

# SocialQ&A: An Online Social Network Based Question and Answer System

Haiying Shen \*, Guoxin Liu\*, Nikhil Vithlanit†

\*Department of Electrical and Computer Engineering, Clemson University, Clemson, USA

†Amazon.com, Seattle, USA

Presenter: Guoxin Liu  
PhD candidate



# Outline

- **Introduction**
- **Background**
- **Social Q&A**
- **Results and Analysis**
- **Conclusion and Future work**

# → Introduction



# Introduction

- **Important source of information**
  - **Yahoo! Answers – 10 million users in first 2 years**
  - **Currently 200 million users**
  - **15 million visits everyday**

# Introduction

- **Scope for improvement**

■ Answered   ■ Unanswered   ■ Unsatisfied

## Yahoo! Answers



## Baidu Zhidao

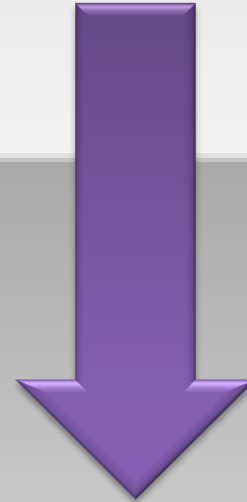


# Introduction



**Quality of answers**

**Wait Time**



# Introduction

## Contributions of this paper

- **Design of SocialQ&A**
- **Implementation of SocialQ&A**
- **Trace-driven experiments and real usage of SocialQ&A**



# Outline

- Introduction
- **Background**
- Social Q&A
- Results and Analysis
- Conclusion and Future work





# Background

## History of search engines

- **Manual Indexing**
- **Automatic Indexing**
- **Google**

# Background

## Information retrieval paradigms

### Library paradigm



### Village Paradigm



# Background

## Q&A systems: Evolution

- **Community based Question & Answer systems**
- **Online social based Question & Answer systems**



# Outline

- Introduction
- Background
- **SocialQ&A**
- Results and Analysis
- Conclusion and Future work

# → SocialQ&A

## Components of SocialQ&A



# SocialQ&A

## Core algorithms

- **User Interest Analyzer**
- **Question Categorizer**
- **Question-User Mapper**

# User Interest Algorithm

**Analyzes the user profile and associates each user with the user interest vector**

# User Interest Algorithm

Tokenization

I love country music



# User Interest Algorithm

Matching and association

Tokens

Synset

User X's  
interest vector

love

rock

music

metal

country

country

reggae

# User Interest Algorithm

	Rock	Metal	Country	Reggae
User x	0	0	1	0

# SocialQ&A

## Core algorithms

- **User Interest Analyzer**
- **Question Categorizer**
- **Question-User Mapper**

# Question Categorizer

**Analyzes the question and associates the question with tags.  
Updates the user interest vector**

# Question Categorizer

Tokenization

Asked by user X

what is the best country band ?

# Question Categorizer

Matching and association

Tokens

Synset

Question  
Tags

what

rock

best

metal

country

country

band

reggae

# SocialQ&A

## Core algorithms

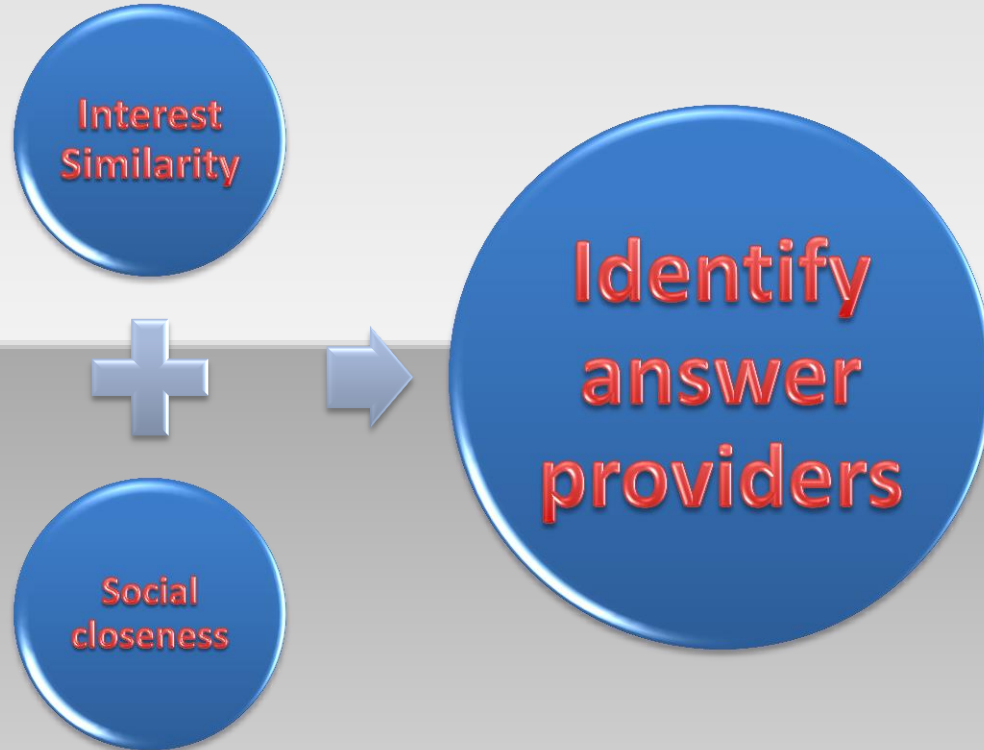
- **User Interest Analyzer**
- **Question Categorizer**
- **Question-User Mapper**

