Spatiotemporal Distribution Change of Online Reference During the Time of COVID-19

Thomas Gerrish and Ningning Nicole Kong

ABSTRACT

The goal of this project was to identify the impact of the COVID-19 pandemic on the spatiotemporal distribution of the library’s online patrons, so that we could assess if the scheduled library reference hours are meeting the needs of the academic community. We collected each online reference patron’s location information via their IP address, as well as the timestamp of each online reference instance. The spatiotemporal distribution patterns were analyzed and compared before and after in-person instruction was suspended due to COVID-19 distance protocols and a closing of the campus in the 2020 spring semester. The results show that the geographic origins of reference questions redistributed after COVID-19 protocols were initially implemented and the university community underwent a temporary geographical redistribution. Reference question origins tended to move away from campus to other areas of the state, other states, and internationally. This population redistribution suggested that the library could adjust the online reference schedule to provide better access and service to patrons.

INTRODUCTION

The library’s online reference service, also referred to as library chat or digital reference, is a synchronous text-based interaction between the library and the patron via an internet connection, though audio/video communications are now also available. This online reference service provides a way to meet the information needs of patrons who cannot access the physical library location or prefer virtual communication. In this way, it expands the library’s reference services from the physical location to a virtual environment. When the university community was encouraged to socially distance due to COVID-19, online reference became a key library function that maintains the library’s connection to the community and their information needs. This connection became vital when most of the library’s physical services were suspended for a short period during the 2020 spring semester. For many libraries, chat became the only possible way to connect with patrons. This increased reliance on online reference and the greater dispersion of the student population led to a looming assessment question with regard to the service: Are the hours of online reference convenient for the populations that may live in time zones other than the library’s local time? That is to say, are we available when our patrons are likely to need us?

To address the above questions, this study recorded the time stamp and IP address associated with every incoming chat reference from the beginning of the 2019 fall semester to the end of the 2020 fall semester. We evaluated and compared the information from the spatiotemporal distribution pattern change of online reference patrons before and during the COVID-19 pandemic. An IP address is a unique string of numeric characters that identifies a particular computer or user over a network, and this number is generally saved in the background with all other information about the chat reference interaction. The IP addresses can be translated to
latitude and longitude information using geocoding services. With the latitude and longitude coordinates derived from the IP address, we determined a patron's location at a city level and time zone. Thus, the information helped to evaluate the user population distribution in the world and compare users' local times to the online reference service's operation hours. The patrons' location information as well as the timestamps were evaluated and analyzed in geographic information systems (GIS) to provide insights about how the online reference service met the user needs and how the library could improve, if possible.

BACKGROUND

Purdue University Libraries serves a large R1 university with a student population of 45,869 as of fall 2020. The online reference service is staffed by approximately 20 professional staff, though this number jumped to 29 when the library’s physical locations closed due to COVID-19. Online reference at our libraries uses the Springshare platform for synchronous chat service with the university community.

COVID-19 has had a measured effect on online reference hours of operation. Prior to COVID-19, online reference operated from 11 a.m. to 9 p.m., Monday through Thursday. On Fridays, online reference operated from 11 a.m. to 5 p.m. Sundays had truncated hours with the service open from 6 p.m. to 10 p.m. Immediately following the move to virtual instruction in March 2020, the administration requested that the online reference service reflect the original hours of the physical library as closely as possible. Thus, online reference hours of operation shifted to 7:30 a.m. to 10 p.m. with staff who normally cover the physical reference desks now covering the online service. Additional hours were added on Saturday afternoon, from 1 p.m. to 5 p.m., and Sunday hours were extended from 1 p.m. to 10 p.m. During the university's virtual instruction phase, only one physical library maintained limited hours of operation, from 8 a.m. to 5 p.m. for local students who needed a computer, Wi-Fi access, or printing. This was the only in-person service available until August 2020, when libraries began opening with COVID-19 restrictions in place. During summer 2020, online reference opening time was moved back to 9 a.m. to allow staff to cover the operations of the library, but the evening and weekend hours were maintained. Finally, in fall 2020, all physical libraries reopened with operating hours from 7 a.m. to 12 a.m. However, online reference hours did not return to the 2019 model. Instead, online reference operated from 9 a.m. to 11 p.m., Monday through Thursday; 9 a.m. to 5 p.m. on Friday; 1 p.m. to 5 p.m. on Saturday; and 2 p.m. to 11 p.m. on Sunday. Figure 1 shows the timeline of the physical libraries operating status and online reference operating hours changes in 2020.

