Enriching Traffic Information with a Spatiotemporal Model based on Social Media

Bruno P. Santos[†], Paulo H. L. Rettore[†], Heitor S. Ramos[‡], Luiz F. M. Vieira[†], and Antonio A. F. Loureiro[†]
[†]Department of Computer Science – Universidade Federal de Minas Gerais (UFMG) – Brazil
[‡]Institute of Computing – Universidade Federal de Alagoas (UFAL) – Brazil
Email: [†]{bruno.ps, rettore, lfvieira, loureiro}@dcc.ufmg.br, [‡]heitor@ic.ufal.br

***Abstract*—In this work, we argue that Location-Based Social Media (LBSM) feeds may offer a new layer to improve traffic and transit comprehension. Initially, we showed the significant correlation between Twitter's feed and traditional traffic sensors. Then, we presented the Twitter MAPS (T-MAPS) a low-cost spatiotemporal model to improve the description of traffic conditions through tweets. T-MAPS enhance traditional traffic sensors by carrying the human lens into the transportation system. We conducted a case study by running T-MAPS and Google Maps route recommendation, in which, we showed T-MAPS viability, as an additional traffic descriptor. As a result, we noticed the median of route similarity reached 62%, and for a quarter of the evaluated trajectories, the similarity achieved between 75% and 100%. Also, we presented three route description services, based on natural language analyzes, Route Sentiment (RS), Route Information (RI), and Area' Tags (AT) aiming to enhance the route information.**

I. Introduction

The transport infrastructure might be able to promote people's movement efficiently, but it also implies in the constant need for planning and management of the transportation system. In this sense, understanding urban mobility (traffic and transit) has been the focus of governments, researchers, and industries [1]. Usually, traffic and transit specialists use traditional raw data sources (e.g., data from inductive loops, traffic cameras, and origin-destination matrix) to perform their analyzes. Unfortunately, the access to these data sources is, in general, limited to those who are connected to governmental entities or large corporations, it covers a limited scope and has a high financial cost to access and use it. This becomes a barrier to understand better urban mobility that asks for other solutions.

In that way, the Location-Based Social Media (LBSM) (e.g., Twitter, Instagram, and Foursquare) becomes an alternative data source to study urban mobility. These platforms allow users to share their thoughts, viewpoints, and activities related to their feelings about almost everything, which include traffic conditions. There are different research issues which benefit from using LBSM as a low-cost data source [1], [2], [3], [4].

According to the *Twitter*, about 330 million users are active every month in their network. This considerable attendance may open up several opportunities. In this work, we investigated the traffic scenario in the lens of LBSM. However, data from those social media also brings issues that can lead to other challenges such as data imprecision, users' bias, and spatiotemporal assignment or inconsistency. Therefore, we should overcome those data issues before making complete use of LBSM's data.

We conducted a study to understand better the relationship between the real traffic scenario and the data provided by Twitter, a very well-known and largely used LBSMs platform. Initially, we focused on the data collection and its characterization. Then, we proposed the Twitter MAPS (T-MAPS), a low-cost spatiotemporal model to improve the description of traffic conditions based on tweets. T-MAPS intends to enhance the current navigation context by connecting LBSM's data in different ways, for example, by evaluating tweets frequency or users' perspective of a region of interest.

Based on that, we collected tweets from New York City (NYC for short) demonstrating its coverage and the traffic factor correspondence. Then, we proposed and evaluated the T-MAPS applicability by showing its route similarity with Google Maps recommendations, also we provided three route description services upon T-MAPS: Route Sentiment (RS), Route Information (RI) and Area' Tags (AT). We highlighted two main contributions: 1) The LBSM data characterization, especially from Twitter, as a data source to better understand and describe the traffic conditions. 2) The proposition of T-MAPS as a model to enrich route description.

An important question emerges from the inherent subjectivity of enriching the traffic description, as we proposed. To the best of our expertise, there is no ground truth for the best route. For that reason, many tools aim to offer their traffic viewpoint like Google Maps, Here Wego, and TomTom maps. The main reason which motivated us to develop the T-MAPS was the desire to demonstrate the potential of using LBSM data, as a traffic data. Also, we aim to encourage the design of new applications, models, and analysis of urban mobility using LBSM.

