GI_Forum 2020, Issue 1
Page: 137 - 152
Full Paper
Corresponding Author:
andreas.petutschnig@sbg.ac.at
DOI: 10.1553/giscience2020_01_s137

Exploratory Spatiotemporal Language Analysis of Geo-Social Network Data for Identifying Movements of Refugees

Andreas Petutschnig[1], Clemens Rudolf Havas[1], Bernd Resch[1,3], Veronika Krieger[1] and Cornelia Ferner[2]

[1]University of Salzburg, Austria
[2]Salzburg University of Applied Sciences, Austria
[3]Harvard University, USA

Abstract

Refugee movements in recent years have caused enormous challenges for relief organizations and public authorities, but especially for refugees themselves. Organizations which have to allocate their resources to regions where large groups of arrivals are expected struggle to prepare the refugees' admission, transfer, care and accommodation in time. Events like the refugee movement of 2015/16 in Austria and Germany in the wake of the Syrian civil war have shown that many of these issues are caused by a lack of up-to-date information about logistical requirements. We evaluate various methods to acquire this information that utilize semantic, spatial and temporal features to analyse geo-social network data. A multimodal analysis of these features leads to information about refugee movements across borders and regions. Approaches based on user trajectories and attempts to identify refugees by the language they used showed little promise, whereas using spatiotemporal aggregation and hotspot analysis of keyword-based filtered data allowed us to retrace refugees' collective movement patterns. Using temporal bins, we were able to detect changes in these patterns caused by external factors such as border closures.

Keywords:

language, refugees, social media, GSND, ESDA

1 Introduction

The phenomenon of refugee movement is inherently geographical (Lewis, 1982), but it has also been studied from a variety of other viewpoints, focusing on the causes of flight (Warner, 2009; Black et al., 2011; Mueller et al., 2014), effects (Jacobsen, 1997; Garfi et al., 2009; Biswas & Tortajada-Quiroz, 1996), and demographic aspects (Randall, 2005; Greenwood, 1997). Because the phenomenon is so multifaceted, it warrants an analysis which incorporates multiple modalities.

The event under investigation is the refugee movement in 2015/16 during which over 2.5 million people[1] fled to Europe from war-torn countries, mostly in northern Africa and western Asia. During that time, it was largely unknown when, where and how many refugees would appear at European borders, which made planning the necessary distribution of goods and personnel challenging for relief organizations and public authorities. This led to authorities at many border regions being overwhelmed by the necessary logistics at such short notice, which in turn led to humanitarian and societal problems (Razum & Bozorgmehr, 2015; Breen, 2016).

One way to mitigate these problems is to provide the information needed for resource planning based on an up-to-date picture of the situation, which in turn necessitates comprehensive, near real-time information. Information about refugee movements, especially in regions around borders, derived from new data sources like social networks, news outlets or crowdsourcing platforms, fits these requirements because of the potentially high spatial and temporal resolution of the data.

The methods and findings presented in this paper are a contribution towards building a system that employs a geo-social network data (GSND) source, processes the data, and provides relevant up-to-date information to relief organizations. For this approach to be useful requires the ability to extract information about refugee movements from the GSND. We outline methods to explore what information we can derive from such data, and present a multimodal analysis approach in which we collected georeferenced posts from the social media platform Twitter and analysed their spatial, temporal and semantic characteristics.

We begin by defining and checking the assumptions based on which we perform the analysis. For example, we need to check whether there are enough data available to make reasonable predictions at all. We further identify the languages used in the text data to examine whether Twitter users in Arabic-speaking countries actually use Arabic as their language of communication. This is critical because we use Arabic-language data as a proxy for populations originating from Arabic-speaking countries, who are the focus of this study. As the text corpus also contains other languages, mostly English, we include a selection of words from other languages in the keyword-based analysis as well. Using this setup, we aim to understand the potential for, and limitations of, detecting collective refugee movement patterns in a multi-disciplinary approach, extracting and combining information from geographic, temporal and semantic space. We also describe the measures we employed to preserve the privacy of individuals represented in the data.

2 Related Work

Besides food and shelter, smartphones are one of the essentials for refugees (Matthew, 2015) on their way to their destinations. Smartphones allow refugees to connect to social media networks where they can gather information and share their own experiences within their network. Refugees connect to various social media networks such as Instagram or the now

---

1 https://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tps00191&plugin=1

defunct Google+, but Facebook and Twitter are the most popular networks in this user group. The information obtained from social media networks is used by refugees for decision-making on their way to their destinations or for planning where to settle (Dekker et al., 2018). Further, they also use smartphones to plan and document their journeys, and to contact friends, family and people who will help them to get to their desired destinations. Although they are useful, smartphones also hold many dangers for refugees, such as surveillance by other actors. Therefore, refugees frequently use encrypted or closed communication channels, like closed Facebook groups or encrypted WhatsApp messages (Gillespie et al., 2018). Although refugees must use social media networks carefully, they are the main resource for communication, which makes them a generally suitable data source for data analysis in the context of refugee movements. This allows us to lean on the principles of collective sensing as described by Resch (2013) for data collection and analysis.

Many challenges associated with the recent refugee movements in Europe have been pointed out, including humanitarian protection (Ostrand, 2015), policy making (Guild et al., 2015), public health systems (Catchpole & Coulombier, 2015) or news coverage (Chouliaraki & Zaborowski, 2017). One of the factors leading to such problems is the lack of real-time information about refugee movements. Curry et al. (2019) discuss the potential and challenges of alternative data sources such as Volunteered Geographic Information or social media data. They present multiple results from analysing data from social media networks such as Flickr or Instagram, which they used to detect refugee-related activities in Europe. They conclude that social media are a crucial data source for information associated with mass migration and can help fill the information gap between authoritative information products and the actual situation. Hübl et al. (2017) performed an exploratory spatial data analysis (ESDA) on a Twitter dataset to demonstrate the value of GSND analysis for refugee movements. In the context of the refugee crisis in 2015, they detected and visualized generalized trajectories from the Middle East and northern Africa to Europe. In order to filter their dataset, they used language-dependent keywords for German, Italian and Greek. They were able to identify only a few potential refugees as Twitter users, and consequently only a limited number of trajectories.

