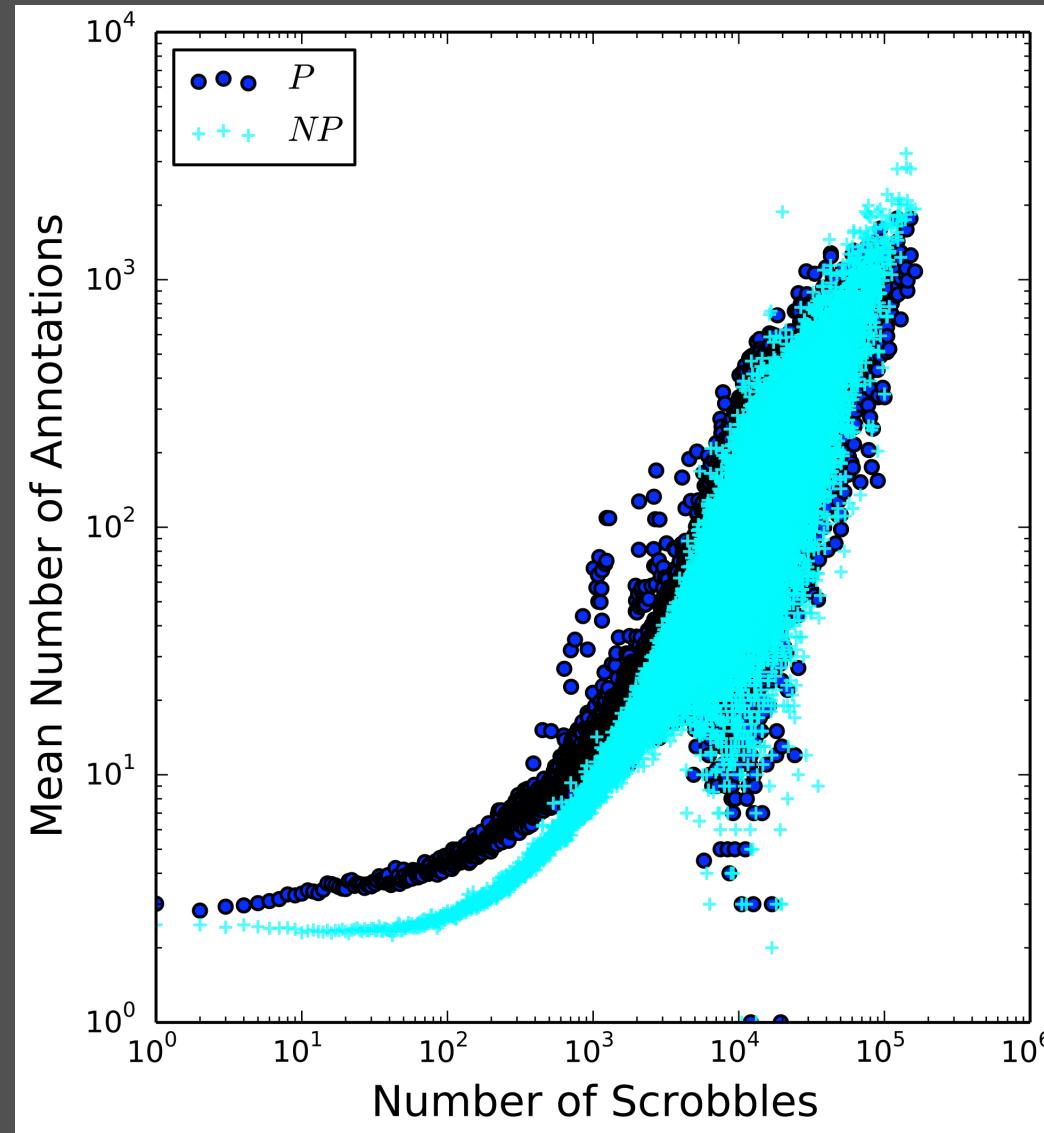
Cosine Similarity: 0.768

Rank Correlation: 0.216



Are the differences in items tagged systematic?

Are the differences in items tagged systematic?



