International Journal of Web Services Research
Volume 13 • Issue 4 • October-December 2016

Understanding Human Behavior in Urban Spaces using Social Network Data:
A Mobility Graph Approach

Victor R. Martinez, Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles, CA, USA & Department of Computer Science, ITAM, Mexico City, Mexico

Miguel A. Escalante, Department of Applied Mathematics, ITAM, Mexico City, Mexico

Mariano Beguerisse-Díaz, Department of Mathematics, Imperial College London, London, UK

Elmer Garduño, Sinnia, Mexico City, Mexico

Victor M. Gonzalez, Department of Computer Science, ITAM, Mexico City, Mexico

ABSTRACT

Public streams of geo-located social media information have provided researchers with a rich source of information from which to draw patterns of urban-scale human-mobility. However, most of the literature relies on assumptions over the spatial distribution of this data (e.g., by considering only a uniform grid division of space). In this work the authors present a method that does not rely on such assumptions. They followed the social media activity of 24,135 Twitter users from Mexico City over a period of seven months (June 2013 - February 2014). The authors' method clusters user's geo-locations into a 19 zone data-driven division of Mexico City. These results can be interpreted from a graph theory-based perspective, by representing each division as nodes, and the edges between them as the number of people traveling between locations. Graph centrality reveals city's infrastructural key points. Without these gateways the authors can argue that mobility would either be radically transformed or break the city apart.

KEYWORDS:

Geo-located Social Networks, Graph Theory, Human Mobility, Social Media

INTRODUCTION

Twitter, is a popular micro-blogging platform in which users can broadcast brief text updates (of 140 characters or fewer) to the public over the Internet (Bollen, Mao, & Pepe, 2011; Sakaki, Okazaki, & Matsuo, 2010). A status update message, called a tweet, is often used as a message to friends and colleagues (Java, Song, Finin, & Tseng, 2007). One user can "follow" other users, in other words, a user is subscribed to the messages generated by the other users he or she follows. An edge between two users is not necessarily reciprocal: a user can follow another user without being followed back. After its launch on July 2006, Twitter users have increased rapidly: as of 2013, the number of monthly active users is estimated around 241 million worldwide and 500 million tweets are sent per day (Twitter Inc., n. d.). In Mexico, 35% of the population is online, 82% use social media. With over 10.7 million active users and 55 million tweets sent per day (from which 18.3 million come from a mobile device), Twitter has become the 2nd most important social network in Mexico after Facebook (De Choudhury, Monroy-Hernández, & Mark, 2014; El Economista, n. d.; Ramírez, 2012). According to

DOI: 10.4018/IJWSR.2016100104

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Mexico's National Institute of Statistics and Geography (INEGI, 2010), 7.879% of all Mexicans live in Mexico City, we estimate the number of Twitter users in Mexico City to be just under one million.

In this work, we present a study of mobility at an individual scale. Our aim is to understand the daily commutes of the Twitter users in the metropolitan region of Mexico City. Our database comprises 4.1 million geo-located tweets obtained from public streams using Twitter's API. Using clustering algorithms allows us to infer the approximate home-location of 24,135 users and to focus on their day-to-day movements inside the city. By translating this information into a graph and analyzing it, we are able to discern relevant areas of the city. In summary, the contributions we present in this work are: (1) proposing a workflow to translate social-network information into a graph-theoretic framework; (3) present a data-driven separation of a specific urban environment, and (3) study mobility patterns inside Mexico City.

In the following sections we introduce related work on spatiotemporal analysis and human mobility using social networks data, specifically those extracted from Twitter. Then, we present our methodology for data acquisition and mobility graph construction, followed by a detailed analysis of the graph's properties and future outlook for this work.

