

What is the Optimal Bar Crawl Route for My Requirements?

GIS Mini Project

Written By
Date of Submission

Monika Świderska, Lucille Ablett, Atsuyoshi Kita, Gosia Lachowska, David Jonathan Garcia
14 November 2016

I. ABSTRACT

Pub crawls, which can be also called bar crawls or barhops, are social activities that involve travelling between multiple pubs over the course of the night, whilst consuming at least one drink in each establishment. They are especially popular amongst students, but can also engage other groups such as tourists. Pub crawls have been a subject of scientific study in the past, however, the topic has never been analysed using scientific approach and geospatial analysis tools.

The aim of this study is to determine the optimal pub crawl route based on the user preferences of budget, pub rating, and walking distance between pubs using geospatial analysis tools.

The necessary information was acquired from the internet-based user review websites Trip Advisor and Yelp using Python code and then, filtered to reflect the desired preference criteria. As this research was conducted by a group of University College London students, for the purpose of this exercise the starting point of all routes was chosen as the UCL main entrance. This can be easily altered to reflect future applications.

A total of 9 routes were created for Fridays, Saturdays, and Sundays using the Greedy Algorithm and ArcMap software, taking into account the opening hours of the pubs. This method was found to be effective, however it was too limiting for this purpose as the next pub is selected based solely on its proximity to the current pub. The Greedy Algorithm does not take into consideration the potential locations of pubs later in the route, nor the density of pubs that are open during the period of the pub crawl. This often resulted in excessively long walking distances between pubs. It was concluded that the best area to pub crawl is around Leicester Square, since many bars are open late into the night. However, none of the bars in this location fall into the lowest price range and so budget would have to be sacrificed.

To further improve the process, additional pub selection algorithms should be investigated to ensure the most appropriate algorithm for the purpose. A new method is also required to influence the route of the generated pub crawl based on the locations with the highest density of pubs that both meet the preference criteria, and are open late in the evening. Assigning pubs a ranking based on proximity to other open pubs, and then using this in the selection criteria could potentially do this. Future applications of this research could result in the creation of a smartphone or web application capable of generating pub crawl routes on demand, according to user preferences.

Key words: *pub crawl, bars, Greedy Algorithm, network analysis, GIS*

II. INTRODUCTION

A. Background

A pub crawl, also known as a bar crawl or barhop, is defined by the Oxford English Dictionary as "a tour taking in several pubs or drinking places, with one or more drinks at each" (Oxford English Dictionary, n.d.). The term 'pub crawl' originated in Britain and has been in use since the early 20th century, with the terms 'gin crawl', 'beer crawl', and 'barhop' becoming popular in America around the same time (Kipfer and Chapman, 2007). Pub crawls can range from groups of friends on an informal outing, to planned events that are advertised worldwide. Organised pub crawls often target specific demographics, such as students and tourists, for example Carnage UK specialises in the organisation of pub crawls for undergraduate students in a number of British university cities (Carnage UK, n.d.).

The Guinness World Record for the largest bar crawl currently stands at 4,885 participants. The record was broken on 1st June 2013 in Missouri, USA, and required all participants to visit a minimum of 10 different bars within 8 hours (Moussaïd et al., 2010) and travelling either by foot or public transport. Participants were also required to consume at least half a pint of any alcoholic or not-alcoholic drink in each bar (Guinness World Records, 2016).

B. Problem Definition and Objectives

As current students, members of this research group recognised that pub crawls are an integral part of students' life. The events, organised by various student organisations and societies, are advertised on campus and in social media and it is a common knowledge that their main purpose is to help students to settle in to their new university environment and socialise with peers. The need for pub crawling as a social activity for other groups such as tourists, has been recognised by a number of research studies, such as those by Tutenges (2015), Thurnell-Read (2011), and

Sönmez et al. (2013).

Although many organised pub crawls advertise themselves as the ‘biggest’ or ‘best’, the definition is somewhat subjective. What some may consider the best bar crawl due to the low cost of alcohol, others may consider average, due to the low quality of the drinking establishments, or the large walking distance between bars.

In light of the need for social pub crawls and the lack of previous methodological approach on spatiotemporal analysis of pub crawling, the aim of this project is, therefore, to determine the optimal route for a pub crawl based entirely upon user preferences for the following criteria:

- Budget;
- Maximum desired walking distance; and
- Minimum visitor rating

C. Literature Review

In every city accessibility to facilities, such as pubs, restaurants, and clubs is an important factor affecting human behaviour and quality of life. Thus, the understanding of drivers of partying, socialising and pub crawling is crucial. A growing number of research studies in this area shows that scientists recognise this pattern and perceive it as a valid target of research. Quigg, et. al (2013) and Dodd, et. al (2012) analysed the alcohol consumption of students during pub crawls and concluded that drinking can have a range of negative impacts, such as: injuries, assaults, unintentional sexual activities, underachieving desired academic goals and financial problems. The study of tourists’ behaviour by Sönmez et al. (2013) also showed that excessive alcohol consumption during parties and pub crawls can lead to undesired social and morally questionable behaviours. However, as pub crawling is a socially acceptable behaviour that is commonly perceived as fun, the need for mapping potential routes for pub crawls can be justified.

Some organisations exist solely for the purpose of organising large-scale pub crawls for commercial purposes. Quigg, et. al (2013) highlighted some common characteristics of these commercial pub crawls; most of them occur on week day evenings and operate from around 7 pm to 2 am, with the three bars being the median number of bars visited per event.

Although a wealth of information on pub locations, visitor ratings, and price range is available from websites such as TripAdvisor, Yelp.co.uk, and Google, there is currently no means of utilising this information to automatically generate a pub crawl based entirely on the budget, proximity, and establishment rating preferences of the participants.

The iPhone and Android application ‘PubRally’ offers a partial solution to this problem in that it allows users to define a bar crawl route by selecting pubs that are listed in the app, and then advertise the routes to other app users (Influence, 2015). The app then allows all participants of the pub crawl to communicate via the integrated messaging feature, and post pictures and updates throughout the event. What the app does not do, however, is allow users to select pubs based on user reviews, budget, or proximity of bars to each other.

Although this project will not result in the creation of an app or website, the knowledge gained from the generation of routes conforming to the above criteria could aid in the future development of an app.

III. DATA DESCRIPTION

A. Data Sources from Social Media

The data required for the creation of the optimal pub crawl route comprised pub locations, user ratings, and price range of items for sale. This information was acquired from the following crowd-sourced data depositories:

- tripadvisor.co.uk
- yelp.co.uk

As a preparatory study, the data from TripAdvisor was downloaded and sample routes created with the resulting data. This sample route was then audited via fieldwork (Section V.C) to ensure validity. This exercise established that a large number of pubs are not listed on TripAdvisor, and also that many establishments are closed on weekends, highlighting the need to consider the opening hours of pubs to when creating a pub crawl route. Although TripAdvisor lists the opening times for many of its pubs and bars, this information was not listed for a significant number of establishments.