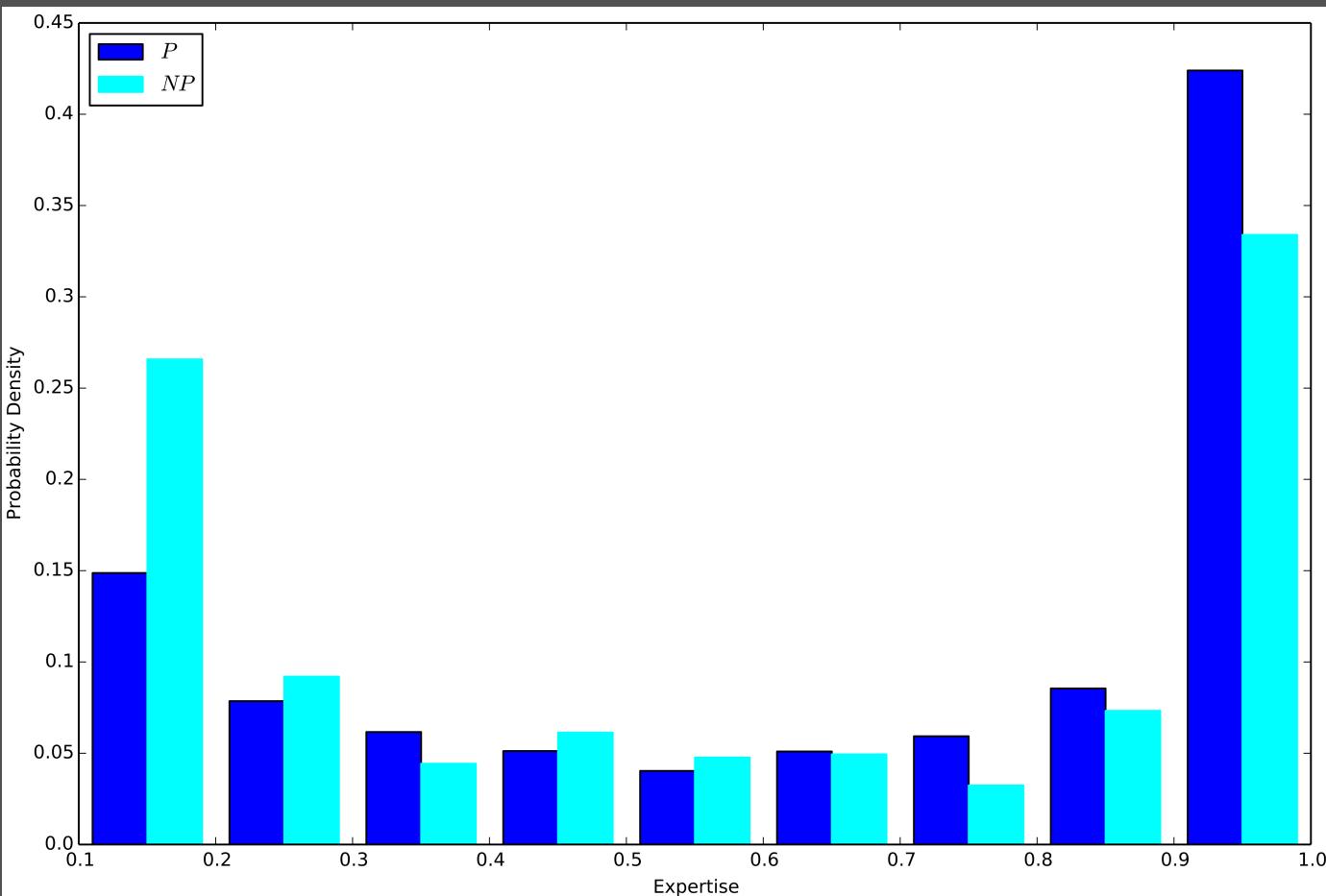
Question 3: What is driving these differences?

Question 3: What is driving these differences?

Expertise effects?
(SPEAR, Yeung, et al.)

