Social Spatial: A Qualitative GIS for Social Big Data Investigations

Michael E. Martin*[1]

[1]University of Auckland
Department of Geography
302-419
23 Symonds St.
Auckland, New Zealand
*Email: Michael.martin@auckland.ac.nz

Abstract

In this paper we introduce Social Spatial, a qualitative GIS for social media and big data research. This software enables GIScience researchers to build social media corpus that reflects a phenomenon being researched and implement methods that analyse that corpus. Natural language processing methods are integrated into Social Spatial, and the code framework has been designed to allow for easy integration of further algorithms. The software builds upon the knowledge of the researcher to identify new ways that phenomena are expressed and see where these posts are geospatially. The software was released open-source with thorough internal documentation to a collaborative code repository that encourages contribution. This program seeks to demystify the analysis process of qualitative social media exploration by use of settings files that expose parameters – word lists, model coefficients, and stop words. These files enhance transparency in qualitative social media research methods and the code/spaces they were enacted within without increasing the burden of research documentation.

Keywords: Qualitative GIS, Natural Language Processing, Software, GISystems, Social Media, Big Data

1. Introduction

There is an increasing abundance of methods originating from computer science that allow GIScientists to push the boundaries of our capacity to capture and integrate big data. They will also potentially help us to process and understand social phenomena. By integrating computing science approaches to big data integration and analysis, we may finally realize the promises of a qualitative GIS. A GIS that is conversant with qualitative data and that reflect the persons who generate the data themselves in time, space, and place.

GIScience is poised to undertake research and produce tools that better reflect the physical and human landscape than perhaps any other discipline, yet there are concerns that if it does not do so soon, it will suffer the fate of being left behind (Kitchin, 2013). This article presents software (Figure 1) that builds upon the work of GIScientists in the field of big data, focusing on workflows and natural language methods. The software is open-source and actively under development. It is our belief that this software will contribute a space for further development for big data geospatial research tools.