This paper is organized as follows. In Sec. II, we presented the related works. In Sec. III, we detailed the collected data and its issues. In Sec. IV, we showed the correlation between LBSM and traffic sensors data. In Sec. V, we presented the T-MAPS modeling process. We showed a case study in Sec. VI. Sec. VII showed the route description services, which aims to improve the information about the route. Finally, in Sec. VIII, we presented the final remarks and future directions.

II. Related Work

In the literature, there are many studies about event detection and diagnostics using LBSMs [5], [6], [7]. Most of them focuses on detect general events and use language pattern recognition to understand events. Ribeiro et al. [3] proposed a technique to detect traffic events and displayed them in near real-time on the web. Similarly, Septiana et al. [8] used a text mining system on RSS feed Facebook E100 (a page to provide transit information) aiming to categorize road condition into six types as floods, traffic jams, congested roads, road damage, accidents, and landslides. They showed an accuracy of 92% in the road condition monitoring.

Some proposals [9], [10] studied sentiment analysis by using LBSMs data. Bertrand et al. [9] studied the sentiment in NYC from a spatiotemporal perspective in a high-resolution. Kim et al. [4] proposed SocRoutes, a safe route recommending system, based on Twitter data. Giridhar et al. [11] focused on explaining unusual traffic events using social media feeds, but their work does not provide ways to recommend routes.

Gu et al. [12] explored tweets text aiming to extract traffic incident information providing a low-cost solution to existing data sources. In that way, they developed a methodology to the data acquisition, process, and filtering. Gu validated the Twitter-based incidents using data from RCRS (Road Condition Report System) incident, 911 Call For Service (CFS) incident, and HERE travel time.

Differently, from the most of above-related works, we are going beyond by providing a model to clarify the traffic condition, adding extra information to the current navigation systems. Besides, T-MAPS model is flexible enough to consider instantaneous and historical data, and text mining techniques. In that sense, we provide three route description services examples over T-MAPS model: Route Sentiment (RS), Route Information (RI), and Area' Tags (AT). These services show T-MAPS viability as a tool to offer extra information about routes.

III. Data Acquisition

One of the most significant challenges to study urban mobility is the absence of open data in such context. Therefore, most of the work in this field lies in the theoretical studies or has a large private data provider (e.g., government agencies, Google, Tomtom among others.). Fortunately, the growing LBSMs adoption allows people to share on online platforms their thoughts, viewpoints, and activities. All this content is related to users' feelings and impressions, including the traffic and transit conditions. With the right tools and code, we were able to collect data from Twitter, where many users periodically share information about traffic and transit events. For this end, we used the Twitter's APIs respecting the restriction terms[1].

The dataset consists of 353.807 tweets from twenty-one manually selected users accounts. Those accounts are maintained by departments of transport, specialists on traffic and transit reports such as news channels or dedicated companies.

[1]https://developer.twitter.com/en/docs

TABLE I
Example of users

Account name	# Tweets
@511NYC	126925
@TotalTrafficNYC	20267
@WazeTrafficNYC	7850
@Traffic4NY	3789
...	...
@NYC_DOT	3680
Total:	≈ 354K

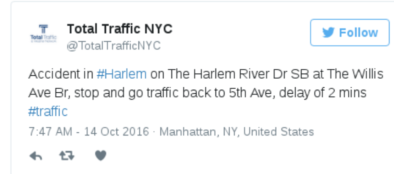
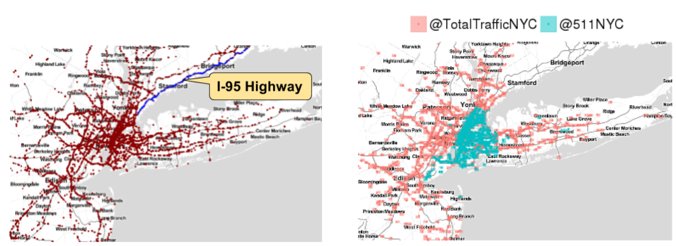


Fig. 1. Tweet example in the dataset

Tab. I shows some accounts and its tweeting frequency. The number of tweets with geotagging is 307.020, most of them in NYC. Here, we explored Manhattan where has 38.112 tweets. The dataset was collected during the last three months of 2016.

