Yelp Review Service

Derek Nelson & Brian Oh

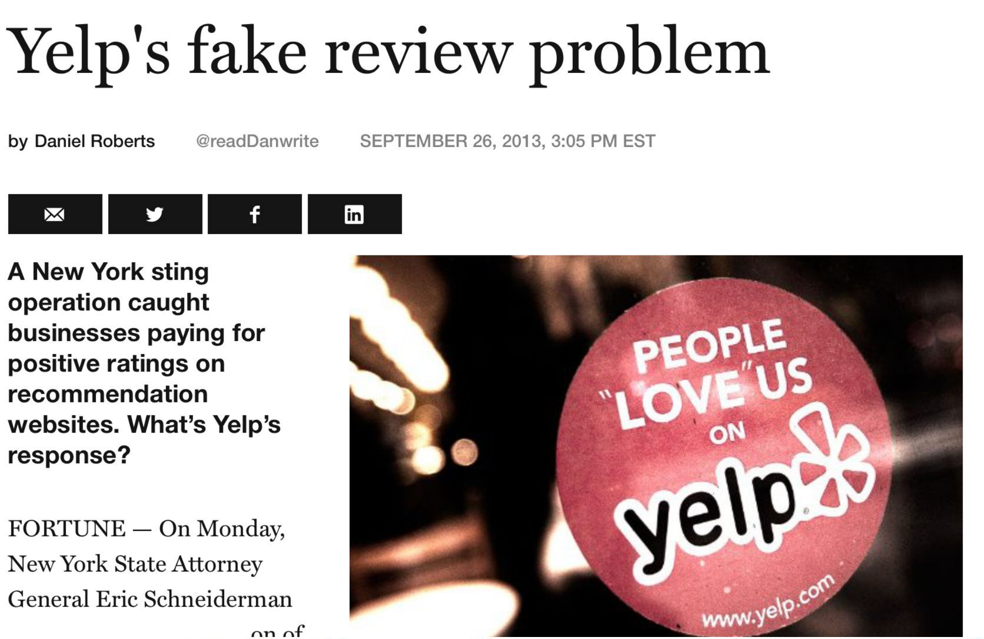
Introduction

- Yelp is one of the biggest review website in US
- They provide valuable local business info.
- As of 2014, Yelp has 135 million monthly visitors and 71 million reviews.

Problem...

Its slogan used to be real people, real reviews.

Now, not so much. About 25 percent of submitted reviews are suspicious or not recommended.



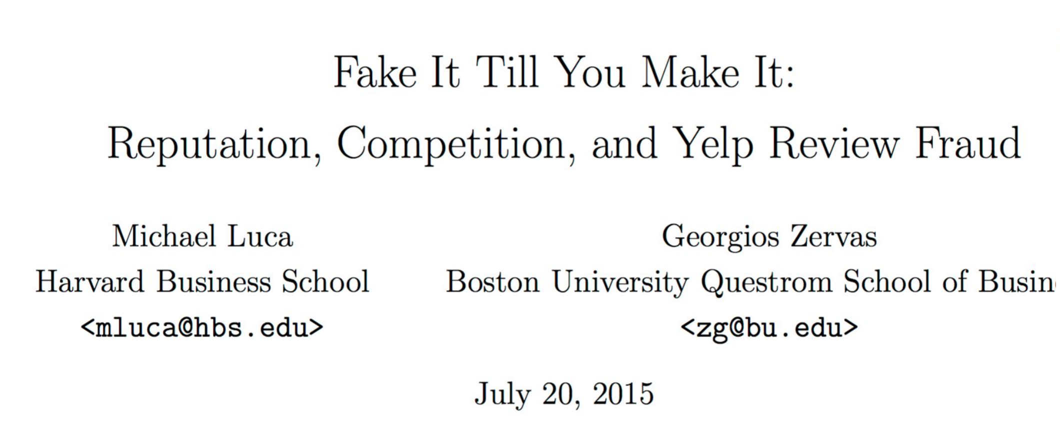
Yelp's fake review problem

by Daniel Roberts @readDanwrite SEPTEMBER 26, 2013, 3:05 PM EST

A New York sting operation caught businesses paying for positive ratings on recommendation websites. What's Yelp's response?