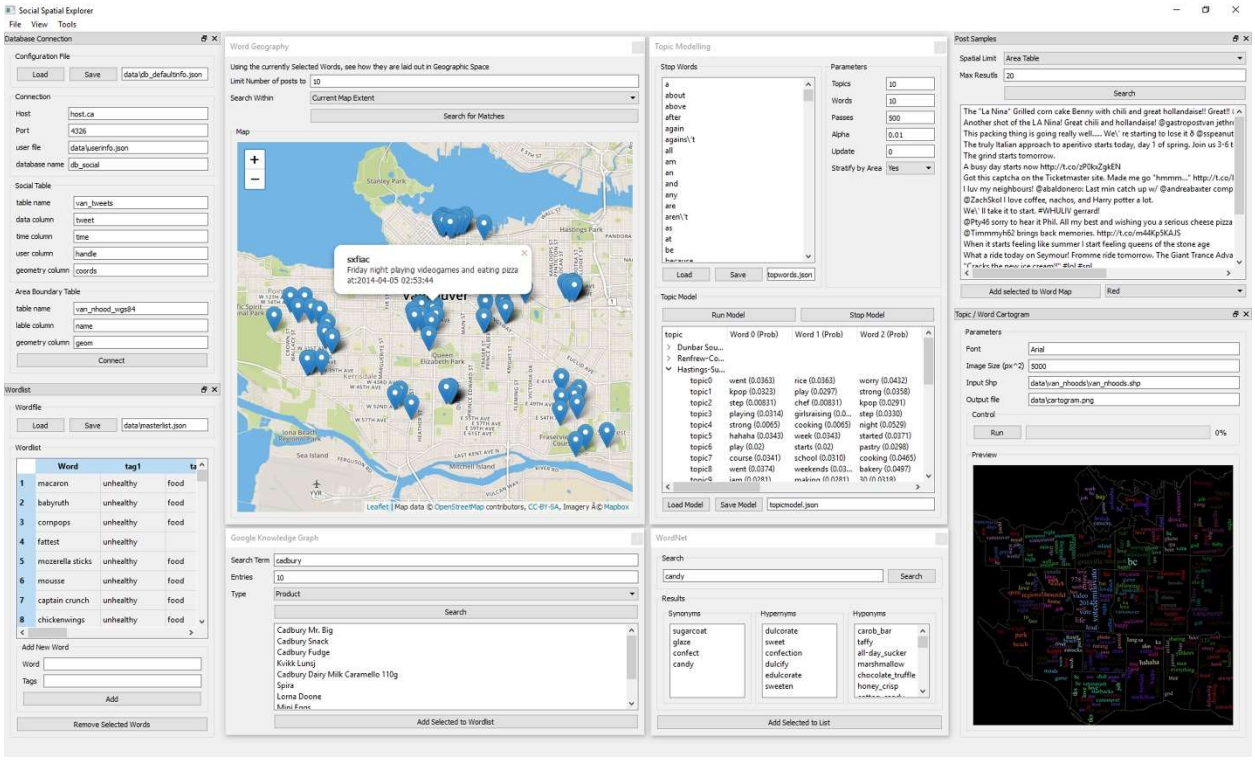

Figure 1: The Social Spatial Software Interface

2. Background

The last five years has seen a burgeoning of big data research in the geographical sciences. Geospatial web research (J. W. Crampton, 2009; Haklay, Singleton, & Parker, 2008) and participatory research (Jeremy W. Crampton et al., 2013; Elwood & Leszczynski, 2011; Johnson et al., 2015) led GIScience directly into the path of social big data. The automatic registration of location metadata through mobile phones has created a deluge of spatial data

and new methods for interrogating this information (Lee & Kang, 2015; Yeager & Steiger, 2013). At the same time, critical and qualitative GIS thinkers have followed these developments and contributed towards a better understanding of how these new data might be used (Elwood, 2008; Elwood, Goodchild, & Sui, 2013; Harvey, Kwan, & Pavlovskaya, 2006). Perhaps ahead of its time, Qualitative GIS (2009) by Cope and Elwood exemplifies how critical-qualitative researchers interacted with humanistic data, pre-twitter and other social media data. In the early 2000s, scholars explained the advantages (Pavlovskaya, 2009; Schuurman & Leszczynski, 2006) and opportunities (Jung, 2007; Kwan & Ding, 2008) qualitative GIS offers, but due to the absence of simple and automated data integration, the burden of data generation and analysis were prohibitive.

As geospatial social data sources have become prevalent, in-depth qualitative research has surfaced. GIScience scholars have used the tools that are already integrated in standard GIS environments to great effect (Crooks, Croitoru, Stefanidis, & Radzikowski, 2013; Stephens, 2013; Zook & Poorthuis, 2014).

- Jung (2015) demonstrates the challenge of integrating qualitative data in the face of the avalanche of information that Twitter provides.

- Stephens (2013) generated maps of hate using data from Twitter, thematically displaying words of hate-speech to heat maps (density) in a web browser.

- Zook (2014) et al. employ the use of odds-ratio analysis of keyworded social media in their investigations of the patterns of various phenomena, including those of American drinking habits.

Computer science has also contributed greatly (Cody, Reagan, Mitchell, Dodds, & Danforth, 2015; Frank, Mitchell, Dodds, & Danforth, 2013; Liu, Ester, Hu, & Cheung, 2015;

Mitchell, Frank, Harris, Dodds, & Danforth, 2013; Wang, Wang, Xie, & Ma, 2007). For qualitative GIScience and phenomenological GIScience researchers however, one computer science discipline stands out as particularly useful, natural language processing (Allen, 2003). Within the field, two important methodological techniques stand out as being pertinent to qualitative GIScience, Topic Modelling (Bauer, Noulas, Seaghdha, Clark, & Mascolo, 2012; Hao et al., 2010; Lansley, Adnan, & longley, 2015) and Sentiment Analysis (Frank et al., 2013; Larsen et al., 2015; Mitchell et al., 2013).

