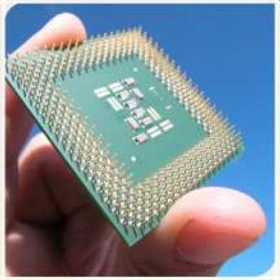


A Few Bad Votes Too Many?
Towards Robust Ranking in Social Media

Jiang Bian	Georgia Tech
Yandong Liu	Emory University
Eugene Agichtein	Emory University
Hongyuan Zha	Georgia Tech

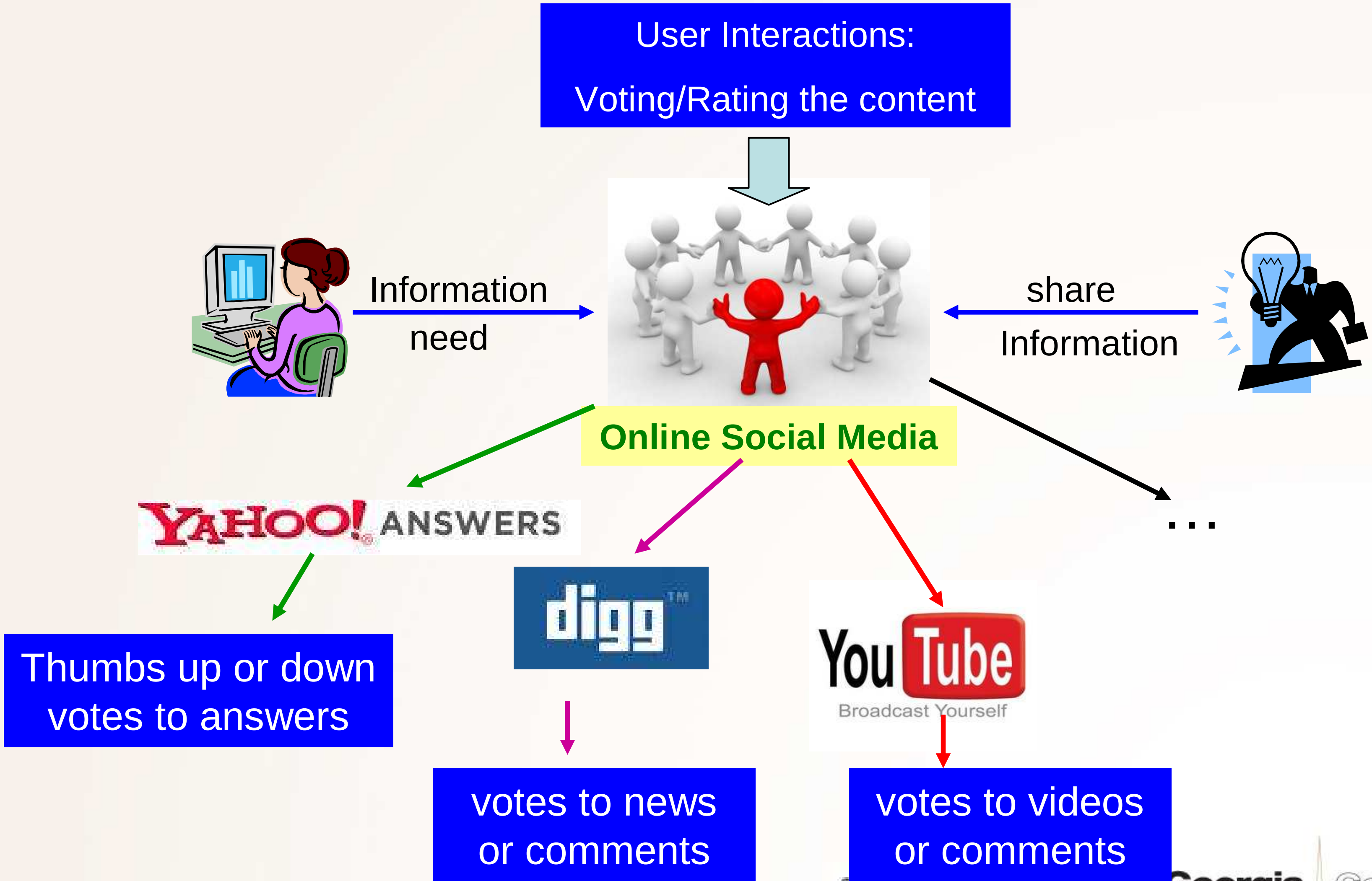
EMORY

Georgia Tech College of Computing
School of Computer Science

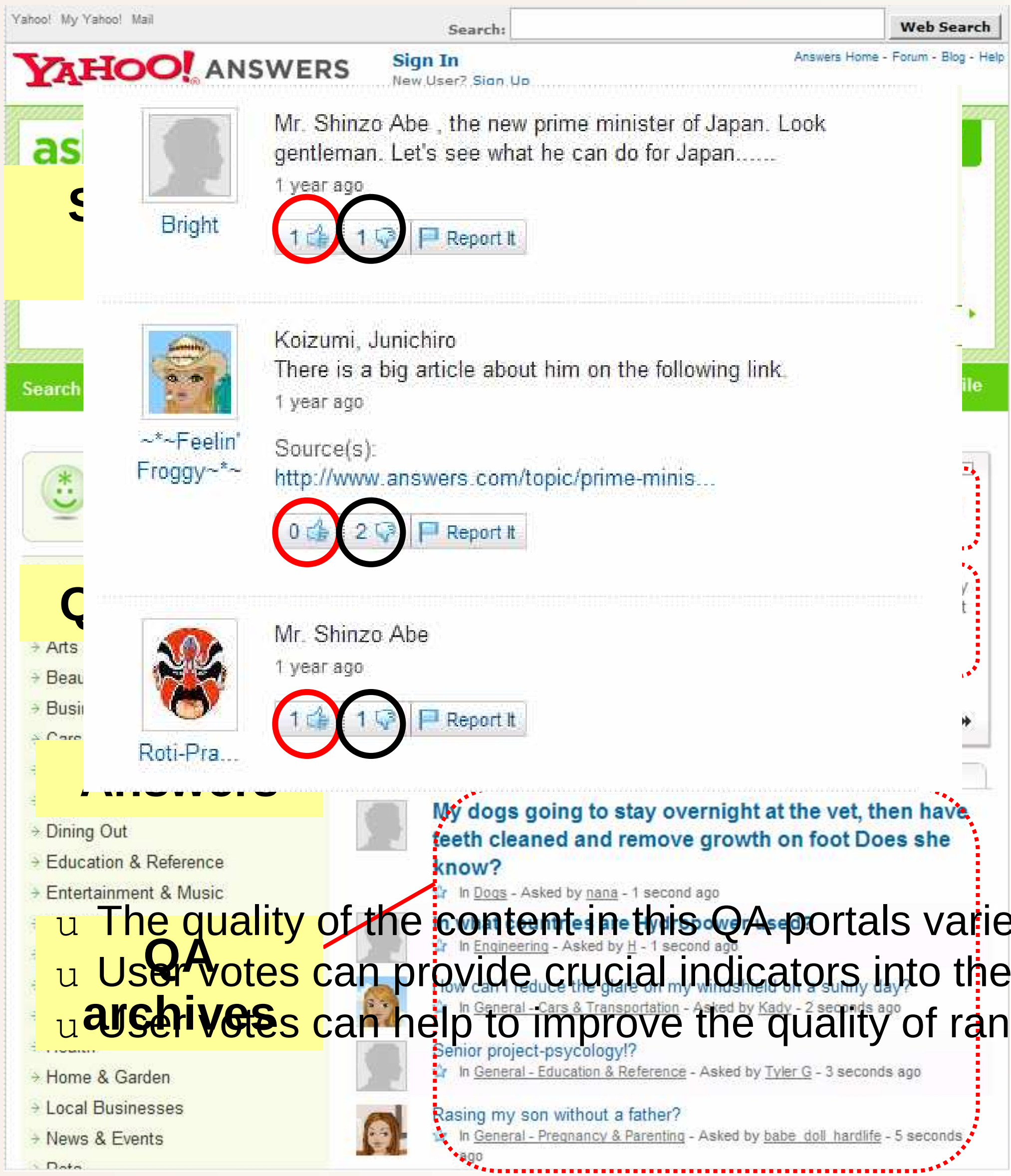
Outline

- Background and Motivation
- Learning Ranking Functions in Social Media
- Vote Spam in Social Media
- Experiments on Community Question Answering

Online Social Media



Community Question Answering (CQA)



- users can express specific information needs by posting questions, and get direct responses authored by other web users.
- Both questions and answers are stored for future use
- Allow searchers to attempt to locate an answer to their question
- Existing answers can be voted on by any users who wants to share her evaluations of the answers

- The quality of the content in this QA portals varies drastically [Agichtein et al. 2008]
- User votes can provide crucial indicators into the quality and reliability of the content
- User votes can help to improve the quality of ranking CQA content [Bian et al. 2008]

