

EVALUATING THE EFFECTIVENESS OF VISUALIZATION  
DESIGN FOR TWITTER CONVERSATIONS ON ACADEMIC  
TOPICS

by

Shali Liu

Submitted in partial fulfillment of the requirements  
for the degree of Master of Computer Science

at

Dalhousie University  
Halifax, Nova Scotia  
December 2015

© Copyright by Shali Liu, 2015

# Table of Contents

<b>List of Tables</b> . . . . .	<b>vii</b>
<b>List of Figures</b> . . . . .	<b>viii</b>
<b>Abstract</b> . . . . .	<b>ix</b>
<b>List of Abbreviations and Symbols Used</b> . . . . .	<b>x</b>
<b>Acknowledgements</b> . . . . .	<b>xi</b>
<b>Chapter 1 Introduction</b> . . . . .	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Objectives . . . . .	3
1.3 Contributions . . . . .	4
<b>Chapter 2 Background and Related Work</b> . . . . .	<b>5</b>
2.1 Data and Structure of Twitter . . . . .	5
2.1.1 Studies on Twitter Data . . . . .	5
2.1.2 Twitter Networks and Conversations . . . . .	5
2.2 Text Mining Methods . . . . .	6
2.2.1 Feature Selection and Evaluation and Classification . . . . .	6
2.3 Visual Analytics and Cognition . . . . .	7
2.3.1 Human Computer Interaction Design Process . . . . .	7
2.3.2 Design Guidelines for Interactive Visualization . . . . .	8
2.4 Design Conventions and Encodings . . . . .	8
2.4.1 Visual Encodings . . . . .	8
2.4.2 Visualization Design Perceptions . . . . .	8
2.5 Task-oriented Visualization Design . . . . .	9
2.5.1 Visual Analytics . . . . .	9
2.5.2 User Tasks for Information Systems and for Twitter . . . . .	10
<b>Chapter 3 Understanding the Features of Tweets</b> . . . . .	<b>11</b>
3.1 Introduction for the Two Phased User Study . . . . .	11
3.2 Data Collection . . . . .	12

3.2.1	Data Domain . . . . .	12
3.2.2	Data Processing . . . . .	13
3.3	Understanding the Features of Twitter Conversations . . . . .	15
3.3.1	Previous Studies on Features of Twitter Conversation . . . . .	15
3.3.2	Candidate Features for the User Study . . . . .	16
3.4	Study Design . . . . .	17
3.4.1	Introduction . . . . .	17
3.4.2	Feature Vector and Data Sampling . . . . .	17
3.4.3	Study Protocol . . . . .	19
3.5	Study Results Analysis . . . . .	22
3.5.1	Introduction to Data Analysis Approaches . . . . .	22
3.5.2	Analysis of Users' preference . . . . .	23
3.5.3	Feature Evaluation . . . . .	25
3.5.4	Results Discussion . . . . .	28
<b>Chapter 4</b>	<b>Designing an Interactive Visualization of Twitter Con- versations . . . . .</b>	<b>29</b>
4.1	Visual Features vs. Automatic Methods Trade-offs . . . . .	29
4.1.1	Features Definition and Examples . . . . .	30
4.1.2	Automatic Methods and Examples . . . . .	30
4.2	Processing Data to Support Visual Representation . . . . .	31
4.2.1	Twitter Conversation Classification . . . . .	31
4.2.2	Score Prediction . . . . .	33
4.2.3	Data Support for Visual Features . . . . .	34
4.3	Design of Visualizations . . . . .	34
4.3.1	Design considerations . . . . .	34
4.3.2	Visual Encodings . . . . .	36
4.3.3	Four Versions of Visualizations . . . . .	38
4.4	Design of User Study . . . . .	42
4.4.1	Grouping Participants . . . . .	42
4.4.2	Data Collection . . . . .	45
4.4.3	Study Protocol . . . . .	45
<b>Chapter 5</b>	<b>Evaluation of Visualization Designs . . . . .</b>	<b>48</b>
5.1	Introduction . . . . .	48
5.2	Research Questions . . . . .	48
5.3	Description of Data . . . . .	49

5.4	Evaluation Variables . . . . .	49
5.4.1	Dependent Variables and Between-subjects Variables in ANOVA	50
5.4.2	Within-subjects Variables and Between-subjects Factors in Repeated Measures ANOVAs . . . . .	51
5.5	One-way Analysis of Variance (ANOVA) . . . . .	52
5.5.1	ANOVA of Dependent Variables with $H_{feature} * H_{score}$ as Between-subjects Variable . . . . .	53
5.5.2	ANOVA and Post Hoc Analysis of Dependent Variables with Version as Between-subjects Variables . . . . .	54
5.5.3	ANOVA and Post Hoc Analysis of Dependent Variables with Task Category as Between-subjects Variables . . . . .	54
5.6	One-way Repeated Measures ANOVAs . . . . .	57
5.6.1	Repeated Measures of $H_{feature} * H_{score}$ . . . . .	57
5.6.2	Repeated Measures and Post Hoc of Version . . . . .	57
5.7	Other Data Analysis . . . . .	59
5.7.1	Function Usage . . . . .	59
5.7.2	Post Study Questionnaires and Questions . . . . .	60
5.8	Search Task vs. Browsing Task . . . . .	60
5.9	Conclusion and Design Guidelines . . . . .	61
5.9.1	Design Guidelines and Procedures . . . . .	61
5.9.2	Future Improvements for the User Study and the Visualization	62
<b>Chapter 6</b>	<b>Conclusion . . . . .</b>	<b>63</b>
6.1	List of Contributions . . . . .	63
6.1.1	Datasets . . . . .	64
6.1.2	Relatedness and Informativeness . . . . .	64
6.1.3	Visual Analysis Tool . . . . .	64
6.1.4	User Study Design for Evaluating Visualization System . . . . .	64
6.1.5	Design Guidelines and Procedures . . . . .	64
6.2	Limitations and Future Work . . . . .	65
<b>Bibliography</b>	<b>. . . . .</b>	<b>67</b>
<b>Appendix A</b>	<b>First Phase User Study - Study Protocol . . . . .</b>	<b>71</b>
<b>Appendix B</b>	<b>First Phase User Study - Recruitment Notice . . . . .</b>	<b>80</b>
<b>Appendix C</b>	<b>First Phase User Study - Informed Consent . . . . .</b>	<b>81</b>

Appendix D	First Phase User Study - Background Questionnaire . . . . .	84
Appendix E	First Phase User Study - On-screen Questionnaires . . . . .	86
Appendix F	First Phase User Study - Post-task Questionnaires . . . . .	87
Appendix G	First Phase User Study - Letter of Approval . . . . .	88
Appendix H	Second Phase User Study - Task Questions . . . . .	91
Appendix I	Second Phase User Study - Study Protocol . . . . .	92
Appendix J	Second Phase User Study - Recruitment Notice . . . . .	102
Appendix K	Second Phase User Study - Informed Consent . . . . .	103
Appendix L	Second Phase User Study - Background Questionnaire . . . . .	106
Appendix M	Second Phase User Study - Post-task Questionnaires . . . . .	108
Appendix N	Second Phase User Study - Post Study Interview Questions . . . . .	109
Appendix O	Second Phase User Study - Letter of Approval . . . . .	110
Appendix P	Second Phase User Study - Post Study Interview Questions Results . . . . .	113
Appendix Q	Hits / Clicks Raw Results . . . . .	120
Appendix R	Time Raw Results . . . . .	131
Appendix S	Precision / Recall Raw Results . . . . .	142

Appendix T	Function Raw Results	153
------------	----------------------	-----

## List of Tables

Table 3.1	Topics and Keywords . . . . .	13
Table 3.2	Data Statistics . . . . .	15
Table 3.3	Candidate Features . . . . .	16
Table 3.4	Feature Vector and the Range of the Values for Each Feature .	18
Table 3.5	First Phase Study Procedures . . . . .	20
Table 4.1	Method Tasks and Type . . . . .	31
Table 4.2	Four Versions Of Visualization System . . . . .	31
Table 4.3	Visual Properties for 1) visual features and 2) automatic methods	35
Table 4.4	Second Phase Study Procedures . . . . .	46
Table 5.1	Values of $H_{feature} * H_{score}$ to Version Number . . . . .	51
Table 5.2	Tests of Between-Subjects Effects (Dependent Variable: TotalClicks)	53
Table 5.3	Total Clicks in Different Versions . . . . .	54
Table 5.4	Multiple Comparisons (Dependent Variable: TotalClicks) . . .	55
Table 5.5	Total Clicks in Different Task Categories . . . . .	57
Table 5.6	Tests of Between-Subjects Effects for Repeated Measures (De- pendent Variable: TotalClicks) . . . . .	58
Table 5.7	Function Usages in Different Versions . . . . .	59
Table 5.8	Function Usages in Different Tasks . . . . .	60

## List of Figures

Figure 3.1	Twitter conversation rating page. A: conversation text area; B: topic tags; C: rating board; D: map for twitter location; E: users profile. . . . .	21
Figure 3.2	User Preferences of Features . . . . .	24
Figure 3.3	Feature Evaluation for (1) Information Score (2) Relatedness Score . . . . .	27
Figure 4.1	System Screenshot (Version 4). A: Conversation Popup; B: Axes; C: Universities and topics. . . . .	37
Figure 4.2	Version one: (1) general view (2) detail view . . . . .	40
Figure 4.3	Version two: (1) general view (2) detail view . . . . .	41
Figure 4.4	Version three: (1) general view (2) detail view . . . . .	43
Figure 4.5	Version four: (1) general view (2) detail view . . . . .	44
Figure 5.1	Significant Interactions Between Dependent Variables and TaskCategory. . . . .	56
Figure 5.2	The Graph of Significant Interactions . . . . .	58



## **Abstract**

Our research focuses on evaluating the efficiency of the combination of the computed scores and visual representations of features. We conducted a two-phased user study. The first phase were designed to determine the important text features. In the first phase, participants were asked to rate a set of Twitter data and prominent features. After the key feature set was determined, we started to evaluate the two sources of results to be visualized: 1) informativeness and relatedness scores and 2) visual features. We built four versions of visualizations to represent the full projection of the two sources. In the second phase of the user study, 48 participants were recruited to do a between-subject user study for evaluating the four versions of visualizations using the same set of information seeking tasks. The study shows that the visualization combining scores and features perform the best in efficiency for browsing tasks.

## List of Abbreviations and Symbols Used

$H_{feature}$	Boolean value for whether the visualization is with or without visualized features
$H_{score}$	Boolean value for whether the visualization is with or without scores
$e_{c\_length}$	Conversation text length
$e_{i\_score}$	Informativeness score
$e_{loc}$	Boolean value for whether in campuses or not
$e_{r\_score}$	Relatedness score
$e_{t\_num}$	Number of tweets in the conversation
$e_{topic}$	Classification result for topics
$e_{u\_num}$	Number of users in the conversation
$e_{unv}$	Classification result for universities
$e_{url}$	Boolean value for whether the text contains url or not
$e_{user}$	User influence score

**ANOVA** One-way Analysis of Variance

**LDA** Latent Dirichlet allocation

**RBF** Radial Based Function

**SVC** Support Vector Classifier

**SVM** Support Vector Machine

**SVR** Support Vector Regression

**TF-IDF** Term Frequency-Inverse Document Frequency

**UGC** User-Generated Content

## Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Dr. Milios and Dr. Hawkey for the continuous support of my master study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better supervisors and mentor for my master study. Also, many thanks to Dr. MacKay for the valuable feedbacks when Dr. Hawkey is away.

Besides my supervisors, I would like to thank Dr. Reilly and Dr. Mackey to be in my defence committee, and for their insights which will help enrich my work.

My sincere thanks also goes to Boeing project for the generous grants and supporting platforms which enabled me to pursue my research goals and conduct various user studies. Also, the exchange of research plans and results in the annual meetings also helped me a lot for setting my directions and goals.

I thank my fellow labmates in for the stimulating discussions, for the time we were working together for presentations, and for all the fun we have had in the past years. Many thanks to them for encouraging me to complete this thesis when I was faced with difficulties. Without them, it would be hard for me to imagine how my work would have ended.

Last but not the least, I would like to thank my parents and relatives for supporting me spiritually throughout writing this thesis and my life in general.

# Chapter 1

## Introduction

### 1.1 Motivation

Twitter is a microblogging service that enables its users to send and read text-based messages of up to 140 characters, known as “tweets”. It is one of the most influential social media with a reported over 340 millions tweets daily in 2012 [1]. Users on Twitter post text, pictures and urls of videos, talking about various topics such as economics, politics, academics, celebrity and lifestyle.

There are several forms of interactivities on Twitter, including retweets, replies, trending topics and messages. One of the major interactive activities is Twitter conversation, which includes Retweet, Reply amongst users. Retweet, abbreviated as RT on Twitter, is used to tweet the content that was posted by another user. A reply is marked as *@username* in Twitter. A reply represents a tweet that responds to a previously posted tweet created by, in most cases, another user. It is important to note that retweet is symbolized as RT *@username* or even simply *@username*, which appears the same as replies. The major difference between Retweet and Reply is that replies usually create new content while retweets simply refer to the same content as the original tweet. Conversation activity is of high importance on Twitter, especially for research purposes. One research [2] shows that conversation interactions have some benefits over standalone tweets. First, conversations have more user involvement in a particular topic than a standalone tweet. Second, the majority of conversations have been used to address some questions raised by earlier tweets in the conversation. Third, people tend to give opinions when they are in a conversation.

There has been increasing interest in using social media for market research. Companies and institutions have started to take advantage of social media communities, especially on some of the more popular social media networks such as Facebook and Twitter. Traditionally, marketers use online questionnaires, email surveys, telephone

interview, postal questionnaires, and face-to-face interviews to investigate their market requirements [3]. However, these traditional approaches have a few limitations. First, these approaches only reach a limited study population. Second, these methods produce low response levels.

Instead of reaching out to invite participants, more institutions are using existing online resources. User-generated Content (UGC), also characterized as “Conversational media”, encourages the publishing of one’s own content and commenting on other people’s content [4]. The role of passive audience has been greatly shifted since the creation of social media, and a growing number of users take advantage of this opportunity to generate independent content. Since UGC is the major form of social media content, Twitter has a great deal of user-generated content and it is still growing. With the amount of user-generated content, there are great opportunities for providing data in information seeking tasks.

Raw Twitter data is too unorganized to be used for market research purposes. Research has been devoted to two major approaches for addressing this problem: 1) automatic machine learning and natural language processing techniques; and 2) visual representations of Twitter data. However, both of them have limitations:

1. Fully computational methods usually have a high computational complexity and require lots of computing time. To get higher accuracy scores, some algorithms require more iterations, such as LDA for topic-modeling. Except for execution time, other algorithms may also need to be tuned for parameters manually or programmatically to fulfill different data requirements.
2. Raw results coming from automatic algorithms are hard to understand from a user’s perspective. Computational results are usually presented as grouped raw data or with computed scores, which are difficult for terminal users to read, manipulate or analyze. Better ways of representing data should be explored to address the difficulties. This leads us to the discussion of how to represent computational results visually.
3. Visualizing fully computation data or relying on user interactions. There is a common question of how much visualizations should depend on computed data. On one hand, since the computed results are not one-hundred percent accurate,

visualizing data fully based on those results may mislead users. On the other hand if the visualization opens all the options for users to manipulate the data on their own, it will be too overwhelming.

4. Visualizations are task-oriented in that one visualization usually serves for a specific task. Better guidelines are needed to suggest how to design different visualizations for different forms of tasks.

## 1.2 Objectives

Motivated by the challenges mentioned above, we developed a visualization system and propose several design guidelines for a visualization system for information tasks.

The thesis aims at exploring the trade-offs in visual analytics of how to combine automatic methods and their visual representation. More specifically, our target is to find the most influential and important Twitter conversations that discuss specific task questions. We designed a two-phased user study to evaluate performance of different combinations of calculated scores and visualized quantified features. The first user study was designed to determine the feature set that corresponds to the informativeness and relatedness of Twitter data. Then we modeled the computation of informativeness and relatedness scores using the determined features. The second user study explored the effectiveness of four different combinations of quantified features and computed scores. Therefore, this thesis addresses the following research questions in regards to the integration of automatic methods and visual features:

1. Which quantified features have more impact on the informativeness and relatedness of Twitter conversations?
2. Which combinations of computed scores and quantified features are more effective for information seeking tasks?
3. How to design the user study for evaluating the effectiveness of visualizing computed scores and quantified features?

### 1.3 Contributions

We present the main contribution of this thesis to tackle the research questions we mentioned at the end of the Objectives section.

1. We designed a two-phased user study to evaluate four versions of visualizations combining automatic methods and visual features. In the first phase, we sampled and labeled experimental data as input to build visualizations. In the second phase of the user study we evaluated and analyzed the performance of four different combinations for automatic methods and visual features for different user tasks.
2. We built an interactive visualization system in four versions that can be used to evaluate visualizations that show a full projection combining automatic methods and visual features.
3. Our study suggests that for browsing tasks, combining automatic methods and visual features performs the best in efficiency.

## Chapter 2

### Background and Related Work

#### 2.1 Data and Structure of Twitter

##### 2.1.1 Studies on Twitter Data

Many studies have been focused on Twitter data since it was created. Institutions and companies explore the potential of Twitter data for promotion and marketing. The benefits of adopting social media for market research and what will be the process of research via social media was described in Poynter's book [3]. Meanwhile, social media started reforming the economy. Studies on marketing and economy involving social media was introduced as "Socialnomics" [5]. This book explains how social media has an influence on people's daily life and economics. Besides business research and marketing, universities and research institutions have also been exploring the potential of using social media data. Researchers from Stanford University show their work on academic social media data [6]. They collected tweets from the top 100 US universities. Their results suggest some interesting statistics on number of applicants, the acceptance rate, the US News and World Report rank, the freshman retention rate and graduation rate. Our work is also based on academic Twitter data, but focuses more on the efficiency of visualizing data in this domain.

##### 2.1.2 Twitter Networks and Conversations

Social media has network structures, induced by the interactions among users. Much research has been done on the network nature of Twitter. The conversation on Twitter is one of the activities that creates Twitter's network structure. Studies has been done on the network structure of Twitter conversations and the major activities on Twitter conversation [7] [8]. Another work [7] explored how interacting with the Twitter network can possibly improve user influence and bring more followers. Their results show that replying(@mention) has the dominant correlation to the number of



new followers and performs much better than the number of tweets one user posted and number of retweets one user gets. Similar research studied the conversations and collaborations via Twitter [2]. Their research suggested that although Twitter may not be a good tool for formal collaborations, the conversational infrastructure enables people to exchange views. Cha et al. [9] measured the influence of Twitter users through following, being followed, conversations and other interactive activities on Twitter. They observed that conversation (retweet and reply) on Twitter implies content value and name value of tweet authors. In our research, we studied several features of Twitter networks in the research mentioned above (e.g. reply network, number of tweets, retweets, tweet author influence) and proposed a general ranking of features based on who has more influence on Twitter conversation content value. We built a few visualizations on the key features and evaluated their performance in a user study.

## **2.2 Text Mining Methods**

### **2.2.1 Feature Selection and Evaluation and Classification**

Feature selection has become the focus in lots of research to deal with dataset with many variables. It potentially facilitate data visualization and data understanding, reduce dimensionally and storage requirements [10]. Many methods have been proposed for selecting the feature set, such as information gain and CHI-square. A previous literature review [11] compared the performance of various feature selections methods for text categorization tasks. During our feature selection process for the data, we experimented on a few effective feature selection methods according to the discussion and analysis in that paper.

Researchers have increasing interests in aspect topic classification [12] [13] [14] [15]. Like many other similar aspect topic classification research, these three papers work on the improvement of topic modeling using popular methods based on Latent Dirichlet Allocation and Latent Semantic Analysis. However, these methods have a relative high computing complexity or even are known to be NP-hard problems [16] [17]. For time-sensitive and real-time data, such as social media data, methods with high complexity have some disadvantages by having longer processing time or require longer

waiting time for users. Along with the growth of social media, interests have grown in classification for social media documents [18]. Twitter, as one of the most popular social media, constantly attracts researchers to study various approaches towards processing its text such as the classification of twitter data. Support Vector Machine (SVM) [19] is one of the most popular and efficient classification methods. Many researchers explored classification using support vector machine on Twitter data [20] [21] [22]. We adopted SVM for our topic classification and score prediction to process our Twitter data, as data input for the visualizations.

## **2.3 Visual Analytics and Cognition**

### **2.3.1 Human Computer Interaction Design Process**

A rich and growing literature considers the area of Human Computer Interactions. Human Computer Interaction is a relatively broad field, so we only review the works that are related to the design of human-centred interactive visualizations. Tory et al. [23] discussed how human factors may affect the design and interactivity of visualizations. They reviewed several visualization principles that support three different approaches for human factor research: user motivated design, user and task-based design and perception and cognition-based design. In order to improve a user's understanding towards the presented data, we adopted the principles of user and task-based design when designing for study and visualization. Those principles include user-centred design flow and cognition support such as parallel processing, grouping, abstraction and organization. Chiang et al. [24] proposed two concepts, interpretation and trust, to assist people in designing and coordinating for interactive visualizations with experts from multiple disciplines. They emphasize user and task centric evaluations during machine learning research. This work mainly focused on analysis and design of visualizations based on user interaction and communicating with users during interactive design phase. In our work, we make use of the ideas from this paper by interacting and communicating with users for requirements, and later focus on how to design the evaluation process of human computer interaction.

### **2.3.2 Design Guidelines for Interactive Visualization**

A related literature introduced the strategies and methodology for study design via a nine-stage design framework for designing user-centred studies [25]. The nine stages fall into three top-level categories: precondition (personal validation), core (inward-facing validation) and analysis (outward-facing validation). In our study, we iteratively practiced this design framework in both building data model phase and building visualization phase.

## **2.4 Design Conventions and Encodings**

### **2.4.1 Visual Encodings**

Researchers on visualizations have long been contributing to visual encodings for visualization tasks. Since colour acts as one of the important visual encoding, Silva et al. [26] reviewed the usage of colour scales in different visualizations. They proposed guidelines for selecting the right colour scale based on data types or task types. In our study and visualization design, colour scale plays an important role for presenting dimensions of data. During our design and implementation, we also considered the rules and resources reviewed by Silva et al. for applying colour scales to our visualization design. Other research studied different forms of visualization encodings, such as glyphs and fonts. A study introduced the procedures of encoding glyphs in visualizations [27]. FatFonts [28] worked as an application for showing how to encode fonts for arts and real life usage. These works described how to encode multi-dimensional data with other visual elements such as glyphs, length, direction and area. In our study, we encode different dimensions of our data with colour, size, brightness and position.

### **2.4.2 Visualization Design Perceptions**

Besides the challenges in how to visually encode data, another issue is that a good designed visualization should meet the existing conventions. Two books [29] [30] described the general perceptions for information visualization. In those books, various visual components, visual patterns and principles have been explained in detail about how to use them in practice from data structure and patterns, display contrast and

constancy to visual encodings like color and shape. Research on a visualization named imMens introduced design conventions via a real-time querying system [31]. Several visual patterns for representing scalable and parallel data were reviewed and then integrated to the design of the real-time visual system. Other research proposed the drawing conventions for sociogram in social network visualization [32]. They explored approaches for the layouts of social network. They inferred the drawing conventions and edge crossings for dealing with network visualizations. During the design and implementation of visual patterns for the visualizations, we referred to these guidelines for displaying data patterns, encoding data in color and shapes and addressing overlapping issues.

## **2.5 Task-oriented Visualization Design**

### **2.5.1 Visual Analytics**

There has been growing interest into the research of visual text analytics. Several visualizations have been proposed by Liu et al., for instance Tiara [33], Storyflow [34] and topic competition [35] around topic visualizations. In the research of Tiara [33], Liu et al. proposed a topic flow visualization for exploring classified topics. This visualization displays time sensitive keywords in a flowchart that uses layers of flow to represent different topics. In the Storyflow project, a visualization named Storyline was built to illustrate the dynamic relationships between entities in a story [34]. They aimed at effective layout storyline visualizations in applications and how to display the relationships among different entities. In the third project of topic competition, three different views were presented for showing topic flow and transition based on Twitter competition data [35]. They presented events in a visualization combining flow view for time data and tag view for context. The three papers by Liu et al. focused on how to improve the performance of data model and layout algorithms. However, their work didn't conduct user studies to validate the performance of visualizations in practice. Other related research includes visualizations for high-dimensional data [36]. In that research, a procedure model of transforming data into visualizations and finally presenting the visualizations to users was proposed. Our work follows on the research agenda proposed by this paper and focus more on the user evaluation.

### 2.5.2 User Tasks for Information Systems and for Twitter

Users tasks act as an important role in user studies on information seeking. Kellar et al. examined how users interact with their web browsers within the context of information seeking task [37]. They conducted studies to ask participants to rearrange a pre-defined task category list. The final refined list of categories includes fact finding, information gathering, transactions and browsing. Besides general information seeking tasks for web browsing, researchers have been exploring the possible user tasks from Twitter data. Studies conducted by Hurlock et al. explored features that may lead to influential Twitter content, which helps to study which features are considered important in Twitter context [38]. User studies were conducted to ask participants to give scores for a group of tweets in three different types of searching tasks: 1) a temporal monitoring task, 2) a subjective choice task, and 3) a location-sensitive planning task. Elswailer et al. did a long term diary study by surveying users for the tasks that motivates users' searching behavior on Twitter [39]. There are eight coding schemes defined for the tasks: standalone informational, monitoring, sentiment/opinion finding, sense-making of conversations/info. filtering, people search, query social network, refinding and boredom motivated. In our study, we extracted four information seeking task categories that focus on academic Twitter conversation data.

## Chapter 3

### Understanding the Features of Tweets

#### 3.1 Introduction for the Two Phased User Study

As mentioned in Section 1.2 of Chapter 1, our research is mainly focus on finding the answers on three research questions:

1. Which visual features have more impacts on the informativeness and relatedness of Twitter conversations?
2. To what degree do users rely on the calculated results (e.g. degree of informativeness)?
3. How to design effective visualizations for information seeking tasks?

To address the questions above, we proposed two phases of user study for our research. In the first phase, participants were asked to rate the relatedness and informativeness scores of Twitter conversations. Based on the analysis of these scores, we built a visualization system in four different versions. In the second phase, participants evaluated these different versions of visualizations. The details of the Phase Two user study will be described in Chapter 4.

In this chapter, we will describe the design of the first phase user study. We will explain how we collected the experiment data. We will then describe how we determined the candidate features of Twitter conversations for this user study. In Section 3.4, we will present the design of the user interface for this study, as well as the study methodology, including the user study protocol for participants, equipment, user tasks and experiment procedures. Finally, we will discuss the study results of this study and the analysis of the data, as well as the research questions related to the design of a visualization system.

## 3.2 Data Collection

### 3.2.1 Data Domain

Academia was the topic domain for our dataset. We collected tweets that were about selected five Canadian universities from Twitter. The five universities are: the University of British Columbia, Simon Fraser University, McGill University, McMaster University and Dalhousie University. We chose these universities for three reasons:

1. Population. These universities have relatively large numbers of faculties and students that ensured there were at least some activities on Twitter.
2. Less Data noise. The search results for these universities are more distinguishable than some other schools. University of Toronto, for example, often results in tweets that talked about events around Toronto, while UBC was more likely to have results specific to the University of British Columbia. Since we are not focusing on improving the searching accuracy, we decided to collect tweets of universities with less noisy data.
3. Location. The five universities are located across Canada, which increases the variety of context in the data.

Our research goal was to explore the visual representations of data for different information seeking tasks. To support an organized, efficient visualization, the first step was to organize the data. Aspect classification [12] [13] is an approach to data organization. Information categorization help humans to memorize and understand knowledge [40]. For instance, users will be more likely to remember and understand UBC with information on scholarship topics than if presented with random information about this university. Therefore, we tried to determine what topics of academic data do people focus on and how to organize and group the information needed for academia.

In order to define the topics for classification of the academic Twitter data, we interviewed domain experts from Dalhousie University. We set up meetings with managers from Dalhousie Communications and Marketing office, Faculty of Graduate Studies and the Registrar’s Office. Our experts supplied us with different tools that

they use to collect data about schools. For example, questionnaires and reports that were designed by the experts from those offices. These materials were designed to evaluate the performance of promotions and enrolments for Dalhousie University. Some reports included statistics such as comparisons among multiple universities across Canada. The original questionnaires and reports are confidential and cannot be published. However, we summarized topics and keywords from those materials and interviews. The summarized topics and keywords are shown in Table 3.1.

Academics	Finance	Campus	Admission
academic	tuition	location	accept
reputation	fee	campus size	admission
graduate	cost	housing	enrollment
undergraduate	scholarship	residency	
research	fellowship	population	
internship		dining	
co-op		living	
course		sports	

Table 3.1: Topics and Keywords

Four topics have been defined: academics, finance, campus and admission. Each of these topics contains a list of keywords. After the topics and keywords were determined, we used the keywords to crawl tweets from Twitter. Twitter conversations were then extracted from the collected tweets as the corpus for our research. We will show the statistics of our data in Section 3.2.2 and describe the techniques for data crawling in Section 3.2.2.

### 3.2.2 Data Processing

The Twitter data was collected from March to the end of April in 2013. The keywords in Table 3.1 were used as input in Twitter search API.

Before we do further experiment and visual presentation, the data needed to be prepared and processed. The procedure of data processing consists of five steps.

1. University keywords crawling. We searched for tweets that contain names or abbreviations of the five universities and stored them in the database.



2. Location tweets crawling. In addition to tweets containing keywords, we also crawled tweets that were posted from each of the campuses. We achieved this by drawing campus boundaries for those schools and collecting any tweets of which geographic coordinates (latitude, longitude) fall into the campus areas. The campus areas include all the campuses for the universities. In each campus, we read the longitude and latitude for the north, south, west and east point, and retrieved all the timeline tweets falling into this rectangle. The idea of collecting geographic related tweets was inspired by the research from Sean [6].
3. User data crawling. After getting tweets data, we started collecting the profile data (i.e. number of followers, number of followings and number of tweets) of Twitter users who posted the tweets in our database.
4. Conversation extraction. For any tweet that is a reply to other tweet, we requested replied-to tweet if it was not saved to our database. In Twitter API, each tweet has a field named *in\_reply\_to\_status\_id* that stores the original Tweets ID. After all the tweets are collected, we found all the root tweets that has more than one replies, and stored all the replies under the data structure of each root tweet in a chronicle sequence. Since we prepared our data in a certain time span, all the further replies were ignored after we stopped crawling data.
5. Aspect classification. We conducted two naive aspect classifications for the conversational data based on the keywords, one for distinguishing universities and the other one for topics. For the first classification, we classified Twitter data into five class, each of which is related to one university. For the second classification, we grouped data based on the topics mentioned in Table 3.1. In each Twitter conversation, we check if the conversation text contains one of keywords relating to the university or topic aspect. If it contains any keyword, the item was labeled as the class of the university or topic aspect. One Twitter conversation can be labeled with multiple universities or topics. These classifications benefit our later procedures of data sampling, which will be described in Section 3.4.

In total, 153,185 tweets were collected during that time with 36,939 users who posted. After data pre-processing and cleaning, 13,460 Twitter conversations were

extracted. The detailed statistics is shown in Table 3.2. Our experiments were conducted based on the 13460 Twitter conversations.

	Before Pre-processing	Data for Conversations
Number of Tweets	153185	-
Number of Users	36939	-
Number of Conversations	-	13460

Table 3.2: Data Statistics

### 3.3 Understanding the Features of Twitter Conversations

#### 3.3.1 Previous Studies on Features of Twitter Conversation

The goal of this phase of our user study is to locate important and influential Twitter conversations. It is crucial to know which features can help to identify significant conversations. One Twitter conversation consists of a set of standalone tweets that come with a time sequence and interactions among them. Therefore, the features of Twitter conversations can be segmented into two parts: 1) features of tweets, and 2) features that reflect conversation structure. Hurlock et al. A user study was conducted to explore features of standalone tweets, whose main goal is to investigate features that can direct users to valuable and relevant information on a range of topics [38]. Three different types of user tasks were defined: 1) temporal, 2) subjective and 3) location-sensitive tasks. The results show that tweets with specific facts are considered useful in temporal and location-sensitive tasks; tweets with trusted and informative links tend to be useful in all tasks; tweets that have location information are useful for location-sensitive tasks. The experiments defined 16 positive features that falls into 5 categories: Twitter content (domain related topics in tweets), reflection on Tweet (entertaining and sentiment), relevant (time and location), trust (user, avatar, link), links and meta tweets (retweeted lots and conversation). Their results show that tweets which have the features of experience, specific information, time, location, trusted author and useful links were more frequently rated as useful tweets. Besides the features of standalone tweets, Twitter conversations have other features because of their network structure, for instance, network-based user influence [9] and the tree structure of Twitter conversations [41].

### 3.3.2 Candidate Features for the User Study

According to the knowledge from previous studies mentioned above and quantifying our own data, we defined 8 features for Twitter conversations. The list of features is shown in Table 3.3.

