16

Localness of Location-based Knowledge Sharing: A Study of Naver KiN "Here"

SANGKEUN PARK, KAIST, South Korea
MARK S. ACKERMAN, University of Michigan, USA
UICHIN LEE, KAIST, South Korea

In location-based social Q&A services, people ask a question with a high expectation that local residents who have local knowledge will answer the question. However, little is known about the locality of user activities in location-based social Q&A services. This study aims to deepen our understanding of location-based knowledge sharing by investigating the following: general behavioral characteristics of users, the topical and typological patterns related to geographic characteristics, geographic locality of user activities, and motivations of local knowledge sharing. To this end, we analyzed a 12-month period Q&A dataset from Naver KiN "Here," a location-based social Q&A mobile app, in addition to a supplementary survey dataset obtained from 285 mobile users. Our results reveal several unique characteristics of location-based social Q&A. When compared with conventional social Q&A sites, users ask and answer different topical/typological questions. In addition, those who answer have a strong spatial locality wherein they primarily have local knowledge in a few regions, in areas such as their home and work. We also find unique motivators such as ownership of local knowledge and a sense of local community. The findings reported in the article have significant implications for the design of Q&A systems, especially location-based social Q&A systems.

CCS Concepts: • **Information systems → Location based services**; **Social networks**; **Crowdsourcing**; *Geographic information systems*; Collaborative and social computing systems and tools;

Additional Key Words and Phrases: Knowledge sharing, mobile applications

ACM Reference format:
Sangkeun Park, Mark S. Ackerman, and Uichin Lee. 2018. Localness of Location-based Knowledge Sharing: A Study of Naver KiN "Here". *ACM Trans. Web* 12, 3, Article 16 (July 2018), 33 pages.
https://doi.org/10.1145/2983645

1 INTRODUCTION

Location-based social Q&A allows users to directly tap into the knowledge of local people to obtain information related to a particular geographic location or area (e.g., insider tips about the best places to go). When compared with traditional social Q&A services such as Yahoo! Answers[1]

[1] https://answers.yahoo.com.

This work was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C4A7065960).
Authors' addresses: S. Park, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, South Korea; email: sk.park@kaist.ac.kr; M. S. Ackerman, University of Michigan, 500 S. State Street, Ann Arbor, MI, 48109, USA; email: ackerm@umich.edu; U. Lee (corresponding author), KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, South Korea; email: uclee@kaist.edu.

https://doi.org/10.1145/2983645

and Naver KiN,[2] one significant distinction of location-based social Q&A is local knowledge sharing—the questions that are related to a local community (i.e., primarily about local places and services) are answered by those who have local knowledge (e.g., current local residents). As our findings indicate, this leads to very different usage and distinct design implications compared to traditional social Q&A services.

Researchers have endeavored to build location-based online social platforms that facilitate social interaction and local knowledge sharing, e.g., SocialSearchBrowser [6] and Micro-Blog [15]. Furthermore, commercial platforms have recently been designed for location-based social Q&A, such as Locql[3] and Naver KiN "Here" (NKH).[4] The key feature of location-based social Q&A services is that questions are classified by regions instead of topics. From this distinction, users ask different types and topics of questions from conventional social Q&A services, and users behavioral patterns are also different. The questions ask people who have local knowledge of a specific region to share their local knowledge rather than topical knowledge (local knowledge may be regional). In this article, we define localness as a geographical social tendency in knowledge sharing. For example, Davenport and Prusak [9] explained localness of knowledge, that is, people tend to share their knowledge with local neighbors because people usually trust those they know in local areas or organizations.

We found that understanding of the localness of knowledge sharing remains limited because prior studies have primarily focused on evaluating prototype systems and drawing design implications with small-scale user studies [6, 7, 15]. We focus on the localness of knowledge sharing, mainly related to the influence of regional characteristics for knowledge sharing (e.g., typological or topical patterns) and the behavioral patterns of user contributions regarding local knowledge (e.g., the number of regions and their extent for local knowledge sharing).

In this article, we deepen the understanding of location-based knowledge sharing through analysis of a large-scale, longitudinal dataset and a survey of current users from Naver KiN "Here," a mobile app for location-based social Q&A in South Korea. The app uses smartphone GPS sensors to track users' locations. When the users ask questions, their specific geographic coordinates are recorded with the question. Asking and answering activities are also possible via a web interface. In the latter case, rough geographic coordinates are recorded based on the IP address instead of GPS. All Q&A data are accessible on the web, but the accuracy of the geographic coordinates of questions differs depending on the interface—app or web. The Q&A data with geographic coordinates allow investigation of the spatial characteristics of location-based social Q&A. In addition, location-based questions are categorized by region (district, city) rather than by topic (technology, travel), so the questions are about local knowledge and the answers include local knowledge corresponding to the specific location. These differences from conventional Q&A led us to focus on this study.

We consider the following research questions to investigate the key aspects of localness:

— First, what are the general usage patterns of location-based social Q&A? (e.g., question and answer distributions, answer rate, answer delay, and interactive map usage)
— Second, are topical/typological patterns related to geographic characteristics, and if so, how?
— Third, how are users geographically focused in their asking and answering activities?
— Fourth, what answer motivations are unique in location-based social Q&A?

[2]http://kin.naver.com.
[3]http://www.locql.com.
[4]NAVER Corp. (http://www.naver.com) Android: https://play.google.com/store/apps/details?id=com.nhn.android.kin (iOS is not supported, http://m.kin.naver.com is recommended for iOS users).

To answer the research questions, we crawled a 12-month period Q&A dataset from Naver KiN “Here,” and we performed supplementary surveys with 285 Naver KiN “Here” users. Using these datasets, we first analyzed general user behavior to understand how behavior patterns are different from those of conventional social Q&A. We then performed topic analyses using different geographic scales (e.g., district/city) and evaluated how the geographic characteristics were reflected in the Q&A activities, and we also performed a type analysis to examine what types of questions are asked (e.g., factual or opinion questions). Moreover, we analyzed the district/city-level activity patterns of users to gain insights into their geographic focus and performed spatial clustering analyses to quantify the geographic locality of user contributions. We further conducted a content analysis of the user survey results to identify user motives of local knowledge sharing that are unique to location-based social Q&A.

Our primary findings are summarized as follows:

— First, we found that users’ overall Q&A frequencies followed a power-law distribution as in conventional Q&A services [30, 37]. However, Naver KiN “Here” had lower answer rates and longer delays, possibly because questions were categorized based on regions instead of topics, and answering requires both local and topical knowledge. Regarding askers’ usage of an interactive map, we found that users were not precise on marking a question’s fine-grained location (e.g., point-of-interest), but they tended to correctly indicate district/city-level areas of interest, which provide relevant contextual information for answering.
— Second, Naver KiN “Here” had very different topical (e.g., travel, education) and typological (e.g., information, suggestion, and opinion) patterns than traditional social Q&A sites. The overall topic distribution had a significantly higher level of travel and lifestyle information topics, which are often localized in nature. We also examined question types such as information, suggestion, and opinion. We found that topical distributions varied widely across different districts, as were appropriate for the areas in question. Factual information seeking was surprisingly high (67.7%), followed by recommendations (17.5%), and the type distribution varied significantly across different topics.
— Third, there was strong spatial locality of user contributions. Askers’ activities typically spanned multiple districts and cities, whereas many answerers focused on very few cities. We discovered that answerers primarily focused on 1–3 spatial clusters using the DBSCAN algorithm [12]. These clusters were closely related to the users’ life experiences (e.g., current/former home as well as work and school); many questions required very specific local knowledge that was difficult to answer without local experience. Furthermore, the city selection strategies of the answerers did not significantly change over time. Nonetheless, there were answerers with many spatial clusters who were likely either factual information experts or state/province-level local experts.
— Fourth, we found unique motivators for local knowledge sharing, i.e., ownership of local knowledge (competence about local knowledge learned over many years) and sense of community (membership and fulfillment of needs). These are all different from those reported for “normal” social Q&A systems.

This article significantly extended our prior work [36] as follows:

— In Section 2, we provide a more comprehensive review of related studies in the field and illustrate how our work is different from prior studies.
— We added a section that reports generic usage behavior analysis results (Section 5). We have also performed statistical and manual analyses to understand question/answer distributions, answer delay/rates, and interactive map usage.

—In Section 6, we investigate topically focused districts/cities to better understand question characteristics. In addition, we have studied topic/type distribution in location-based social Q&A by conducting content analysis of 1,000 randomly selected questions.
—In Section 7, we have significantly extended our analysis on spatial locality. We have analyzed askers' cluster distributions and cluster-level entropy and identified how their activities were different from answerers' activities. Furthermore, we have identified the characteristics of those answerers who did not demonstrate strong spatial locality.
—In Section 9, we additionally discuss how we can leverage user behavior patterns to design location-based social Q&A systems.

The rest of the article is organized as follows: Section 2 presents prior studies related to this work. Section 3 describes how users interact on Naver KiN "Here" for location-based question and answer. Section 4 includes a description of how we collected data and what methods we used to analyze the dataset. Section 5 provides our analysis of general user behaviors, compared with those of conventional social Q&A services. Section 6 presents topical and typological characteristics of questions. Section 7 considers geographical focus of asker/answerer activities. Section 8 presents our survey result analysis regarding answer motivations in location-based social Q&A. In Section 9, we discuss design implications and limitations. Finally, Section 10 concludes the article.

2 BACKGROUND AND RELATED WORK

In this section, we describe prior work on online knowledge sharing. This includes asking and answering behavior in social Q&A services or in social network services and prior research on location-based mobile information seeking and location-based social Q&A. We explain how our research differs from previous work.

2.1 Conventional Social Q&A

In social Q&A services, such as Yahoo! Answers and Naver KiN, people can ask and answer questions, as well as rate or vote on these answers and questions. Alternatively, people can also use social network services, such as Twitter and Facebook, SMS, and mobile apps to ask/answer questions. The characteristics of these conventional social Q&As, in general, have been extensively studied. Three types of findings stand out. First, many researchers focused on what kinds of questions are asked and answered. Kim et al. [22] classified questions in Yahoo! Answers as soliciting facts, opinions, or suggestions. Morris et al. [32] showed that the most popular question topics are technology and entertainment, and the most popular question types are recommendations and opinions in status message Q&A behavior on Twitter and Facebook. In Mobile Q&A, Lee et al. [26] revealed that *lifestyle* topics and *information* types of questions dominate, and the distribution of question types across various topics is significantly different. In addition, Forte et al. [14] found that American teenagers mostly sought factual knowledge via social network services such as Facebook, Twitter, and Google+. In an analysis of Yahoo! Answers, Adamic et al. [1] found that user participation varied widely (and is skewed) depending on the topic, and that knowledge sharing patterns across different topic categories existed (e.g., experts in different domains help one another). Similarly, Nam et al. [33] demonstrated that Naver KiN users' level of participation is highly skewed and intermittent, and that their expected level of expertise is lower than that found in other online help forums.

Second, previous studies have examined the response time in Q&A services or communities. Zhang et al. [52] showed that in a specialized Q&A site for theJava programming language, the average time to receive an answer was around 9h. However, Mamykina et al. [30] reported that 50% of all questions in StackOverflow received the first answer within 12min. In addition, Lee et al. [26] revealed that the average time to receive an answer is 15.5min in Naver Mobile (SMS) Q&A. Furthermore, Hsieh et al. [19] reported that the average time to receive the first answer was 2h and 53min, and 20% of the questions never received an answer in Microsoft Live Q&A. Further, Lampe et al. [25] revealed that only 56% of mobile posts such as requests for help, information, or other types of support received at least one comment on Facebook.

Finally, studies have examined why people answer questions. The motivation behind answering questions is largely dependent on a mixture of intrinsic factors (e.g., enjoyment, feelings of gratitude and respect) and extrinsic factors (e.g., monetary rewards, reputation systems, etc.) [38]. Nam et al. [33] demonstrated that the motivation for answering in Naver KiN comes from both intrinsic and extrinsic factors; altruism is the leading factor, followed by business motives, learning, hobbies, and earning points. In pay-for-answer Q&A sites, researchers have reported that financial incentives and social factors are the key motivators [10, 20, 27, 39, 41].

Our work differs from prior studies in that we investigated the localness of knowledge sharing by analyzing user behavioral patterns, topical/typological patterns related to geographic characteristics, and geographic locality of user activities. We also extended the prior studies by identifying unique motivators for local knowledge sharing.

2.2 Mobile Information Seeking and Local Searches

Mobile information seeking has received significant attention. *PewResearchCenter* recently reported that 68% of Americans owned a smartphone in the United States in December 2015 [18]. And *KT Corporation's Research Center* in Korea reported that Korea has the 4th highest smartphone penetration; 83.0% of Koreans owned a smartphone [4]. Sohn et al. [42] found that approximately 38% of mobile information needs involve local intent and 72% were prompted by explicit contextual factors, including activity, location, time, and conversation. Similarly, Church et al. [8] demonstrated that over 40% of entries were location-based and 67% of entries were generated when users were located away from familiar contexts. Henrich and Ludecke [17] analyzed the key properties of geographic information needs from the perspective of geographic information retrieval, and they found intentions for geo-reference (e.g., to perform activities and obtain facts in a given location), geo-coverage of relevant documents, the shape of geo information needs (point/region, near/within), and the current location of a user.