EMORY Georgia Tech College of Computing School of Computer Science

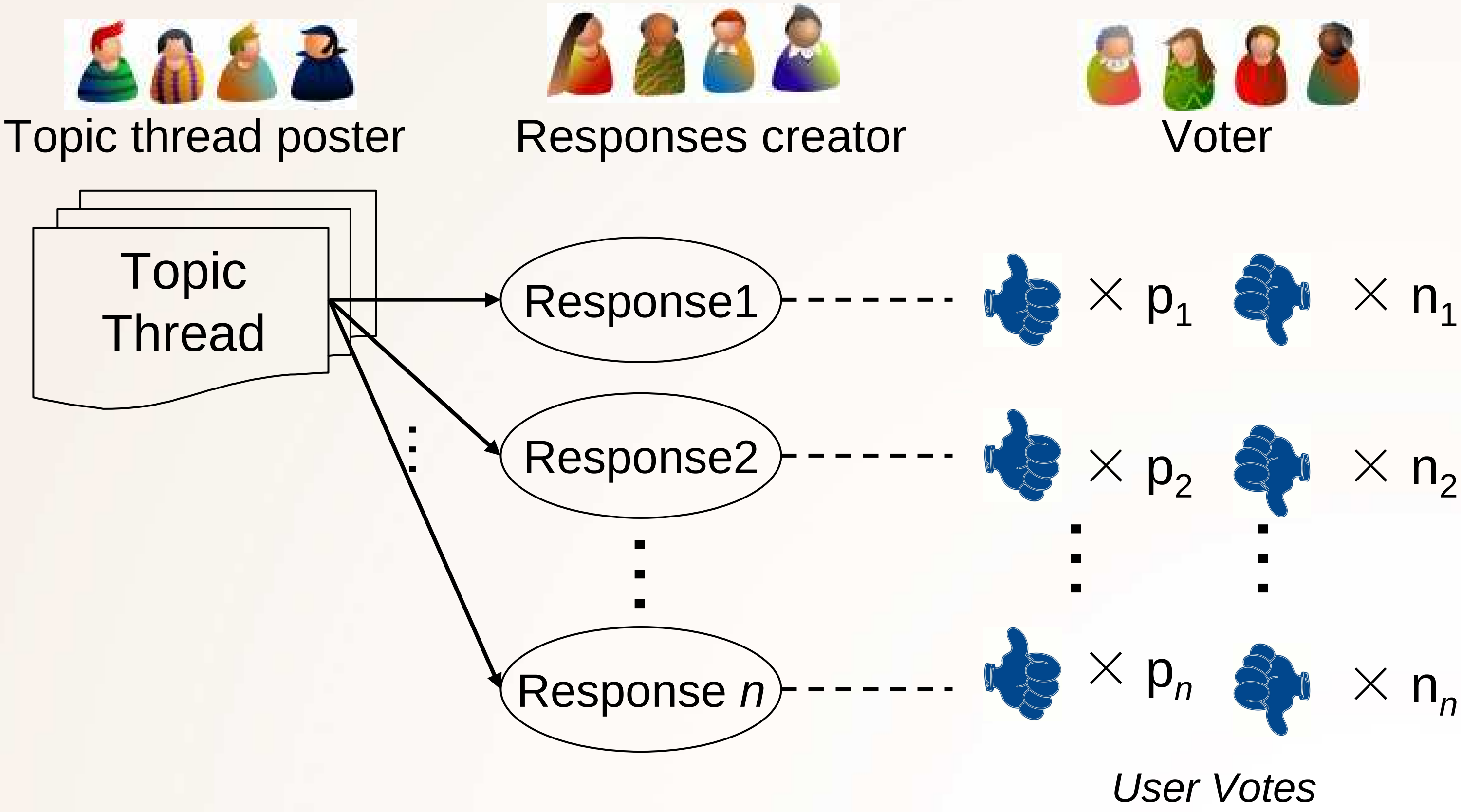
Vote Spam

- Not all user votes are reliable
  - Many “thumbs up” or “thumbs down” votes are generated without much thought
  - In some cases, users intend to game the system by promoting specific answers for fun or profit
  - We refer those bad or fraudulent votes as ***vote spam***
- How to handle vote spam for robust ranking of social media content?
  - Yahoo! Team semi-automatically removes some of more obvious vote spam after the fact
  - It is not adequate
    - The amount and the patterns of vote spam evolve
    - Vote spam methods can change significantly due to varying popularity of content, specifics of media and topic
- Challenge
  - A robust method to train a ranking function that remains resilient to evolving vote spam attacks