# Question-User Mapper

**Generates a list of potential answer providers by leveraging the asker's and answers' interest vectors and the tags associated with the question.**



# Question-User Mapper



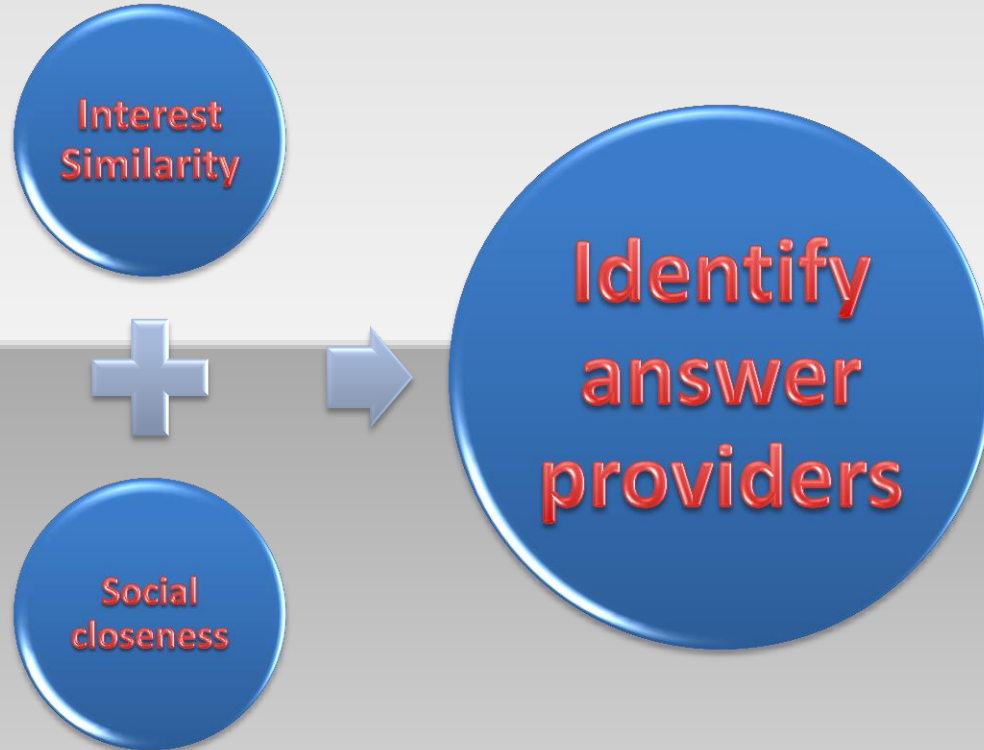
# Question-User Mapper

Factors considered for calculating the **interest similarity**

- **User interest vector**
  - **User's profile data**
  - **Type of questions asked by the user**
  - **Answers provided by the user**
- **Question interest vector**

