Opinion Spam and Analysis

Nitin Jindal and Bing Liu

Department of Computer Science

University of Illinois at Chicago

Motivation

- Opinions from reviews
  - Used by both consumers and manufacturers
  - Significant impact on product sales
- Existing Work
  - Focus on extracting and summarizing opinions from reviews
  - Little knowledge about characteristics of reviews and behavior of reviewers
  - No study on trustworthiness of opinions
  - No quality control

spam reviews

Review Spam

- Fake/untruthful review to promote or damage a product’s reputation
- Different from finding usefulness of reviews

- Increasing mention in blogosphere
- Articles in leading news media
  - CNN, NYTimess

- Increasing number of customers vary of fake reviews (biased reviews, paid reviews)

by leading PR firm Burson-Marsteller

Different from other spam types

- Web Spam (Link spam, Content spam)

  In reviews

  – not much links

  – adding irrelevant words of little help

- Email Spam (Unsolicited commercial advertisements)

  – In reviews, advertisements not as frequent as in emails

    - relatively easy to detect

Overview

- Opinion Data and Analysis
  - Reviews, reviewers and products
  - Feedbacks, ratings
- Review Spam
  - Categorization of Review Spam
  - Analysis and Detection

Amazon Data

- June 2006
  - 5.8mil reviews, 1.2mil products and 2.1mil reviewers.
- A review has 8 parts
  - *<Product ID> <Reviewer ID> <Rating> <Date> <Review Title> <Review Body> <Number of Helpful feedbacks> <Number of Feedbacks> <Number of Helpful Feedbacks>*
- Industry manufactured products "*mProducts*"

  e.g. electronics, computers, accessories, etc
  - 228K reviews, 36K products and 165K reviewers.