FORTUNE — On Monday, New York State Attorney General Eric Schneiderman

on of



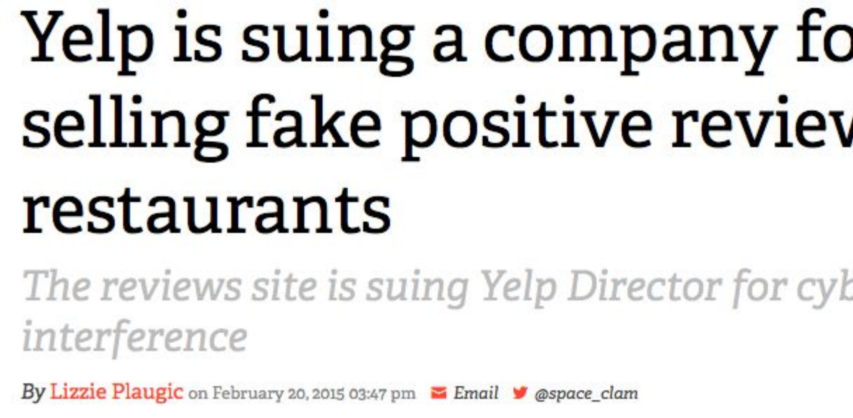
Fake It Till You Make It:

Reputation, Competition, and Yelp Review Fraud

Michael Luca
Harvard Business School
<mluca@hbs.edu>

Georgios Zervas
Boston University Questrom School of Busin
<zg@bu.edu>

July 20, 2015

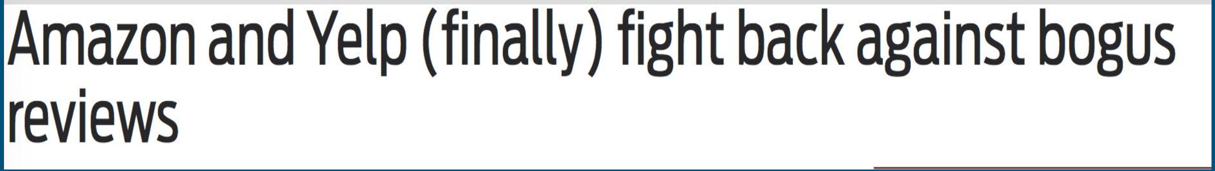


Yelp is suing a company fo selling fake positive reviev restaurants

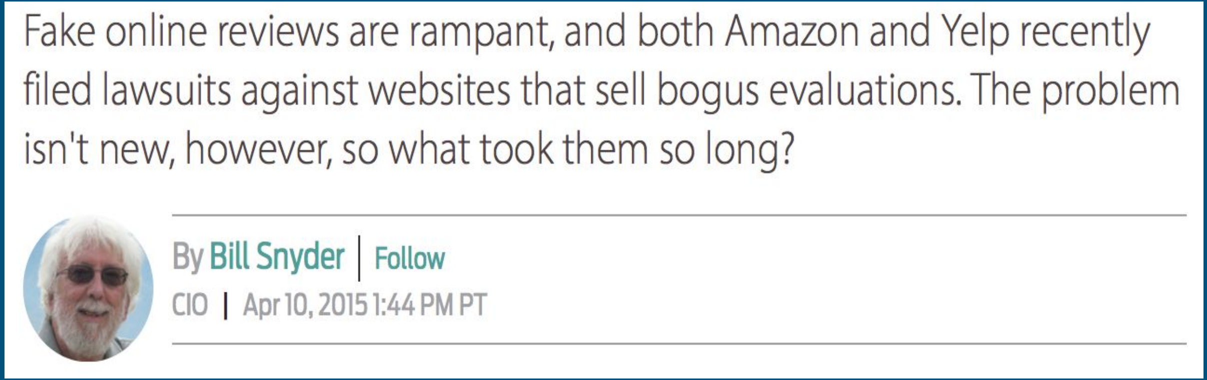
The reviews site is suing Yelp Director for cyb interference

By Lizzie Plaugic on February 20, 2015 03:47 pm ✉ Email @space_clam

Anything We Can Do?



