

Evaluating participants in group-discussions on Twitter

Utkarsh Upadhyay Nakoi Yoshinaga Satoko Itaya
Rie Tanaka Taku Konishi Shinchin Doi Keiji Yamada
Peter Davis



Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis
Scoring
Reliability

Results

Applications

1 Introduction

2 Approach

3 Results

4 Applications

What are Group discussions?

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis
Scoring
Reliability

Results

Applications



Like:

Short problems solving meetings
Brainstorming

Not like:

Mailing lists
Extended Monologues

What are Group discussions?

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis
Scoring
Reliability

Results

Applications



- Conversation is between multiple members
- Happens in real-time

Where do they happen online?

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected

First-use analysis

Scoring

Reliability

Results

Applications



- **IRC:** e.g. Globally distributed software teams
- **Twitter:** TweetChats, new but with low barrier to entry

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected

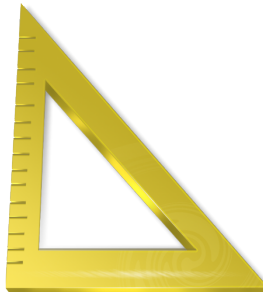
First-use analysis

Scoring

Reliability

Results

Applications



- For rewards incentives
- Monitoring *burn-outs* and *health* of groups
- Promoting collaboration (eg. CSCL)

How is it currently done?

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected

First-use analysis

Scoring

Reliability

Results

Applications



Informally

- Subjective
- Bias, first impressions
- Good reasons



Theoretical assessment

- Expert knowledge
- Manual labor
- No agreed criteria

How is it currently done?

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected

First-use analysis

Scoring

Reliability

Results

Applications



Can we do better?

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis
Scoring
Reliability

Results

Applications

1 Introduction

2 Approach

3 Results

4 Applications

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis
Scoring
Reliability

Results

Applications



Transcript



First-use Content
Analysis



Final Scores





- Microblogging platform
- 140 character long messages (called *tweets*)
- Viewing tweets:
 - Timelines (user-specific)
 - Real time search (keyword of #tag specific)
- Explicit @username mentions and ReTweeting



#TheOnlineMom



#KidLitArt



#u30pro



#GenYchat

Some TweetChats

Some properties:

- Wide variety
- ≈ 1.5 hour long
- Weekly
- For ≈ 15 weeks

Ideal group-discussions:

- Real time
- Conversational
- Simple English
- Explicit references

Sample Tweet

“Agree with ^{mention} @PersonX, no ^{keyword} little kids on ^{keyword} Facebook.”

Keywords : Noun phrases (w/ adj.) with IDF = 1

Reference : Explicit references to other participants

TweetChat	Keywords chosen	Keywords rejected
#TheOnlineMom	voice command, Sim City	kids, Kinect , game
#GenYChat	etiquette, Beiber	people, blog
#u30pro	long term goal, role model	job, company
#KidLitArt	Wordpress, Dropbox	work, book, art