It was therefore necessary to use an alternative data source. Yelp was chosen as the alternative as it lists 8,710 pubs and bars in its database in the Greater London area (Yelp, 2016), compared to the 857 locations listed by TripAdvisor (TripAdvisor, 2016). In addition, the opening hours are listed on Yelp for the majority of the pubs and bars.

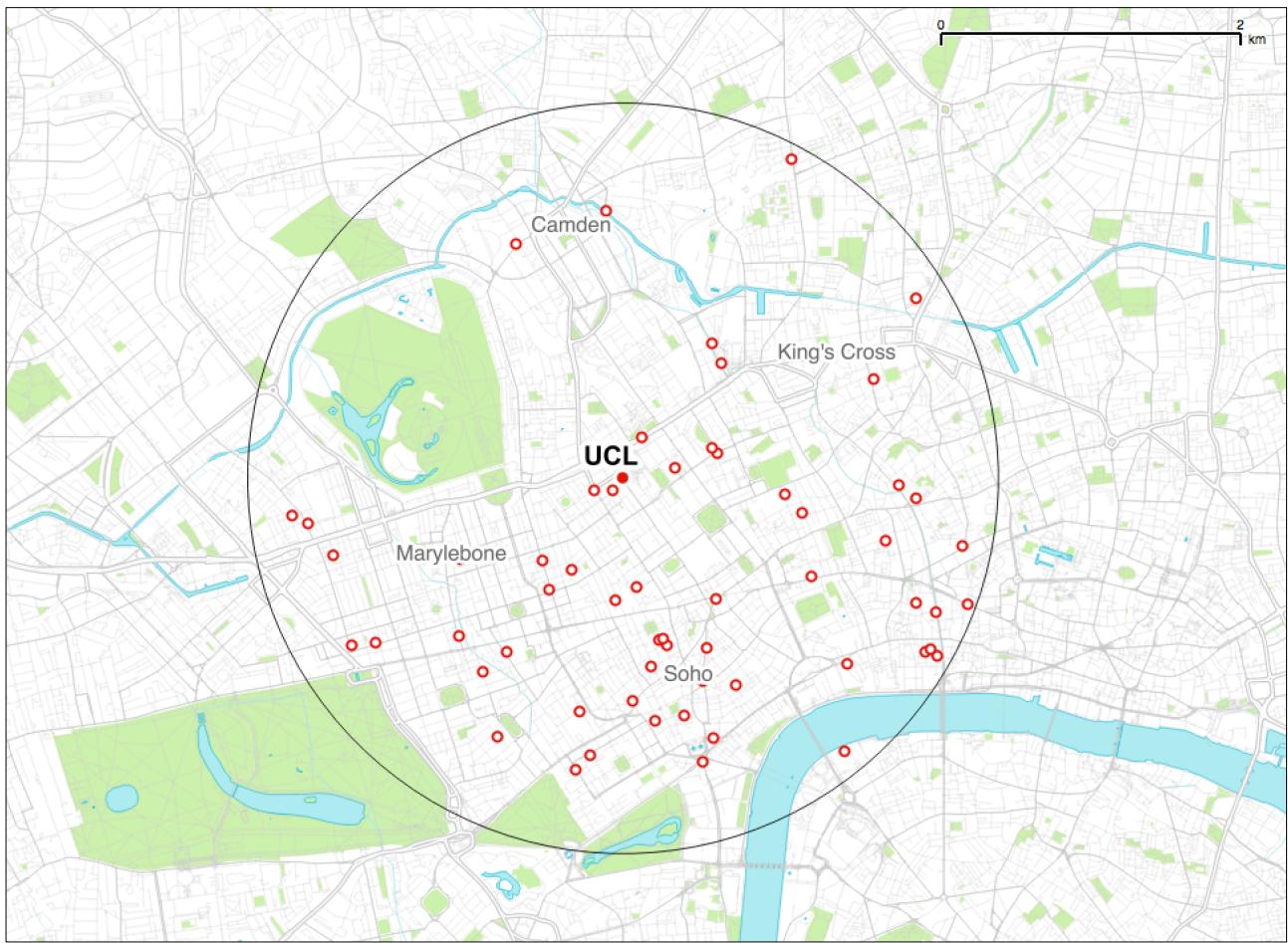


FIGURE 1. Map showing location of subject pubs from TripAdvisor (red circles) and with respect to the road network from OpenStreetMap (gray lines). The large circle represents an area covered by a 2.5 km radius from University College London (UCL).

B. OpenStreetMap

The research utilized road shapefiles that were downloaded from Geofabrik, which is website that hosts the global datasets of OpenStreetMap (Karlsruhe, 2016). OpenStreetMap is a project that relies on volunteered geographic information (OpenStreetMap Foundation, 2016). It is a very useful source of data due to an active community of practitioners who populate, validate, and curate the GIS data.

The unaltered version of the vector files (file name: gis.osm_roads_free_1.shp) was clipped using a bounding circle, which is discussed in the next section.

C. Survey

In addition to the TripAdvisor and Yelp reviews, a survey was carried out to better determine student bar preferences, and to aid in the creation of a “student” bar

crawl route. The survey was created using the SurveyMonkey platform and distributed to students in the Department of Civil, Environmental, and Geomatic Engineering at University College London (UCL) via email. The survey comprised one open question: “What is your favourite pub/bar in the UCL neighbourhood? Please state the name”. Due to the low number of responses (17) the results could not be utilised independently of the TripAdvisor and Yelp datasets, and were instead considered as supporting evidence to test the validity of the pub crawl routes designed for students, comprising pubs only in the lowest possible price range.

The results of this survey were useful in comparing the results of the network analysis with actual preferences of UCL students.

IV. METHODOLOGY

A. Data Acquisition

The required information was obtained from both TripAdvisor and Yelp via a Python script written for this purpose. The code enabled the data to be downloaded in HTML format.

Due to the large size of the Yelp dataset, a set of keywords was required to filter the dataset to find the establishments that matched requirements of the study.

The following table shows following information that the Python code obtained for each establishment:

TABLE 1. Information on pubs that were obtained using a Python Code

INFORMATION	DESCRIPTION
Name	e.g. The Old Bell Tavern
Rating	1 – 5
Number of Reviews	e.g. 26
Price Range	£ - ££££ (1 – 4)
Type	e.g. Pubs, bars, British restaurants etc.
Location	Latitude and Longitude
Opening Hours	Times per day of the week

Having obtained the data, the datasets were clipped to within 2.5 km of UCL and cleaned to remove duplicates and incomplete entries. This resulted in the following total number of pubs and bars from each source:

- TripAdvisor: 80
- Yelp: 259

B. Variables

1. Bar Crawl Logistics

For the purpose of this exercise, it is assumed that 50 minutes is spent in each pub visited during the pub crawl. This amount of time allows for the following activities at each location:

- Queuing for drinks
- Drinking
- Toilet breaks
- Dancing

It is also assumed that the pub crawl occurs between the hours of 19:00 and 02:00, which is a total of 7 hours. Assuming a total of 50 minutes in each bar, with a maximum walk of 10 minutes between bars, this allows for a maximum of 7 bars in an evening, with 60 minutes spent in the final bar.