$$\Psi_{I,U_k} = \text{Sum}(W_{I_j}^{U_k})$$

# Question-User Mapper



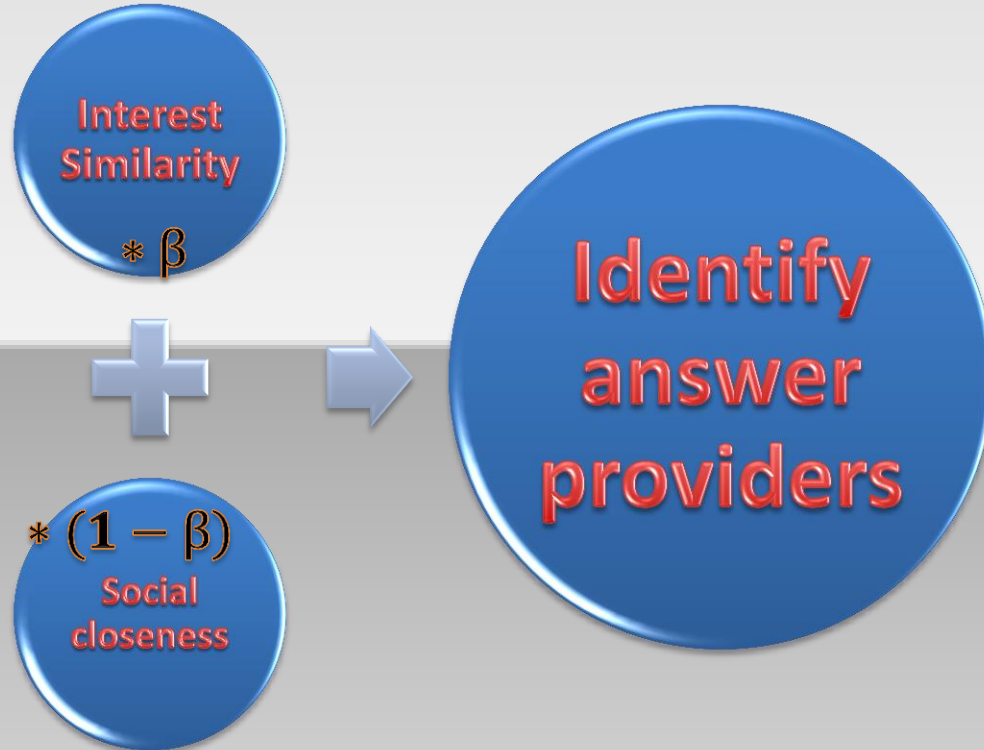
# Question-User Mapper

## Factors considered for calculating the social closeness factor

- Interest similarity between two users ( $P_S$ )
- Past interactions between the users ( $P_A$ )
- Number of common friends to both the users ( $P_C$ )

$$\Psi_{C,U_k} = \gamma_S * P_S + \gamma_A * P_A + \gamma_C * P_C$$

# Question-User Mapper





# Outline

- Introduction
- Background
- Social Q&A
- **Results and Analysis**
- Conclusion and Future work



# Results and Analysis

## Statistics

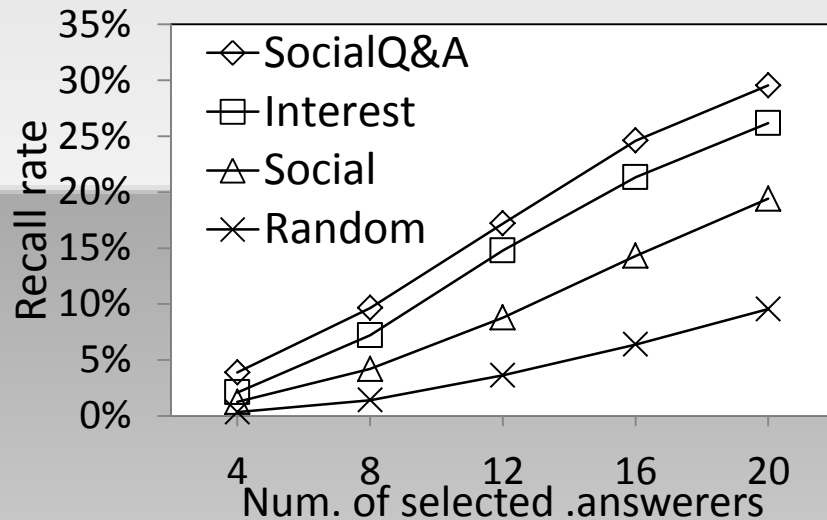
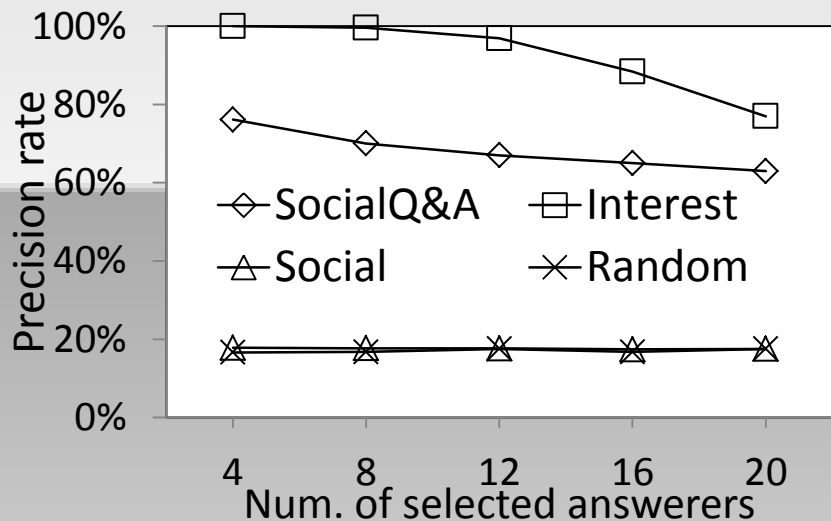
- 1000 Facebook users
- 9419 Yahoo! Answer questions with best answers

