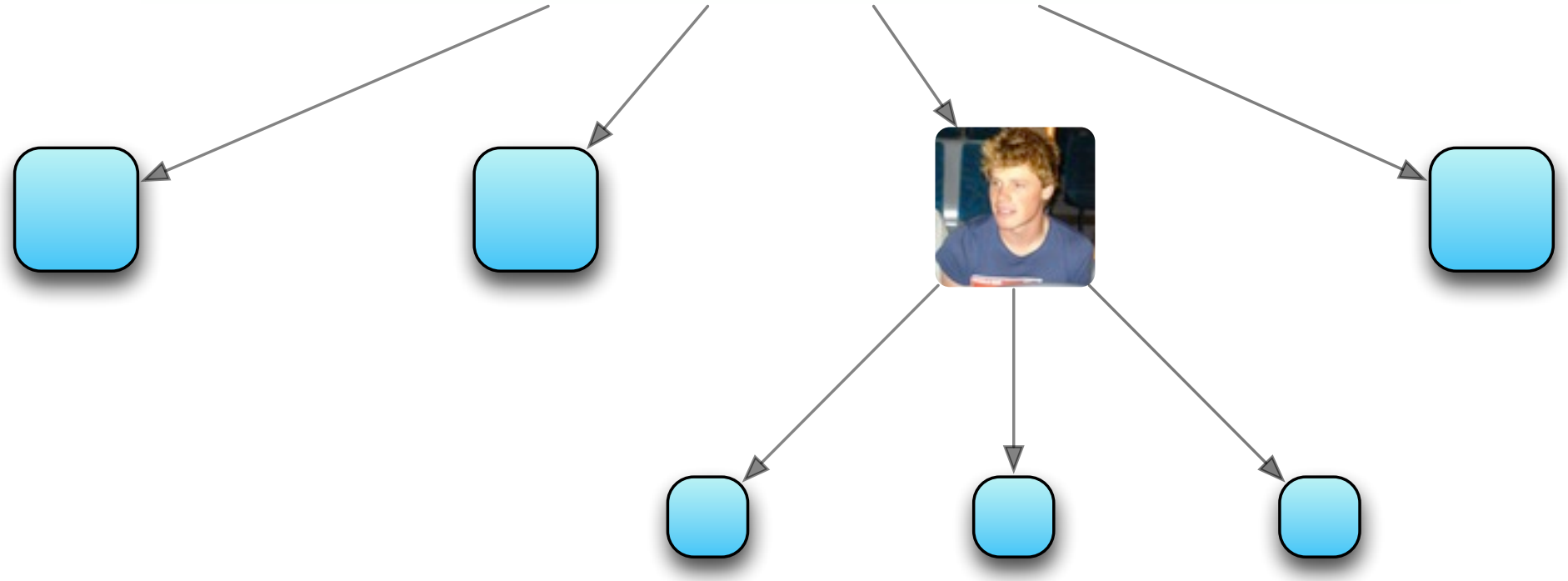
Inferring the Interesting Tweets in your Network