<b>Content Features</b>
1. Topics (e.g. UBC, campus)
2. Location
3. Urls
<b>Conversation Features</b>
4. User influence score
5. Character length of conversation content
6. Number of users in a conversation
7. Number of tweets in a conversation
<b>User Features</b>
8. User description

Table 3.3: Candidate Features

These candidate features fall into three groups: the features of tweet contents, the features of conversation structure, and user profile.

As a simplified feature set, number of users and numbers of tweets were selected to represent the structure of conversations. In addition to those features we mentioned from previous studies, we added two other features: conversation character length and user profile. We consider conversation character length as a feature because longer conversation tends to contain more information. For example, “2013 International Conference on Social Media and Society Dalhousie University, Halifax, Canada” contains more information than “I’m at Dalhousie”. In the latter example, people can hardly even tell if it means Dalhousie University. The decision on adding user profile as a feature is based on the functionality of Twitter. User profile includes users’ bio information, number of followers and number of followings.

## 3.4 Study Design

### 3.4.1 Introduction

We defined two concept to determine the importance of Twitter conversations: informativeness and relatedness. The informativeness defines as whether this conversation is helpful or informative to present some aspects of university such as research, course or scholarship. For example, I am here in Dalhousie University is less informative than Introducing a new scholarship for graduate students at Dalhousie University. The relatedness defines as whether one conversation is related to topics about universities. For example, Dalhousie Street in Calgary is not related to academic topics although it has the same keyword as Dalhousie University.

The purpose of the Phase One user study is to determine which conversation features have more influence on a user’s perceptions of relatedness and informativeness of a Twitter conversation. It is the first phase of our two-phased user study. In this phase, participants were asked to rate the informativeness and relatedness of a set of twitter conversations to a given topic. We varied the characteristics of the set of tweets according to the selected features that we hypothesize may impact the perception of informativeness and relatedness of those tweets. Based on these results we developed visualizations that employ the most prominent features (Section 3.3) allowing users to identify the information of their interests from a set of tweets.

### 3.4.2 Feature Vector and Data Sampling

As we have a twitter conversation corpus with a size of 13460 conversations, it is not feasible to label all the data. So we sampled a subset of data for labeling. To ensure a balanced distribution of values for each feature in the sample sets, we created a feature vector for each conversation. The set of features in this vector was extracted by quantifying features described in Table 3.3 The feature vectors contains all the calculable features, including conversation character length, has location, number of tweets, number of users, and user influence score. The detailed description of the variables and the value range for each vector is shown is Table 3.4.

For each Twitter conversation, the vector is defined as  $V(f)$ :

$$V(f) = [e_{topic}, e_{univ}, e_{loc}, e_{url}, e_{user}, e_{c\_length}, e_{u\_num}, e_{t\_num}, e_{i\_score}, e_{r\_score}] \quad (3.1)$$

Feature	Description
Topic	1: Academics 2: Finance 3: Campus 4: Admission 0: Other
University	1: Dalhousie 2: UBC 3: SFU 4: McGill 5: McMaster 0: Other
Has location	1: Has location 0: None
Has urls	1: Has urls 0 : None
User influence score	1: Ranging from 1 to 10
Character length of conversation content	Number of characters in an conversation
Number of users in a conversation	2 or more
Number of tweets in a conversation	2 or more
Informativeness score	scale: -2 to 2
Relatedness score	scale: -2 to 2

Table 3.4: Feature Vector and the Range of the Values for Each Feature

$e_{topic}, e_{univ}, e_{loc}, e_{url}, e_{user}, e_{c.length}, e_{u.num}, e_{t.num}, e_{i.score}, e_{r.score}$  correspond to the features described in Table 3.4. Take the following Twitter conversation as an example,

*“McGill students use garden as an ashtray. @ McLennan Library (Humanities and Social Sciences Library) <http://t.co/ibsMjOZI1n>  
@loriekloda To be fair, we can’t be sure it was all due to students. How about ‘McGill smokers’?”*

the feature vector is

$$[e_{topic} = 0, e_{univ} = 4, e_{loc} = 1, e_{url} = 1, e_{user} = 2.64E-5, e_{c.length} = 287, e_{u.num} = 2, e_{t.num} = 2, e_{i.score} = 0, e_{r.score} = 0] \quad (3.2)$$

In this vector,  $e_{topic}$  and  $e_{univ}$  are the results of keywords classification. In this example, there is no matching topic keywords so the value is other ( $e_{topic} = 0$ ). It

took place in McGill ( $e_{univ} = 4$ ).  $e_{loc}$  and  $e_{url}$  were the boolean values of the meta data.  $e_{loc} = 1$  and  $e_{url} = 1$  stand for there are location meta data and url in this Twitter conversation.  $e_{user}$  was computed using page rank [42] algorithm, using following relationship between users as in-links and out-links. The following user graph was built by traversing all the users in our database and connecting any two which have a following relationship.  $e_{c\_length}$  is the count of characters ( $e_{c\_length} = 287$ ) in this text.  $e_{u\_num}$  represents two users ( $e_{u\_num} = 2$ ) involved in the conversation.  $e_{t\_num}$  stands for that there are two tweets ( $e_{t\_num} = 2$ ) in this example conversation.  $e_{i\_score}$  and  $e_{r\_score}$  were targets of classifications which remained zeros during data processing.

In order to evaluate which features from Table 3.4 have most impact on the relatedness and informative of Twitter data, we needed to sample 200 Twitter conversations for four groups of participants to evaluate. First, we manually sampled and labeled 1000 Twitter conversations with informative scores ( $e_{i\_score}$ ) and relatedness scores ( $e_{r\_score}$ ) to ensure that we can sample a random distribution for informative and relatedness scores, avoiding extreme cases which all the samples results have zero informative and relatedness scores. Second, we sampled four sets of Twitter conversations from the labeled 1000 conversations, each containing 50 conversations. These samples were generated by assigning 200 random vectors of features and retrieving the conversations that have the most similar vector values to ensure a randomized distribution.

When the data samples were ready, the next step was designing a user interface that displays all the feature values for Twitter conversations.

### 3.4.3 Study Protocol

#### Study Process

The task is to ask participants to rate twitter conversations on screen. We recruited 20 participants for rating one of the four sampled sets of Twitter conversations. Each set contains 50 Twitter conversations. 20 participants were grouped into 4 groups with 5 people in each group. Since we selected conversational features through a regression model, 200 (4 groups \* 50 twitter conversations) data points were needed to make a credible statistical analysis to counteract bias. Therefore, participants were assigned to four groups which each participant in the same group was presented a same set

of Twitter conversations. Each participant was shown Twitter conversations talking about the five universities. The participant rated the informativeness and relatedness of the conversations to help us determine the impact features. The detailed process and estimated time is described in Table 3.5.

Steps	Description	Approximate time
1. Consent Form(Appendix C)		5 minutes
2. Background Questionnaire(Appendix D)	Participants knowledge about five Canadian universities and campus life.	5 minutes
3. User Study		30 minutes
1) Task description	Researcher introduces the task	5 minutes
2) On-screen Questionnaire (Appendix E)	Rate relevance and relatedness of 50 twitter conversations	40 minutes
3) Post-task Questionnaire (Appendix F)	After rating all the twitter conversations, participants answer questions about the importance of various features.	5 minutes

Table 3.5: First Phase Study Procedures

Step 1, each participants signed consent form that they understand the benefits and risks of the studies.

Step 2, participants filled in the background questionnaires on their basic age, education and English fluency background, and also about their knowledge on the five Canadian universities and Twitter data. This background information was used to classify participants into different research group, so that within each group participants has relatively similar background.

Step 3, each participants was presented with 50 Twitter conversations, and was asked to rate each conversation as well as answering on-screen and post-task questionnaires.

### Study User Interface

The user interface for Phase One user study is shown in Figure 3.1:

Information Panel

### Twitter Conversation

[@loriekloda](#) -> McGill students use garden as an ashtray. @ McLennan Library (Humanities and Social Sciences Library) <http://t.co/lbsMjOZlIn> | at McGill

[@edbilodeau](#) -> @loriekloda To be fair, we can't be sure it was all due to students. How about "McGill smokers"?

**A**

### Location of Conversation

Map Satellite

Map data ©2014 Google, INEGI Terms of Use

### User Profiles

**Lorie Kloda (@loriekloda following:229 followers:568)** : Assessment librarian; Associate editor for @EBLIP journal, PhD

**Ed Bilodeau (@edbilodeau following:450 followers:524)** : Librarian @ McGillU // Web, libraries, academia, technology & other miscellanea

**E**

### Rating Panel

#### Academic Topics

- Research
- Courses
- Funding
- Scholarship
- Campus Life
- Admission
- Career
- Reputation
- Dalhousie
- UBC
- SFU
- McGill
- McMaster

#### Rate Relatedness to Academic Topics

(e.g. "Dalhousie University" is related; "Dalhousie Street in Calgary is not related")

Not related <- -> Completely related

○ -2 ○ -1 ○ 0 ○ 1 ○ 2

#### Rate Informativeness of Topics (Regarding to academic topics)

Not informative <- -> Very informative

○ -2 ○ -1 ○ 0 ○ 1 ○ 2

**Your notes:**

Figure 3.1: Twitter conversation rating page. A: conversation text area; B: topic tags; C: rating board; D: map for twitter location; E: users profile.

21



The conversations text is shown in the top left (Figure 3.1A). Conversations are selected from academic twitter data that we collected on four different topics (academic, finance, campus, admission). Participant were asked to rate the relatedness and informativeness of conversations to academic topics in the top right (Figure 3.1C). Relatedness has a scale from -2 to 2, in which -2 means not related, 2 for extremely related. Scale of informativeness is from -2 to 2 (not informative to extremely informative). A cloud of pre-defined tags are displayed in Figure 3.1B, and participant can click to select all the tags that are related to the current conversation. Figure 3.1D shows the locations of tweets in conversation or users location from their profile. Information of tweet user profiles are displayed below the tags (Figure 3.1E). A text box for note is located on the bottom of page. If participants have any comments they can write down there. A post-task questionnaire of the task (Appendix F) will be shown after participant finishes rating all the tweets.

Each participant performed a set of tasks interacting with the interface in Figure 3.1. They were firstly given a set of twitter conversations to rate the informativeness and relatedness of each conversation to query academic topics. Then, they were asked to fill in a questionnaire on their preference of which information is more useful while making decisions.

Since we selected conversational features through a regression model, 200 (4 samples \* 50 twitter conversations) data points were needed to make a credible statistical analysis. For each sample, there were five participants in order to eliminate bias results from individual participant. Participants were Dalhousie University students, staff or faculty. We required participants to have some knowledge about Canadian universities and Twitter.

## 3.5 Study Results Analysis

### 3.5.1 Introduction to Data Analysis Approaches

We collected three different types of user data for this study: 1) rating scores of informativeness and relatedness; 2) tags of universities and topics; and 3) participants' preference on different features.

1. Rating scores of informative and relatedness. The participants labeled these

scores on the user interface of this study. In each page, participants read the conversation content and checked information on other features of conversation (e.g. user profile, location, urls and so on). Then they gave a rating about how informative or related the selected conversation is to the academic domain.

2. Tags of universities or topics. On the user interface, the participants tagged universities or topics that the Twitter conversation is related to. This data was used for university classification and topic classification in the second phase of the user study.
3. Participants preference for different features. This data was collected from the post study questionnaire. In the questionnaire, the participants were asked to rate the usefulness of each selected features from their own perspective.

We performed two types of data analysis for the collected data: analysis of user preference and feature evaluation.

### **3.5.2 Analysis of Users' preference**

Every participant was asked to rate for the usefulness of each feature for determining informative and related Twitter conversations. Each feature was voted in a 1 to 5 scale (1 - not useful, 5 - very useful). We summed up the votes to calculate a total score for each feature. In our current design, participants voted each feature only once for informativeness and relatedness of conversations at the same time. We should consider in the future design to vote for how each feature impacts informativeness and relatedness respectively. The results of the total scores for each feature is shown in Figure 3.2.

As seen in Figure 3.2, we can conclude that the most useful features from a user's perspective are urls, topic tags and Twitter user profile description. Twitter conversations that contain urls to extended content are considered more useful. Conversations that talk about academic topics are useful for our domain specific tasks. In addition, conversations with a trusted Twitter user, for example, a professor working at UBC, is also likely to carry useful information.

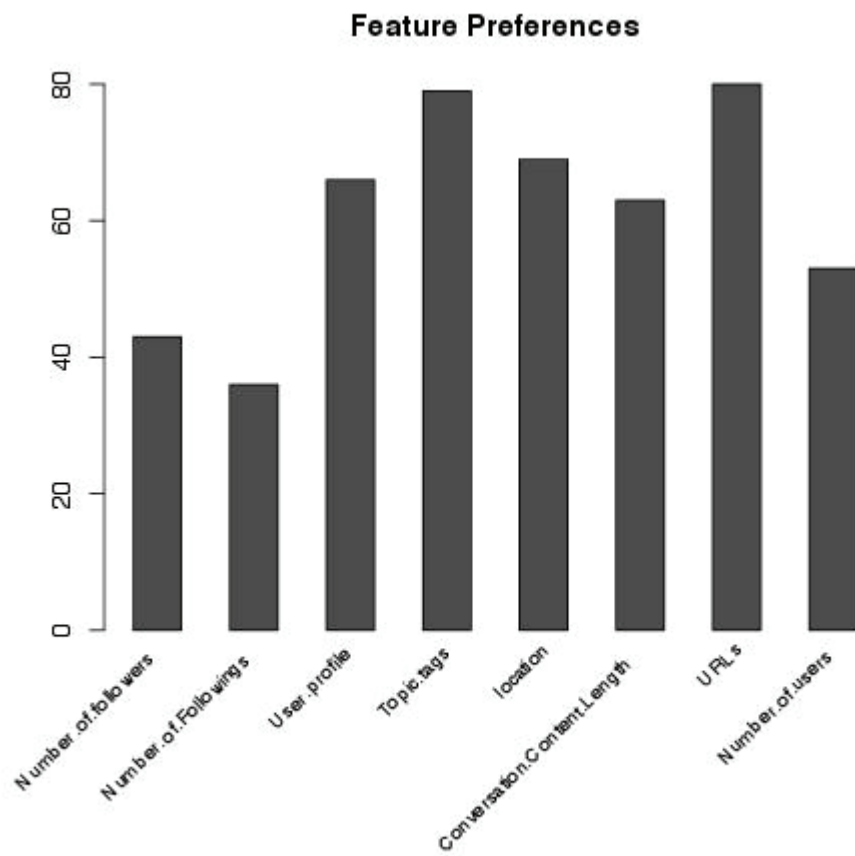


Figure 3.2: User Preferences of Features

### 3.5.3 Feature Evaluation

Besides users' preference in feature set, we computed the performance of feature sets using different feature selection algorithms as an alternative approach to determine the feature set. We wanted to select a subset of features that performs the best in distinguishing different classes for the classification tasks. The input of these tasks were feature vectors (Equation 3.1 without  $e_{topic}$  and  $e_{univ}$ ) of the 200 sampled Twitter conversations with labeled informativeness scores and related scores by participants in the user study. Topics and university features were excluded because that these two features are nominal variables which do not have an intrinsic order. Some feature evaluation algorithms such as gain ratio and chi-square, rank the features by calculating a score to evaluate their performance. Therefore, evaluation methods can be also used to evaluate the performance of individual features. In our study, we used different feature evaluation methods for calculating the performance of the selected features.

Three common methods of feature ranking and evaluation have been used in the data analysis: information gain ratio attribute evaluation, gain ratio attribute evaluation and, chi-square attribute evaluation.

#### Information Gain

Information Gain method evaluates features by calculating their information gain. This is a decision tree based approach. Information gain is defined as the difference in the entropy between one parent node to one of its child nodes. For example, we have a dataset  $D$  and want to evaluate its attribute set  $A$ . We can get the information entropy as:

$$info(D) = - \sum_{i=1}^k p_i \log_2(p_i)$$

Then we dividing attribute set  $A$  and calculate entropy for each of  $A$ 's sub attribute, we can get a set of entropy as:

$$info_A(D) = \sum_{j=1}^k \frac{D_j}{D} info(D_j)$$

Finally the information gain is calculated by:

$$gain(A) = info(D) - info_A(D)$$

Then we can calculate the information gain for each of the features iteratively.

### Information Gain Ratio

Information Gain Ratio evaluates the worth of an attribute by measuring the gain ratio with respect to the class. The computation for gain ratio is one step further to information gain. The difference is that gain ratio considers the balance of attribute distribution. The gain ratio is computed as:

$$GR = gain(A)/H(v)$$

where  $H(v)$  is an intrinsic value that was defined as the entropy of distribution of attributes into decision tree branches. The formula for  $H(v)$  is shown as below.

$$H(v) = - \sum_j p(v_j) \log p(v_j)$$

Previous studies show that gain ratio performances better than information gain [11], yet we still compared results in both methods.

### Chi-square

Chi-square method evaluates features by computing the value of the chi-squared statistic with respect to the class. The formula to calculate chi-square values is defined as follows:

$$\chi^2 = \sum_{i=1}^k \frac{(A_i - np_i)^2}{np_i}$$

Chi-square evaluates the ability of an attribute to distinguish classes.

In our study, the classification tasks are the prediction for informativeness and relatedness score. The training set with labeled score by participants from this study is used to evaluate the selected features. We used Weka [43] to evaluate the performance of features. Weka implemented the three evaluation methods and provides an user interface for manipulating input, output and parameters. The results of the three evaluation methods after normalization are shown in Figure 3.3. As shown in Figure 3.3, three methods give similar results. For informativeness, whether the Twitter conversation has school location information, character length of conversation and number of tweets in the conversation are more important. For relatedness, whether

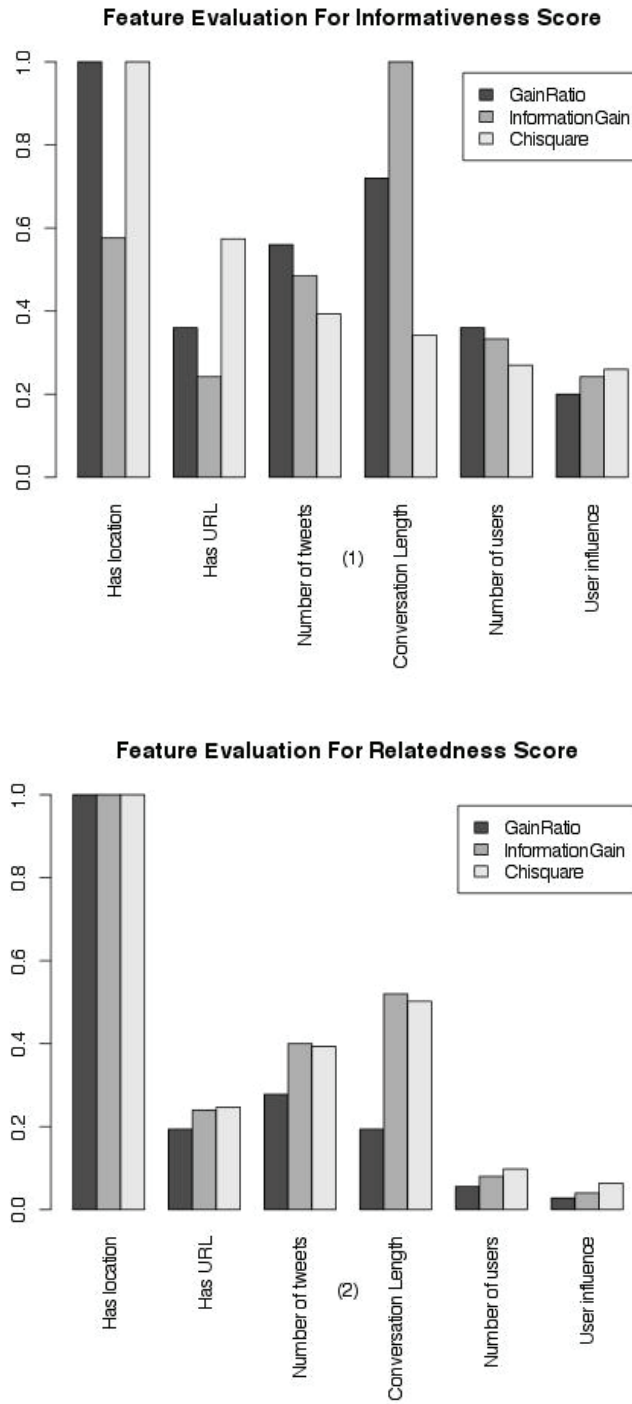


Figure 3.3: Feature Evaluation for (1) Information Score (2) Relatedness Score

the conversation was tweeted within campus area is more important than the rest features.

#### **3.5.4 Results Discussion**

In conclusion, from the participants' perspective, Twitter conversations with url, conversations that talk about specific universities or topics, conversations that have nearby locations (relative to the campuses of five universities) and trusted Twitter user profile are more helpful features than the number of tweets, number of users. Based on featured evaluation algorithms, whether the Twitter conversations show its located in and around those school campuses and length of conversation content (characters or number of tweets) are more helpful for predicting the informativeness of one conversation on academic topics. The conversations that are within campus areas are more related to academic topics. The output of the first phase user study provided the candidate features, the classification and score results to support the two sources of data which can help users determine the importance of Twitter data: 1) the quantified features that can be presented and visualized to display to users and 2) the informativeness and relatedness scores that can be calculated based on the quantified features.

We summarized three research questions related to the design of a visualization system for academic Twitter conversations.

1. Is it necessary to present the first source of data (quantified features) to users while the second source of data (informativeness and relatedness) was calculated based on the first one? If it's going to be presented, how to visualize it?
2. How much would users trust the calculated informativeness and relatedness scores?
3. Which source of data or combination of both sources should be presented to users to help them with information seeking tasks effectively?

## Chapter 4

# Designing an Interactive Visualization of Twitter Conversations

As we mentioned in Chapter 3, we split our user study into two phases. The first phase described in Chapter 3.3 explored key features for identifying informative and relative Twitter conversations. The second phase of the user study, which we describe in this chapter, is to evaluate different versions of visual representations for user tasks. In Section 4.1, we describe the main research question whether we should present to users the quantified features or calculated relatedness and informativeness scores. We define the scopes of visual features and computed results respectively and gives some examples about different cases.

We introduce the procedures of data processing in Section 4.2. The methods we used for classification and scores prediction will be described and we will elaborate how the data was used to support our visualizations. Section 4.3 is the description for the design of our visualizations. Details about how we transformed visualizations from the research challenges and results are presented. In Section 4.4, we are going to describe the design of our user study to evaluate the visualization system we built.

### 4.1 Visual Features vs. Automatic Methods Trade-offs

At the end of Chapter 3, we summarized a few questions from an analysis of the results for the phase one user study. The major question we want to investigate is what are the trade-offs of representing the visualization of quantified features or representing the computed results from automatic methods. To be more specific, there are two sources of results which users can rely on to determine the importance of Twitter data. One source consists of a set of information such as number of users in a Twitter conversation, user influence of the blogger who tweets. Another source is computed scores of informativeness and relatedness which are not 100 percent trusted. Users have the options of not using any of the two sources, using one of the sources, or



using both sources. The problem came to which option is more effective.

As we all know, there is no guarantee for automatic methods, such as classifications, score prediction and clustering, to perform with a one-hundred percent accuracy. This is especially true when dealing with text mining problems. Text mining methods have to consider the ambiguity in languages and bias in individual perspective. As a result, representing the data computed by automatic methods may be misleading. On the other hand, if we put aside the computed results completely, and display only some selected features of data, it may end up requesting an overwhelming work for the user to manipulate and get insights of the data. Therefore, our task was to investigate the trade-offs between representing the visual features and computed data, and explore the guidelines of how different combinations work in different task scenarios.

#### **4.1.1 Features Definition and Examples**

We define visual features as the built-in properties of data or the attribute values we can get through simple calculations. In our data, some example features are the length of a Twitter conversation, the number of users in a conversation, user influence score and whether a conversation contains url information. In this case, we can get the length of a conversation through observation, while calculating user influence score simply based on the network properties. By representing these features, we mean representing the values in some straightforward visual scales. For instance, the scale of numeric results can be represented visually as different sizes, distances or brightness. There are some conventions [26] [29] [30] for visual encodings, and we will introduce how we adopted those conventions and encoded our visualizations in Section 4.3.

#### **4.1.2 Automatic Methods and Examples**

The definition of automatic methods is the machine learning or Natural Language Processing methods we used for text mining tasks. Some typical methods are text classification, score prediction and text clustering. In our study, we used the automatic methods to classify universities and topics, as well as predicting for the informativeness and relatedness score of Twitter conversations. The methods types and

corresponding tasks are shown in Table 4.1.

Tasks	Type
Classification for the Universities	classification
Classification for Topics	classification
Score Prediction of Informativeness	score prediction
Score Prediction of Relatedness	score prediction

Table 4.1: Method Tasks and Type

Now we have two aspects of attributes to evaluate: visual features and automatic methods as defined in Section 4.1. The way we evaluate the effects of these two aspects is to integrate different combinations of the two aspects into four versions of the interface, which is shown in Table 4.2. We will describe implement of these four

Version	Description
Version One	NO visual features + NO automatic methods
Version Two	NO visual features + automatic methods
Version Three	Visual features + NO automatic methods
Version Four	Visual features + automatic methods

Table 4.2: Four Versions Of Visualization System

versions in Section 4.3 and the evaluations of four versions in Section 4.4.

## 4.2 Processing Data to Support Visual Representation

In this chapter, we are going to describe how we process our data for different data tasks (Table 4.1) using classification and scores prediction.

### 4.2.1 Twitter Conversation Classification

It is difficult to help performing information seeking tasks on unprocessed Twitter conversations alone. A common way of processing data is data classification. In our data, two classifications were performed on the data. First, we classify it into five universities. For each university, there are four sub-categories of topics: academics, finance, campus and admission. The goal of the second classification is to classify data of each university into four topic sub-categories. Classification is a very important task in data mining area. The aim for a classification is to learn the model or function

for projecting data into a certain class. In that way, the model can be used for future prediction. Support Vector Machine (SVM) [19] is one of the popular supervised state-of-the-art classification methods. It achieved the classification by determine the hyperplane which has the biggest margins to the nearest support vectors. Our training set is the labeled data from the first phase of user study, in which participants were asked to tag universities and topics for every conversation. The inputs are the feature vectors of conversations that were computed previously for the data sampling in the first user study, combining values of the term frequency-inverse document frequency (TF-IDF) of each conversation. Output results are the classes for universities or topics.

### C-Support Vector Classifier

We were using c type SVC to classify universities and topics. Giving training set  $x_i, i = 1, \dots, l$ , and target vector  $y_i (y \in \{1, 2, 3, 4, 5\}$  for universities,  $y \in \{1, 2, 3, 4\}$  for topics), C-SVC solves the optimization problem as follows [19]

$$\begin{aligned} \min_{\omega, b, \xi} \quad & \frac{1}{2} \omega^T \omega + \sum_{i=1}^l C \xi_i \\ \text{subject to} \quad & y_i (\omega^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, l, \end{aligned} \tag{4.1}$$

Usually we solve the following problem (Equation 4.2) as well due to high dimensionality of  $\omega$ .

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T Q \alpha + e^T \alpha \\ \text{subject to} \quad & y^T \alpha = 0, \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, l, \end{aligned} \tag{4.2}$$

where  $Q_{ij} \equiv y_i y_j K(x_i, x_j)$ , and  $K(x_i, x_j)$  is the kernel function. The kernel function we used to classify universities and topics were Radial Based Function (RBF) and linear respectively. RBF is defined as in Equation 4.3.

$$K(x_i, x_j) = \exp(\gamma \|x_i - x_j\|^2) \tag{4.3}$$

$\gamma$  was set to 0.091 for university classification. After Problem 4.2 is solved, the

optimal  $\omega$  satisfies

$$\omega = \sum_{i=1}^l y_i \alpha_i \phi(x_i) \quad (4.4)$$

and the decision function is

$$\text{sgn}(\omega^T \phi(x_i) + b) = \text{sgn}\left(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b\right) \quad (4.5)$$

#### 4.2.2 Score Prediction

We calculated two types of scores: informativeness and relatedness. The training sets for these two scores are as labeled for informativeness and relatedness by our participants in the previous user study. We reused the training sets and feature vectors as input for SVM to perform a regression task for scores prediction.

#### $\epsilon$ -Support Vector Regression

We were using  $\epsilon$  type SVR to classify universities and topics. Given training set  $x_i, i = 1, \dots, l$ , and target vector  $z_i$ , the standard form of Support Vector Regression [44] is as follows

$$\begin{aligned} \min_{\omega, b, \xi, \xi^*} \quad & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^* \\ \text{subject to} \quad & \omega^T \phi(x_i) + b - z_i \leq \epsilon + \xi_i, \\ & z_i - \omega^T \phi(x_i) - b \leq \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1, \dots, l, \end{aligned} \quad (4.6)$$

Usually the following problem (Equation 4.7) becomes dual problem.

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \sum_{i=1}^l (\alpha + \alpha^*) + \sum_{i=1}^l z_i (\alpha - \alpha^*) \\ \text{subject to} \quad & e^T (\alpha - \alpha^*) = 0, \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l, \end{aligned} \quad (4.7)$$

where  $Q_{ij} \equiv K(x_i, x_j)$ , and  $K(x_i, x_j)$  is the kernel function. The kernel function we used to classify universities and topics were RBF (Equation 4.3).  $\gamma$  was set to 1 for both classifications.

After Problem 4.2 is solved, the approximate function is

$$\sum_{i=1}^l (-\alpha + \alpha^*) K(x_i, x_j) + b \quad (4.8)$$

### 4.2.3 Data Support for Visual Features

In Chapter 3, we selected a list of eight features (Table 3.4 except for the two scores) to investigate their performance both from users' perspective and the feature evaluation. The list of features and the calculating methods are described as follows:

1. Number of characters in the conversation. This doesn't include mention notation (@username), urls or tag notations (#tag).
2. Number of tweets in one conversation.
3. Number of users in one conversation.
4. User influence score. This score is calculated based on the page rank value from the network of our conversation data.
5. University. One of the five universities appears in the conversation.
6. Topic. One of the four topics (or other topics) the conversation belongs to.
7. Location. Whether conversation has location information in and around university campuses.
8. URLs. Whether the conversation contains url to external website.

## 4.3 Design of Visualizations

### 4.3.1 Design considerations

From the results and discussion from the user study on key features to identify important Twitter conversations, we summarized several design questions:

1. *What are the properties that need to be visualized?*

In general, we divided the properties into two groups: 1) properties for visual

Visual Features	
Numbers of characters	
User influence score	Automatic methods
Number of tweets	Informativeness score
Number of users	Relatedness score
URLs	
User profile	

Table 4.3: Visual Properties for 1) visual features and 2) automatic methods

features; 2) properties for automatic methods. The details of properties in each group are shown in Table 4.3. The finalized visual features are slightly different from the results of the previous user study to serve for a more focused and simplified study model. There are three reasons for choosing these features: 1) they can be easily quantified; 2) the results of the first phase study suggest the importance of these features; 3) a simplified list of features is less confusing for users. Users may feel confused if they have to manipulate the combinations of over ten different features. How many features would be the appropriate size and effective could be interesting to study as future work. Except for the informativeness and relatedness scores, we did classifications for universities and topics which are also automatic methods. The results were used to help refine the information, and were not evaluated or compared with visual features in our current design.

2. *How to visually represent the values for each property?*

Each property should have a unique visual interpretation so that they can be distinguishable from each other. We will describe the visual encodings for each property later in this chapter.

3. *What are the task scenarios of the visualizations?*

No visualization can be universal for all the tasks. We categorized information seeking tasks into four categories: 1) Information gathering, 2) opinion, 3) fact searching and 4) user profile related tasks.

4. *How many visualizations do we need?*

We need four sets of visualizations to evaluate different combinations of visual

features and automatic methods. Two variables needs to be evaluated: visual feature and automatic methods. In each combination the two variables have boolean values (with or without). Therefore, four sets of visualizations were needed to make an universal set.

### 4.3.2 Visual Encodings

#### Visual Encodings for Features

In the list of visual features, the number of characters, number of tweets and number of users have numeric values, while urls and user profile are text format. For the two textual features, we decided to represent them as they are originally in the Twitter conversation. The context of Twitter conversations is accessible for users, and they can click web links of users' profile and urls to check profiles on Twitter or browse page from the url of a website (Figure 4.1A). For the four numeric features, we project them onto a two-dimensional axis, each time with two features (Figure 4.1B). There are two reasons for this visual projection. On one hand, it is difficult for humans to understand multi-dimensional data, and low dimensions of data is easier to perceive and suit for visual scalability [31]. On the other hand, the combinations of features may provide extra insights. For example, a long conversation posted by two users back and forward may be more personal than a conversation with equivalent length posted by multiple irrelevant users. To display information in a straightforward way while showing the collaborations between features, two-dimensional projection is a good way to benefit on both sides. Number of characters for a conversation was set to be the default selection, and users could switch to number of tweets, number of users and user influence score alternatively. The order followed the feature selection results from Chapter 3.3 which suggests number of characters for a conversation has more impact on the informativeness and relatedness of Twitter data.