Fig. 1 displays a tweet in the dataset. The tweet consists of a rich explanation of the traffic event (textual address, the cause of the event and even delays). Also, the meta-data contains post time, a geotag signature, counters (e.g., retweets and likes). The dataset does not contain regular users due to the high user bias in their tweets regarding traffic feelings. Next, we analyze some aspects which involve the use of LBSM' data, and then the spatiotemporal characteristics of data which provided some initial insights.

a) Data issues: Often data from Twitter has aspects that lead to issues in its use on traffic context. In [13], [14], the authors classify a variety of data aspects. Here, we highlight four of them: i) *Data imprecision:* presents incomplete data, vagueness or granularity effects, usually the inherent heterogeneity of the data sources and "freedom" of data input on online platforms promote such aspect; ii) *User bias:* LBSM users can interpret the traffic congestion in different ways introducing bias into the data; iii) *Spatiotemporal assignment:* the geolocation and temporal tagging allow traffic specialists to study and characterize a region at any instant or time interval, and iv) *Inconsistencies:* appears when two or more data sources diverge about a specific event or when traffic and transit information are out of sequence into the system.

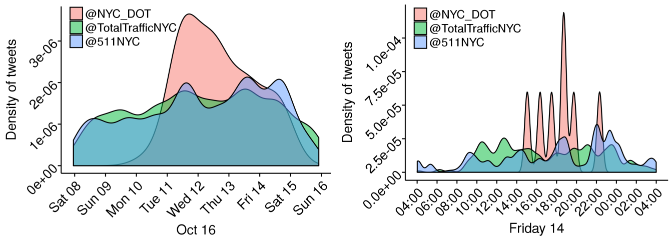


(a) Tweets on NYC map

(b) Spatial coverage of two users

Fig. 2. Dataset coverage in New York City and neighborhood

b) Spacial coverage: Fig. 2(a) shows tweets with geotagging in the dataset. Most tweets are over the road network, i.e., if we do zoom in, it is possible to see the I-95 highway with tweets along its extension. Also, the central region presents higher tweets density than non-central ones, which can indicate a tendency of the user's preferred region to report information. Fig. 2(b) shows the filtered dataset with tweets from @TotalTrafficNYC and @511NYC. As expected, we note that different accounts have contrasting coverage. While @511NYC focused on reporting traffic information within NYC boundaries, @TotalTrafficNYC exhibited broader coverage. Aware of this characteristic, one might use as many as possible spatial complementary accounts to cover a region.



(a) Weekly tweeting density (b) Hourly tweeting density

Fig. 3. Temporal coverage of three accounts for one week

c) Temporal coverage: In Fig. 3(a), we show the tweets' density along the week for @NYC_DOT, @TotalTrafficNYC, and @511NYC users. As expected, different accounts have disparate behavior in their posting rate. Although @NYC_DOT posts mainly on working days, @TotalTrafficNYC and @511NYC do postages every day. However, they still have a different tweeting rate behavior as shown in Fig. 3(b) displaying the hourly tweeting density. The most of the @NYC_DOT posts occur during business hours, while @TotalTrafficNYC and @511NYC post along the day. Note that some peaks of tweets appear during rush times. For example, @TotalTrafficNYC presents high post volume from 7:00 AM to 10:00 AM; this suggests that traffic events occur while people are starting their daily activities. The peaks also appear in @511NYC's curve, one at 12:30 PM – 3:00 PM, another in 5:30 PM – 8:00 PM, and in 9:00 AM – 12:00 PM suggesting high posting rate at lunchtime, another when people are finishing their business day, and when people are starting their nightlife, respectively. Thus, one might use complementary accounts to increase the temporal coverage.

