Eurographics Conference on Visualization (EuroVis) 2017
B. Kozlíková, T. Schreck, and T. Wischgoll (Editors)

TypoTweet Maps: Characterizing Urban Areas through Typographic Social Media Visualization

Alex Godwin, Yongxin Wang, and John T. Stasko

School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA

Abstract

Analyzing the important events and news stories that have captured the public interest in a city can be useful for determining the topics that are vital to the people that live there. Social media data, such as tweets, provides a useful and ever-churning feed of data to analyze for this purpose. For even a moderately-sized city, however, individual neighborhoods can have very different characteristics from one another. Geotagged tweets can be a rich data source for determining what people are saying online about the location they are in. Relating the text data to spatial location, however, presents a unique challenge in representation and layout. In this paper, we introduce TypoTweet Maps: a technique for constructing representations of neighborhood topics as typographic maps. TypoTweet Maps show differences in neighborhood topics using only text, avoiding the channel interference of feature labels that are unnecessary for residents who are familiar with the shape of the city. We describe the process of mapping geotagged tweets to the shape of neighborhoods and streets, and present a case study applying the technique to the city of Atlanta.

1. Introduction

Analyzing the important events and news stories that have captured the public interest in a city can be useful for determining the topics that are vital to the people that live there. Social media data, such as tweets, provides a useful and ever-churning feed of data to analyze for this purpose. For even a moderately-sized city, however, individual neighborhoods can have very different characteristics from one another. The residents of one neighborhood may be interested in the highest quality nightlife and dining, while residents of another may be focused on the home. For business owners and city planners looking to make changes to the city, these characteristics can be used to plan new opportunities and features that will resonate with the people who already live in each area. Similarly, a resident may find enjoyment in reading what the other inhabitants of her neighborhood are saying and how it differs from the chatter of other areas.

Geotagged tweets can be a rich data source for determining what people are saying online about the location they are in. Relating the text data to spatial location, however, presents a unique challenge in representation and layout. The distribution of social media topics can be used to identify clusters of similar residents that differ from the statistically defined areas defined by the city [JTC12]. The text can be analyzed to determine sentiment, which can then be compared to spatial distributions of historical crime data to determine the effects of long-term crime in each neighborhood on residents speech patterns [VEDC15] or to find correlations with demographic distribution and sites of segregation [Lin14]. Many of these efforts have not conveyed the tweet topics and sentiments in a geospatial representation that maintains the context of relative neighborhood locations. A geospatial visualization, however, would readily fit the mental maps of residents of the city who are familiar with the general shape of the city, allowing them to quickly lookup and browse areas that they are interested in [BM13].

Maps typically use text to relay the name of locations and alter the typography of feature names to represent attributes. For example, major highways can be rendered with thick lines and bold text, while secondary roads and footpaths use thin lines and less salient text [KO11]. Text for relevant topics can be overlaid on the map as pins, but this hides the content of the tweets. The text of tweets can be overlaid on the map, but will compete with the ink used to display map features for salience. Residents who are already familiar with the city, however, already know the names of many of the streets and neighborhoods. In this paper, we introduce TypoTweet Maps: a technique for constructing representations of neighborhood topics as typographic maps. TypoTweet Maps show differences in neighborhood topics using only text, avoiding the channel interference of feature labels that are unnecessary for residents who are familiar with the shape of the city. We describe the process of mapping geotagged tweets to the shape of neighborhoods and streets, and present a case study applying the technique to the city of Atlanta.

DOI: 10.2312/eurovisshort.20171128

2. Related Work

Spatial analysis of social media data can be useful for understanding the distribution of topics and sentiment to serve a variety of domain analysis questions. For example, Hawelka et al. used globally sourced geotagged tweets to reveal patterns of travel and cohesive regions of tourism between groups of countries [HSB*14]. Geotagged tweets can also reveal momentary hotspots as people tweet from or refer to locations, indicating places which currently occupy a high interest within online communities [SCR13]. By combining this notion with interactive spatial representations of data, researchers have been able to design systems that can quickly identify trending hotspots and provide information about what types of topics are contributing to those trends. Chae et al., for example, utilized seasonal-trend decomposition (STL) to identify unusual peaks and outliers in social discourse against the usual popular topics and pinpoint their location on a map [CTB*12]. Similarly, MacEachren et al. have presented multiple systems for conducting spatial tweet trend analysis in the context of crisis management [MJR*11, MRJ*11]. These approaches, however, do not support the representation of all the available twitter data on the map concurrently. Similarly, these systems are not intended for the use or enjoyment by casual users who live in a city and want to understand it more fully.