For the purpose of this project, the starting point of the bar crawl was fixed at UCL. This could be easily changed if the concept were to be developed further, such as a mobile application.

2. Budget

The price range for each pub is recorded on both TripAdvisor and Yelp. Both TripAdvisor use a symbolic ranking system of £ - ££££, resulting in four possible price brackets. This has been translated to a numerical rating system of 1 – 4 for the purpose of analysis.

3. Walking Distance

The average walking speed of humans is 5 km/h (British Heart Foundation, n.d.), however research into human crowd dynamics (Moussaïd et al., 2010) determined that walking speeds decrease linearly with growing group size according to the following model, where y is the walking speed estimated in m/s. The formula is on the next page.

$$y = -0.08x + 1.24$$

As the study only analysed the walking speed of groups comprising up to a maximum of 4 persons, this model loses validity once the group size increases significantly above this figure, with walking speed estimates supposedly becoming negative once the group size reaches 16 people. For the purpose of this study, it is assumed that the average sub-group size of pub crawlers within the total pub crawling group is 5 people. Using the above model, the average walking speed of a group of 5 people is estimated to be around 3 km/h. Another factor affecting the walking speed of pub crawlers is the level of intoxication. The likelihood of people getting distracted whilst walking and requiring assistance between pubs increases as level of intoxication increases,

For this exercise, it was decided that the maximum desired walking distance between pubs is 10 minutes. Walking at a speed of 3 km/h, this gives a maximum desired distance between bars of 500 m.

Since there will be a maximum of 6×10 minute walks throughout the evening, this gives a total maximum range of 3 km from the starting location, although this was reduced to 2.5 km to prevent the creation of pub crawls in a straight line from the starting point.

4. Rating

The rating obtained for each pub represents an average of the total number of reviews submitted by the public. Both TripAdvisor and Yelp use a scale of 1 – 5, however only Yelp rounds the ratings to the nearest 0.5.

C. Running the Network Analysis

1. Method for Making a Route

In this research, a simple algorithm called Greedy Method (Cormen et al., 2001) (Cormen et al., 2001) was applied to create a route from the Yelp dataset. In this method, a pub is added to the route one by one until the required number of pub is contained in the route. The method can create a route with n pubs within n times of iteration. The detailed steps of the algorithm are as follows:

- Select the starting point (in this case, UCL).
- Set all the pubs as the target point.
- Go to the closest pub from starting point and spend 50 minutes there.

- For the next iteration, remove the current pub and the pubs that will be closed in the next one hour from the target group. Set the current pub as the new starting point.
- Repeat the steps 3 and 4 until a preferred number of stops have been obtained.

For each iteration, the pubs that have been already visited or closed are removed from the target group. Therefore, the amount of pubs to choose from becomes limited with time.

These steps are available in the “find closest facility” function under the Network Analyst toolbar of ArcMap 10.3 (ESRI, 2016)

2. *Criteria for Choosing Pubs for Budget Route and Quality Route*

Initially, a pub crawl route was generated using all pubs within a 2.5 km radius of UCL. The route did not take into account any budget or pub quality criteria. In order to create a route based on budget, establishment quality, or both, the pubs that satisfied the desired criteria were first selected and then the other pubs are removed from the list. The above algorithm was then applied using the remaining pubs.

To generate a route for participants with the lowest budget, the pubs in price range 1 (the lowest price range) were selected. For a route where quality is favoured over cost, the pubs with more than 10 reviews that are rated 4 or higher were selected. By filtering the pubs before applying the Greedy Heuristics, it is guaranteed that all the pubs in the route are within the desired price range or rating group. This method, however, does not take into account the walking distance between locations, and so the pubs are not guaranteed to be within 500 m or less of the previous location. If the walking distance is greater than 10 minutes, then less than 50 minutes is available to spend at the next location.

An illustration of the Greedy Method is shown in Figure 2.

3. *Method for Making a Route with the Data from Trip Advisor*

As described in Section 3A, in the dataset obtained from TripAdvisor did not contain sufficient information on opening hours. The route shown in the Section 4.5 was created without incorporating the opening hours, and therefore assumed that all the visited pubs are open until 2 am. The TripAdvisor route was created using the same algorithm and criteria as previously described.

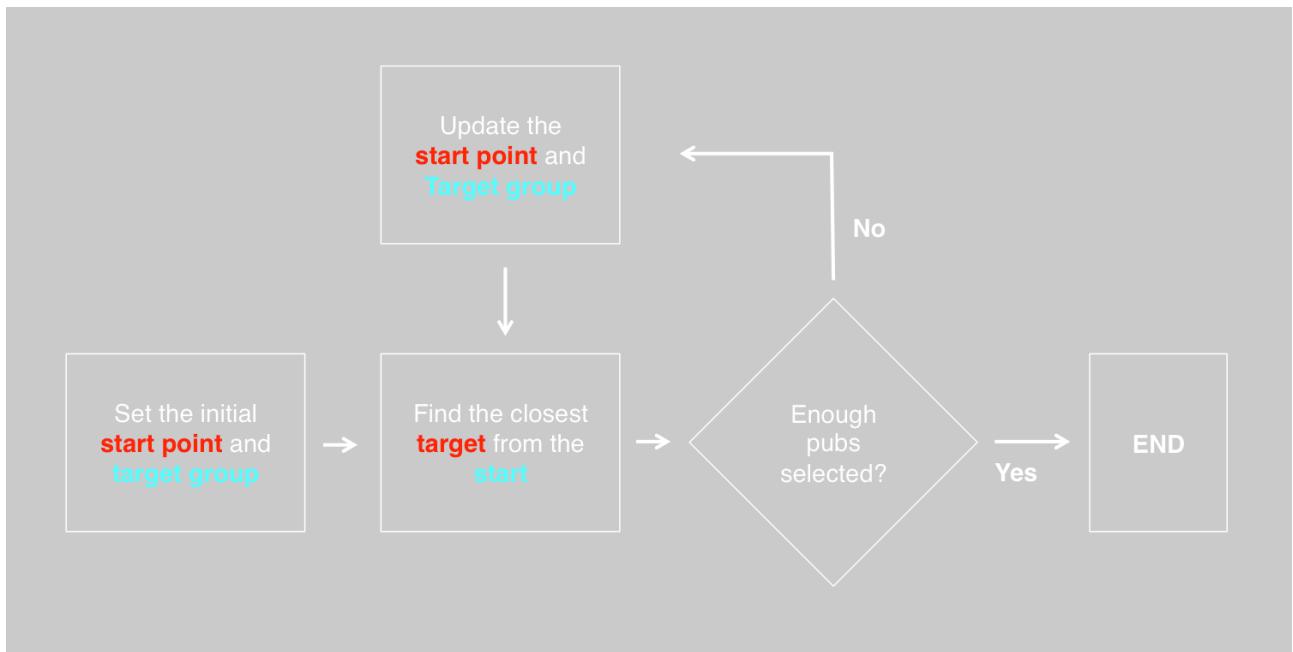


FIGURE 2. Flow chart of the Greedy Method.