Related Work

Several studies have focused on understanding Twitter from a multitude of perspectives. In 2007, Java (2007) found out the three main uses for Twitter: (1) daily chatter, posts dedicated to daily routine; (2) conversations, replies to another user using the @ symbol, and (3) sharing of information, posts containing some URL, weather reports or news story. Huberman (2008) analyzed the interaction between Twitter users following a friendship model, where each user was defined as a friend if they had at least one conversation between them in the past. Considerable work has been done on understanding how the information flows given a particular network structure (Cha, Haddadi, Benevenuto, & Gummadi, 2010; Kwak, Lee, Park, & Moon, 2010; Naveed, Gottron, Kunegis, & Alhadi, 2011; Romero, Galuba, Asur, & Huberman, 2011) and by understanding the individual motivational aspects and social impact (Beguerisse-Díaz, Garduño-Hernández, Vangelov, Yaliraki, & Barahona, 2014; Choi & Park, 2014; Signorini, Segre, & Polgreen, 2011; Tremayne, 2014). Several other works have centered on the sentiment depicted in tweets collected from a population (Bollen, Mao, & Zeng, 2011; Chen & Lazer, 2013; Fan, Zhao, Chen, & Xu, 2013; Go, Bhayani, & Huang, 2009; O'Connor, Balasubramanyan, Routledge, & Smith, 2010).

The idea of using the location of tweets (i.e., the coordinates from where the message was broadcasted) as a spatial indicator for event detection and mobility inference has been addressed with a good level of detail. Sakaki (2010) proposed a particle-filtering algorithm to detect earthquake centers in Japan by using tweets collected during the event. Takahashi (2011) regarded Twitter as a sensor of real-world phenomena including natural phenomena such as hay fever. Becker (2011) explored approaches for analyzing the stream of Twitter messages to distinguish between messages about real-world events and non-event messages. Pontes (2012) studied simple methods to infer the user home location using a collection of tweets. Hecht (2011) found that although a huge percentage of users did not specify their location beyond a city level scale, it was possible to deduce their state and city using machine-learning techniques. Bora, Zaytsev, Chang and Maheswaran (2013) studied 10 million geo-tagged tweets as observations of human movement to understand relationships of geographical regions, neighborhoods, and gang territories. Their results were threefold: (1) they were able to accurately predict the rival and non-rival nature of gang links, by using a graph representation with gang territories as vertices and gang interactions as edges; (2) showed that the travel distance of habitants of Los Angeles follows a power-law distribution, complying with several previous work on the topic, and (3) identifying major US events, popular holidays and few smaller events across the city by looking at the tweet frequency distribution.

On the topic of understanding human mobility in Mexico City, we were to identify three main related studies. Bajardi (2011) studied the role of travel restrictions in halting and delaying the 2009 H1N1 pandemic, which first reported outbreak occurred in Mexico, by using a model that integrates air travel and short-range mobility data. A study by the FIMEVIC (n. d.) (roughly translated: Trusteeship for the Improvement of Roads in Mexico City) found that Mexico City's mobility is composed of two main corridors: North to South, and West to East. Their study registered 20.57 million trips across these directions, with 33% between 6 a.m. and 9 a.m. Furthermore, the most important boroughs in terms of incoming travels are: Cuauhtémoc, Gustavo A. Madero, Benito Juárez and Miguel Hidalgo. Finally, Pérez (2014) used geo-location of tweets to understand mobility in Mexico City. He starts from the idea of dividing the city into a grid, and creating a graph where nodes represent each cell and edges count the interactions between different cells. To find the division within the city, he proposes the use of Markov Clustering Algorithm (MCL).

In this work, we extend on the results of Bora et al. (2013) in two ways. First, just as Bora did, we use a graph to approximate human-mobility using Twitter users' activity; however, in our graphs the nodes represent zones within the city and weighted edges represent the number of people moving between locations. Second, we present a data-driven division based on home locations and neighborhoods of similar density, whereas Bora used a pre-existing division of Los Angeles. Additionally, given our definition of mobility graph, we are able to present an analysis based on centrality measurements to highlight areas of vital importance to the city's human mobility. Without these areas, the mobility graph would become disconnected, and the flow pattern of people would be destroyed.