Search engines are frequently used as information channels for specific locations; this usage is called a *local search*. Examples include searches for local restaurants, local stores, and local transportation information. In the past, researchers have analyzed search engine log data to understand the usage of local searches. In recent years, the overall portion of geo-queries has increased sharply (possibly due to increased smartphone use), where 79% of mobile phone owners used their devices to search for local information in 2012 [43]. Jones et al. [21] analyzed Yahoo! query logs and reported the characteristics of their geo queries: the distance between the home location and queried location (30% were within 100km) and query reformulation patterns. Teevan et al. [45] conducted a survey of local search usage and found that local searches tended to be highly contextual for a current location and that time constraints and social factors had a significant impact on usage behavior.

In this work, we complement earlier studies on geographic information needs [6, 7, 17, 42] using a real-world dataset, as existing studies examined limited datasets such as web search logs [17] and small-scale user logs (e.g., diaries and user studies) [6, 7, 42].

2.3 Location-based Social Q&A

Table 1 summarizes existing location-based social Q&A systems based on their design features, i.e., message access (remote, *in situ*), delivery (push, pull), message content (text, multimedia:

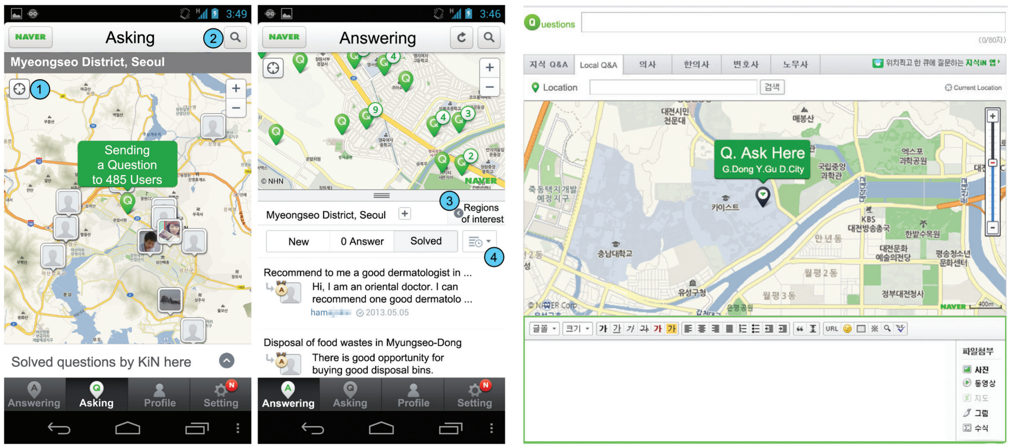
Table 1. Classification of Location-based Social Q&A

Service	Message Access	Question Delivery	Message Content	Delivery Target	Filtering
GeoNote [11]	Remote/ *In-situ*	Pull Push (geocast)	Text/MM/Map	Area/Thing	Topic/ Proximity
Micro-Blog [15]	Remote/ *In-situ*	Pull Push (geocast)	Text/MM/Map	Area	Topic/ Proximity
CityFlocks [3]	Remote/ *In-situ*	Pull Push (unicast)	Text/Map	Area/Person	Proximity
Twitter (TSA Tracker) [34]	Remote/ *In-situ*	Pull Push (multicast)	Text/Map	Area/Person	N/A
SocialSearchBrowser [6]	*In-situ*	Pull	Text/Map	Area	Topic/ Proximity
Glaucus [5]	Remote	Push (unicast)	Text	Topic	Interest Areas
Locql	Remote	Pull	Text/Map	POI/City	Interest Areas
Yahoo! Answers	Remote	Pull	Text/Map	City	Topic
Naver KiN "Here"	Remote/ *In-situ*	Pull Push (geocast)	Text/MM/Map	Area	Interest Areas

MM), user interface (text, map), message destination (location, thing, Point of Interest (POI), city), and message filtering (topic/area subscription, proximity). In this article, we define POI as a specific point location such as a supermarket, a hardware store, or a university. In existing systems, message access modes are the key differentiating factor. Remote access means that a user can post and retrieve messages in a remote location; *in situ* access means that a user can retrieve relevant messages only when the user approaches the area of interest. *In situ* access is suitable for facilitating information sharing among local residents, such as in SocialSearchBrowser. Questions can be delivered to target users in various ways: relaying questions to people in a regional area as in Naver KiN "Here" and Micro-Blog (geocast) or sending questions directly to a person as in CityFlock (unicast). For location-based social Q&A services, text and interactive maps can be used for question/answer navigation [7]. Existing systems support various filtering mechanisms ranging from place/area/topic subscription to proximity sensitive filtering. In this article, we study the localness of local knowledge sharing by studying Naver KiN "Here," which has representative design features of location-based social Q&A services.

2.4 Localness of Information Sharing Behavior

Prior studies have explored the localness of information sharing, such as influence of regional characteristics on social media usage [24] and user contribution behavior of local content in online social systems [16, 28, 49]. Kulshrestha et al. [24] demonstrated that shared national, linguistic, and cultural backgrounds had a significant impact on Twitter usage (e.g., social links and information exchanges). Yardi and Boyd [49] investigated how local community members used Twitter to share and exchange information about local events. A large percentage of Wikipedia users only edit a few geo-pages (e.g., city, school), and the geo-locations of the edited geo-pages tend to be localized within a 100km radius [28]. Hecht and Gergle [16] compared the mean contribution distance between the specified location of a contributor and locations of each Flickr posts or Wikipedia edit (geo-page) by the contributor. They demonstrated that the Wikipedia edits had a longer contribution distance than Flickr posts. In addition, Sen et al. [40]



(a) Naver KiN "Here". The numbered labels are (1) GPS, (2) search, (3) region of interest, and (4) sort by distance/time.

(b) Naver KiN Local Q&A

Fig. 1. User interface (the original text was translated into English for ease of understanding).

found extensive geographic inequalities in localness based on the socioeconomic status of the local population and the health of the local media.

Prior studies have provided valuable insights into the localness of knowledge sharing. However, knowledge sharing in location-based social Q&A differs significantly from that in Twitter and Wikipedia, because it is designed to resolve everyday life geographic information needs with the help of people with local knowledge. Our goal was to study the localness of location-based social Q&A by conducting spatial clustering analyses of a large-scale real-world Q&A dataset.

3 USER INTERACTION IN NAVER KIN "HERE"

We investigated NKH, which is a mobile app for location-based social Q&A. NKH was released on December 3, 2012 to provide a mobile interface for existing location-based social Q&A, called Naver Local Q&A[5] (see Figure 1(b)). The asking and answering activities are compatible in both interfaces. For example, the same question-and-answer data are shared regardless of whether NKH or Naver Local Q&A is used. However, there are several differences between the app and web interfaces of the service regarding interaction methods.

NKH uses smartphone GPS sensors to track users current locations (only when the app is launched). The current location is set as a region to ask a question by default. Users can freely change the location on the interactive map by zooming in or out, moving by dragging, or searching the city or district name to ask a question The questions are then organized by proximity to the current location. When a user pans or zooms on the map, the question list is automatically updated depending on the zoom level; the questions within a nearby geographic area are aggregated into a question bubble with a number inside indicating the number of questions (Figure 1(a)).

The web interface also provides an interactive map. By default, the location to ask a question is automatically set based on the IP address. Users can freely change the location on the interactive map by zooming in or out, moving by dragging, or searching the city or district name to ask a

[5] http://kin.naver.com/qna/list.nhn?dirId=12.

question. The questions, however, are organized only by region categories (for example, cities and districts) regardless of the user's current location.

In the mobile app, active app users, who have answered more than five questions in a given area on the map, are shown on the location of the last question they answered on the map so that users can locate active app users. Figure 1(a) shows an interactive map and a list of questions regarding the current location based on GPS. If a mobile user asks a question, then active app users on the map receive notifications about the new question. On the web interface, however, active users are not displayed on the map.

Questions and answers via NKH are marked with a mobile icon next to the title of the questions or answers, allowing other people to recognize that those questions or answers have been posted via NKH.

Except for these differences, other location-based social Q&A mechanisms are the same in both web and mobile interfaces. In both interfaces, users can subscribe to the regions of interest (that is, district or city) for question filtering so that the system automatically provides new questions based on their subscription information. Users receive points for asking and answering questions. Furthermore, users can acquire various badges by meeting certain conditions, such as providing fast responses, highly accurate answers, or a large number of answers for a specific region.

Before detailing our results, we provide an explanation of the administrative divisions in Korea, which is important, because they differ from other countries. The administrative divisions have four levels: province ("Do"; there are nine provinces in Korea), city ("Si"; typical size of 100–1,000km^2), sub-city ("Gu"; typical size of 10–100km^2), and district ("Dong"; typical size of 1–10km^2). The levels are more fine-grained than those of western countries, e.g., the USA, which are typically composed of three levels: state, county/shire, and city/town/village. People typically refer to city/sub-city/district names when searching for or speaking about places in South Korea.

4 METHODOLOGY

In this section, we first explain how we collected data (i.e., the Q&A dataset and the survey dataset). Second, we describe various approaches we used to deepen understanding of knowledge sharing in location-based social Q&A via Naver KiN "Here."

4.1 Data Collection

4.1.1 Q&A Dataset. The Naver Local Q&A web site shows only the questions from within the past month. Every week, therefore, we downloaded all question-listing web pages (see Figure 19), which included all the URLs of each question page, from the most recent month using Wget[6] from the main page of Naver Local Q&A from December 17, 2012 to December 31, 2013. This ensured that we did not miss any question pages.

In January 2014, we extracted the URLs of each question from the downloaded question-listing web pages, and downloaded all of the HTML files from each question page (see Figure 2), naming each file with its own URL address.[7] This was downloaded onto our server, using the Python built-in library urllib2[8] to overwrite duplicated question pages into a single question page. (Because we downloaded the recent months questions each week, there was some duplication from week to week.)

We then extracted detailed Q&A information from the question pages using BeautifulSoup, which is a Python library for parsing HTML/XML files. We used four servers in parallel to

[6] https://www.gnu.org/software/wget.
[7] For example, a file name was http://kin.naver.com/qna/detail.nhn?d1id=12&dirId=120929&docId=288100480.
[8] https://docs.python.org/2/library/urllib2.html.

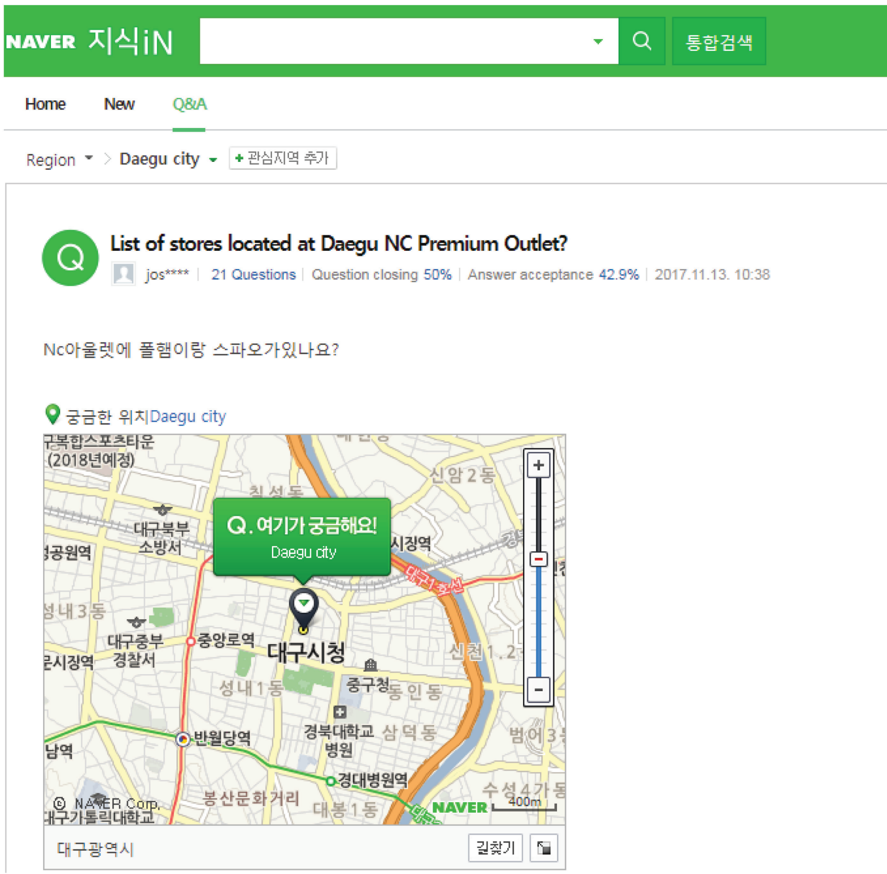


Fig. 2. A question asking page (the original text was translated into English for ease of understanding).

complete this work as quickly as possible, which required just two days. The question pages include (1) question title, content, posting time, categorized region, posted coordinates (latitude and longitude), and posting device (i.e., mobile or PC), (2) answer information such as answerer ID, posted time, answer content, answer status information (i.e., if it is accepted or not), and posting device (i.e., mobile or PC), and (3) user information (for example, asker's ID, the question closing rate, and answer acceptance rate).