IV. Twitter as a Traffic Sensor

To reveal the potential of LBSM data to enhance and complement the conventional ways to see traffic and transit, it is fundamental the understanding of how related the tweets are to the traditional traffic sensor. For example, if a conventional traffic sensor detects an anomalous event, can tweets explain such atypical event? This section presents directions to answer questions like this.

First, it is required to get access to classic traffic measurement data, such as inductive loop detector counts, traffic cameras, vehicle GPS traces on road network, or origin-destination matrices, among others. With these data, traffic specialists can study demand and supply aspects. Demand can be seen as vehicles and pedestrians while supply is related to streets, highways, sensors and control devices [1]. Thus, it is possible to study the interactions between demand and supply, and eventually develop efficient transportation systems, which optimize urban mobility and decrease transit congestion.

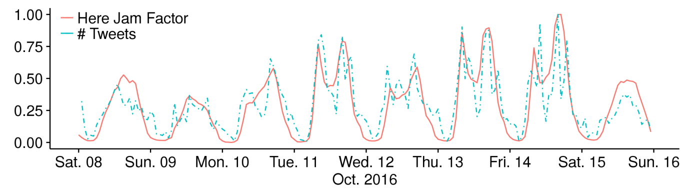


Fig. 4. Tweets frequency and Here Jam Factor time series

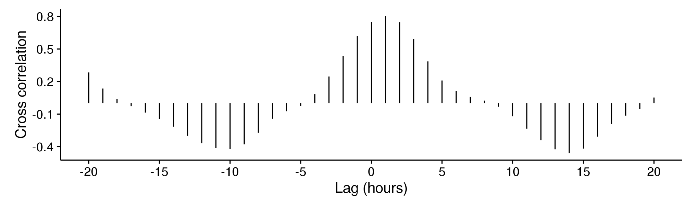


Fig. 5. Cross-correlation between Jam Factor and # tweets time series

Unfortunately, the access to raw traditional sensor data is a challenge for the regular community. Raw traffic data are kept locked by government entities or large companies. Usually, the traditional sensors sense three variables of interest: velocity, density, and flow. These quantities relate to each other allowing traffic behavior analyses and visualizations [1], [15]. On the other hand, LBSMs data is more accessible, which allows urban mobility studies [1], [2]. Also, it is common that users share their thoughts, viewpoints, and activities on LBSM platforms. It expands the sensing capacity by capturing the users' perspective about the situation.

Naturally, raw data holders perform some data fusion process and present the result in their services or statistics. For example, Google gathers heterogeneous data such as GPS traces, cameras, and inductive loops. Thus it makes a data fusion process and presents the results of traffic conditions in colors over the map. In that way, companies like Google, Here, and TomTom allow access to the resulting data fusion process. In this work, we use the Jam Factor (JF) from HERE API as aggregated traditional traffic sensor data. According to the Here documentation, the JF is a fused representation of traditional heterogeneous data. JF ranges from 0 to 1 (from free to congested). We chose Here JF since no other company provides such kind of data.

Fig. 4 shows the correlation between Here JF and tweets in the dataset along a week in Oct. 2016. The time series in blue is the aggregated Here JF, and the orange one corresponds to the number of tweets. We re-scale tweet time series to lie between 0 and 1, and we aggregated each series hourly. Then, we observe that the curves are similar. We compute the Spearman's rank (ρ) a nonparametric correlation coefficient to identify relationships between two variables. The ρ has a value between -1 and +1, where -1 means the observations are entirely dissimilar and +1 the opposite. We apply Spearman's rank in the time series resulting in $\rho = +0.81$. It is possible to interpret that the #tweets tend to increase when the JF increases.

Applying the cross-correlation technique, it is possible to figure out where time series match [16]. Fig. 5 shows on the y-axis the cross-correlation between JF and #tweets, and on the x-axis the lag between the time series, we use JF as the test waveform. The highest correlation (0.8) appears when the lag is +1 meaning that #tweets curve is 1 hour ahead of JF. One can interpret that tweets appear on the platform before JF increases, but note that the time series were hourly aggregated.