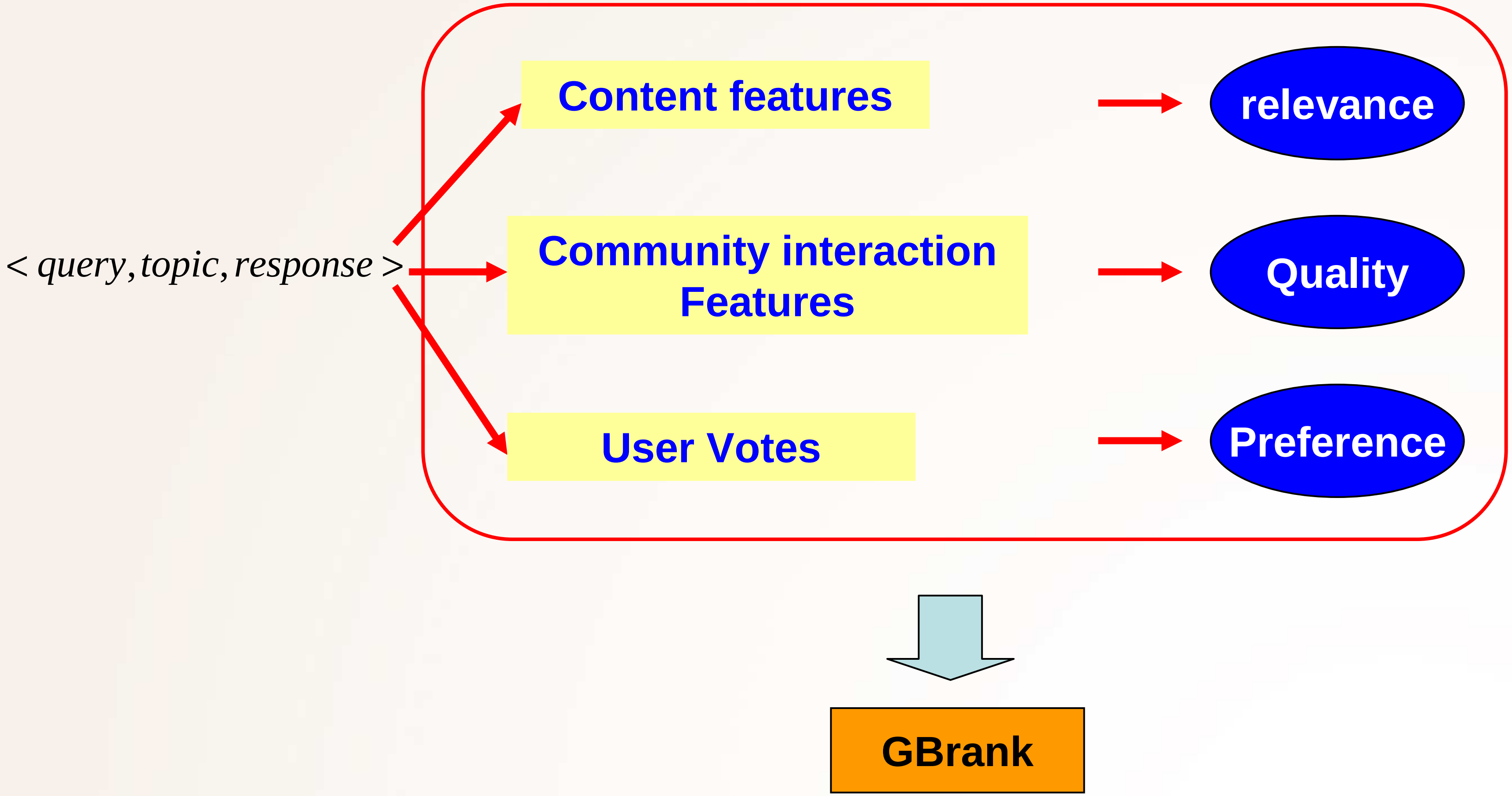
Outline

- Background and Motivation
- Learning Ranking Functions in Social Media
- Vote Spam in Social Media
- Experiments on Community Question Answering