One limitation of the data collection is that there were some deleted questions when we accessed the downloaded question pages. The questions existed when we downloaded the question-listing web pages, which included the URLs of each question from the main page of Naver Local Q&A; but they were deleted when we accessed the URLs of the question pages. This is mainly because of question policy violation (for example, sexual content) or because the askers did not wish to keep the question posted. The URLs of deleted question pages returned the message, "This page is removed" when we accessed the URL to download the HTML document. In these cases, we excluded the URLs from our Q&A dataset for the analysis. However, we assumed that these deleted questions did not affect our result of analysis because of their rarity (i.e., fewer than 1,000 questions).

Each question page also includes a maximum of 10 answers. If the question has more than 10 answers, then a page navigation appears at the bottom of the page so that people can navigate to other answers. Thus, when we crawled the answer information for each question, if the question page included a page navigation, we accessed the pages and downloaded the HTML file of the additional answers and extracted the answer information.

At the beginning of the crawling process, the crawling servers IP address was blocked, because it may have been regarded as a DDoS attack, which is an attempt to make an online service unavailable by overwhelming it with traffic. Thus, we started crawling with 10min breaks every 30min. This allowed us to crawl without being blocked. As a result of the crawl, we obtained a total of 508,334 question pages and 567,156 answers were obtained for the questions. We saved all the questions, answers, and user information into a database using MySQL for data analysis.

4.1.2 Survey Dataset. We supplemented the question and answer data with results from an online survey. The online survey was created using Survey Monkey.[9] First, analyzing the question and answer data revealed that 4,557 users had answered at least one question via NKH. We then sent those 4,557 users requests to participate in the survey, with a questionnaire link, via Naver email (that is, user ID@naver.com) in August 2013. The online survey was open for one week, and the total number of responses was 293. After removing duplicate and erroneous responses, 285 responses remained. The low response rate may have been caused by the fact that Naver emails may not be users' primary email accounts. At the end of the survey, we randomly selected 50 participants and rewarded an online gift voucher worth $10.

The survey questions asked about demographics (e.g., age, gender, and occupation), strategy for formulating reference location, regions of interest, the number of answerable questions in the region of interest, and answer motivations. The survey data indicated that 59.7% of participants were males and 40.3% were females. Participants were predominantly in their teens (10–19 years; 30%) and in their 20s (43%); the remainder were in their 30s (14.4%), 40s (10.8%), and 50s (1.8%). Their occupations were quite diverse: middle/high school students (29.8%), college/graduate students (23.3%), financial/service workers (12.1%), lawyers (0.3%), designers/artists (3%), homemakers (2.3%), architects (2.3%), engineers (7.2%), miscellaneous area workers (11.8%), and unemployed (6.2%).

4.2 Data Analysis Methods

We used a mixed method of quantitative and qualitative analyses to understand the localness of knowledge sharing, i.e., the influence of regional characteristics for knowledge sharing (e.g., typological/topical patterns) and behavioral patterns of user contributions regarding local knowledge (e.g., the number of regions and their extent for local knowledge sharing).

Quantitative analyses: We analyzed all the crawled dataset to characterize overall regional characteristics and user behaviors of NKH. Our analyses included general usage patterns, topical diversity across different regions, and spatial locality of knowledge sharing behaviors.

Qualitative analyses: We additionally performed qualitative analyses of a sampled dataset to analyze geographic relevancy and question topics/types, and a survey dataset to categorize the strategies for formulating reference location and answer motivations.

Note that due to their privacy policy, Naver only revealed the first three characters of the askers ID, followed by three asterisks as in "abc****." For this reason, we tried to uniquely identify each asker using the question closing rate and answer acceptance rate (see Figure 2). "Question closing rate" indicates how often the user selected the best answer to questions the user asked (e.g., 50.0%). "Answer acceptance rate" indicates how often the users answers were selected as the best answers (also expressed as a percentage, such as 42.9%). As we crawled user information

[9]http://surveymonkey.com.

simultaneously (within two days), we assumed that some users could be considered as unique users if they had the same first three characters, question closing rate, and answer acceptance rate at any given moment. For example, in Figure 2, the askers ID is “jos****,” question closing rate is 50.0%, and answer acceptance rate is 42.9%. We assumed that if the askers whose ID was “jos****” had the same question closing rate, and answer acceptance rate, then the askers were actually one unique asker, shown as “jos500429” (e.g., jos + 50.0 + 42.9) in our analysis.

4.2.1 Geographic Relevancy. Church et al. [6] showed that a map interface is popular for location queries and appears to be more appropriate for representing locations in location-based queries than a text interface. As Table 1 shows, map interfaces have been widely employed in location-based Q&A systems. In addition, Henrich and Luedecke [17] demonstrated that users prefer to use POI/district/city names along with constraints such as “at,” “within,” and “near” in geographic queries. However, there remains a lack of analysis on how users accurately pinpoint the location of a question in location-based social Q&A.

To judge the geographic relevancy of 1,000 questions, we divided the geographic scope into three levels: POI, district, and city. If a user asked about a POI or places near a POI, then we checked the map to determine whether this was correctly pinned (if it was within 500m from the stated POI). Otherwise, we checked whether it belonged to the correct city. If both cases failed, then we classified it as incorrectly marked. Similarly, if a user asked about a district or other places near the district, this question belonged to the district level, and thus, we used the same criteria for classification. Note that route questions (A to B) were excluded due to their relative inconsistency in representation, because, for transportation, questions are primarily in the form of a source to a destination.

4.2.2 Question Topic/Type Categorization.

Question topics: We performed content analyses in order to characterize the topical patterns (overall distributions and geographic/typological differences). We used the following topic categories from Naver KiN: information and communication technology (ICT), games, entertainment/art (e.g., TV, radio, movies), economy (e.g., banking, tax, real estate), shopping, society (e.g., laws, politics, culture, governments), health, lifestyle (e.g., transportation, food, cars), travel, sports, and education. Unlike traditional social Q&A, NKH does not have a topic field, because questions are categorized by location. We automatically classified the topic categories as follows: for a given question, we extracted key words using a Korean parser called Kokoma Korean Morpheme Analysis (KKMA).[10] Then, we searched the extracted keywords using Naver KiN in which the question askers manually select the relevant topics when posting questions. From the top 100 search results, the most frequent topic category was selected as the topic category for the question. To confirm whether the automatic classification provided accurate results, we compared the automatically selected topics with the manually coded topics for the 1,000 questions and measured the interrater agreement between manual and automatic classifications using Cohens Kappa statistic. The measured value for the topic classification was $k = 0.87$, which indicates substantial agreement.

Question types: In addition to studying at question topics, we classified the types of the 1,000 questions to understand typological patterns and relationships between topics and types of questions in location-based social Q&A. To this end, the questions were classified using existing categories, such as information, suggestion, opinion, request, and monologue [26]. Information questions are used to find specific facts (e.g., “*Are there any games in this stadium today?*”); suggestion questions are used to seek recommendations or advice (e.g., “*What's a good hospital in Jinju City that specializes in temporomandibular joint (TMJ) problems?*”); opinion questions are

[10]http://kkma.snu.ac.kr.

used to survey other peoples thoughts or preferences (e.g., "*I will go to Jeju Island tomorrow. Do you think the snow will melt before then?*"); request questions are for tasks, resources, or services (e.g., "*Please park (your car) in the apartment complex*"); monologue questions are for a single person speaking alone, yet actually asking nothing (e.g., "*People in Seoul have poker faces*").

For typological categorization, two authors manually coded 200 questions from the 1,000 questions together, and then separately coded the remaining 800 questions (i.e., 400 questions each). These common questions were used to measure the inter-rater agreement using Cohens Kappa statistic. The measured value for the type classification was $k = 0.84$, which indicate substantial agreement.

4.2.3 Spatial Locality Analysis. To better understand spatial locality of user activities, we conducted two analyses: i.e., an *administrative division* based spatial locality analysis, and a *clustering* based spatial locality analysis. The administrative division level analysis allows us to intuitively understand the overall topical and typological patterns and easily compare their differences across various divisions. This analysis, however, does not reveal geographic locality across divisions; for example, a user may have answered for a few districts, but the division-level analysis does not indicate whether these districts are nearby. To address this limitation, we additionally conducted a cluster based spatial locality analysis by clustering the geographic coordinates of a users activities. This analysis allows us to answer the following questions: "Does a user answer mostly for his/her district and nearby districts?" and "How many spatial clusters are there, and what are the actual sizes of such spatial clusters?"

Entropy analysis: We used an entropy measure to capture the degree of topical or regional focus by an asker/answerer. For regional focus, we considered two region types: i.e., the administrative divisions (i.e., district/city-level) and any uncovered spatial clusters.

Considering a user i who made p_k percentage of questions/answers for a topic/region k, the users' entropy is given as $-\sum_k p_k \log_2 p_k$. The entropy measure has the following property: the lower the entropy, the higher the level of a users focus on certain topics/regions. For example, let us consider regional entropy values. If a user only answered for a single district, e.g., district j (i.e., $p_j = 1$ and $p_i = 0$ for all i other than j), then the entropy value is zero. If a user intensively answered in one district, then the entropy value is close to zero even if the user occasionally answered in other districts. The entropy is maximized when any region is equally likely to be asked about by a user (i.e., a uniform distribution). We considered users who asked/answered more than ten questions/answers for the entropy calculation; this resulted in 3,287 answerers (top 2.33%) and 1,387 askers (top 1.54%).

Spatial clustering analysis: To cluster each asker/answerers activities based on their geographic coordinates (i.e., GPS points in latitude and longitude), regardless of the administrative divisions, we ran a density-based spatial clustering algorithm (called DBSCAN) [13], which requires two parameters, *eps* (epsilon in distance units) and the minimum number of points (*minPts*) to form a cluster. The algorithm begins with an arbitrary starting point that has not been visited previously. Other points within the *eps* neighborhood are then retrieved; and if the number of point(s) is above *minPts*, then these point(s) form a cluster and the cluster expands by repeatedly adding the other points, which are within the *eps* of the cluster. Otherwise, the point is disregarded as noise.

We implemented the DBSCAN algorithm using Python programming language. We then set *eps* to 10km and *minPts* to 1, 5, and 10 to see the users local knowledge distributions based on the size and number of clusters. Only one revision from the DBSCAN algorithm was used to calculate the distance of two points. Because we used geographic coordinates (latitude and longitude) where the questions and answers were posted as points for the DBSCAN algorithm, we used the great-circle distance,[11] which is the shortest distance between two latitude and longitude points on the surface of a sphere, for the calculation of *eps* instead of the Euclidean distance, because there are no straight lines on a sphere.

For the spatial cluster analysis, we only considered heavy users whose number of answers/ questions was equal to or greater than 30 due to the requirements of the minimum points; this resulted in 1,492 answerers (top 1%) and 298 askers (top 0.5%).

Contribution distance analysis: In addition, we quantified the localness of knowledge sharing in location-based social Q&A using the mean contribution distance (MCD), as in Hecht and Gergle [16], to understand the scope of the local knowledge of answerers. For given primary clusters (e.g., the cluster with the largest number of answers) of answerers, we calculated the MCD value, defined as $\sum_{i=1}^{n} d(C,\ c_i)/n$, where C is the centroid of the cluster, c_i is the location of an answered question, $d(C,\ c_i)$ is the great-circle distance between two latitude and longitude points of questions (C and c_i), and n is the total number of answers in a cluster. For the contribution distance analysis, we considered the same dataset used for spatial clustering analysis.

5 GENERAL USER BEHAVIOR

One of the key distinctions of location-based social Q&A is the manner in which questions are categorized: that is, in location-based social Q&A, questions are categorized based on region, whereas in conventional social Q&A services they are categorized based on topics, such as computers and life. Because of this structural difference, there may be significant dissimilarities among users behavioral patterns, such as in users asking and answering behaviors. Thus, we were curious about how dissimilarity leads to different user behaviors. If user behavior is different from behavior in conventional social Q&A, then it may be possible to find unique design implications to leverage user participation.

In this section, first, we present the distributions of Q&A activities and compare them with those of conventional social Q&A. Second, we examine answering patterns (e.g., number of answers per question) and answer delays. Third, as interactive map usage is crucial in location-based social Q&A, we examine users map usage patterns by analyzing questions texts and location specifications on the map.

5.1 Question and Answer Distribution

As this research is the first large-scale study of location-based social Q&A, we first compared the number of questions per asker and the number of answers per answerer to check whether there is a considerable difference in terms of distribution. Earlier work [30] revealed that most users had very little answering activity, but a few users had higher activity on StackOverflow. Likewise, Paul et al. [37] documented that the activities of users in Quora followed a power law distribution. As shown in Figure 3, we found that distribution of the number of answers/questions per asker/answerer also follows the power law distribution in location-based Q&A. This means that heavy users roles are as critical as those in conventional social Q&As.