D. Field Exercise

To cross check the dataset and the results of network analysis, the designed pub crawl route was checked in practice by the members of the research group on Saturday 29th October 2016. It was consequently established that a significant number of bars were closed on weekends, a factor that had not previously been considered when creating pub crawl routes.

V. RESULTS AND ANALYSIS

A. Data Overview

1. Price Range

The total number of pubs within a 2.5 km radius of UCL according to Yelp is 259. The price range of these pubs is distributed from 1 to 4, with mean of 1.89 and standard deviation of 0.44, indicating that the majority of the pubs fall into the second cheapest pub range (Figure 3).

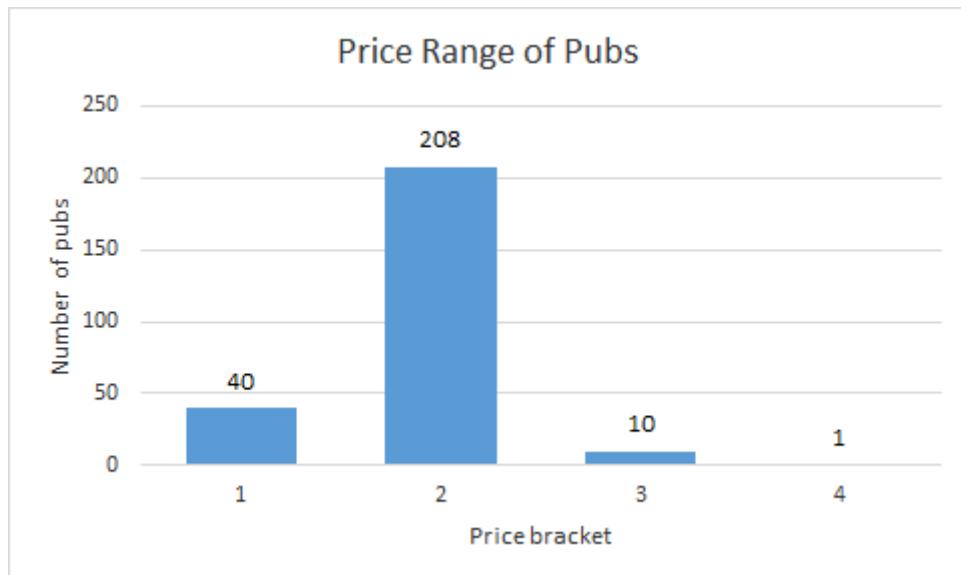


FIGURE 3. Price range of pubs within 2.5 km of UCL (Yelp, 2016)

2. Ratings

The rating of pubs within a 2.5 km radius of UCL ranges between 3 and 5, with mean of 3.71, and standard deviation of 0.49. Figure 4 shows that the majority of pubs are rated between 3.5 and 4, with a limited number rated above 4.5.

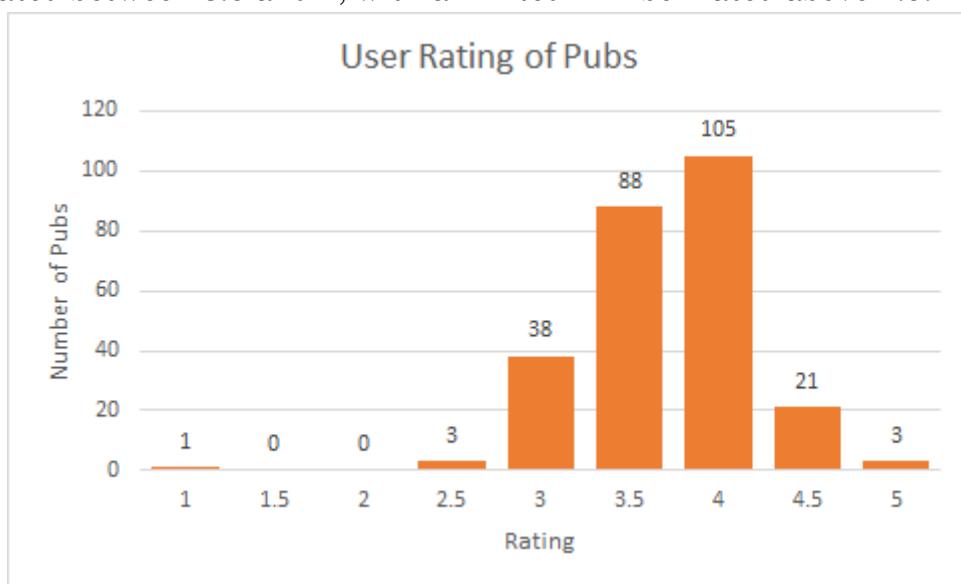


FIGURE 4. Distribution of user rating pubs within 2.5 km of UCI (Yelp, 2016).

3. User Reviews

The number of reviews per establishment ranges between 1 and 252, with mean of 29.27. The median number of reviews is 23. The distribution is shown in Figure 5

below.

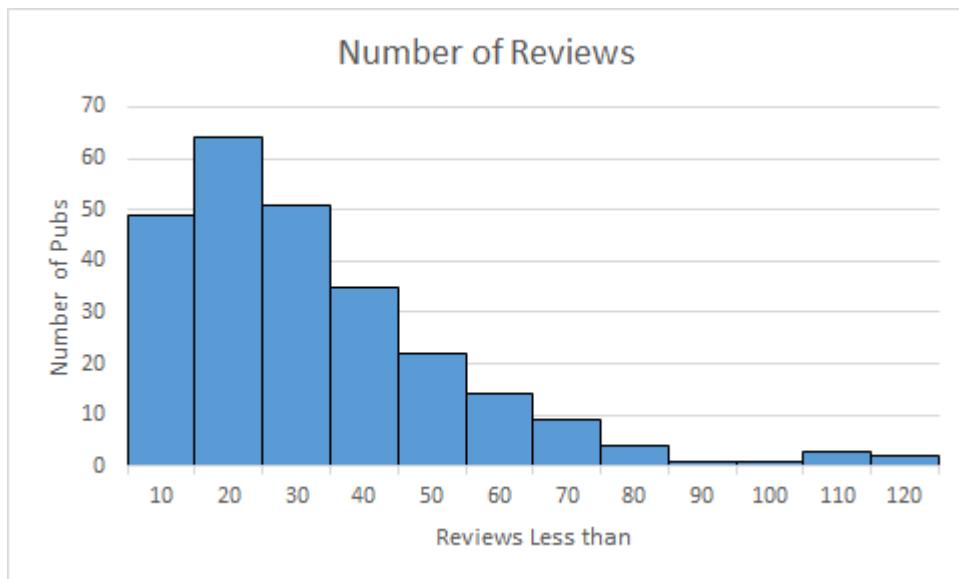


FIGURE 5. Number of reviews per pub (Yelp, 2016)

4. Opening Hours

The number of pubs that are open decreases over the course of the evening. After 12 pm it becomes difficult to find a pub that is open, with only a few pubs open later than 1 am. On Saturdays, 26 pubs are closed all day, therefore the number of pubs that are open between 8 pm and 10 pm are fewer than on weekdays. After midnight, however, there are still more pubs open than there are on weekdays. On Sunday, 78 pubs are closed all day, which is more than on any other day of the week, including after midnight.