First use analysis

Evaluating participants in group-discussions

Introduction

Group discussions

Approach

Data collected

First-use analysis

Scoring

Reliability

Results

Applications

Primary User 1



Secondary User 1



Primary User 2



Secondary User 2



Timeline of usage of keywords

First use analysis

Evaluating participants in group-discussions

Introduction

Group discussions

Approach

Data collected

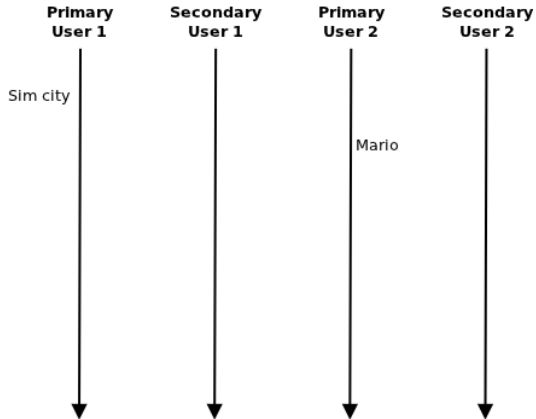
First-use analysis

Scoring

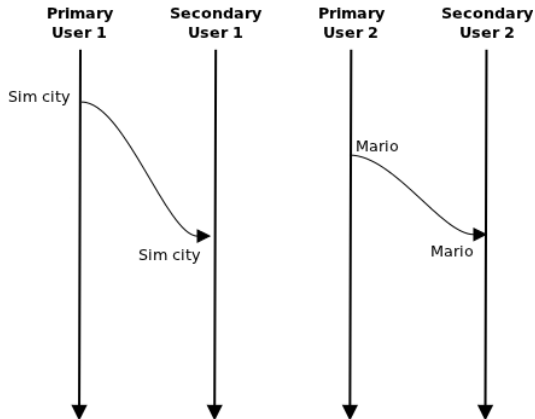
Reliability

Results

Applications



First use of keywords defines *Primary users* for each word



Second use of keywords marks *Secondary users* for each word

Evaluating participants in group-discussions

Introduction

Group discussions

Approach

Data collected

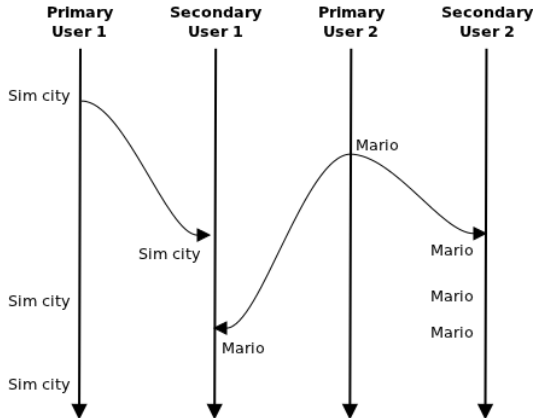
First-use analysis

Scoring

Reliability

Results

Applications



Subsequent usages does not alter the relationship, but can form new ones



Generators



Interactors



Propagators

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected

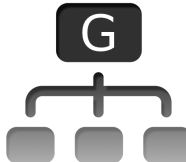
First-use analysis

Scoring

Reliability

Results

Applications



Role : Initiate discussions, introduce new interesting topics.

Score : The number of **secondary users**.

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis

Scoring
Reliability

Results

Applications



Role : Social pillars, involve others in discussion.

Score : The number of **mentions exchanged**.

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected

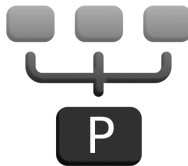
First-use analysis

Scoring

Reliability

Results

Applications



Role : Critics, early adopters.

Score : The number of **secondary uses**.



Robustness via repetition:

- Multiple chats
- Regular users
- Boot strapping (resampling)

Peer review:

- Generators
- Interactos

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis
Scoring
Reliability

Results

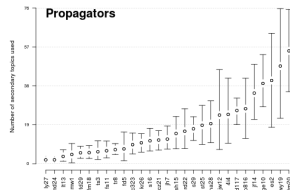
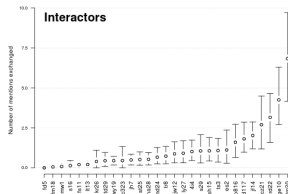
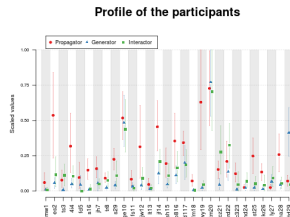
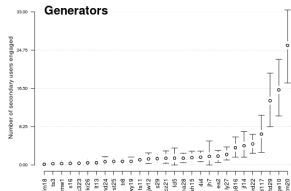
Applications