5.2 Answer Rate and Delay Characteristics

In Figure 4, we plot the cumulative distribution of the number of answers received per question. The results indicate that the mean value was 1.16 (SD: 1.43) and, surprisingly, 33.0% of the questions did not received any answer. This rate appears to be much higher than that of other conventional Q&A sites such as Quara (20%) and Stackoverflow (10%) [46]. One possible reason is that answering location-based questions requires local knowledge such as living experience, which is not easily

[11] https://en.wikipedia.org/wiki/Great-circle_distance.

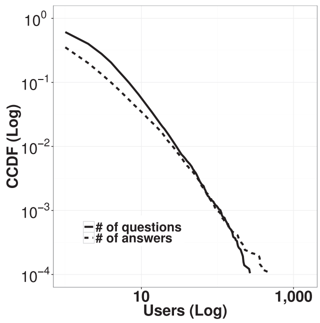


Fig. 3. Distribution of the number of answers and questions per user.

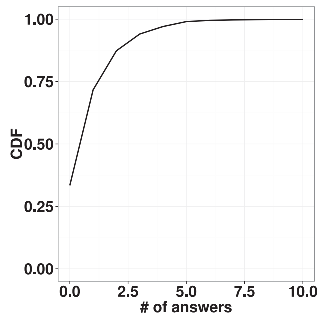


Fig. 4. The number of answers per question.

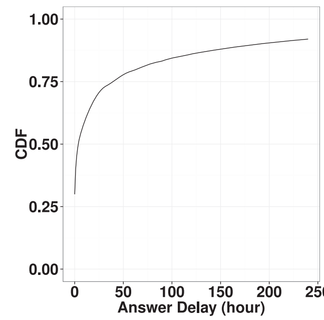


Fig. 5. Answer delay.

acquired or may be difficult to find on the web; as shown later in our spatial locality analyses, answerers mostly answer questions posted only within a few kilometers centered around home and work places. Furthermore, as questions are categorized by many regions (i.e., districts/cities), there may be regions with fewer answerers, which could lead to low answering rates.

To deepen our understanding of how answers were received over time, we evaluated the answer delay distribution as depicted in Figure 5. We found that 50% of all questions received the first answer within 4h. 70.0% of questions received answers within 24h (a day), and 81.2% of questions received answer within 72h (three days). Overall, the answer delay was longer than in other Q&A sites. For example, 50% of all questions in StackOverflow and Live QnA received the first answer within 12min and 3h, respectively [19, 30]. To understand why some questions received answers after some time, we manually investigated the questions that received answers after a month (i.e., top 1% of the questions in terms of answer delay). The manual investigation revealed that many of those questions were answered by local experts and business promoters. For example, one user answered not only recently posted questions but also unanswered, old questions in a city of interest. Another example is business promoters whose nicknames are company names, and they answer local questions, hoping to advertise their companies in a local area. More detailed spatial activity patterns will be presented in Section 7.

5.3 Interactive Map Usage: Geographic Relevancy

In a location-based social Q&A, questions are categorized by regions. However, since the category is only given as district/city-level, the map interface helps users pinpoint a specific place: for example, in the case when asking about a specific POI. Map usage can help the user easily browse questions around his or her location. Church et al. [7] compared effectiveness of map-based interface and text-based interface in a location-based mobile Q&A service; they showed that people choose map-based interfaces over text-based interfaces to ask a question in a location-based mobile Q&A service. They also showed that a map-based interface is effective in helping people to understand physical proximity/distance, and where the question originates.

As shown in the Table 2, we found that a mere 13.5% of POIs were correctly pinned at the POI level, and 16.0% of the questions pinpointing specific POIs were wrongly pinned. However, for the district level questions, users mostly specified the correct district (87.7%), and such usage behavior occurred with the city level questions as well. Most of the POIs (70.5%) were correctly marked at the district and city levels. According to our survey results on how participants selected positions, there are several explanations as to why people prefer to specify less precise locations in POI level

Table 2. Map Relevancy

Geographic Scope	Map Relevancy	
POI (275)	Correct POI	13.5% (37)
	Correct District/City	70.5% (194)
	Wrongly pinned	16.0% (44)
District (179)	Correct District	87.7% (157)
	Correct City	4.5% (8)
	Wrongly pinned	7.8% (14)
City (195)	Correct City	93.8% (183)
	Wrong City	6.2% (12)

questions. First, it requires considerable effort to interact with the map to find the exact reference point using a mobile device. One participant stated, *“I first click the current location and then zoom in to move around to find the location. If I cannot find the location that I want, then I just select a public place like a town hall.”* In addition, users may feel that they do not need to specify exact locations as they assume that local residents probably already know these, for example, by saying, *“I click approximate location. People are around the location so I don’t need to find exact location.”*

6 TOPICAL CHARACTERISTICS IN KNOWLEDGE SHARING

Location-based social Q&A service is for asking and answering questions about local knowledge of a specific region. The questions are classified by regions instead of topics. Through investigating the types of questions asked and answered in location-based social Q&A, we can examine possible topic-based design implications to leverage location-based social Q&A systems, such as topic-based question routing to expert users.

In location-based social Q&A, questions are classified by regions instead of topics. In this section, we investigate topical characteristics in location-based knowledge sharing. First, we analyze the overall topical characteristics (e.g., topic distribution and topical focus of users). Second, we study the topical characteristics across different regions to see whether geographic characteristics are reflected in questions for a region. Third, we investigate topic distributions across different types to reveal the relationships between topics and types in location-based questions.

Topic distributions: The topic distributions of location-based (Naver KiN “Here”) and conventional Q&A (Naver KiN) are presented in Figure 6. Significant differences in the rates of questions asked were found in the ICT, games, lifestyle, and travel topic categories. The lifestyle (23.7%) and travel (11.7%) categories were dominant in the location-based social Q&A. For example, people asked many questions about transportation (e.g., how to go from point A to point B) in the lifestyle category and tourism questions (e.g., where beautiful sight-seeing locations are, which restaurants are good) in the travel category. However, these two topic categories were not popular in conventional Q&A (lifestyle accounted for 8.6% of questions, travel accounted for 1.9% of questions).

However, the ICT (2.5%) and games (1.5%) categories had minimal representation in location-based social Q&A. For example, people asked ICT-related questions (e.g., where the nearest computer repair shop is) and game-related questions (e.g., what computer game trading is occurring in the user’s vicinity), whereas those types of questions made up 13.9% and 10.9% in conventional social Q&A as users asked questions about topics related to their interest (e.g., computer problems or online game strategies).

In contrast to the categories that were significantly more popular in one type of Q&A or the other, the education topic was similarly popular in both location-based social Q&A (17.6%)

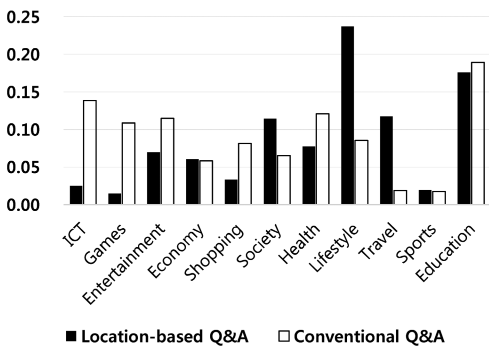


Fig. 6. Topic distribution.

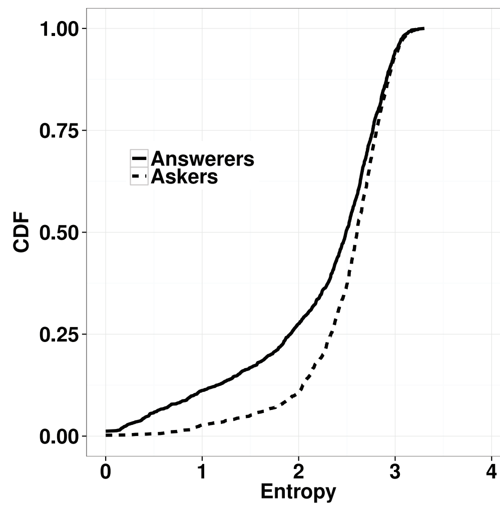


Fig. 7. Topic distribution of users.

and conventional social Q&A (18.9%). We manually examined why this particular category was popular in both types of Q&A. In location-based social Q&A, questions usually referred to a specific school (e.g., admission consulting), whereas there was a greater variety of educational questions asked in conventional social Q&A (e.g., mathematical problems, English proofreading). Similarly, the economy topic was also similarly popular in both location-based social Q&A (6.1%) and conventional social Q&A (5.8%). Economy questions in location-based social Q&A often regarded real estate (e.g., real estate market prices, real estate agency information), whereas questions in conventional Q&A were more general (e.g., taxes, finance, economic policy). A third pattern emerged with regard to the sports category, which was rare in both types of Q&A. In the location-based social Q&A, 2.0% of questions were about a sports center or a sports club, whereas in conventional Q&A they were mainly general sports questions (e.g., sport skills and tips).

Our results confirm that there is a significant difference in the topic distributions between location-based social Q&A sites and conventional social Q&A sites. Although the distribution of some topics (e.g., education, sports, and economy) was similar between the two types of services, the content of the questions themselves was quite different, and the location-based questions were primarily related to local businesses/services.

Topical focus of users: To understand how users are topically focused, we evaluated users topic entropy as shown in Figure 7. Askers' topic entropy value was 2.6 (SD: 0.52) on average, and answerers' topic entropy value was 2.4 (SD: 0.77) on average. This range of entropy values indicates that users are likely to focus on four to six topics when uniform topical distribution is assumed. The mean values of askers and answerers look quite similar, but there is a clear distribution difference. Only 7.36% of askers have a topic entropy value lower than 2, whereas 27.5% of answerers have a topic entropy value lower than 2. This means that askers are less topically focused than answerers.

6.1 Topical Characteristics of Regions

Many studies have aimed to discover the geographic topic distribution in location-based social networks. For example, Yin et al. [51] showed that different topics distributions varied between across different locations by using text description and location of a geo-tagged photo dataset from Flickr. Wang et al. [48] confirmed that user interests are different across cities in the US by

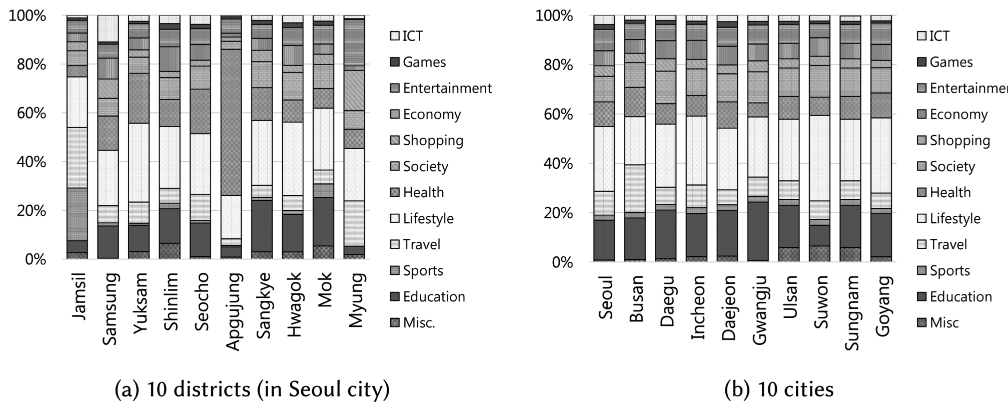


Fig. 8. Topic distribution of most frequently asked 10 districts/cities.

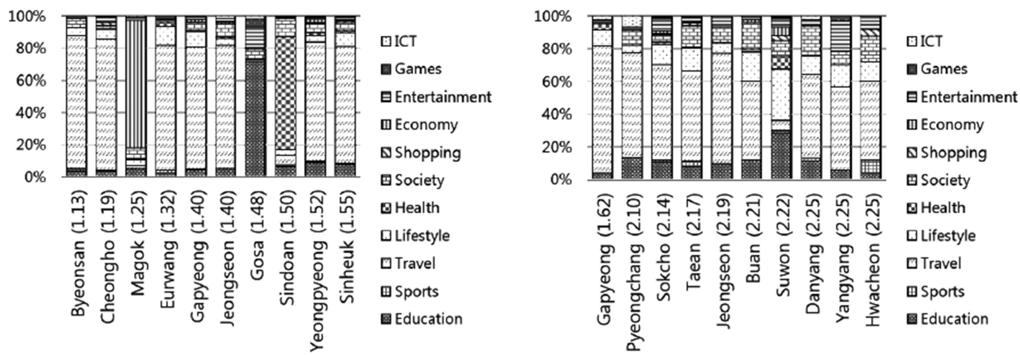
using a check-in dataset from Foursquare. Yin et al. [50] showed that top event types are different across cities in China by using a dataset from DoubanEvent,[12] which is the largest Chinese local event-based social network.

Our work extended these studies in that we also investigated topical patterns across different geographic regions in location-based social Q&A. We investigated whether the characteristics of a region were reflected in the location-based social Q&A service. Identifying such characteristics has been beneficial when building geographic information retrieval systems. We found that, in general, the geographic characteristics were well reflected and that some patterns of topical locality existed. In addition, the topical distributions were largely dependent on the size and functional complexity of the region.