Log-log plot

Reviews, Reviewers and Products

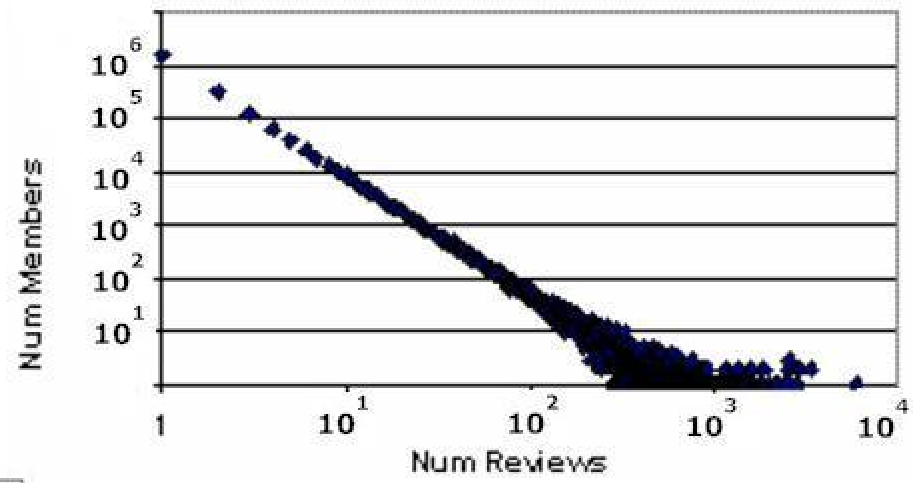


Fig. 1 reviews and reviewers

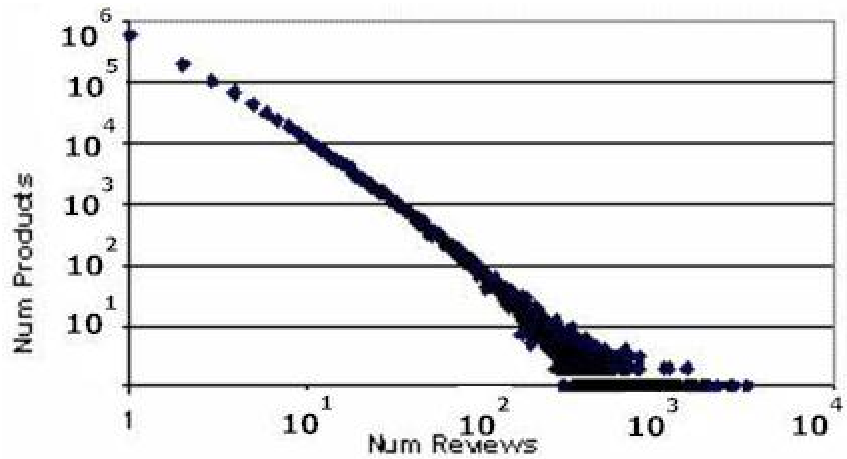


Fig. 2 reviews and products

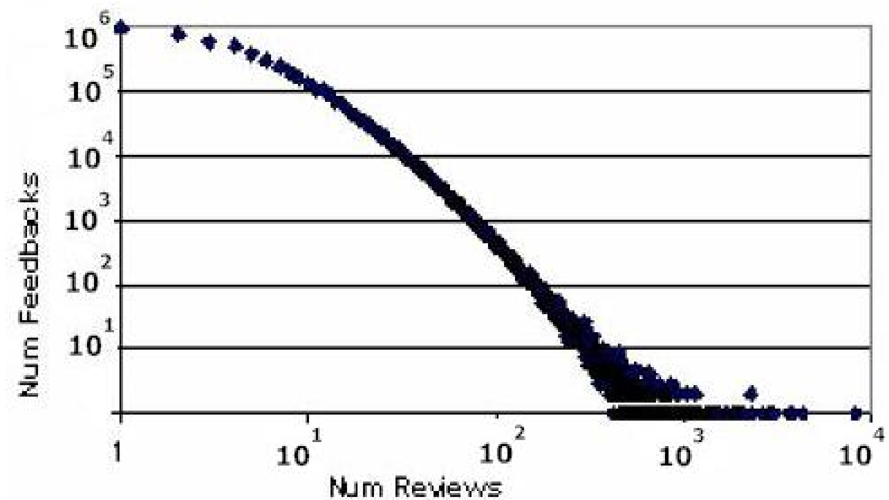


Fig. 3 reviews and feedbacks

Observations

Reviews & Reviewers

- 68% of reviewers wrote only one review
- Only 8% of the reviewers wrote at least 5 reviews

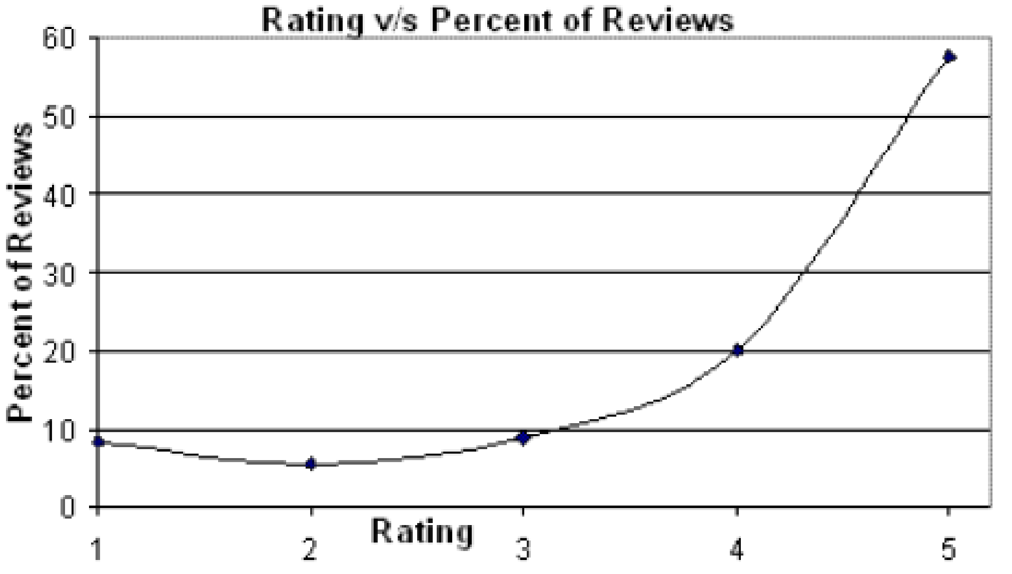
Reviews & Products

- 50% of products have only one review
- Only 19% of the products have at least 5 reviews