Prior to COVID-19, this reference service mirrored the traditional in-person reference desk in the hours of operation and staff support. Indeed, it was originally conceived, structured, and promoted as a supplement to the in-person reference desk, which was stationed at each library. After campus-wide COVID-19 restrictions went into place in March 2020, on-campus students were asked to return home. All of the physical libraries except one were closed, and patrons were actively directed to online reference services. As a result, this service underwent changes in its availability and how users accessed it. While the service was already increasing in use, the post-COVID-19 period observed a 30% increase as it became more widely used by not only the student population but also the faculty and staff. At this point, online reference became the primary connection between the geographically distributed community and the library. On March 23, 2020, all classes became virtual, and students largely departed campus. This dispersion was not
only applied to students. Members of the faculty and staff communities also relocated once it was clear that work would need to be done remotely. In the fall 2020 semester, university classes resumed with a hybrid online/in-person structure. Of the total 45,869 students enrolled as of fall 2020, approximately 4,900 students (or 10.7% of the student population) elected online only classes for at least one semester. Given that online reference would still have a schedule of 9 a.m. to 11 p.m., Sunday through Friday, the question became does this serve the students who were potentially located in time zones geographically distant from the university.

**LITERATURE REVIEW**

Online reference service has been offered in academic libraries for more than three decades. Projects that evaluate online reference service, along with technology changes and user community needs, have been conducted ever since. For example, McClure et al. provided a guideline of statistics, measures, and quality standards for evaluating online reference services. However, not many studies have been done that assess the needs from patrons’ geographic perspective in order to improve the service.

Applying GIS analysis to improve the library’s services has a comparatively long history. Most of those studies are about mapping the interior spaces of libraries and understanding both space and facility use by patrons. Among the representative articles, Xia used GIS to map the physical location of library materials against a user’s self-height to gain a better understanding of patron browsing habits. Weessies used GIS to evaluate the likelihood of a computer station’s use in relation to the distance to the library entrance, windows, printers, number of neighboring computers, and library service locations. Given and Mandel combined the traditional library metric of user counts from “visual sweeps” with GIS to visualize library space usage, patron preferences, and traffic flow. Complementing this, Stoddart mapped library spaces against the expected use to visualize how library space was used. Shen takes the space model a step further...
by creating a library-space information model that can direct patrons to the shelf location of a given book while also giving the library data on circulation stats vs. book location and shelf height. Characteristic for this broad grouping of research is the use of GIS to analyze patrons, patron behavior, and library resources within the extent of the brick and mortar library building.

Spaces outside of library buildings have been less examined. Historically, the external space of the library has been important with regard to neighborhood service areas, wayfinding, distribution of reference within a consortium, and coverage over larger geographic areas. While many studies focused on the community and a library’s immediate locality and community, Donnelly mapped library geographic dispersion on the national level against existing United States populations to examine the variation in local library coverage.