As recent history has shown, there has been an increasing coupling of contemporary computer science methods with GIScience. This coupling, along with the availability of geographic qualitative data, increases the capacity of qualitative GIScience to move beyond the conceptual stage of research and into the primetime methods of everyday research. However, a key ingredient is missing. At the present time, no clear methodologies or tools exist for aspiring qualitative GIScientists to employ.

3. Software Design

Social Spatial was designed around two specific goals:

1. Allow researchers to explore and analyse data using advanced methods and to create a program that is modular, flexible (figure 2 and 3)..
2. Provide access to methods that otherwise require an understanding of computer science to utilize.

While this article is not an in-depth summary of the capabilities of Social-Spatial, a brief explanation of its modules (figure 3) are offered here.

- Data Acquisition. Social Spatial currently assumes data is already gathered and stored in a PostGIS database.

- Data exploration. The data exploration tools facilitate building an extensive list of words used for generating a corpus of relevant postings. Social Spatial implements this via four modules:
    - The Wordlist provides functionality to grow and organize a set of keywords for specific thematic areas of inquiry for the qualitative study. Selection from the list allows the researcher to perform later analysis methods on matching social media posts.
    - The Google Knowledge Graph and Wordnet modules integrate semantic word discovery, enabling expansion of the wordlist beyond those they already know.
    - Topic modelling. This module acts as a mechanism for discovery of additional words. The tool can search for all postings within a single geographical boundary (i.e. a city) or iterate through sub-areas (i.e. neighbourhoods). Generating topics and topic words can identify new items to add to the wordlist that the researcher may be unaware of, be it because they are regionally, culturally, age, or otherwise specific beyond their knowledge.
    - The Post Samples and Word Geography. These provide a link between the wordlist and the source material. By selecting different words from the wordlist the user can use the post samples module to see the raw data and geospatial distribution of that data. The map interface can be exported as html/javascript written for Leaflet[1].
- Analysis. Analysis methods currently available and can be organized into three

[1] www.leafletjs.com

groups: social science, GIScience, and computer science.

- Social Science. Social Spatial familiar methods for social science authors by creating an interface to identify themes and sub-themes based on keywords. The Post samples module offers SQL output, making augmentation of conditional statements straightforward.

- GIScience. Currently, Social Spatial does not support many of the functions associated with geospatial analysis of social media, aside from overlay and data manipulation. However, it may not be advantageous to reimplement GIScience methods in Social Spatial, as export of the SQL statements used to generate posts is easy to do and can be copied directly to existing GISystems such as QGIS[2].

- Computer Science. Currently, Social Spatial offers topic modelling as a first step, via the Topic Modelling module. The settings and parameters of this module can be imported and exported, as well as the topics it produces. Sentiment modelling is currently under development.

- Visualization and Output. Social Spatial is currently able to generate two visual outputs and several data outputs.

  - Visually, Social Spatial can generate area-based topic models similar to those in Martin and Schuurman (2017). These area-based topic models produce PNGs that are easily integrated with existing GISystems. The Word Geography module produces html and javascript as described above.

[2] www.qgis.org

Comprehensive layout functionality is not currently in development.

- Social Spatial makes use of json objects to store and operate on all parameters, settings, and data. This format makes interoperability easy with many other programs, and it enables researchers to open access to all the internally configured elements of their research, aiding in replicability.