V. Modeling Process of Twitter MAPS (T-MAPS)

The T-MAPS is a low-cost spatiotemporal model which aims to clarify traffic events through tweets. This model allows the representation of the traffic scenario in different aspects by considering instantaneous or historical data, and its text mining. Following, we presented the three steps of the modeling process.

1. *Data acquisition*: this step consists of segmenting the area of interest and retrieving data from the LBSMs platforms. It is possible to segment the region of interest in several resolutions, ranging from micro (at the level of roadways and streets segments) to macro resolutions (those at the level of entire roads, boroughs, city). The segmentation resolution may be adjusted to fit the spatial coverage of the data.

2. *Filtering and data fusion process*: this step aims to filter and bind LBSM's data to the segmented region. We propose the use of a weighted time-varying digraph as a model to map these areas and data. The time-varying digraph is represented as a series of static networks, one for each time step. Formally, let R be the set of segments of the region, then a snapshot digraph is defined as $D_t = (V, E, m)$, where $V = \{r | r \in R\}$ denotes the segmented region, and $E = \{(u, v) \in V | u \text{ is adjacent to } v \text{ in } R \text{ segmentation}\}$ denotes the directed edges between physically connected regions, and m is the weights (discussed below). The T-MAPS's time-varying digraph is a sequence of snapshot digraphs, thus $\text{T-MAPS}(D) = \{D_{t=t_{\min}}, D_{t+\Delta}, \ldots, D_{t_{\max}}\}$, where $t_{\min}$ and $t_{\max}$ are the start and end time of the available dataset, and Δ can be adjusted conveniently.

3. *Metrics*: it consists of assigning cost weights to the directed edges. Formally $m(u, w) : E \rightarrow \textit{value}$, where $m(u, w)$ is a function mapping the directed edges to a metric cost. The metric function represents the analyzed traffic scenario using the LBSM data. Fig. 6 illustrates a simple example of the T-MAPS modeling process. First, we segmented the NYC map into five regions of interest, then we collected LBSM available data. Next, we obtained the digraph $G = (V, E, m)$, where V are the regions, and E are directed edges between adjacent regions. Then, we bound Twitter's traffic data to the resulting regions graph. Finally, the weights are assigned to the edges using different metric functions. The resulting time-varying digraph allows us to analyze the traffic scenario condition and description. We present some metric functions below:

Instant: this metric function considers all tweets in a given time t on a day by fusing and filtering them properly. This strategy corresponds to a snapshot view of the traffic at that moment. The smallest t must agree with the configured Δ of T-MAPS model. Usually, instantaneous data are sparse and poorly cover the region of interest. However, this data may highlight a particular event at a given time.

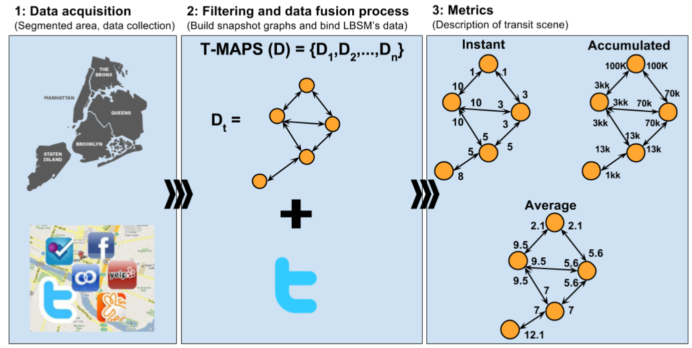


Fig. 6. T-MAPS's modeling process

Accumulated: this metric considers all previously available data for a given time. It requires two parameters, t_{start} and $t_{\text{reference}}$, where $t_{\text{start}} < t_{\text{reference}}$ and it must respect the temporal dataset availability. It accumulates all data between t_{start} and $t_{\text{reference}}$. One can interpret this metric as a historical metric looking to the past until the reference time point. In our experiments $t_{\text{start}} = t_{\min}$.