Many commercial typographic maps are created by hand using a vector-based drawing program, and distributed primarily as enjoyable art pieces for purchase online. Axis Maps, for example, sells typographic maps for many major cities, including Boston, London, and Chicago [axi]. These maps encode only the place names and layout but do not indicate any additional information about the city. Maharik et al. explored the use of text-based representations of abstract shapes that could be used to create artistic renderings of all forms, and presented a technique for constructing text paths that conveyed the shapes of elements within an image [MBS*11]. Instead of aligning the text to lines within each neighborhood, we base our approach on word clouds in order to fill the available space and preserve separability of adjacent neighborhoods without altering text orientation. Thom et al. used a tag cloud representation in combination with a spatial clustering algorithm to identify spatiotemporal anomalies in social media data, but did not align the analysis with political boundaries or streets [TBK*12]. Buchin et al. proposed a word cloud method for filling the interior of political regions with tagged social media topics while preserving the boundary silhouette of the region and the interior proximity of topics to location [BCL*16]. In contrast, previous work by Afzal et al. explored methods for automatically creating typographic maps [AMJ*12], and speculated that they may also be used to convey thematic distributions of underlying spatial data. They did not, however, explore the substitution of labels with descriptive topics drawn from geotagged social media data.

3. TypoTweet Maps

To create typographic maps of twitter data, we first collected geospatial data that could be used to define the shapes of neighborhoods and streets within the city. These features serve as a guiding mesh for the contouring of text within the display so that it follows familiar routes and boundaries. This is to support the user as they attempt to lookup the topics that are represented in familiar neighborhoods to determine if topics they believe are important are present. This representation also makes it possible for a user to browse the visualization, finding familiar neighborhoods, to discover what people are saying in those areas.

To facilitate rapid data querying, we compiled tweets from the Twitter API using a collection of sample points spaced at least 1km apart. These locations were determined using Bridson's poisson disk sampling algorithm [Bri07] within a bounding box for the city of Atlanta, yielding 471 locations from which to draw tweets. This allowed us to create a somewhat uniform distribution of sampling locations throughout the city. For each sampling point, we queried the Twitter API for the most recent 100 tweets at that location. Tweets were then filtered for common stop words and links. We did not, however, remove emoji, foul language, insults, or slurs, many of which are present in the resulting visualization (see Section 3.3).

3.1. Neighborhoods

We obtained neighborhood statistical area shapefiles from the Atlanta Regional Commission (ARC). For each neighborhood, we determined the sampling points that were contained within the boundaries. The tweets obtained from these sampling locations were used to create a bag of words model, identifying the keywords that were commonly found in the tweets. For each neighborhood, we create a word cloud using its bounding box in which common words are located in the center. We do not, however, map the term frequency to the style or size of the type. Though this can be used to convey information about the tweets [BGN08], we wanted to instead convey a more uniform appearance within each neighborhood rather than having a heavy and salient centroid with a diminishing outer perimeter. We then used the boundary of the neighborhood as a clipping path, retaining the visible shape of the area while excluding the tweets that are less characteristic of the conversations occurring in that location. Each neighborhood was assigned a color to easily distinguish it from adjacent regions.

3.2. Roads

We obtained road data from OpenStreetMap (OSM) for the major highways located in the city. This data consisted of the motorways, trunks, primary, and secondary roads located within the city limits. We then used the waypoints along the path of each street to query the sampling locations, assembling the collection of tweets that were most descriptive of the route taken by the road through the city. We then used the sequential waypoints that define the path of each road to create a textpath. We placed keywords at the beginning of each street, gradually adding more terms until the road terminates. The salience of the type for the displayed road is mapped to the importance of that road for transportation: larger, more heavily utilized highways are rendered in dark, bold lettering while less important roads are rendered more lightly. To help visually distinguish the roads from the underlying neighborhoods, a white line is rendered beneath them.