Social Content and User Votes in Social Media



Learning-based Approach

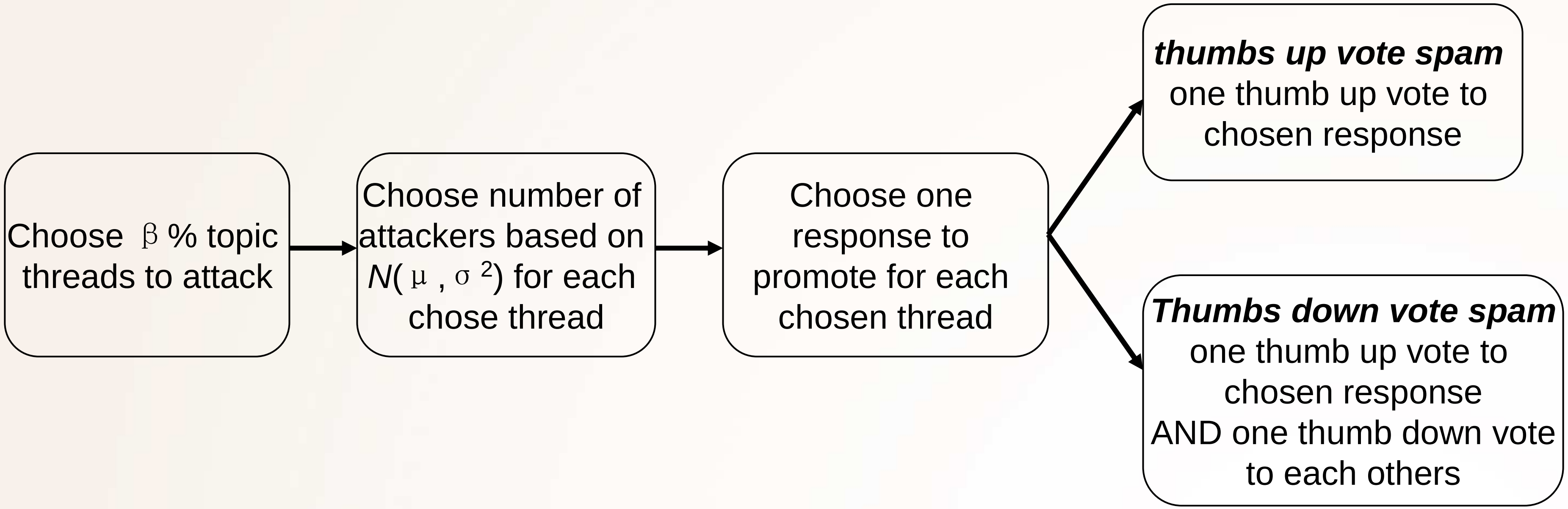


Outline

- Background and Motivation
- Learning Ranking Functions in Social Media
- Vote Spam in Social Media
- Experiments on Community Question Answering

Vote Spam Attack Models

- Two main types of vote spam
  - Incorrect votes – not an expert
  - Malicious votes – promote some specific responses



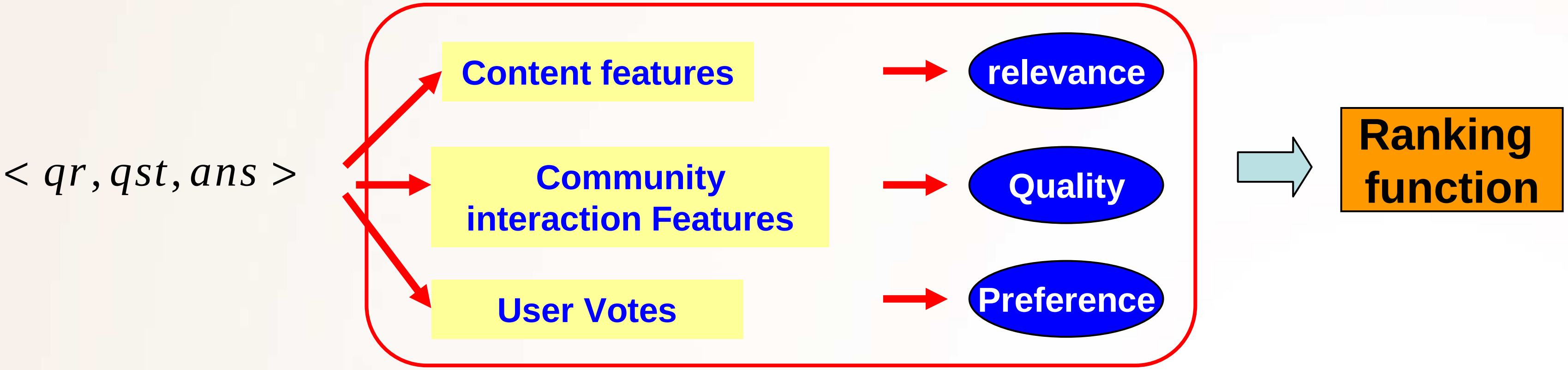
Outline

- Background and Motivation
- Learning Ranking Functions in Social Media
- Vote Spam in Social Media
- Experiments on Community Question Answering