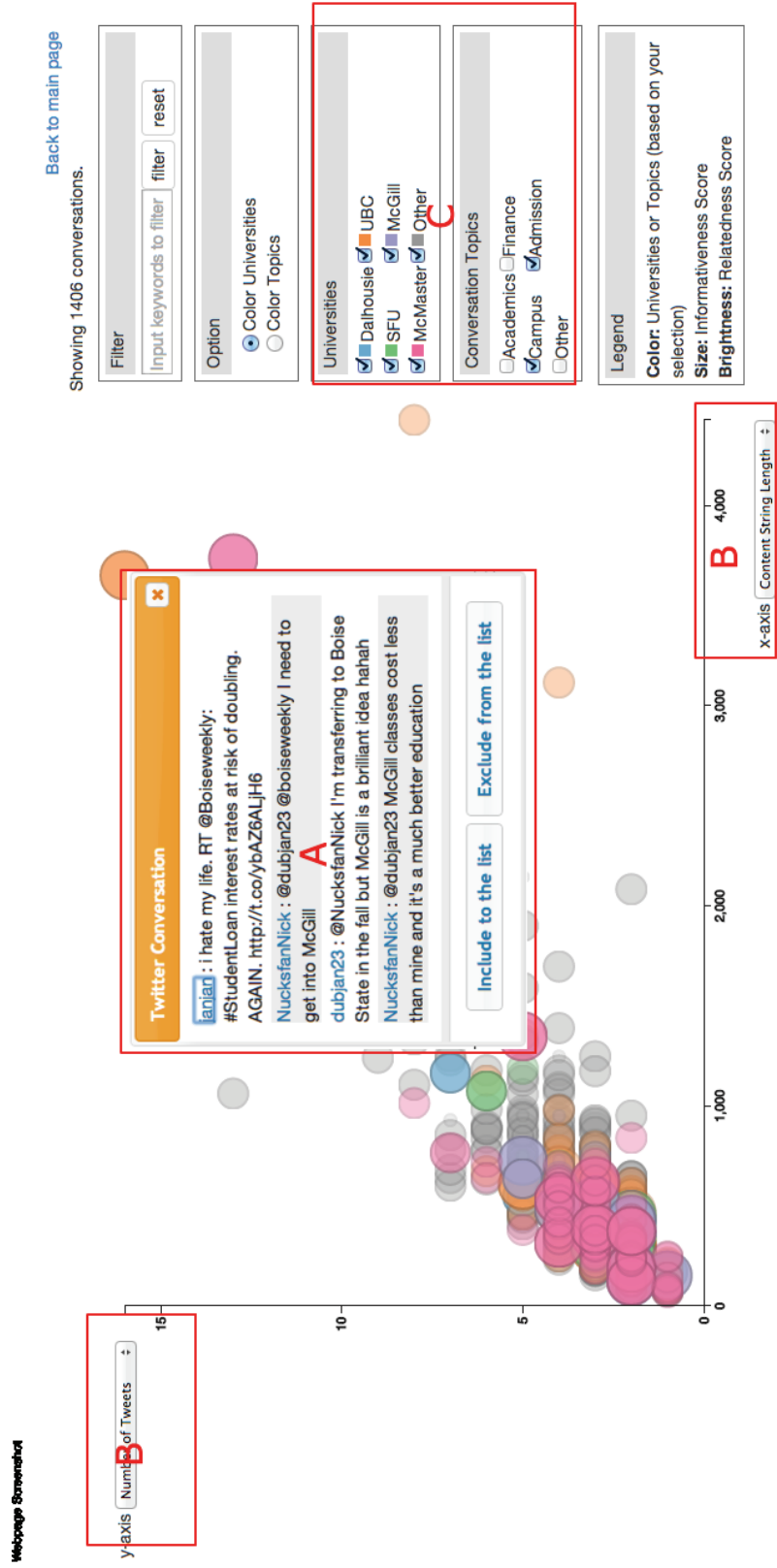


Figure 4.1: System Screenshot (Version 4). A: Conversation Popup; B: Axes; C: Universities and topics.



## Visual Encodings for Computed Results

For the classifications of universities or topics, we used different colors to display the classes. Color is conventionally encoded for grouping [30]. Since there are two categories of classes, we cannot simply distinguish classes using different colors. Instead we need an extra approach to represent which type the colors stand for. We propose two alternative approaches.

1. First clustering, then coloring. We firstly do a clustering data using a palette for their classes. For example, clustering the data into clusters representing academics, finance, campus, admission and other topics respectively, then in each cluster, displaying different universities with different colors.
2. Color one type at a time. In this approach, the data is colored with one type at a time, and users have the option to decide which type to color. Users have the option to color universities or topics at a time in the cases that whether they want to distinguish universities or topics. The consideration for using color for both universities and topics instead of using colors with shapes is that in two of the versions, the size of the nodes varies with informativeness score and using different shape would add complications for comparing the size differences.

The two approaches were adopted in all versions of our visualization system.

The unit of this data is Twitter conversation, so the visual unit is also a conversation. Each conversation is represented as a node with two attributes for each node: size and brightness. Size and brightness are conventionally encoded for numeric values [30]. In our visualizations, informativeness score is encoded as the size of a node. The bigger the size, the more informative the conversation is. Relatedness is encoded as brightness. The brighter the node represents the more related conversation.

### 4.3.3 Four Versions of Visualizations

Because of the combinations of visual features and results computed by automatic methods, we built four versions of visualization system. The variable properties for visual features are: number of characters for a conversation, user influence score,

number of tweets and number of users. We didn't consider those the textual features (urls, user profile, topics and universities) as variables for the current system. Urls and user profiles were presented the same in every version as meta data for Twitter conversation. The information of universities and topics keep the same in every version for the current stage of the study as filter functions. We may consider the visual representations of non-numeric features as a future work. The variable properties for automatic methods are: informativeness and relatedness scores. Each version consists of two views: the general view and the detail view. The general view shows the number of Twitter conversations for each category in a matrix-like bar chart. The detail view displays conversations as a collection of nodes. For each question, participant was first shown general view with general statistics on the dataset, and then detail view was displayed by clicking through one of the category block.

### **Version One**

The visualizations of version one system is shown in Figure 4.2. It doesn't contain any information about the four variable visual features or scores of informativeness and relatedness. The classifications of universities and topics are represented by the colour of nodes and the colour of cluster stroke respectively. The size of nodes stays the same for each conversation.

### **Version Two**

the visualizations of version two system is shown in Figure 4.3. This version adds the information for informativeness and relatedness on top of version one. In the general view, the brightness of colour for each bar shows the score of informativeness. The darker the colour, the higher the informativeness scores are. Since general view is showing statistics for a set of conversation data, we didn't come with a solution for displaying relatedness score at the same time. In the detail view, size and brightness are added to represent the informativeness and relatedness respectively.

### **Version Three**

Version three doesn't have information for scores. It adopted two axes to show the values of visual features. Users are given the options to choose the combinations

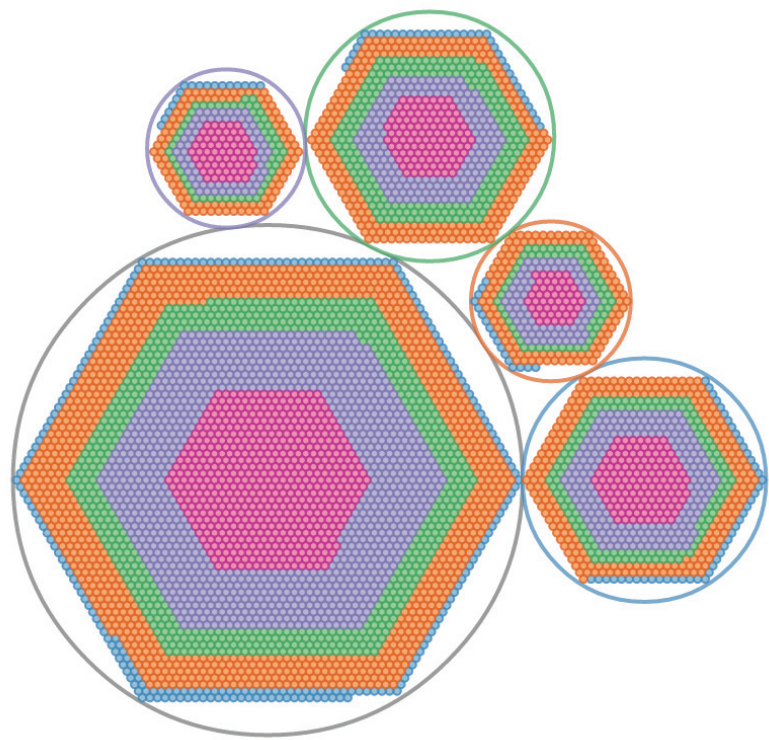
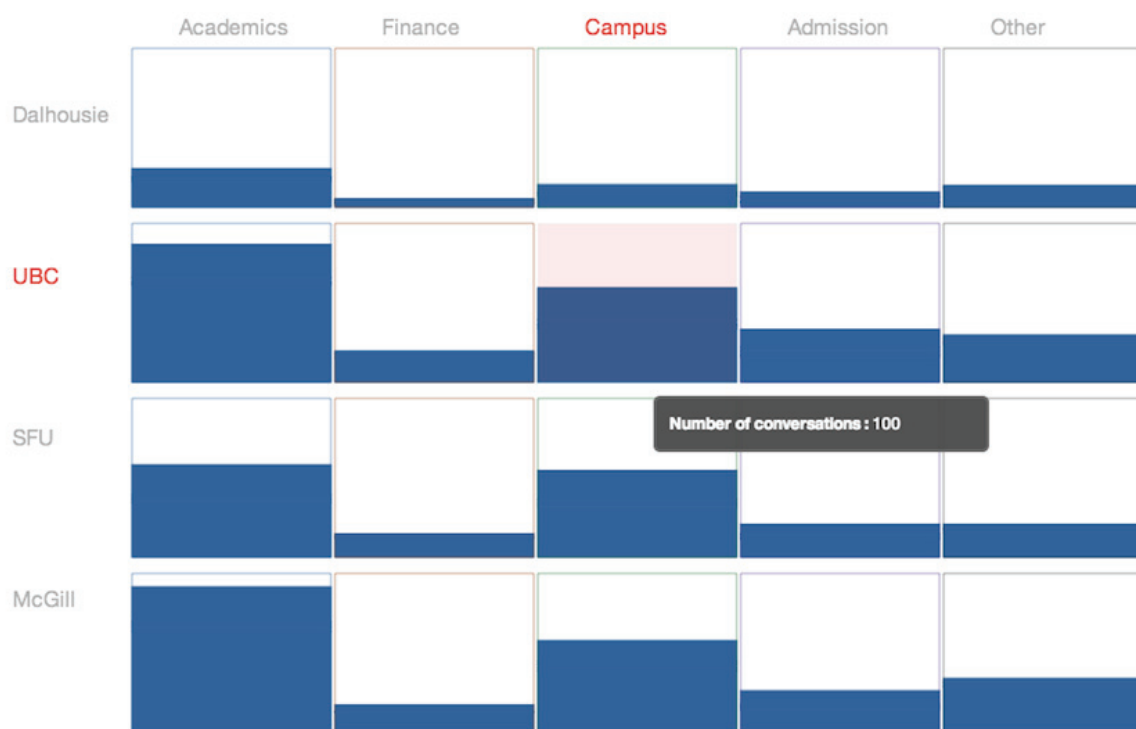


Figure 4.2: Version one: (1) general view (2) detail view



Figure 4.3: Version two: (1) general view (2) detail view

of which pair of features to observe. The interface for version three is shown in Figure 4.4.

## **Version Four**

Similar to version two, the fourth version added the information of the informativeness and relatedness scores on top of version three visualizations. The see interface for version four is shown in Figure 4.5.

### **4.4 Design of User Study**

#### **4.4.1 Grouping Participants**

We recruited 48 participants for this user study. Most participants were recruited within Dalhousie University. Among them, more than half of the participants were younger than 25, 30 people have bachelor degree and above, mostly study computer science, 30 of the participants were male students and more than half of them used Twitter at least once.

Since we have four different versions of system, the minimum number of people needed to possibly show all the cases is  $\text{factorial}(4) = 24$ . We doubled this number to leave spaces for more cases. Participants included Dalhousie University students, staff or faculty. We required participants to have some knowledge about Canadian universities, Twitter and have experience using interactive systems. We assigned each participant to a version of system so that we can have 12 participants for each version of our system. We asked our participant to fill in a background questionnaire so that we can group participants with similar background in one group (which contains four people who have been assigned four different versions) to balance the performance between groups. For example, we grouped people with similar English proficiency in one group so that the performance will not likely to be affect too much by their reading speed. In this way, for instance, version one from better English group might performs better than version one with less fluent English group, but the relative difference between version one and two within the same group might be similar. The final results were sums for each version. The purpose of this grouping is to reduce variations of personal factors that may affect the performance (such as discipline,

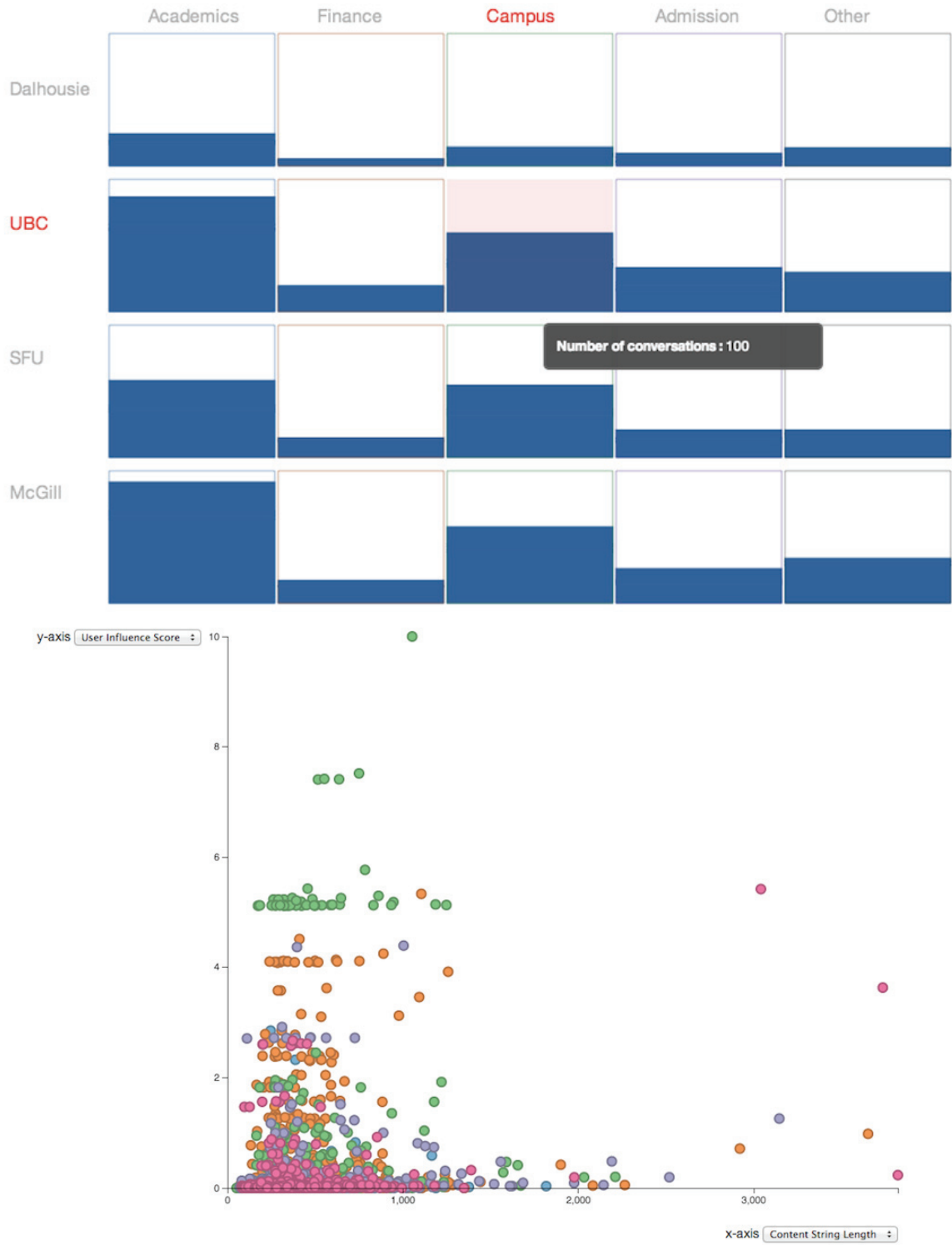


Figure 4.4: Version three: (1) general view (2) detail view

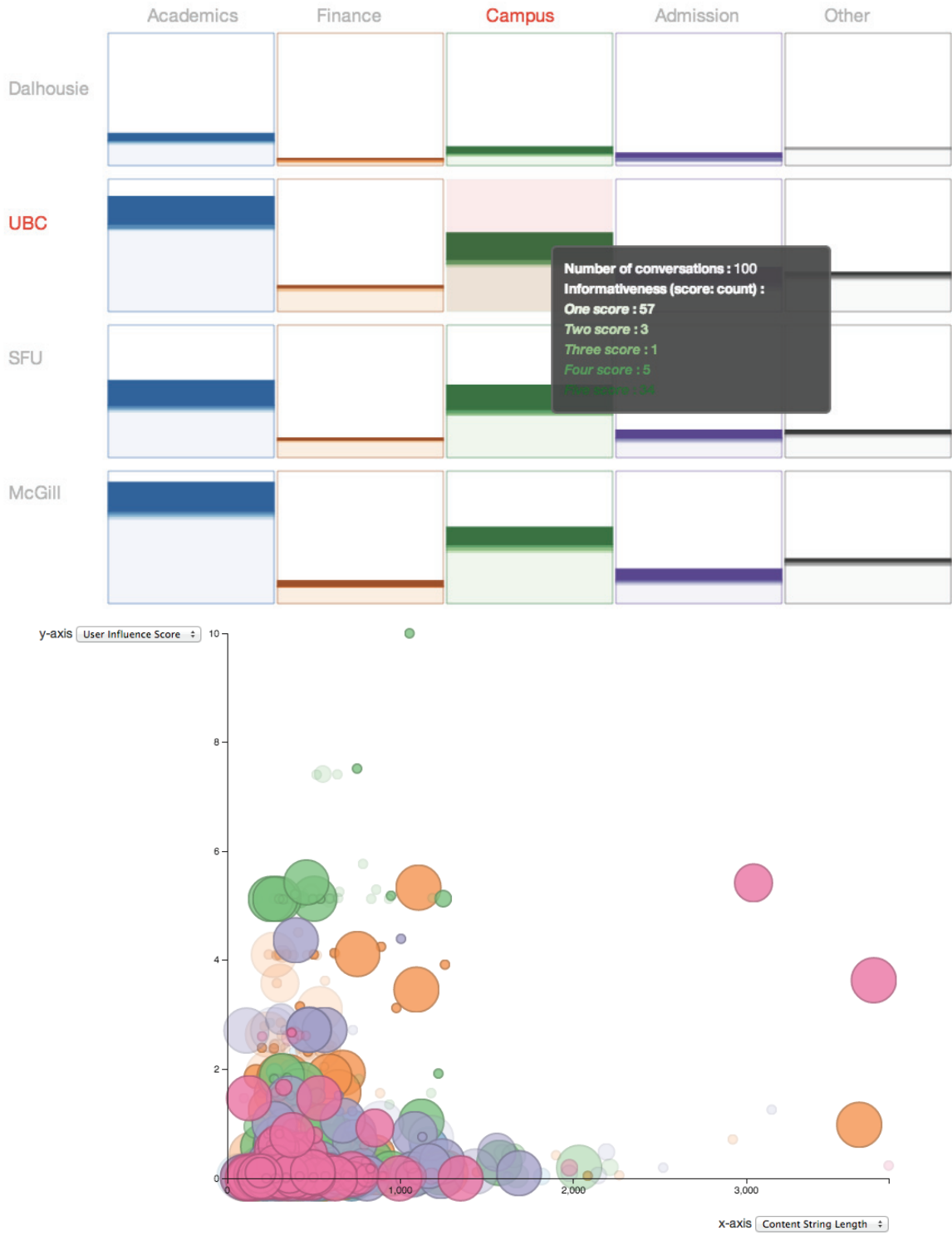


Figure 4.5: Version four: (1) general view (2) detail view

education level, language proficiency and so on) so that we can focus on the factors we want to evaluate (response speed, accuracy, moves of interaction and so on).

#### 4.4.2 Data Collection

Three data collection methods were used during this study: system log, background questionnaires, and post-task questionnaires.

1. *System Logs*

We logged the position of mouse clicks and positions, and all the interactivity participants did when they performing tasks.

2. *Background Questionnaire*

The questions (Appendix L) ask the participants about their knowledge of top universities in Canada, twitter usage and experience in interactive systems. These questions help us to know more about participants experiences and preferences.

3. *Post-task Questionnaire*

The post-task questions (Appendix M) ask the participants to rate the tweets they selected. This will help evaluate our system.

#### 4.4.3 Study Protocol

The procedure of study is described in Table 4.4.

#### The Tasks

We performed a between subjects user study to evaluate the efficiency of the different versions for our system. Each participant performed tasks on one assigned version out of four versions of our system.

#### Four Different Versions of Our System

The aim of our system is to evaluate the two factors that we defined for our system: 1) calculated relatedness and informativeness scores; 2) selected visualized features. We have calculated relatedness and informativeness scores for each of the Twitter



Steps	Description	Approximate time
1. Consent Form		5 minutes
2. Background Questionnaire	Participants knowledge about Canadian universities, campus life and experience with interactive systems.	5 minutes
3. Training Sessions		10 minutes
4. Tasks		95 minutes
1) Task description	Researcher introduces the tasks	5 minutes
2) Interacting with one assigned version of the proposed system	Collecting Twitter conversations with our system for different tasks.	90 minutes
5. Post Study Interview Questions		5 minutes

Table 4.4: Second Phase Study Procedures

conversation in our database. The select features includes content length of the conversation, number of tweets in one conversation, number of users in one conversation, user influence scores, whether the conversation has urls. Our four versions of the system are developed based on a full factorial of these two factors.

1. *Basic View (Version One)*

In the basic version, participants are shown a basic interface that doesnt have any information about scores or features.

2. *Scores Only (Version Two)*

In this version, the interface shows score related views.

3. *Features Only (Version Three)*

In this version, the interface only shows features related views.

4. *Scores + Features (Version Four)*

In this version, participants are shown interfaces that combines scores and features views.

Each participant was only shown one version during the whole study. Then they performed the same set of tasks aimed at answering questions on academic topics with the help of our proposed system and Twitter conversation data. We pre-selected tasks that could be finished within 15 minutes time window.

### **Between Subjects Design**

We adopted between subjects design for this user study. Four participants perform for a set of tasks. Each of the participants only dealt with one version of our system. The participants were grouped according to their background information so that we can balance variations of the results due to the differences between participants.

### **Task Definition**

Based on the previous studies of information seeking tasks for web [37] and Twitter [39], we decided to evaluate on four different categories of tasks for information seeking on academic Twitter data:

1. Information Gathering (e.g. Top 5 useful conversations about course in UBC)
2. Opinion (e.g. School ranking based on admission conversations)
3. Fact Searching (e.g. Undergrad tuition fee at McGill)
4. User Profile Related (e.g. 3 active users talking about research news)

Participants were asked to find answers to 8 task questions from these four categories (see Appendix H) interacting with our system. These tasks were performed under a time limit (15 minutes per each task). For each task, participant were asked to search and browse the informative and related Twitter conversations they found, and include the conversations in the answer list by clicking the “include to the list” button shown in Figure 4.1A.

The evaluations of tasks performance were based on time of completion, amount of interactions (clicking, dragging, and so forth) and the content of questionnaire answers. The analysis of study data will be introduced in the following chapter.

## Chapter 5

### Evaluation of Visualization Designs

#### 5.1 Introduction

In this chapter we present the analysis conducted for our Phase Two user study (Evaluating Effectiveness of an Interactive Visualization System on Twitter Data). This chapter starts from research questions that our following analyses needs to answer in Section 5.2. Section 5.3 describes the data we use for analysis. Section 5.4 shows the variables we defined for the analyses. In Section 5.5, we will show the results of One-way Analysis of Variance (ANOVA). We will show some further analysis in repeated measures ANOVAs in Section 5.6. Other data analysis will be described in Section 5.7. Discussions on four categories will be described in Section 5.8. In Section 5.9, we will discuss designs insights based on this research, and possible improvements for the user study and visualizations.

#### 5.2 Research Questions

In the Phase Two user study, we evaluated four systems in complete combinations of visual features and predicted scores under four categories of information seeking tasks. The analysis of this study results were trying to answer the following research questions:

1. Which combinations of features and scores are more efficient and perform the best in information seeking tasks?
2. What are performance of these combinations on individual category of tasks?
3. Which tasks have better performance and why?

### 5.3 Description of Data

The data we use in the analysis comes from two major sources: 1) user interactions with our visual system, and 2) post study questionnaires. Different types of data are used to evaluate the effectiveness and efficiency of four versions of visual system. There was a set of Twitter conversations as gold standard for each task question. The gold standard Twitter conversations were manually selected by going over all the conversations in the corpus and selecting all the conversations that are related to the question. The types of data we use for evaluating efficiency are as follows:

1. Total clicks. The number of clicks on each Twitter conversation node that participant performed to find answer to each question. Total clicks were collected by recording the node click events during participants' interaction with the system.
2. Time. The time each participant spent on each question.

The types of data we use for evaluating effectiveness are as follows:

1. Hits. The number of clicks when participant hit the answer from gold standard.
2. Precision. The precision of answers comparing to gold standard. In our analysis, precision suggests how many correct results out of participants' selection were returned.
3. Recall. The recall of answers comparing to gold standard. In our analysis, recall suggests how many correct results out of gold standard set were returned.

The raw results we collected from the user study are shown in Appendix Q, R and S. We performed normalization test on each type of data. All of the variables are normally distributed and can be used for further analysis.

### 5.4 Evaluation Variables

Several ANOVAs and repeated measure ANOVAs have been conducted to evaluate our study results and answer the research questions. ANOVA analyses were used to analyze the overall performance difference among four versions of system and

difference in four task categories. Repeated measure ANOVAs were conducted to show the interactions of four versions and four task categories.

In ANOVA, Analysis data are grouped into two types: dependent variables and between-subjects variables. Dependent variables are usually defined as being closely connected to the measuring instrument, or in other words, variables that can show performance. Between-subjects are the variables that set the differences between participants.

In repeated measures ANOVAs, besides for between-subjects variables, each dependent variable is divided into  $n$  levels to explore the variations for one participant. Therefore, repeated measures ANOVAs are conducted to study the differences of performance between participants and within each participant.

#### **5.4.1 Dependent Variables and Between-subjects Variables in ANOVA**

There were different dependent variables:

1. Total clicks. The number of clicks on each Twitter conversation node that participant performed to find answer to each question.
2. Hits. The number of clicks when participant hit the answer from gold standard.
3. Precision. The precision of answers comparing to gold standard.
4. Recall. The recall of answers comparing to gold standard.
5. Time. The time each participant spent on each question.

Among these five variables, total clicks and time are measurements for efficiency, while hits, precision and recall are measurements for effectiveness. Total clicks are measured by how many clicks one participant performs on clicking nodes to open Twitter conversations. Time is the time span for each participant to complete each task. Hits stands for the clicks when participant hit nodes that contain answers. Precision and recall are the performance of participants answers comparing to gold standards.

We have conducted our measurements with three different between-subjects factors:

Table 5.1: Values of  $H_{feature} * H_{score}$  to Version Number

		$H_{score}$	
		0	1
$H_{feature}$	0	Version One	Version Two
	1	Version Three	Version Four

1.  $H_{feature} * H_{score}$ .  $H_{feature}$  ( $H_{feature} \in \{0, 1\}$ ) is a boolean value shows whether the visualization is with or without visualized features.  $H_{score}$  ( $H_{score} \in \{0, 1\}$ ) is a boolean value shows whether the visualization is with or without scores.  $H_{feature} * H_{score}$  ( $H_{feature} \in \{0, 1\}, H_{score} \in \{0, 1\}$ ) is the production of these two values.
2. Version. One of the four versions of visualizations.
3. Task Category. One of the four task categories.

In order to study both the production effects and individual effects, we analyzed the combining effects and individual effects of  $H_{feature} * H_{score}$  at the same time. The relationships of  $H_{feature} * H_{score}$  to versions is shown in Table 5.1. Version refers to four different versions of our visualizations. Task category refers to the four different categories of task questions. Task category was used to measure the difference between tasks, which is different from the former two variables.

#### 5.4.2 Within-subjects Variables and Between-subjects Factors in Repeated Measures ANOVAs

For repeated measures ANOVAs, the between-subjects factors consist of 1)  $H_{feature} * H_{score}$  and 2) Version. The projection of  $H_{feature}$  and  $H_{score}$  resulted in four versions. However, since we were also interested how  $H_{feature}$  and  $H_{score}$  would perform individually, we did analysis to both  $H_{feature} * H_{score}$  and version respectively. The task category was used to determine within-subjects levels. All the dependent variables are split into four levels based on the task categories. For instance, total clicks are split into total clicks\_category\_1, clicks\_category\_2, clicks\_category\_3 and clicks\_category\_4 to represent clicks performed by one participant in each task category respectively.

## 5.5 One-way Analysis of Variance (ANOVA)

The aim for ANOVA is to measure the difference in efficiency (clicks, time) and effectiveness (hits, precision, recall) between versions or different categories of tasks. We conducted different ANOVAs with dependent variables and between-subjects variables as described above.

The results show that the effectiveness variables had similar performances among versions and tasks, thus they are not much valuable for further analyses. There are two possible reasons to explain this issue.

One reason is that we gave each participant sufficient amount of time so they tried their best to find the best answers. As one participant mentioned in the post-study questionnaire:

*“You might consider reducing time (from 15min to 10min) and increase the number of tasks”.*

This suggests that participants have extra time to refine their answers to be close to the gold standard. The consideration for an adequate time span was to enable participants to forward to next question so we can measure the time differences. We may need a more intuitive design in interface and tasks in the future to reach this goal while be able to measure effectiveness.

Another reason is that since we have a relatively small corpus, participants could narrow down to a small set merely by using search interfaces to help filtering results. This may also lead to statistically similar performance in effectiveness.

As for efficiency, all the three between-subjects variables impact significantly on total clicks. There’s no significantly difference in time between versions since time could be affected by many other factors such as network speed, and although we have control on the background of participants on the same group, differences in computer skills could still affects how long it takes to complete a task.

Though we need to work on fixing these issues to be able to evaluate and observe possible differences in effectiveness variables and time in the future, in this thesis we narrow down the analyses to efficiency variable, specifically, total clicks.

Table 5.2: Tests of Between-Subjects Effects (Dependent Variable: TotalClicks)

Source	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3	4357.829	4.277	.005	.033
Intercept	1	464051.565	455.413	.000	.545
$H_{score}$	1	3907.878	3.835	<b>.051</b>	.010
$H_{feature}$	1	8020.898	7.872	<b>.005</b>	.020
$H_{feature} * H_{score}$	1	1144.711	1.123	.290	.003
Error	380	1018.968			
Total	384				
Corrected Total	383				

### 5.5.1 ANOVA of Dependent Variables with $H_{feature} * H_{score}$ as Between-subjects Variable

To explore the differences among the  $H_{feature} * H_{score}$ , we conducted a one-way analysis of variance (ANOVA) with total clicks as the dependent variables and ( $H_{feature} * H_{score}$ ) as the between-subjects variable. The between-subjects effect is shown in Table 5.2.

ANOVA of TotalClicks in different  $H_{score}$  or  $H_{feature}$  situations indicates that there is a significant differences in TotalClicks between groups that  $H_{feature} = 1$  and groups that  $H_{feature} = 0$ ,  $p > 0.05$ . There is a marginally significance between groups that  $H_{score} = 1$  and  $H_{score} = 0$ ,  $p = 0.051 < 0.1$ . Both  $H_{score}$  ( $PartialEtaSquared = 0.01$ ) and  $H_{feature}$  ( $PartialEtaSquared = 0.02$ ) are considered having a small effect size according to rules of thumb [45]. However,  $H_{feature}$  impact the results slightly more than  $H_{score}$ .

Therefore, whether the visualization has visualized features or not impact the number of total clicks performed. Version 1 and Version 3 have the same  $H_{score} = 0$  value but different  $H_{feature}$ . Same difference can be found between Version 2 and Version 4. These two sets of versions have different results in the number of total clicks (see Table 5.3). Meanwhile, whether the visualization has informativeness and relatedness scores or not have only a marginally impact on the number of total clicks. The differences show in between of Version 1 and Version 2, or Version 3 and Version 4 (Table 5.3). To have a better understanding for the difference in versions (e.g. Version 2 vs. Version 3, Version 1 vs. Version 4), we started ANOVA and post hoc analysis in versions.



Table 5.3: Total Clicks in Different Versions

Version	Total Clicks
Version 1 ( $H_{feature} = 0, H_{score} = 0$ )	4248
Version 2 ( $H_{feature} = 0, H_{score} = 1$ )	3304
Version 3 ( $H_{feature} = 1, H_{score} = 0$ )	3039
Version 4 ( $H_{feature} = 0, H_{score} = 1$ )	2758

### 5.5.2 ANOVA and Post Hoc Analysis of Dependent Variables with Version as Between-subjects Variables

ANOVA results shown that version has a significant impact on TotalClicks ( $p = 0.005 < 0.05$ ) as well. To analyze which version performs better, we conducted a post hoc study on the impact of version on TotalClicks using Turkey and LSD evaluation. Results are shown in Table 5.4.