## Metrics

- Precision
- Recall
- Wait time

# Results and Analysis

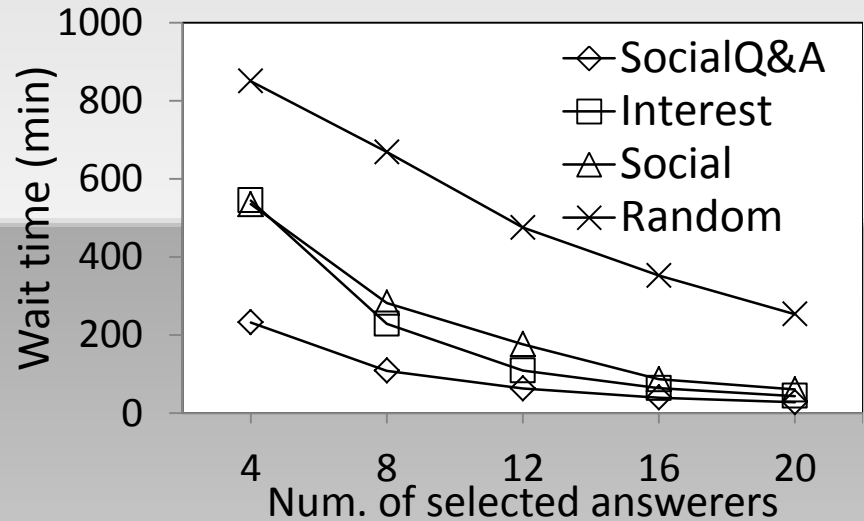
High precision + high recall = high answer quality





# Results and Analysis

## Short wait time





# Prototype usage of SocialQ&A

**Number of users - 124**

**Number of questions - 163**

**Number of answers - 282**

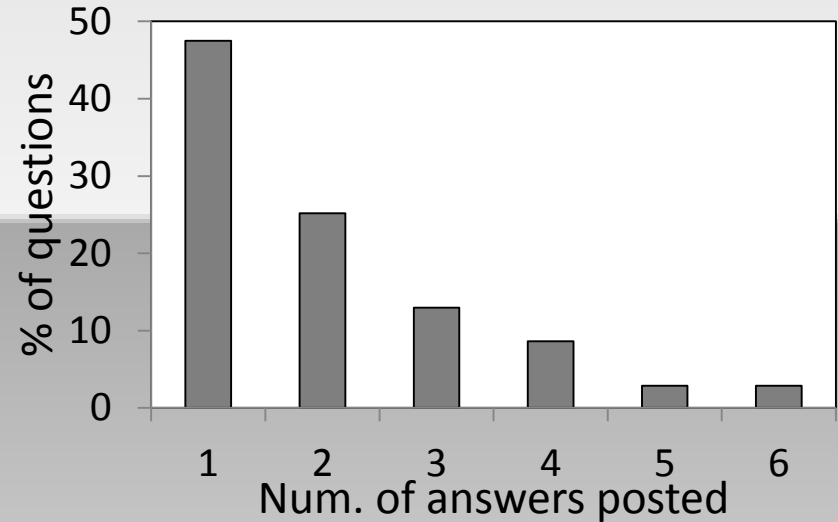
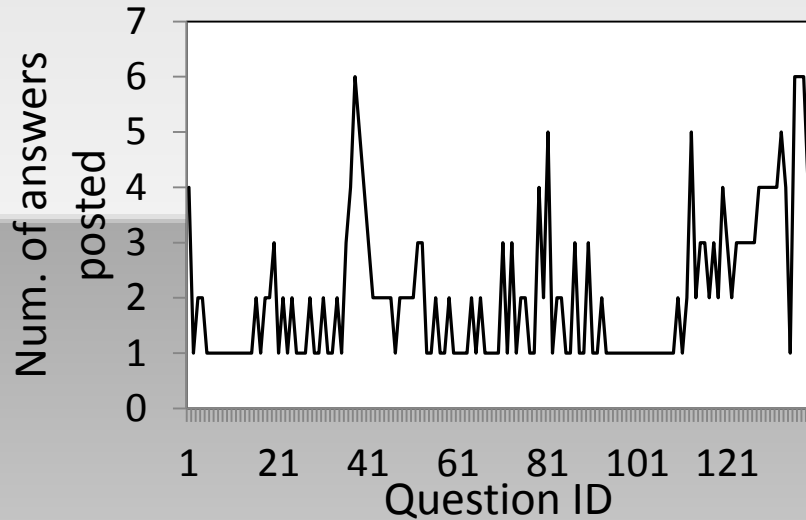
**Number of users from India - 70**