To understand how regional characteristics are reflected in Q&A usage, we first analyzed the topic distributions of the top 10 most frequently asked for districts in Seoul, which is the capital city of South Korea (Figure 8(a)). Overall, the regional characteristics were well reflected in the district level questions. For example, the Jamsil district, which is well known for its sports stadium, had a large percentage of sports questions (22.2%). The Apgujeong district had a high percentage of health questions (60.4%), because it is famous as a medical area that is densely populated by plastic surgery clinics. Therefore, we hypothesized that such observations were partly related to the concept of zoning in urban planning; zoning is a method of urban planning that prevents new developments from interfering with the existing residents or businesses and preserves the "character" of a community [2]. This type of planning is widely adopted in most developed nations, and local municipalities in Korea abide by the national zoning guidelines.

Furthermore, we examined whether there were significant differences for the topical distributions in different geographic scales. In Figure 8(b), we plotted the topic distributions of the top 10 cities ranked in terms of the number of questions. Unlike the questions for districts, there were only minor variations in categories across different cities. Note that there were a few cities that had significant tourist attractions and therefore have distinctive topic distributions with higher percentages of shopping, entertainment, and travel categories. For example, Jeju Island, which was designated as a UNESCO World Heritage location in 2007, had a significantly high number of questions/answers related to travel.

[12] http://www.douban.com/location.



(a) 10 districts (b) 10 cities

Fig. 9. Topic distribution of most topically focused 10 districts/cities.

Topically focused districts and cities: We also investigated the top 10 most topically focused districts and cities. As shown in Figure 9, most districts and cities were topically focused on travel, as they include famous tourist attractions. To examine how/what types of questions are asked in the regions, we manually investigated the questions in the districts and cities. We found that many similar questions were repeatedly posted. For example, in districts and cities with a focus on the travel category (e.g., Jeju Island), we found many questions asking for recommendations on good accommodations and good restaurants. Another example is the case of Ma-gok district, which was designated as a residential development district; many people enthusiastically asked housing-related questions.

6.2 Topical and Typological Patterns

Regarding type classification, we found that information and suggestion question types were dominant, and they comprised 67.7% and 17.5% of all questions, respectively, followed by opinion questions (10.0%), request questions (2.8%), and monologue questions (2.0%). This type distribution differs significantly from other social Q&A services [22, 26, 32]. Kim et al. [22] showed that in Yahoo! Answers, opinion is the most popular question type (39%), followed by information (35%). Morris et al. [32] revealed that the most popular question types are recommendation and opinion in social networks such as Facebook and Twitter. In addition, in mobile SMS social Q&A, information questions comprised 51.1% of all questions, followed by suggestion (23.4%) and opinion (18.8%) questions [26]. Thus, location-based social Q&A had a significantly higher percentage of information questions and a lower percentage of opinion questions compared with conventional social Q&A. Considering the high percentage of factual questions, it would be beneficial to archive them for local searches.

As shown in Table 3, the type distribution across different topics varies significantly. For instance, information questions are more pronounced in life (76.2%) (e.g., *"Way to go to Goo-ro digital complex station to Nam Choon cheon Hak Gok Li. What should I take? How long does it take? How much is it?"*), shopping (79.2%) (e.g., *"Where can I buy clothes near Shinwol-dong?"*), entertainment (79.5%) (e.g., *"Please tell me about the location of Excalibur internet cafe in Daegu-si Sooseong-gu Joong-dong, or nearest buildings"*), and society (70.1%) (e.g., *"How can I volunteer at Incheon Soobong library?"*). Suggestion questions are more notable in health (40.3%) (e.g., *"I want to infuse fillers in my nose and to have surgical jaw muscle reduction. Can you tell me which*

Table 3. Topic and Typological Distribution of Randomly Selected 1,000 Questions

	Type	ICT	Game	Ent.	Eco	Shopping	Society	Health	Life	Travel	Sport	Edu.	Misc.
Information	67.7%	65.6%	55.6%	54.1%	66.7%	79.2%	70.1%	46.8%	76.2%	67.9%	67.9%	69.7%	67.7%
Suggestion	17.5%	9.4%	22.2%	17.6%	8.3%	12.5%	14.0%	40.3%	12.1%	27.1%	17.9%	12.6%	22.6%
Opinion	10.0%	18.8%	11.1%	21.6%	16.7%	4.2%	10.3%	11.7%	6.7%	3.4%	3.6%	12.6%	8.1%
Request	2.8%	6.3%	0.0%	2.7%	4.2%	4.2%	2.8%	0.0%	3.3%	0.0%	10.7%	3.4%	1.6%
Monologue	2.0%	0.0%	11.1%	4.1%	4.2%	0.0%	2.8%	1.3%	1.7%	0.8%	0.0%	1.7%	0.0%
Count	1,000	32	9	74	48	24	107	77	239	125	28	175	62

place offers reasonable prices? I want to know the price. No ads please."), travel (26.4%) (e.g., *"Please recommend valleys in Chooncheon or Gwangyang, since I am a student I dont have enough money to book a mountain cabin."*). Opinion questions are remarkable in entertainment (21.6%) (e.g., *"How is the atmosphere of Club Harlem?"*) and computer (18.8%) (e.g., *"Is WiBro fast enough for web surfing and video streaming in an apartment in Amsa district?"*). Request and monologue questions were rarely asked in location-based social Q&A.

7 SPATIAL LOCALITY OF USER ACTIVITIES

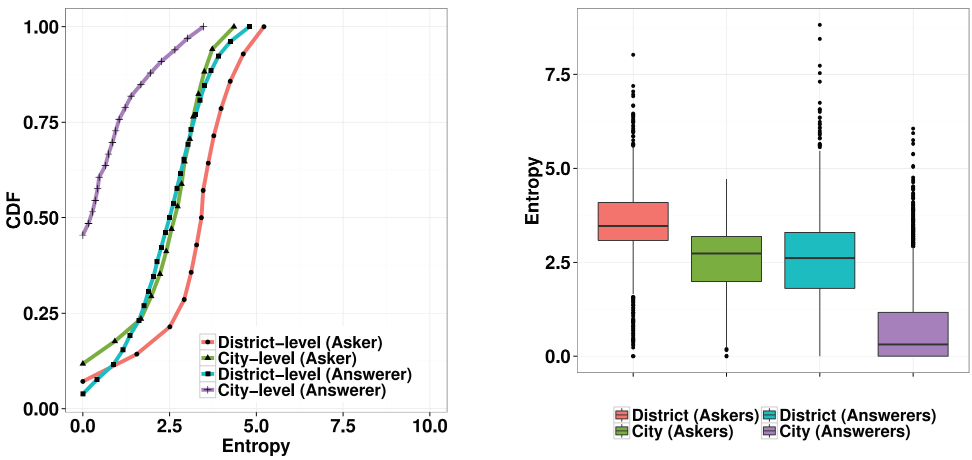
Because users can contribute to various regions, they tend to have their own spatial locality based on their asking and answering activities. For example, some users answer only for one region, whereas others answer for more than 10 regions. Also, the regions may be adjacent or scattered. By understanding spatial locality, we can determine if there are local experts in the location-based social Q&A services as there are topical experts in conventional social Q&A services. We then can provide possible design implications to improve location-based social Q&A services, by routing questions to the relevant local experts.

To understand spatial locality, we first analyzed district/city-level geographical focus of both askers and answerers. We also performed fine-grained analysis of spatial locality by applying spatial clustering algorithms to deepen the spatial locality of user activities (e.g., size of clusters). We then investigated the answerers regional selection patterns, by (1) examining how their regions of interest were related to the locus of their life and experience, (2) analyzing temporal variation of answerers regional selection, and (3) identifying the characteristics of those answerers who did not show strong spatial locality.

7.1 Spatial Locality Analysis

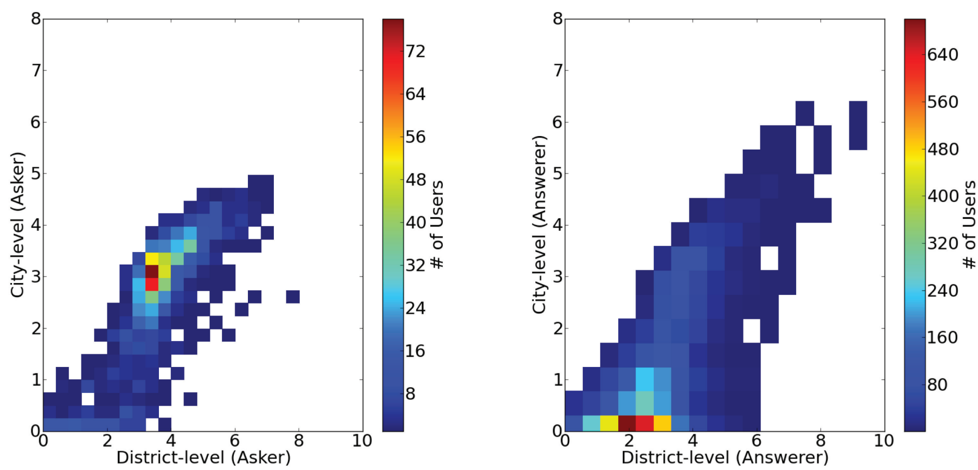
District/city-level spatial locality analysis: We analyzed the dataset to understand the geographic focus of the users' activities (asking/answering). In Figure 10, we present the cumulative distributions and corresponding boxplots in parallel. The results demonstrated that, in general, the asking activities had higher entropy values than the answering activities: asking at the district/city levels had means of 3.5/2.5 and answering at the district/city levels had means of 2.6/0.8.

We then analyzed the geographic containment (for example, answering in many districts but the same city) using each users district- and city-level entropy values. In Figure 11, we arranged a pair of entropy values (district, city) and drew a heat map for both district- and city-level activities to evaluate user distributions. As shown in In Figure 11(a), for the asking activity, many users were spread primarily on a city-level entropy value of 3.0, which represents eight cities (if an asker posted questions equally across cities) and district-level entropy value of 4.0, which represents 16 cities (if an asker posted questions equally across districts). This means that many askers posted questions on various cities and districts. However, as shown in Figure 11(b), answerers were



(a) CDF (b) Boxplot

Fig. 10. Spatial entropy of askers/answerers.



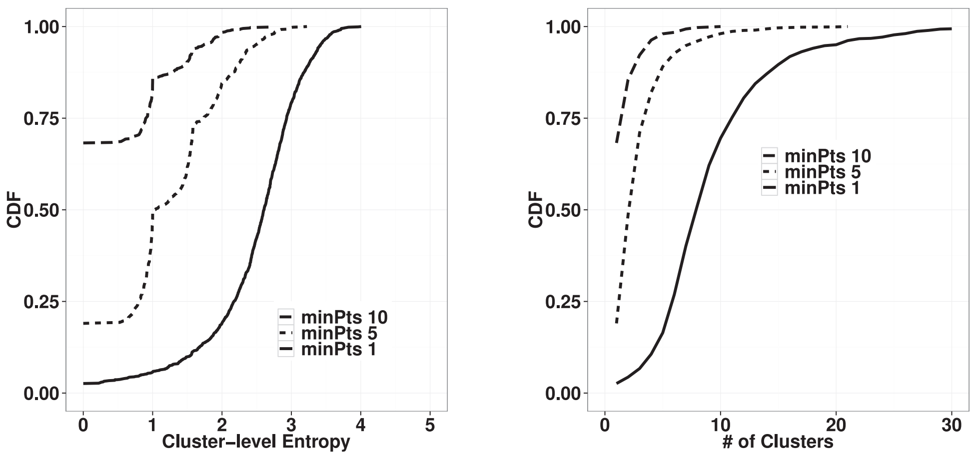
(a) Asker (b) Answerer

Fig. 11. Heat map of entropy.

spread primarily on a city-level entropy value of 0 and district-level entropy value of 2, which means that answerers mostly answered in several districts but in one city of the districts.

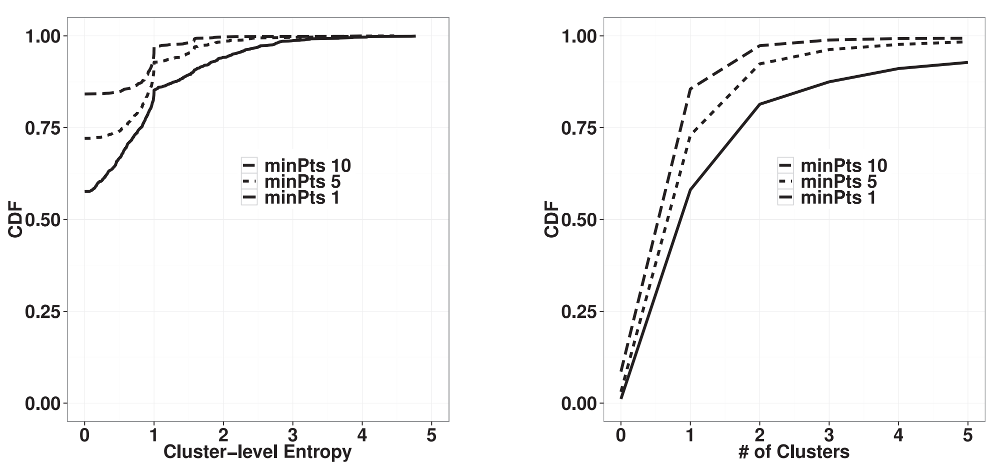
Coordinate based spatial locality analysis: Because the size of administrative divisions in South Korea is heterogeneous, and user activities typically span multiple districts, our spatial locality analysis may not correspond with the true geographic scope of users activities. Therefore, we additionally conducted spatial cluster analysis based on the geographic coordinates (i.e., latitude and longitude) of a user's Q&A data regardless of the administrative divisions.