Version 4 has the most impact on TotalClicks and Version 3 also has a great impact on TotalClicks. Version 2 still perform significantly better than Version 1. There are no significantly differences among Version 2, Version 3 and Version 4.

### 5.5.3 ANOVA and Post Hoc Analysis of Dependent Variables with Task Category as Between-subjects Variables

We conducted ANOVAs on the difference in each dependent variable with different TaskCategory. The results are plotted in Figure 4.2.

Table 5.4: Multiple Comparisons (Dependent Variable: TotalClicks)

	I Version	J Version	Mean Diff I-J	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Turkey HSD	1	2	9.83	4.607	.144	-2.06	21.72
		3	12.59	4.607	<b>.033</b>	.70	24.48
		4	15.52	4.607	<b>.005</b>	3.63	27.41
	2	1	-9.83	4.607	.144	-21.72	2.06
		3	2.76	4.607	.932	-9.13	14.65
		4	5.69	4.607	.605	-6.20	17.58
	3	1	-12.59	4.607	<b>.033</b>	-24.48	-.70
		2	-2.76	4.607	.932	-14.65	9.13
		4	2.93	4.607	.921	-8.96	14.82
	4	1	-15.52	4.607	<b>.005</b>	-27.41	-3.63
		2	-5.69	4.607	.605	-17.58	6.20
		3	-2.93	4.607	.921	-14.28	8.96
LSD	1	2	9.83	4.607	<b>.033</b>	.77	18.89
		3	12.59	4.607	<b>.007</b>	3.53	21.65
		4	15.52	4.607	<b>.001</b>	6.46	24.58
	2	1	-9.83	4.607	<b>.033</b>	-18.89	-.77
		3	2.76	4.607	.549	-6.30	11.82
		4	5.69	4.607	.218	-3.37	14.75
	3	1	-12.59	4.607	.007	-21.65	-3.53
		2	-2.76	4.607	.549	-11.82	6.30
		4	2.93	4.607	.526	-6.13	11.99
	4	1	-15.52	4.607	<b>.001</b>	-24.58	-6.46
		2	-5.69	4.607	.218	-14.75	3.37
		3	-2.93	4.607	.526	-11.99	6.13

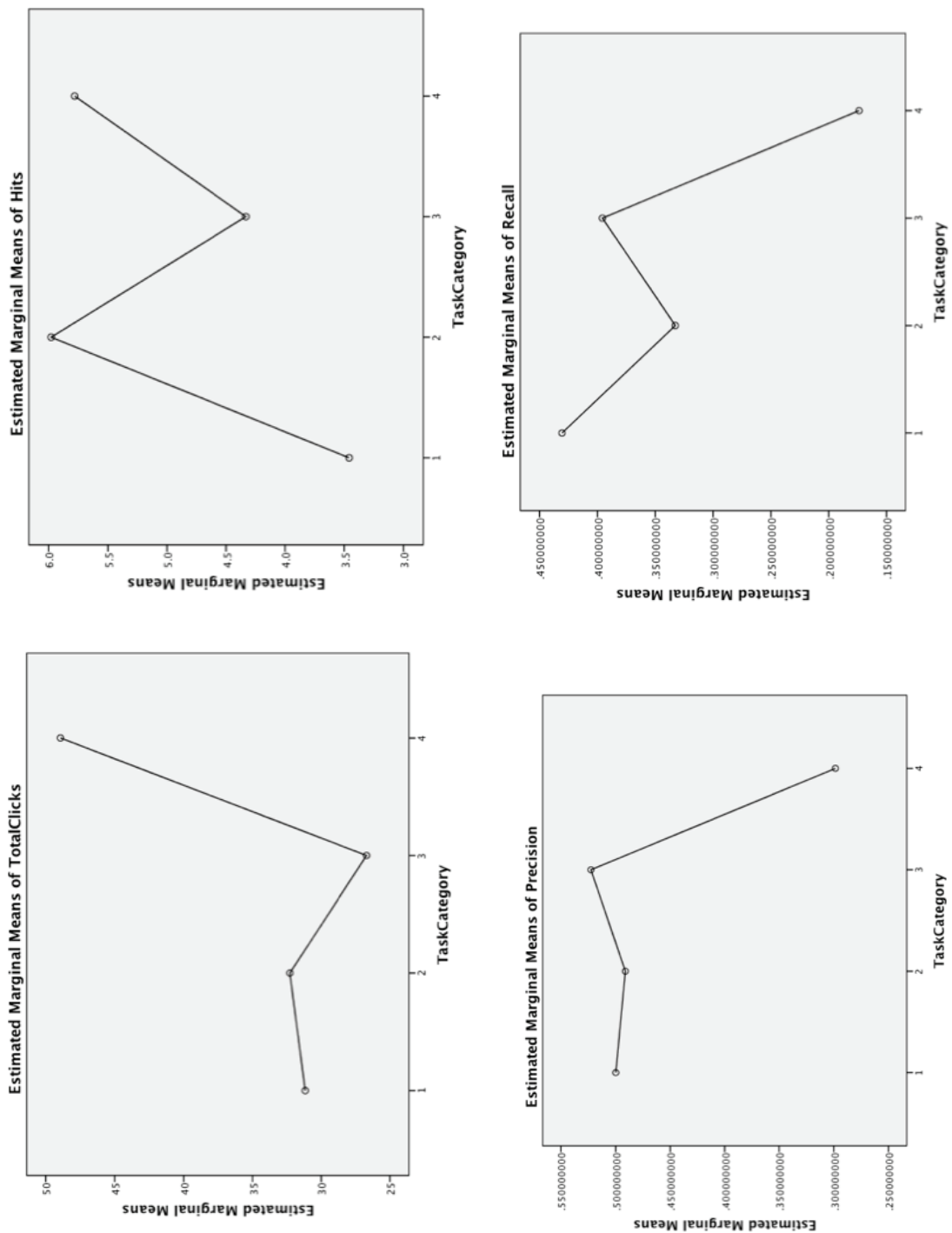


Figure 5.1: Significant Interactions Between Dependent Variables and TaskCategory.

Table 5.5: Total Clicks in Different Task Categories

Task Category	Total Clicks
Task Category 1 (Information Gathering)	2991
Task Category 2 (Opinion)	3098
Task Category 3 (Fact Searching)	2562
Task Category 4 (User Profile Related)	4698

Both precision and recall drops dramatically in TaskCategory 4 (Figure 5.1). Further post hoc analysis indicates that TaskCategory 4 is significantly different from the other three categories while among other categories the difference is not very significant. TaskCategory 4 also has significant more TotalClicks (Table 5.5). The hits are more in category two and four, and less in task category one and three. As discussed before, the differences in precision, recall, and hits are not significant, though they still suggests a slightly different in TaskCategory 4.

## 5.6 One-way Repeated Measures ANOVAs

### 5.6.1 Repeated Measures of $H_{feature} * H_{score}$

For repeated measures ANOVAs, the ANOVA analyses had been done on four task categories repeatedly to get average effects.

Two 2 ( $H_{feature}$ ,  $H_{score}$ ) X 4 (TotalClicks) repeated measures ANOVA design was used to evaluate the relationship between participants using different versions of system that how many clicks they need to complete different categories of questions. Main effects of clicks for different type of questions are shown in Table 5.6, where  $p = 0.001$ , and  $H_{score}$  ( $p = 0.225$ ),  $H_{feature}$  ( $p = 0.085$ ),  $H_{feature} * H_{score}$  ( $p = 0.508$ ). The interaction between clicks and  $H_{feature}$  is statistically significant ( $0.085 < 0.05$ ). Therefore, versions which  $H_{feature} = 1$  performs differently than versions that  $H_{feature} = 0$ . To reveal more details, we conducted repeated measures and post hoc on versions.

### 5.6.2 Repeated Measures and Post Hoc of Version

The graph of significant interactions between four versions on four task categories is shown in Figure 5.2. We can see from the graph that for TaskCategory 1, 2

Table 5.6: Tests of Between-Subjects Effects for Repeated Measures (Dependent Variable: TotalClicks)

Source	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	1	232025.783	180.154	.000	.804
$H_{score}$	1	1953.939	1.517	.225	.033
$H_{feature}$	1	4010.449	3.114	<b>.085</b>	.066
$H_{feature} * H_{score}$	1	572.355	.444	.508	.010
Error	44	1018.968			

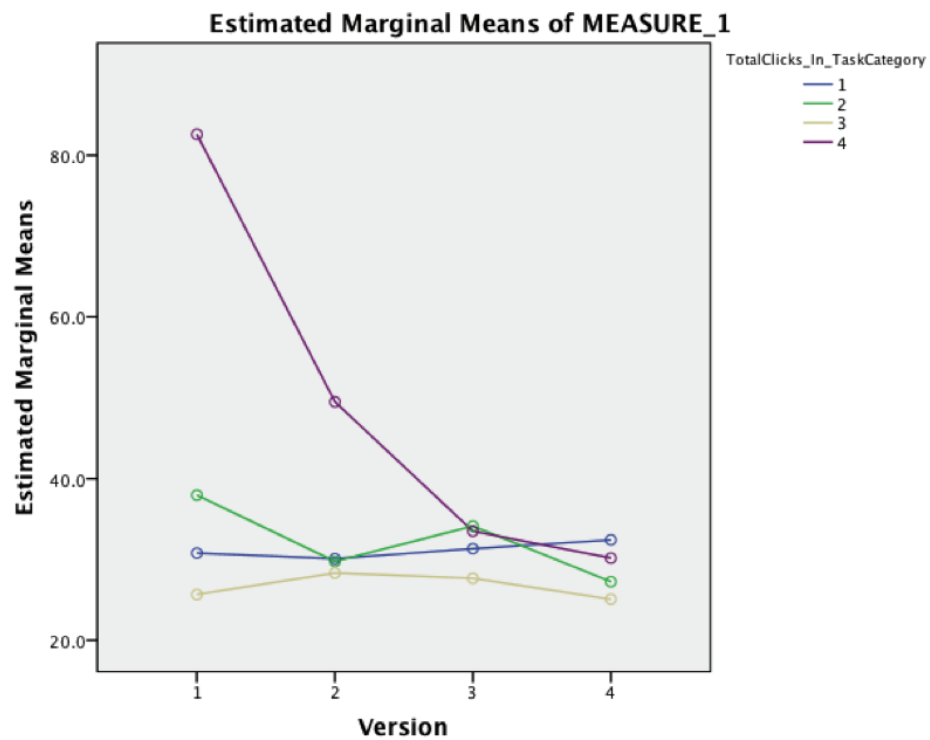


Figure 5.2: The Graph of Significant Interactions

Table 5.7: Function Usages in Different Versions

<b>Version</b>	deselect _checkbox	include _node	select _checkbox	click _block	click _node	exclude _node	search
<b>1</b>	407	525	115	29	4248	792	356
<b>2</b>	455	505	212	38	3304	733	380
<b>3</b>	543	510	265	64	3039	640	407
<b>4</b>	406	494	178	23	2758	529	331

and 3, there's not very much difference in TotalClicks between the four versions. For TaskCategory = 4, there's a significant difference between version 1 and other versions which the post hoc study shows a significance value  $p < 0.05$  (version 2  $< 0.05$  when using LSD but  $> 0.05$  using Turkey evaluation). No big difference found among version 2, 3 and 4.

From repeated measure analysis results, we conclude that: 1) category 4 were completed with lower precision and recall score, which suggests that this category of task is more difficult to complete the four versions of visualizations; 2) version 4 that combines visual features and computed scores ( $H_{feature} = 1, H_{score} = 1$ ) performs the best in TaskCategory 4; 3) in general,  $H_{feature} = 1$  performs slightly better than  $H_{score} = 1$ . We will discuss later in Section 5.8 for the possible explanations to the patterns we observed.

## 5.7 Other Data Analysis

We collected other forms of data to help improve our results and the future design in user studies and visualizations. We collected the data for usage of different functions, keywords that users input for narrowing down results and comments from post-study questionnaires.

### 5.7.1 Function Usage

The statistics of function usage in different versions and different task categories are shown in Table 5.7 and Table 5.8 (See Appendix T for the raw data we collected from each participant).

Participants who used version four relied on search functions less and clicked more nodes before they got answers.

Table 5.8: Function Usages in Different Tasks

<b>Task</b>	deselect _checkbox	include _node	select _checkbox	click _block	click _node	exclude _node	search
<b>1</b>	401	376	189	54	2991	603	655
<b>2</b>	400	620	159	37	3098	826	396
<b>3</b>	449	508	254	34	2562	589	487
<b>4</b>	561	530	168	29	4698	726	236

### 5.7.2 Post Study Questionnaires and Questions

After the user study, each participant had the option to write about their experiences on the scores, features, design of user study and system. 19 participants thought that informativeness and relatedness scores are helpful, yet 3 of them said that they don't trust some part of the results. 20 participants felt that features were helpful for finding the answers. We observed the network latency of the system and informed the participants to expect that during the study, and two of the participants still talked about the speed in the post study questions.

*“The speed can be improved.”*

*“Maybe the system could be faster and more robust.”*

Many participants suggested that they would like a better filtering and searching function. Although search function is just an assistant function in our system and was beyond the scope that we wanted to evaluate, we should consider the impacts of search functions to both the study results and system design in the future.

## 5.8 Search Task vs. Browsing Task

As we mentioned in the chapter Designing an Interactive Visualization of Twitter Conversations, we categorized information seeking tasks into four categories: 1) Information gathering, 2) opinion, 3) fact searching and 4) user profile related tasks.

We observed that in task 1, 2 and 3, participants used more search functions than categories four. Participants tended to click to read text in task 4 more frequently than the rest. These observations show that there are two patterns in these tasks: searching and browsing. Task 1, 2 and 3 shown more like a searching pattern, while users tended to search through keywords. Task 4 took more clicks and browsing text to complete, thus it suggested more as a browsing pattern. Our questionnaire and

interview results (Appendix P) show that browsing tasks are considered more difficult to complete than searching tasks.

## 5.9 Conclusion and Design Guidelines

In this chapter, we described the variables that have been used in our study analysis and presented analysis results of ANOVAs and Repeated Measures. We summarize the answers to some research questions as follows:

1. *Which combinations of features and scores are more efficient and perform the best overall in information seeking tasks?*

Our analyses show that combining both features and scores in the visualization performs the best in terms of total clicks. Version 4 has the most impact on total clicks and Version 3 also has an impact on total clicks. Version 2 still performs significantly better than Version 1.

2. *What are performances of these combinations on individual category of tasks?*

For information browsing tasks (task category 4), combining both features and scores performs the best.  $H_{feature}$  performs slightly better than  $H_{score}$ .

3. *Which tasks have better performance and why?*

Searching tasks (task category 1, 2 and 3) have better performance than browsing task (task category 4). Participants relied heavily on searching functions besides the visualized results in searching tasks.

From the analyses, we distill some process-oriented recommendations and future improvements for both how to conduct user study and implementing better visualizations.

### 5.9.1 Design Guidelines and Procedures

After conducting analysis on various measurements for our visualization study, we propose the following guidelines and procedures for designing interactive visualization system for information seeking tasks.



First, determine whether the tasks are browsing tasks or searching tasks. Adapt visualizations to focus on browsing or searching tasks.

Second, if they are searching tasks, more focus should be put into how to improve searching algorithms and searching experience.

Third, if they are browsing task, experiment with appropriate visual features that may potentially help lead to the answers and also assist the user by suggesting a computed score of possibility to answer. Choosing an appropriate set of visual features potentially improves performance more than suggesting imperfect computing algorithms.

### **5.9.2 Future Improvements for the User Study and the Visualization**

For both the user study and visualization, our first future consideration will be having a bigger dataset. Through this we may be able to study more possible difference among participants behaviors.

It is also a challenge to combine the interface for both searching and browsing tasks. Some tasks may need both searching and browsing skills. One approach could be progressive disclosure. For instance, implementing an interface that helps users with searching procedures first. Once searching part is complete, disclose detailed interface to enable users to browse interesting visual patterns.

For implementing the visualizations, improvements can be done on both the data side and visual side. Support interface with better algorithms will always help, while how to display dominant visual features will be an interesting research topic. As we came across difficulties trying to measure the performance of time, how to design a better user study that can adopt time as a reliable measurement remains a challenge.

## Chapter 6

### Conclusion

In this thesis, I presented an evaluation process for visualizing solutions of information seeking tasks on Twitter data. This process consists of four major procedures: 1) define information seeking tasks and features for categorizing tasks; 2) study the impact features for information seeking tasks; 3) implement visualization systems; 4) study and evaluate the effectiveness and performance of visualization systems.

First, four categories were defined for common information seeking tasks. The four categories are information gathering, opinion, fact searching, and user profile related tasks. Our research uses the Twitter data as an experiment domain. The goal of completing information seeking tasks on Twitter conversation data, is to find the most important pieces of data that relates to certain topic. In our case, we focus on academic related topics.

Second, we designed a study to find the prominent features of important Twitter data. We defined two measurements for evaluating how important one Twitter conversation is. The two measurements are informativeness and relatedness. Meanwhile, we listed several features that may possibly affect the informativeness and relatedness of Twitter data, as the variables of our study. 20 participants were recruited to label the score of Twitter conversation data.

Third, we built four versions of visualization systems on the basis of the ranking of influential features to compare and evaluate the performance of those features. Finally, we evaluated four versions of visualizations systems under different tasks.

In this chapter, I will review the list of contributions in my thesis and later discuss about the limitations of my work and possible future works.

#### 6.1 List of Contributions

Based on the studies and analysis described before, my thesis work can be summarized in a few aspects: dataset, study for relatedness and informativeness of Twitter

conversation, visualizations system and the design of user studies for evaluating visualization systems.

### **6.1.1 Datasets**

We collected 13460 Twitter conversation data from March 2013 to April 2013. The Twitter dataset includes 36939 users and 153185 tweets. That Twitter data comes from five Canadian academic institutions: University of British Columbia, Simon Fraser University, McGill University, McMaster University and Dalhousie University. The domain of this data was restricted in academic topics.

### **6.1.2 Relatedness and Informativeness**

Considering the features of text document and social media, we defined two measurements for the importance of each Twitter conversation: relatedness and informativeness. A user study was designed to study the influence of selected features that may affect the relatedness and informativeness of Twitter conversation.

### **6.1.3 Visual Analysis Tool**

We built four visual systems based on the selected features from previous study. These four different versions of system focus on two variables: score and quantified visual features. Performance and efficiency of those visual systems were evaluated by a following user study.

### **6.1.4 User Study Design for Evaluating Visualization System**

We designed and conducted a human-centered user study to evaluate four different versions of system under different information seeking tasks.

### **6.1.5 Design Guidelines and Procedures**

We propose the following procedures and guidelines for designing interactive visualization system for information seeking tasks.

First, determine whether the tasks are browsing tasks or searching tasks. Adapt visualizations to focus on browsing or searching tasks by showing more data visualization patterns or more refined search functions.

Second, if they are searching tasks, more focus should be put into how to improve searching algorithms and searching experience.

Third, if they are browsing task, experiment with appropriate visual features that may potentially help lead to the answers and also assist the user by suggesting a computed score of possibility to answer. Choosing an appropriate set of visual features potentially improves performance more than suggesting imperfect computing algorithms.

## 6.2 Limitations and Future Work

There are limitations around the algorithms we use for the classification of our data set, and score predictions. In the future, there are possibilities to explore more details on these questions:

1. What learning algorithms should we apply?
2. How does it adoptable for other domains other than academic corpus?
3. How can the algorithms scale to a bigger corpus?

Other than computing algorithms, there are also questions about how to explore a better task definition, how we evaluate and select appropriate sets of measurements for a given analytic task and whether those measurements can be adapted to other domains.

The evaluation of visual system design remains challenging and needs a lot of efforts in the future studies. It is a challenge to combine the interface for both searching and browsing tasks. Some tasks may need both searching and browsing skills. One possible approach to address this issue could be progressive disclosure. Searching function could be used to narrow down the results and browsing design offers the insights to explore interesting patterns in the data. For implementing the visualizations, improvements can be done on both the data side and visual side. For designing the data support of text visualization system, improving the performance

of text mining algorithms will direct users to be closer to the desired results, while how to display dominant visual features to enable exploring potential data patterns would be an interesting research topic. The last but not the least, as we came across difficulties trying to measure the performance of time, how to design a better user study that can adopt time as a reliable measurement remains a challenge for the future as well.

## Bibliography

- [1] Wikipedia, “Twitter — wikipedia, the free encyclopedia,” 2015. [Online; accessed 22-November-2015].
- [2] C. Honey and S. Herring, “Beyond microblogging: Conversation and collaboration via twitter,” in *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, pp. 1–10, 2009.
- [3] R. Poynter, *The Handbook of Online and Social Media Research: Tools and Techniques for Market Researchers*. Chichester: Wiley, 2010.
- [4] Wikipedia, “User-generated content — wikipedia, the free encyclopedia,” 2015. [Online; accessed 22-November-2015].
- [5] E. Qualman, *Socialnomics: How Social Media Transforms the Way We Live and Do Business*. Business professional collection, Hoboken, N.J.: Wiley, 2010.
- [6] S. Choi, E. Grewal, and K. Wen, “Social data and college statistics,” in *Proceedings of #Influence12: Symposium & Workshop on Measuring Influence on Social Media*, (Dalhousie University, Halifax, Nova Scotia, Canada), September 2012.
- [7] L. Rossi and M. Magnani, “Conversation practices and network structure in twitter,” in *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*, 2012.
- [8] D. Boyd, S. Golder, and G. Lotan, “Tweet, tweet, retweet: Conversational aspects of retweeting on twitter,” in *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pp. 1–10, Jan 2010.
- [9] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, “Measuring user influence in twitter: The million follower fallacy,” in *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.
- [10] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [11] Y. Yang and J. O. Pedersen, “A comparative study on feature selection in text categorization,” in *ICML*, vol. 97, pp. 412–420, Morgan Kaufmann Publishers, 1997.
- [12] M. J. Paul and R. Girju, “A two-dimensional topic-aspect model for discovering multi-faceted topics,” in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010* (M. Fox and D. Poole, eds.), AAAI Press, 2010.

- [13] Y. Wu and D. W. Oard, “Bilingual topic aspect classification with a few training examples,” in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, (New York, NY, USA), pp. 203–210, ACM, 2008.
- [14] A. Zubiaga, D. Spina, V. Fresno, and R. Martínez, “Classifying trending topics: A typology of conversation triggers on twitter,” in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, (New York, NY, USA), pp. 2461–2464, ACM, 2011.
- [15] T. Finin, W. Murnane, A. Karandikar, N. Keller, J. Martineau, and M. Dredze, “Annotating named entities in twitter data with crowdsourcing,” in *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, CSLDAMT '10, (Stroudsburg, PA, USA), pp. 80–88, Association for Computational Linguistics, 2010.
- [16] D. Sontag and D. Roy, “Complexity of inference in latent dirichlet allocation,” in *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain.*, pp. 1008–1016, 2011.
- [17] E. Bingham and H. Mannila, “Random projection in dimensionality reduction: Applications to image and text data,” in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, (New York, NY, USA), pp. 245–250, ACM, 2001.
- [18] S. Huang, W. Peng, J. Li, and D. Lee, “Sentiment and topic analysis on social media: A multi-task multi-label classification approach,” in *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, (New York, NY, USA), pp. 172–181, ACM, 2013.
- [19] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [20] A. Montejo-Ráez, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña-López, “Ranked wordnet graph for sentiment polarity classification in twitter,” *Computer Speech & Language*, vol. 28, no. 1, pp. 93–107, 2014.
- [21] J. Lin and A. Kolcz, “Large-scale machine learning at twitter,” in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, (New York, NY, USA), pp. 793–804, ACM, 2012.
- [22] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” *CS224N Project Report, Stanford*, vol. 1, p. 12, 2009.
- [23] M. Tory and T. Moller, “Human factors in visualization research,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 10, pp. 72–84, Jan 2004.

- [24] J. Chuang, D. Ramage, C. Manning, and J. Heer, “Interpretation and trust: Designing model-driven visualizations for text analysis,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 443–452, ACM, 2012.
- [25] M. Sedlmair, M. D. Meyer, and T. Munzner, “Design study methodology: Reflections from the trenches and the stacks,” *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2431–2440, 2012.
- [26] S. Silva, J. Madeira, and B. Santos, “There is more to color scales than meets the eye: A review on the use of color in visualization,” in *Information Visualization, 2007. IV '07. 11th International Conference*, pp. 943–950, July 2007.
- [27] C. Wittenbrink, A. Pang, and S. Lodha, “Glyphs for visualizing uncertainty in vector fields,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 2, pp. 266–279, Sep 1996.
- [28] M. Nacenta, U. Hinrichs, and S. Carpendale, “Fatfonts: Combining the symbolic and visual aspects of numbers,” in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, (New York, NY, USA), pp. 407–414, ACM, 2012.
- [29] C. Ware, *Information Visualization: Perception for Design*. Interactive Technologies, Waltham, MA: Elsevier Science, 2004.
- [30] U. Fayyad, A. Wierse, and G. Grinstein, *Information Visualization in Data Mining and Knowledge Discovery*. The Morgan Kaufmann series in data management systems, San Francisco, CA, USA: Morgan Kaufmann, 2002.
- [31] Z. Liu, B. Jiang, and J. Heer, “immens: Real-time visual querying of big data,” *Computer Graphics Forum*, vol. 32, no. 3pt4, pp. 421–430, 2013.
- [32] W. Huang, S. hee Hong, and P. Eades, “Effects of sociogram drawing conventions and edge crossings in social network visualization,” *Journal of Graph Algorithms and Applications*, pp. 397–429, 2007.
- [33] S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian, “Tiara: Interactive, topic-based visual text summarization and analysis,” *ACM Trans. Intell. Syst. Technol.*, vol. 3, pp. 25:1–25:28, Feb. 2012.
- [34] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu, “Storyflow: Tracking the evolution of stories,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, pp. 2436–2445, Dec 2013.
- [35] P. Xu, Y. Wu, E. Wei, T.-Q. Peng, S. Liu, J. Zhu, and H. Qu, “Visual analysis of topic competition on social media,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, pp. 2012–2021, Dec 2013.



- [36] E. Bertini, A. Tatu, and D. Keim, “Quality metrics in high-dimensional data visualization: An overview and systematization,” *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, pp. 2203–2212, Dec 2011.
- [37] M. Kellar, C. Watters, and M. Shepherd, “A field study characterizing web-based information-seeking tasks,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, pp. 999–1018, May 2007.
- [38] J. Hurlock and M. L. Wilson, “Searching twitter: Separating the tweet from the chaff,” in *ICWSM* (L. A. Adamic, R. A. Baeza-Yates, and S. Counts, eds.), The AAAI Press, 2011.
- [39] D. Elswiler and M. Harvey, “Engaging and maintaining a sense of being informed: Understanding the tasks motivating twitter search,” *JASIST*, vol. 66, no. 2, pp. 264–281, 2015.
- [40] G. C. Bowker and S. L. Star, *Sorting Things Out: Classification and Its Consequences*. Inside technology, Cambridge, Mass.: MIT Press, 1999.
- [41] B. Goncalves, N. Perra, and A. Vespignani, “Modeling users’ activity on twitter networks: validation of Dunbar’s number,” *PLoS ONE*, vol. 6, no. 8, p. e22656, 2011.
- [42] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web.,” Technical Report 1999-66, Stanford InfoLab, November 1999.
- [43] M. A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.
- [44] V. Vapnik, *Statistical learning theory*, vol. 1. New York: Wiley, 1998.
- [45] J. Miles and M. Shevlin, *Applying Regression and Correlation: A Guide for Students and Researchers*. London: SAGE Publications, 2001.

## Appendix A

### First Phase User Study - Study Protocol

This study protocol below excerpts from Research Ethics Application (REB # 2013-2954) submitted to Dalhousie University under this research. Some of the references have been modified according to the architecture of this thesis.

#### **2.3 STUDY DESIGN** – In this section

- 2.3.1 state the hypotheses or the research questions or research objectives
- 2.3.2 describe the general study design and how it will address the hypotheses / questions / objectives
- 2.3.3 describe how many participants are needed and how this was determined
- 2.3.4 describe the plan for data analysis in relation to the hypotheses/questions/objectives
- 2.3.5 if a phased review is being requested, describe why this is needed for this study and which phases are contained in this application

#### **2.3.1 Research Question**

Two primary research questions are addressed in Phase one of this study:

- 1) What features matter to users when determine the importance of twitter conversations?
- 2) Which is more valuable, tweets in conversations or standalone tweets?

#### **2.3.2 General Study Design**

##### **Process**

We will conduct a laboratory study using members of the Dalhousie University community as participants (primarily students, but recruitment is also open to faculty and staff). We will use a desktop computer with a screen recording system to perform tasks. The procedure of study is described in the following table.

<b>Steps</b>	<b>Description</b>	<b>Approximate time</b>
1. Consent From(Appendix A.3)		5 minutes
2. Background Questionnaire(Appendix A.4)	Participants' knowledge about U15 universities and campus life.	5 minutes
3. Task		30 minutes
1) Task description	Researcher introduces the task	5 minutes
2) On-screen Questionnaire (Appendix A.5)	Rate relevance and relatedness of 40 twitter conversations	20 minutes
3) Post-task Questionnaire (Appendix A.6)	After rating all the twitter conversations, participants answer questions about the importance of various features.	5 minutes

At the end of the study the researcher will conduct a brief unstructured interview, asking participants about improvements of user study design.

### **The Tasks**

The task the experimental session is rating twitter conversations on screen. Each participant will be shown approximately conversations from twitter talking about U15. The participant will rate the relevance and relatedness of the conversations to help us determine the impact features.

The layout page for the on-screen task is shown in Figure 1:

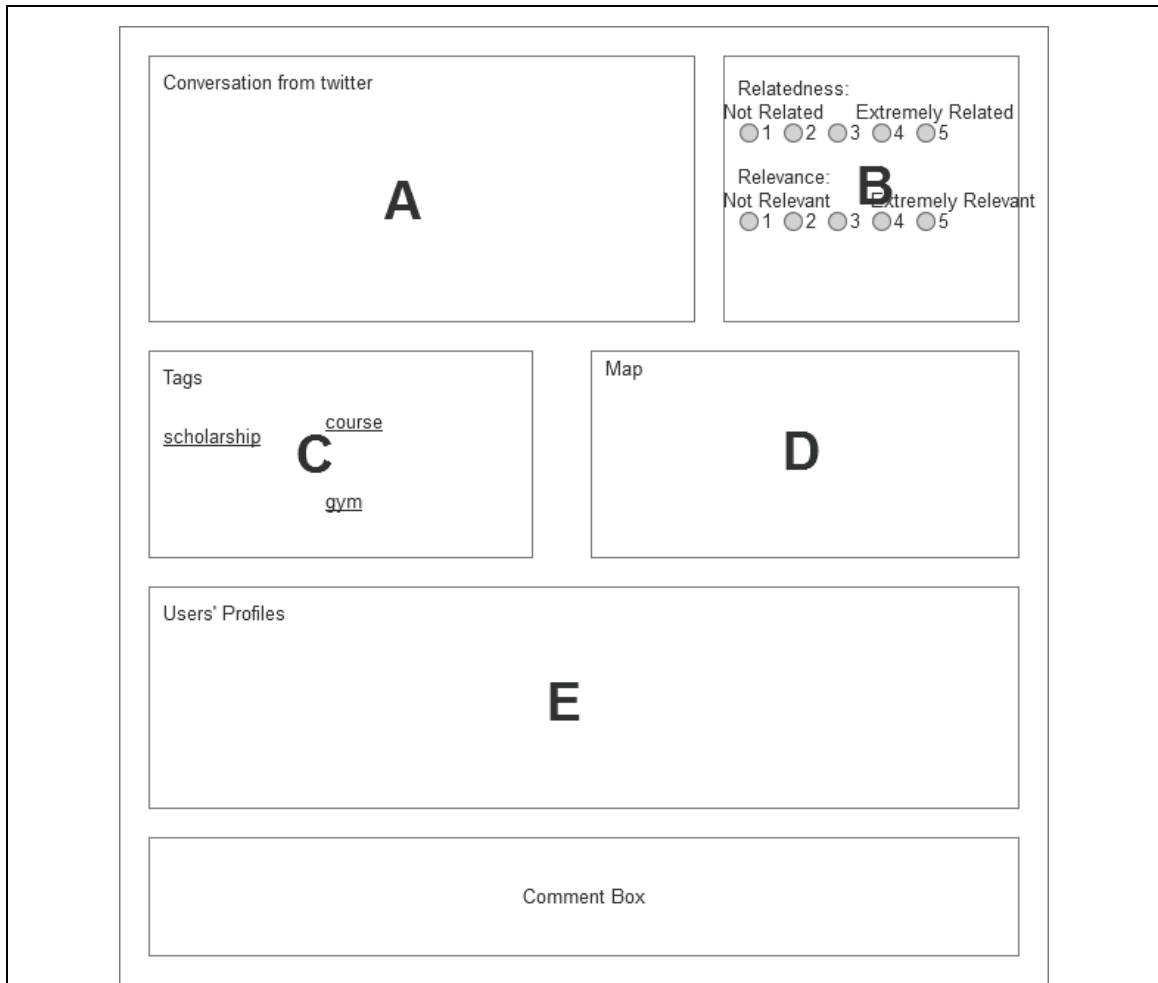


Figure 1: Twitter conversation rating page. A: conversation text area; B: rating board; C: topic tags; D: map for twitter location; E: users' profile.