Average: it uses the same approach of *Accumulated*. However, the value assigned to the edges are the average of tweets' occurrences over the time, such as day, week and year. This information must be passed as a parameter to the metric function. One can interpret it as a typical traffic condition metric, putting into the account the historical information.

VI. A Case Study

We conducted a case study to demonstrate the T-MAPS potential. In that sense, we first compare the T-MAPS, and Google Direction (GD) routes recommendation similarity. Afterward, we presented three route description services demonstrating the T-MAPS' potential as well as others opportunities to enhance and clarify the traffic scenario description. The following results corresponded to the Manhattan region segmentation and the data collected (Sec. III). The region was segmented into 29 official neighborhoods. Consequently, the T-MAPS' digraph snapshot contains 29 vertices. We removed all tweets outside the Manhattan region. Besides, the minimum time interval between two consecutive T-MAPS graphs corresponds to a $\Delta = 1$ hour. Although T-MAPS was designed to accommodate both data resolution (micro and macro), in the case study was used a macro viewpoint due to data coverage limitation.

A. T-MAPS Applicability

We evaluated the T-MAPS applicability by comparing its similarity, in recommend routes, with GD. Note that the T-MAPS route suggestion considers a macro resolution of the regions on the map, but our model is flexible enough to encompass fine-grained resolution as well if there is enough data for this. From a macro resolution, T-MAPS aims to recommend regions which have the best conditions regarding the applied metrics.

We query the T-MAPS and GD 812 to recommend routes in Manhattan neighborhoods. The routes were derived from the combination $2 \times C_k^n$, where $n = 29$ (Manhattan neighborhoods) and $k = 2$ (origins and destinations). Note that we considered routes like $A \rightarrow B$ and $B \rightarrow A$. The routes start and end at the center of the region. Also, we rule out routes that start and end at the same local. We query the routes in three different moments (7:00 am, 3:00 pm and 7:00 pm) of a day along one week. Those moments were chosen purposely, based on its rush hour representation, and its higher volume of tweets in the dataset. Besides, the Here Jam Factor increases on those moments as well as the frequency of tweets per hour in the dataset, see Fig. 7.

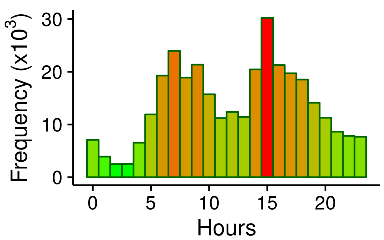


Fig. 7. The frequency of tweets per hour in the dataset

The similarity technique measured the matched areas where the recommended routes by T-MAPS (using Dijkstra's algorithm) and GD passed through. Fig. 8 displays the similarity between routes along eight days in the dataset, considering three metric functions. The box-plots summarize 58.464 routes analyzed. T-MAPS with *Instant* metric showed a high variation of similarity rate, its median range from 50% up to 66.7%, while *Accumulated* metric shows 60% to 70% and *Average* metric 60% to 66.7%. It means that more than a half of the evaluated routes overlapped the GD. We expected that *Instant* metric would pose the lowest similarity due to its intrinsic disparity with other metrics since it does not consider the historical data. As a global evaluation, the median of route similarity reached 62% with Google Directions. Note that T-MAPS uses a macro view, while GD does not, thus implies in fewer regions per route by T-MAPS than GD. The upper quartile (1/4 of the routes) until the maximum value exhibited a similarity between 75% and 100% between the T-MAPS and GD suggested routes.

VII. Route Description Services

Based on the applicability results, which demonstrated a possibility to aggregate extra information to a current route recommendation services, we move on to explore the tweet's texts. Initially, we performed the cleaning phase in the tweet (lowercase transformation, accents removal, tokens extraction, and filtering stops words, links, and special characters). Then, we applied three type of text mining to build the descriptions