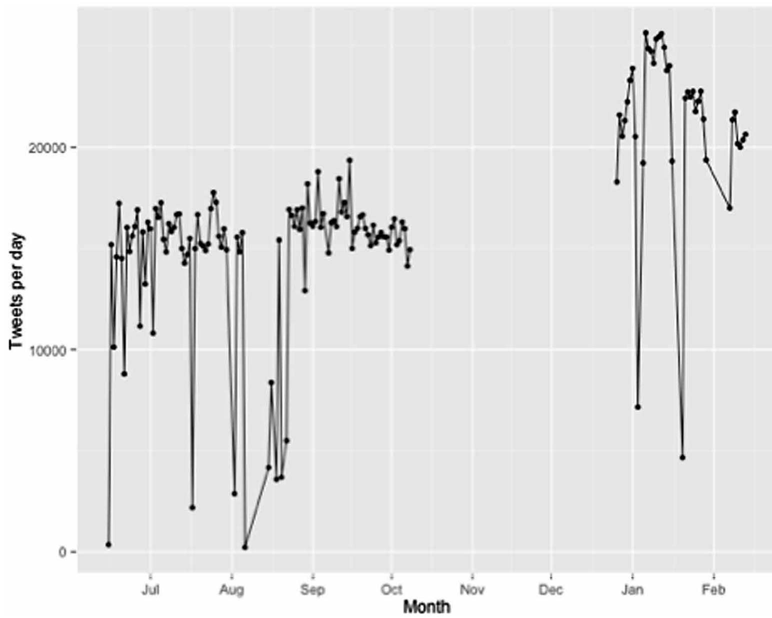
On the topic of mobility in Mexico City, we obtained similar results to those of the FIMEVIC and to Pérez (2014), specifically those regarding the importance of certain boroughs to the overall mobility, and the data-driven division of Mexico City, respectively. However, the results presented in this work differ from Pérez in the following: (1) they do not depend on an artificially constructed grid, since we cluster our data with a method that accounts for differently shaped groups, and (2) can be more easily interpreted, given that our results only rely on a distance parameter ε (with units in meters) whereas Pérez considers both the number of cells in the grid and the inflation parameter for the MCL (i.e., the Hadamard power of the random walk matrix followed by a scaling step).

DATA AND METHODOLOGY

We obtained a collection of public tweets recorded from June 15th 2013 to February 13th 2014 (4,100,000 tweets from 1,805 distinct locations around Mexico City). Each captured tweet has a unique identifier, date-time of submission, coordinates of submission (i.e., latitude and longitude), and the content of message. We were able to capture tweets in two distinct date blocks: from June 15th to October 8th, and from December 26th to February 13th, 2014. Figure 1 shows the number of captured tweets in each of the days in our dataset. This multi-date approach reduces the influence of seasonality in our mobility model, allowing for a better generalization of the results.

In order to infer the home-location of a user, we employed a density-based clustering algorithm DBSCAN (Ester, Kriegel, Sander, & Xu, 1996). The DBSCAN algorithm finds clusters of arbitrary shape, is robust to noise, and scales well to large databases. This method works by defining the notion of density-reachable. We say that two points in space p, q are directly density-reachable if there are less than ε units apart; q is density-reachable from p if there is a sequence of points $p = p_1, p_2, \ldots, p_n = q$, such that p_{i+1} is directly density-reachable from p_i. We apply DBSCAN to infer the home-location of each user. By re-applying the clustering algorithm on the cluster of home-locations, we group these locations into zones with approximately the same population density. Finally, we used these zones and the people traveling between them to construct a mobility graph.

Figure 1. Number of captured tweets per day. The capturing procedure ran from July 15th to October 8th. It ran a second time from December 26th until February 13th.



Inference of a User's Home and Urban Density Zones

In his work, Bora (2013) bases his home-location inference method on the assumptions that users generally tweet from home during the night. He retrieves a subset of tweets by time of submission, from 7 p.m. to 4 a.m. We believe this criterion is not representative of the typical working habits of Mexico where working hours can extend until much later (Booth, 2011). Hence, we use tweets sent between 10 p.m. and 4 a.m. to estimate the location of the users' homes; in our database there are 508,549 geo-located messages generated between these times of the night.