EMORY Georgia Tech College of Computing School of Computer Science

Robust Ranking Method

- *GBrank* in QA retrieval [Bian et al. 2008]
  - Promising performance
  - User vote information provides much contribution to the high accuracy (no vote spam)
- Robust ranking method – *GBrank-robust*
  - Apply the general vote spam model to generate vote spam into unpolluted QA data
  - Train the ranking function based on new polluted data
  - Transfer more weight to other content and community interaction features



EMORY Georgia Tech College of Computing School of Computer Science

Experimental Setup

- Dataset
  - Factoid questions from the TREC QA benchmarks:
    - Total question set: 3000 factoid questions from 1999 to 2006
    - 1250 factoid questions from total question set—have at least one similar question in the Yahoo! Answers archive
  - Question-answer collection dataset
    - To simulate a user's experience with a community QA site
    - Submit each TREC query to Yahoo! Answers and retrieve up to 10 top-ranked questions according to Yahoo! Answer ranking
    - For each of Yahoo! Questions, we retrieve all of its answers
    - 89642 <*qr, qst, ans*> tuples
  - Relevance Judgments
    - Automatically labels using the TREC factoid answer patterns
    - 17711 tuples (19.8%) are labeled as "relevant"
    - 71931 tuples (81.2%) are labeled as "non-relevant"

Experimental Setup

- Evaluation Metrics
  - Precision at *K*
    - For a given query, *P(K)* reports the fraction of answers ranked in the top *K* results that are labeled as relevant
  - Mean Reciprocal Rank (MRR)
    - The MRR of each individual query in the reciprocal of the rank at which the first relevant answer was returned
  - Mean Average of Precision (MAP)
    - The mean of average precision of all queries in the test set

Ranking Methods Compared

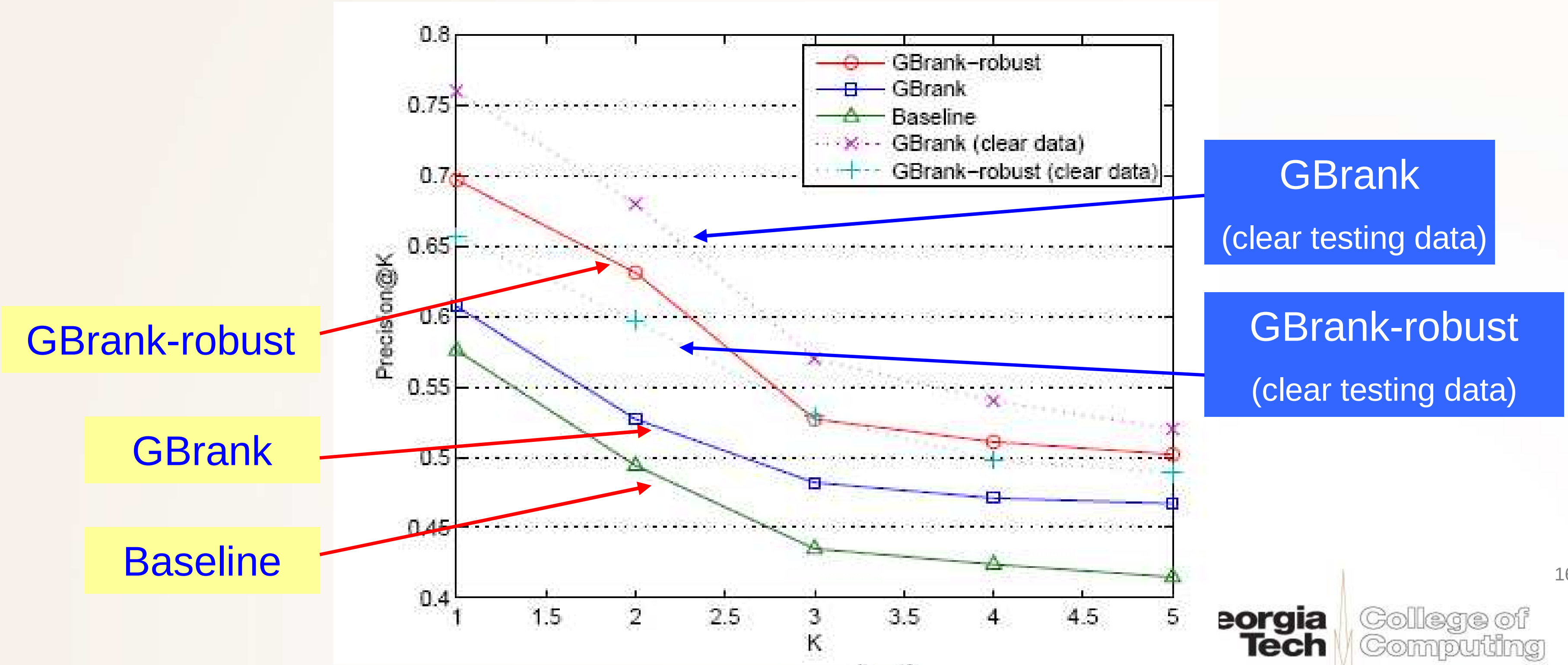
- Baseline:
  - Let "*best answer*" always be on top
  - Following answers are ranked in decreasing order by number of (thumbs up votes – thumbs down votes)