In many cases, the triggers for the refugee movements we are observing here were terror and war. Consequently, ethical and responsible handling of data, and the presentation of results in such a way as to protect the research subjects are paramount, as most refugees fear surveillance and surveillance by others (Gillespie et al., 2016). Kounadi & Resch (2018) give an overview of potential privacy threats that come with GSND analysis, while Kounadi et al. (2018) have drawn up a set of practical guidelines and design principles for researchers who work with GSND to mitigate these threats.

3 Methods

The original dataset consists of 354,116,330 georeferenced tweets in the area between 8.0°E - 43.2°E and 28.2°N - 50.0°N, covering the period January 2014 to January 2020. Spatially and temporally the data cover the 'Balkan route', an informal land route connecting Greece and central Europe, which many refugees took during the refugee movement in 2015/16. The time frame of the dataset exceeds the actual event in order to provide a baseline before and after the event as context for interpretation. All our Twitter data are georeferenced (i.e., each tweet is linked to a geographic coordinate). This geolocation is only possible if the user chose to make their location public before tweeting. No additional georeferencing was carried out, because this would have introduced a bias. With this reliable reference in place, we can observe the number of tweets, the proportions of tweets using the different languages for each country, as well as the shift of languages or of relevant tweets in time. Besides coordinates, the data we used include a numeric user ID generated by Twitter, a timestamp, and the text of the message.

To be able to draw meaningful conclusions from the data, we used different methods for our analysis. We aggregated the Twitter data such that the data were grouped semantically by language, spatially by hexagonal bin or by the country from which the tweet was sent, and temporally using weekly bins for the timestamps. The aggregated data then served as input for further analysis. Language detection was carried out by applying the language detection module of the Python library Polyglot[2] on the tweets' texts. Polyglot can detect up to 196 languages, including Arabic, modern Greek, Turkish, German and English, which are the predominant languages used in tweets from the area between the Near East and Central Europe. The result of the language detection for each tweet is a probability distribution of the most probable languages used in the text. Due to a tweet's short length, there is usually just one language that exhibits dominant probability, and this is the language which we assigned to the tweet.

To learn about the spatial and temporal distribution of the data, we counted the yearly number of tweets within each country. This step was necessary to determine which assumptions about the data we could make in our analysis. Similarly, we counted the number of tweets grouped by language in each country. This way we could determine the composition of languages used within each country and whether this composition changes over time, which may in turn reflect changes in the population. As most of the refugees originated from predominantly Arabic-speaking countries, we designed part of our study to focus on the movement patterns of Arabic-speaking Twitter users. Our approach was similar to the one used by Bulbul et al. (2018) to identify Arabic-speaking refugees in Turkey. For this, we counted the number of Arabic tweets grouped by country and aggregated them in weekly bins. The resulting time series show signatures of changing use of the Arabic language, and therefore, by proxy, changes in the dynamics of the local Arabic-speaking population. For the keyword-based part of the study, we used refugee-related keywords. The study area covers a large region, which includes a number of language borders. In order to take into account the linguistic diversity in the Twitter data, we defined keywords in a set of locally used languages, which were defined and translated by language experts or native speakers who manually examined the text contents. The keywords are listed in Table 1. We matched tweets and keywords based on case-insensitive approximate string matching in the tweets' texts. We prepared the keyword-matched tweets for analysis by grouping them in hexagonal weekly bins and generating a series of maps for visual interpretation. The proportion of keyword-relevant tweets was 0.47% for German, 0.14% for English, 0.14% for Hungarian, 0.33% for Serbian or Croatian, 0.25% for Greek, and 0.58% for Arabic.

2 http://polyglot.readthedocs.org

Table 1: Keyword overview

German	English	Hungarian	Serbian or Croatian	Greek	Arabic
Flüchtling	refugee	migráns	Sirija	πρόσφυγες	لاجئ
Fluechtling	migra	migrans	Сирија	προσφύγων	لاجئين
Flucht	syria	Szíria	izbeglice	πρόσφυγας	لاجئون
Migrant	border*cross	sziria	izbjeglice	μετανάστης	مهاجر
Syrien	cross*border	Soros	избеглице	μεταναστες	مهجر
Grenz		StopSoros	migranti	μετανάστες	مهجرون
Asyl		brüsszel	мигранти	Συρία	مهجرين
unbegleitete Minderjährige		bruesszel	migracije	Βρυξέλλες	لاجئة
unbegleitete Minderjaehrige		Brusszel	миграције	Prosfuga	لاجئات
Schlepper		O1G	азил	Prosfiga	مهاجرة
Willkommenskultur		kvóta	granica	Prosfyga	مهجرة
Erstaufnahmezentrum		kvota	граница	Prosfuges	مهجرات
Balkanroute		fidesz	Brisel	Prosfiges	سوريا
		bevandorlas	Брисел	Prosfyges	لاجئ
		bevándorlás	Батровци	Prosfugwn	لاجئين
		keritesepites	Batrovci	Prosfigwn	لاجئون
		kerítésépítés	Хоргош	Prosfygwn	ملجأ
		ahataron	Horgoš	Metanastis	مأوى
		ahatáron	azilanti	Metanasths	مهاجر
		migráncs		Metanastes	مهجر
		bevandored		Metanastwn	مهجرون
		menekült		Brixelles	مهجرين
		dzsihadista		Bruxelles	حدود
		dzsihadistak		Synora	حدودية طة
		terrorista		Sunora	حدودي
		határainkat		Πρόσφυγ	معبر
		hatarainkat		Προσφύγ	لجوء
		védyük meg		Μετανάστ	أزمة
		muszlim		Μεταναστ	السورية زمة
		muzulman		Prosfug	السورية حرب
				Prosfig	البلقان ق
				Prosfyg	مخيم
				Metanast	الركبان
					الزعتري