Will Webberley

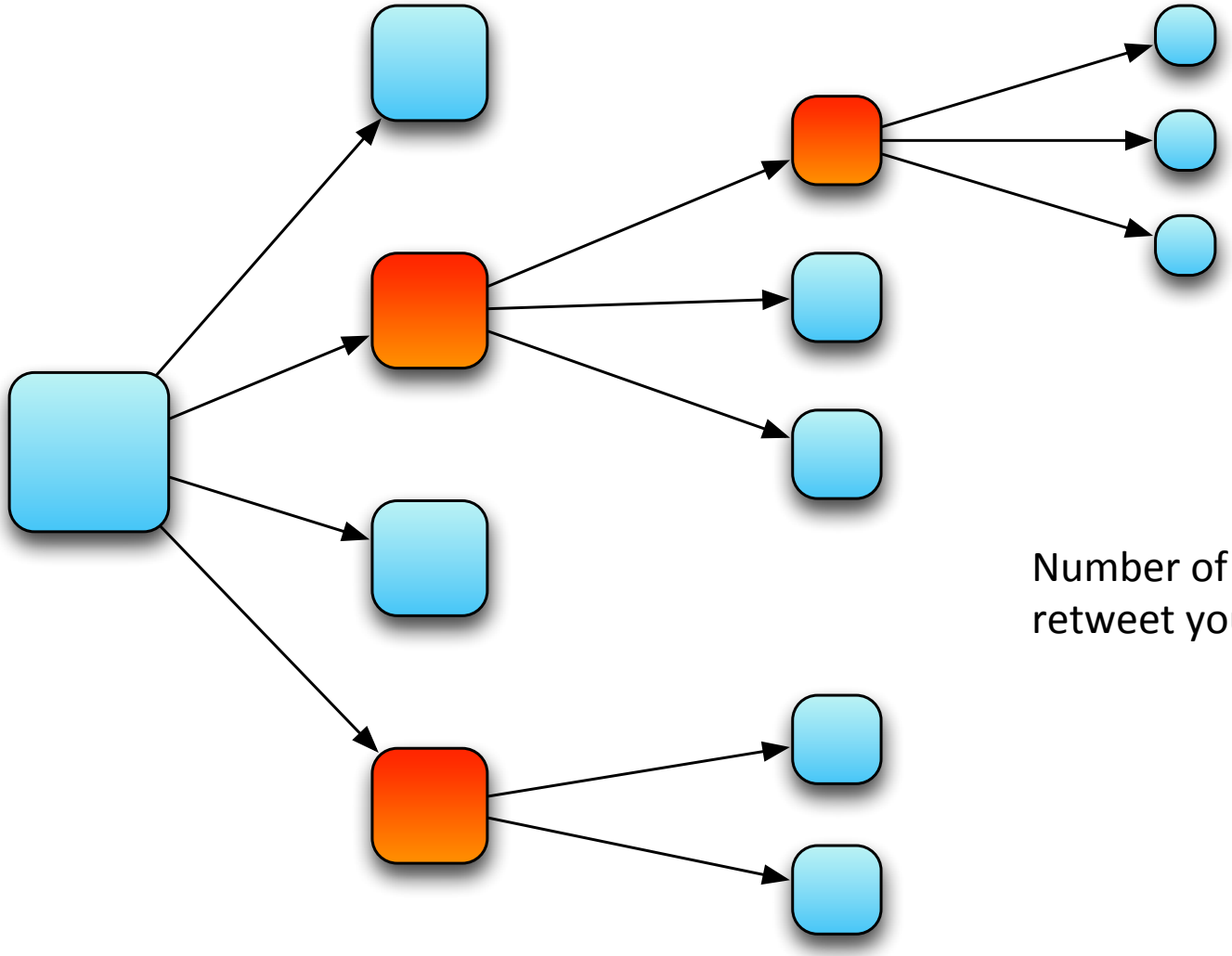
Retweeting

Karissa McKelvey @karissamck

"We didn't read half the papers we cite because they are behind a paywall." #ACM #overlyhonestmethods



‘Retweet volume’



Number of users who retweet your Tweet

Interestingness

measure of how much '**affective stimulation**'

'global' interest, enjoyment and relevance

Motivation and the problem with interestingness 

1. Interestingness ≠ retweet volume

BBC News (World) @BBCWorld 22 May

"Sweetie I'm coming - I'm gonna get you out of here" - man rescues children from school flattened by tornado.

56	23
RETWEETS	FAVORITES

Justin Bieber @justinbieber 10 Jun

all love

108,485	64,503
RETWEETS	FAVORITES

But also, Interestingness ≠ Tweet features

	Tweet 1	Tweet 2
author	Cardiff University (@cardiffuni)	Cardiff University (@cardiffuni)
contains hashtag	✔	✔
contains URL	✔	✔
contains smiley	✗	✗
Retweet volume:	238	11

Motivation - assessing interestingness

This means…

Something else portrays interestingness

So, how do we find this?