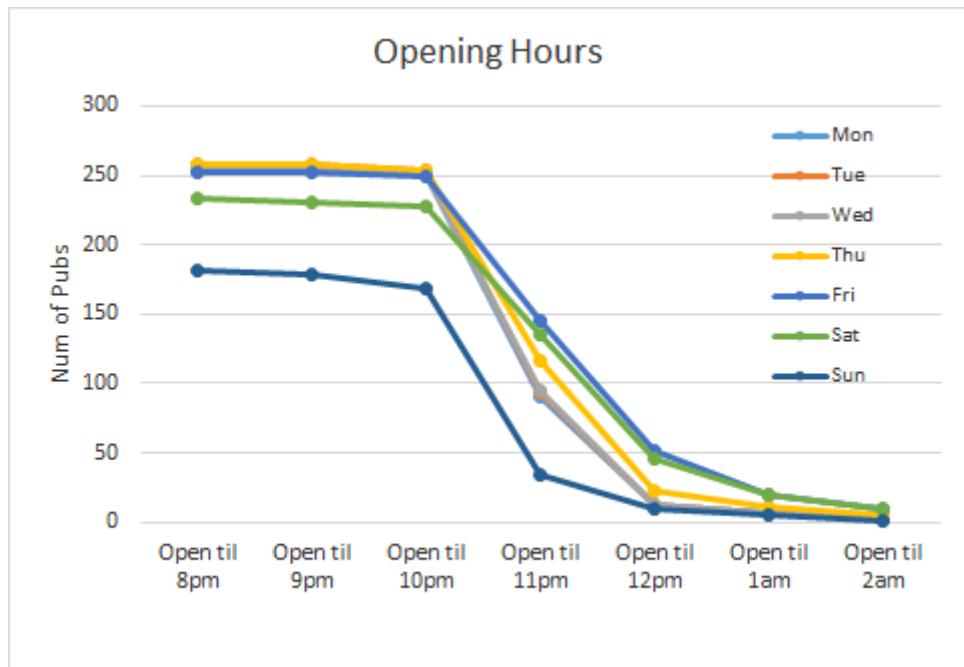


FIGURE 6. Opening hours of pubs within 2.5 km of UCL (Yelp, 2016).

B. Sample Routes

1. Route Generated Using No Preference Criteria

Using all the 259 pubs in the Yelp dataset, the Greedy Method was applied to create three test pub crawl routes in ArcMap 10.3. The routes were generated for Friday, Saturday, and Sunday, taking into account the different opening hours on each day. The routes and auxiliary information are shown in Figure 7 and Table 2, respectively. The following observations were made when comparing the three routes:

- All three routes initially travel north east from the main entrance of University College London.
- On Friday and Saturday, the routes terminate at a pub "Central Station" near King's Cross Station.
- On Sunday, since the pub "Central Station" is closed, the generated route selects the pub the "Red Lion" near Oxford Street, which is almost 3.0 km from the previous location. This long walking distance resulted from the limited number of open pubs open on Sundays.



FIGURE 7. Sample routes on Fridays, Saturdays, and Sundays using no preference criteria.

LEGEND

The green lines represent the road network within a 2.5 kilometer radius of UCL.

Yellow lines show the route segments that are shorter than 500m (<10 mins walk at 3km/h).

The red lines show the route segments that are longer than 500m (>10 mins walk at 3km/h).

TABLE 2. The names of pubs for the “no-preference” routes on selected days, and their distances (meters) from the previous pub stop.

FRIDAY		SATURDAY		SUNDAY	
<i>name</i>	<i>distance(m)</i>	<i>name</i>	<i>distance(m)</i>	<i>name</i>	<i>distance(m)</i>
Jeremy Bentham	153.7	The Court	267.0	The Court	267.0
The Court	113.4	Carpenters Arms	132.6	Carpenters Arms	132.6
Carpenters Arms	132.6	Mortimer Arms	124.8	Fitzrovia Belle	128.0
Mortimer Arms	124.8	Fitzrovia Belle	3.1	Euston Tap	768.6
Euston Tap	765.5	Euston Tap	768.6	The Rocket	319.3
The Rocket	319.3	The Rocket	319.3	Old China Hand	1706.8
Central Station	940.4	Central Station	940.4	Red Lion	2984.8
average	364.2	average	365.1	average	901.0

2. Cheapest Budget Route

Using the 40 pubs in the lowest price range from the Yelp dataset, the Greedy Method was applied to create an additional three pub crawl routes, again for Friday, Saturday, and Sunday. The routes and auxiliary information are shown in Figure 8 and Table 3, respectively. The following observations were made:

- On Friday and Saturday, the routes stretch towards South West from University College London main entrance. They all terminate in the area between Oxford Station and Tottenham Court Road Station.

- On Friday and Saturday, after the 6th pub, there are no more pubs left to visit that meet the pub crawl criteria, with the only remaining possibilities to either accept a higher price range, or terminate the pub crawl early and remain at the final bar for two hours. By accepting a higher price range, there are pubs that fall within an acceptable walking distance of the 6th bar.
- On Sunday, there are no open pubs to choose from after visiting 5th pub. The best option is to terminate the route at this point as even the pubs that fall within the higher price ranges are located far away.