4 Results

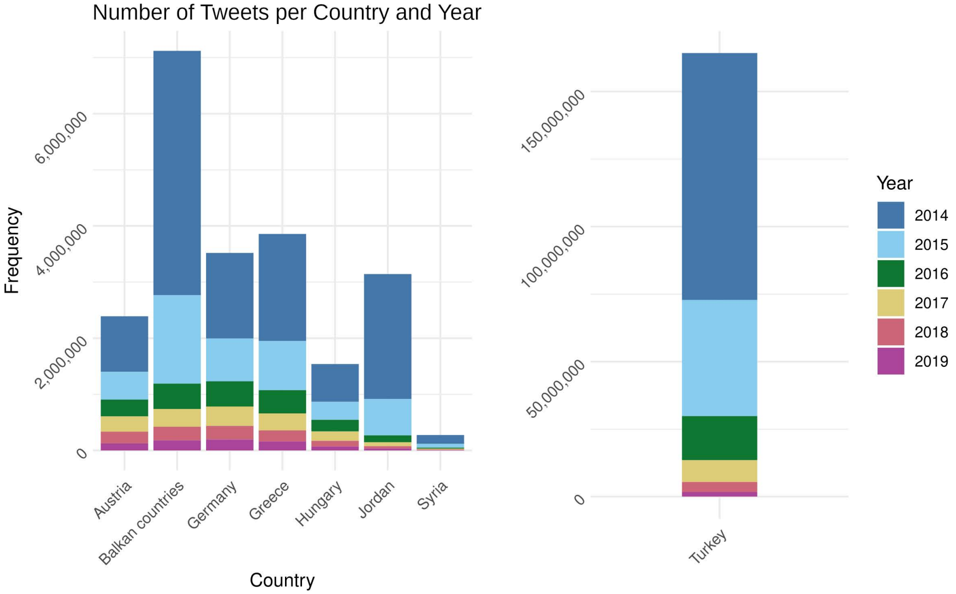

Figure 1: Number of tweets per country and year

The focus in this analysis is on the countries through which people passed on their way from Syria to Germany. Because the Balkan countries of Albania, Bosnia and Herzegovina, Croatia, Kosovo, North Macedonia, Montenegro, Serbia and Slovenia individually gave very little data, we aggregated them in some results under the term 'Balkan countries', for readability. The results show that in general the number of tweets collected is higher in 2014/15 than in later years. This is not necessarily attributable to changes in user behaviour: it may be affected by Twitter's data-sharing policy and the data collection process. However, one can see a change in the pattern for the different countries in Figure 1. For better readability, the chart for Turkey is presented separately, as significantly more tweets were collected in Turkey than in the other countries.

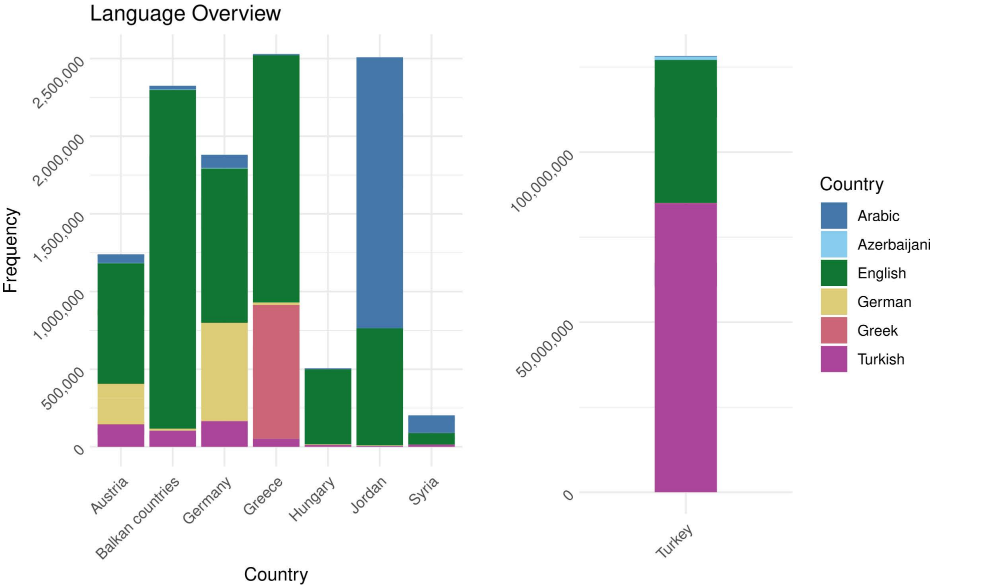

Figure 2: Number of tweets per language and country for 2014–2020

Figure 2 shows the number of tweets in Arabic, Azerbaijani, English, German, Greek and Turkish, which are overall the most dominant languages across the listed countries. All countries contain Arabic tweets, but in Turkey, Greece and Hungary their numbers are marginal.

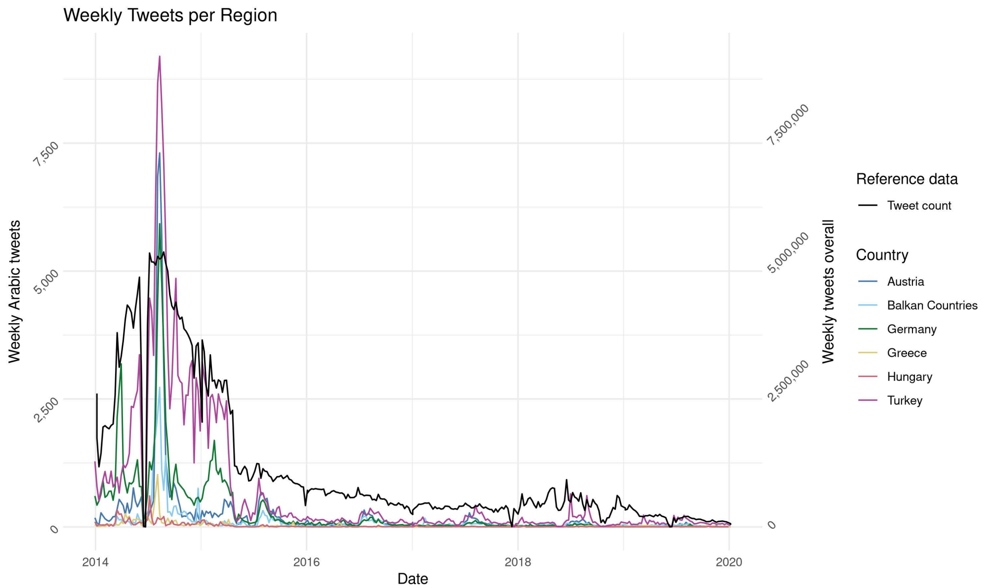

Figure 3: Time series of Arabic language tweets and overall tweet counts from 2014 to 2020 per country (note the secondary y-axis for total tweet counts)

Figure 3 shows a development of the number of Arabic tweets between 2014 and 2016. The peaks in the data mostly appear simultaneously in a seemingly seasonal interval throughout the countries with few exceptions. The grey line shows the total number of weekly tweets, providing an indication for how strongly the peaks deviate from the baseline. Within the group of the Balkan countries, the temporal signatures appear to be similar, but substantiated conclusions are difficult due to the relatively sparse data.