Reviews & Feedbacks

- Closely follows power law

Review Ratings



•Rating of 5

60% reviews

45% of products

59% of members

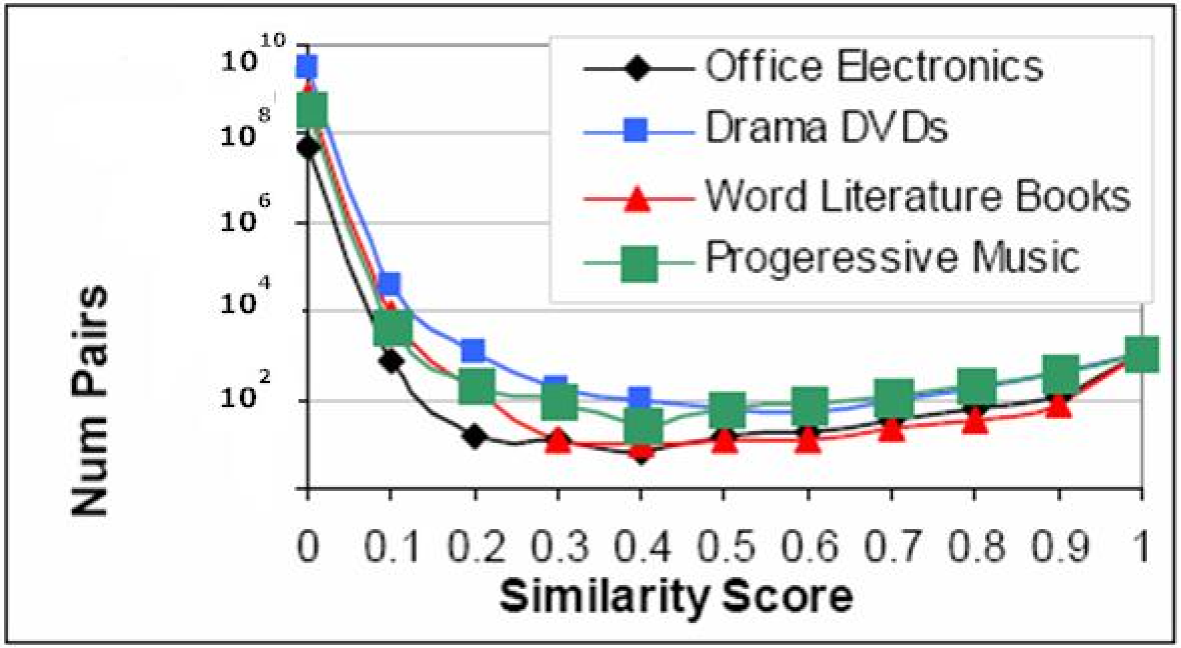
Reviews and Feedbacks

1st review – 80% positive feedbacks

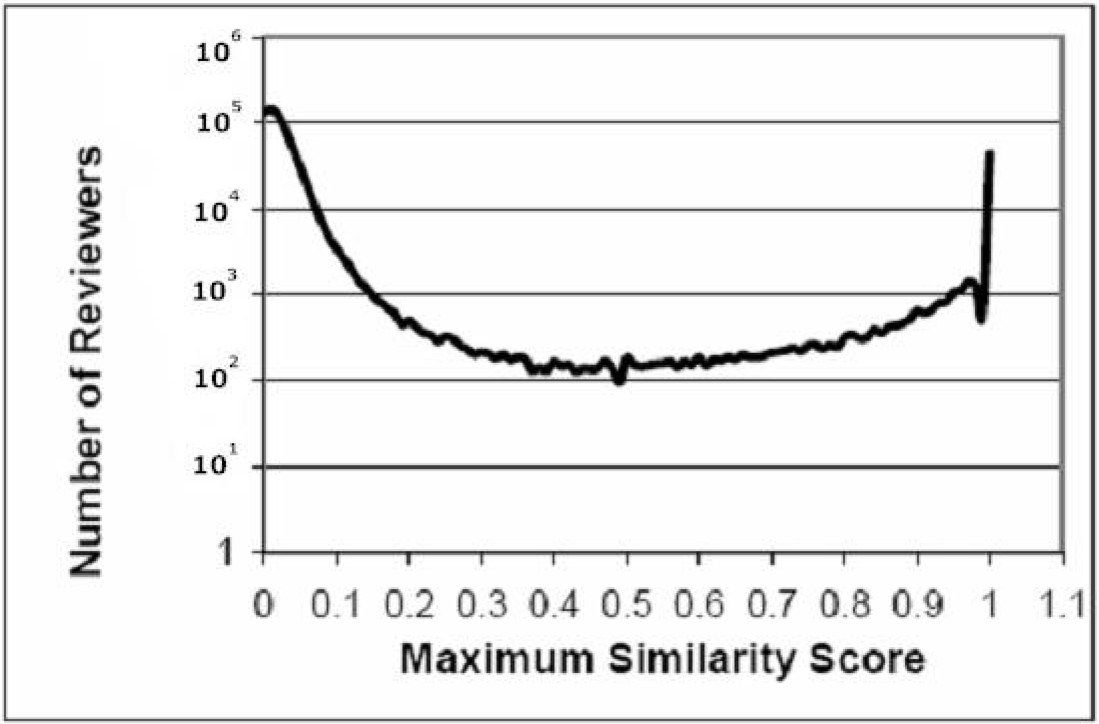
10th review – 70% positive feedbacks

Duplicate Reviews

Two reviews which have similar content are called duplicates



Members who duplicated reviews



- 10% of reviewers with more than one review (~650K) wrote duplicate reviews
- 40% of the times exact duplicates

Types of Duplicate Reviews

Type of duplicates

1. Same userid, same product
2. Different userid, same product
3. Same userid, different product
4. Different userid, different product

	Spam Review Type	Num Reviews (*mProducts*)
1	Different userids on the same product	3067 (104)
2	Same userid on different products	50869 (4270)
3	Different userids on different products	1383 (114)
	Total	55319 (4488)

Categorization of Review Spam

- Type 1 (Untruthful Opinions)

  Ex:
- Type 2 (Reviews on Brands Only)

  Ex: "*I don't trust HP and never bought anything from them*"
- Type 3 (Non-reviews)
  - Advertisements

    Ex: "*Detailed product specs: 802.11g, IMR compliant, …*"

    "*…buy this product at: compuplus.com*"
  - Other non-reviews

    Ex: "*What port is it for*"

    "*The other review is too funny*"

    "*Go Eagles go*"

Spam Detection

- Type 2 and Type 3 spam reviews
  - Supervised learning

- Type 1 spam reviews
  - Manual labeling very difficult
  - Propose to use duplicate and near-duplicate reviews

Detecting Type 2 & Type 3 Spam Reviews

- Binary classification
  - Logistic Regression
    - Probabilistic estimates
    - Practical applications, like give weights to each review, rank them, etc
- Poor performance on other models
    - naïve Bayes, SVM and Decision Trees

Features Construction

- Three types
  - Review centric, reviewer centric and product centric

- Total 32 features
  - Rating related features
    - Average rating, standard deviation, etc
  - Feedback related features
    - Percentage of positive feedbacks, total feedbacks, etc
  - Textual Features
    - Opinion words [Hu, Liu ‘04], numerals, capitals, cosine similarity, etc
  - Other features
    - Length and position of review
    - Sales rank, price, etc