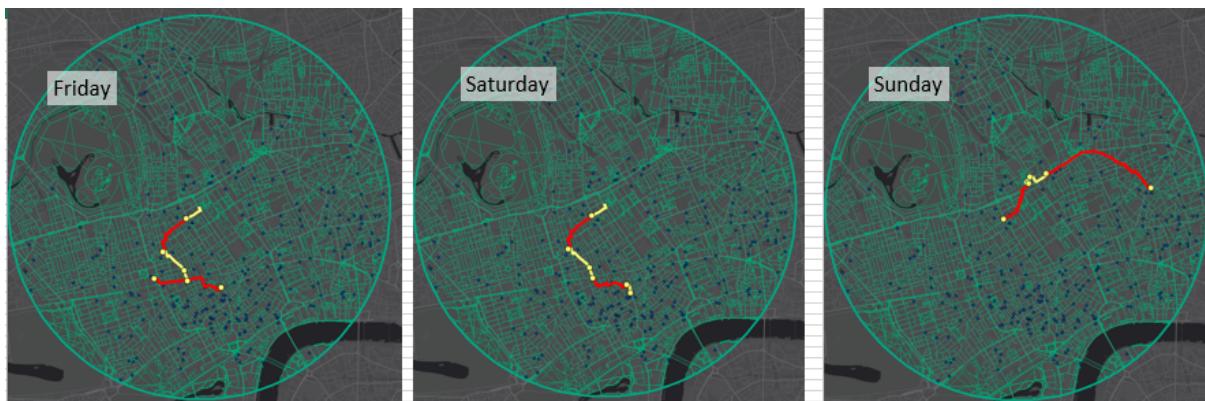


FIGURE 8. Sample routes on Fridays, Saturdays, and Sundays using the cheapest budget criteria.

LEGEND

The green lines represent the road network within a 2.5 kilometer radius of UCL.

Yellow lines show the route segments that are shorter than 500m (<10 mins walk at 3km/h).

The red lines show the route segments that are longer than 500m (>10 mins walk at 3km/h).

TABLE 3. The names of pubs for the “cheapest budget” routes on selected days, and their distances (meters) from the previous pub stop.

FRIDAY		SATURDAY		SUNDAY	
<i>name</i>	<i>distance(m)</i>	<i>name</i>	<i>distance(m)</i>	<i>name</i>	<i>distance(m)</i>
The Court	267.0	The Court	267.0	The Court	267.0
The Yorkshire Grey	662.3	The Yorkshire Grey	662.3	Cider Tap	710.3
The Champion	436.0	The Champion	436.0	The Doric Arch	154.6
Green Man	119.7	Green Man	119.7	The Rocket	288.5
Old Explorer	496.8	The Crobar	617.6	Old China Hand	1706.8
The Crobar	1086.0	Molly Moggs	133.0	Terminate at 5th pub	
Terminate at 6th pub		Terminate at 6th pub			
average	511.3	average	372.6	average	625.4

3. Highest Quality Route

Using the 107 pubs from the Yelp dataset with at least 10 reviews at 4 or above, the Greedy Method was applied to create three pub crawl routes for Friday, Saturday, and Sunday.

As shown in Figure 9, all three routes advance towards the north to visit a pub near Camden Road Station. The routes then all return toward the south to a pub near Oxford Circus Station, resulting in a long walking distance of almost 3.0 km between pubs. The average walking distance between pubs is 1.2 km for Friday and Saturday, increasing to 1.4 km on Sunday. If the routes headed towards the south from the beginning, there would be several pubs open until 2 am that could be included in the route. This would also have considerably reduced the average walking distance between pubs. Auxiliary information is also shown in Table 4.

User rating cannot, therefore, be used exclusively in the generation of pub crawl routes using the Greedy Method, although this is likely due to the inefficiency and limitations of the algorithm used. This issue is further explored in section A1 in the Discussions and Conclusions.



FIGURE 9. Sample routes on Fridays, Saturdays, and Sundays using the highest quality criteria.

LEGEND

The green lines represent the road network within a 2.5 kilometer radius of UCL.

Yellow lines show the route segments that are shorter than 500m (<10 mins walk at 3km/h).

The red lines show the route segments that are longer than 500m (>10 mins walk at 3km/h).

TABLE 4. The names of pubs for the “highest quality” routes on selected days, and their distances (meters) from the previous pub stop.

FRIDAY		SATURDAY		SUNDAY	
<i>name</i>	<i>distance(m)</i>	<i>name</i>	<i>distance(m)</i>	<i>name</i>	<i>distance(m)</i>
Cider Tap	443.3	Cider Tap	443.3	Cider Tap	443.3
Euston Tap	48.7	Euston Tap	48.7	Euston Tap	48.7
Exmouth Arms	404.0	Exmouth Arms	404.0	Exmouth Arms	404.0
The Somers Town Coffee House	735.8	The Somers Town Coffee House	735.8	The Good Mixer	1587.2
The Colonel Fawcett	1668.0	The Colonel Fawcett	1668.0	Red Lion	3221.0
Old China Hand	2941.1	Old China Hand	2941.1	Old China Hand	2984.8
The Crobar	2161.8	The Crobar	2161.8	Terminate at 6th pub	
average	1200.4	average	1200.4	average	1448.2

C. Survey Data

Despite the limited response rate from the students’ survey, the data acquired can be discussed. The most popular pub amongst students was "The Court", which was included in the dataset. "The Court" is a pub to be visited while taking on a normal or a budget route. The other popular venues, "Phineas Bar" and "Richard Mully's Basement Bar" (referred to as "Mully's") are bars run by University College London students' union, therefore, the accessibility to those places is limited to holders of student identification cards. Both bars are located on the University College London campus and are within a short walking distance from the starting point of bar crawls. They are also in close vicinity to all the previously generated routes. Therefore, if the generated pub crawl is to be created for students, and attended solely by students, those places make excellent additions to the pub dataset.

TABLE 5. Summary of results of quick survey.

Name	Number of responses
The Court	5
Phineas	4
Mully's	3

All of the routes that were produced from the analysis are shown in Figure 10 on the next page.

FIGURE 10

Pub Crawl Route Options

○ ————— 5 KM
N

ABOUT THE ROUTES

High quality routes are for reaching pubs with 4- to 5-star rating.
Low budget routes are for reaching pubs in the lowest price class.