The conversations text will be shown in the top left (Figure 1A). Conversations are selected from academic twitter data that we collected talking about four different topics (academic, finance, campus, admission). Participant will be asked to rate the relatedness and relevance of conversations to academic topics in the top right (Figure 1B). The relatedness defines as whether one conversation is related to topics about U15 universities. For example, “Dalhousie Street in Calgary” is not related to U15 although it has the same keyword as “Dalhousie University”. Relatedness has a scale from 1 to 5, in which 1 means “not related”, 5 for “extremely related”. The relevance defines as whether this conversation is helpful or informative to present some aspects of university such as research, course or scholarship. For example, “I am here in Dalhousie University” is less informative than “Introducing a new scholarship for graduate students at Dalhousie University”. Scale of relevance is from 1 to 5 (“not relevant” to “extremely

relevant”). A cloud of pre-defined tags will be displayed in Figure 1C, and participant can click to select all the tags that are related to the current conversation. Figure 1D shows the locations of tweets in conversation or users’ location from their profile. Information of tweet users profile will be displayed below the tags (Figure 1E). A text box for note is located on the bottom of page. If participants have any comments they can write down there. A post-task questionnaire (Appendix A.6) will be shown after participant finishes rating all the tweets.

### **2.3.3 Participants**

We will recruit 20 participants. Since we are going to select conversational features through a regression model, 800 (20 participants \* 40 twitter conversations) input is needed to make a credible statistical analysis. Participants should be Dalhousie University students, staff or faculty. We will require participants to have some knowledge about U15 universities and Twitter.

### **2.3.4 Data Analysis**

After the study, we will use the questionnaires to help training the model for features selection. Through screen records, we will keep track of any difficulties (response time, mouse movement) in performing any of the tasks to help determine the effectiveness of the options. Interview notes are taken by interviewing participants. These notes will be used as reference of feature selection and improvement for future study design.

This screen records and the source records will be encrypted and placed on a password-protected computer account accessible only to the investigators. The records will be cleared after study results come out.

The study results will be reported as research paper built from qualitative and quantitative analysis of interaction data combined with questionnaire data, aggregate information, charts and visualizations, and statistics. We will adopt a regression model to select significant conversational features from rated twitter conversations. Results from the study will be submitted to appropriate conferences or journals in the area of Information Visualization, Human Computer Interaction or Social Media.

### **2.3.5**

A phased review is required. We propose two phases for our research. In the first phase described in this application, participants will be asked to rate the relatedness and relevance of conversations as the first task. Then they will perform comparisons between conversations and standalone tweets. The visualizations in the second phase will be designed according to the study results from phase one. In the second phase, participants will be shown a visualization tool and asked to interact with it. We will evaluate the effectiveness of visualizations.

**2.4 RECRUITMENT** – In this section, for each type of participant to be recruited, describe

- 2.4.1 the study population
- 2.4.2 any social / cultural / safety considerations
- 2.4.3 and justify all specific inclusion / exclusion criteria of participants
- 2.4.4 any recruitment instruments (attach copies)
- 2.4.5 who will be doing the recruitment and what actions they will take
- 2.4.6 any screening measures, and how they will be used (attach copies)
- 2.4.7 any permissions that are needed and attach letters

#### **2.4.1 Study Population**

The study population will be Dalhousie University students, staff, and faculty.

#### **2.4.2 Considerations**

N/A

#### **2.4.3**

Participants should be familiar with names of top universities in Canada (U15), campus life and Twitter. Since we are going to hire participants within school communities, most candidates are supposed to have some knowledge about academic institutions and campus life.

#### **2.4.4-2.4.5 Recruitment Procedures and Instruments**

Shali Liu will recruit Dalhousie University students, staff and faculty to take part in

this study. All participants will be recruited by email recruitment notice through events calendars (i.e. Notice Digest, notice.digest@dal.ca) and through mailing lists (i.e. cs.all@dal.ca), which is also a monitored mailing list. We may also post recruitment notice through public websites such as Facebook, twitter. In the recruitment notice, participants will be asked to email their interest to participate to the listed researcher. The participant and researcher will then communicate to find an appropriate time for the participant to do the study. The recruitment notice script is shown in Appendix A.2.

#### **2.4.6 Screening Measures**

Screening will occur through self-selection by participants. Participants will be asked to complete a background questionnaire before study which asks questions about U15 universities and Twitter usage.

#### **2.4.7 Permissions**

N/A

### **2.5 INFORMED CONSENT PROCESS – In this section**

- 2.5.1 describe the informed consent process (attach a copy of all consent forms)
- 2.5.2 if oral consent is desired, describe why it is necessary and how it will be done (attach a  
copy of the script)
- 2.5.3 if a waiver of informed consent is sought, explain why and describe how the four criteria  
needed for this are met
- 2.5.4 for third party consent (with or without assent), describe how this will be done
- 2.5.5 describe plans (if any) for on-going consent
- 2.5.6 if community consent is needed, describe how it will be obtained

#### **2.5.1 Informed Consent Process**

All participants involved in the study will sign an informed consent form (see Appendix A.3). This will be administered by a researcher at the initial meeting of the study. The informed consent outlines the risks and benefits associated with

the study, a description of the study, the participant's right to withdraw without consequence, and assurances of confidentiality and anonymity of personal data. As well, the informed consent makes it clear to participants that they can withdraw from the study at any time without loss of compensation.

**2.5.2-2.5.6**

N/A.



## **2.6 DETAILED METHODOLOGY** - In this section describe

2.6.1 where the research will be conducted

2.6.2 what participants will be asked to do and the time each task will take (plus total time)

2.6.3 what data will be recorded and what research instruments will be used (attach copies)

2.6.4 the roles and qualifications of the study investigators / research staff

2.6.5 how long the participants will be involved in each part of the study

### **2.6.1 Location of Research**

The study will be conducted in a quiet private meeting room in Dalhousie University.

### **2.6.2 Study Details**

We will meet with participants at selected location. This study will take about 60 minutes. Before study, we will get informed consent (Appendix A.3) from participants. We will then discuss the study and answer any questions prior to starting the study. After that, we will have the participants perform the study, and fill in questionnaires.

During the study, participants will perform two different sections of tasks, using a dedicated desktop, which participants will fill in an on-screen questionnaire (Appendix A.5). After the participants finish a task, they will fill in a post-task questionnaire (see Appendix A.6). In the questionnaire, we will ask the participants to compare the different ways to perform the tasks and for any feedback that they have that may be helpful for the design of the application.

### **2.6.3 Data Recorded and Instruments Used**

We will use five main data collection methods during this study: Screen recording, interview notes, and background questionnaires, on-screen questionnaires, and post-task questionnaires.

#### **Screen recording**

We will record screen during on-screen questionnaire. Through screen records,

we will keep track of any difficulties in performing any of the tasks to help determine the effectiveness of the options.

### **Interview notes**

Interview notes are taken by interviewing participants. These notes will be used as reference of feature selection and improvement for future study design.

### **Background Questionnaire**

The questions (Appendix A.4) will ask the participants about their knowledge of top universities in Canada. These questions help us to know more about participants' experiences and preferences.

### **On-screen Questionnaire**

The on-screen questions (Appendix A.5) will ask user to rate the relevance and relatedness of text shown on screen. This is one of the major parts for collecting data to help analysing results.

### **Post-task Questionnaire**

The post-condition questions (Appendix A.6) will ask the participants about how they feel about different features of twitter conversations and their preference of choosing different type of tweets. This will help analyse research results.

## **2.6.4 Roles/qualifications of Study Investigators**

Primary investigator Shali Liu will be responsible for recruitment, for administering the study and analysis of the results. The investigators have designed the study, including the questionnaires, and prototype tools used during the study. Shali will be the primary researcher responsible for reviewing and accessing results. Dr. Evangelos Milios and Kirstie Hawkey will provide guidance and assistance during analysis.

## **2.6.5 Length of Participation**

This user study should take each participant about an hour to complete.

## Appendix B

### First Phase User Study - Recruitment Notice

We are recruiting participants to take part in a research study examining which features determine the importance of twitter conversations. We are looking for members of the Dalhousie University community, who have knowledge about Canadian universities, campus life and twitter.

The study will be conducted in a quiet private meeting room in Dalhousie University. This study will take about one hour to complete. You will first meet with a researcher to go over the study details, give consent to do the study and fill in a background questionnaire. You will then complete a set of tasks with the application that will be provided, and you will fill in questionnaires after doing each task set. Compensation is 15\$ for participation in the study.

If you are interested in participating, please contact Shali Liu ([shali@dal.ca](mailto:shali@dal.ca)).

## Appendix C

### First Phase User Study - Informed Consent



#### **Measuring relatedness and relevance of topic oriented twitter conversation**

Principal Investigators: Shali Liu, Faculty of Computer Science, [shali@dal.ca](mailto:shali@dal.ca)

Dr. Evangelos Milios, Faculty of Computer Science,  
[eem@cs.dal.ca](mailto:eem@cs.dal.ca)

Dr. Kirstie Hawkey, Faculty of Computer Science,  
[hawkey@cs.dal.ca](mailto:hawkey@cs.dal.ca)

Contact Person: Shali Liu, Faculty of Computer Science, [shali@dal.ca](mailto:shali@dal.ca)  
(902-999-2151)

We invite you to take part in a research study being conducted by Shali Liu at Dalhousie University. This description tells you about the risks, inconvenience, or discomfort which you might experience. Participating in the study might not benefit you, but we might learn things that will benefit others. You should discuss any questions you have about this study with Shali Liu.

The purpose of the study is to help us learn which features determine the importance of tweets. To be eligible to participate in the study, you must be a Dalhousie University student, staff or faculty, have knowledge about Canadian universities, campus life and twitter. This study takes about one hour to complete. You will be compensated \$15 for participating in this study; you can withdraw from the study at any time without consequence. A researcher is always available over the study period by email or to meet in person to answer any questions you may have or address any problems that you may experience with the tasks.

For each user study, at the beginning of the study, you will meet with an investigator (in a quiet, private meeting room in Dalhousie University). At this initial meeting you will be asked to give consent to do the study and to fill in a background questionnaire detailing knowledge about Canadian universities, campus life and twitter usage. You will perform a set of tasks with the application using a desktop. You will be asked to participate in a set of user study. You will be firstly given a set of twitter conversations to rate their relevance and relatedness to query topics. Then, you will fill in a questionnaire asking you about your preference of which information is more useful while making decisions. During the user study, the time

and mouse interactions will be recorded with screen recording software. Note that this record data will not identify you. You will fill in a few questionnaires before and after each phase of user study. We will also have a short interview with you after study. We may take notes during the interview. At the end of the study, you will be asked to briefly describe your study experience.

There are very low risks associated with this study. There is a low risk that some participants may become frustrated or embarrassed if they experience some difficulties performing the tasks during the study but the researcher will always be available during the study to answer any questions.

All personal and identifying data will be kept anonymous and confidential. The informed consent form and all research data will be kept in a secure location under confidentiality in accordance to University policy for 5 years post publication. You have the option of allowing direct quotes included in any final reports.

In the event that you have any difficulties with, or wish to voice concern about, any aspect of your participation in this study, you may contact Catherine Connors, Director, Office of Research Ethics Administration at Dalhousie University's Office of Human Research Ethics for assistance: phone: (902) 494-1462, email: Catherine.connors@dal.ca.

*"I have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I hereby consent to take part in the study. However, I understand that my participation is voluntary and that I am free to withdraw from the study at any time."*

**Participant**

**Researcher**

Name: \_\_\_\_\_ Name: \_\_\_\_\_

Signature: \_\_\_\_\_ Signature: \_\_\_\_\_

**Date:** \_\_\_\_\_ **Date:** \_\_\_\_\_

*"I agree to let you directly quote any comments or statements made in any written reports without viewing the quotes prior to their use and I understand that the anonymity of textual data will be preserved by using pseudonyms."*

**Participant**

**Researcher**

Name: \_\_\_\_\_ Name: \_\_\_\_\_

Signature: \_\_\_\_\_ Signature: \_\_\_\_\_

**Date:** \_\_\_\_\_ **Date:** \_\_\_\_\_

If you are interested in seeing the results of this study, please check below and provide your email address. We will contact you with publication details that describe the results.

*"I would like to be notified by email when results are available via a publication."*

[if this option is chosen, please include a contact email address:  
\_\_\_\_\_]

## Appendix D

### First Phase User Study - Background Questionnaire

#### PART I - PLEASE FILL IN THE FOLLOWING INFORMATION:

Identification number: \_\_\_\_\_

1. Age:

- 15-25                       25-35                       35-45                       over 45

2. Gender:

- Male                               Female

3. What is your role in school?

- Graduate                       Undergraduate                       Faculty and staff

4. Which of the Canadian top universities are you familiar with (familiar with names and some abbreviation such as UBC)?

Here is the list of universities that we focus on:

- Dalhousie University*  
 *McGill University*  
 *McMaster University*  
 *University of British Columbia*  
 *Simon Fraser University*

5. On the average, how much time do you spend per day on a computer?

- Less than 1 hour                       1 to less than 5 hours                       5 to less than 10 hours                       Over 10 hours

6. Do you have an account in Twitter?

Yes  No

7. How frequently do you use Twitter?

At Least once per day       A few times per week

A few times per month       Seldom use

Not at all



## Appendix E

### First Phase User Study - On-screen Questionnaires

**Please respond to the following statements using the given scale (circle response):**

The **relatedness** means whether one conversation is related to topics about U15 universities.

For example, “Dalhousie Street in Calgary” is not related to U15 although it has the same keyword as “Dalhousie University”.

The **relevance** is whether this conversation is helpful or informative to present some aspects of university such as research, course or scholarship.

For example, “I am here in Dalhousie University” is less informative than “Introducing a new scholarship for graduate students at Dalhousie University”.

	1	2	3	4	5		
	<i>Strongly not Related/Relevant</i>	<i>Somewhat not Related/Relevant</i>	<i>Neutral</i>	<i>Somewhat Related/Relevant</i>	<i>Strongly Related/Relevant</i>		
<b>1.</b> Relatedness			<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
			Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
<b>2.</b> Relevance			<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
			Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree

## Appendix F

### First Phase User Study - Post-task Questionnaires

*Please rate your general impression about the helpfulness of following features:*

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
	<i>Completely helpless</i>	<i>Somewhat helpless</i>	<i>Neutral</i>	<i>Somewhat helpful</i>	<i>Very helpful</i>
<b>1.</b> User profiles	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
	Completely helpless	Somewhat helpless	Neutral	Somewhat helpful	Very helpful
<b>2.</b> The number of user followers	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
	Completely helpless	Somewhat helpless	Neutral	Somewhat helpful	Very helpful
<b>3.</b> The number of user following	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
	Completely helpless	Somewhat helpless	Neutral	Somewhat helpful	Very helpful
<b>4.</b> Description of user	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
	Completely helpless	Somewhat helpless	Neutral	Somewhat helpful	Very helpful
<b>5.</b> Mentions(@) and hashtags(#) contributes to the value of conversation	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
	Completely helpless	Somewhat helpless	Neutral	Somewhat helpful	Very helpful
<b>6.</b> Whether the conversation hits some tags	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
	Completely helpless	Somewhat helpless	Neutral	Somewhat helpful	Very helpful
<b>7.</b> Location information	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
	Completely helpless	Somewhat helpless	Neutral	Somewhat helpful	Very helpful

## Appendix G

### First Phase User Study - Letter of Approval



**Social Sciences & Humanities Research Ethics Board**  
**Letter of Approval**

June 03, 2013

Ms Shali Liu  
Computer Science\Computer Science

Dear Shali,

**REB #:** 2013-2954  
**Project Title:** Measuring Relatedness and Relevance of Topic Oriented Twitter Conversation  
**Effective Date:** June 03, 2013  
**Expiry Date:** June 03, 2014

The Social Sciences & Humanities Research Ethics Board has reviewed your application for research involving humans and found the proposed research to be in accordance with the Tri-Council Policy Statement on *Ethical Conduct for Research Involving Humans*. This approval will be in effect for 12 months as indicated above. This approval is subject to the conditions listed below which constitute your on-going responsibilities with respect to the ethical conduct of this research.

Sincerely,

Dr. Sophie Jacques, Chair

---

Post REB Approval: On-going Responsibilities of Researchers

After receiving ethical approval for the conduct of research involving humans, there are several ongoing responsibilities that researchers must meet to remain in compliance with University and Tri-Council policies.

#### 1. Additional Research Ethics approval

Prior to conducting any research, researchers must ensure that all required research ethics approvals are secured (in addition to this one). This includes, but is not limited to, securing appropriate research ethics approvals from: other institutions with whom the PI is affiliated; the research institutions of research team members; the institution at which participants may be recruited or from which data may be collected; organizations or groups (e.g. school boards, Aboriginal communities, correctional services, long-term care facilities, service agencies and community groups) and from any other responsible review body or bodies at the research site

#### 2. Reporting adverse events

Any significant adverse events experienced by research participants must be reported **in writing** to Research Ethics **within 24 hours** of their occurrence. Examples of what might be considered “significant” include: an emotional breakdown of a participant during an interview, a negative physical reaction by a participant (e.g. fainting, nausea, unexpected pain, allergic reaction), report by a participant of some sort of negative repercussion from their participation (e.g. reaction of spouse or employer) or complaint by a participant with respect to their participation. The above list is indicative but not all-inclusive. The written report must include details of the adverse event and actions taken by the researcher in response to the incident.

#### 3. Seeking approval for protocol / consent form changes

Prior to implementing any changes to your research plan, whether to the protocol or consent form, researchers must submit them to the Research Ethics Board for review and approval. This is done by completing a Request for Ethics Approval of Amendment to an Approved Project form (available on the website) and submitting three copies of the form and any documents related to the change.

#### 4. Submitting annual reports

Ethics approvals are valid for up to 12 months. Prior to the end of the project’s approval deadline, the researcher must complete an Annual Report (available on the website) and return it to Research Ethics for review and approval before the approval end date in order to prevent a lapse of ethics approval for the research. Researchers should note that no research involving humans may be conducted in the absence of a valid ethical approval and that allowing REB approval to lapse is a violation of University policy, inconsistent with the TCPS (article 6.14) and may result in suspension of research and research funding, as required by the funding agency.

#### 5. Submitting final reports

When the researcher is confident that no further data collection or analysis will be required, a Final Report (available on the website) must be submitted to Research Ethics. This often happens at the time when a

manuscript is submitted for publication or a thesis is submitted for defence. After review and approval of the Final Report, the Research Ethics file will be closed.

#### 6. Retaining records in a secure manner

Researchers must ensure that both during and after the research project, data is securely retained and/or disposed of in such a manner as to comply with confidentiality provisions specified in the protocol and consent forms. This may involve destruction of the data, or continued arrangements for secure storage. Casual storage of old data is not acceptable.

It is the Principal Investigator's responsibility to keep a copy of the REB approval letters. This can be important to demonstrate that research was undertaken with Board approval, which can be a requirement to publish (and is required by the Faculty of Graduate Studies if you are using this research for your thesis).

Please note that the University will securely store your REB project file for 5 years after the study closure date at which point the file records may be permanently destroyed.

#### 7. Current contact information and university affiliation

The Principal Investigator must inform the Research Ethics office of any changes to contact information for the PI (and supervisor, if appropriate), especially the electronic mail address, for the duration of the REB approval. The PI must inform Research Ethics if there is a termination or interruption of his or her affiliation with Dalhousie University.

#### 8. Legal Counsel

The Principal Investigator agrees to comply with all legislative and regulatory requirements that apply to the project. The Principal Investigator agrees to notify the University Legal Counsel office in the event that he or she receives a notice of non-compliance, complaint or other proceeding relating to such requirements.

#### 9. Supervision of students

Faculty must ensure that students conducting research under their supervision are aware of their responsibilities as described above, and have adequate support to conduct their research in a safe and ethical manner.

## Appendix H

### Second Phase User Study - Task Questions

#### **Information Gathering**

1. Top three conversations about online courses
2. Top three conversations talking about weather around universities

#### **Opinion**

3. What do Twitter users think are good places to study in and around SFU campus
4. What do Twitter users think are good places to eat around UBC

#### **Fact Searching**

5. Ranking of universities based on conversations talking about admissions.
6. What is the average range of tuition fee per year?

#### **User Profile Related**

7. Active Twitter users that post research related information for McGill.
8. Active applicants of (UBC|McGill|Dalhousie)

## Appendix I

### Second Phase User Study - Study Protocol

This study protocol below excerpts from Research Ethics Application (REB # 2013-3063) submitted to Dalhousie University under this research. Some of the references have been modified according to the architecture of this thesis.

<p><b>2.3 STUDY DESIGN</b> – In this section</p> <p>2.3.1 state the hypotheses <u>or</u> the research questions <u>or</u> research objectives</p> <p>2.3.2 describe the general study design and how it will address the hypotheses / questions / objectives</p> <p>2.3.3 describe how many participants are needed and how this was determined</p> <p>2.3.4 describe the plan for data analysis in relation to the hypotheses/questions/objectives</p> <p>2.3.5 if a phased review is being requested, describe why this is needed for this study and which phases are contained in this application</p>		
Steps	Description	Approximate time
1. Consent Form (Appendix B.3)		5 minutes
2. Background Questionnaire (Appendix B.4)	Participants' knowledge about Canadian universities, campus life and experience with interactive systems.	5 minutes
3. Training Sessions		10 minutes
4. Tasks		95 minutes
1) Task description	Researcher introduces the tasks	5 minutes
2) Interacting with one	Collecting Twitter conversations with our	90 minutes

assigned version of the proposed system	system for different tasks.	
5. Post Study Interview Questions (Appendix B.6)		5 minutes

### 2.3.1 Research Question

We have two goals for this study:

- 1) Evaluate the effectiveness of the improvements in our system interface;
- 2) Study efficient ways of representing Twitter data.

### 2.3.2 General Study Design

#### Process

We will conduct a laboratory study using members of the Dalhousie University community as participants (primarily students, but recruitment is also open to faculty and staff). We will use a computer with a screen recording system to perform tasks. The procedure of study is described in the following table.

At the end of the study the researcher will conduct a brief unstructured interview, asking participants about possible future improvements of our system and user study design.

#### The Tasks

We will perform a between subjects user study to evaluate the efficiency of the different versions for our system. Each participant will perform tasks on one assigned version of our system. There are four different versions of system, and we will group all the participants into groups with four participants per each group.

#### *Four Different Versions of Our System*

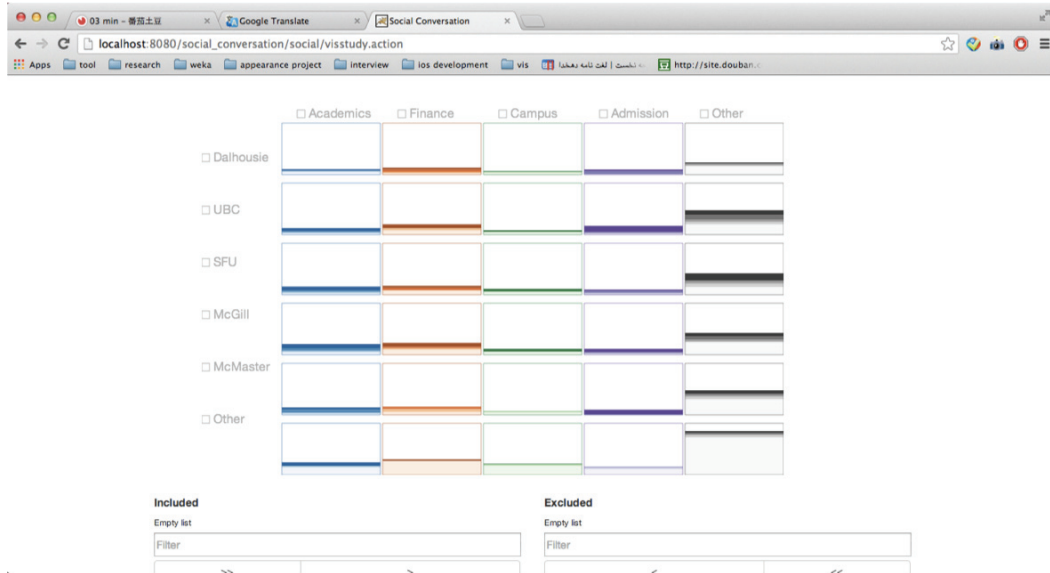
The aim of our system is to evaluate the two factors that we defined for our system: 1) calculated relatedness and informativeness scores; 2) selected features. We have calculated relatedness and informativeness scores for each of the Twitter conversation in our database. The select features includes content



length of the conversation, number of tweets in one conversation, number of users in one conversation, user influence scores, whether the conversation has urls. Our four versions of the system are developed based on a full factorial of these two factors.

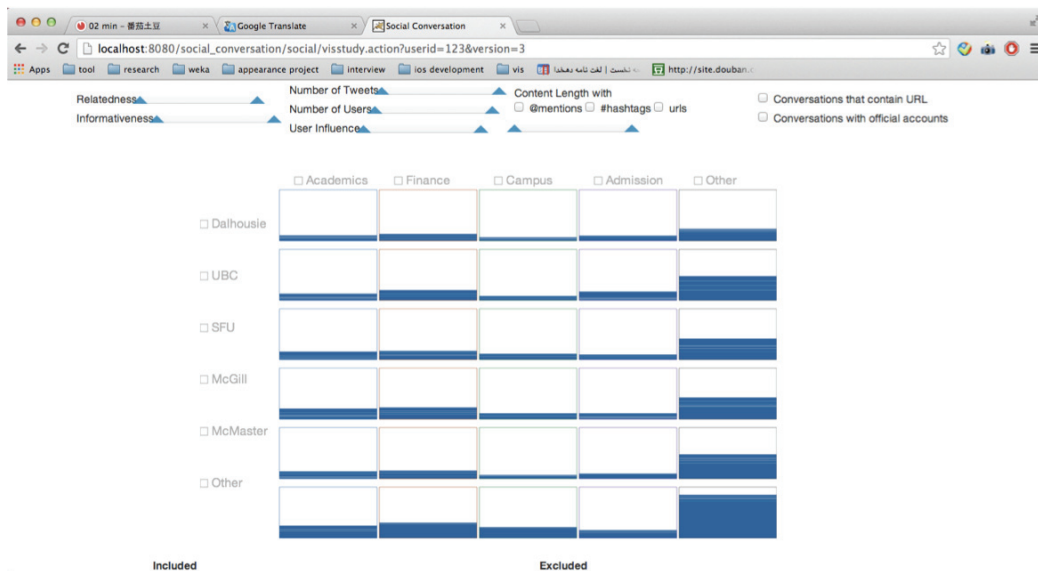
### 1) Scores Only

In this version, the interface will only show score related views.



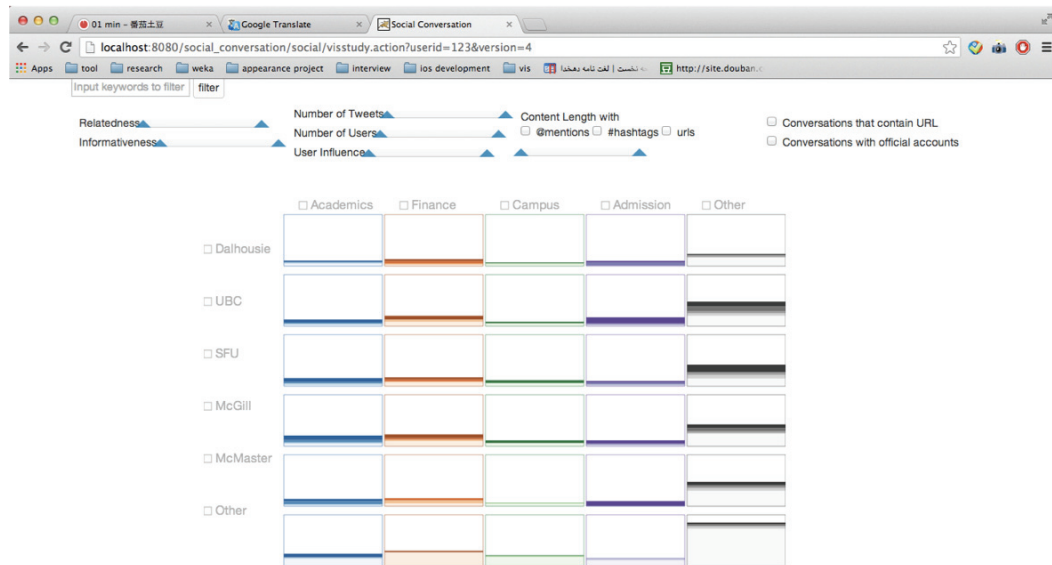
### 2) Features Only

In this version, the interface will only show features related views.



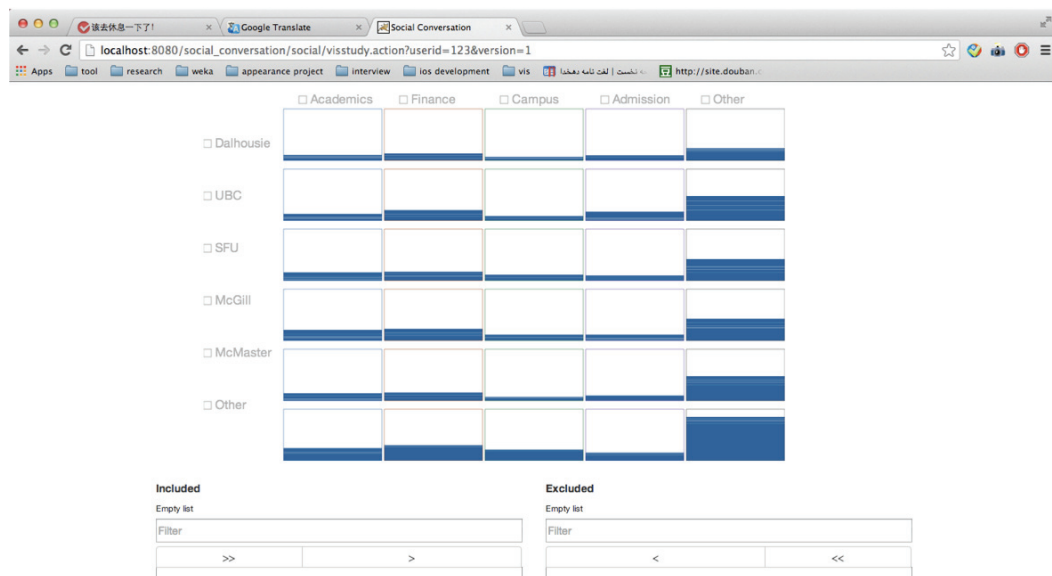
### 3) Scores + Features

In this version, participants will be shown interfaces that combines scores and features views.



#### 4) Basic View

In the basic version, participants will be shown a basic interface that doesn't have any information about scores or features.



Each participant will only be shown one version during the whole study. They will then perform a same set of tasks aiming at answering questions on academic topics with the help of our proposed system and Twitter conversation data.

#### ***Between Subjects Design***

We will adopt between subjects design for this user study. We will ask four participants to perform the same set of tasks at the same time. Each of the participants will only deal with one version of our system. We will group the participants according to their background information so that we can balance variations of the results due to the differences between participants.

### ***Tasks Definition***

We have defined 6 different categories of tasks:

- Information Seeking (e.g. Top 5 useful conversations about course in UBC)
- Opinion / Decision-making (e.g. School ranking based on admission conversations)
- Fact Searching (e.g. Undergrad tuition fee at McGill)
- User Profile Related (e.g. 3 active users talking about research news)
- Statistics (Ranking by number of admissions).

Participants will look for answers to these task questions interacting with our system. These tasks will be performed under a time limit (e.g. 10 minutes per each task). We will pilot for the length of each task and the amount of tasks under each category.

The evaluation of tasks performance will base on time of completion, amount of interactions (clicking, dragging, and so forth) and the context of answers.

### **2.3.3 Participants**

We will recruit up to 50 participants to have a statistics result and check agreements between participants. Since we have four different versions of system, we are considering 12 participants for each version, and the rest 2 participants will be treated as evaluated for our system. We will pilot for variations of possible results, and then decide the exact number of participants. Participants should be Dalhousie University students, staff or faculty. We will require participants to have some knowledge about Canadian universities, Twitter and have experience using interactive systems.

We will assign each participant to a version of system so that we can have 12

participants for each version of our system. Participant will be assigned a version randomly. We will ask our participant to fill in a background questionnaire so that we can group participants with similar background in one group (which contains 4 people who have been assigned 4 different versions) to balance their performance. To be more specific, for example, we will try to group people with similar English proficiency in one group so that the performance will not likely to be affect too much by their reading speed. Our purpose of this grouping is to reduce variations of personal factors that may also affect the performance (such as discipline, education level, language proficiency and so on) so that we can focus on the factors we want to evaluate (response speed, accuracy, moves of interaction and so on).