1 Introduction

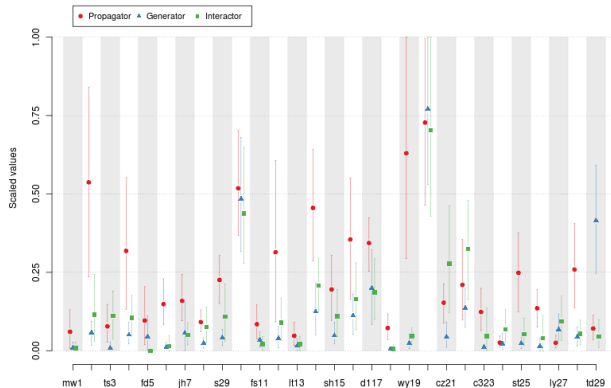
2 Approach

3 Results

4 Applications



All properties of the participants



Profile of participants

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis
Scoring
Reliability

Results

Applications

1 Introduction

2 Approach

3 Results

4 Applications

Evaluating participants in group-discussions

Introduction

Group discussions

Approach

Data collected

First-use analysis

Scoring

Reliability

Results

Applications



Extreme overall performers

Evaluating participants in group-discussions

Introduction

Group discussions

Approach

Data collected

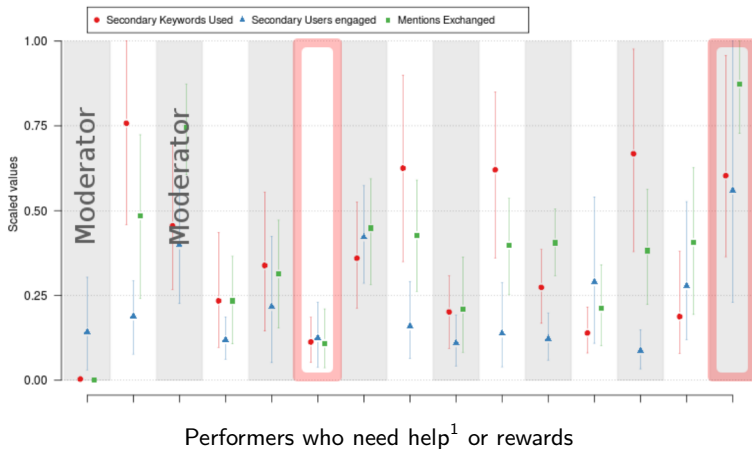
First-use analysis

Scoring

Reliability

Results

Applications



¹voluntary participation

Monitoring health of the group

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected

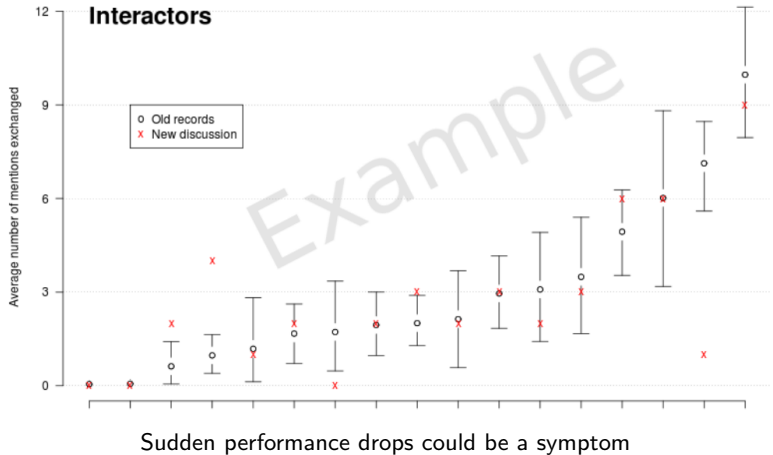
First-use analysis

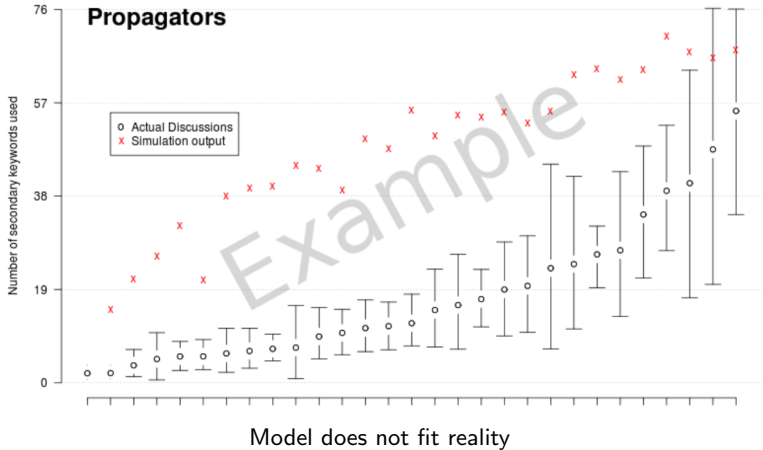
Scoring

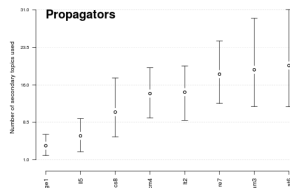
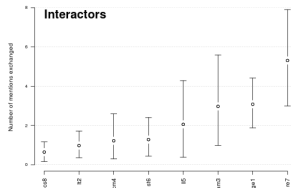
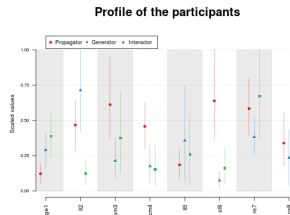
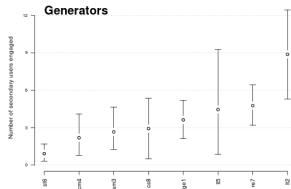
Reliability

Results

Applications







Less intra-group specialization in #LitChat than #TheOnlineMom

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected
First-use analysis
Scoring
Reliability

Results

Applications



- Objective robust performance evaluation
- Monitoring health of participants (and the group)
 - Behavioral anomalies
- Evaluation criteria of models and inter-group comparisons

Evaluating
participants in
group-discussions

Introduction

Group
discussions

Approach

Data collected

First-use analysis

Scoring

Reliability

Results

Applications



- Restrictive definition of keywords
- Anaphoric references
 - To people (esp. in transcripts from oral discussions)
 - To old topics
- Extending to other Group-discussions

Thank you

Questions?

Utkarsh Upadhyay