Assess on a per-Tweet & per-user basis

Interestingness scores

interestingness of a Tweet, t :

$$TS(t) = \frac{\text{observed retweet volume of } t}{\text{expected retweet volume of } t}$$

Obtaining observed/predicted volumes

Observed retweet volume, $T_O(t)$

✔

done

Returned from Twitter's REST API

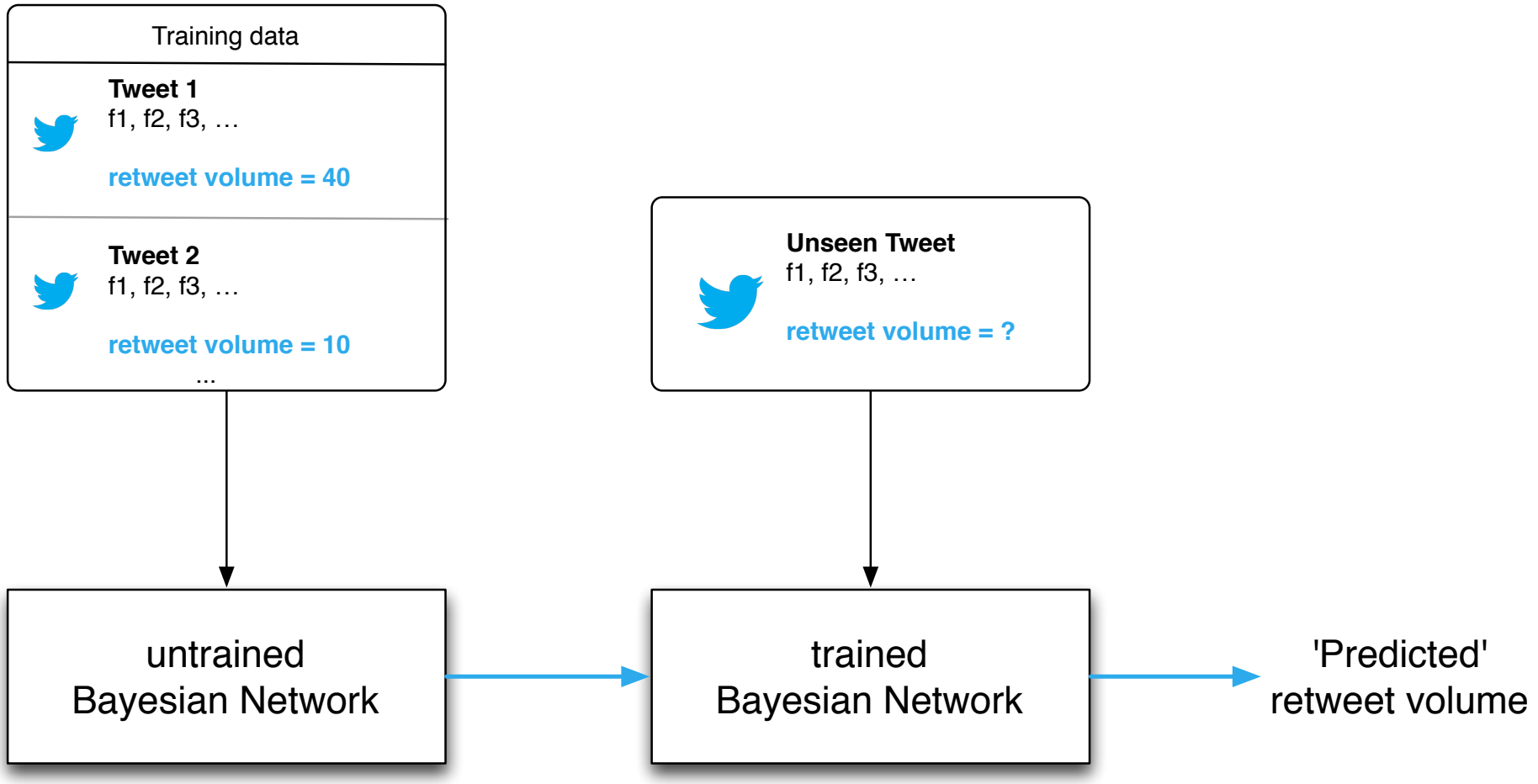
'Predicted' retweet volume, $T_P(t)$

✗

to do

?