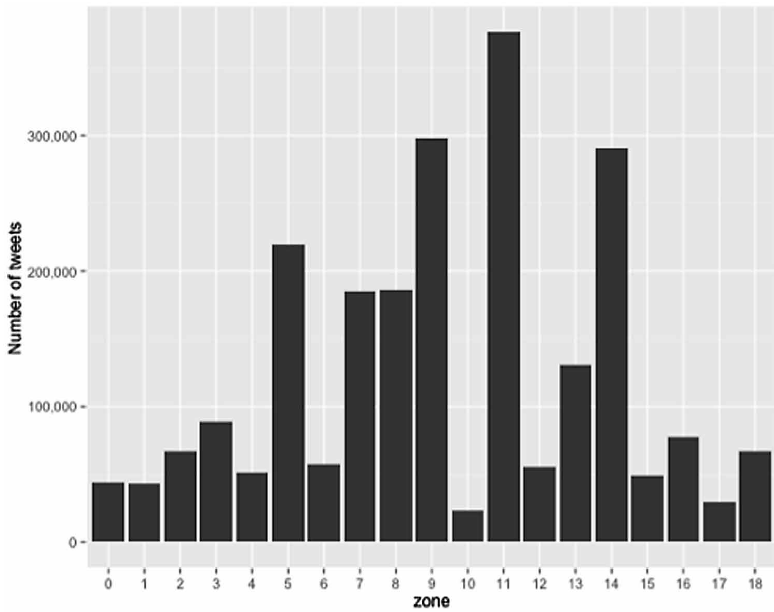
We grouped messages according to their user, and clustered their geographical information using DBSCAN with parameter $\varepsilon = 0.3$ degrees (33.396 km). As in Bora (2013), we assumed that the centroid of the biggest obtained cluster is a good approximation for the coordinates of the user's home location. We retrieved the home locations of 24,135 users (2.9% of the estimated number of Mexico City's Twitter users).

A successive application of DBSCAN (with parameter $\varepsilon = 0.0001$ degrees, roughly corresponding to 10 meters) produces a partition of Mexico City into 19 zones of similar density. Figure 2 shows the result of this procedure. Each user was assigned to a zone according to the location of their inferred home. In a similar way, each tweet was assigned to a zone according to the coordinates of submission.

Mobility Graph

As mentioned previously, we represent the flow of people between the different zones using a weighted, directed graph where each node represents a zone and each edge-weight is proportional to the number of persons from one zone tweeting in another. The mobility graph in this work is a fully connected asymmetric graph (i.e., a clique) including self-loops with19 nodes and 361 edges.

Figure 2. Number of captured tweets in each zone



RESULTS

Mobility Patterns in Mexico City

Figure 2 shows the number of tweets broadcast in each zone. We obtained the mean traveled distance (MTD) for each user (i.e., the average of the haversine formula between home-location coordinates and the coordinates of each captured tweet for the same user). Figure 4a shows the distribution of the MTD for the users in our data. Previous works on understanding human mobility (González, Hidalgo, & Barabási, 2008; Song, Qu, Blumm, & Barabási, 2010) suggested that a Lévy flight model, commonly used to approximate movement and migration patterns in animals, could approximate human mobility. Under this model, motion is the result of a sequence of random steps (random walk) whose step size follows a power-law distribution (Newman, 2005). We investigated if such a model could explain our measurements in MTD. This hypothesis was rejected using the test proposed by Clauset (2009), as evinced by the cumulative probability distribution of the users' MTD in Figure 4Figure 4b. This is not surprising if we consider the fact that we are measuring a distribution of means in a very limited geographical region (we do not consider journeys outside of the metropolitan area of Mexico City).

We obtained summary statistics for each zone (shown in Table 1 in the appendix). Users from Zone 7, centered in Mexico City's International Airport, attain the highest MTD. Those with the lowest MTD (Zones 8, 13 and 15) belong to the northeast part of the city; the data indicates that users in these regions travel shorter distances that the rest of the city's inhabitants and tend to stay closer to their home location. If we restrict ourselves only to intra-metropolis movement, Zones 1, 13 and 16 attain the minimum traveled distance, due to their central location and the fact that these areas contain a huge number of the touristic attractions, city parks, and commercial establishments all within a short distance. In contrast, Zones 14, 12 and 4 do not contain enough services and attractions forcing their habitants to travel longer distances into the city in order to secure them.