HIGH QUALITY ROUTE
friday



HIGH QUALITY ROUTE
saturday



HIGH QUALITY ROUTE
sunday



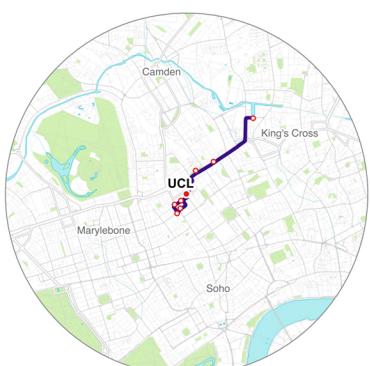
LOW BUDGET ROUTE
friday



LOW BUDGET ROUTE
saturday



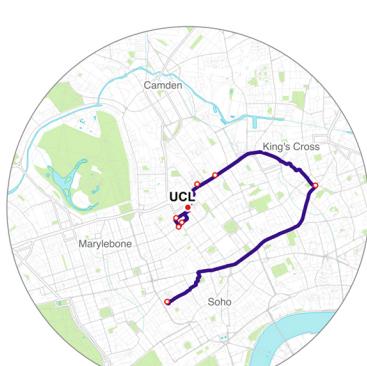
LOW BUDGET ROUTE
sunday



ALL PUBS ROUTE
friday



ALL PUBS ROUTE
saturday



ALL PUBS ROUTE
sunday

DATA SOURCES

Road network from OpenStreetMap
Pub data from Yelp

MAP ANALYSIS AND CARTOGRAPHY

Lucille Ablett, Atsuyoshi Kita, Gosia Lachowska
Monika Swiderska, David Garcia

UNDER THE GIS MINIPROJECT REQUIREMENT

For the GIS Principles and Technology course
CEGEG082

VI. DISCUSSION AND CONCLUSIONS

To determine the optimal bar crawl route taking into account budget, desired pub quality, and maximum walking distance preferences, a total of 9 routes were generated using the data obtained from social media portals. It was found that, using Yelp data alone, it is not possible to generate any pub crawl routes that contain pubs only within 500 m of each other, and that also fall precisely within the desired criteria of either budget or quality. Even when generating a pub crawl route with no budget or quality constraints, the resulting routes still contained segments of the route where the maximum walking distance exceeded the desired 500 m. The pub crawl routes with the shortest average walking distance were generated for Fridays and Saturdays, with an average walking distance between pubs of 365 m. Even so, the final walk of the night was calculated to be almost double the desired distance, at 940 m. This was due to a significant number of pubs closing after midnight.

Restricting the route to pubs falling within the cheapest budget range resulted in the generation of routes with acceptable walking distances between pubs, with only two of the segments being greater than 500 m. However, due to the reduced number of pubs available to choose from, the bar crawls on all three days had to be cut short since no open pubs were open later in the evening. This was particularly evident on Sundays when the generated route was stopped after the 5th pub. Selecting pubs based on the high quality of their reviews results in even longer walking distances of up to almost 3 km on Sunday evenings, and even then it was necessary to end the pub crawl early due to lack of open establishments.

Due to the fact that some of the pubs close earlier than others, it is worth checking which area the pub crawls route should lead to in order to avoid closed venues. After analysing the data and creating routes on the map, it appeared that the best area to continue pub crawls late at night, would be central area located South of the initial start point at University College London main entrance. This direction of a planned pub crawl would be especially good, if the requirement is that the route is a budget one and pub crawlers are looking for cheap drinks.

Initial maps were created using the data from TripAdvisor. The route was then tested on 29th October 2016 as described in Field Exercise section (ref). The test exercise showed that the data from TripAdvisor is incomplete, and thus, invalid for the purpose of this research as a stand alone dataset. Therefore, new data from Yelp were acquired in order to verify initial results of TripAdvisor analysis. The possible cause of the limitation of TripAdvisor data is that not all pubs have profiles on the portal and as it is edited by volunteers and can be accessed by everyone, the data might contain biased or misleading information.

The dataset obtained from Yelp seemed to be fit for the purpose of this research,

however, it still might not be complete due to the filtering and extraction process used. Although possible pub crawls routes were generated for several scenarios, it seems that organizing this type of events on Sunday is not a good idea. Most of the venues are closing early or do not open at all. The obvious reason for that would be that Monday is a working day and people would not like to engage in drinking before work. Students or tourists might not be affected by these time constraints, however, even for them the option of Friday or Saturday pub crawls would be better, simply due to the fact that more venues are open on these days.

A. Errors and Limitations

1. Inefficiency of Algorithm

Due to its simplicity, the Greedy Method often results in routes with long walking distances between pubs. Figure 11 shows the route generated from highest quality establishments on Friday and Saturday, after visiting the 4th pub. The pubs available for the next one hour (from 11 pm to 12 am) are shown with light blue dots. The available pubs are categorized into three areas, Camden, East, and Central. All the four pubs in the Camden area close at midnight, whilst three of the pubs in East area close at midnight, with one remaining open until 1 am. In the Central area, three of the pubs close at 12 am and other three remain open until 2 am.

Considering the situation, the best strategy would be to head towards the Central area. However, the algorithm chooses to go to Camden Area, because one of the pubs there is the closest from current position. The algorithm is therefore perhaps not the most suitable algorithm for this purpose, or should, most likely, be used in conjunction with alternative algorithms.

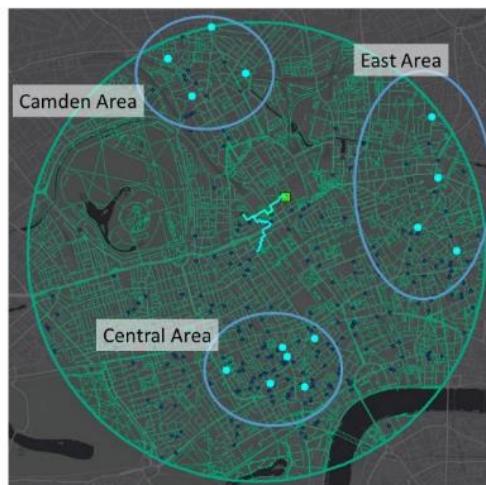


FIGURE 11. The actual clustering of pubs (encircled dots) compared with a sample route (blue green line) from the algorithm.

Preliminary knowledge of the spatial distribution and density of pubs open late at night provides a simple solution to overcome limitation of the Greedy Method. In that case, it is possible to guide the algorithm towards the desired region by removing the pubs in the undesirable areas before applying the algorithm. In the example case in Figure 12, it is possible to create a better route by deleting all the pubs to the north and east of the starting point. The improved route generated by this method for a Friday pub crawl to high quality pubs is shown in Table 6. The improved route drastically reduces the average walking distance between pubs from 1200.4 m to 505.7 m, which is almost within the maximum desired walking distance of 500 m.

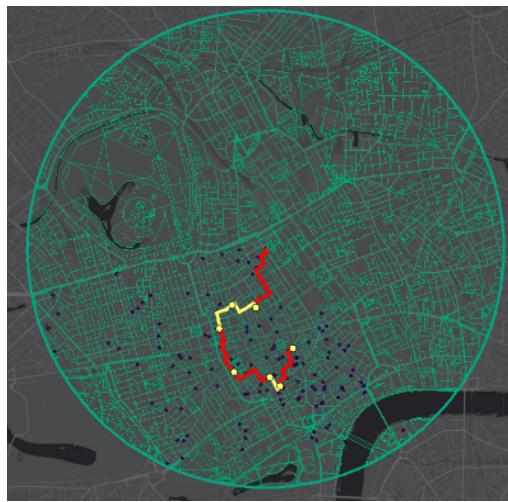


FIGURE 12. The improved pub crawl route after deleting pub clusters that are far from the starting point.

LEGEND

The green lines represent the road network within a 2.5 kilometer radius of UCL.

Yellow lines show the route segments that are shorter than 500m (<10 mins walk at 3km/h).

The red lines show the route segments that are longer than 500m (>10 mins walk at 3km/h).