The next method we used aimed at the identification of trajectories to derive routes that are shared by larger numbers of refugees. A trajectory consists of a minimum of two tweets from the same user ID. To identify trajectories that meet our criteria for country of origin and time frame, we filtered our dataset using spatial and temporal constraints. We removed bots from the dataset by restricting the average number of tweets per day to 15, which we defined based on empirical evidence. We defined the start of the trajectory to be east of Greece, and the average distance between two tweets was chosen with the aim of eliminating 'extreme' values – i.e. to filter out small-scale as well as unrealistically long distances, such as tourists using planes to reach their destinations. The remaining data were assessed manually. The resulting trajectories during the major refugee movement period are shown in Figure 4. Visual analysis of the results did not yield any usable results because of the large number of trajectories with very low sampling rate. In addition, upon visual inspection of the text contents, we found that none of the trajectories were likely to belong to a refugee, but rather to either previously undetected bots or users such as tourists, business people, or journalists, who were not the focus of this study.

Trajectories (2015/2016) after Rule-based Filtering

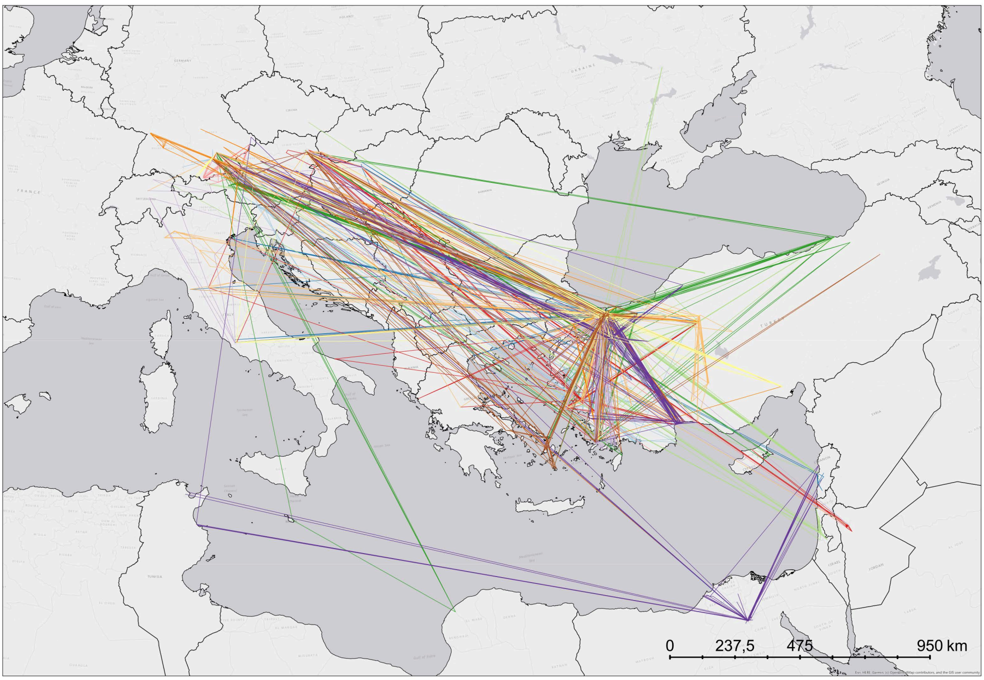

Figure 4: Coloured trajectories differentiating between individual users

Another possibility to detect refugee movements is to visualize the number of refugee-related tweets in an area of interest. The hypothesis is that if an unusually high number of refugees are passing through an area, people will refer to this on social media networks. We used the refugee-related keywords from Table 1 to identify relevant tweets and aggregated them spatially in hexagonal cells in the area of interest. We also binned the data temporally by week to detect temporal changes.

We observed that at the beginning of 2015 the refugee-related tweets were mostly sent from near the Greek coast. Later, more activity was observed at the border between Turkey and Greece, as shown in Figure 5. Figure 6 shows how, over time, the tweet frequency increased at the Greek and Bulgarian borders with Turkey. From August onwards, the number of refugee-related tweets rose in Serbia and Hungary. Figure 7 shows that between Belgrade through Hungary and Austria to Munich, the number of refugee-related tweets was high, which is in line with the actual path the refugees used as the primary route to central Europe during this specific time frame.[3] Furthermore, the tweet activity along the coast of Turkey and at the border with Greece is still high, comparable to the preceding months. This observation concords with articles that report consistent numbers of refugees arriving in Greece in 2015.[4]

Weekly Number of Refugee-related Tweets (2015-06-01 to 2015-06-08)