Graph Analysis

As we mentioned previously, our fully connected (including self-loops) mobility graph consists of 19 nodes; Figure 5a shows the graph's spatial embedding. For each node, we calculated the in and out weighted degree, that is, the number of journeys arriving and departing from each zone. Figure 5b and Figure 5c show the obtained weighted degree distributions of all nodes.

We compute the betweenness centrality of each node in the mobility graph. The betweenness centrality is proportional to the number of times appears in a shortest path between two nodes (Freeman, 1977); we denote this centrality by $g(v)$ as given by

$$g\left(v\right)=\sum_{s\neq v\neq t}\frac{\sigma_{st}\left(v\right)}{\sigma_{st}}$$

In a mobility context, areas with a high centrality score correspond to parts of the city where people must transit every day (see Figure 3).

Figure 6 shows the mobility graph where the size of the nodes is in direct relation to each node's betweenness centrality. We find that Zone 15 is the most central node in the graph, as much of the traffic generated in the densely populated north, northwest zones and even Mexico City's satellite town must transit through it. Similarly, Zone 10 represents the meeting point of the city's most important avenues, including Insurgentes Avenue, a 28.8 km north-south axis, Miguel Aleman's freeway, a crosscutting freeway that runs east-west across the middle of the city and Circuito Interior boulevard, a 42 km-long boulevard that forms a loop around the city. We attribute the importance of this zone to its role as a meeting point for north and south and east and west.

CONCLUSION AND FUTURE WORK

The aim of this paper was to establish a data-handling protocol to process geo-referenced tweets into a dynamic graph to provide insights into Mexico City's mobility patterns. This graph and the analyses that can be performed on it have considerable potential value to policy makers and urban planners. We obtained a partition of Mexico City's metropolitan region into 19 zones of equal population density. We used the location of the tweets to infer the number of people traveling between these zones.

Our work provides an approach to producing data-driven separation of an urban environment based on information density beyond externally imposed boundaries. The results show that boroughs in Mexico City can only partially explain the activity defined by social-networks records; and our approach provides new ways to understand the urban dynamics. Furthermore, our approach can be used to perform analyses with different levels of granularity based on information density. The application of this approach is part of our future work that we will conduct with researchers from other disciplines.

Our centrality analysis reveals certain key points of the city, which coincide with known locations of the city's transport infrastructure. Without these gateways we can argue that mobility would either be radically transformed or break the city apart. This confirms popular perception in terms of the critical role of some points of the city for transit.

While consistent with prior research on the subject, the conclusions of this paper are limited in scope by a sample that is not representative of the general population, only of Twitter users with geo-tagging enabled. However, given that our results are consistent with key features of the city that are extraneous to the data, we have confidence that they are a useful indication of the mobility patterns in the metropolitan region. Furthermore, we believe that our results point to interesting questions for future research such as measurement of distance traveled by individual "tweeters" or the proximity between users via non-spatial proxy variables that substitute and complement geo-spatial variables.