3.3. Atlanta Tweets

We used the shapes of the neighborhoods and roads of Atlanta to visualize Twitter data for the city (Figure 2). Many of the topics for each neighborhood contain relatively routine chatter, discussing the importance of food in that neighborhood or talking about restaurants and events. Neighborhoods vary in the importance of nighttime activities (e.g., cocktails, beer, sweat) to daytime (e.g., brunch, job offers). Certain topics may have no local referent, and are likely important topics of general national interest that locals are buzzing about (e.g., Russia, oscars). Foul language is more prominent in some neighborhoods than others, as are variations on racial slurs that may or may not be appropriate depending on who is doing the tweeting. In some places the roads reflect the characteristics of the regions they are passing through, though many of them refer specifically to road-based topics (e.g., the name of the road or "accident"). That there is a clear connection in topics about roads in the location of the road is concerning, as it may indicate that many of these tweets are originating from moving cars.

Some of the most interesting areas to focus on within the map are the transitions from one neighborhood to another or areas where several neighborhoods converge. At these locations, it is far easier to perceive the differences in topics between locations. For example, in Figure 1, several neighborhoods are shown in close proximity. At the moment the tweets were captured, the topics were very different, moving from runners in the area around the primary Atlanta park to discussion of theater towards primary theater in the midtown area. Discussion of bars and nightlife is heavier north of those neighborhoods, including several references to women, bars, and adult entertainment. The roads in this area largely reflect place names, such as midtown Atlanta and the location of a highly visible building (i.e., Spire).

It is admittedly challenging to determine, in this view, whether the visible topics indicate a referent that is located within the neighborhood. While topics that can be tied directly to locations of interest are more obvious, other topics (e.g., histrology in the northwest corner of Figure 1) are more challenging to consider. This indicates that a useful next step to consider would be to conduct a term frequency inverse document frequency (*tf-idf*) analysis of the topics among the neighborhoods, weighting the topics that more uniquely define each neighborhood more highly than those that are more common. This would make it difficult, however, to determine the characteristics in common between neighborhoods.

Figure 1: *TypoMap zoomed into the area around midtown in Atlanta. In the area around Piedmont park (red text) topics generally focus on the park and fitness activities. To the west, discussion is more centered around theater, stage, and tickets largely due to that area's proximity to the Fox Theater, a local landmark. Moving north, there is frequent discussion of nightlife.*

4. Discussion and Future Work

In this paper, we have presented TypoTweet Maps: our approach for capturing representative social media topics about neighborhoods in typographic maps. In its current format, the visualization is most easily inspected as a high-resolution printout. It's no surprise that the current business model for typographic maps is distribution through printed posters, as the small text is challenging to read on a computer screen or even on a single page high resolution printout. The current implementation is also very slow to render, though the use of sampling points does allow for a server to gradually update the backing twitter data as time passes. Regardless, the process of generating the visual representation of the text currently takes quite a bit of time (30-45 minutes for the image in Figure 2). This could be improved, however, with further work to parallelize the process of querying map features from the sampled tweets.

There are several promising directions for future work based on this technique. The current technique contains no interaction, owing largely to the extremely processor-intensive task of displaying the text without additional constraints. Nevertheless, it would be extremely useful to include the capability for a user to query a term or topic and highlight the regions and roads that include that topic. Similarly, it would be useful for a user to be able to select a road or region on the map and see a secondary visual representation of the topics that comprise that feature. Additional, more complex interactions with the proposed visualization technique would be interesting to consider. For example, a user could select a route along the roads on the map and review the change in topics from one location to another. The topics for the road could be contrasted within the context of the surrounding regions that the user would be passing as they move from one area to another.