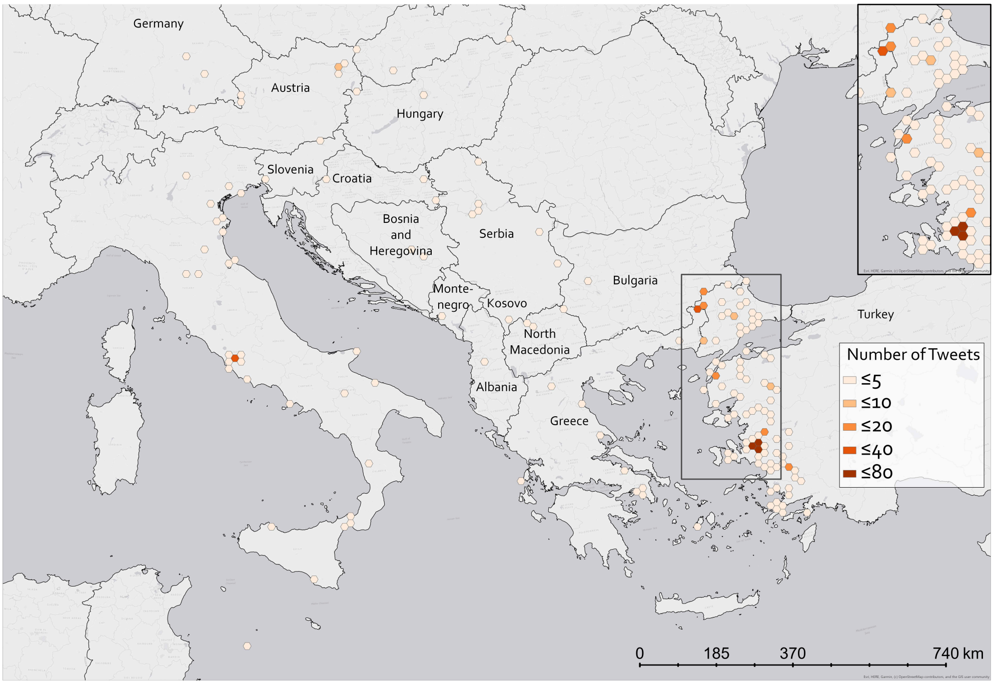

Figure 5: Weekly hexagonal aggregation of relevant tweets from 2015-06-01 to 2015-06-08

3 https://www.unhcr.org/news/latest/2015/9/55f0230e6/frustrated-refugees-migrants-serbia-hungary-border-seek-escape-poor-reception.html
4 https://www.unhcr.org/news/latest/2015/10/560e63626/refugee-sea-arrivals-greece-year-approach-400000.html

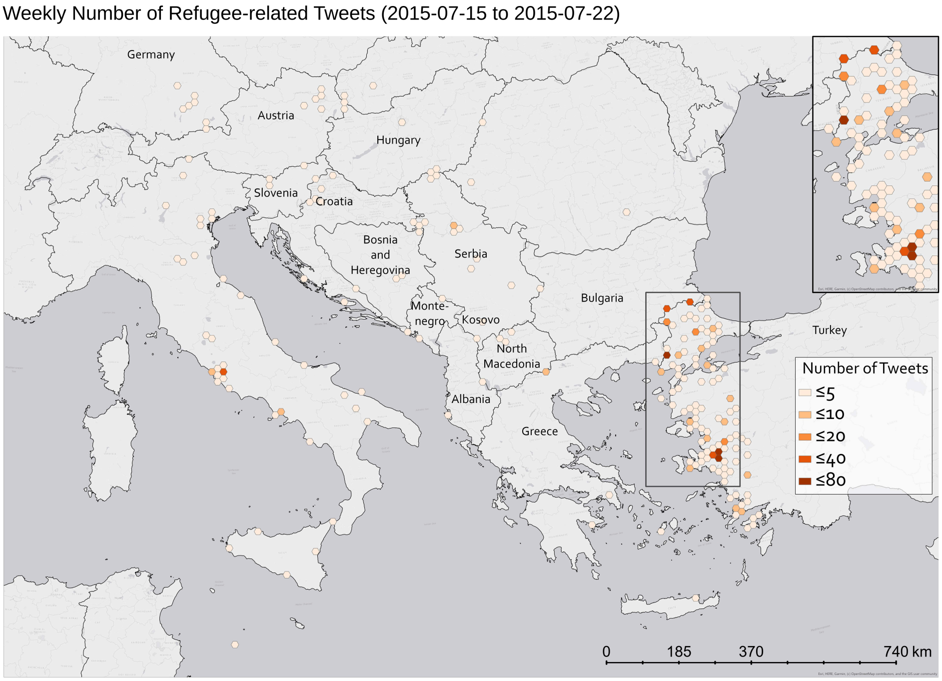

Figure 6: Weekly hexagonal aggregation of relevant tweets from 2015-07-15 to 2015-07-22

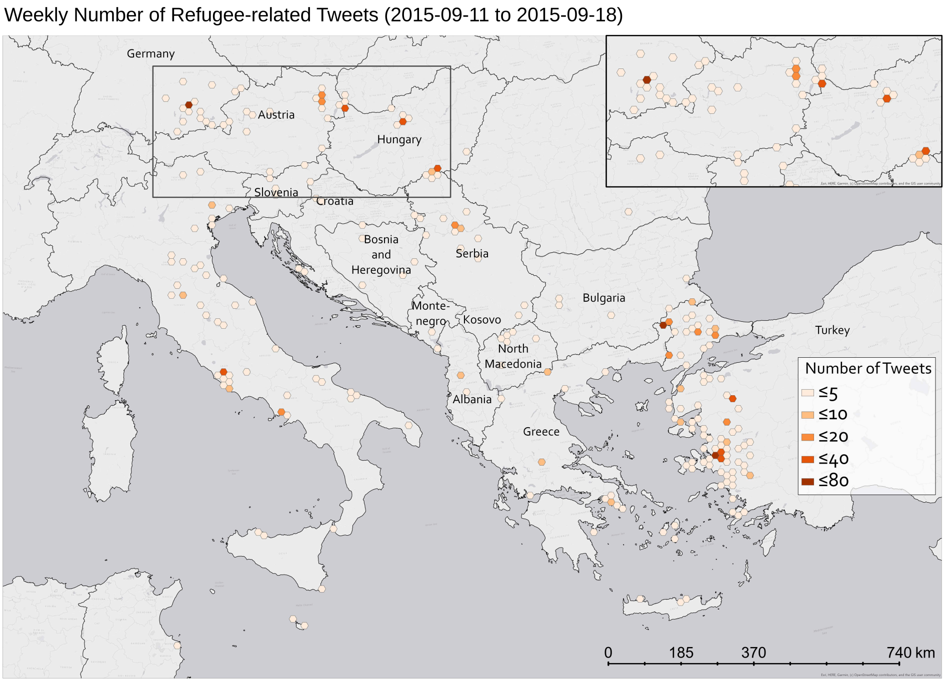

Figure 7: Weekly hexagonal aggregation of relevant tweets from 2015-09-11 to 2015-09-18

Spatiotemporal binning of the data provides a reasonable basis for drawing some conclusions from the data. However, there are several limitations to this approach. Firstly, it does not consider the cell's spatial neighbourhood, which results in outliers being visually very prominent on the map. Secondly, because social media activity in urban areas is higher than in rural areas, the raw counts for the two types of area cannot be compared meaningfully.

