What is CS?

The design and creation of effective computing systems

- artificial intelligence
- network science
- theory & algorithms
- computational science & engineering
- graphics
- game design
- systems
- programming languages
- data-intensive computing
- human-language technologies
- security & trustworthy computing
- systems
- game design
Example: A CUCS system that *learns* important cities and landmarks therein from terabytes of raw Flickr photos & data

(algorithms, computer vision, machine learning, massive parallel computing, mobile devices, social computing, etc.)
Broader implications: sociology/social psychology

What opinions are influential?

→ proxy question: which Amazon reviews are rated helpful?

[Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee ’09]
Broader implications: sociology/social psychology

What opinions are influential?
   → proxy question: which Amazon reviews are rated helpful?
   [Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee ’09]

Prior work has focused on features of the text of the reviews, and has not been in the context of sociological inquiry. [Kim et al. ’06, Zhang and Varadarajan ’06, Ghose and Ipeirotis ’07, Jindal and B. Liu ’07, J. Liu et al ’07].

Our focus: how about non-textual features (social aspects, biases)?
Our corpus: millions of Amazon book reviews.
Some social factors boosting helpfulness scores

- using “real name”
Some social factors boosting helpfulness scores

- using “real name”
- being from New Jersey (for science books)
Some social factors boosting helpfulness scores

- using “real name”
- being from New Jersey (for science books)
- not being from Guam
Some social factors boosting helpfulness scores

- using “real name”
- being from New Jersey (for science books)
- not being from Guam

Our focus: What about the review’s star rating in relationship to others?

Theories from social psychology:

- conform (to the average rating) [Bond and Smith ’96]
- “brilliant but cruel” [Amabile ’83]
New observation: effect of variance

As variance among reviews increases, be slightly above the mean.
New observation: effect of variance

As variance among reviews increases, be \textit{slightly above} the mean

\[
\sigma^2 = 0
\]
\[
\sigma^2 = 0.5
\]
\[
\sigma^2 = 1
\]
\[
\sigma^2 = 1.5
\]
\[
\sigma^2 = 2
\]
\[
\sigma^2 = 2.5
\]
\[
\sigma^2 = 3
\]
\[
\sigma^2 = 3.5
\]
\[
\sigma^2 = 4
\]

... except in Japan, where it’s best to be \textit{slightly below}.

Example: \( \sigma^2 = 3 \):
Are the social effects just textual correlates?

We would like to control for the actual quality of a review’s text. (Maybe people from NJ inherently write better reviews about science books?)

How should we determine the ”real” helpfulness, in order to control for it?

- manual annotation? Tedious, subjective.
- automatic classification? Need extremely high accuracy guarantees.
Are the social effects just textual correlates?

We would like to control for the actual quality of a review’s text. (Maybe people from NJ inherently write better reviews about science books?)

How should we determine the ”real” helpfulness, in order to control for it?

- manual annotation? Tedious, subjective.
- automatic classification? Need extremely high accuracy guarantees.

It turns out that 1% of Amazon reviews are plagiarized! (see also David and Pinch ['06]).

Our social-effects findings regarding position relative to the mean hold on plagiarized pairs, which by definition have the same textual quality.