Amazon and Yelp (finally) fight back against bogus reviews



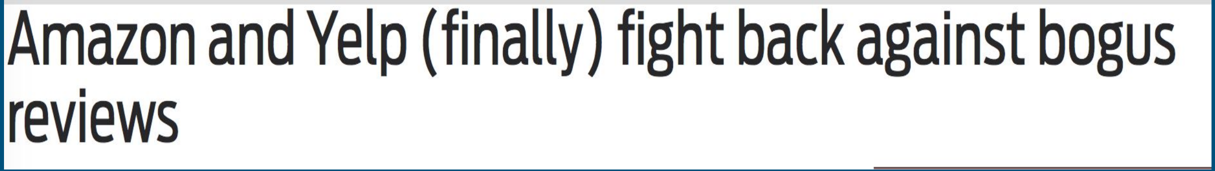
Fake online reviews are rampant, and both Amazon and Yelp recently filed lawsuits against websites that sell bogus evaluations. The problem isn't new, however, so what took them so long?

By Bill Snyder | Follow

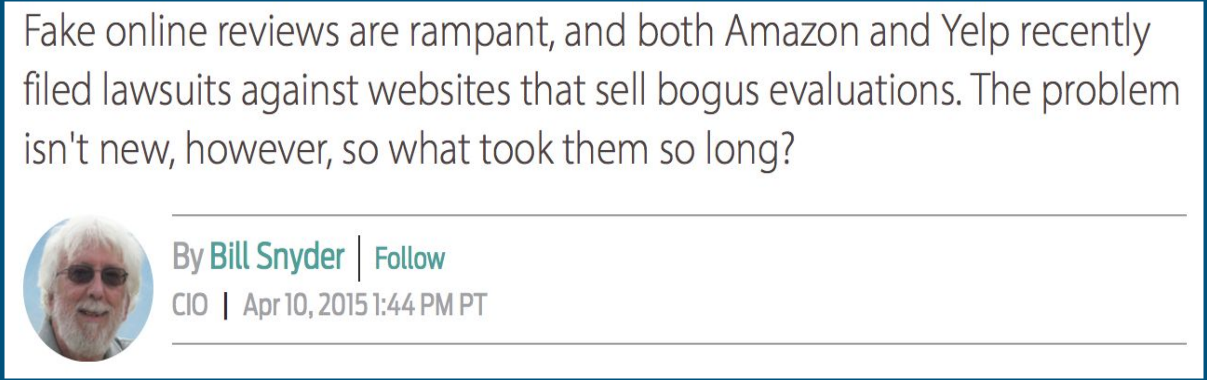
CIO | Apr 10, 2015 1:44 PM PT

Anything We Can Do?

- Lawsuit
- Human filtering
- Filter out using NLP
- Reviewer Rank



Amazon and Yelp (finally) fight back against bogus reviews



Fake online reviews are rampant, and both Amazon and Yelp recently filed lawsuits against websites that sell bogus evaluations. The problem isn't new, however, so what took them so long?

By Bill Snyder | Follow

CIO | Apr 10, 2015 1:44 PM PT

Reviewer Rank

$$\begin{array}{cc} & \text{User} \\ \text{User} & \begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 1 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \end{array}$$

$$\mathbf{v}' = \beta M \mathbf{v} + (1-\beta)\mathbf{e}/n$$

Modification to Each Review Score

Review Score: range from 1 to 5 stars

(Review Score - 3): range from -2 to 2

Modified Review Score = (Review Score - 3) x Reviewer Rank

Ex) Papa John's got 1 star from high rank reviewer: (1 - 3) x (0.02) = - 0.04

Olive Garden got 5 stars from row rank reviewer: (5 - 3) x (0.000013) = + 0.000026

So, Which Restaurant is the Best in My Town?