The spatial neighbourhood can be included by applying a hotspot analysis to the dataset. A hotspot analysis based on Getis-Ord $G_i^\star$ (Ord & Getis, 1995) identifies statistically significant cold- and hotspots in a dataset by including the values of the spatial neighbourhood of each cell. For the hotspot analysis, we defined the ratio between the refugee-related and the unfiltered social media as values in a fishnet grid. We chose the cell size based on the surface of the study area A and number of Tweets n (Wong & Lee, 2005) and adapted it based on the visual interpretation of results, resulting in cells with a side length of $l = \frac{1}{12} \cdot \sqrt{2\frac{A}{n}}$. We selected weekly bins experimentally because they are short enough to capture the large-scale refugee movement events under investigation and long enough to be reasonably robust against outliers. Figures 8 and 9 show two hotspot maps derived from the Twitter data which capture the situation before and after the Hungarian government closed their border with Serbia,[5] which in turn led to refugees using alternative routes via Croatia and Slovenia.[6] This results in a shift of hotspots from Hungary to the neighbouring countries, as seen on the maps.

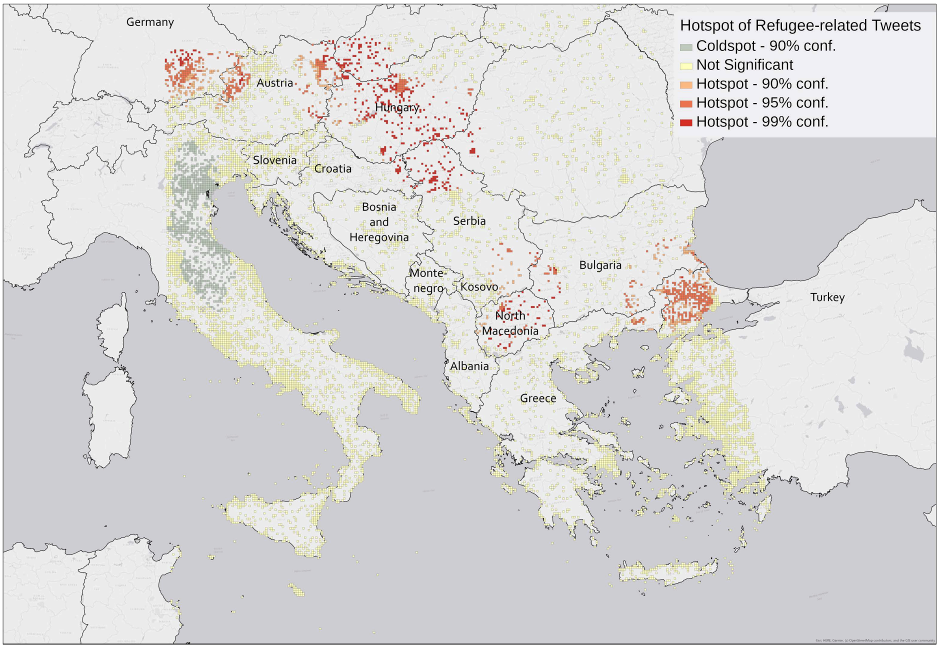

Figure 8: Weekly hotspots of aggregated relevant tweets from 2015-09-04 to 2015-09-11

5 https://www.bbc.com/news/world-europe-34252812

6 https://www.theguardian.com/world/2015/sep/19/young-migrants-trailblazers-hungary-croatia-serbia

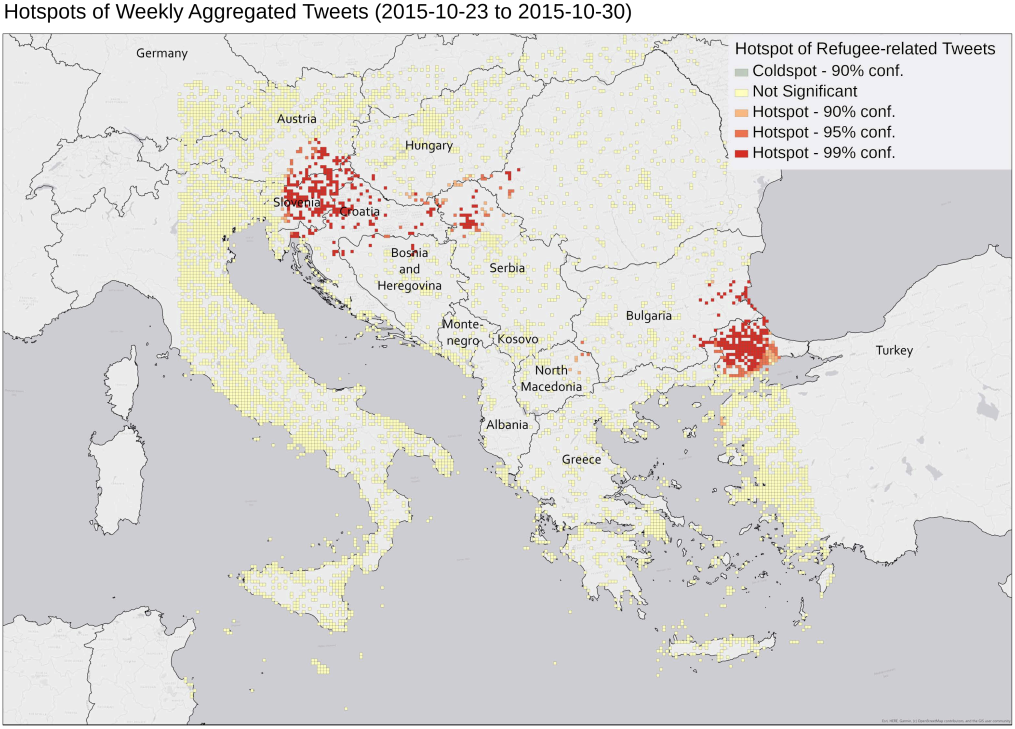

Figure 9: Weekly hotspots of aggregated relevant tweets from 2015-10-23 to 2015-10-30