- GBrank:
  - Ranking function with textual and community interaction features and preference extracted from voting information

- GBrank-robust:
  - Similar to GBrank
  - The training data is polluted according to the chosen spam model

Experimental Results

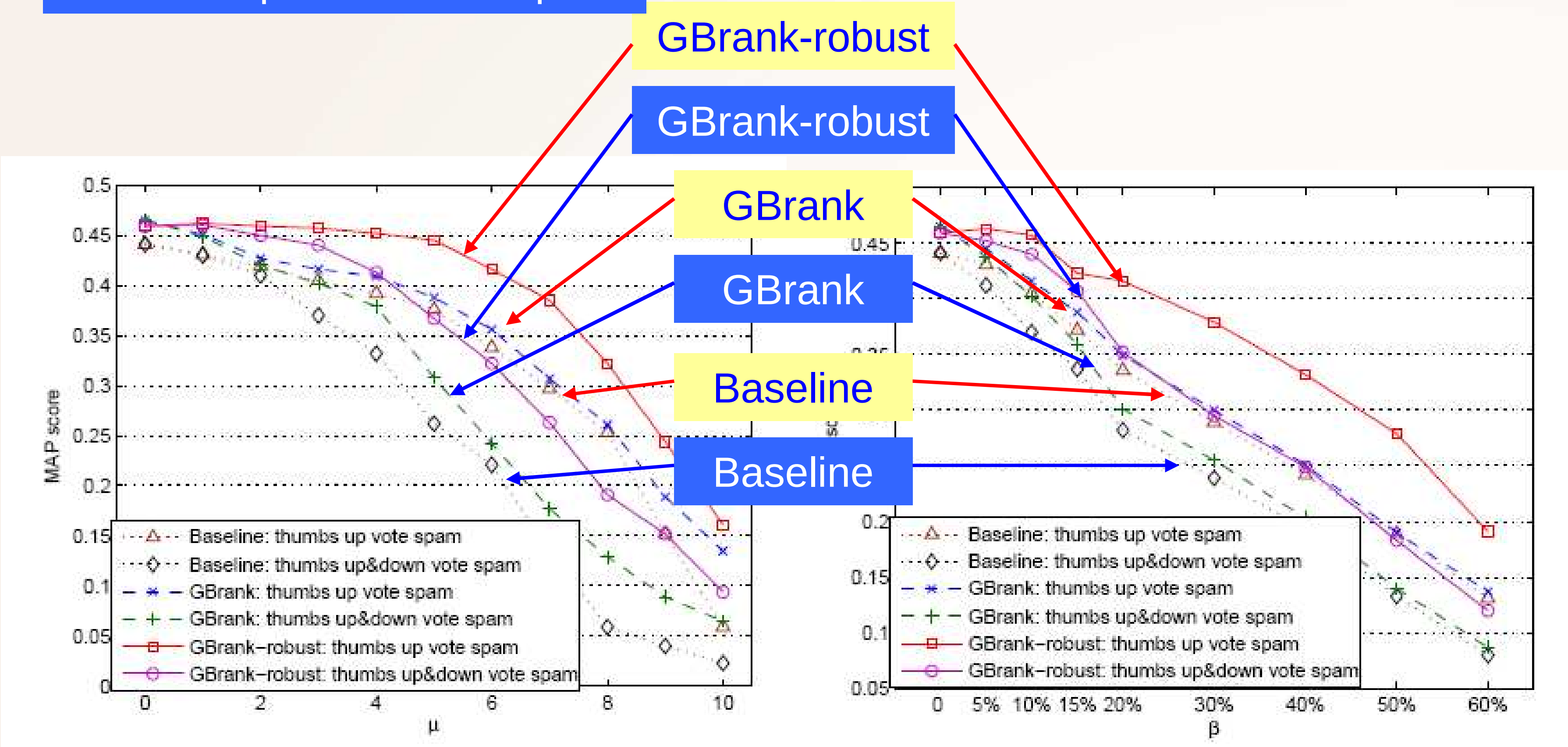
- QA Retrieval
  - Vote spam model: $\beta\% = 10\%; N(\mu, \sigma^2) = N(3, 1^2)$
  - Training data: randomly select 800 TREC queries and all related QA
  - Testing data (polluted): remainder 450 TREC queries and all related QA