Figure 3. Cluster of home-locations into density-zones

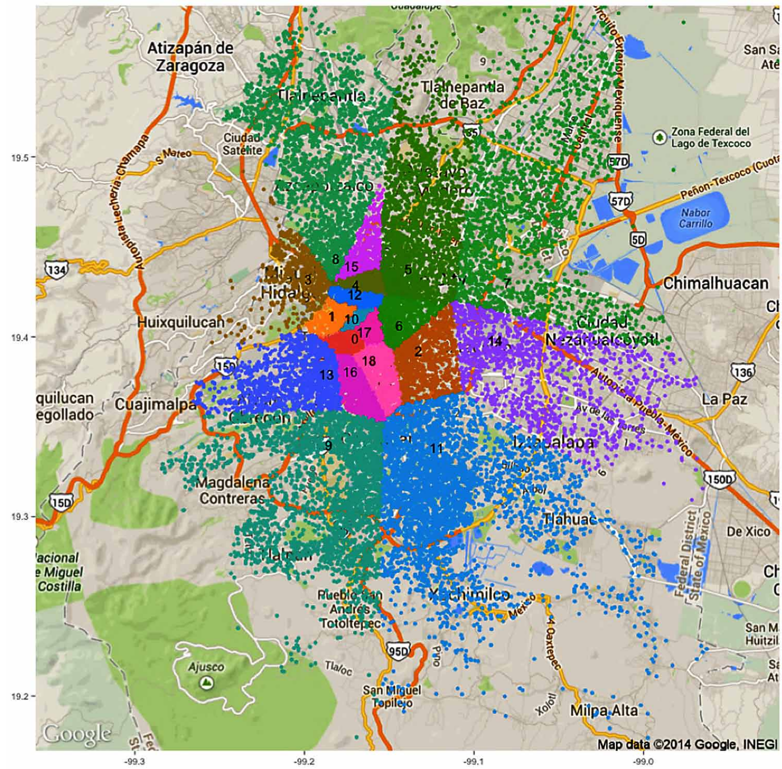
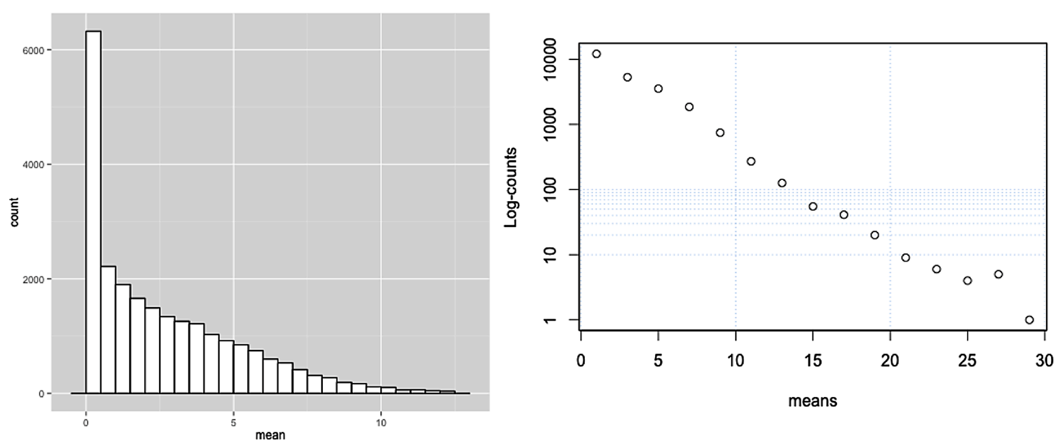


Figure 4. (a) Mean traveled distance distribution. (b) Same distribution plotted over log scale. If the hypothesis of power law holds, (b) should be a straight line.



(a) (b)

Figure 5. (a) Graph representation of the obtained density zones. (b) In-degree distribution. (c) Out-degree distribution.