5 Discussion and Conclusion

We conclude that Twitter data can be used as an indicator for refugee movements in the above scenario. Our findings are based on the large amount of GSND available in our area of interest and on the assumption that many users from Arabic-speaking countries use the Arabic language on social media, which we were able to confirm. We also found that the results vary strongly depending on the methods used. While reliable refugee movement trajectories would provide a good basis for the identification of collective movement patterns in given areas of interest, we were not able to derive the patterns from our data. Upon checking the contents of the tweets from which we constructed the trajectories, we found the authors to be business people, tourists, reporters or bots, instead of the expected refugees. One explanation for the lack of refugee trajectories in the results is the fact that many refugees avoid public communication channels in order to remain inconspicuous for safety reasons. They avoid sharing their location or do not use public social networks at all, opting instead for private communication platforms like WhatsApp or Facebook's Messenger. Findings such as this underline the importance of not relying solely on initial assumptions about the data, and the need to carefully check what new aspects of the data are revealed during the ESDA. The fact that our manual checking revealed some bot-generated messages mixed in with our results tells us that our threshold-based bot-filter approach would benefit from being combined with other more stringent bot-detection methods.

The time series analysis of Arabic-language tweets showed that even though we were able to identify large numbers of Arabic-language tweets in the observed countries, we were not able to clearly identify refugee traces from them. We noticed a series of peaks that appeared almost simultaneously across countries. Overlaying the overall tweet counts of the time period as a baseline, we ruled out the possibility that the peaks were a result of sampling irregularities. Upon manual inspection of the Arabic tweets in question, we found that most of them were posted by tourists who were visiting the regions concerned. This observation is also in line with the time of year, and the interval of approximately one year between peaks. Possible explanations for the lack of refugee-related data are, as already stated above, that refugees do not use public communication channels much, or at least refrain from using the Arabic language on these media.

Some of the most promising results were achieved by grouping the keyword-relevant tweets in weekly hexagonal bins and creating a series of maps. These offer the advantage of high spatial and temporal resolution, which makes them valuable for relief organizations and public authorities who might benefit from this study. The use of keywords for the identification of refugee-relevant tweets also means that we are not restricting our search to tweets created by refugees themselves, but are analysing tweets by a broader range of users engaging in the topic.

The hotspot maps based on the keyword-related tweet counts showed promising results as well. Because they not only report the absolute difference in numbers between rural and urban areas but also consider a cell's neighbourhood, they made some less frequent but locally strong hotspots stand out visually. This allows us to observe the small changes in keyword-related tweet counts that we are interested in. The maps in Figures 8 and 9 show good examples of this effect in North Macedonia, where only a few relevant tweets were located, but because of their homogeneous neighbourhoods the cells appear as hotspots. The opposite effect is visible in parts of Turkey, where the overall numbers are consistently high, but despite that, large areas do not contain significant hotspots. Such differences can also be observed on the temporal scale. The maps show the shift of refugee indicators around Hungary following the closure of the border between Serbia and Hungary.

From this we were able to retrace movements with a very good fit to what we were expecting based on reports from news sources or NGO reports about refugee arrivals, including specific destinations like cities in which many refugees arrived. Our approach also allowed us to partly retrace the paths that many people used in relatively remote regions. However, because the spatial and temporal resolution of our results is high, we have no reference data available for direct comparison, as statistics agencies and NGOs provide their data at a lower resolution. We therefore had to rely on anecdotal evidence in the form of news articles to confirm our findings.

For the end-users of information about refugee movements, it is important to have information available as early as possible and in a format that is easy to understand. Because the information is derived from data that are available as a constant stream, we can meet the time requirement by automating the data processing steps and operating a constantly running service. Based on our results, we believe that providing an interactive web map interface that displays spatially and temporally aggregated keyword-based tweet counts and hotspot maps along with additional relevant context information as map layers would be a good way to communicate an overview of the situation in an easily accessible and interpretable manner.

It is essential that individuals must not be identifiable from the results. Therefore personally identifying characteristics were not stored in the data, following the principle of data economics and the guidelines developed in Kounadi & Resch (2018). Additional privacy protection measures are the spatial and temporal aggregation of results, with the intention of obscuring individual movement traces.

As the results are largely exploratory, the next steps for our work will focus on the development of a more sophisticated methodology for semantic information extraction to identify refugees in social media and messages that contain information about refugee movements. Potential approaches include supervised learning methods like convolutional (Dos Santos & Gatti, 2014) or recurrent neural networks (Lee & Dernoncourt, 2016), or unsupervised techniques like Latent Dirichlet Allocation (Qiang et al., 2016), as used in Resch et al. (2018), or dependency parsing (Di Caro & Grella, 2013). Furthermore, semantic analysis needs to be more dynamically integrated with spatial and temporal analysis methods such as spatiotemporal autocorrelation, and techniques for spatiotemporal analysis of semantically homogeneous user networks, which may be an indicator for collective refugee movements.

When using GSND as a proxy for an underlying population, because of the heterogeneous usage of these platforms we cannot assume that this population is represented uniformly (Duggan & Maeve Brenner, 2013; Tufekci, 2014). This affects the interpretation of our results insofar as we have to be aware that they only apply to a specific subset of users. Despite this limitation, GSND is still a valuable data source for many spatiotemporal data analysis applications, many of which were catalogued by Steiger et al. (2015). There are numerous possibilities to extend our research. Comparing our results with equally highly sampled official migration time series data from the Balkan countries, which are not available at this point, would be an important step towards statistical validation. The language and keywords a person uses are presumably not the only identifiers by which we can determine whether a person might be a refugee. Spatiotemporal patterns or specific regions that a person visits may also contain such information. Identifying, extracting and integrating such information in our models might significantly improve the validity of this research. Further, the language-based approach presented here could potentially be improved by including more languages that are being used in the area of interest. A related concern is the fact that Arabic can be written in both the Arabic and Latin alphabets. Our current method of language detection only recognizes texts written using Arabic letters, which potentially eliminates useful data. Lastly, applying our methods to similar situations in other parts of the world would give us insights into the spatial transferability of our approach.