#### **2.3.4 Data Analysis**

After the study, we will use the questionnaires to help evaluate the output results of the system and collect comments on possible future improvements.

The screen records, audio records will be encrypted and placed on a password-protected computer account accessible only to the investigators. All the data will be kept anonymous and safe.

The study results will be reported as a research paper based on qualitative and quantitative analysis of interaction data combined with questionnaire data, aggregate information, charts and visualizations, and statistics. We will use the study results to evaluate and improve our system. Results from the study will be submitted to appropriate conferences or journals in the area of Information Visualization, Human Computer Interaction, Social Media or Artificial Intelligence.

#### **2.3.5**

A phased review is required. We propose two phases for our research. In the Phase 1 (which has been approved with a reference number REB 2013-2954), participants were asked to rate the relatedness and relevance of conversations. In the Phase 2 described in this application, participants will be shown a visualization tool built from results of the first study and asked to interact with it. We will evaluate the effectiveness of visualizations.

**2.4 RECRUITMENT** – In this section, for each type of participant to be recruited, describe

- 2.4.1 the study population
- 2.4.2 any social / cultural / safety considerations
- 2.4.3 and justify all specific inclusion / exclusion criteria of participants
- 2.4.4 any recruitment instruments (attach copies)
- 2.4.5 who will be doing the recruitment and what actions they will take
- 2.4.6 any screening measures, and how they will be used (attach copies)
- 2.4.7 any permissions that are needed and attach letters

#### **2.4.1 Study Population**

The study population will be Dalhousie University students, faculty and staff.

#### **2.4.2 Considerations**

N/A

#### **2.4.3**

Participants should be familiar with names of top universities in Canada, Twitter and have experience in interactive systems. Since we are going to recruit participants within school communities, most candidates are expected to have some knowledge about academic institutions and campus life.

#### **2.4.4-2.4.5 Recruitment Procedures and Instruments**

Shali Liu will recruit Dalhousie University students, staff and faculty to take part in this study. All participants will be recruited by email recruitment notice through events calendars (i.e. Notice Digest, notice.digest@dal.ca) and through mailing lists (i.e. cs.all@dal.ca), which is also a monitored mailing list. We may also post recruitment notice through public websites such as Facebook and Twitter. In the recruitment notice, participants will be asked to email their interest in participating in the listed researcher. The participant and researcher will then communicate to find an appropriate time for the participant to do the study. The recruitment notice script is shown in Appendix B.2.

#### **2.4.6 Screening Measures**

Anybody within the study population, have knowledge on universities and Twitter, and have basic skills about using computer can participate.

#### **2.4.7 Permissions**

N/A

### **2.5 INFORMED CONSENT PROCESS – In this section**

2.5.1 describe the informed consent process (attach a copy of all consent forms)

2.5.2 if oral consent is desired, describe why it is necessary and how it will be done (attach a  
copy of the script)

2.5.3 if a waiver of informed consent is sought, explain why and describe how the four criteria  
needed for this are met

2.5.4 for third party consent (with or without assent), describe how this will be done

2.5.5 describe plans (if any) for on-going consent

2.5.6 if community consent is needed, describe how it will be obtained

#### **2.5.1 Informed Consent Process**

All 10 participants involved in the study will sign an informed consent form (see Appendix B.3). This will be administered by a researcher at the initial meeting of the study. The informed consent outlines the risks and benefits associated with the study, a description of the study, the participant's right to withdraw without consequence, and assurances of confidentiality and anonymity of personal data. As well, the informed consent states to participants that they can withdraw from the study at any time without loss of compensation.

#### **2.5.2-2.5.6**

N/A.

### **2.6 DETAILED METHODOLOGY - In this section describe**

- 2.6.1 where the research will be conducted
- 2.6.2 what participants will be asked to do and the time each task will take (plus total time)
- 2.6.3 what data will be recorded and what research instruments will be used (attach copies)
- 2.6.4 the roles and qualifications of the study investigators / research staff
- 2.6.5 how long the participants will be involved in each part of the study

### **2.6.1 Location of Research**

The study will be conducted in a quiet private meeting room in Dalhousie University.

### **2.6.2 Study Details**

We will meet with participants at selected location. This study will take about 120 minutes. Before study, we will get informed consent (Appendix B.3) from participants. We will then discuss the study and answer any questions prior to starting the study. After that, we will have the participants perform the study, and fill in questionnaires.

During the study, participants will perform two tasks, using a dedicated desktop or laptop computer, with which participants will interact with our system. After the participants finish the first task, we will also have a brief interview with participants at the end of the study (Appendix B.6).

### **2.6.3 Data Recorded and Instruments Used**

We will use five main data collection methods during this study: Screen recording, interview notes, and background questionnaires, and post-task questionnaires.

#### **Screen recording**

We will record screen during participants' interacting with the system. We will be using screen recorder software to record mouse moving and clicking. Through screen records, we will keep track of any difficulties in performing any of the tasks to help determine the effectiveness of the options.

#### **Interview notes and Audio Recording**

Interview notes are taken by interviewing participants (See interview questions in

Appendix B.6). We will also record the audio during the interview using phone, laptop or other electronic equipment that has audio function. The audio will be deleted after we copy important information to our interview notes. These notes will be used for the improvements of our system and future study design.

### **Background Questionnaire**

The questions (Appendix B.4) will ask the participants about their knowledge of top universities in Canada, twitter usage and experience in interactive systems. These questions help us to know more about participants' experiences and preferences.

### **Post-task Questionnaire**

The post-condition questions (Appendix B.5) will ask the participants to rate the tweets they selected. This will help evaluate our system.

### **2.6.4 Roles/qualifications of Study Investigators**

Primary investigator Shali Liu will be responsible for recruitment, for administering the study and analysis of the results. The investigators have designed the study, including the questionnaires, and prototype tools used during the study. Shali will be the primary researcher responsible for reviewing and accessing results. Dr. Evangelos Milios and Dr. Kirstie Hawkey will provide guidance and assistance during analysis.

### **2.6.5 Length of Participation**

This user study should take each participant about two hours to complete.



## Appendix J

### **Second Phase User Study - Recruitment Notice**

We are recruiting participants to take part in a research study that evaluates an interactive visualization system. We are looking for members of the Dalhousie University community, who have knowledge about Canadian universities, twitter and experience in using interactive systems.

The study will be conducted in a quiet private meeting room in Dalhousie University. This study will take about two hours to complete. You will first meet with a researcher to go over the study details, give consent to do the study and fill in a background questionnaire. You will then complete two sections of tasks with the application that will be provided, and you will fill in questionnaires after doing each task. Compensation is \$30 for participation in the study.

If you are interested in participating, please contact Shali Liu ([shali@dal.ca](mailto:shali@dal.ca)).

## Appendix K



### Second Phase User Study - Informed Consent

#### **Evaluating Effectiveness of an Interactive Visualization System on Twitter Data**

Principal Investigators: Shali Liu, Faculty of Computer Science, shali@dal.ca

Dr. Evangelos Milios, Faculty of Computer Science,  
eem@cs.dal.ca

Dr. Kirstie Hawkey, Faculty of Computer Science,  
hawkey@cs.dal.ca

Contact Person: Shali Liu, Faculty of Computer Science, [shali@dal.ca](mailto:shali@dal.ca)  
(902-999-2151)

We invite you to take part in a research study being conducted by Shali Liu at Dalhousie University. This description tells you about the risks, inconvenience, or discomfort which you might experience. Participating in the study might not benefit you, but we might learn things that will benefit others. You should discuss any questions you have about this study with Shali Liu.

The purpose of this study is to evaluate the effectiveness of our proposed interactive visualization system on academic Twitter data. To be eligible to participate in the study, you must be a Dalhousie University student, staff or faculty, have knowledge about Canadian universities, and twitter, as well as have some experience with interactive visualization systems (i.e. interacting with networking visualizations). Participating in this study will take about two hours to complete. You will be compensated \$30 for participating in this study; you can withdraw from the study at any time without loss of compensation. A researcher will be accompanying you during the study period to answer any questions you may have or address any problems that you may experience with the tasks.

You will be asked to fill in a background questionnaire detailing your knowledge about your age, education level, English fluency, nationality, Canadian universities, your twitter usage and experience with visualization systems. This information will be used to assign you a group. We will only keep your personal background information together with your name and contact information till you get the assignment for a group. Then your background information will only be marked with a unique identity (without name or contact information) that cannot infer to you in the future.

At the beginning of the study, you will meet with an investigator (in a quiet, private meeting room in Dalhousie University). You will be asked give consent to do the study. You will then perform two tasks interacting with our system using a desktop or laptop computer. You will feel the same as using other normal monitors. In addition, the software used by participants will log and timestamp all interactions. You will perform a set of tasks that ask you to find out Twitter conversations towards perspectives, facts or statistics on academic using our system. Sample tasks include finding the top five Twitter conversations under a specific condition (e.g. courses in Dalhousie) and statistics about the admission of students to different universities. We will also have a short interview with you about your study experience and we will ask about possible future improvements to our system. The result tweets that you select for our study tasks from your interacting with our system will be stored. We will also take notes during the interview at the end and audio taping the interview to help complete out notes.

There are very low risks associated with this study. There is a low risk that you may become frustrated or embarrassed if you experience some difficulties performing the tasks during the study. The researcher will always be available during the study to answer any questions you may have.

All the data mentioned above will be kept confidential and will be de-identified eventually. The informed consent form and all research data will be kept in a secure location under confidentiality. You have the option of allowing direct quotes to be included in any final reports.

In the event that you have any difficulties with, or wish to voice concern about, any aspect of your participation in this study, you may contact Catherine Connors, Director, Office of Research Ethics Administration at Dalhousie University's Office of Human Research Ethics for assistance: phone: (902) 494-1462, email: Catherine.connors@dal.ca.

*"I have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I hereby consent to take part in the study. However, I understand that my participation is voluntary and that I am free to withdraw from the study at any time."*

**Participant**

**Researcher**

Name: \_\_\_\_\_ Name: \_\_\_\_\_

Signature: \_\_\_\_\_ Signature: \_\_\_\_\_

**Date:** \_\_\_\_\_ **Date:** \_\_\_\_\_

*"I agree to let you directly quote any comments or statements made in any written reports without viewing the quotes prior to their use and I understand that the anonymity of textual data will be preserved by using pseudonyms."*

**Participant**

**Researcher**

Name: \_\_\_\_\_ Name: \_\_\_\_\_

Signature: \_\_\_\_\_ Signature: \_\_\_\_\_

**Date:** \_\_\_\_\_ **Date:** \_\_\_\_\_

If you are interested in seeing the results of this study, please check below and provide your email address. We will contact you with publication details that describe the results.

*"I would like to be notified by email when results are available via a publication."*

[if this option is chosen, please include a contact email address:  
\_\_\_\_\_]

## Appendix L

### Second Phase User Study - Background Questionnaire

#### PART I - PLEASE FILL IN THE FOLLOWING INFORMATION:

**Identification number:** \_\_\_\_\_

1. Age:

- 15-25                       25-35                       35-45                       over 45

2. Gender:

- Male                               Female

3. What is your role in school?

- Graduate                       Undergraduate                       Faculty                       Staff

4. What is your education level?

- High School                       Bachelor                       Master                       PhD and  
above

5. What do you think about your English proficiency?

- Native speaker                       Second Language and quite fluent  
 Second Language and OK                       Second Language and less fluent

6. Your nationality is \_\_\_\_\_

7. What is your major? \_\_\_\_\_

8. Which of the Canadian top universities are you familiar with (familiar with names and some abbreviation such as UBC)?

Here is the list of universities that we focus on:

- Dalhousie University*  
 *McGill University*  
 *McMaster University*  
 *University of British Columbia*  
 *Simon Fraser University*

9. On the average, how much time do you spend per day on a computer?

- Less than 1 hour                       1 to less than 5 hours                       5 to less than 10 hours                        
Over 10 hours

10. Do you have an account in Twitter?

Yes       No

11. How frequently do you use Twitter?

At Least once per day       A few times per week

A few times per month       Seldom use

Not at all

12. Do you have previous experience of using an interactive system (e.g. network visualizations, google maps)?

Yes       No

## Appendix M

### Second Phase User Study - Post-task Questionnaires

**Please respond to the following statements using the given scale (circle response):**

The **informativeness** is whether this conversation is helpful or informative to present some aspects of university such as research, course, scholarship, campus life or admission.

For example, "I am here in Dalhousie University" is less informative than "Introducing a new scholarship for graduate students at Dalhousie University [http:...](#)".

	1	2	3	4	5
	<b>NOT</b> Informative at All	Somewhat <b>NOT</b> Informative	Neutral	Somewhat Informative	Very Informative
<b>1.</b> Selected tweet 1 (results varies from different participants)	<b>1</b> NOT Informative at All	<b>2</b> Somewhat NOT Informative	<b>3</b> Neutral	<b>4</b> Somewhat Informative	<b>5</b> Very Informative
<b>2.</b> Selected tweet 2	<b>1</b> NOT Informative at All	<b>2</b> Somewhat NOT Informative	<b>3</b> Neutral	<b>4</b> Somewhat Informative	<b>5</b> Very Informative

## Appendix N

### Second Phase User Study - Post Study Interview Questions

1. Do you think the calculated informativeness and relatedness scores help you find the answers? Do you trust the scores? (For versions that contain scores information)
2. Do you think the features views (filters, features axes) help you locating the answers? Why? (For versions that contain features information)
3. Can you give any comments on improvement of our user study design?

Any comments on improvement of our system?



## Appendix O

### Second Phase User Study - Letter of Approval



**Social Sciences & Humanities Research Ethics Board**  
**Letter of Approval**

June 03, 2013

Ms Shali Liu  
Computer Science\Computer Science

Dear Shali,

**REB #:** 2013-2954  
**Project Title:** Measuring Relatedness and Relevance of Topic Oriented Twitter Conversation  
**Effective Date:** June 03, 2013  
**Expiry Date:** June 03, 2014

The Social Sciences & Humanities Research Ethics Board has reviewed your application for research involving humans and found the proposed research to be in accordance with the Tri-Council Policy Statement on *Ethical Conduct for Research Involving Humans*. This approval will be in effect for 12 months as indicated above. This approval is subject to the conditions listed below which constitute your on-going responsibilities with respect to the ethical conduct of this research.

Sincerely,

Dr. Sophie Jacques, Chair

---

Post REB Approval: On-going Responsibilities of Researchers

After receiving ethical approval for the conduct of research involving humans, there are several ongoing responsibilities that researchers must meet to remain in compliance with University and Tri-Council policies.

#### 1. Additional Research Ethics approval

Prior to conducting any research, researchers must ensure that all required research ethics approvals are secured (in addition to this one). This includes, but is not limited to, securing appropriate research ethics approvals from: other institutions with whom the PI is affiliated; the research institutions of research team members; the institution at which participants may be recruited or from which data may be collected; organizations or groups (e.g. school boards, Aboriginal communities, correctional services, long-term care facilities, service agencies and community groups) and from any other responsible review body or bodies at the research site

#### 2. Reporting adverse events

Any significant adverse events experienced by research participants must be reported **in writing** to Research Ethics **within 24 hours** of their occurrence. Examples of what might be considered “significant” include: an emotional breakdown of a participant during an interview, a negative physical reaction by a participant (e.g. fainting, nausea, unexpected pain, allergic reaction), report by a participant of some sort of negative repercussion from their participation (e.g. reaction of spouse or employer) or complaint by a participant with respect to their participation. The above list is indicative but not all-inclusive. The written report must include details of the adverse event and actions taken by the researcher in response to the incident.

#### 3. Seeking approval for protocol / consent form changes

Prior to implementing any changes to your research plan, whether to the protocol or consent form, researchers must submit them to the Research Ethics Board for review and approval. This is done by completing a Request for Ethics Approval of Amendment to an Approved Project form (available on the website) and submitting three copies of the form and any documents related to the change.

#### 4. Submitting annual reports

Ethics approvals are valid for up to 12 months. Prior to the end of the project’s approval deadline, the researcher must complete an Annual Report (available on the website) and return it to Research Ethics for review and approval before the approval end date in order to prevent a lapse of ethics approval for the research. Researchers should note that no research involving humans may be conducted in the absence of a valid ethical approval and that allowing REB approval to lapse is a violation of University policy, inconsistent with the TCPS (article 6.14) and may result in suspension of research and research funding, as required by the funding agency.

#### 5. Submitting final reports

When the researcher is confident that no further data collection or analysis will be required, a Final Report (available on the website) must be submitted to Research Ethics. This often happens at the time when a

manuscript is submitted for publication or a thesis is submitted for defence. After review and approval of the Final Report, the Research Ethics file will be closed.

#### 6. Retaining records in a secure manner

Researchers must ensure that both during and after the research project, data is securely retained and/or disposed of in such a manner as to comply with confidentiality provisions specified in the protocol and consent forms. This may involve destruction of the data, or continued arrangements for secure storage. Casual storage of old data is not acceptable.

It is the Principal Investigator's responsibility to keep a copy of the REB approval letters. This can be important to demonstrate that research was undertaken with Board approval, which can be a requirement to publish (and is required by the Faculty of Graduate Studies if you are using this research for your thesis).

Please note that the University will securely store your REB project file for 5 years after the study closure date at which point the file records may be permanently destroyed.

#### 7. Current contact information and university affiliation

The Principal Investigator must inform the Research Ethics office of any changes to contact information for the PI (and supervisor, if appropriate), especially the electronic mail address, for the duration of the REB approval. The PI must inform Research Ethics if there is a termination or interruption of his or her affiliation with Dalhousie University.

#### 8. Legal Counsel

The Principal Investigator agrees to comply with all legislative and regulatory requirements that apply to the project. The Principal Investigator agrees to notify the University Legal Counsel office in the event that he or she receives a notice of non-compliance, complaint or other proceeding relating to such requirements.

#### 9. Supervision of students

Faculty must ensure that students conducting research under their supervision are aware of their responsibilities as described above, and have adequate support to conduct their research in a safe and ethical manner.

## Appendix P

### Second Phase User Study - Post Study Interview Questions Results

Scores	Features	Study Design	System
	Yes, if keywords are used in filter, it is much easier for us to locate the valuable targets	Not Really	Maybe the system could be faster and more robust
Yes, defintely	Yes, the filter is very useful. But I did not being asked to do axes	Your user study design is already very nice. But it can be a shorter time.	No.
Yes. Yes.	Filters really helpful	None	Advanced Filtering
Yes, I trust the scores.	Yes, it can filter the useless information and help me find the usefule information faster.	None	None
Yes	Filters: Fast and convenient		

<p>Yes. All the decisions were based on the scores.</p>	<p>Yes. Filters can filter some options to save the time.</p>		
<p>Yes.</p>	<p>I did, but for some questions I looked for every possible conversations.</p>	<p>Yes, because there were too many conversations there.</p>	<p>* Improve your searching functionality like logical AND, OR, NOT in queries.          * User different colors, I had problems with faded colors.          * I've lost I'm looking at circles of universities or circles of aspects</p>
	<p>Yes, because there are so many reviews and filtering out them would help a lot.</p>		
			<ol style="list-style-type: none"> <li>1. The keywords (Filters) could have been highlighted</li> <li>2. More information could be given about the user.</li> </ol>

They did. Bigger circles and stronger colors contained more interesting information.		Sometimes the questions were a bit open-ended	- User information on the interface and filtering based on user influence.
		It is OK.	Filtering can be improved.
	Yes. There was still too much data for me to look at every conversation by hand - the filters helped cut it down to a manageable size.	It may have been a little long. (If performance drops toward the end, it may be because of user fatigue and not because the visualization is unhelpful for the last tasks)	- There seemed to be times when I would include/exclude a conversation, change the filtering, and see the same conversation in the graph again. - The general filter (matrix) didn't seem to filter the conversations. - Sometimes the animation was too fast for me to stay ** when the axes changed. - I would have it very useful to see my search term text highlighted in the conversation text.
Yes, two out of 8 questions were completely based on two scores.			
Yes. Not really trust.			
Yes, they did. I trust the scores to some degree.	Yes. Especially the filter view helps a lot.	It's better to offer a guideline in detail for the system.	There exist overlap problems.

	<p>Yes, the filters do. The filters may not work really well because of the scales and large overlapping.</p> <p>But it selects relevant information and saves a lot of time.</p>		<p>The function of the filter, the axes scale. Separate graphs could be applied for some extreme situation.</p>
Yes.		Yes, it helped me narrow down the answers.	
Yes, I always click on higher score to find information. I am not sure if I am trusting the scores.			
	Yes, I like the color difference.		If the axel can be changed, it will be better
Yes, I do. I think feature trust them to some degree.	<p>The filters can help me located the keywords.</p> <p>However, I don't think feature axes can help me too much.</p>		It is better to highlight the keyword that the user searched in the conversation.

<p>Not really - I did not understand the score and it did not seem to help whether I selected the informative or less informative conversations. I did not quite trust the scores.</p>			<p>Being able to keep the conversation under pinned in a certain location.  Being able to see which conversations in the graph I've already seen.  As the location of circles did not really have any meaning (location within given topic / university group), I would like better were list-like arrangement so that there is an order to go through the conversation.</p>
	<p>Filters were quite useful, but axes not so much.</p>	<p>There were a lot of tweets that are irrelevant. Pruning the tweets maybe?</p>	<p>System works quite nicely. No comment comes to mind.</p>
<p>Yes, it helps and I did trust.</p>		<p>Good.</p>	<p>Works well.</p>
<p>Yes I do. Although scores are not 100% correct, they were helpful.</p>	<p>Filter option was very helpful, as well as axis (but I didn't use axis features as much as search filter)</p>		<p>Maybe add "Deselect All" button which can be used in cases when only one university / or topic needs to be looked up.</p>
<p>Yes.</p>			<p>Overall it is interesting.</p>
	<p>Yes. Narrowing the scope</p>		



I think it helps me find the answer. I trust the score.	Help me find the answer. It decreases the time to find answers.		The speed can be improved.
			Should be able to search within conversations that already included or excluded
	Trying to click on the tiny nodes can be hard, when the graph is really dense.		Provide functionality to enumerate through queries.
Yes, however the score may not effect the results.	Yes, by filtering it allows me to find out the related topic that I am looking for.		
No, I didn't find it quite informative		Demo Video	Better search features Most influential can have a separate option Different shapes can distinctively show the difference between university or topics The general view was not useful, try to add some features like drill down to the specific categories.

<p>It's helpful, but don't trust some of the result.</p>	<p>Axes is helpful only in case that users want to know the popularity</p>	<p>Can randomly arrange the order of questions to acquire more accurate results.</p>	
			<p>Use arrows to show next tweet Nodes change place after included them into including list</p>
	<p>Yes. Butter visualization of search results.</p>		
	<p>Yes. They saved time. Also, they contributed to the accuracy of the answers.</p>	<p>You might consider reducing time (from 15min to 10min) and increase the number of tasks to 10</p>	

## Appendix Q

### Hits / Clicks Raw Results

Participant	Task	TaskCategory	Version	HasScore	HasFeature	TotalClicks	Hits	HitRatio
10	1	1	1	0	0	30	2	0.066667
10	2	1	1	0	0	30	5	0.166667
10	3	2	1	0	0	41	5	0.121951
10	4	2	1	0	0	47	10	0.212766
10	5	3	1	0	0	54	5	0.092593
10	6	3	1	0	0	11	4	0.363636
10	7	4	1	0	0	39	1	0.025641
10	8	4	1	0	0	55	6	0.109091
11	1	1	2	1	0	36	2	0.055556
11	2	1	2	1	0	54	5	0.092593
11	3	2	2	1	0	60	7	0.116667
11	4	2	2	1	0	45	11	0.244444
11	5	3	2	1	0	49	15	0.306122
11	6	3	2	1	0	15	4	0.266667
11	7	4	2	1	0	37	3	0.081081
11	8	4	2	1	0	45	9	0.2
12	1	1	4	1	1	7	0	0
12	2	1	4	1	1	15	0	0
12	3	2	4	1	1	9	0	0
12	4	2	4	1	1	16	0	0
12	5	3	4	1	1	7	0	0
12	6	3	4	1	1	5	0	0
12	7	4	4	1	1	11	0	0
12	8	4	4	1	1	25	1	0.04
13	1	1	3	0	1	41	3	0.073171
13	2	1	3	0	1	22	5	0.227273
13	3	2	3	0	1	38	6	0.157895
13	4	2	3	0	1	51	15	0.294118
13	5	3	3	0	1	89	7	0.078652
13	6	3	3	0	1	28	4	0.142857

13	7	4	3	0	1	75	14	0.186667
13	8	4	3	0	1	87	18	0.206897
14	1	1	4	1	1	29	3	0.103448
14	2	1	4	1	1	52	2	0.038462
14	3	2	4	1	1	37	5	0.135135
14	4	2	4	1	1	16	8	0.5
14	5	3	4	1	1	23	1	0.043478
14	6	3	4	1	1	25	4	0.16
14	7	4	4	1	1	13	1	0.076923
14	8	4	4	1	1	22	8	0.363636
15	1	1	2	1	0	9	0	0
15	2	1	2	1	0	17	4	0.235294
15	3	2	2	1	0	14	5	0.357143
15	4	2	2	1	0	14	6	0.428571
15	5	3	2	1	0	24	0	0
15	6	3	2	1	0	26	5	0.192308
15	7	4	2	1	0	9	0	0
15	8	4	2	1	0	7	0	0
16	1	1	1	0	0	35	0	0
16	2	1	1	0	0	41	0	0
16	3	2	1	0	0	47	0	0
16	4	2	1	0	0	14	4	0.285714
16	5	3	1	0	0	13	3	0.230769
16	6	3	1	0	0	29	3	0.103448
16	7	4	1	0	0	58	0	0
16	8	4	1	0	0	140	0	0
17	1	1	4	1	1	34	3	0.088235
17	2	1	4	1	1	25	2	0.08
17	3	2	4	1	1	8	2	0.25
17	4	2	4	1	1	13	4	0.307692
17	5	3	4	1	1	11	1	0.090909
17	6	3	4	1	1	6	4	0.666667
17	7	4	4	1	1	6	1	0.166667
17	8	4	4	1	1	33	8	0.242424
18	1	1	3	0	1	30	2	0.066667
18	2	1	3	0	1	46	4	0.086957
18	3	2	3	0	1	49	2	0.040816
18	4	2	3	0	1	36	10	0.277778

18	5	3	3	0	1	31	0	0
18	6	3	3	0	1	49	8	0.163265
18	7	4	3	0	1	45	13	0.288889
18	8	4	3	0	1	149	9	0.060403
19	1	1	4	1	1	14	2	0.142857
19	2	1	4	1	1	9	4	0.444444
19	3	2	4	1	1	28	5	0.178571
19	4	2	4	1	1	11	4	0.363636
19	5	3	4	1	1	14	5	0.357143
19	6	3	4	1	1	11	4	0.363636
19	7	4	4	1	1	25	1	0.04
19	8	4	4	1	1	18	2	0.111111
1	1	1	4	1	1	41	2	0.04878
1	2	1	4	1	1	49	7	0.142857
1	3	2	4	1	1	37	5	0.135135
1	4	2	4	1	1	20	6	0.3
1	5	3	4	1	1	19	3	0.157895
1	6	3	4	1	1	11	4	0.363636
1	7	4	4	1	1	29	7	0.241379
1	8	4	4	1	1	23	7	0.304348
20	1	1	1	0	0	42	2	0.047619
20	2	1	1	0	0	15	4	0.266667
20	3	2	1	0	0	42	9	0.214286
20	4	2	1	0	0	31	11	0.354839
20	5	3	1	0	0	20	2	0.1
20	6	3	1	0	0	21	5	0.238095
20	7	4	1	0	0	135	9	0.066667
20	8	4	1	0	0	70	11	0.157143
21	1	1	2	1	0	27	2	0.074074
21	2	1	2	1	0	56	14	0.25
21	3	2	2	1	0	40	5	0.125
21	4	2	2	1	0	21	6	0.285714
21	5	3	2	1	0	88	24	0.272727
21	6	3	2	1	0	16	5	0.3125
21	7	4	2	1	0	116	15	0.12931
21	8	4	2	1	0	138	12	0.086957
22	1	1	3	0	1	11	1	0.090909
22	2	1	3	0	1	17	5	0.294118

22	3	2	3	0	1	34	2	0.058824
22	4	2	3	0	1	16	5	0.3125
22	5	3	3	0	1	12	4	0.333333
22	6	3	3	0	1	12	4	0.333333
22	7	4	3	0	1	22	0	0
22	8	4	3	0	1	8	0	0
23	1	1	4	1	1	17	2	0.117647
23	2	1	4	1	1	24	7	0.291667
23	3	2	4	1	1	47	2	0.042553
23	4	2	4	1	1	24	8	0.333333
23	5	3	4	1	1	19	2	0.105263
23	6	3	4	1	1	35	4	0.114286
23	7	4	4	1	1	69	17	0.246377
23	8	4	4	1	1	85	16	0.188235
24	1	1	1	0	0	78	5	0.064103
24	2	1	1	0	0	68	7	0.102941
24	3	2	1	0	0	134	5	0.037313
24	4	2	1	0	0	64	11	0.171875
24	5	3	1	0	0	47	1	0.021277
24	6	3	1	0	0	54	5	0.092593
24	7	4	1	0	0	71	8	0.112676
24	8	4	1	0	0	223	6	0.026906
25	1	1	2	1	0	25	3	0.12
25	2	1	2	1	0	73	4	0.054795
25	3	2	2	1	0	35	5	0.142857
25	4	2	2	1	0	52	17	0.326923
25	5	3	2	1	0	32	3	0.09375
25	6	3	2	1	0	25	7	0.28
25	7	4	2	1	0	12	2	0.166667
25	8	4	2	1	0	102	14	0.137255
26	1	1	3	0	1	35	3	0.085714
26	2	1	3	0	1	37	4	0.108108
26	3	2	3	0	1	39	7	0.179487
26	4	2	3	0	1	23	7	0.304348
26	5	3	3	0	1	30	7	0.233333
26	6	3	3	0	1	24	4	0.166667
26	7	4	3	0	1	25	8	0.32
26	8	4	3	0	1	65	0	0

27	1	1	2	1	0	17	2	0.117647
27	2	1	2	1	0	14	4	0.285714
27	3	2	2	1	0	39	5	0.128205
27	4	2	2	1	0	17	6	0.352941
27	5	3	2	1	0	19	0	0
27	6	3	2	1	0	14	4	0.285714
27	7	4	2	1	0	55	4	0.072727
27	8	4	2	1	0	95	20	0.210526
28	1	1	1	0	0	41	1	0.02439
28	2	1	1	0	0	9	4	0.444444
28	3	2	1	0	0	12	3	0.25
28	4	2	1	0	0	14	5	0.357143
28	5	3	1	0	0	2	0	0
28	6	3	1	0	0	12	5	0.416667
28	7	4	1	0	0	12	1	0.083333
28	8	4	1	0	0	13	5	0.384615
29	1	1	4	1	1	42	5	0.119048
29	2	1	4	1	1	35	10	0.285714
29	3	2	4	1	1	34	8	0.235294
29	4	2	4	1	1	23	6	0.26087
29	5	3	4	1	1	69	0	0
29	6	3	4	1	1	25	6	0.24
29	7	4	4	1	1	48	10	0.208333
29	8	4	4	1	1	27	10	0.37037
2	1	1	3	0	1	12	1	0.083333
2	2	1	3	0	1	14	4	0.285714
2	3	2	3	0	1	29	3	0.103448
2	4	2	3	0	1	83	0	0
2	5	3	3	0	1	30	0	0
2	6	3	3	0	1	10	3	0.3
2	7	4	3	0	1	15	5	0.333333
2	8	4	3	0	1	10	4	0.4
30	1	1	1	0	0	30	1	0.033333
30	2	1	1	0	0	26	3	0.115385
30	3	2	1	0	0	31	5	0.16129
30	4	2	1	0	0	5	2	0.4
30	5	3	1	0	0	17	4	0.235294
30	6	3	1	0	0	11	4	0.363636

30	7	4	1	0	0	16	2	0.125
30	8	4	1	0	0	18	0	0
31	1	1	2	1	0	13	1	0.076923
31	2	1	2	1	0	28	6	0.214286
31	3	2	2	1	0	15	2	0.133333
31	4	2	2	1	0	17	3	0.176471
31	5	3	2	1	0	13	3	0.230769
31	6	3	2	1	0	12	4	0.333333
31	7	4	2	1	0	9	1	0.111111
31	8	4	2	1	0	22	2	0.090909
32	1	1	1	0	0	28	2	0.071429
32	2	1	1	0	0	38	7	0.184211
32	3	2	1	0	0	25	3	0.12
32	4	2	1	0	0	41	14	0.341463
32	5	3	1	0	0	83	7	0.084337
32	6	3	1	0	0	20	6	0.3
32	7	4	1	0	0	150	0	0
32	8	4	1	0	0	243	20	0.082305
33	1	1	1	0	0	44	2	0.045455
33	2	1	1	0	0	27	7	0.259259
33	3	2	1	0	0	42	8	0.190476
33	4	2	1	0	0	41	11	0.268293
33	5	3	1	0	0	26	9	0.346154
33	6	3	1	0	0	11	4	0.363636
33	7	4	1	0	0	58	6	0.103448
33	8	4	1	0	0	204	10	0.04902
34	1	1	3	0	1	12	3	0.25
34	2	1	3	0	1	10	2	0.2
34	3	2	3	0	1	14	2	0.142857
34	4	2	3	0	1	14	5	0.357143
34	5	3	3	0	1	10	0	0
34	6	3	3	0	1	9	2	0.222222
34	7	4	3	0	1	9	0	0
34	8	4	3	0	1	6	0	0
35	1	1	4	1	1	41	1	0.02439
35	2	1	4	1	1	34	4	0.117647
35	3	2	4	1	1	44	3	0.068182
35	4	2	4	1	1	39	13	0.333333