With the implementation of COVID-19 procedures and the ubiquity of online reference services, the external location of patrons and their change in distribution over time has increasingly become an important question that GIS can address. There are limited studies using patron information, including IP addresses, to track patron location within an online reference model. In one example, Clark geocoded patron addresses to visualize the library’s external service area. In an academic library setting, Ruttenberg used the IP address to locate on-campus patrons when they asked a question to online reference. Kinkin studied patron locations in the world and used GIS to determine which populations were using the library and its branches while also informing decision makers on areas of low library coverage. These ideas were expanded in Mon’s study, which geolocated IP addresses for the physical location of a patron who asked a question within the statewide Florida Electronic Library collaborative chat service. Building on this, Bishop compared the originating location of a question against which librarian in a geographically disperse network asked the question as a measure of the utility of local knowledge in the reference process. In our study, we expanded upon the previous methods to compare patrons’ spatiotemporal pattern changes before and after the COVID-19 pandemic.

METHODS

Purdue University Libraries online reference service uses Springshare’s LibAnswers platform as an interface for chat. The system records each patron’s message text, timestamp, as well as the associated IP address. An initial data set of all online reference questions dated from fall semester 2019 until the end of fall semester 2020 (inclusive August 19, 2019, to December 15, 2020) was downloaded from the LibChat administration function as a .csv file. This initial data set included fields for IP address, date, time, interaction transcript, patron email (if provided), and name (if provided). This initial data set included 8,754 chat interactions excluding emails, SMS text transactions, and questions asked via Twitter. All reference interactions were initiated from the library’s website, the Ask-A-Librarian page, or through the proactive chat widget, which is located on all Purdue Library webpages. We do not require patrons to identify their relation to the university, their email address, or their status in the university (i.e., undergraduate student, faculty, staff, etc.), so this information is generally unavailable unless the patron self-identifies through the course of the reference interaction. Similarly, unless a patron identifies their physical location, reference staff generally do not know where in the world an online reference patron is located. In practice, most incoming online reference questions are anonymous with no indicators of identity or location.
The database was geocoded using the IPinfoDB tool to get latitude and longitude information.\textsuperscript{16} Per IPinfoDB, the platform’s accuracy is “99.5% on a country level and around 60% on a city level for the US within a 25-mile radius.” The date and timestamps were also associated with the output file. The latitude and longitude coordinates were translated to city, region (i.e., state), and country. Time zone information was added to each record according to its geographic location. The final data set contained separate fields for latitude, longitude, state or region, country, time zone, and timestamp, which includes both date and time.

At this point, all personal information embedded in the original data set was de-identified, though certain conversations could potentially identify users. One potential limitation to the data set is that if a patron used the university’s VPN network, the reference interaction would be georeferenced with an IP address on campus. Indeed, any VPN network would not report the reference interaction’s originating location correctly.

The data from each semester was broken down and sorted separately. In addition, records in spring semester 2020 for both pre- and post-COVID-19 restrictions were compared to measure the redistribution of a question point of origin due to COVID-19. The data were spatially plotted and analyzed using ArcGIS Pro.

RESULTS

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Total digital reference transactions per semester.}
\end{figure}
The digital reference transaction data were compiled and analyzed on a semester basis except for spring 2020, when all classes moved to 100% online on March 23, 2020. For the spring 2020 semester, the data were split into pre-March 23 (i.e., pre-COVID-19 restrictions) and post-March 23 (i.e., post-COVID-19 restrictions). The total number of chat interactions generally grew for each semester starting with fall 2019 (fig. 2). The number of summer chat interactions was relatively fewer because it was a shorter semester with fewer students taking classes as compared to the fall and spring semesters. We analyzed the spatial and temporal distribution of patrons before and after the COVID-19 pandemic in the following sections.

**Spatial Distribution of the Patrons before and after Implementation of COVID-19 Protocol**

The spatial distributions of digital reference patrons before and after the pandemic are mapped in figure 3. There is a trend showing that after the start of the pandemic, patrons were more geographically distributed within the Unites States and around the world. We mapped the international distribution of patrons in fall semester of 2019 and 2020 (see fig. 3(a) and 3(b)), as most of the international students make their travel plans by semester. There is a significant increase in questions coming from India, several European countries, and South America. We compared the spatial distribution of patrons before and after the implementation of COVID-19 protocols in spring 2020 within the United States as the time frame is more suitable for domestic travel plans. Figure 3(c) and 3(d) also show an increase of patrons around the country other than the campus area.