Figures 12 and 13 present distributions of cluster-level entropy values and the number of clusters with different *minPts* thresholds of askers/answerers. We found that answerers focused on a very few regions for their answering activities. The majority of answerers had a single cluster; in



(a) Cluster-level entropy (b) The number of clusters

Fig. 12. Askers’ spatial analysis.



(a) Cluster-level entropy (b) The Number of clusters

Fig. 13. Answerers’ spatial analysis.

particular when *minPts* = 1, more than 50% of answerers had a single cluster, and more than 75% of users had less than two clusters. The mean values of *minPts* = 1, 5, and 10 were 2.3, 1.5, and 1.2, respectively. The maximum values of the *minPts* = 1, 5, and 10 were 101, 61, and 40, respectively.

In contrast, askers widely post questions to various regions. A majority of askers had more than nine clusters; in particular when *minPts* = 1, 50% of askers had more than 8 clusters. The mean values of the *minPts* = 1, 5, and 10 were 9.47, 3.17, and 1.63, respectively. The maximum values of the *minPts* = 1, 5, and 10 were 44, 21, and 10, respectively. We found that askers mean values of the number of clusters were larger than answerers, whereas their maximum values of the number of clusters were lower than answerers. We believe there are two reasons. First, the number of answers of heavy answerers was much larger than the number of questions of heavy askers. The second reason is that some heavy answerers answered regardless of their local knowledge.

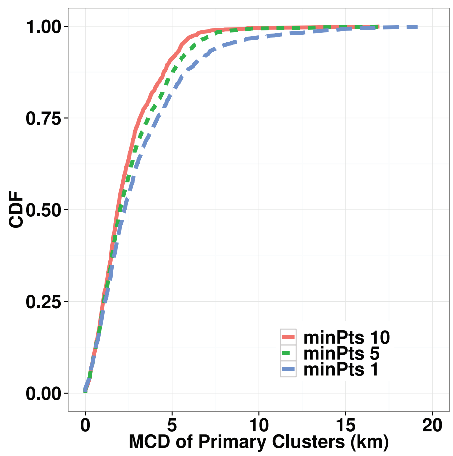


Fig. 14. MCD of heavy answerers.

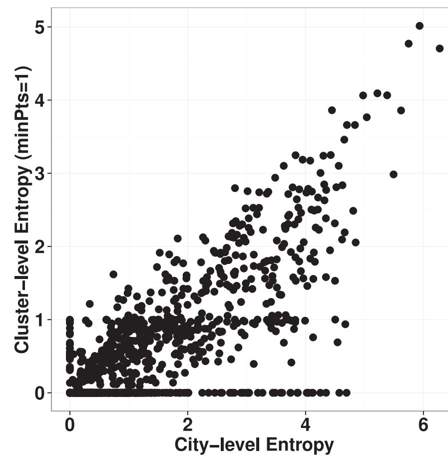


Fig. 15. City-level entropy and cluster-level entropy of answerers (minPts = 1).

To investigate the scope of local knowledge, we measured the mean contribution distance (MCD) of each cluster of answerers. Here, the MCD value represents the size of the cluster in which the user primarily answered. In Figure 14, we plot the CDF of the MCD values from the users' primary clusters (i.e., the cluster with the largest number of answers). The mean MCD values for *minPts* = 1, 5, and 10 were 3.1, 2.6, and 2.3 km, respectively. This indicates that the primary clusters MCD values covered a few nearby districts, e.g., home and nearby home, and work and nearby work.

Because the number of clusters was very small, a majority of users had a zero entropy value. Then, we compared the city-level entropy and cluster-level entropy values; Figure 15 presents the scatter plots with *minPts* = 1. The figure shows the relationship between city-level spatial locality and cluster-level spatial locality. A Pearson correlation coefficient was computed to assess the relationship between the two variables. We found that there was a positive correlation between the two variables, $r = 0.783$ ($p < 0.001$). The cluster-level entropy is smaller than the city-level entropy, because the DBSCAN removed the noisy clusters that had points less than or equal to the *minPts* and adjacent cities/districts can be clustered together. Interestingly, those answerers whose cluster entropy values were zero, but had high city-level entropy values, were mostly business promoters. In contrast, the number of clusters of askers was larger than the number of clusters of answerers. (*minPts*: 1; Mean: 2.5, SD: 0.65). This means that askers post questions to various regions (the result of city/district-level spatial locality analysis).

7.2 Answerers Regional Selection Patterns

Regional selection characteristics: To understand where users focused their activities in detail, we asked the survey participants to report (1) the number of their selected regions of interests in the app, (2) a list of those names as well as the reason for each choice, and (3) how many questions they could answer in the regions of interests. The number of selected regions of interests is presented in Figure 16. The mean number of districts was 2.9 (SD: 3.2) and the maximum number was 35. Regarding the second part, there were 142 valid answers with a detailed list of interest regions and the reasons for their choices, and this led to 317 annotated regions of interest. The major categories of areas included home, work/school, and downtown areas. The manual classification results indicated the following: home (93.7%), nearby home (16.2%), previous home

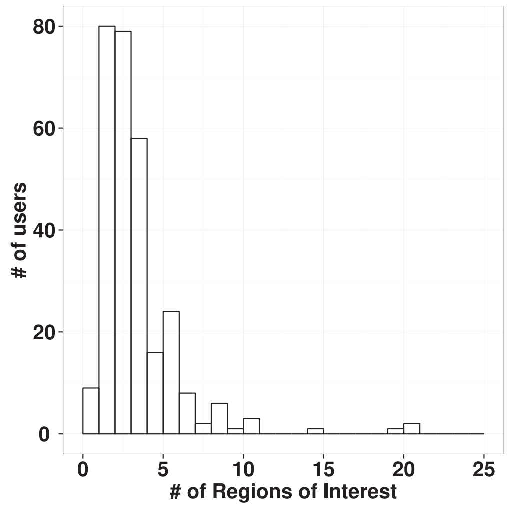


Fig. 16. Number of regions of interest per user.

(14.8%), school (23.9%), nearby school (0.7%), previous school (0.7%), work (28.9%), nearby work (2.1%), previous work (3.5%), and downtown (24.6%). When respondents had their home and work in the same region, they were counted twice. Miscellaneous regions of interest (25 regions) included churches, hobby places, parents houses, relatives houses, and tutoring institutes. The results aided in understanding the relationship between a users local connection and knowledge. However, despite local familiarity (home, work, school), the survey participants reported that the perceived percentage of answerable questions was 37% on average (SD: 24%). As shown later, many of the questions required very specific knowledge based on local experiences, e.g., *"In Daejeon, are there any places that I can buy big dumplings after midnight?"* and *"In Wonju, please let me know where I can buy less expensive medium and large size vases."*

Temporal variation of regional selection: Users geographical answering pattern might change over time as their local familiarity changes (e.g., getting a job in another city). In Figure 17, we plotted the city-level entropy changes of the top 10 answerers. For a series of answers by a user, we calculated the entropy in each block of 50 consecutive questions. Then, we plotted the magnitude of the entropy differences between two consecutive blocks. The figures clearly demonstrate that changes in the entropy values were quite small. This indicates that the city selection strategies did not significantly change over time: those who had low/high entropy values would continue to have low/high values.

Answerers with weak spatial locality: While users were mostly geographically focused, there remained some users whose activities were geographically scattered. We divided the heavy users into two groups based on city-level entropy values. If the city-level entropy value was greater than 2.0 (i.e., typically more than four cities), then we assumed that the users activities were geographically scattered (GS group = 19% of the heavy users); otherwise, we assumed that the users activities were geographically focused (GF group). In Figure 18, we present a scatter plot: each dot represents a users city-level entropy and topic entropy.

We manually investigated the answerers in the GS group to understand what types of local knowledge they provide. We then further took a look at those with high and low topic entropy values. First, those who had high topic entropy values were mostly top-ranked answerers in Naver KiN. Our manual investigations revealed two types of answerers: factual information

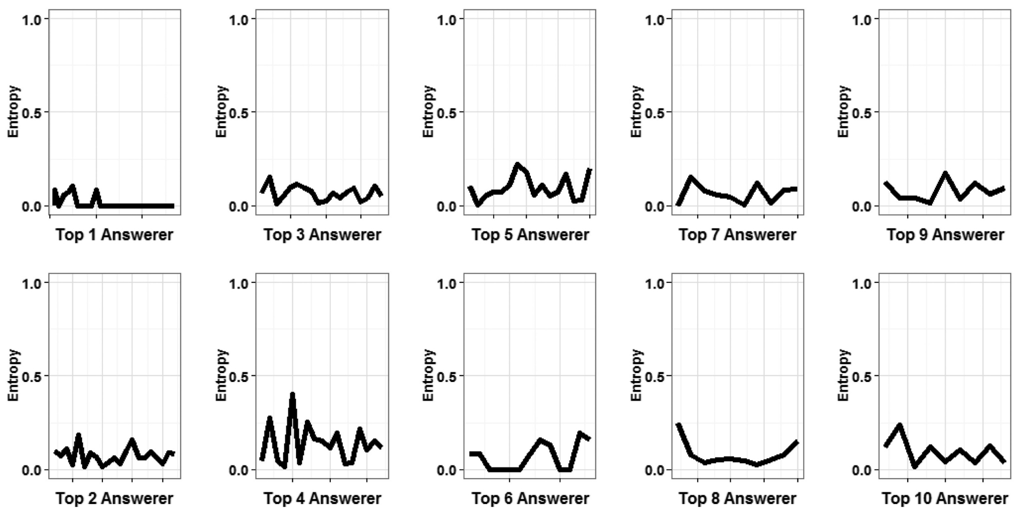


Fig. 17. City-level entropy changes of top 10 answerers.

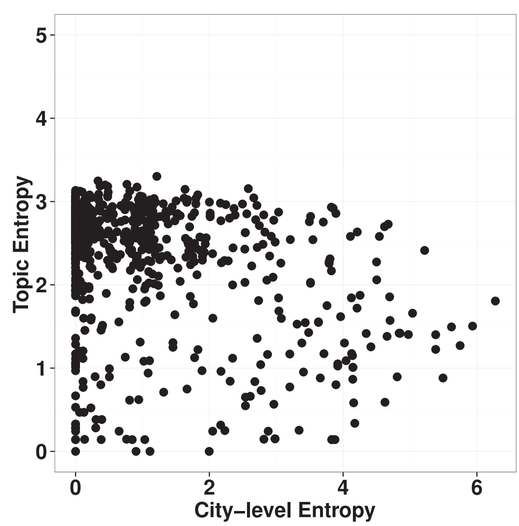


Fig. 18. City-level entropy vs. topic entropy of the answerers.

experts and province-level local experts. The factual information experts typically answered local questions that were easily searchable on the web (mostly factual) such as transportation costs, traffic status, and local facilities. The province-level experts were actively contributing to a number of cities within a province; however, detailed local answers were more skewed to a few familiar cities. Second, those who had low topic entropy values mostly promoted their business (e.g., clinics, lawyers, online shopping malls) or online communities (e.g., volunteering). It appears that these users tended to copy and paste similar, but lengthy, general answers about specific topics without any local relevance. For example, a medical doctor mostly answered questions on recommending local clinics that offer a specific cosmetic surgery, but the doctor provided only general information about that surgery regardless of askers location of interest.

8 ANSWER MOTIVATIONS

By understanding the unique motivations of participants in location-based social Q&A, we can identify opportunities to encourage participation. To understand the answer motivation in NKH, we asked the participants to write in detail why they answered questions on the online survey using an open-ended question. From the answers, we extracted all motivators and performed affinity diagramming. The following major themes were derived: knowledge exchange (24.9%), altruism (18.2%), ownership of local knowledge (10.1%), points (9.8%), pastime (9.2%), and sense of community (7.0%). Miscellaneous themes included business promotion and learning. We found some common motivators as in previous studies [32, 33, 38]. However, unlike the existing results on social Q&A that reported intrinsic (altruism, enjoyment), external (points), and social (knowledge exchange) motivators, we found two unique motivators in the location-based social Q&A: ownership of local knowledge (competence about local knowledge learned over many years) and sense of community (feeling of belonging and serving the information needs of other community members).