‘Predicted’ retweet volumes

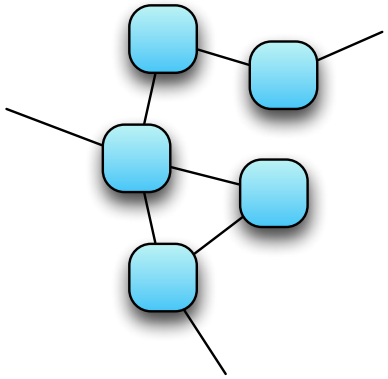


Feature overview

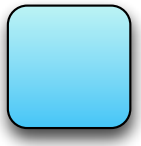
genetics	'Tweet-etics'
genome	*Tweet features*
eye colour, height, personality, etc.	text length, contains URL, hashtags, etc.
environment	*network features*
temperature, humidity, peers, etc.	author followers, community interest, etc.

Total = 31

TWO predictions for t...



‘global’ prediction



‘single-user’ prediction

.. Which gives 2 interestingness scores

$TS_G(t)$, $TS_U(t)$

- > 1 t is interesting
- ≤ 1 t is non-interesting

Are the scores valid?

Randomised controlled trial

- 750 Tweets tested

- 9 assessments in total for each tested Tweet

What did we ask?

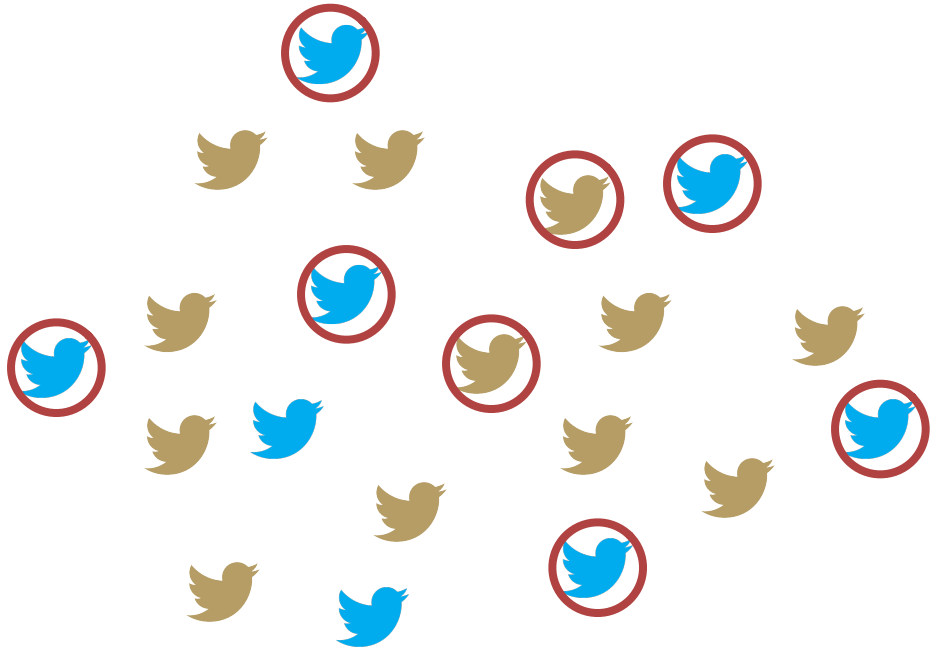
Select the Tweet(s) that you find the *most* interesting