Business Score = ∑ modified score / number of reviews

1. Filter to Restaurants in my City
2. Sort them by Business score

Concerns...

1. Is User to User “Useful vote” data available?

2. Isn’t the matrix too sparse?

Concerns...

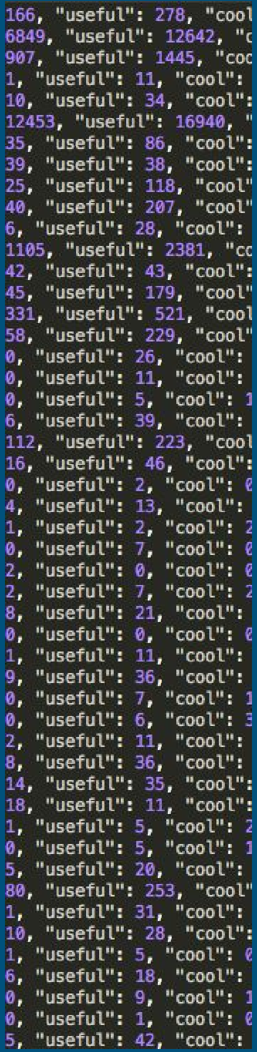
1. **Is User to User “Useful vote” data available?**

   Not available, but number of Useful votes that each user got is available.

   We generated the Transition Matrix according to the data

2. **Isn’t the matrix too sparse?**

   Approximately, 90% of user has multiple “Useful vote”



```
166, "useful": 278, "cool
6849, "useful": 12642, "c
907, "useful": 1445, "coo
1, "useful": 11, "cool":
10, "useful": 34, "cool":
12453, "useful": 16940, "
35, "useful": 86, "cool":
39, "useful": 38, "cool":
25, "useful": 118, "cool"
40, "useful": 207, "cool"
6, "useful": 28, "cool":
1105, "useful": 2381, "co
42, "useful": 43, "cool":
45, "useful": 179, "cool"
331, "useful": 521, "cool
58, "useful": 229, "cool"
0, "useful": 26, "cool":
0, "useful": 11, "cool":
0, "useful": 5, "cool": 1
6, "useful": 39, "cool":
112, "useful": 223, "cool
16, "useful": 46, "cool":
0, "useful": 2, "cool": 0
4, "useful": 13, "cool":
1, "useful": 2, "cool": 2
0, "useful": 7, "cool": 0
2, "useful": 0, "cool": 0
2, "useful": 7, "cool": 2
8, "useful": 21, "cool":
0, "useful": 0, "cool": 0
1, "useful": 11, "cool":
9, "useful": 36, "cool":
0, "useful": 7, "cool": 1
0, "useful": 6, "cool": 3
2, "useful": 11, "cool":
8, "useful": 36, "cool":
14, "useful": 35, "cool":
18, "useful": 11, "cool":
1, "useful": 5, "cool": 2
0, "useful": 5, "cool": 1
5, "useful": 20, "cool":
80, "useful": 253, "cool"
1, "useful": 31, "cool":
10, "useful": 28, "cool":
1, "useful": 5, "cool": 0
6, "useful": 18, "cool":
0, "useful": 9, "cool": 1
0, "useful": 1, "cool": 0
5, "useful": 42, "cool":
```

Our Goals...

How effective is the Reviewer Rank system?