8.1 Ownership of Local Knowledge

Researchers have demonstrated that when people believe they own knowledge, they are more likely to contribute to knowledge sharing [47]. This result can be attributed to individuals' internal satisfaction derived from sharing their knowledge with others. As they have lived in that location for a long time, we expect that the local residents gained a strong felling of ownership of local knowledge, ranging from specific goods/services to distinctions between places and their popularity [29]. In our survey, the ownership of local knowledge was clearly noted. One user stated, *"Because I know everything about my town as I have lived in my town for a long time"*; another user stated his experience by saying, *"I have been living here for 20 years and went to schools in this region. With this knowledge, I'm sure I can help answer other peoples questions. Thats why I started answering here."*

8.2 Sense of Community

We found that another key motivator was the sense of community. McMillan and Chavis [31] defined the sense of community as a feeling that members have of belonging, a feeling that members matter to one another and to the group, and a shared faith that members needs will be met through their commitment to be together. They also proposed four elements for the sense of community, which are membership, influence, integration and fulfillment of needs, and shared emotional connection. In our survey, users responses about sense of community were relevant to membership and fulfillment of needs among the four elements. Note that the sense of community in location-based Q&A differs significantly from that in social network services (and the Q&A therein), because the interpersonal relationships among answerers are weak in the latter.

Membership: Membership is a feeling of belonging and of becoming a member. We found that many users stated, "my regions," which is a good indicator that they regard themselves as members of their local community. Furthermore, for those who have a strong membership attachment, they are eager to invest their time in the local community. Membership becomes more meaningful by personal investment of members in a community [31]. There were some users who were concerned about lack of answers in their regions or the reputation of their regions. One user stated, *"I have been answering questions in the Laws section. But after I noticed that the areas that I live in and know well did not receive answers, I decided to answer questions. I subscribed to question delivery from those areas and started answering questions."* Another user commented, *"I think kind and sincere answering is one of the representative images of the area, and I want to build a good image of my region."*

Fulfillment of needs: Our participants concurred that they wanted to serve the information needs within their communities (for both members and visitors). A user expressed their feeling as follows: "*Knowing that I can help other local community members makes me feel really great.*" In addition, responsibility as a community member was expressed: "*I was born here, and I know [this region] very well. I feel fresh, because people from other areas often ask questions. I feel like I should take care of them just like taking care of a baby.*" Some users fulfilled needs with the expectation of receiving useful help in return later, e.g., "*By exchanging questions, I can receive help from other people when I visit other areas; I can also give help to other people when they visit ours.*"

In summary, we found two unique answering motivators, ownership of local knowledge and sense of community. This is because questions were categorized by regions as oppose to topics in location-based social Q&A. This key difference exposed the unique motivators. As mentioned in Section 7.2, the results of the survey showed that users major categories of area include home, work/school, and downtown as the regions of interest. It enables people to have local knowledge in their regions. In addition, it enables them to feel that they are members of their regions.

9 DISCUSSION

In this section, we discuss the implications of our findings for the design of improved location-based social Q&A services and the limitations of this work.

9.1 Enhancing Interactive Map Usage

A map-based interface is provided in location-based social Q&A to support better usability. However, our investigation of users' map usage revealed that they usually correctly marked when they posted questions at a district/city-level, but 16.0% of questions that referred to specific POIs were pinned at incorrect locations. This behavior may be due to the fact that they did not want to invest extra effort or did not feel the need for providing detailed information. The problem is that such incorrectly pinned questions may confuse the answerers, because they will try to relate question texts with incorrectly pinned locations on the map. We can resolve this issue by automatically recommending related regions/POIs. Our manual investigation found that question texts typically included the name of a POI, district, or city regardless of the length of the question. Thus, it is beneficial to automatically extract location information from the question text to recommend a more accurate region category. We can further supplement the automatic location extraction with hyperlocal crowdsourcing. Local residents are thus able to easily distinguish whether questions are well-posted in appropriate regions in their regions of interest. Therefore, it might be valuable to give users an opportunity to correct/report incorrectly posted questions.

9.2 Leveraging Topical/Typological Patterns

Our work analyzed topical and typological patterns of location-based social Q&A; and we used fine-grained regional activity analysis based on administrative division labels (i.e., both cities and districts). In particular, we found that regional characteristics of topical distributions were more remarkable in district-level knowledge sharing than in city-level knowledge sharing. In the following, we illustrate how topical/typological locality can be leveraged to improve location-based Q&A services.

Topic based question filtering: Recall that NKH users are asked to subscribe to regions of interest. The users will be automatically notified of questions sent to these regions via push messages. Because our participants reported that 37% of the questions were answerable on average, it would be beneficial to include an additional option for selecting topical categories based on topics of interest of each user for push notification even in location-based social Q&A. This will assist to lower the interruption overhead caused by immoderate push notification.

Question routing: Our results showed that, compared to conventional Q&A services, NKH had lower answer rates (33% of the questions received no answer), longer delays before first answers, and some questions answered late (even several months later) by local experts in the regions. This may be due to the questions being categorized by region instead of topic. Answering location-based questions demands local knowledge of the region, which is difficult to acquire without living experience. To find relevant local experts in a specified region, we can automatically classify the user's topical interest and level of expertise on those topics [44, 52] and then route questions to topical local experts. In some cases, a user might post a local question through traditional social Q&A services. Then, we can automatically recognize that it is a local question (for example, questions including the name of a place or region) and automatically recommend a possible regional category.

Local search: We found that many location-based questions were related to local business/service. And our typological analyses demonstrated that factual information seeking was high (67.7%), followed by recommendations (17.5%). We also demonstrated that the topic distributions varied widely across different districts. This observation implies that a location-based social Q&A dataset could be effectively archived for local searches. For some areas, similar questions were posted repeatedly by many users (e.g., asking about good accommodations around tourist attractions). Furthermore, topic clustering algorithms could be applied such that the regional topic characteristics could automatically be extracted and utilized for local search optimization. Therefore, local search enables to understand regional characteristics such as what are the current issues in the region and what are the most popular topics based on the questions in the region.

9.3 Leveraging the Spatio-temporal Activity Analyses

Our spatial locality analyses revealed that there was strong spatial locality of user contributions. The answerers primarily focused on 1–3 spatial clusters that were closely related to their life experiences (e.g., current/former home, work, and school), and a cluster typically spanned a few neighboring districts. Furthermore, the mean contribution distance of their primary cluster was approximately 2–3km on average. It is interesting to note that the mean contribution distance in Wikipedia was much greater in scale (hundreds of kilometers) [16]. Our results imply that while a users subscribed districts are currently considered for push notifications, the radius of question geo-casting could be extended slightly to a few neighboring districts, and neighboring districts could be recommended to the user as additional regions of interest to leverage more local knowledge.

9.4 Motivating User Contributions

We found additional motivators that are unique in location-based social Q&A, i.e., ownership of local knowledge and sense of community. These motivators could be leveraged to encourage user contributions and increase user commitment. One immediate method is to use community-level symbols. A key element of membership in a sense of community is a common symbol system [31].

Considering that the existing location-based services often employ badges, we could create community badges and award them to those who are actively participating in that community. Awarding badges and levels act as a reward for achievement and social recognition (at least with the community that a user belongs to), thereby reinforcing contributive behaviors [23, 35].

Another method is providing individual/regional competition, which could elicit more contributions; for example, the system could provide scoreboards displaying user ranking by the number of answers, or region ranking by the number of answer/selection rates at different geographic scales (e.g., district- and city-level).

Another method is providing for individual/regional competition based on sense of community (i.e., membership), which could elicit more contributions. Recall that some users were concerned about lack of answers in their regions or the reputation of their regions. To this end, for example, we can provide regional scoreboards by displaying user rankings based on the number of answers, or regional rankings by the number of answer/selection rates at different geographic scales (e.g., district- and city-level). Since NKH allows users to maintain regions-of-interest, region-specific ranking would be fairly straightforward to implement in the system. In NKH, some users allow receipt of push notifications when questions are posted. Various regional Q&A statistics could be published along with the list of questions in the notifications. This could increase regional members awareness of regional Q&A activities, thereby encouraging contributions. Furthermore, highlighting the most popular questions (e.g., those which receive the highest percentage of views or answers) in a location-based social Q&A would help people to understand specific local characteristics (e.g., knowing "hot topics" in my area). This would not only provide significant insights for the local residents about their regions, but it would also assist people from other regions in understanding the topical characteristics of a region.

9.5 Generalizability

As with any qualitative or single-site work, the generalizability of this work is limited so that additional work on similar sites is necessary. Despite this limitation, we believe that our major findings provide the foundations on understanding the localness of knowledge sharing in location based social Q&A.

We demonstrated that topical and typological locality exists; and that locality is closely related to geographic characteristics. Another major finding is that strong spatial locality of user contributions exists in social Q&A services. It is very likely that our major findings on the localness of knowledge sharing also appear in other location-based social Q&A services.

In general, additional work on similar sites is necessary to improve the generalizability of this work, but overall, due to the difficulties in acquiring a similar dataset, we were not able to perform further analysis. At the time of writing, we found that location-based Q&A services such as Localmind, LocalUncle, and Locql were recently closed. Yahoo! Answers has topic categories related to geographic locations such as Travel and Local Business. However, we found that it only supports coarse-grained geographic categories based on text-based country and city names; and interactive maps are not supported. Topics were already preset to travel and business, which makes it difficult to analyze topical patterns across different regions. In addition, questions do not contain geographic coordinates, which are necessary for fine-grained spatial locality analysis. Due to these difficulties, we alternatively investigated social network services such as Twitter and Foursquare. Prior studies showed that Twitter and Facebook can be used for Q&A [32, 34]. We examined public a Twitter dataset, because geo-tagged tweets can be used for analysis. However, we faced several challenges: (1) Twitter only supports city-level location tagging, and only city-level topic analysis is feasible; and (2) in the dataset, questions are rare (and difficult to sift out automatically), and thus, it is very difficult to perform similar analyses to those presented in this article. As an alternative, we checked Foursquare, as it allows users to check-in and rate specific places as well as leave tips/comments on visited places. Like Twitter, questions were rarely asked because users mostly leave tips about visited places in Foursquare. Despite this limitation, analyzing this dataset will help us to understand spatial locality of check-in behaviors as in prior studies [48]. Note that each check-in instance includes anonymized user id, venue id, latitude, longitude, and venue-categories. For similar spatial locality analyses, we can map venue categories into topical categories and acquire administrative divisions of each check-in instance via open postal address mapping APIs. We downloaded a Foursquare dataset that includes long-term global-scale check-in data[13] (about 18 months from April 2012 to September 2013). Our preliminary analysis showed that there were 51,148 check-in instances from three major Korean cities such as Seoul, Pusan, and Incheon in the dataset; and district-level check-in behaviors were highly related to geographic characteristics. For example, similar to Q&A actvities (top 10 districts in Figure 8(a)), Jamsil and Apgujung have many check-in activities for a sports stadium and medical places, respectively. However, spatial locality of check-in behaviors is not directly related to the localness of knowledge sharing in location-based social Q&A, and thus, we did not perform further analysis in this direction. It would be interesting to study how mobility of users is related to the spatial locality of information seeking and sharing behaviors, by comparing these two datasets. Readers can download our NKH dataset from the following link: https://zenodo.org/record/46018.

10 CONCLUSION

We investigated the characteristics of location-based knowledge sharing by analyzing general users' behavioral characteristics, the topical and typological patterns related to geographic characteristics, geographic locality of user activities, and motivations of local knowledge sharing. We collected a large-scale real-world dataset from NKH and conducted a complementary survey of 285 mobile app users.

From the analyses, we found that NKH had different patterns of asking and answering questions, such as a lower answer rate and longer answering delays, compared with conventional social Q&A sites. This could be due to questions are being categorized by region instead of topic. Answering location-based questions requires local knowledge about the regions. Users focus on four to six topics, on average, to answer the questions. Regarding the map usage for question asking, we found that users preferred to pinpoint the question to district/city-level rather than to pinpoint a specific POI on the map, because they assume that people living in the region already know the POI. They also seem to not want to expend any considerable effort in using the map. Furthermore, location-based social Q&A has a unique topical and typological distribution. Lifestyle and travel topics are dominant, whereas ICT and game topics were popular in conventional Q&As. Although some topics such as education and economy were similarly popular in both types of services, the contents of the questions were different. The type of distribution across different topics also significantly varies in location-based social Q&A. A strong spatial locality of contributions exists around a few spatial clusters; more than 75% of answers focus mainly on fewer than two spatial clusters related to users' life experiences. These span a few neighboring districts such as home, work, and downtown. We also found that these characteristics of local knowledge sharing are facilitated by unique motives such as ownership of local knowledge and a sense of community.

Our results contribute to providing several practical system design implications such as enhancing location-related interfaces, leveraging topical and spatio-temporal activity patterns (e.g., question routing, local searches), and motivating user contributions (e.g., badges, request framing, hyperlocal news).