![Figure 3](image-url)

**Figure 3.** Spatial distribution of patrons before and after initial COVID-19 protocols closed the campus.
To help us further understand this spatial distribution change, we divided the geographic regions into four categories: local areas (West Lafayette and Lafayette areas of Indiana, where the Purdue University main campus is located), other Indiana areas, other states apart from Indiana, and areas outside of the United States. The transactions in these four regions are shown in table 1, and the percentages of transactions from each region were summarized in figure 4.

In fall 2019, which is considered the last “normal” semester prior to the COVID-19 response, 60% of reference questions originated in the immediate local area to campus (defined as the local area). The beginning of spring 2020 followed closely to this number with 56% of the total questions originating in the local area of the university. This proportion dropped at the March 23 boundary with only 29% of reference questions originating from the immediate local area to the university. During this period, there was an increase in questions originating in the state of Indiana as well as other regions of the United States and the world. During Summer 2020, 35% of the total number of chats originated in the campus area. In fall 2020, classes were offered as a combination of hybrid, in-person, and virtual formats; however, the proportion of questions originating in the immediate area did not return to fall 2019 levels. Instead, only 43% of questions originated from the campus area.

Figure 4. The percentage of transactions before and after the population redistribution due to the implementation of COVID-19 protocols in four geographic areas: local area (campus), Indiana, United States outside of Indiana, and international.
Table 1. Digital reference transactions coming from different geographic regions

<table>
<thead>
<tr>
<th>Semester</th>
<th>Total Trans.</th>
<th>Local Area</th>
<th>Indiana not local</th>
<th>Outside Indiana</th>
<th>Outside US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2019</td>
<td>1928</td>
<td>1156</td>
<td>203</td>
<td>493</td>
<td>76</td>
</tr>
<tr>
<td>2020 Spring Pre-COVID-19</td>
<td>1168</td>
<td>656</td>
<td>205</td>
<td>273</td>
<td>34</td>
</tr>
<tr>
<td>2020 Spring Post-COVID-19</td>
<td>1119</td>
<td>323</td>
<td>349</td>
<td>389</td>
<td>58</td>
</tr>
<tr>
<td>Summer 2020</td>
<td>1909</td>
<td>676</td>
<td>552</td>
<td>600</td>
<td>81</td>
</tr>
<tr>
<td>Fall 2020</td>
<td>2630</td>
<td>1118</td>
<td>670</td>
<td>714</td>
<td>128</td>
</tr>
</tbody>
</table>

In general, the geography of origin was redistributed primarily to the state of Indiana, followed by the United States as a whole, and then to international locations. Fall 2019 saw 11% of virtual reference questions originating in Indiana but outside of the local area. This increased to 18% in spring 2020 prior to the implementation of campus COVID-19 restrictions. After March 23, 2020, when all classes went online, the percentage of questions in Indiana outside of our campus area rose to 31%. This proportion remained steady during summer 2020, when 29% of questions originated in Indiana outside of the local area. In fall 2020, when in-person classes resumed, the proportion dropped to 26%, but this remains more than two times the proportion measured during fall 2019.

This pattern of redistribution of geographic origin was repeated in the data points outside the United States, though the fluctuation due to COVID-19, while similar, occurred to a lesser degree. In fall 2019, 3.9% of questions arrived from geographic origins outside of the United States. In spring 2020, prior to COVID-19 restrictions, this number dropped to 2.9%. After classes went virtual and students moved off campus, the percentage of questions increased to 5.2%. In fall 2020, the proportion dropped to 4.9%, but this is still not a return to fall 2019 levels.