35	5	3	4	1	1	65	2	0.030769
35	6	3	4	1	1	16	5	0.3125
35	7	4	4	1	1	37	8	0.216216
35	8	4	4	1	1	54	13	0.240741
36	1	1	3	0	1	35	3	0.085714
36	2	1	3	0	1	10	4	0.4
36	3	2	3	0	1	35	3	0.085714
36	4	2	3	0	1	30	6	0.2
36	5	3	3	0	1	8	6	0.75
36	6	3	3	0	1	11	4	0.363636
36	7	4	3	0	1	15	8	0.533333
36	8	4	3	0	1	32	2	0.0625
37	1	1	3	0	1	87	8	0.091954
37	2	1	3	0	1	117	3	0.025641
37	3	2	3	0	1	38	5	0.131579
37	4	2	3	0	1	23	7	0.304348
37	5	3	3	0	1	54	1	0.018519
37	6	3	3	0	1	17	5	0.294118
37	7	4	3	0	1	14	10	0.714286
37	8	4	3	0	1	8	0	0
38	1	1	1	0	0	49	0	0
38	2	1	1	0	0	23	7	0.304348
38	3	2	1	0	0	36	9	0.25
38	4	2	1	0	0	46	11	0.23913
38	5	3	1	0	0	39	7	0.179487
38	6	3	1	0	0	20	5	0.25
38	7	4	1	0	0	67	5	0.074627
38	8	4	1	0	0	158	2	0.012658
39	1	1	4	1	1	72	5	0.069444
39	2	1	4	1	1	48	9	0.1875
39	3	2	4	1	1	60	10	0.166667
39	4	2	4	1	1	49	13	0.265306
39	5	3	4	1	1	116	11	0.094828
39	6	3	4	1	1	12	5	0.416667
39	7	4	4	1	1	54	18	0.333333
39	8	4	4	1	1	43	8	0.186047
3	1	1	2	1	0	15	2	0.133333
3	2	1	2	1	0	9	3	0.333333

3	3	2	2	1	0	21	2	0.095238
3	4	2	2	1	0	13	6	0.461538
3	5	3	2	1	0	32	0	0
3	6	3	2	1	0	23	9	0.391304
3	7	4	2	1	0	16	3	0.1875
3	8	4	2	1	0	23	3	0.130435
40	1	1	2	1	0	19	2	0.105263
40	2	1	2	1	0	125	3	0.024
40	3	2	2	1	0	38	5	0.131579
40	4	2	2	1	0	41	11	0.268293
40	5	3	2	1	0	31	9	0.290323
40	6	3	2	1	0	6	1	0.166667
40	7	4	2	1	0	207	0	0
40	8	4	2	1	0	48	4	0.083333
41	1	1	4	1	1	50	6	0.12
41	2	1	4	1	1	43	8	0.186047
41	3	2	4	1	1	21	1	0.047619
41	4	2	4	1	1	18	4	0.222222
41	5	3	4	1	1	20	3	0.15
41	6	3	4	1	1	11	6	0.545455
41	7	4	4	1	1	14	9	0.642857
41	8	4	4	1	1	12	0	0
42	1	1	1	0	0	15	2	0.133333
42	2	1	1	0	0	14	4	0.285714
42	3	2	1	0	0	35	5	0.142857
42	4	2	1	0	0	45	10	0.222222
42	5	3	1	0	0	20	5	0.25
42	6	3	1	0	0	13	4	0.307692
42	7	4	1	0	0	10	0	0
42	8	4	1	0	0	18	1	0.055556
43	1	1	2	1	0	10	0	0
43	2	1	2	1	0	38	7	0.184211
43	3	2	2	1	0	27	3	0.111111
43	4	2	2	1	0	34	10	0.294118
43	5	3	2	1	0	51	1	0.019608
43	6	3	2	1	0	68	19	0.279412
43	7	4	2	1	0	31	2	0.064516
43	8	4	2	1	0	29	1	0.034483

44	1	1	3	0	1	19	3	0.157895
44	2	1	3	0	1	13	5	0.384615
44	3	2	3	0	1	25	3	0.12
44	4	2	3	0	1	23	5	0.217391
44	5	3	3	0	1	79	0	0
44	6	3	3	0	1	11	4	0.363636
44	7	4	3	0	1	9	5	0.555556
44	8	4	3	0	1	11	0	0
45	1	1	4	1	1	33	5	0.151515
45	2	1	4	1	1	33	5	0.151515
45	3	2	4	1	1	42	5	0.119048
45	4	2	4	1	1	22	6	0.272727
45	5	3	4	1	1	39	9	0.230769
45	6	3	4	1	1	13	3	0.230769
45	7	4	4	1	1	9	4	0.444444
45	8	4	4	1	1	14	2	0.142857
46	1	1	3	0	1	65	0	0
46	2	1	3	0	1	12	2	0.166667
46	3	2	3	0	1	28	3	0.107143
46	4	2	3	0	1	36	12	0.333333
46	5	3	3	0	1	15	4	0.266667
46	6	3	3	0	1	13	3	0.230769
46	7	4	3	0	1	7	6	0.857143
46	8	4	3	0	1	51	17	0.333333
47	1	1	2	1	0	33	2	0.060606
47	2	1	2	1	0	4	2	0.5
47	3	2	2	1	0	38	5	0.131579
47	4	2	2	1	0	38	10	0.263158
47	5	3	2	1	0	0	0	#NULL!
47	6	3	2	1	0	11	4	0.363636
47	7	4	2	1	0	17	0	0
47	8	4	2	1	0	4	2	0.5
48	1	1	1	0	0	20	2	0.1
48	2	1	1	0	0	10	5	0.5
48	3	2	1	0	0	31	5	0.16129
48	4	2	1	0	0	21	4	0.190476
48	5	3	1	0	0	13	0	0
48	6	3	1	0	0	36	4	0.111111

48	7	4	1	0	0	21	0	0
48	8	4	1	0	0	74	5	0.067568
4	1	1	1	0	0	14	0	0
4	2	1	1	0	0	12	4	0.333333
4	3	2	1	0	0	28	4	0.142857
4	4	2	1	0	0	38	12	0.315789
4	5	3	1	0	0	37	2	0.054054
								0.2857
4	6	3	1	0	0	7	2	14
4	7	4	1	0	0	57	6	0.105263
4	8	4	1	0	0	72	6	0.083333
5	1	1	4	1	1	21	1	0.047619
5	2	1	4	1	1	10	2	0.2
5	3	2	4	1	1	23	3	0.130435
5	4	2	4	1	1	13	3	0.230769
5	5	3	4	1	1	19	4	0.210526
5	6	3	4	1	1	11	4	0.363636
5	7	4	4	1	1	23	5	0.217391
5	8	4	4	1	1	30	9	0.3
6	1	1	3	0	1	22	1	0.045455
6	2	1	3	0	1	14	3	0.214286
6	3	2	3	0	1	25	3	0.12
6	4	2	3	0	1	11	6	0.545455
6	5	3	3	0	1	19	6	0.315789
6	6	3	3	0	1	11	3	0.272727
6	7	4	3	0	1	15	3	0.2
6	8	4	3	0	1	10	2	0.2
7	1	1	2	1	0	22	3	0.136364
7	2	1	2	1	0	14	6	0.428571
7	3	2	2	1	0	18	1	0.055556
7	4	2	2	1	0	9	5	0.555556
7	5	3	2	1	0	30	2	0.066667
7	6	3	2	1	0	4	4	1
7	7	4	2	1	0	24	5	0.208333
7	8	4	2	1	0	24	0	0
8	1	1	2	1	0	49	2	0.040816
8	2	1	2	1	0	15	4	0.266667
8	3	2	2	1	0	26	3	0.115385

8	4	2	2	1	0	42	11	0.261905
8	5	3	2	1	0	35	7	0.2
8	6	3	2	1	0	56	8	0.142857
8	7	4	2	1	0	60	7	0.116667
8	8	4	2	1	0	58	6	0.103448
9	1	1	3	0	1	38	2	0.052632
9	2	1	3	0	1	33	7	0.212121
9	3	2	3	0	1	80	6	0.075
9	4	2	3	0	1	39	11	0.282051
9	5	3	3	0	1	79	9	0.113924
9	6	3	3	0	1	13	4	0.307692
9	7	4	3	0	1	48	23	0.479167
9	8	4	3	0	1	68	18	0.264706

## Appendix R

### Time Raw Results

Participant	Question	Version	HasScore	HasFeature	Seconds
1	1	4	1	1	768
1	2	4	1	1	892
1	3	4	1	1	418
1	4	4	1	1	348
1	5	4	1	1	548
1	6	4	1	1	284
1	7	4	1	1	462
1	8	4	1	1	839
2	1	3	0	1	345
2	2	3	0	1	319
2	3	3	0	1	506
2	4	3	0	1	900
2	5	3	0	1	300
2	6	3	0	1	144
2	7	3	0	1	373
2	8	3	0	1	252
3	1	2	1	0	541
3	2	2	1	0	281
3	3	2	1	0	354
3	4	2	1	0	574
3	5	2	1	0	895
3	6	2	1	0	422
3	7	2	1	0	307
3	8	2	1	0	539
4	1	1	0	0	383
4	2	1	0	0	302
4	3	1	0	0	436
4	4	1	0	0	563
4	5	1	0	0	651
4	6	1	0	0	315
4	7	1	0	0	898

4	8	1	0	0	900
5	1	4	1	1	625
5	2	4	1	1	701
5	3	4	1	1	846
5	4	4	1	1	269
5	5	4	1	1	496
5	6	4	1	1	415
5	7	4	1	1	479
5	8	4	1	1	606
6	1	3	0	1	600
6	2	3	0	1	409
6	3	3	0	1	623
6	4	3	0	1	416
6	5	3	0	1	385
6	6	3	0	1	341
6	7	3	0	1	849
6	8	3	0	1	494
7	1	2	1	0	602
7	2	2	1	0	529
7	3	2	1	0	508
7	4	2	1	0	324
7	5	2	1	0	600
7	6	2	1	0	211
7	7	2	1	0	377
7	8	2	1	0	532
8	1	2	1	0	900
8	2	2	1	0	416
8	3	2	1	0	468
8	4	2	1	0	770
8	5	2	1	0	627
8	6	2	1	0	803
8	7	2	1	0	743
8	8	2	1	0	869
9	1	3	0	1	900
9	2	3	0	1	864
9	3	3	0	1	898
9	4	3	0	1	714
9	5	3	0	1	900

9	6	3	0	1	528
9	7	3	0	1	614
9	8	3	0	1	752
10	1	1	0	0	478
10	2	1	0	0	900
10	3	1	0	0	566
10	4	1	0	0	597
10	5	1	0	0	693
10	6	1	0	0	321
10	7	1	0	0	714
10	8	1	0	0	900
11	1	2	1	0	767
11	2	2	1	0	888
11	3	2	1	0	634
11	4	2	1	0	705
11	5	2	1	0	652
11	6	2	1	0	250
11	7	2	1	0	526
11	8	2	1	0	900
12	1	4	1	1	524
12	2	4	1	1	392
12	3	4	1	1	360
12	4	4	1	1	466
12	5	4	1	1	492
12	6	4	1	1	445
12	7	4	1	1	833
12	8	4	1	1	832
13	1	3	0	1	552
13	2	3	0	1	440
13	3	3	0	1	506
13	4	3	0	1	565
13	5	3	0	1	594
13	6	3	0	1	354
13	7	3	0	1	291
13	8	3	0	1	527
14	1	4	1	1	574
14	2	4	1	1	639
14	3	4	1	1	488



14	4	4	1	1	302
14	5	4	1	1	572
14	6	4	1	1	501
14	7	4	1	1	557
14	8	4	1	1	378
15	1	2	1	0	479
15	2	2	1	0	761
15	3	2	1	0	494
15	4	2	1	0	303
15	5	2	1	0	792
15	6	2	1	0	517
15	7	2	1	0	442
15	8	2	1	0	354
16	1	1	0	0	820
16	2	1	0	0	875
16	3	1	0	0	900
16	4	1	0	0	804
16	5	1	0	0	537
16	6	1	0	0	626
16	7	1	0	0	610
16	8	1	0	0	646
17	1	4	1	1	825
17	2	4	1	1	645
17	3	4	1	1	240
17	4	4	1	1	316
17	5	4	1	1	709
17	6	4	1	1	169
17	7	4	1	1	325
17	8	4	1	1	451
18	1	3	0	1	900
18	2	3	0	1	900
18	3	3	0	1	794
18	4	3	0	1	900
18	5	3	0	1	900
18	6	3	0	1	382
18	7	3	0	1	805
18	8	3	0	1	900
19	1	4	1	1	460

19	2	4	1	1	432
19	3	4	1	1	741
19	4	4	1	1	423
19	5	4	1	1	380
19	6	4	1	1	231
19	7	4	1	1	365
19	8	4	1	1	201
20	1	1	0	0	781
20	2	1	0	0	421
20	3	1	0	0	286
20	4	1	0	0	352
20	5	1	0	0	424
20	6	1	0	0	207
20	7	1	0	0	520
20	8	1	0	0	267
21	1	2	1	0	900
21	2	2	1	0	900
21	3	2	1	0	717
21	4	2	1	0	678
21	5	2	1	0	900
21	6	2	1	0	316
21	7	2	1	0	734
21	8	2	1	0	791
22	1	3	0	1	650
22	2	3	0	1	382
22	3	3	0	1	729
22	4	3	0	1	236
22	5	3	0	1	428
22	6	3	0	1	171
22	7	3	0	1	736
22	8	3	0	1	779
23	1	4	1	1	774
23	2	4	1	1	778
23	3	4	1	1	854
23	4	4	1	1	490
23	5	4	1	1	548
23	6	4	1	1	641
23	7	4	1	1	613

23	8	4	1	1	824
24	1	1	0	0	582
24	2	1	0	0	745
24	3	1	0	0	900
24	4	1	0	0	431
24	5	1	0	0	704
24	6	1	0	0	442
24	7	1	0	0	828
24	8	1	0	0	860
25	1	2	1	0	557
25	2	2	1	0	900
25	3	2	1	0	487
25	4	2	1	0	729
25	5	2	1	0	593
25	6	2	1	0	331
25	7	2	1	0	112
25	8	2	1	0	713
26	1	3	0	1	555
26	2	3	0	1	672
26	3	3	0	1	523
26	4	3	0	1	409
26	5	3	0	1	900
26	6	3	0	1	343
26	7	3	0	1	264
26	8	3	0	1	900
27	1	2	1	0	426
27	2	2	1	0	567
27	3	2	1	0	811
27	4	2	1	0	542
27	5	2	1	0	469
27	6	2	1	0	274
27	7	2	1	0	413
27	8	2	1	0	438
28	1	1	0	0	900
28	2	1	0	0	556
28	3	1	0	0	624
28	4	1	0	0	642
28	5	1	0	0	268

28	6	1	0	0	404
28	7	1	0	0	389
28	8	1	0	0	546
29	1	4	1	1	353
29	2	4	1	1	892
29	3	4	1	1	557
29	4	4	1	1	900
29	5	4	1	1	900
29	6	4	1	1	535
29	7	4	1	1	838
29	8	4	1	1	596
30	1	1	0	0	861
30	2	1	0	0	900
30	3	1	0	0	900
30	4	1	0	0	413
30	5	1	0	0	738
30	6	1	0	0	590
30	7	1	0	0	609
30	8	1	0	0	470
31	1	2	1	0	778
31	2	2	1	0	800
31	3	2	1	0	755
31	4	2	1	0	776
31	5	2	1	0	726
31	6	2	1	0	461
31	7	2	1	0	658
31	8	2	1	0	839
32	1	1	0	0	490
32	2	1	0	0	817
32	3	1	0	0	455
32	4	1	0	0	587
32	5	1	0	0	900
32	6	1	0	0	180
32	7	1	0	0	780
32	8	1	0	0	749
33	1	1	0	0	709
33	2	1	0	0	737
33	3	1	0	0	757

33	4	1	0	0	808
33	5	1	0	0	512
33	6	1	0	0	429
33	7	1	0	0	862
33	8	1	0	0	646
34	1	3	0	1	896
34	2	3	0	1	401
34	3	3	0	1	369
34	4	3	0	1	543
34	5	3	0	1	819
34	6	3	0	1	479
34	7	3	0	1	181
34	8	3	0	1	135
35	1	4	1	1	831
35	2	4	1	1	900
35	3	4	1	1	792
35	4	4	1	1	561
35	5	4	1	1	770
35	6	4	1	1	340
35	7	4	1	1	773
35	8	4	1	1	699
36	1	3	0	1	723
36	2	3	0	1	350
36	3	3	0	1	681
36	4	3	0	1	356
36	5	3	0	1	224
36	6	3	0	1	182
36	7	3	0	1	242
36	8	3	0	1	341
37	1	3	0	1	833
37	2	3	0	1	900
37	3	3	0	1	468
37	4	3	0	1	278
37	5	3	0	1	403
37	6	3	0	1	357
37	7	3	0	1	115
37	8	3	0	1	170
38	1	1	0	0	900

38	2	1	0	0	769
38	3	1	0	0	532
38	4	1	0	0	886
38	5	1	0	0	519
38	6	1	0	0	360
38	7	1	0	0	221
38	8	1	0	0	457
39	1	4	1	1	705
39	2	4	1	1	782
39	3	4	1	1	788
39	4	4	1	1	537
39	5	4	1	1	746
39	6	4	1	1	435
39	7	4	1	1	365
39	8	4	1	1	456
40	1	2	1	0	583
40	2	2	1	0	877
40	3	2	1	0	664
40	4	2	1	0	744
40	5	2	1	0	363
40	6	2	1	0	169
40	7	2	1	0	683
40	8	2	1	0	500
41	1	4	1	1	900
41	2	4	1	1	650
41	3	4	1	1	328
41	4	4	1	1	443
41	5	4	1	1	464
41	6	4	1	1	358
41	7	4	1	1	204
41	8	4	1	1	353
42	1	1	0	0	324
42	2	1	0	0	463
42	3	1	0	0	535
42	4	1	0	0	523
42	5	1	0	0	317
42	6	1	0	0	223
42	7	1	0	0	369

42	8	1	0	0	306
43	1	2	1	0	277
43	2	2	1	0	900
43	3	2	1	0	252
43	4	2	1	0	347
43	5	2	1	0	578
43	6	2	1	0	351
43	7	2	1	0	485
43	8	2	1	0	251
44	1	3	0	1	544
44	2	3	0	1	271
44	3	3	0	1	750
44	4	3	0	1	600
44	5	3	0	1	787
44	6	3	0	1	236
44	7	3	0	1	183
44	8	3	0	1	189
45	1	4	1	1	652
45	2	4	1	1	238
45	3	4	1	1	789
45	4	4	1	1	389
45	5	4	1	1	311
45	6	4	1	1	431
45	7	4	1	1	226
45	8	4	1	1	144
46	1	3	0	1	818
46	2	3	0	1	335
46	3	3	0	1	341
46	4	3	0	1	370
46	5	3	0	1	363
46	6	3	0	1	216
46	7	3	0	1	195
46	8	3	0	1	608
47	1	2	1	0	334
47	2	2	1	0	122
47	3	2	1	0	459
47	4	2	1	0	201
47	5	2	1	0	66

47	6	2	1	0	140
47	7	2	1	0	186
47	8	2	1	0	824
48	1	1	0	0	430
48	2	1	0	0	291
48	3	1	0	0	438
48	4	1	0	0	246
48	5	1	0	0	256
48	6	1	0	0	279
48	7	1	0	0	279
48	8	1	0	0	300



## Appendix S

### Precision / Recall Raw Results

Participant	TaskCategory	Task	Version	HasScore	HasFeature	Precision	Recall
1	1	1	4	1	1	1	0.666667
1	1	2	4	1	1	1	0.25
1	2	3	4	1	1	1	0.8
1	2	4	4	1	1	0.6	0.5
1	3	5	4	1	1	0	0
1	3	6	4	1	1	1	1
1	4	7	4	1	1	0.625	0.384615
1	4	8	4	1	1	0.571429	0.666667
2	1	1	3	0	1	0.333333	0.333333
2	1	2	3	0	1	0.4	0.5
2	2	3	3	0	1	0.6	0.6
2	2	4	3	0	1	0	0
2	3	5	3	0	1	0	0
2	3	6	3	0	1	0.75	0.75
2	4	7	3	0	1	0.666667	0.307692
2	4	8	3	0	1	0.666667	0.333333
3	1	1	2	1	0	0.666667	0.666667
3	1	2	2	1	0	0.333333	0.25
3	2	3	2	1	0	0.666667	0.4
3	2	4	2	1	0	0.5	0.25
3	3	5	2	1	0	0	0
3	3	6	2	1	0	0.6	0.75
3	4	7	2	1	0	0	0
3	4	8	2	1	0	0.25	0.166667
4	1	1	1	0	0	0	0
4	1	2	1	0	0	0.333333	0.25
4	2	3	1	0	0	0.5	0.2
4	2	4	1	0	0	0.7	0.583333
4	3	5	1	0	0	0.125	0.033333
4	3	6	1	0	0	1	0.25

4	4	7	1	0	0	0.142857	0.230769
4	4	8	1	0	0	0.25	0.666667
5	1	1	4	1	1	0.25	0.333333
5	1	2	4	1	1	0.666667	0.5
5	2	3	4	1	1	0	0
5	2	4	4	1	1	1	0.25
5	3	5	4	1	1	0	0
5	3	6	4	1	1	0.6	0.75
5	4	7	4	1	1	0.333333	0.076923
5	4	8	4	1	1	0.5	0.5
6	1	1	3	0	1	0	0
6	1	2	3	0	1	0.666667	0.5
6	2	3	3	0	1	0.5	0.2
6	2	4	3	0	1	0.571429	0.333333
6	3	5	3	0	1	1	0.066667
6	3	6	3	0	1	0.75	0.75
6	4	7	3	0	1	0.666667	0.153846
6	4	8	3	0	1	0	0
7	1	1	2	1	0	0.666667	0.666667
7	1	2	2	1	0	0.75	0.75
7	2	3	2	1	0	0.666667	0.4
7	2	4	2	1	0	0.5	0.166667
7	3	5	2	1	0	0.142857	0.033333
7	3	6	2	1	0	1	1
7	4	7	2	1	0	0.333333	0.153846
7	4	8	2	1	0	0	0
8	1	1	2	1	0	0.25	0.666667
8	1	2	2	1	0	0.5	0.75
8	2	3	2	1	0	0.428571	0.6
8	2	4	2	1	0	0.357143	0.416667
8	3	5	2	1	0	0.4	0.133333
8	3	6	2	1	0	0.75	0.75
8	4	7	2	1	0	0.555556	0.384615
8	4	8	2	1	0	0	0
9	1	1	3	0	1	0.5	0.666667
9	1	2	3	0	1	0.333333	0.25
9	2	3	3	0	1	0.428571	0.6
9	2	4	3	0	1	0.25	0.166667

9	3	5	3	0	1	0.125	0.033333
9	3	6	3	0	1	0.6	0.75
9	4	7	3	0	1	0.5	0.230769
9	4	8	3	0	1	0.5	0.333333
10	1	1	1	0	0	1	0.666667
10	1	2	1	0	0	0.333333	0.25
10	2	3	1	0	0	0.5	0.4
10	2	4	1	0	0	0.25	0.166667
10	3	5	1	0	0	0.166667	0.066667
10	3	6	1	0	0	1	0.75
10	4	7	1	0	0	0.333333	0.076923
10	4	8	1	0	0	0.25	0.166667
11	1	1	2	1	0	0.5	0.666667
11	1	2	2	1	0	0.5	0.25
11	2	3	2	1	0	1	0.2
11	2	4	2	1	0	0.4	0.333333
11	3	5	2	1	0	0.4	0.133333
11	3	6	2	1	0	0.666667	1
11	4	7	2	1	0	0.333333	0.230769
11	4	8	2	1	0	0.6	0.5
12	1	1	4	1	1	0	0
12	1	2	4	1	1	0	0
12	2	3	4	1	1	0	0
12	2	4	4	1	1	0	0
12	3	5	4	1	1	0	0
12	3	6	4	1	1	0	0
12	4	7	4	1	1	0	0
12	4	8	4	1	1	0	0
13	1	1	3	0	1	0.333333	1
13	1	2	3	0	1	0.25	0.25
13	2	3	3	0	1	0.4	0.8
13	2	4	3	0	1	0.222222	0.166667
13	3	5	3	0	1	0	0
13	3	6	3	0	1	1	0.75
13	4	7	3	0	1	0.625	0.384615
13	4	8	3	0	1	0.5	0.5
14	1	1	4	1	1	1	0.666667
14	1	2	4	1	1	0.5	0.5

14	2	3	4	1	1	0.5	0.4
14	2	4	4	1	1	0.428571	0.25
14	3	5	4	1	1	0.111111	0.033333
14	3	6	4	1	1	1	0.75
14	4	7	4	1	1	0.333333	0.076923
14	4	8	4	1	1	0.5	0.5
15	1	1	2	1	0	0	0
15	1	2	2	1	0	0.5	0.75
15	2	3	2	1	0	1	0.2
15	2	4	2	1	0	0.5	0.416667
15	3	5	2	1	0	0	0
15	3	6	2	1	0	1	0.75
15	4	7	2	1	0	0	0
15	4	8	2	1	0	0	0
16	1	1	1	0	0	0	0
16	1	2	1	0	0	0	0
16	2	3	1	0	0	0	0
16	2	4	1	0	0	0.5	0.166667
16	3	5	1	0	0	0.285714	0.066667
16	3	6	1	0	0	0.6	0.75
16	4	7	1	0	0	0	0
16	4	8	1	0	0	0	0
17	1	1	4	1	1	0	0
17	1	2	4	1	1	0.333333	0.25
17	2	3	4	1	1	0.333333	0.2
17	2	4	4	1	1	1	0.25
17	3	5	4	1	1	0.333333	0.033333
17	3	6	4	1	1	0.666667	0.5
17	4	7	4	1	1	0	0
17	4	8	4	1	1	1	0.333333
18	1	1	3	0	1	0.666667	0.666667
18	1	2	3	0	1	0.333333	0.25
18	2	3	3	0	1	0.666667	0.4
18	2	4	3	0	1	0.444444	0.333333
18	3	5	3	0	1	0	0
18	3	6	3	0	1	1	1
18	4	7	3	0	1	0	0
18	4	8	3	0	1	0.176471	0.5

19	1	1	4	1	1	0.5	0.666667
19	1	2	4	1	1	1	0.25
19	2	3	4	1	1	0.5	0.8
19	2	4	4	1	1	0.4	0.333333
19	3	5	4	1	1	0.333333	0.066667
19	3	6	4	1	1	1	1
19	4	7	4	1	1	0.333333	0.076923
19	4	8	4	1	1	0.5	0.166667
20	1	1	1	0	0	0.666667	0.666667
20	1	2	1	0	0	0.5	0.25
20	2	3	1	0	0	1	0.2
20	2	4	1	0	0	0.5	0.166667
20	3	5	1	0	0	0.333333	0.033333
20	3	6	1	0	0	1	0.25
20	4	7	1	0	0	0	0
20	4	8	1	0	0	0.333333	0.333333
21	1	1	2	1	0	1	0.666667
21	1	2	2	1	0	0.5	0.75
21	2	3	2	1	0	0	0
21	2	4	2	1	0	0.454545	0.416667
21	3	5	2	1	0	0.5	0.2
21	3	6	2	1	0	0.5	0.75
21	4	7	2	1	0	0.75	0.230769
21	4	8	2	1	0	0	0
22	1	1	3	0	1	0.333333	0.333333
22	1	2	3	0	1	0.333333	0.25
22	2	3	3	0	1	0	0
22	2	4	3	0	1	0.333333	0.083333
22	3	5	3	0	1	0.5	0.066667
22	3	6	3	0	1	1	0.25
22	4	7	3	0	1	0	0
22	4	8	3	0	1	0	0
23	1	1	4	1	1	0.4	0.666667
23	1	2	4	1	1	0.5	0.25
23	2	3	4	1	1	0.25	0.2
23	2	4	4	1	1	0.714286	0.416667
23	3	5	4	1	1	0	0
23	3	6	4	1	1	0.75	0.75

23	4	7	4	1	1	0.714286	0.384615
23	4	8	4	1	1	0	0
24	1	1	1	0	0	1	0.333333
24	1	2	1	0	0	1	0.25
24	2	3	1	0	0	0.25	0.2
24	2	4	1	0	0	1	0.333333
24	3	5	1	0	0	0	0
24	3	6	1	0	0	1	0.75
24	4	7	1	0	0	0.166667	0.076923
24	4	8	1	0	0	0	0
25	1	1	2	1	0	0.666667	0.666667
25	1	2	2	1	0	0	0
25	2	3	2	1	0	0.666667	0.4
25	2	4	2	1	0	0.142857	0.083333
25	3	5	2	1	0	0.111111	0.033333
25	3	6	2	1	0	1	0.25
25	4	7	2	1	0	0	0
25	4	8	2	1	0	0.5	0.166667
26	1	1	3	0	1	0.666667	0.666667
26	1	2	3	0	1	0.5	0.5
26	2	3	3	0	1	0.333333	0.2
26	2	4	3	0	1	0.5	0.166667
26	3	5	3	0	1	0.5	0.033333
26	3	6	3	0	1	1	0.75
26	4	7	3	0	1	0.833333	0.384615
26	4	8	3	0	1	0	0
27	1	1	2	1	0	1	0.666667
27	1	2	2	1	0	0	0
27	2	3	2	1	0	0.333333	0.2
27	2	4	2	1	0	0.555556	0.416667
27	3	5	2	1	0	0	0
27	3	6	2	1	0	1	1
27	4	7	2	1	0	0.25	0.076923
27	4	8	2	1	0	0.6	0.5
28	1	1	1	0	0	0.5	0.333333
28	1	2	1	0	0	0.6	0.75
28	2	3	1	0	0	0.5	0.2
28	2	4	1	0	0	0.6	0.25

28	3	5	1	0	0	0	0
28	3	6	1	0	0	1	1
28	4	7	1	0	0	0.333333	0.076923
28	4	8	1	0	0	1	0.333333
29	1	1	4	1	1	0.666667	0.666667
29	1	2	4	1	1	0.333333	0.25
29	2	3	4	1	1	1	0.2
29	2	4	4	1	1	0.461538	0.5
29	3	5	4	1	1	0	0
29	3	6	4	1	1	1	0.75
29	4	7	4	1	1	0.875	0.538462
29	4	8	4	1	1	0.428571	0.5
30	1	1	1	0	0	0.333333	0.333333
30	1	2	1	0	0	0.5	0.25
30	2	3	1	0	0	1	0.4
30	2	4	1	0	0	0.666667	0.166667
30	3	5	1	0	0	0.125	0.033333
30	3	6	1	0	0	0.8	1
30	4	7	1	0	0	0.333333	0.153846
30	4	8	1	0	0	0	0
31	1	1	2	1	0	0.333333	0.333333
31	1	2	2	1	0	1	0.5
31	2	3	2	1	0	1	0.4
31	2	4	2	1	0	0.6	0.25
31	3	5	2	1	0	0	0
31	3	6	2	1	0	1	1
31	4	7	2	1	0	0.333333	0.076923
31	4	8	2	1	0	0	0
32	1	1	1	0	0	0.666667	0.666667
32	1	2	1	0	0	0.6	0.75
32	2	3	1	0	0	0.6	0.6
32	2	4	1	0	0	0.1	0.083333
32	3	5	1	0	0	0.111111	0.033333
32	3	6	1	0	0	1	0.75
32	4	7	1	0	0	0	0
32	4	8	1	0	0	0	0
33	1	1	1	0	0	0.4	0.666667
33	1	2	1	0	0	0.5	0.5