Experimental Results

- Evaluation criteria
  - Area Under Curve (AUC)
  - 10-fold cross validation

Table 3. AUC values for different types of spam

Spam Type	Num reviews	AUC	AUC – text features only	AUC – w/o feedbacks
Types 2 & 3	470	98.7%	90%	98%
Type 2 only	221	98.5%	88%	98%
Type 3 only	249	99.0%	92%	98%

- High AUC -> Easy to detect
- Equally well on type 2 and type 3 spam
- text features alone not sufficient
- Feedbacks unhelpful (feedback spam)

Type 1 Spam Reviews

•Hype spam – promote one's own product
Defaming spam – defame one's competitors product

Table 4. Spam reviews vs. product quality

	Positive spam review	Negative spam review
Good quality product	1	**2**
Bad quality product	**3**	4
Average quality product	**5**	**6**

Harmful Regions

Very hard to detect manually

Predictive Power of Duplicates

- Representative of all kinds of spam
- Only 3% duplicates accidental
- Duplicates as positive examples, rest of the reviews as negative examples

Table 5. AUC values on duplicate spam reviews.

Features used	AUC
All features	78%
Only review features	75%
Only reviewer features	72.5%
Without feedback features	77%
Only text features	63%

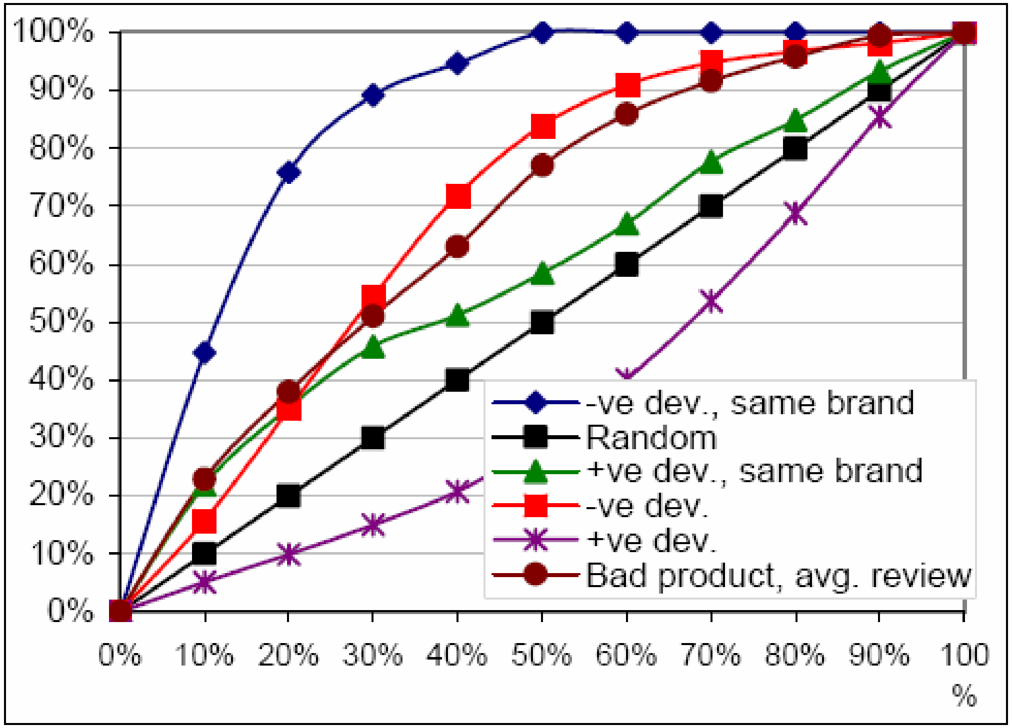
  - good predictive power
  - How to check if it can detect type 1 reviews? (outlier reviews)

Outlier Reviews

  - Reviews which deviate from average product rating
  - Necessary (but not sufficient) condition for harmful spam reviews