Results summary

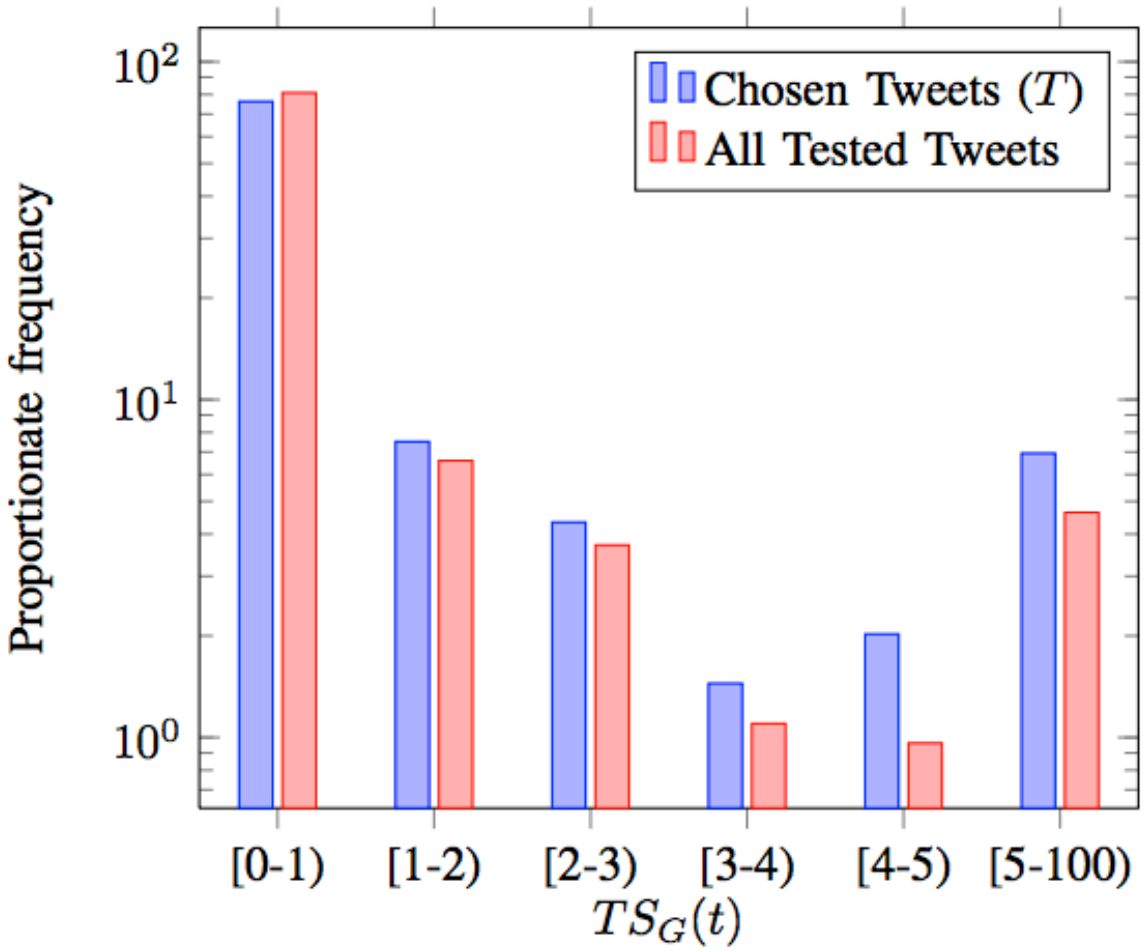
325 / 450 answers with high confidence (> 66%)

91 unique workers (diverse opinion)