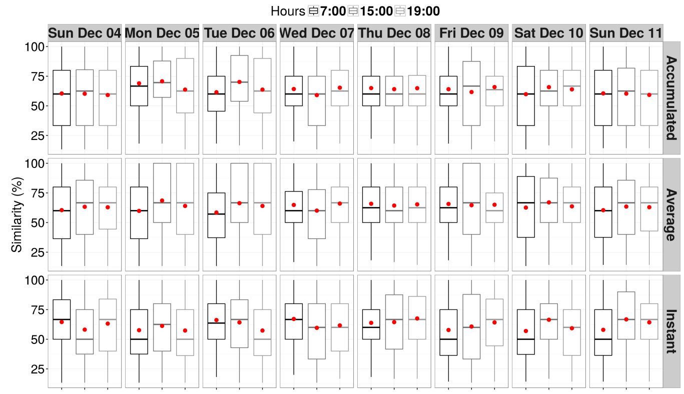


Fig. 8. Route recommendation similarity between T-MAPS and Google Directions (dots represent the mean)

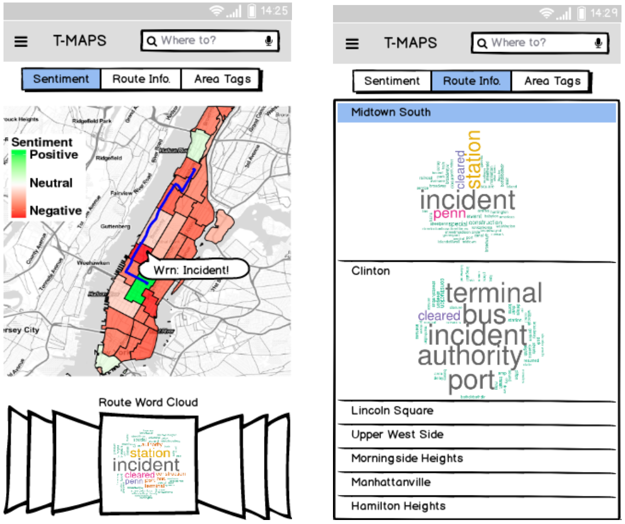
services over the T-MAPS model which are Route Sentiment (RS), Route Information (RI), and Area' Tags (AT). In Fig. 9, we presented a prototype to offer the T-MAPS services.

The RS service showed in Fig. 9(a) allows us to observe the feelings that LBSM users have about the region, in which they will pass through. To derive the sentiment from tweet's text, we used a dictionary of words and its associated feelings [17]. Then the sentiment depends on the number of word/feelings occurrences. Lastly, a score is calculated, and we can associate a sentiment (positive, neutral or negative) to the tweet.

We also presented the RI service which exhibits for each area a word cloud. The word size indicates the word frequency over the tweets. RI allows the T-MAPS users to see what are the most spoken words along with their route. In Fig. 9(b), we can see relevant words like "cleared", "incident" and "station" highlighted in the cloud. These may indicate why the Midtown South has a positive sentiment, and oppositely, Clinton with words like "incident" derived a negative sentiment.

Finally, we developed the AT service. For that service, we used the Term Frequency–Inverse Document Frequency (TF-IDF) method to measure how important a word is to a set of tweets in given area of Manhattan. The Fig. 10 shows the area' tags of each region of the path. This technique allowed us to find words which are unique per area. For instance, the *Authority Port/Terminal* is only in the *Clinton* area, as well as the *Upper West Side* is one of the references with *Parks* to visit. In both examples, the T-MAPS AT service may guide the users to find places which are characteristic of a given area.

The T-MAPS services developed used the *Accumulated* metric, aiming to characterize the Manhattan region, based on our window observation. Any other metric can be applied to provide a different description, achieving a different goal. With these services (sentiment, route information and area' tags), the T-MAPS can enrich the current route recommendation systems, indicating to the users an extra path description or even providing routing based on these descriptions. For instance, the user may choose a route which expresses good feelings and beautiful environment. Alternatively, even, routes which there are cultural activities.