1. Simulate writing fake reviews, and analyze impact on Business Ranking.

2. Compare to other Rank Model.

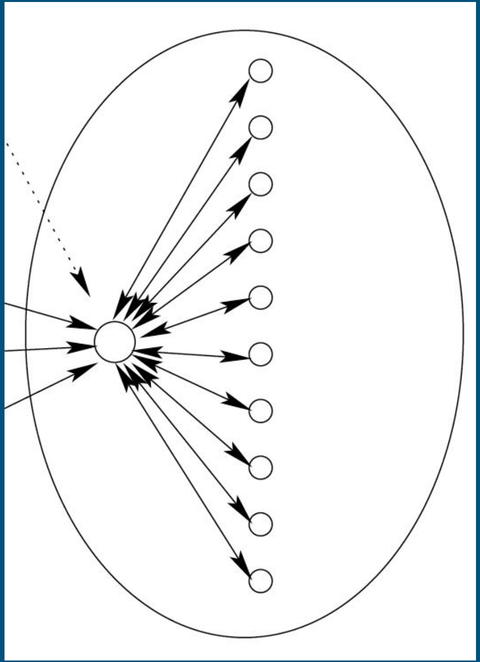
Yelp Data

- In .json format
- About 2 GB
- The data needed to be simplified
- Data we used
    - Business Data: (Business ID, Business Name, Categories, City, State, Stars, Review Count)
    - Review Data: (Review Date, Review Score, User ID, Business ID)
    - User Data: (User ID, Number of Reviews, Number of Helpful)

- We obtained Yelp data from http://www.yelp.com/dataset_challenge.

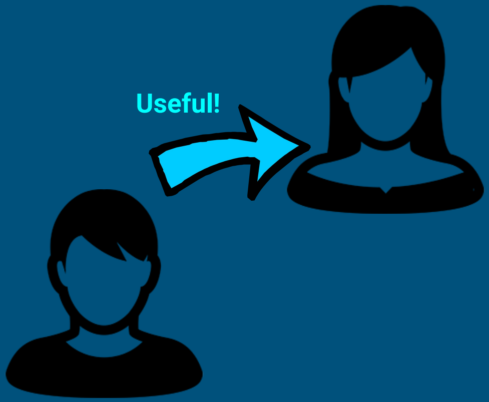
Simulation of Fake Reviews (Modifications to Data)

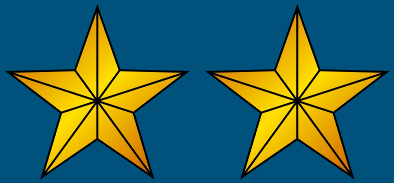
- 100 Accounts generated
- Each Account wrote 5 star review to the chosen restaurant
- Two different cases:
  - The fake reviewer does not use “useful vote” on their account
  - The fake reviewer does utilize spam farm architecture



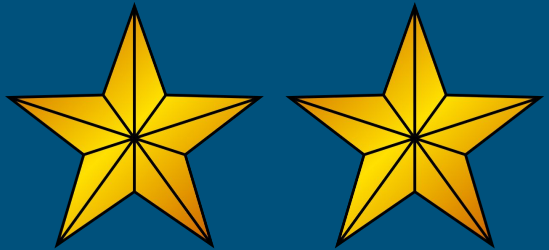
Methods

- Three methods
  - Average of the Stars
  - Having more useful votes gives more weight
  - Reviewer Rank System