The Distance of Digital Reference Patrons to Main Campus
To analyze the spatial distribution change of digital reference patrons, we calculated the distance of each patron to our main campus. A small portion of IP addresses (less than 4%) which couldn’t be correctly located was excluded from the analysis. Figure 5 represents the distance distributions in a box and whisker diagram. The horizontal lines within the boxes show the median of the data sets. In both fall 2019 and early spring 2020, the median distances are about 400 miles or less, which is within the local area. This indicates that most of the digital reference questions come from patrons who live around the main campus. The median distance increased to 1,000 miles after classes were moved online in spring 2020, which is about the typical distance of traveling within the state of Indiana. The maximum value was extremely high coming from international countries. This indicates that a significant portion of the patrons moved outside of the local area. In fall 2020, although the maximum distance dropped to a similar range as in normal semesters, the median and average values of the distance dataset are still much higher than the time before the pandemic.
To test the statistical significance of the distance values in different time periods, we conducted ANOVA tests for the distances in spring 2020 semester comparing before and after the pandemic, as well as a comparison between fall 2019 and fall 2020. The test results are shown in table 2. Both tests show there are significant differences before and after the classes were moved online. This means the pandemic situation significantly changed the patron distances to the main campus with p-values < 0.05. We likewise compared the distance between spring 2020 post-pandemic and fall 2020. There was no significance found. Although the university started to offer in-person classes in fall 2020, and most of the students were back to campus, there were still quite a few questions coming from students/faculty who were not living in the geographic area around campus.

Figure 5. The box and whisker diagram shows the distance from IP addresses to campus (in miles).
Table 2. The ANOVA test results of the patrons’ distance to the main campus before and after the pandemic protocols were implemented (in miles)

<table>
<thead>
<tr>
<th></th>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring 2020 Pre</td>
<td>1134</td>
<td>323,361</td>
<td>285.15</td>
<td>8E+05</td>
<td></td>
</tr>
<tr>
<td>Spring 2020 Post</td>
<td>1061</td>
<td>592,238</td>
<td>558.19</td>
<td>2E+06</td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>4.1E+07</td>
<td>1</td>
<td>4.1E+07</td>
<td>27.54942</td>
<td>0.0000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>3.3E+09</td>
<td>2193</td>
<td>1.4E+06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.3E+09</td>
<td>2194</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SUMMARY**

<table>
<thead>
<tr>
<th></th>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2020</td>
<td>2502</td>
<td>1E+06</td>
<td>500.95</td>
<td>2E+06</td>
<td></td>
</tr>
<tr>
<td>Fall 2019</td>
<td>1852</td>
<td>7E+05</td>
<td>385.43</td>
<td>2E+06</td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1.4E+07</td>
<td>1</td>
<td>1.4E+07</td>
<td>7.104865</td>
<td>0.0077</td>
</tr>
<tr>
<td>Within Groups</td>
<td>8.7E+09</td>
<td>4352</td>
<td>2.0E+06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8.7E+09</td>
<td>4353</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Temporal Distribution of the Questions**

While the spatial distribution of reference questions allowed us to understand where patrons were located, the temporal distribution of these questions helped us to plan the digital reference service hours to better meet patrons’ needs. We analyzed the temporal distribution of questions by the day of the week for fall 2019 and fall 2020. Figure 6 shows the median distances of the questions for each day of the two semesters. The distance was broken down into six ranges differentiated by color. From the nearest distance range to the ranges above, the analysis covers the questions coming around campus, in the local area within Indiana, around eastern United States (with mostly Eastern and Central time zones), the entire United States, and international locations. In fall 2019, we provided digital reference service Sunday through Friday, and in fall 2020, the service was provided every day except holidays. In fall 2019, the Monday to Friday range showed that most digital reference questions came from the campus area, especially in the first half of the semester. Questions from further distances within Indiana started to occur more often after November, probably due to holiday related travels. Relatively remote questions came more often on Sundays. One possible explanation for this difference is that students/faculty might travel away from campus during weekends. Interestingly, the fall 2020 weekly distribution of questions shows a different pattern. First, most median distances are further than the fall 2019 semester, which means there were a lot of questions coming from people living off campus, no matter if it was at the beginning of the semester or later. Second, there was no obvious difference between the median distances during the weekdays and weekends.
Figure 6. The median distance of reference questions in miles by weekdays (left: fall 2019, right: fall 2020).