Acknowledgements

This study was carried out in the HUMAN+ project, funded by the Austrian security research programme KIRAS of the Federal Ministry of Agriculture, Regions and Tourism (BMLRT), project number 865697. We would like to thank Harvard University's Center for Geographic Analysis for their support in providing us with the Twitter data for our study.

Preferences

Biswas, A. K., & Tortajada-Quiroz, H. C. (1996). Environmental impacts of the Rwandan refugees on Zaire. *Ambio*, *25*(6), 403–408. https://doi.org/10.2307/4314504

Black, R., Adger, W. N., Arnell, N. W., Dercon, S., Geddes, A., & Thomas, D. (2011). The effect of environmental change on human migration. *Global Environmental Change*, *21*(SUPPL. 1), S3–S11. https://doi.org/10.1016/j.gloenvcha.2011.10.001

Breen, D. (2016). Abuses at Europe's Borders. *Forced Migration Review*, *51*(January), 21–23. https://www.fmreview.org/destination-europe/breen

Bulbul, A., Kaplan, C., & Ismail, S. H. (2018). Social media based analysis of refugees in Turkey. *CEUR Workshop Proceedings*, *2078*, 35–40.

Catchpole, M., & Coulombier, D. (2015). Refugee crisis demands European Union-wide surveillance! *Eurosurveillance*, *20*(45), 30063. https://doi.org/10.1086/520454

Chouliaraki, L., & Zaborowski, R. (2017). Voice and community in the 2015 refugee crisis: A content analysis of news coverage in eight European countries. *International Communication Gazette*, *79*(6-7), 613–635. https://doi.org/10.1177/1748048517727173

Curry, T., Croitoru, A., Crooks, A., & Stefanidis, A. (2019). Exodus 2.0: crowdsourcing geographical and social trails of mass migration. *Journal of Geographical Systems*, *21*(1), 161–187. https://doi.org/10.1007/s10109-018-0278-1

Dekker, R., Engbersen, G., Klaver, J., & Vonk, H. (2018). Smart Refugees: How Syrian Asylum Migrants Use Social Media Information in Migration Decision-Making. *Social Media and Society*, *4*(1). https://doi.org/10.1177/2056305118764439

Di Caro, L., & Grella, M. (2013). Sentiment analysis via dependency parsing. *Computer Standards and Interfaces*, *35*(5), 442–453. https://doi.org/10.1016/j.csi.2012.10.005

Dos Santos, C. N., & Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis of short texts. *COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers*, 69–78.

Duggan, Maeve Brenner, J. (2013). *The Demographics of Social Media Users —2012* (Vol. 14). Pew Research Center's Internet & American Life Project. https://doi.org/10.1002/cd.23219957004

Garfi, M., Tondelli, S., & Bonoli, A. (2009). Multi-criteria decision analysis for waste management in Saharawi refugee camps. *Waste Management*, *29*(10), 2729–2739. https://doi.org/10.1016/j.wasman.2009.05.019

Gillespie, A. M., Ampofo, L., Cheesman, M., Faith, B., Iliadou, E., Issa, A., Osseiran, S., & Skleparis, D. (2016). *Mapping Refugee Media Journeys: Smartphones and Social Media Networks* (May).

Gillespie, M., Osseiran, S., & Cheesman, M. (2018). Syrian Refugees and the Digital Passage to Europe: Smartphone Infrastructures and Affordances. *Social Media and Society*, *4*(1). https://doi.org/10.1177/2056305118764440

Greenwood, M. J. (1997). *Chapter 12 Internal migration in developed countries* (pp. 647–720). https://doi.org/10.1016/S1574-003X(97)80004-9

Guild, E., Costelle, C., Garlick, M., & Moreno-Lax, V. (2015). The 2015 Refugee Crisis in the European Union. *CEPS Policy Brief*, *332*, 1–6. https://www.ceps.eu/system/files/CEPS PB332 Refugee Crisis in EU{_}0.pdf

Hübl, F., Cvetojevic, S., Hochmair, H., & Paulus, G. (2017). Analyzing refugee migration patterns using geo-tagged tweets. *ISPRS International Journal of Geo-Information*, *6*(10). https://doi.org/10.3390/ijgi6100302

Jacobsen, K. (1997). Refugees' environmental impact: The effect of patterns of settlement. *Journal of Refugee Studies*, *10*(1), 19–36. https://doi.org/10.1093/jrs/10.1.19

Kounadi, O., & Resch, B. (2018). A Geoprivacy by Design Guideline for Research Campaigns That Use Participatory Sensing Data. *Journal of Empirical Research on Human Research Ethics*, *13*(3), 203–222. https://doi.org/10.1177/1556264618759877

Resch, B. (2013). People as Sensors and Collective Sensing-Contextual Observations ComplementingGeo-Sensor Network Measurements. *Progress in Location-Based Services*, 373–388. https://doi.org/10.1007/978-3-642-34203-5

Resch, B., Usländer, F., & Havas, C. (2018). Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science*, *45*(4), 362–376 https://doi.org/10.1080/15230406.2017.1356242

Steiger, E., Albuquerque, J. P. de, & Zipf, A. (2015). An Advanced Systematic Literature Review on Spatiotemporal Analyses of Twitter Data. *Transactions in GIS*, *19*(6), 809–834. https://doi.org/10.1111/tgis.12132

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *Eighth International Aaai Conference on Weblogs and Social Media.*

Warner, K. (2009). *In Search of Shelter Mapping the Effects of Climate Change on Human Migration and Displacement Koko Warner , Charles Ehrhart , Alex de Sherbinin, Susana Adamo , and Tricia Chai-Onn. January*.

Wong, D. W. S., & Lee, J. (2005). *Statistical analysis of geographic information with ArcView GIS and ArcGIS*. Hoboken, NJ: Wiley.