33	2	3	1	0	0	0.5	0.8
33	2	4	1	0	0	0.444444	0.666667
33	3	5	1	0	0	0.375	0.1
33	3	6	1	0	0	0.666667	1
33	4	7	1	0	0	0.192308	0.384615
33	4	8	1	0	0	0	0
34	1	1	3	0	1	0.666667	0.666667
34	1	2	3	0	1	1	0.25
34	2	3	3	0	1	0.333333	0.2
34	2	4	3	0	1	0.454545	0.416667
34	3	5	3	0	1	0	0
34	3	6	3	0	1	1	0.5
34	4	7	3	0	1	0	0
34	4	8	3	0	1	0	0
35	1	1	4	1	1	0.5	0.333333
35	1	2	4	1	1	0.571429	1
35	2	3	4	1	1	1	0.4
35	2	4	4	1	1	0.3	0.25
35	3	5	4	1	1	0.1	0.033333
35	3	6	4	1	1	1	1
35	4	7	4	1	1	0	0
35	4	8	4	1	1	0.333333	0.333333
36	1	1	3	0	1	1	0.333333
36	1	2	3	0	1	0.666667	0.5
36	2	3	3	0	1	0.5	0.2
36	2	4	3	0	1	0.833333	0.416667
36	3	5	3	0	1	0.8	0.133333
36	3	6	3	0	1	1	0.75
36	4	7	3	0	1	1	0.384615
36	4	8	3	0	1	1	0.333333
37	1	1	3	0	1	1	0.666667
37	1	2	3	0	1	0.2	0.25
37	2	3	3	0	1	0.666667	0.4
37	2	4	3	0	1	0.416667	0.416667
37	3	5	3	0	1	0.142857	0.033333
37	3	6	3	0	1	0.666667	1
37	4	7	3	0	1	0.333333	0.076923
37	4	8	3	0	1	0	0



38	1	1	1	0	0	0	0
38	1	2	1	0	0	0.4	0.5
38	2	3	1	0	0	0.5	0.6
38	2	4	1	0	0	0.222222	0.166667
38	3	5	1	0	0	0.4	0.133333
38	3	6	1	0	0	1	1
38	4	7	1	0	0	0	0
38	4	8	1	0	0	0.333333	0.166667
39	1	1	4	1	1	0.666667	0.666667
39	1	2	4	1	1	0.4	0.5
39	2	3	4	1	1	0.444444	0.8
39	2	4	4	1	1	0.3	0.25
39	3	5	4	1	1	0.2	0.066667
39	3	6	4	1	1	0.8	1
39	4	7	4	1	1	0.333333	0.076923
39	4	8	4	1	1	0	0
40	1	1	2	1	0	0.666667	0.666667
40	1	2	2	1	0	0.333333	0.25
40	2	3	2	1	0	0.1875	0.6
40	2	4	2	1	0	0.347826	0.666667
40	3	5	2	1	0	0.3	0.1
40	3	6	2	1	0	1	0.25
40	4	7	2	1	0	0	0
40	4	8	2	1	0	0.333333	0.333333
41	1	1	4	1	1	1	0.666667
41	1	2	4	1	1	0.666667	0.5
41	2	3	4	1	1	0.333333	0.2
41	2	4	4	1	1	0.363636	0.333333
41	3	5	4	1	1	0.222222	0.066667
41	3	6	4	1	1	0.666667	0.5
41	4	7	4	1	1	0.8	0.307692
41	4	8	4	1	1	0	0
42	1	1	1	0	0	0.5	0.666667
42	1	2	1	0	0	0.4	0.5
42	2	3	1	0	0	0.5	0.6
42	2	4	1	0	0	0.3	0.25
42	3	5	1	0	0	0.5	0.166667
42	3	6	1	0	0	0.8	1

42	4	7	1	0	0	0	0
42	4	8	1	0	0	0.25	0.166667
43	1	1	2	1	0	0	0
43	1	2	2	1	0	1	0.25
43	2	3	2	1	0	0.75	0.6
43	2	4	2	1	0	0.5	0.166667
43	3	5	2	1	0	0	0
43	3	6	2	1	0	0.5	0.25
43	4	7	2	1	0	0	0
43	4	8	2	1	0	0	0
44	1	1	3	0	1	0.6	1
44	1	2	3	0	1	0.5	0.25
44	2	3	3	0	1	0	0
44	2	4	3	0	1	0.625	0.416667
44	3	5	3	0	1	0	0
44	3	6	3	0	1	1	1
44	4	7	3	0	1	0.5	0.307692
44	4	8	3	0	1	0	0
45	1	1	4	1	1	0.25	0.333333
45	1	2	4	1	1	0.5	0.25
45	2	3	4	1	1	0.5	0.2
45	2	4	4	1	1	0.428571	0.25
45	3	5	4	1	1	0.625	0.166667
45	3	6	4	1	1	0.666667	0.5
45	4	7	4	1	1	0.75	0.230769
45	4	8	4	1	1	0	0
46	1	1	3	0	1	0	0
46	1	2	3	0	1	0.2	0.25
46	2	3	3	0	1	0.333333	0.2
46	2	4	3	0	1	0.222222	0.166667
46	3	5	3	0	1	0.3	0.1
46	3	6	3	0	1	1	0.75
46	4	7	3	0	1	1	0.307692
46	4	8	3	0	1	0.833333	0.833333
47	1	1	2	1	0	1	0.666667
47	1	2	2	1	0	0.333333	0.25
47	2	3	2	1	0	0.666667	0.8
47	2	4	2	1	0	0.9	0.75

47	3	5	2	1	0	0	0
47	3	6	2	1	0	1	1
47	4	7	2	1	0	0	0
47	4	8	2	1	0	0.666667	0.333333
48	1	1	1	0	0	0.666667	0.666667
48	1	2	1	0	0	0.4	0.5
48	2	3	1	0	0	1	0.6
48	2	4	1	0	0	0.4	0.333333
48	3	5	1	0	0	0	0
48	3	6	1	0	0	0.8	1
48	4	7	1	0	0	0	0
48	4	8	1	0	0	0.5	0.166667

## Appendix T

### Function Raw Results

Parti cipan t	Ver sion	Has Score	HasF eature	desele ct_che ckbox	chang e_y_a xis	incl ude_ node	change _x_axi s	restor _e_node	select _check box	click _block	click_ node	excl ude_ node	fil ter	
10	1	1	0	0	4	0	2	0	0	0	0	30	8	3
10	2	1	0	0	6	0	3	0	0	2	0	30	7	3
10	3	1	0	0	0	0	4	0	0	0	0	41	26	1
10	4	1	0	0	4	0	8	0	0	0	0	47	4	2
10	5	1	0	0	5	0	12	0	0	1	1	54	19	0
10	6	1	0	0	0	0	3	0	0	0	0	11	8	1
10	7	1	0	0	4	0	2	0	0	0	0	39	10	1
10	8	1	0	0	8	0	4	0	0	2	0	55	48	0
16	1	1	0	0	0	0	2	0	0	0	0	35	33	3
16	2	1	0	0	4	0	3	0	0	4	0	41	35	9
16	3	1	0	0	1	0	0	0	0	0	0	47	46	2
16	4	1	0	0	4	0	4	0	0	0	0	14	10	1
16	5	1	0	0	5	0	7	0	0	1	0	13	5	0
16	6	1	0	0	4	0	5	0	0	0	0	29	24	1
16	7	1	0	0	4	0	2	0	0	0	0	58	29	2
16	8	1	0	0	4	0	4	0	0	0	0	140	1	8
20	1	1	0	0	8	0	3	0	1	1	1	42	1	6
20	2	1	0	0	4	0	2	0	0	0	1	15	0	6
20	3	1	0	0	3	0	1	0	0	2	1	42	0	2
20	4	1	0	0	8	0	4	0	0	2	1	31	0	1
20	5	1	0	0	4	0	3	0	0	0	1	20	0	2
20	6	1	0	0	4	0	1	0	0	0	1	21	0	1
20	7	1	0	0	4	0	2	0	0	0	1	135	0	6
20	8	1	0	0	20	0	10	0	1	4	2	70	0	4
24	1	1	0	0	9	0	1	0	0	10	1	78	0	18
24	2	1	0	0	0	0	1	0	0	1	1	68	2	12
24	3	1	0	0	3	0	4	0	0	2	1	134	0	8
24	4	1	0	0	4	0	4	0	0	1	1	64	0	10
24	5	1	0	0	7	0	1	0	1	6	1	47	0	19

24	6	1	0	0	1	0	4	0	1	3	1	54	0	6
24	7	1	0	0	4	0	6	0	0	0	1	71	0	3
24	8	1	0	0	13	0	0	0	0	9	1	223	0	2
28	1	1	0	0	4	0	2	0	0	3	0	41	0	6
28	2	1	0	0	0	0	5	0	0	0	0	9	0	6
28	3	1	0	0	4	0	4	0	2	4	0	12	0	8
28	4	1	0	0	6	0	5	0	0	2	0	14	0	1
28	5	1	0	0	6	0	1	0	0	4	0	2	0	1
28	6	1	0	0	0	0	4	0	0	0	0	12	0	3
28	7	1	0	0	0	0	3	0	0	0	0	12	9	1
28	8	1	0	0	9	0	4	0	1	2	0	13	5	4
30	1	1	0	0	8	0	3	0	0	4	0	30	22	9
30	2	1	0	0	4	0	2	0	0	1	1	26	21	7
30	3	1	0	0	0	0	2	0	0	0	0	31	28	6
30	4	1	0	0	9	0	3	0	0	1	0	5	2	13
30	5	1	0	0	4	0	8	0	0	2	0	17	8	1
30	6	1	0	0	5	0	6	0	1	1	0	11	5	12
30	7	1	0	0	5	0	6	0	0	1	0	16	10	1
30	8	1	0	0	6	0	15	0	0	2	0	18	1	0
32	1	1	0	0	4	0	3	0	0	0	1	28	0	13
32	2	1	0	0	0	0	5	0	0	0	1	38	0	8
32	3	1	0	0	0	0	5	0	0	0	1	25	0	1
32	4	1	0	0	4	0	10	0	0	0	1	41	0	1
32	5	1	0	0	4	0	9	0	0	0	1	83	0	0
32	6	1	0	0	0	0	3	0	0	0	1	20	0	1
32	7	1	0	0	4	0	3	0	0	0	1	150	0	0
32	8	1	0	0	7	0	2	0	0	0	1	243	0	0
33	1	1	0	0	4	0	6	0	1	0	0	44	13	12
33	2	1	0	0	0	0	9	0	6	0	0	27	11	6
33	3	1	0	0	0	0	4	0	0	0	0	42	33	1
33	4	1	0	0	6	0	20	0	0	2	0	41	18	1
33	5	1	0	0	6	0	8	0	0	2	0	26	14	1
33	6	1	0	0	0	0	6	0	0	0	0	11	5	1
33	7	1	0	0	4	0	26	0	0	0	0	58	31	1
33	8	1	0	0	13	0	0	0	0	5	0	204	1	3
38	1	1	0	0	0	0	17	0	2	0	0	49	12	1
38	2	1	0	0	0	0	8	0	3	0	0	23	0	2
38	3	1	0	0	0	0	6	0	0	0	0	36	0	8

38	4	1	0	0	4	0	17	0	8	0	0	46	5	1
38	5	1	0	0	13	0	12	0	2	5	0	39	0	0
38	6	1	0	0	0	0	4	0	0	0	0	20	0	5
38	7	1	0	0	0	0	2	0	0	0	0	67	0	1
38	8	1	0	0	6	0	3	0	0	2	0	158	0	1
42	1	1	0	0	7	0	4	0	0	0	0	15	0	3
42	2	1	0	0	4	0	5	0	0	0	0	14	5	3
42	3	1	0	0	3	0	6	0	0	1	0	35	6	1
42	4	1	0	0	5	0	7	0	2	0	0	45	5	1
42	5	1	0	0	5	0	11	0	1	0	0	20	3	2
42	6	1	0	0	4	0	4	0	0	1	0	13	7	2
42	7	1	0	0	5	0	4	0	0	1	0	10	0	8
42	8	1	0	0	7	0	4	0	0	0	0	18	0	0
48	1	1	0	0	0	0	7	0	4	0	0	20	0	7
48	2	1	0	0	0	0	7	0	2	0	0	10	0	1
48	3	1	0	0	0	0	3	0	0	0	0	31	0	6
48	4	1	0	0	4	0	10	0	0	0	0	21	0	1
48	5	1	0	0	13	0	9	0	1	8	0	13	0	4
48	6	1	0	0	0	0	5	0	0	0	0	36	0	8
48	7	1	0	0	4	0	3	0	0	0	0	21	0	3
48	8	1	0	0	11	0	2	0	0	4	0	74	0	2
4	1	1	0	0	7	0	0	0	0	0	1	14	10	0
4	2	1	0	0	0	0	3	0	0	0	0	12	8	2
4	3	1	0	0	1	0	2	0	0	0	0	28	26	1
4	4	1	0	0	5	0	10	0	0	0	1	38	27	1
4	5	1	0	0	7	0	8	0	0	3	0	37	27	2
4	6	1	0	0	2	0	1	0	0	2	0	7	6	6
4	7	1	0	0	4	0	21	0	0	1	0	57	36	1
4	8	1	0	0	7	0	16	0	0	0	0	72	56	0
11	1	2	1	0	4	0	4	0	0	0	0	36	11	12
11	2	2	1	0	3	0	2	0	0	3	0	54	5	9
11	3	2	1	0	0	0	1	0	0	0	0	60	1	6
11	4	2	1	0	4	0	14	0	4	0	0	45	5	1
11	5	2	1	0	4	0	10	0	0	3	0	49	0	0
11	6	2	1	0	0	0	6	0	0	0	0	15	0	1
11	7	2	1	0	0	0	9	0	0	0	0	37	0	1
11	8	2	1	0	7	0	5	0	0	2	0	45	0	0
15	1	2	1	0	4	0	8	0	7	0	0	9	1	1

15	2	2	1	0	0	0	6	0	0	0	0	17	5	6
15	3	2	1	0	0	0	1	0	0	0	0	14	1	11
15	4	2	1	0	4	0	10	0	0	0	0	14	0	1
15	5	2	1	0	10	0	1	0	0	10	0	24	0	3
15	6	2	1	0	0	0	3	0	0	0	0	26	0	2
15	7	2	1	0	0	0	3	0	0	0	0	9	0	2
15	8	2	1	0	3	0	3	0	0	0	0	7	0	3
21	1	2	1	0	9	0	2	0	0	3	0	27	0	14
21	2	2	1	0	8	0	6	0	0	10	0	56	0	7
21	3	2	1	0	4	0	3	0	0	0	0	40	0	1
21	4	2	1	0	8	0	11	0	0	5	0	21	0	1
21	5	2	1	0	21	0	13	0	0	29	0	88	0	5
21	6	2	1	0	16	0	6	0	0	9	0	16	0	1
21	7	2	1	0	13	0	5	0	0	6	0	116	1	4
21	8	2	1	0	19	0	2	0	0	13	0	138	0	3
25	1	2	1	0	4	0	5	0	2	0	1	25	0	7
25	2	2	1	0	0	0	2	0	0	2	0	73	10	9
25	3	2	1	0	0	0	3	0	0	0	1	35	4	4
25	4	2	1	0	4	0	7	0	0	0	1	52	2	1
25	5	2	1	0	4	0	9	0	0	0	1	32	0	0
25	6	2	1	0	5	0	1	0	0	5	1	25	0	4
25	7	2	1	0	4	0	0	0	0	4	0	12	0	1
25	8	2	1	0	17	0	6	0	2	2	1	102	37	0
27	1	2	1	0	4	0	2	0	0	0	0	17	10	7
27	2	2	1	0	0	0	1	0	0	0	0	14	13	3
27	3	2	1	0	0	0	3	0	0	0	0	39	34	3
27	4	2	1	0	4	0	9	0	0	0	0	17	8	1
27	5	2	1	0	7	0	7	0	0	4	0	19	3	1
27	6	2	1	0	0	0	4	0	0	0	0	14	10	2
27	7	2	1	0	0	0	4	0	0	0	0	55	0	1
27	8	2	1	0	18	0	7	0	0	4	0	95	11	3
31	1	2	1	0	0	0	4	0	1	0	1	13	0	16
31	2	2	1	0	0	0	2	0	0	0	1	28	0	14
31	3	2	1	0	0	0	2	0	0	0	1	15	0	8
31	4	2	1	0	4	0	5	0	0	0	1	17	0	7
31	5	2	1	0	0	0	1	0	0	0	1	13	0	3
31	6	2	1	0	0	0	4	0	0	0	1	12	0	7
31	7	2	1	0	0	0	4	0	1	0	1	9	0	14

31	8	2	1	0	3	0	2	0	0	0	1	22	0	15
3	1	2	1	0	8	0	3	0	0	0	0	15	0	1
3	2	2	1	0	3	0	3	0	0	3	0	9	0	1
3	3	2	1	0	2	0	3	0	0	2	0	21	0	1
3	4	2	1	0	6	0	6	0	0	0	0	13	0	1
3	5	2	1	0	4	0	9	0	0	3	0	32	0	2
3	6	2	1	0	3	0	5	0	0	0	0	23	0	7
3	7	2	1	0	4	0	2	0	0	1	0	16	0	1
3	8	2	1	0	8	0	4	0	0	1	0	23	0	0
40	1	2	1	0	4	0	6	0	3	0	1	19	11	1
40	2	2	1	0	16	0	3	0	0	17	1	125	3	23
40	3	2	1	0	0	0	16	0	0	0	1	38	21	1
40	4	2	1	0	5	0	27	0	4	1	1	41	12	3
40	5	2	1	0	0	0	10	0	0	0	1	31	16	1
40	6	2	1	0	8	0	2	0	1	3	1	6	3	7
40	7	2	1	0	0	0	9	0	6	0	1	207	179	5
40	8	2	1	0	3	0	9	0	3	0	1	48	28	2
43	1	2	1	0	7	0	4	0	0	1	0	10	0	2
43	2	2	1	0	4	0	1	0	0	3	0	38	0	6
43	3	2	1	0	0	0	4	0	0	0	0	27	0	1
43	4	2	1	0	7	0	4	0	0	2	2	34	0	1
43	5	2	1	0	8	0	2	0	0	5	0	51	0	3
43	6	2	1	0	0	0	2	0	0	0	0	68	0	4
43	7	2	1	0	4	0	2	0	0	1	0	31	0	1
43	8	2	1	0	4	0	3	0	0	0	0	29	0	0
47	1	2	1	0	0	0	2	0	0	0	1	33	31	7
47	2	2	1	0	0	0	3	0	0	0	0	4	0	4
47	3	2	1	0	0	0	7	0	1	0	1	38	31	1
47	4	2	1	0	0	0	10	0	0	0	1	38	28	1
47	5	2	1	0	13	0	0	0	0	5	0	0	0	0
47	6	2	1	0	0	0	4	0	0	0	1	11	7	2
47	7	2	1	0	4	0	3	0	0	0	1	17	10	0
47	8	2	1	0	8	0	3	0	0	1	1	4	0	0
7	1	2	1	0	8	0	3	0	0	1	0	22	1	18
7	2	2	1	0	3	0	4	0	0	0	0	14	1	1
7	3	2	1	0	3	0	3	0	0	3	0	18	6	0
7	4	2	1	0	6	0	4	0	0	1	0	9	4	1
7	5	2	1	0	7	0	7	0	0	3	0	30	19	2



7	6	2	1	0	0	0	3	0	0	0	0	4	0	1
7	7	2	1	0	4	0	6	0	0	2	0	24	4	1
7	8	2	1	0	7	0	4	0	0	5	0	24	16	6
8	1	2	1	0	0	0	8	0	0	0	1	49	5	2
8	2	2	1	0	4	0	6	0	0	0	1	15	2	3
8	3	2	1	0	5	0	7	0	0	1	1	26	12	2
8	4	2	1	0	4	0	14	0	0	1	1	42	23	1
8	5	2	1	0	11	0	10	0	0	3	1	35	20	1
8	6	2	1	0	26	0	4	0	0	26	1	56	39	20
8	7	2	1	0	4	0	21	0	12	0	1	60	8	1
8	8	2	1	0	11	0	2	0	0	3	1	58	21	4
13	1	3	0	1	4	0	9	0	0	3	1	41	26	8
13	2	3	0	1	0	0	5	0	1	0	0	22	7	3
13	3	3	0	1	0	0	10	0	0	1	0	38	27	2
13	4	3	0	1	4	0	19	0	10	1	1	51	8	1
13	5	3	0	1	42	0	6	0	0	35	0	89	0	4
13	6	3	0	1	0	0	3	0	0	0	1	28	25	4
13	7	3	0	1	0	0	8	0	0	0	0	75	0	9
13	8	3	0	1	10	0	11	0	5	6	0	87	7	2
18	1	3	0	1	4	0	3	0	0	0	1	30	0	11
18	2	3	0	1	0	0	14	0	11	0	0	46	2	8
18	3	3	0	1	8	0	3	0	0	5	0	49	5	13
18	4	3	0	1	4	0	9	0	0	0	1	36	10	1
18	5	3	0	1	2	0	7	0	0	0	0	31	12	3
18	6	3	0	1	0	0	6	0	0	0	0	49	32	4
18	7	3	0	1	2	0	4	0	0	0	0	45	19	9
18	8	3	0	1	17	0	31	0	6	6	0	149	17	3
22	1	3	0	1	8	0	6	0	0	0	6	11	0	3
22	2	3	0	1	0	0	3	0	0	0	1	17	3	4
22	3	3	0	1	3	2	3	0	0	2	1	34	3	7
22	4	3	0	1	5	0	3	0	0	1	1	16	9	1
22	5	3	0	1	0	0	5	0	0	0	1	12	3	3
22	6	3	0	1	0	0	1	0	0	0	1	12	9	1
22	7	3	0	1	0	2	4	0	0	0	1	22	0	0
22	8	3	0	1	5	0	5	0	0	1	1	8	0	4
26	1	3	0	1	7	0	4	0	1	0	1	35	1	6
26	2	3	0	1	4	0	4	0	0	4	4	37	1	9
26	3	3	0	1	5	0	3	0	0	6	1	39	0	1

26	4	3	0	1	9	0	4	0	0	4	1	23	8	1
26	5	3	0	1	4	0	2	0	0	0	0	30	4	2
26	6	3	0	1	0	0	3	0	0	0	1	24	2	3
26	7	3	0	1	4	0	6	0	0	0	1	25	0	1
26	8	3	0	1	12	0	2	0	0	5	0	65	17	0
2	1	3	0	1	7	0	3	0	0	1	1	12	0	6
2	2	3	0	1	2	0	5	0	0	0	1	14	0	1
2	3	3	0	1	7	0	5	0	0	5	1	29	0	1
2	4	3	0	1	7	0	7	2	0	1	1	83	10	5
2	5	3	0	1	8	0	10	0	0	0	1	30	0	0
2	6	3	0	1	0	0	4	0	0	0	1	10	0	1
2	7	3	0	1	4	0	6	1	0	0	1	15	0	0
2	8	3	0	1	7	0	3	1	0	0	1	10	0	0
34	1	3	0	1	7	0	3	2	0	1	0	12	6	7
34	2	3	0	1	1	0	1	0	0	1	0	10	7	5
34	3	3	0	1	1	0	3	0	0	1	0	14	6	3
34	4	3	0	1	4	0	11	0	0	0	0	14	2	1
34	5	3	0	1	8	0	2	0	0	5	0	10	2	5
34	6	3	0	1	4	0	3	0	0	1	0	9	5	3
34	7	3	0	1	0	0	5	1	0	0	0	9	0	0
34	8	3	0	1	3	0	3	1	0	2	0	6	0	0
36	1	3	0	1	4	0	2	0	1	1	3	35	28	16
36	2	3	0	1	0	0	3	0	0	0	1	10	6	1
36	3	3	0	1	5	0	3	0	0	1	1	35	23	1
36	4	3	0	1	4	0	6	0	0	0	1	30	23	2
36	5	3	0	1	8	0	5	0	0	0	2	8	3	0
36	6	3	0	1	0	0	3	0	0	0	1	11	8	1
36	7	3	0	1	0	0	5	0	0	0	1	15	5	2
36	8	3	0	1	3	0	3	0	1	0	1	32	7	4
37	1	3	0	1	17	5	2	5	0	11	1	87	0	10
37	2	3	0	1	7	4	7	5	2	6	1	117	68	4
37	3	3	0	1	9	0	3	0	0	6	1	38	4	1
37	4	3	0	1	12	0	12	0	0	8	1	23	2	5
37	5	3	0	1	0	1	7	6	0	0	2	54	0	4
37	6	3	0	1	1	0	6	0	0	0	0	17	5	3
37	7	3	0	1	4	2	3	1	0	0	1	14	0	0
37	8	3	0	1	9	0	6	1	0	1	1	8	0	0
44	1	3	0	1	4	0	5	0	0	0	1	19	11	4

44	2	3	0	1	0	0	2	0	0	0	1	13	9	2
44	3	3	0	1	2	0	1	0	0	4	1	25	22	3
44	4	3	0	1	8	0	8	0	0	0	1	23	11	4
44	5	3	0	1	9	0	5	0	0	1	1	79	25	4
44	6	3	0	1	0	0	4	0	0	0	1	11	7	1
44	7	3	0	1	0	0	8	0	0	0	1	9	1	1
44	8	3	0	1	4	0	2	0	0	0	1	11	0	0
46	1	3	0	1	7	0	2	0	0	3	0	65	0	8
46	2	3	0	1	5	0	5	0	0	5	0	12	0	6
46	3	3	0	1	2	0	3	0	0	2	0	28	0	2
46	4	3	0	1	12	0	9	0	0	6	0	36	0	3
46	5	3	0	1	4	1	10	5	0	0	0	15	0	2
46	6	3	0	1	0	0	3	0	0	0	0	13	0	5
46	7	3	0	1	3	0	4	0	0	0	0	7	0	1
46	8	3	0	1	16	0	9	0	3	8	0	51	0	4
6	1	3	0	1	5	3	0	2	0	2	0	22	3	11
6	2	3	0	1	4	2	3	1	0	1	0	14	1	2
6	3	3	0	1	3	0	2	1	0	1	0	25	22	2
6	4	3	0	1	4	1	6	1	0	0	0	11	3	1
6	5	3	0	1	4	0	2	1	0	1	0	19	13	0
6	6	3	0	1	4	0	4	0	0	1	0	11	6	1
6	7	3	0	1	4	1	3	1	0	2	0	15	7	3
6	8	3	0	1	3	2	4	1	0	0	0	10	4	1
9	1	3	0	1	4	0	4	0	0	1	1	38	0	19
9	2	3	0	1	11	0	3	0	0	8	0	33	4	24
9	3	3	0	1	20	0	9	0	2	13	0	80	0	20
9	4	3	0	1	21	0	9	0	1	18	1	39	4	6
9	5	3	0	1	19	1	8	0	0	15	0	79	5	17
9	6	3	0	1	12	2	6	0	0	12	0	13	5	8
9	7	3	0	1	18	1	6	1	0	18	0	48	0	7
9	8	3	0	1	15	1	5	1	0	11	0	68	0	12
12	1	4	1	1	5	4	1	3	0	4	0	7	1	9
12	2	4	1	1	0	0	2	0	0	0	2	15	2	4
12	3	4	1	1	4	0	2	0	0	4	1	9	0	4
12	4	4	1	1	0	0	5	0	1	1	1	16	4	4
12	5	4	1	1	0	0	3	0	0	0	4	7	4	7
12	6	4	1	1	0	0	3	0	0	0	0	5	2	5
12	7	4	1	1	0	0	5	0	0	3	0	11	2	16

12	8	4	1	1	3	0	3	0	3	0	1	25	2	5
14	1	4	1	1	10	0	3	0	1	10	2	29	0	11
14	2	4	1	1	11	0	4	0	0	11	2	52	0	9
14	3	4	1	1	9	0	4	0	0	5	0	37	0	3
14	4	4	1	1	4	0	7	0	0	0	1	16	0	2
14	5	4	1	1	5	3	9	2	0	5	0	23	0	1
14	6	4	1	1	8	0	3	0	0	8	0	25	0	5
14	7	4	1	1	4	0	3	0	0	0	0	13	1	0
14	8	4	1	1	9	0	6	0	0	5	0	22	9	1
17	1	4	1	1	5	0	1	0	0	2	0	34	1	9
17	2	4	1	1	8	0	3	0	0	8	0	25	0	4
17	3	4	1	1	0	0	3	0	0	0	0	8	0	4
17	4	4	1	1	4	2	3	3	0	0	0	13	0	1
17	5	4	1	1	0	0	3	0	0	0	0	11	0	1
17	6	4	1	1	0	0	4	0	1	0	0	6	0	2
17	7	4	1	1	4	0	2	3	0	0	0	6	0	4
17	8	4	1	1	9	0	2	0	0	1	0	33	11	0
19	1	4	1	1	6	0	4	0	0	0	0	14	8	1
19	2	4	1	1	3	0	1	0	0	0	0	9	8	1
19	3	4	1	1	1	0	8	0	0	0	0	28	20	1
19	4	4	1	1	6	0	10	0	0	0	0	11	1	1
19	5	4	1	1	3	0	6	0	0	0	0	14	8	1
19	6	4	1	1	0	0	4	0	0	0	0	11	6	1
19	7	4	1	1	0	0	5	1	2	0	0	25	2	0
19	8	4	1	1	5	0	2	1	0	1	0	18	0	0
1	1	4	1	1	10	0	2	0	0	7	1	41	31	0
1	2	4	1	1	4	2	1	0	0	1	1	49	42	0
1	3	4	1	1	2	0	4	0	0	2	0	37	33	0
1	4	4	1	1	4	0	10	0	0	1	0	20	9	0
1	5	4	1	1	6	1	2	0	0	2	1	19	4	0
1	6	4	1	1	0	0	4	0	0	0	0	11	7	0
1	7	4	1	1	4	3	8	4	0	0	0	29	16	0
1	8	4	1	1	8	1	7	1	0	1	1	23	12	0
23	1	4	1	1	0	0	5	0	0	0	1	17	1	5
23	2	4	1	1	4	0	2	0	0	0	1	24	8	9
23	3	4	1	1	5	0	5	0	1	5	0	47	17	7
23	4	4	1	1	4	0	7	0	0	0	0	24	7	2
23	5	4	1	1	6	0	10	0	0	2	0	19	4	3

23	6	4	1	1	1	0	8	0	4	0	0	35	11	6
23	7	4	1	1	9	0	8	0	1	5	0	69	1	2
23	8	4	1	1	8	0	3	0	2	3	0	85	0	7
29	1	4	1	1	8	0	6	0	0	0	0	42	0	6
29	2	4	1	1	3	0	3	0	0	3	0	35	0	10
29	3	4	1	1	4	0	1	1	0	2	0	34	0	15
29	4	4	1	1	5	0	15	0	2	1	0	23	0	1
29	5	4	1	1	4	0	9	5	1	2	0	69	7	3
29	6	4	1	1	0	0	3	0	0	0	0	25	0	1
29	7	4	1	1	0	0	15	0	0	0	0	48	0	3
29	8	4	1	1	7	0	8	0	0	0	0	27	3	0
35	1	4	1	1	9	0	3	2	1	3	1	41	2	6
35	2	4	1	1	1	0	7	0	0	2	0	34	1	4
35	3	4	1	1	3	0	4	0	2	0	0	44	0	1
35	4	4	1	1	8	0	10	1	0	4	0	39	2	2
35	5	4	1	1	16	0	10	0	0	10	0	65	1	6
35	6	4	1	1	0	0	4	0	0	0	0	16	0	2
35	7	4	1	1	0	0	3	1	0	0	0	37	4	3
35	8	4	1	1	9	0	7	1	1	1	0	54	11	2
39	1	4	1	1	4	0	3	0	0	0	1	72	5	10
39	2	4	1	1	1	0	7	1	2	1	0	48	0	9
39	3	4	1	1	0	0	12	1	3	0	0	60	28	4
39	4	4	1	1	4	0	13	0	3	0	0	49	5	1
39	5	4	1	1	4	3	11	5	1	0	0	116	27	0
39	6	4	1	1	0	0	5	0	0	0	0	12	3	1
39	7	4	1	1	4	0	3	1	0	4	0	54	1	5
39	8	4	1	1	5	2	6	3	0	1	0	43	1	0
41	1	4	1	1	8	0	6	0	2	5	0	50	0	22
41	2	4	1	1	6	0	3	0	0	11	0	43	2	9
41	3	4	1	1	6	0	3	0	0	5	0	21	0	1
41	4	4	1	1	10	0	11	0	0	7	0	18	0	1
41	5	4	1	1	4	7	9	14	0	0	0	20	0	0
41	6	4	1	1	5	0	3	0	0	7	0	11	0	2
41	7	4	1	1	4	0	5	1	0	1	0	14	0	1
41	8	4	1	1	4	3	3	1	0	3	0	12	0	3
45	1	4	1	1	4	0	4	0	0	0	0	33	0	5
45	2	4	1	1	0	0	7	0	0	0	0	33	11	5
45	3	4	1	1	2	0	2	0	0	2	0	42	34	6

45	4	4	1	1	4	0	7	0	0	0	0	22	15	1
45	5	4	1	1	4	0	7	0	0	0	0	39	12	0
45	6	4	1	1	4	0	3	0	0	0	0	13	9	2
45	7	4	1	1	8	0	4	0	0	0	0	9	2	0
45	8	4	1	1	3	0	3	0	0	0	0	14	7	0
5	1	4	1	1	5	1	4	1	0	2	0	21	0	3
5	2	4	1	1	4	1	3	1	0	2	1	10	0	6
5	3	4	1	1	0	2	3	1	0	0	0	23	16	8
5	4	4	1	1	4	4	3	1	0	1	0	13	5	1
5	5	4	1	1	4	1	7	1	0	0	0	19	9	0
5	6	4	1	1	0	1	5	1	0	0	0	11	6	1
5	7	4	1	1	4	3	5	3	2	0	0	23	1	1
5	8	4	1	1	9	4	6	3	0	1	0	30	4	0