- Predicting outlier reviews
  - Run logistic regression model using duplicate reviews
    (without rating related features)
  - Lift curve analysis

Lift Curve for outlier reviews



Biased reviewer -> all good or bad reviews on products of a brand

- -ve deviation reviews more likely to be spams
  - Biased reviews most likely
- +ve deviation reviews least likely to be spams

except,

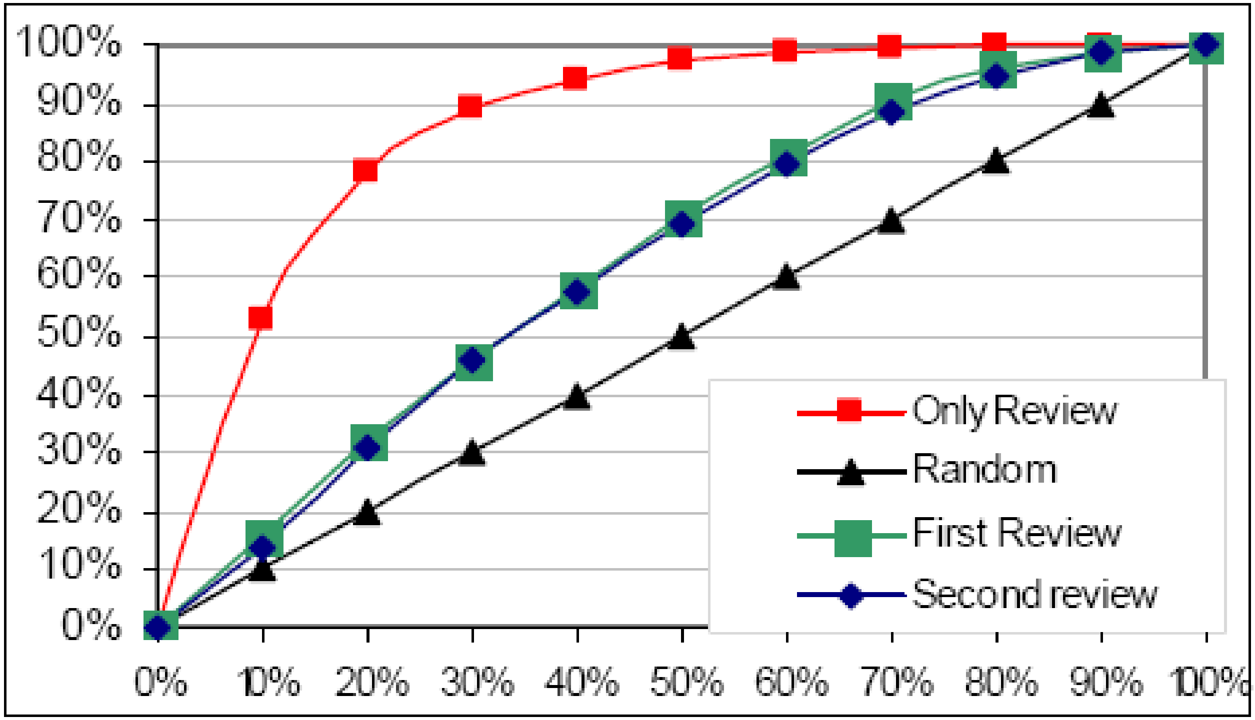
  - average reviews on bad products
  - Biased reviewers

“If model able to predicts outlier reviews, then with some degree of confidence we can say that it will predict harmful spam reviews too”