TABLE 6.

The names of pubs for the improved route on a Friday and their distances (meters) from the previous pub stop.

<i>name</i>	<i>distance(m)</i>
Duke Of York	868.8
The Green Man	369.0
The Cock	389.7
Red Lion	582.3
The Yard Bar	571.8
O'Neill's Wardour Street	186.1
The Crobar	572.3
average	505.7

2. Data Limitations

After initial analysis of the TripAdvisor dataset, the dataset was ascertained to be incomplete, with a significant number of establishments lacking information on opening hours. In addition, some establishments were eliminated from the downloaded dataset since they lacked information on ratings, price ranges, and location.

Although Yelp listed more locations when compared to TripAdvisor, it does not comprise a complete list of all establishments and associated data, and the data that is listed is not guaranteed to be correct. It was found that some establishments had been misclassified, for example restaurants classified as pubs and vice versa. Since only the establishments classified as "Pubs" were used in our analysis, some locations could be lost because of the misclassification. Since both TripAdvisor and Yelp are editable by the public, and each establishment must register their business before they can be listed, it is possible that the information listed could contain biased or incorrect information. It is also possible that certain types of establishments are not listed as they do not believe their rating to be necessary for good business, or in the case of student union bars, not relevant to the general public.

When retrieving the information, the most significant limitation found was that only a maximum of 1,000 records from Yelp could be obtained for each search keyword using the Python code. It was therefore necessary to use several keywords to ensure all establishments falling within the search requirements were downloaded. The necessary data was obtained using the following keywords: "Pubs + London" and "Pubs + Russell Square", to ensure all pubs closest to UCL were included in the dataset.

As mentioned previously, the Greedy Algorithm is also limiting as it often results in the generation of pub crawl routes with long walks between pubs, since it selects pubs based solely on proximity from the previous location. In some of the generated routes, these distances are easily large enough to discourage people from continuing the pub crawl. Although it was possible to manually influence the desired orientation of the bar crawl by analysing the locations with higher densities of pubs, and then removing the unwanted pubs from the dataset, this method requires significant human input and is not suitable for the automatic generation of pub crawl routes based on pre-selected preference criteria.

B. Suggestions and Future Research

Considering the limitations of this research there is much potential for further study. Firstly, the data obtained from social media channels could be improved using qualitative methods regularly applied in research by social scientists (WHAT METHODS?). A simple survey was conducted for this study, however, it had a very low response rate. Future research on the topic could benefit from an extended survey amongst students, tourists, and other potential groups of interest, to better determine the preferences of potential pub crawlers. Instead of using just one open question, the questionnaire could include multiple questions that would investigate the specific preferences of people in a quantifiable manner. Moreover, the information collected from both TripAdvisor and Yelp could potentially be merged into one dataset and further improved by obtaining and incorporating the data available on Google, Facebook, Twitter, and other internet-based review sites.

Another potential improvement would be to use different algorithms to create pub crawl routes. As discussed previously, the Greedy Algorithm was appropriate for this research due to its simplicity. However, in some cases it created long routes that could potentially discourage people from continuing the pub crawl. Therefore, a more complex algorithm, or a combination of algorithms, could be used improve the pubs selection process, taking into account the number of routes available from potential future stops in order to avoid reaching a dead end, where no further pubs are available within an acceptable walking distance. In addition, a new method is required to influence the route of the generated pub crawl based on the locations with the highest density of pubs that both meet the preference criteria, and are open late in the evening. Assigning pubs a ranking based on proximity to other open pubs, and then using this in the selection criteria could potentially do this.

VII. REFERENCES

- British Heart Foundation *Walks and treks - FAQs*, [Online], Available: <https://www.bhf.org.uk/get-involved/events/training-zone/walking-training-zone/walking-faqs> [09 Nov 2016].
- Carnage UK *Carnage UK*, [Online], Available: <https://carnageuk.com/> [06 Nov 2016].
- Cormen, T., Leiserson, C., Rivest, R. and Stein, C. (2001) *Introduction to Algorithms*, MIT Press.
- Dictionary.com Unabridged *Pub-Crawl*, [Online], Available: <http://www.dictionary.com/browse/pub-crawl> [04 Nov 2016].
- Dodd, V.J., Khey, D.N. and Miller, E.M. (2012) 'Intoxication Levels of Bar Patrons at an Organized Pub Crawl in a College Campus Community', *American Journal of Criminal Justice*, vol. 37, no. 2, June, pp. 246-257.
- ESRI (2016) *About the ArcGIS Network Analyst extension tutorial*, [Online], Available: <http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/network-analyst/about-the-network-analyst-tutorial-exercises.htm> [14 October 2016].
- Guinness World Records (2016) *Most people on a pub crawl*, [Online], Available: <http://www.guinnessworldrecords.com/world-records/most-people-on-a-pub-crawl> [04 Nov 2016].
- Influence, S.o. (2015) *PubRally*, [Online], Available: <http://pubrally.com/> [11 November 2016].
- Karlsruhe, G.G. (2016) *Geofabrik*, [Online], Available: <http://www.geofabrik.de/> [14 October 2016].
- Kipfer, B.A. and Chapman, R.L. (2007) *The Dictionary of American Slang*, 4th edition, New York, USA: HarperCollins.
- Moussaïd, M., Perozo, N., Garnier, S., Helbing, D. and Theraulaz, G. (2010) '2010', *PLoS ONE*.
- OpenStreetMap Foundation (2016) *About OpenStreetMap*, [Online], Available: <https://www.openstreetmap.org/about> [14 October 2016].
- Oxford English Dictionary, [Online], Available: https://en.oxforddictionaries.com/definition/pub_crawl [04 Nov 2016].
- Quigg, Z., Hughes, K. and Bellis, M.A. (2013) 'Student Drinking Patterns and Blood Alcohol Concentration on Commercially Organized Pub Crawls in the UK', *Addictive Behaviors*, vol. 38, no. 12, December, pp. 2924-2929.
- Sönmez, S., Apostolopoulos, Y., Theocharous, A. and Massengale, K. (2013) 'Bar crawls, foam parties, and clubbing networks: Mapping the risk environment of a Mediterranean nightlife resort', *Tourism Management Perspectives*, vol. 8, October, pp. 49-59.
- Thurnell-Read, T. (2011) "Common-Sense" Research: Senses, Emotions and Embodiment in Researching Stag Tourism in Eastern Europe', *Methodological Innovations Online*, vol. 6, no. 3, pp. 39-49.
- TripAdvisor (2016), [Online], Available: www.tripadvisor.co.uk [21 Oct 2016].

Tutenges, S. (2015) 'Pub Crawls at a Bulgarian Nightlife Resort: A Case Study Using Crowd Theory', *Tourist Studies*, vol. 15, no. 3, November, pp. 283-299.

Yelp (2016), [Online], Available: www.yelp.co.uk [31 Oct 2016].