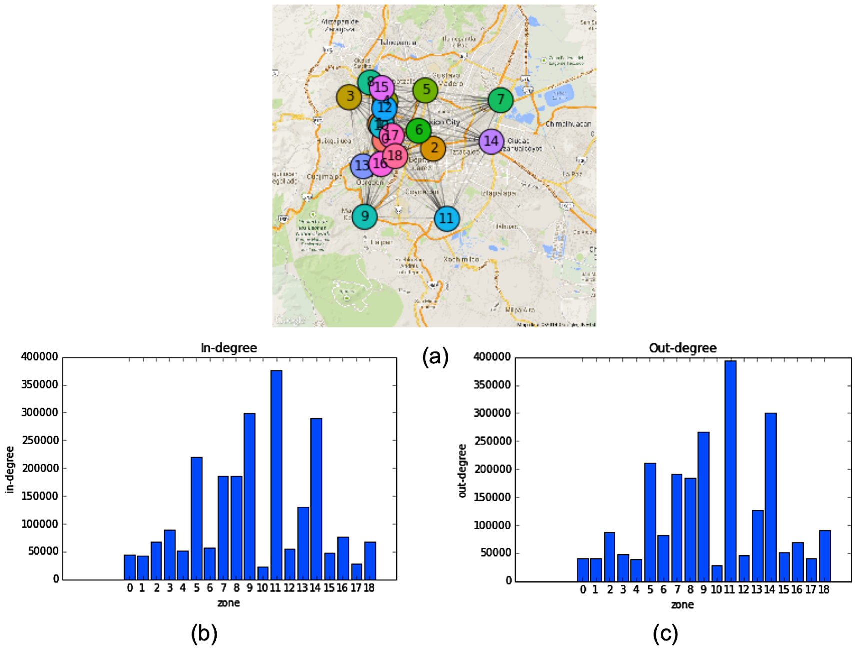
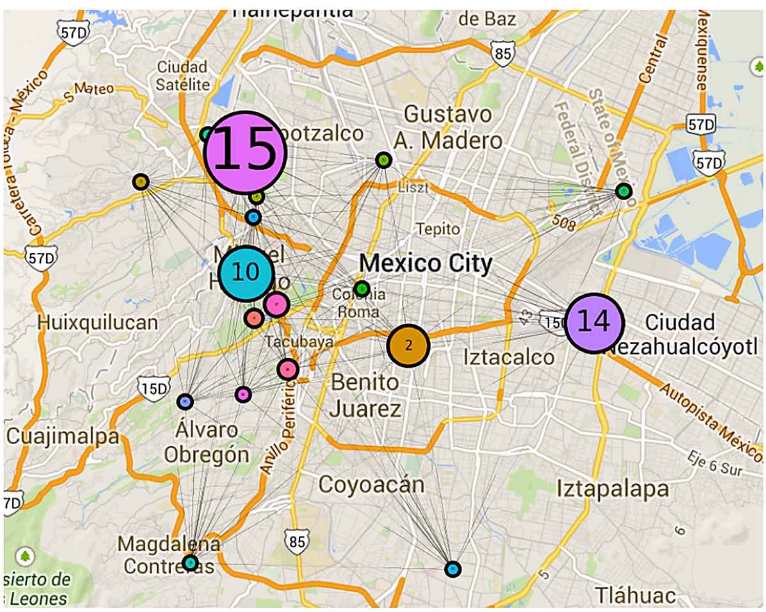


Figure 6. Betweeness centrality. Each node's size is in direct relation to its betweeness centrality. The three most important zones are 15, 14 and 10. The first two representing entering traffic from the north and east respectively and the last connects the south of the city to the north.



ACKNOWLEGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. We would also like to thank everyone involved during the course of this study, specially Jesús D. Arroyo for reviewing our statistical analysis; Consejo Nacional de Ciencia y Tecnología (CONACYT), and to the Asociación Mexicana de Cultura A.C. for all their support. MBD acknowledges support from the James S. McDonnell Foundation Postdoctoral Program in Complexity Science/Complex Systems Fellowship Award (#220020349-CS/PD Fellow).

REFERENCES

Bajardi, P., Poletto, C., Ramasco, J. J., Tizzoni, M., Colizza, V., & Vespignani, A. (2011). Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic. *PLoS ONE*, *6*(1), e16591. doi:10.1371/journal.pone.0016591 PMID:21304943

Becker, H., Mor, N., & Gravano, L. (2011, July 17-21). Beyond Trending Topics: Real-World Event Identification on Twitter.*Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain (pp. 438-441).

Beguerisse-Díaz, M., Garduño-Hernández, G., Vangelov, B., Yaliraki, S. N., & Barahona, M. (2014). Interest communities and flow roles in directed networks: The Twitter network of the UK riots. *Journal of the Royal Society, Interface*, *11*(101), 20140940. doi:10.1098/rsif.2014.0940 PMID:25297320

Bollen, J., Mao, H., & Pepe, A. (2011, July 17-21). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena.*Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. doi:10.1016/j.jocs.2010.12.007

Booth, W. (2011). Siesta? What siesta? Mexican work longest hours in world. The Washington Post. Retrieved from http://www.washingtonpost.com/world/siesta-what-siesta-mexican-work-longest-hours-in-world/2011/04/27/AF3O0yTF_story.html