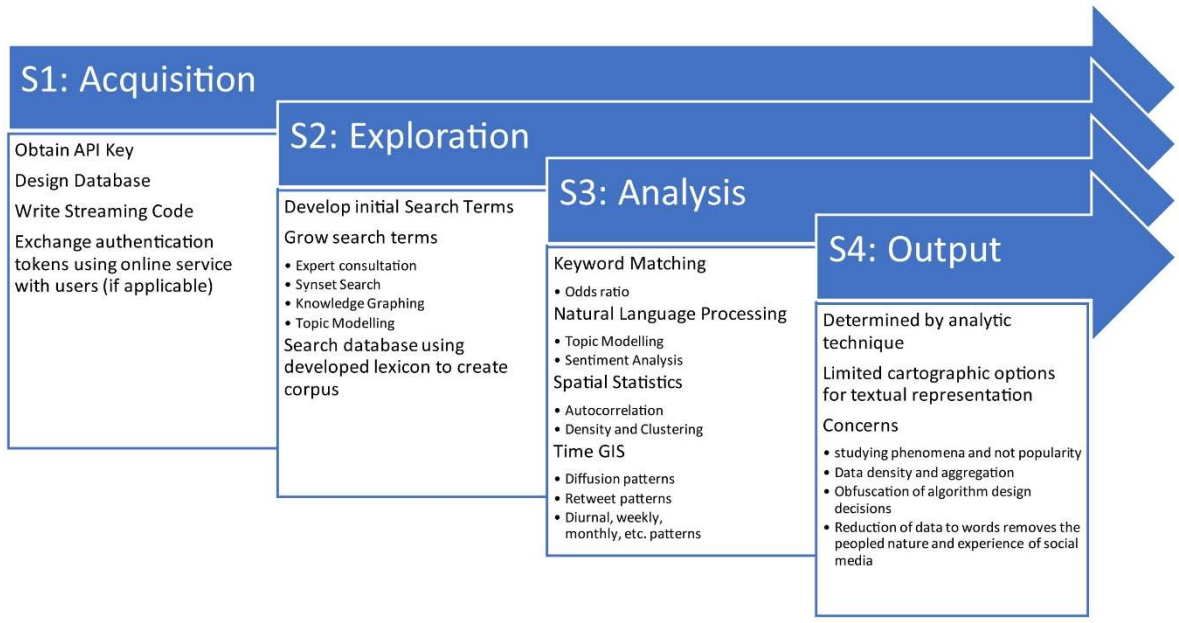

Figure 2:Typical workflow for use of Social Spatial. Using social spatial, researchers can progress linearly through the processes of exploration, analysis and output, or flow forwards and backwards as new lines of inquiry become evident

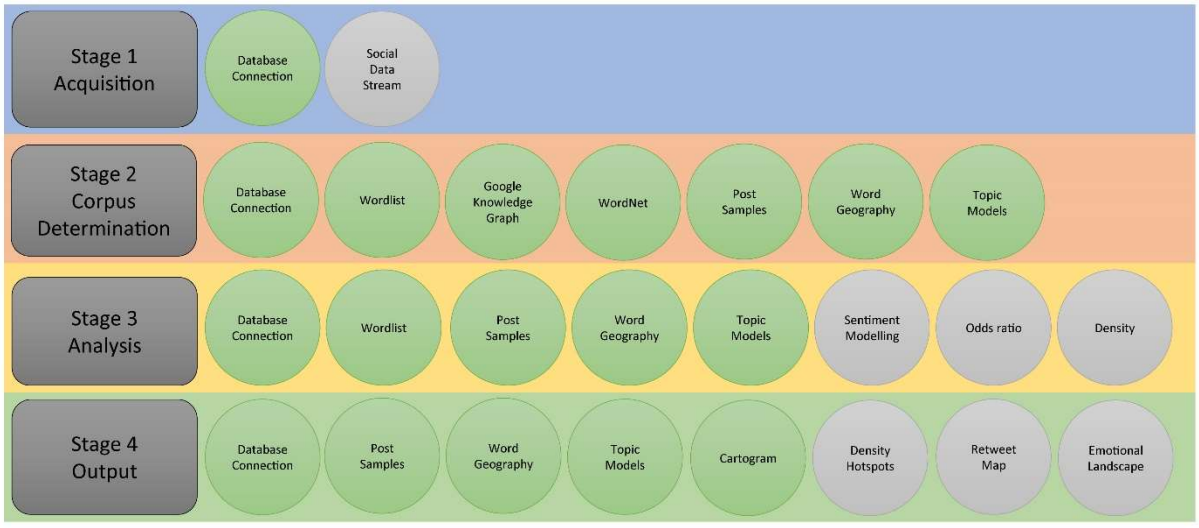

Figure 3: Modules of Social Spatial currently developed (green circles) and those expected to be developed (grey circles). The stages conform to the analysis workflow shown in figure 2.