65% agreement



Score distribution

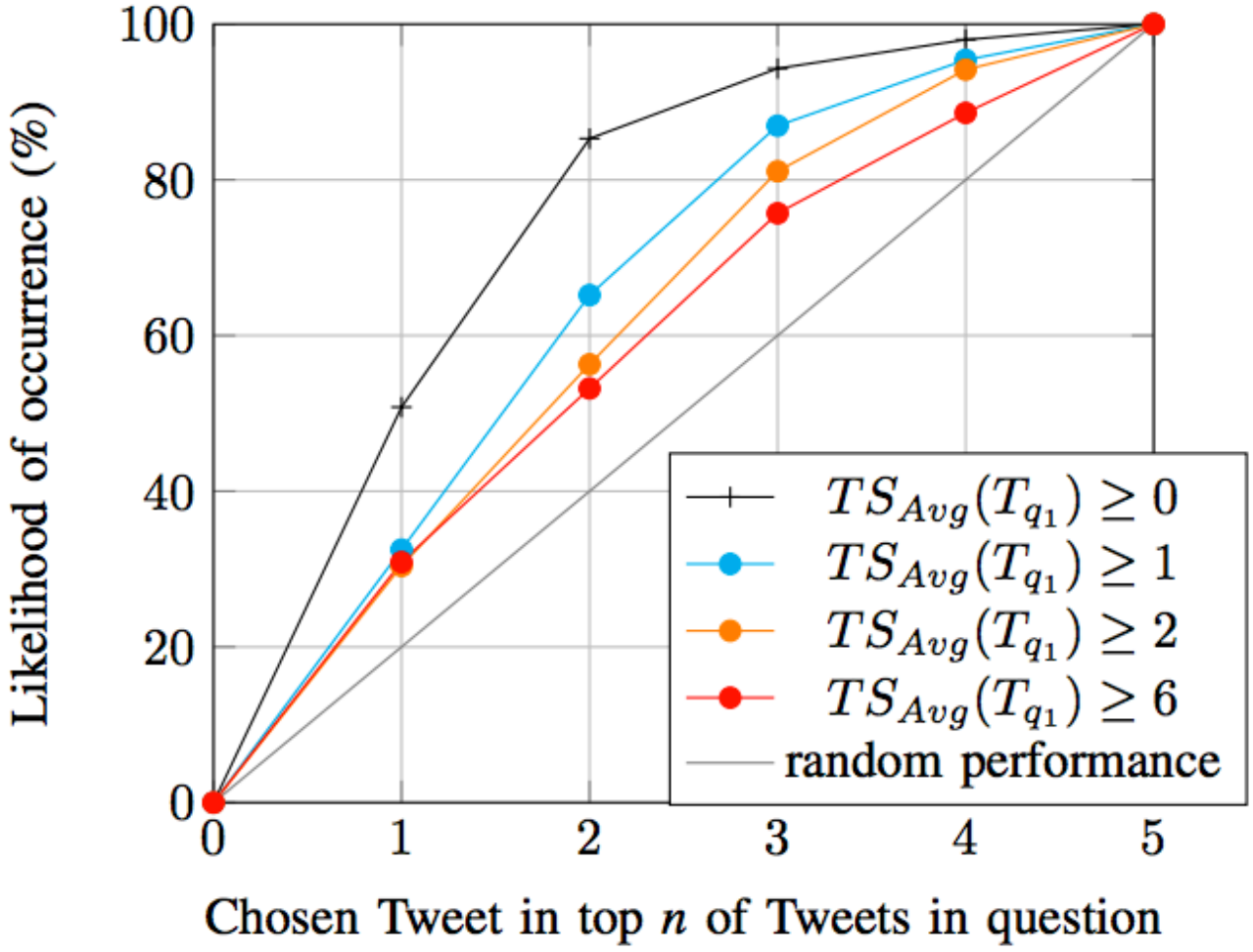


Chosen Tweets scores

>

Average scores

‘Question performance’



70% chance of picking one of the top two Tweets

Summary

Why do we need interestingness scores? ✔

How do we calculate the scores? ✔

Are the scores valid? ✔

What's next?

information relevance

Tweet 1 ✔
Tweet 2 ✗
Tweet 3 ✗
Tweet 4 ✔

Find

interesting and relevant

information *without*

- looking for it
- knowing about it