Other Interesting Outlier Reviews

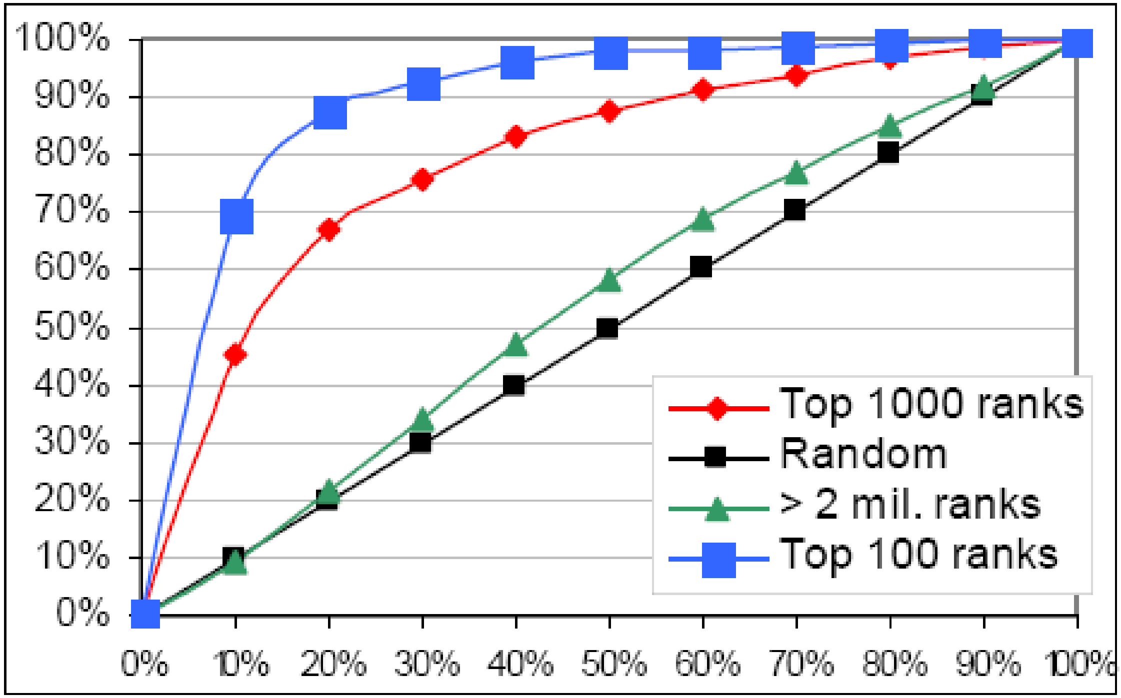
- Only reviews
- Reviews from top ranked members
- Reviews with different feedbacks
- Reviews on products with different sales ranks

Only Reviews



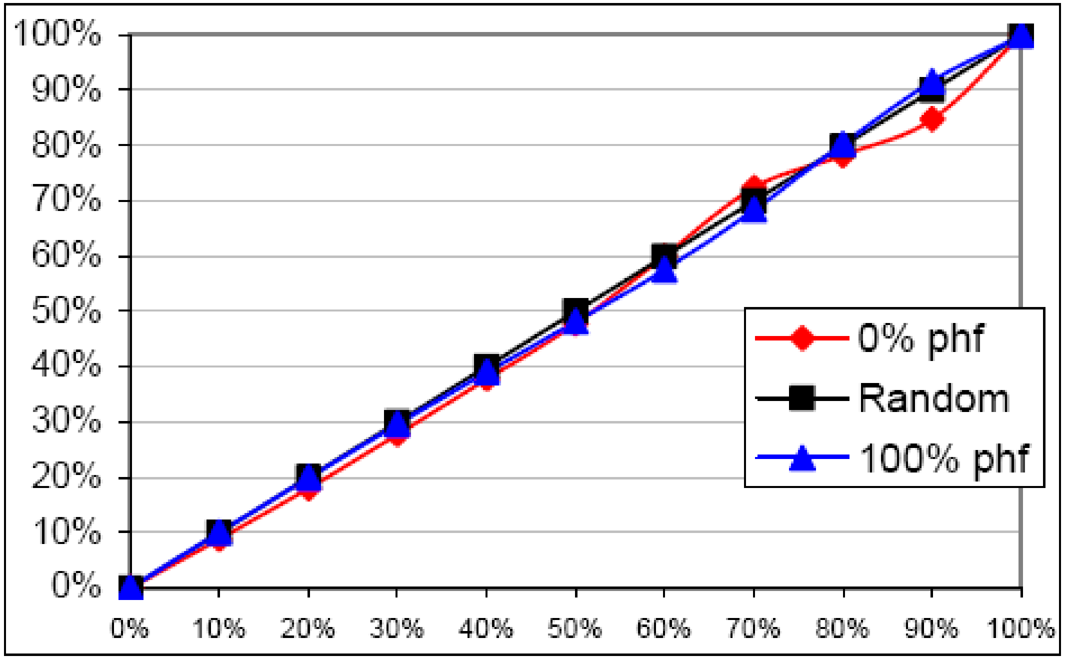
- 46% of reviewed products have only one review
- Only reviews have high lift curve