Robustness to Vote Spam

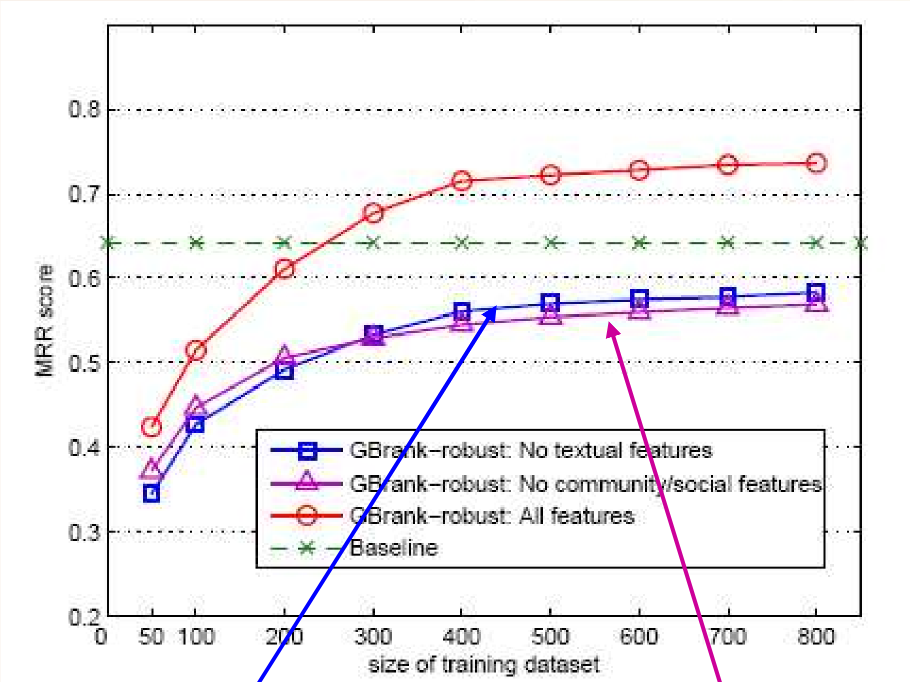
Thumbs up vote spam

Thumbs up&down vote spam



EMORY Georgia Tech College of Computing School of Computer Science

Analyzing Feature Contribution



No textual features

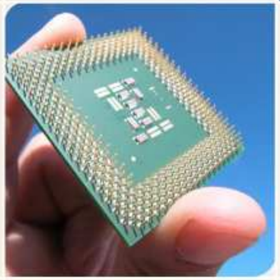
No community interaction features

Info Gain	Feature Name
0.048	Similarity between query and question
0.045	Number of resolved questions of the answerer
0.043	Length ratio between query and answer
0.003	Number of thumbs down vote
0.029	Number of stars for the answerer
0.002	Number of thumb up vote
0.026	Similarity between query and qst+ans
0.018	Number of answer terms

Contribution and Future Work

- Contributions
  - A parameterized vote spam model to describe and analyze some common forms of vote spam
  - A method for increasing the robustness of ranking by injecting noise at training
  - A comprehensive evaluation on ranking performance for community question answering under a variety of simulated vote spam attacks, demonstrating robustness of our ranking

- Future work
  - Explore further the different spam strategies and corresponding robust ranking methods



Thank you!

EMORY

Georgia Tech College of Computing
School of Computer Science

Related Work

- Robustness of web search ranking to click spam
  - [Jansen 2006] reveal the influence of malicious clicks on online advertising
  - [Radlinkski 2006] present how click spam bias the ranking results
  - [Immorlica et al. 2005] demonstrate that a particular class of learning algorithm are resistant to click fraud in some sense

- Ranking the content in social media site [Bian et al. 2008]
  - Present a ranking framework to utilize user interaction information (including user votes) to retrieve high quality relevant content in social media