**Number of users from US - 35**

**Number of users from UK - 1**

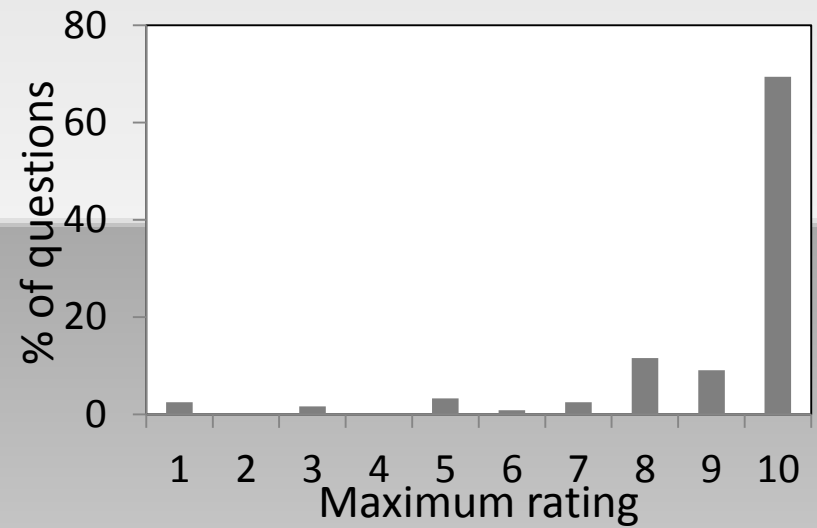
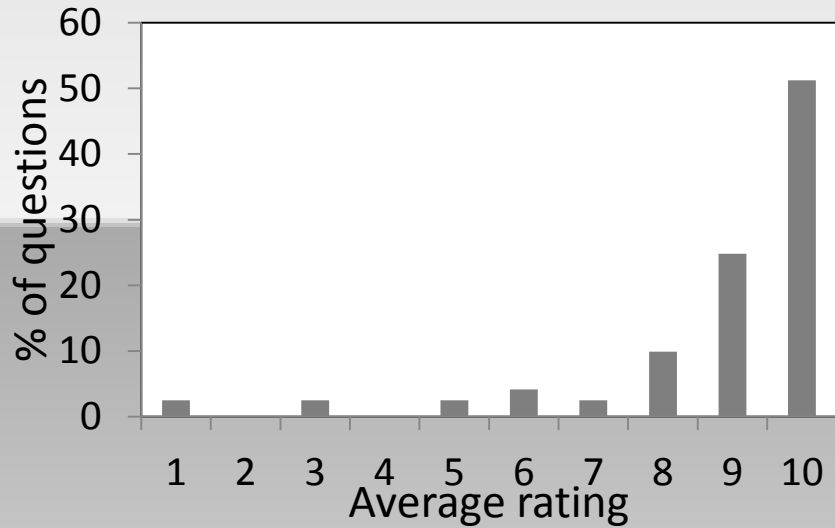
# Results and Analysis

## Questioning and answering activity



# Results and Analysis

Answer quality: high ranked



# Results and Analysis

## User questioning and answering activity

- **Percentage of unanswered questions (SocialQ&A) – Approximately 15%**
- **Percentage of unanswered questions (Yahoo! Answers) – Approximately 16%**
- **Percentage of unanswered questions (Baidu Zhidao) – Approximately 40%**



# Outline

- Introduction
- Background
- Social Q&A
- Results and Analysis
- Conclusion and Future work



# Conclusion and future work

## SocialQ&A

- Social network based answerer selection
- High quality potential answerer selection

## Achievement

- High Quality
- Short wait time

## Future work

- Questions outside the social community

# Outline

Introduction

- Background
- Social Q&A
- Results and Analysis
- Conclusion and Future work







**Thank You**

**Questions** ←

If you have any questions and comments  
please contact [guoxinl@g.clemson.edu](mailto:guoxinl@g.clemson.edu)