Averaging

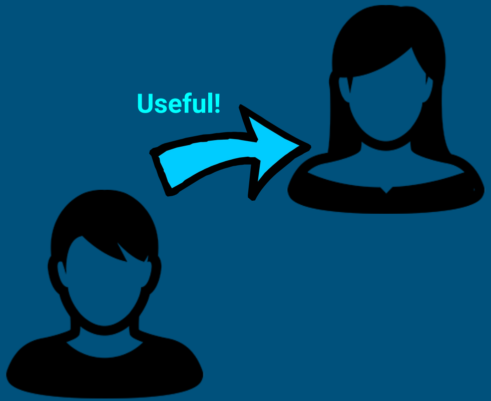


$$ReviewScore = Star\ Count\ for\ Review$$

$$RC = Review\ Count\ for\ Business = \#\ of\ Reviews$$

$$Rank = \frac{\Sigma(ReviewScore)}{RC}$$

Useful Weight



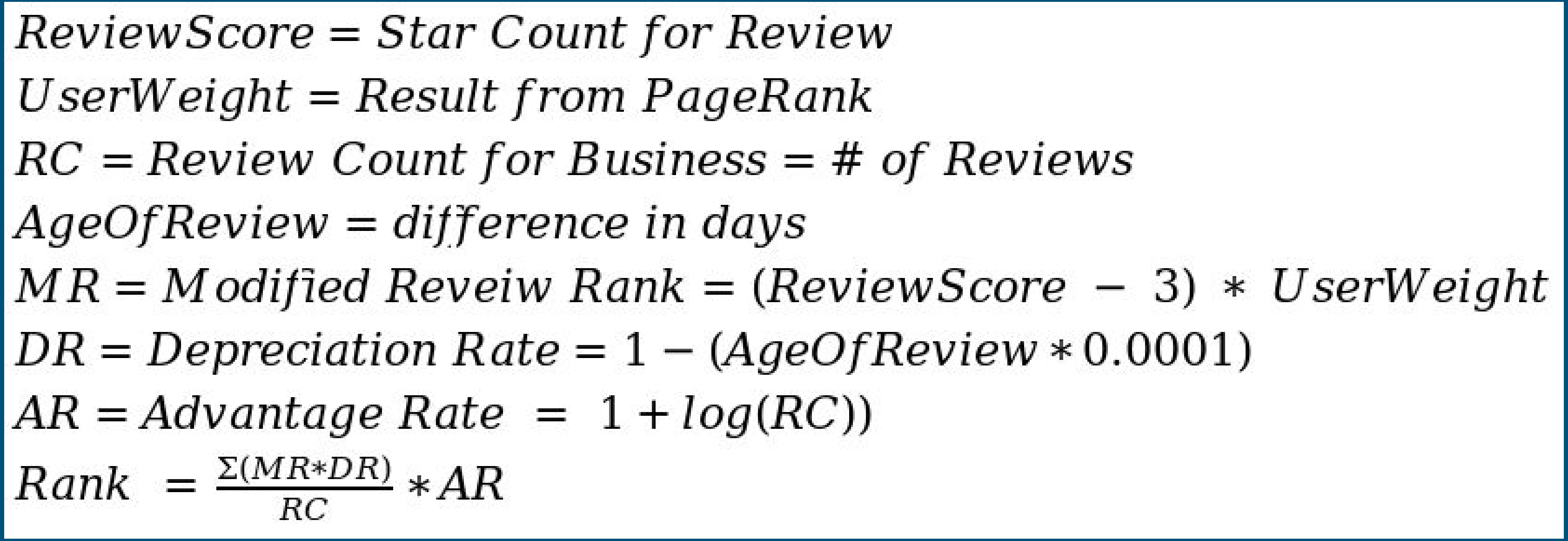
$$ReviewScore = Star\ Count\ for\ Review$$

$$UserWeight = \frac{User\ Useful\ Votes}{Total\ Sum\ of\ Useful\ Votes}$$

$$RC = Review\ Count\ for\ Business = \#\ of\ Reviews$$

$$Rank = \frac{\Sigma((ReviewScore - 3)UserWeight)}{RC}$$

Reviewer Rank



$$ReviewScore = Star\ Count\ for\ Review$$

$$UserWeight = Result\ from\ PageRank$$

$$RC = Review\ Count\ for\ Business = \#\ of\ Reviews$$

$$AgeOfReview = difference\ in\ days$$

$$MR = Modified\ Reveiw\ Rank = (ReviewScore - 3) * UserWeight$$

$$DR = Depreciation\ Rate = 1 - (AgeOfReview * 0.0001)$$

$$AR = Advantage\ Rate = 1 + log(RC))$$

$$Rank = \frac{\Sigma(MR*DR)}{RC} * AR$$

Results

- Focus on one business
    - Business ID: -sV52FN-D-I808tyRPEvwg
    - Business Name: Papa John's Pizza
    - Category: Restaurants
    - Stars: 1.0
    - City: Las Vegas
    - State: NV
    - Number of Reviews: 19

Results of Papa John’s

	Score/Rank	Modified	User Weight of 1	Calculated User Weight
Averaging	4081	✗	--	--
Useful Weight	3700	✗	--	--
Reviewer Rank	4117	✗	✓	✗
Reviewer Rank	46	✓	✓	✗
Reviewer Rank	1196	✓	✗	✓
Reviewer Rank	3763	✗	✗	✓