The representation could also be modified to incorporate multiple data sets concurrently. The typographic style of the topics could be altered to reflect an underlying spatial distribution, as proposed by Afzal et al. [AMJ*12]. This would convey both the spatial distribution of topics within the map but also a secondary channel of point-based data distribution. To accommodate this, however, our technique would have to more closely tie the location of topics within each region to the sampling point rather than constructing tag clouds from all the terms within a region. Similarly, the sentiment of important topics within each neighborhood could be mapped to the salience of typographic features so that words that are more closely aligned with strong positive or negative affect stand out prominently in the display.

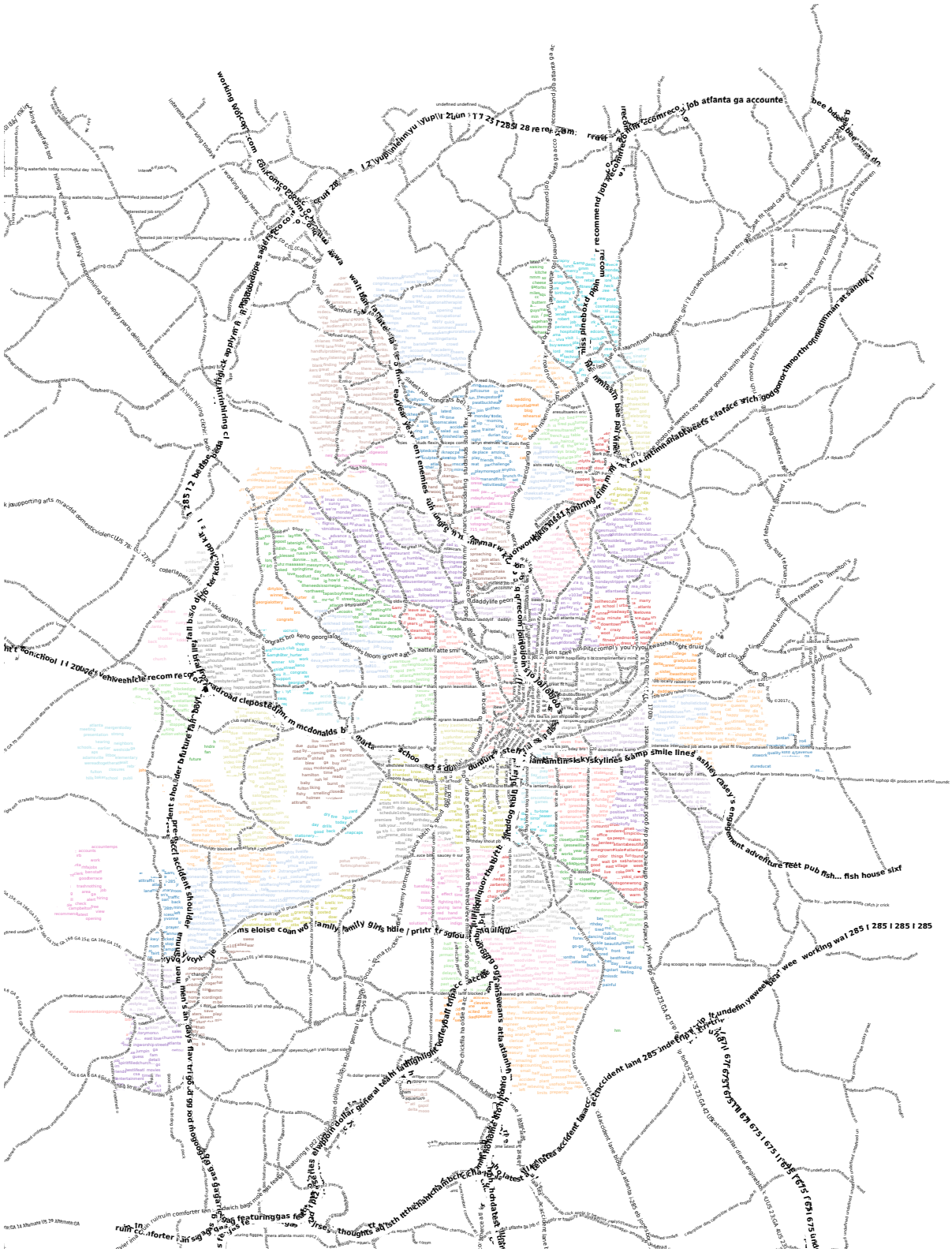

Figure 2: *TypoTweet map of the city of Atlanta. Roads and neighborhoods are rendered using important social media topics in that area. Major highways are rendered in bold text, while secondary and local roads are in smaller type. Neighborhoods are printed in different hues for legibility. Only neighborhoods within the city limits are included, while some neighborhoods inside the city are blank due to the lack of available tweets.*

References

[AMJ*12] AFZAL S., MACIEJEWSKI R., JANG Y., ELMQVIST N., EBERT D. S.: Spatial text visualization using automatic typographic maps. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2556–2564. 2, 3

[axi] AxisMaps. http://www.axismaps.com/. Accessed: 2017-04-24. 2

[BCL*16] BUCHIN K., CREEMERS D., LAZZAROTTO A., SPECKMANN B., WULMS J.: Geo word clouds. In *2016 IEEE Pacific Visualization Symposium (PacificVis)* (2016), IEEE, pp. 144–151. 2

[BGN08] BATEMAN S., GUTWIN C., NACENTA M.: Seeing things in the clouds: the effect of visual features on tag cloud selections. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia* (2008), ACM, pp. 193–202. 2

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2376–2385. 1

[Bri07] BRIDSON R.: Fast poisson disk sampling in arbitrary dimensions. In *SIGGRAPH sketches* (2007), p. 22. 2