Reviews from Top-Ranked Reviewers



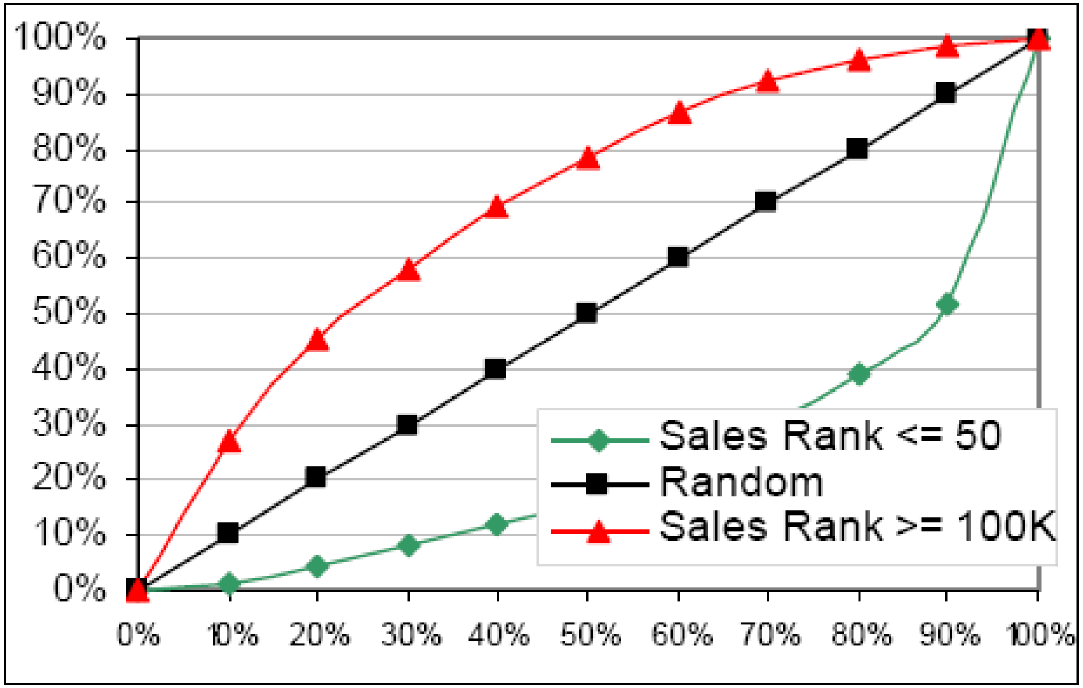
- Reviews by top ranked reviewers given higher probabilities of spam
  - Top ranked members write larger number reviews
  - Deviate a lot from product rating, write a lot of only reviews

Reviews with different levels of feedbacks



- Random distribution
  - Spam reviews can get good feedbacks

Reviews of products with varied sales ranks



- Product sales rank
  - Important feature
- High sales rank – low levels of spam
- Spam activities linked to low selling products

Conclusions

- Review Spam and Detection
- Categorization into three types
- Type 2 and 3 easy to detect
- Type 1 difficult to label manually
  - Proposed to use duplicate reviews for detecting type 1 spam
  - Predictive power on outlier reviews
  - Analyze other interesting outlier reviews

Questions?