3. Conclusion

Social Spatial is a progression from earlier efforts made by GIScience scholars to fuse qualitative methods with traditional GIS environments, such as those of Kwan and Ding (2008), and Jung (2007).The work of these scholars on ArcGIS extensions and CAQ-GIS were important steps that illustrated that it was indeed possible to integrate qualitative data while providing the tools required for analysis.

Social Spatial is a first step towards a qualitative GIS for big social geospatial data. It embraces modern programming techniques, languages, data formats, and methods. It is modular and allows for researchers to act iteratively, testing new thematic hypothesis as they develop. It provides a pathway for researchers to publish the settings, keyword lists, and results of their research alongside research publications, increasing research transparency and trust. As methods become increasingly complex and unsupervised, it is important that research transparency remains an import goal for researchers to strive for. Social Spatial is available on its git repository at www.github.com/mikedotonline/socialspatial.

3. References

Allen, J. (2003). Natural Language Processing. In *Encyclopedia of computer science.* (4th ed., pp. 1218–1222). Chichester, UK: Wiley. Retrieved from http://dl.acm.org/citation.cfm?id=1074630&CFID=737058752&CFTOKEN=58477774

Bauer, S., Noulas, A., Seaghdha, D. O., Clark, S., & Mascolo, C. (2012). Talking Places: Modelling and Analysing Linguistic Content in Foursquare. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing* (pp. 348–357). IEEE. https://doi.org/10.1109/SocialCom-PASSAT.2012.107

Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate change sentiment on Twitter: An unsolicited public opinion poll, 11. Physics and Society; Computers and Society. Retrieved from http://arxiv.org/abs/1505.03804

Crampton, J. W. (2009). Cartography: maps 2.0. *Progress in Human Geography*, *33*(1), 91–100. https://doi.org/10.1177/0309132508094074

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, *40*(2), 130–139. https://doi.org/10.1080/15230406.2013.777137

Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). #Earthquake: Twitter as a Distributed

Sensor System. *Transactions in GIS*, *17*(1), 124–147. https://doi.org/10.1111/j.1467-9671.2012.01359.x

Elwood, Sarah; Cope, M. (2009). *Qualitative GIS: A Mixed Methods Approach*. SAGE Publications. Retrieved from https://books.google.com/books?hl=en&lr=&id=vnAbJ8spyeYC&pgis=1

Elwood, S. (2008). Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal*, *72*(3–4), 173–183. https://doi.org/10.1007/s10708-008-9186-0

Elwood, S., Goodchild, M. F., & Sui, D. (2013). Prospects for VGI Research and the Emerging Fourth Paradigm. In *Crowdsourcing Geographic Knowledge* (pp. 361–375). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-4587-2_20

Elwood, S., & Leszczynski, A. (2011). Privacy, reconsidered: New representations, data practices, and the geoweb. *Geoforum*, *42*(1), 6–15. https://doi.org/10.1016/j.geoforum.2010.08.003

Frank, M. R., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2013). Happiness and the Patterns of Life: A Study of Geolocated Tweets. https://doi.org/10.1038/srep02625

Haklay, M., Singleton, A., & Parker, C. (2008). Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*, *2*(6), 2011–2039. https://doi.org/10.1111/j.1749-8198.2008.00167.x

Hao, Q., Cai, R., Wang, C., Xiao, R., Yang, J.-M., Pang, Y., & Zhang, L. (2010). Equip tourists with knowledge mined from travelogues. In *Proceedings of the 19th international conference on World wide web - WWW '10* (p. 401). New York, New York, USA: ACM Press. https://doi.org/10.1145/1772690.1772732

Harvey, F., Kwan, M.-P., & Pavlovskaya, M. (2006, October 10). Introduction: Critical GIS. *Cartographica: The International Journal for Geographic Information and Geovisualization*. International Cartographic Association/Association Cartographique internationale. Retrieved from http://www.utpjournals.press/doi/abs/10.3138/04L6-2314-6068-43V6?journalCode=cart