In addition, we analyzed the hourly distributions of digital reference questions for fall 2019 and fall 2020 (fig. 7). In fall 2019, the median distance of the questions were mostly from campus, especially during the peak hours. If the median distances measured were not from the campus area, at most they were located in the local area where most of the off-campus students, faculty, staff, and research community live, or within greater Lafayette. Remote questions came from different time zones, such as international time zones or the Pacific time zone, and usually came either in the first hour or the last hour of digital reference service operating hours. In fall 2020, this distribution pattern changed. Most of the median distance ranged 200 miles, which meant that a large portion of the questions came from off-campus populations. There were additional time slots with median distances above 2,000 miles, which came from a time zone with at least 2 hours difference from our campus. Again, these questions were most often observed in either early or late reference service hours, i.e., 8 a.m. to 9 a.m., or 10 p.m. to 11 p.m.
DISCUSSION AND CONCLUSION

COVID-19 and the protocols developed in response to it had a redistributing effect on the geographic origin of reference questions in academic libraries. As the university closed and moved to virtual classes in response to the early pandemic, the geographic origin of reference questions redistributed away from campus. In our case, the geographic origins tended to move away from the campus to nearby areas within the state and neighbor states, though there was redistribution away from the campus at the state, regional, national, and international level. As of fall 2020, when the campus partially opened, these numbers have begun reversing themselves, but there is still a
significant population beyond the campus and the local time zone. Fall 2020 distribution numbers still show some of the redistribution effects observed early in the pandemic.

There were a surprising lack of questions coming from the Russian Federation, China, and Central Asian countries given that our university does have students from these countries. This may be due to the use of a VPN by users in these countries when accessing library resources. If a user were to use VPN, the question would be recorded as having a geographic origin in the VPN provider’s location, rather than in the country of origin. This is one possible explanation for the lower number observed in China and the non-existent users in the Russian Federation, Eastern Europe, and Central Asia.

This study demonstrated the broadening of our library’s geographic footprint in response to COVID-19 protocols. Students, faculty, and staff were not bound to campus and were free to study and work anywhere with internet access. In this regard, with populations distributed around the country and world, the expansion of reference hours was necessary. Prior to COVID-19, online reference operated from 11 a.m. to 9 p.m. This would mean students studying virtually in the Pacific time zone would experience effective reference desk hours of 8 a.m. to 6 p.m., which eliminates access during the evening hours. When the library extended online reference hours during the COVID-19 lockdown from 7:30 a.m. to 10 p.m., this somewhat improved the accessibility for patrons in the Pacific time zone. For those students studying online in the Pacific time zone, this creates effective reference hours of 4:30 a.m. to 7:00 p.m. PST.

The contrast becomes even starker when examining international students studying online in much more distant countries. Many students from India returned to their home country in spring 2020 following the move to virtual learning. For students studying in India, the online reference desk in pre-COVID-19 times would have effective hours 8:30 p.m. to 6:30 a.m. IST (India standard time), which forced this population to interact with the library during their evenings and nights. The expanded reference hours improved this access to 5 p.m. to 7:30 a.m. IST. While this is better, it still forces students in this part of the world to interact with the library during the evenings and nights and excludes daytime hours. Interestingly, the data from fall 2019 seems to indicate that there was an international population prior to COVID-19. These were likely students studying abroad, taking some of the early online classes, or simply traveling. Thus, the distributed online reference user population is nothing new but has been exacerbated by COVID-19 and the expansion of online classes. In this regard, the number of international reference interactions can be predicted to decrease as COVID-19 restrictions are gradually relaxed, but the number will not go to zero.

ENDNOTES