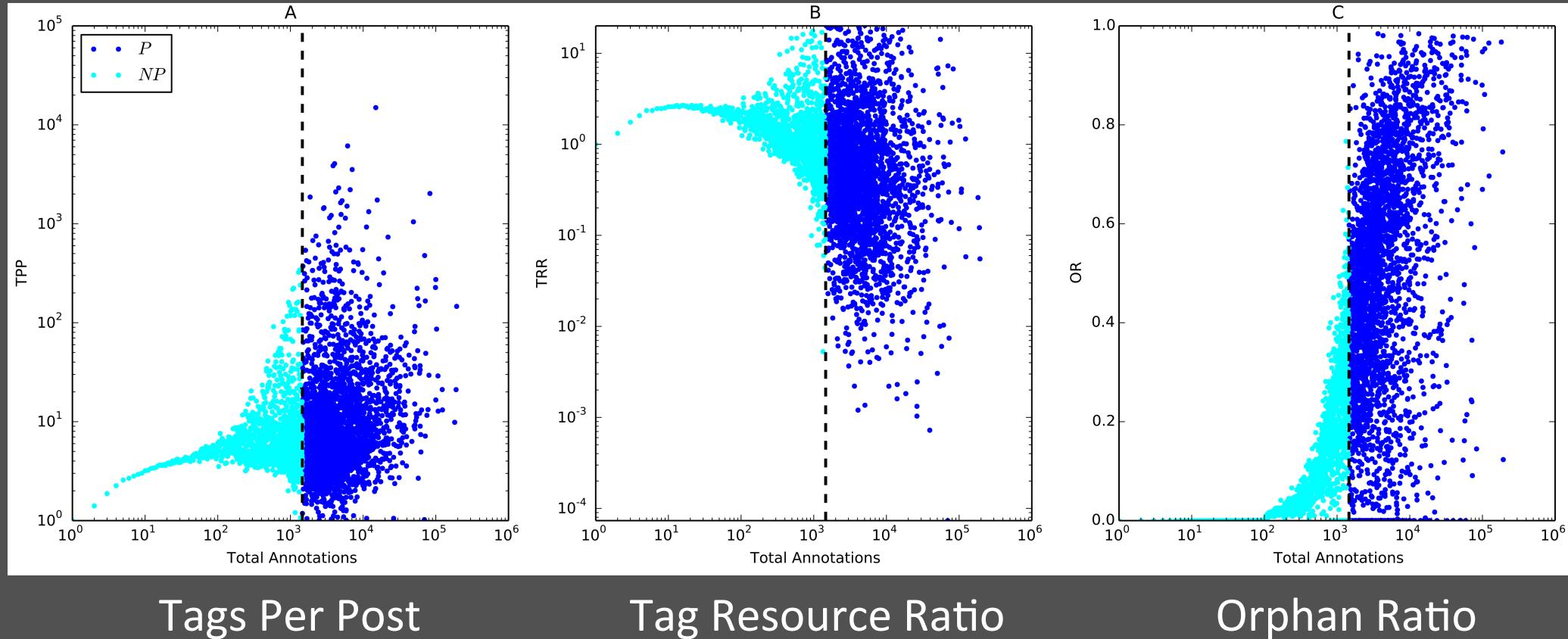
Motivational effects?
(Categorizers and Describers, Körner, et al.)

Expertise:



SPEAR (SPamming-resistant Expertise Analysis and Ranking)
Yeung, et al. *Computational Intelligence*, 27(3):458–488, 2009.

Motivation:



Categorizers vs. Describers

Körner, et al. *Proc. Hypertext 2010*, pp. 157–166, 2010.

Summary:

“Supertaggers” don’t just tag *more*, they tag *differently*.

Most notably, they disproportionately tag content in the long tail
of less popular items.

Established measures suggest they demonstrate greater
expertise and descriptive tagging habits.

Coda: What do we make of all this?

Is this division of labor “good” for tagging systems?

Identifying “supertaggers” and differentially interpreting their tags could improve the quality of tagging systems.

jlorince@indiana.edu
<https://pages.iu.edu/~jlorince>



Ψ