Johnson, P. A., Corbett, J. M., Gore, C., Robinson, P., Allen, P., & Sieber, R. (2015). A Web of Expectations: Evolving Relationships in Community Participatory Geoweb Projects. *ACME: An International Journal for Critical Geographies*, *14*(3), 827–848. Retrieved from http://142.207.145.31/index.php/acme/article/view/1235

Jung, J.-K. (2007, January 1). *Computer-aided qualitative gis (caq-gis) for critical researchers: an integration of quantitative and qualitative research in the geography of communities*. State University of New York at Buffalo. Retrieved from http://dl.acm.org/citation.cfm?id=1329707

Jung, J.-K. (2015). Code clouds: Qualitative geovisualization of geotweets. *The Canadian Geographer / Le Géographe Canadien*, *59*(1), 52–68. https://doi.org/10.1111/cag.12133

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, *3*(3), 262–267. https://doi.org/10.1177/2043820613513388

Kwan, M.-P., & Ding, G. (2008). Geo-Narrative: Extending Geographic Information Systems for Narrative Analysis in Qualitative and Mixed-Method Research∗. *The Professional Geographer*, *60*(4), 443–465. https://doi.org/10.1080/00330120802211752

Lansley, G., Adnan, M., & longley, P. (2015, April 21). The geography of topics from geo-referenced social media data in London. *Http://Meridian.Aag.Org/Callforpapers/Program/AbstractDetail.Cfm?AbstractID=66607*. Retrieved from http://discovery.ucl.ac.uk/1468265/

Larsen, M. E., Boonstra, T. W., Batterham, P. J., ODea, B., Paris, C., & Christensen, H. (2015). We Feel: Mapping Emotion on Twitter. *IEEE Journal of Biomedical and Health Informatics*, *19*(4), 1246–1252. https://doi.org/10.1109/JBHI.2015.2403839

Lee, J.-G., & Kang, M. (2015). Geospatial Big Data: Challenges and Opportunities. *Big Data Research*, *2*(2), 74–81. https://doi.org/10.1016/j.bdr.2015.01.003

Liu, Y., Ester, M., Hu, B., & Cheung, D. W. (2015). Spatio-Temporal Topic Models for Check-in Data. In *2015 IEEE International Conference on Data Mining* (pp. 889–894). IEEE. https://doi.org/10.1109/ICDM.2015.45

Martin, M. E., & Schuurman, N. (2017). Area-Based Topic Modeling and Visualization of Social Media

for Qualitative GIS. *Annals of the American Association of Geographers*, 1–12. https://doi.org/10.1080/24694452.2017.1293499

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE*, *8*(5), e64417. https://doi.org/10.1371/journal.pone.0064417

Pavlovskaya, M. (2009). Non-quantitative GIS. *Qualitative GIS: A Mixed Methods Approach*. Retrieved from https://books.google.ca/books?hl=en&lr=&id=vnAbJ8spyeYC&oi=fnd&pg=PA13&dq=related:PzvHw4maVZwJ:scholar.google.com/&ots=gQLctEzBKU&sig=chxuEwnLgm19IEmDsXtGXXhFy8M

Schuurman, N., & Leszczynski, A. (2006). Ontology-Based Metadata. *Transactions in GIS*, *10*(5), 709–726. https://doi.org/10.1111/j.1467-9671.2006.01024.x

Stephens, M. (2013). The Geography of Hate. Retrieved October 2, 2017, from http://www.floatingsheep.org/2013/05/hatemap.html

Wang, C., Wang, J., Xie, X., & Ma, W.-Y. (2007). Mining geographic knowledge using location aware topic model. In *Proceedings of the 4th ACM workshop on Geographical information retrieval - GIR '07* (p. 65). New York, New York, USA: ACM Press. https://doi.org/10.1145/1316948.1316967

Yeager, C. D., & Steiger, T. (2013). Applied geography in a digital age: The case for mixed methods. *Applied Geography*, *39*, 1–4. https://doi.org/10.1016/j.apgeog.2012.12.001

Zook, M., & Poorthuis, A. (2014). Offline Brews and Online Views: Exploring the Geography of Beer Tweets. In *The Geography of Beer* (pp. 201–209). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7787-3_17