(a) The Route Sentiment (RS) (b) The Route Information (RI)

Fig. 9. Route sentiment based on the tweets text analysis

Fig. 10. The Area' Tags (AT) of each region of the path

VIII. Conclusions

In this work, we have presented the T-MAPS, a low-cost spatiotemporal model to enhance traffic and transit navigation context. T-MAPS bring to the navigation system, the LBSM users lens about the traffic and transit. To do that, T-MAPS uses time-varying digraphs, which models the area of interest to attach LBSM data. We showed the model applicability through a case study, where we compare the similarity between our model and Google Direction route recommendation. The results showed the median of route similarity reached 62%, where T-MAPS uses region granularity while GD uses street granularity. For a quarter of the evaluated trajectories, the similarity achieved up to 100%. Also, we presented three route description services, based on natural language analyzes, Route Sentiment (RS), Route Information (RI), and Area' Tags (AT) aiming to enhance the route information of current navigation tools.

As future work, we intend to extend the T-MAPS routes description by applying strategies to process the data and offer more valuable information. For instance, how to extract, bind and exhibit semantic information using T-MAPS? Besides that, we aim to employ regular users accounts from LBSM and uses reputation models to handle conflicting information.

Acknowledgment

We would like to thank the research agencies, CAPES, CNPq, FAPEMIG, and grant #15/24494-8, São Paulo Research Foundation (FAPESP).

References

[1] A. L. Bazzan and F. Klügl, *Introduction to intelligent systems in traffic and transportation*. Morgan & Claypool Publishers, 2013.

[2] J. Yin and Z. Du, "Exploring Multi-Scale Spatiotemporal Twitter User Mobility Patterns with a Visual-Analytics Approach," *ISPRS International Journal of Geo-Information*, vol. 5, 2016.

[3] S. S. Ribeiro Jr, C. A. Davis Jr, D. R. R. Oliveira, and W. Meira Jr, "Traffic observatory: a system to detect and locate traffic events and conditions using Twitter," in *5th ACM SIGSPATIAL*, 2012.

[4] J. Kim, M. Cha, and T. Sandholm, "SocRoutes: safe routes based on tweet sentiments," in *23rd International Conference on WWW*, 2014.

[5] A. Crooks, A. Croitoru, A. Stefanidis, and J. R, "# Earthquake: Twitter as a distributed sensor system," *Transactions in GIS*, 2013.

[6] J. Weng and B. Lee, "Event detection in twitter." *ICWSM*, 2011.

[7] M. Hasan, M. Orgun, and R. S, "A survey on real-time event detection from the Twitter data stream," *Journal of Information Science*, 2017.

[8] I. Septiana, Y. Setiowati, and A. Fariza, "Road condition monitoring application based on social media with text mining system: Case Study: East Java," in *2016 International Electronics Symposium (IES)*, 2016.

[9] K. Z. Bertrand, M. Bialik, K. Virdee, A. Gros, and Y. Bar-Yam, "Sentiment in new york city: A high resolution spatial and temporal view," *arXiv preprint arXiv:1308.5010*, 2013.

[10] A. Giachanou and F. Crestani, "Like it or not: A survey of twitter sentiment analysis methods," *ACM Computing Surveys (CSUR)*, 2016.

[11] P. Giridhar, M. Amin, T. Abdelzaher, D. Wang, L. K, J. George, and R. Ganti, "ClariSense+: An enhanced traffic anomaly explanation service using social network feeds," *Pervasive and Mobile Computing*, 2017.

[12] Y. Gu, Z. S. Qian, and F. Chen, "From Twitter to detector: Real-time traffic incident detection using social media data," *Transportation Research Part C: Emerging Technologies*, 2016.

[13] P. H. Rettore, B. P. Santos, A. B. Campolina, L. A. Villas, and A. A. Loureiro, "Towards Intra-Vehicular Sensor Data Fusion," *19th International Conference on ITS*, 2016.

[14] B. Khaleghi, A. Khamis, F. Karray, and S. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, 2013.

[15] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Exploiting the fundamental diagram of urban networks for feedback-based gating," *Transportation Research Part B: Methodological*, 2012.

[16] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.

[17] M. Jockers. (2017) syuzhet: Extracts sentiment and sentiment-derived plot arcs from text. [Online]. Available: https://cran.r-project.org/web/packages/syuzhet/