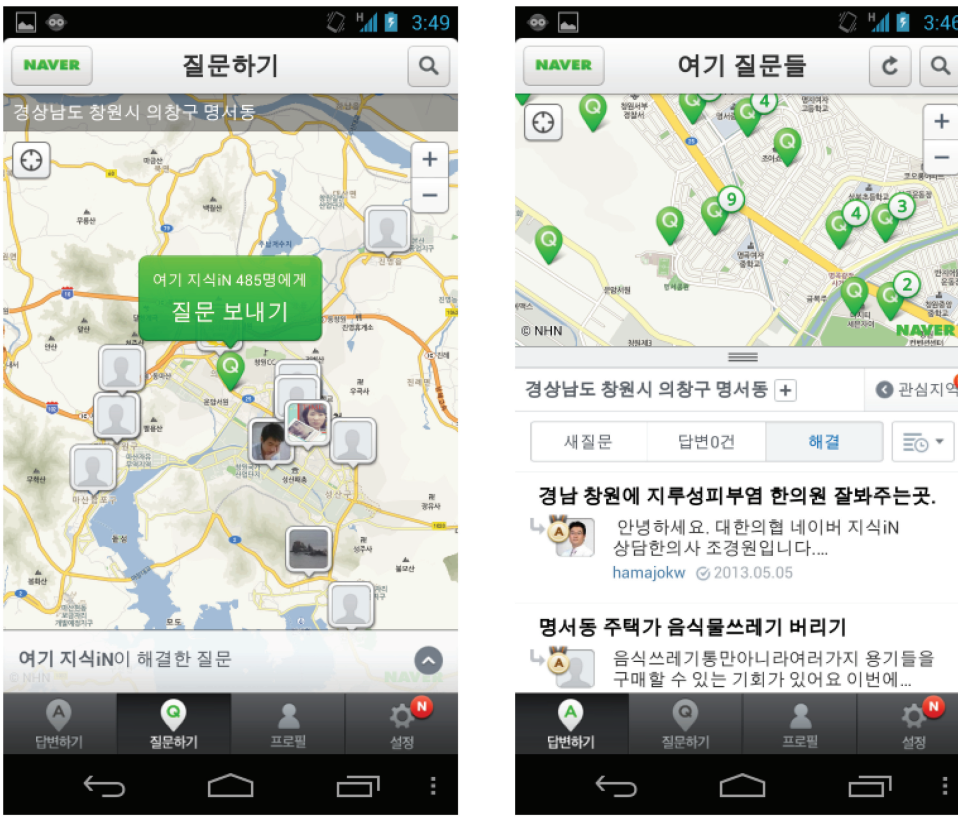
[13] https://sites.google.com/site/yangdingqi/home/foursquare-dataset.

APPENDIXES

A THE MAIN PAGE OF NAVER LOCAL Q&A

Fig. 19. The main page of Naver Local Q&A. (1) Region category, (2) User profile, and (3) Question list.

B NAVER KIN "HERE" USER INTERFACE



(a) (b)

Fig. 20. Naver KiN "Here" user interface (original Korean version of Figure 1).

REFERENCES

[1] Lada A. Adamic, Jun Zhang, Eytan Bakshy, and Mark S. Ackerman. 2008. Knowledge sharing and yahoo answers: Everyone knows something. In *Proceedings of the International World Wide Web Conferences Steering Committee (WWW'08)*. ACM, 665–674.

[2] Philip Berke and Edward John Kaiser. 2006. *Urban Land Use Planning*. University of Illinois Press.

[3] Mark Bilandzic, Marcus Foth, and Alexander De Luca. 2008. CityFlocks: Designing social navigation for urban mobile information systems. In *Proceedings of the ACM Conference on Designing Interactive (DIS'08)*. ACM, 174–183.

[4] KT Corporation's Research Center. 2015. Mobile Trend 2015. (7 2015). Retrieved from http://www.digieco.co.kr/KTFront/report/report_issue_trend_view.action?board_seq=10349&board_id=issue_trend.

[5] Minsoo Choy, Jae-Gil Lee, Gahgene Gweon, and Daehoon Kim. 2014. Glaucus: Exploiting the wisdom of crowds for location-based queries in mobile environments. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

[6] Karen Church, Joachim Neumann, Mauro Cherubini, and Nuria Oliver. 2010. SocialSearchBrowser: A novel mobile search and information discovery tool. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI'10)*. ACM, 101–110.

[7] Karen Church, Joachim Neumann, Mauro Cherubini, and Nuria Oliver. 2010. The "map trap"?: An evaluation of map versus text-based interfaces for location-based mobile search services. In *Proceedings of the International World Wide Web Conferences Steering Committee (WWW'10)*. ACM, 261–270.

[8] Karen Church and Barry Smyth. 2009. Understanding the intent behind mobile information needs. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI'09)*. ACM, 247–256.

[9] Thomas H. Davenport and Laurence Prusak. 1998. *Working Knowledge: How Organizations Manage What They Know*. Harvard Business Press.

[10] Benjamin Edelman. 2012. Earnings and ratings at Google Answers. *Econ. Inquiry* 50, 2 (2012), 309–320.

[11] Fredrik Espinoza, Per Persson, Anna Sandin, Hanna Nyström, Elenor Cacciatore, and Markus Bylund. 2001. Geonotes: Social and navigational aspects of location-based information systems. In *Proceedings of the International Conference on Ubiquitous Computing (UbiComp'01)*. Springer-Verlag, 2–17.

[12] M. Ester, H. P. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'96)*, Vol. 96. ACM, 226–231.

[13] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Vol. 96. 226–231.

[14] Andrea Forte, Michael Dickard, Rachel Magee, and Denise E. Agosto. 2014. What do teens ask their online social networks?: Social search practices among high school students. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 28–37.

[15] Shravan Gaonkar, Jack Li, Romit Roy Choudhury, Landon Cox, and Al Schmidt. 2008. Micro-blog: Sharing and querying content through mobile phones and social participation. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys'08)*. ACM, 174–186.

[16] Brent J. Hecht and Darren Gergle. 2010. On the "localness" of user-generated content. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'10)*. ACM, 229–232.

[17] Andreas Henrich and Volker Luedecke. 2007. Characteristics of geographic information needs. In *Proceedings of the ACM Workshop on Geographical Information Retrieval (GIR'07)*. ACM, 1–6.

[18] John B. Horrigan and Maeve Duggan. 2015. Home Broadband 2015. Retrieved from http://www.pewinternet.org/2015/12/21/home-broadband-2015/.

[19] Gary Hsieh and Scott Counts. 2009. Mimir: A market-based real-time question and answer service. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. ACM, 769–778.

[20] Gary Hsieh, Robert E. Kraut, and Scott E. Hudson. 2010. Why pay?: Exploring how financial incentives are used for question & answer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 305–314.

[21] Rosie Jones, Wei V. Zhang, Benjamin Rey, Pradhuman Jhala, and Eugene Stipp. 2008. Geographic intention and modification in web search. *Int. J. Geogr. Info. Sci.* 22, 3 (2008), 229–246.

[22] Soojung Kim, Jung Sun Oh, and Sanghee Oh. 2007. Best-answer selection criteria in a social Q&A site from the user-oriented relevance perspective. *Proc. Amer. Soc. Info. Sci. Technol.* 44, 1 (2007), 1–15.

[23] Robert E. Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. 2012. *Building Successful Online Communities: Evidence-based Social Design*. Mit Press.

[24] Juhi Kulshrestha, Farshad Kooti, Ashkan Nikravesh, and P. Krishna Gummadi. 2012. Geographic dissection of the twitter network. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM'12)*. The AAAI Press.

[25] Cliff Lampe, Rebecca Gray, Andrew T. Fiore, and Nicole Ellison. 2014. Help is on the way: Patterns of responses to resource requests on facebook. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM, New York, NY, 3–15.

[26] Uichin Lee, Hyanghong Kang, Eunhee Yi, Mun Yi, and Jussi Kantola. 2012. Understanding mobile Q&A usage: An exploratory study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*. ACM, 3215–3224.

[27] Uichin Lee, Jihyoung Kim, Eunhee Yi, Juyup Sung, and Mario Gerla. 2013. Analyzing crowd workers in mobile pay-for-answer Q&A. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'13)*. ACM, 533–542.

[28] Michael D. Lieberman and Jimmy Lin. 2009. You are where you edit: Locating wikipedia contributors through edit histories. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM'09)*. The AAAI Press.

[29] Pamela J. Ludford, Reid Priedhorsky, Ken Reily, and Loren Terveen. 2007. Capturing, sharing, and using local place information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, 1235–1244.

[30] Lena Mamykina, Bella Manoim, Manas Mittal, George Hripcsak, and Björn Hartmann. 2011. Design lessons from the fastest Q&A site in the west. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, 2857–2866.

[31] David W. McMillan and David M. Chavis. 1986. Sense of community: A definition and theory. *J. Commun. Psychol.* 14, 1 (1986), 6–23.

[32] Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. What do people ask their social networks, and why?: A survey study of status message Q&A behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'10)*. ACM, 1739–1748.

[33] Kevin Kyung Nam, Mark S. Ackerman, and Lada A. Adamic. 2009. Questions in, knowledge in?: A study of Naver's question answering community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*. ACM, 779–788.

[34] Jeffrey Nichols and Jeon-Hyung Kang. 2012. Asking questions of targeted strangers on social networks. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'12)*. ACM, 999–1002.

[35] Nathaniel Ostashewski and Doug Reid. 2015. *Gamification in Education and Business.* Springer International Publishing, 187–200.

[36] Sangkeun Park, Yongsung Kim, Uichin Lee, and Mark Ackerman. 2014. Understanding localness of knowledge sharing: A study of Naver KiN "here." In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services (MobileHCI'14)*. 13–22.

[37] Sharoda A. Paul, Lichan Hong, and Ed H. Chi. 2012. Who is authoritative? Understanding reputation mechanisms in quora. *arXiv Preprint arXiv:1204.3724* (2012).

[38] Daphne R. Raban and F. Maxwell Harper. 2008. New media and innovative technologies. *Motivations for Answering Questions Online*, Tal Samuel-Azran and Dan Caspi (Eds.). Ben Gurion University of the Negev Press, 73–91.

[39] Daphne Ruth Raban. 2008. The incentive structure in an online information market. *J. Amer. Soc. Info. Sci. Technol.* 59, 14 (2008), 2284–2295.

[40] Shilad W. Sen, Heather Ford, David R. Musicant, Mark Graham, Oliver S. B. Keyes, and Brent Hecht. 2015. Barriers to the localness of volunteered geographic information. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, 197–206.

[41] Chirag Shah, Jung Sun Oh, and Sanghee Oh. 2008. Exploring characteristics and effects of user participation in online social Q&A sites. *First Monday* 13, 9 (2008).

[42] Timothy Sohn, Kevin A. Li, William G. Griswold, and James D. Hollan. 2008. A diary study of mobile information needs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. ACM, 433–442.

[43] G Sterling. 2012. Study: 43 percent of total Google search queries are local. Retrieved from http://searchengineland.com/study-43-percent-of-total-google-search-queries-have-local-intent-135428.

[44] Juyup Sung, Jae-Gil Lee, and Uichin Lee. 2013. Booming up the long tails: Discovering potentially contributive users in community-based question answering services. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM'13)*. The AAAI Press, 13–22.

[45] Jaime Teevan, Amy Karlson, Shahriyar Amini, A. J. Bernheim Brush, and John Krumm. 2011. Understanding the importance of location, time, and people in mobile local search behavior. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI'11)*. ACM, 77–80.

[46] Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y. Zhao. 2013. Wisdom in the social crowd: An analysis of Quora. In *Proceedings of the International World Wide Web Conferences Steering Committee (WWW'13)*. International World Wide Web Conferences Steering Committee, 1341–1352.

[47] Sheng Wang and Raymond A. Noe. 2010. Knowledge sharing: A review and directions for future research. *Human Resource Manage. Rev.* 20, 2 (2010), 115–131.

[48] Weiqing Wang, Hongzhi Yin, Ling Chen, Yizhou Sun, Shazia Sadiq, and Xiaofang Zhou. 2015. Geo-SAGE: A geographical sparse additive generative model for spatial item recommendation. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15)*. ACM, New York, NY, 1255–1264. Retrieved from http://doi.acm.org/10.1145/2783258.2783335.

[49] Sarita Yardi and Danah Boyd. 2010. Tweeting from the town square: Measuring geographic local networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM'10)*. AAAI Press, 194–201.

[50] Hongzhi Yin, Bin Cui, Ling Chen, Zhiting Hu, and Chengqi Zhang. 2015. Modeling location-based user rating profiles for personalized recommendation. *ACM Trans. Knowl. Discov. Data* 9, 3, Article 19 (Apr. 2015). DOI: http://dx.doi.org/10.1145/2663356

[51] Zhijun Yin, Liangliang Cao, Jiawei Han, Chengxiang Zhai, and Thomas Huang. 2011. Geographical topic discovery and comparison. In *Proceedings of the 20th International Conference on World Wide Web (WWW'11)*. ACM, New York, NY, 247–256. DOI: http://dx.doi.org/10.1145/1963405.1963443

[52] Jun Zhang, Mark S. Ackerman, Lada Adamic, and Kevin Kyung Nam. 2007. QuME: A mechanism to support expertise finding in online help-seeking communities. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*. ACM, 111–114.

Received June 2015; revised November 2017; accepted February 2018