[CTB*12] CHAE J., THOM D., BOSCH H., JANG Y., MACIEJEWSKI R., EBERT D. S., ERTL T.: Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST),* (2012), IEEE, pp. 143–152. 2

[HSB*14] HAWELKA B., SITKO I., BEINAT E., SOBOLEVSKY S., KAZAKOPOULOS P., RATTI C.: Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science 41*, 3 (2014), 260–271. 2

[JTC12] JOSEPH K., TAN C. H., CARLEY K. M.: Beyond local, categories and friends: clustering foursquare users with latent topics. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (2012), ACM, pp. 919–926. 1

[KO11] KRAAK M.-J., ORMELING F.: *Cartography: visualization of spatial data*. Guilford Press, 2011. 1

[Lin14] LIN Y.-R.: Assessing sentiment segregation in urban communities. In *Proceedings of the 2014 International Conference on Social Computing* (2014), ACM, p. 9. 1

[MBS*11] MAHARIK R., BESSMELTSEV M., SHEFFER A., SHAMIR A., CARR N.: Digital micrography. *ACM Transactions on Grpahics(Proc. SIGGRAPH 2011) 30*, 4 (2011), 100:1–100:12. 2

[MJR*11] MACEACHREN A. M., JAISWAL A., ROBINSON A. C., PEZANOWSKI S., SAVELYEV A., MITRA P., ZHANG X., BLANFORD J.: Senseplace2: Geotwitter analytics support for situational awareness. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2011), IEEE, pp. 181–190. 2

[MRJ*11] MACEACHREN A. M., ROBINSON A. C., JAISWAL A., PEZANOWSKI S., SAVELYEV A., BLANFORD J., MITRA P.: Geotwitter analytics: Applications in crisis management. In *25th International Cartographic Conference* (2011), pp. 3–8. 2

[SCR13] STEFANIDIS A., CROOKS A., RADZIKOWSKI J.: Harvesting ambient geospatial information from social media feeds. *GeoJournal 78*, 2 (2013), 319–338. 2

[TBK*12] THOM D., BOSCH H., KOCH S., WÖRNER M., ERTL T.: Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *2012 IEEE Pacific Visualization Symposium (PacificVis)* (2012), IEEE, pp. 41–48. 2

[VEDC15] VALDES J. M. D., EISENSTEIN J., DE CHOUDHURY M.: Psychological effects of urban crime gleaned from social media. In *Ninth International AAAI Conference on Web and Social Media* (2015). 1