Bora, N., Vadim, Z., Yu-Han, C., & Rajiv, M. (2013, September 8-14). Gang Networks, Neighborhoods and Holidays: Spatiotemporal Patterns in Social Media.*Proceedings of the 2013 International Conference on Social Computing*, Washington D.C., USA (pp. 93-101). doi:10.1109/SocialCom.2013.21

Chen, R., & Marius, L. (2013). Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement. *stanford.edu.* Retrieved from http://cs229.stanford.edu/proj2011/ChenLazer-SentimentAnalysisOfTwitterFeeds ForThePredictionOfStockMarketMovement.pdf

Choi, S., & Park, H. W. (2013). An exploratory approach to a Twitter-based community centered on a political goal in South Korea: Who organized it, what they shared, and how they acted. *New Media & Society*, *16*(1), 129–148. doi:10.1177/1461444813487956

Clauset, A., Cosma, R. S., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703. doi:10.1137/070710111

De Choudhury, M., Monroy-Hernandez, A., & Mark, G. (2014, April 26-31). Narco emotions: affect and desensitization in social media during the mexican drug war. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, Canada (pp. 3563-3572). doi:10.1145/2556288.2557197

El Economista. (n. d.). Twitter alcanza en México 10.7 millones de usuarios | El Economista. Retrieved from http://eleconomista.com.mx/tecnociencia/2012/03/21/twitter-alcanza-mexico-107-millones-usuarios

Ester, M., Hans-Peter, K., Jörg, S., & Xiaowei, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, USA, August 2–4, 226-231.

Fideicomiso para el Mejoramiento de las Vías de la Comunicación del Distrito Federal. (n. d.). Diagnóstico de la movilidad de las personas en la Ciudad de México [Diagnostic of human-mobility in Mexico City]. Retrieved from http://www.fimevic.df.gob.mx/problemas/1diagnostico.htm

Freeman, L. C. (n. d.). A set of measures of centrality based on betweenness. In *Sociometry* (pp. 35-41).

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. doi:10.1038/nature06958 PMID:18528393

Hecht, B., Hong, L., Suh, B., & Chi, E. H. (2011, May 7-12). Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles.*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada (pp. 237-246). doi:10.1145/1978942.1978976

Huberman, B. A., Romero, D. M., & Wu, F. (2008). *Social networks that matter: Twitter under the microscope*. Social Science Research Network.

Instituto Nacional de Estadística y Geografía. (2010). Censo de población y vivienda 2010 [Population and housing census 2010]. Retrieved from http://www3.inegi.org.mx/sistemas/mexicocifras/default.aspx?e=9

Java, A., Song, X., Finin, T., & Tseng, B. (2007, August 12-15). Why we twitter: understanding microblogging usage and communities, *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, San Jose, CA, USA (pp. 56-65). doi:10.1145/1348549.1348556

Kwak, H., Changhyun, L., Park, H., & Moon, S. (2010, April 26 - 30). What is Twitter, a social network or a news media?*Proceedings of the 19th international conference on World wide web*, Raleigh, NC, USA (pp. 591 – 600). doi:10.1145/1772690.1772751

Lonely Planet. (n. d.). Mexico City. Retrieved from http://www.lonelyplanet.com/mexico/mexico-city

Meeyoung, C., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010, May 23-26). Measuring User Influence in Twitter: The Million Follower Fallacy.*Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, USA (p. 30).

Naveed, N., Gottron, T., Kunegis, J., & Alhadi, A. C. (2011, June 15 - 17). Bad news travel fast: A content-based analysis of interestingness on twitter. *Proceedings of the 3rd International Web Science Conference*, Koblenz, Germany. doi:10.1145/2527031.2527052

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351. doi:10.1080/00107510500052444

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199206650.001.0001

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010, July 17-21). *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain (pp. 122–129).

Pérez, L. E. (2014). Uso de Twitter como red de sensores para describir el movimiento de personas en la Ciudad de México [Describing human-mobility using Twitter as a sensor network]. Unpublished undergraduate thesis, Instituto Tecnológico Autónomo de México (ITAM), Mexico City, Mexico.

Pontes, T., Magno, G., Vasconcelos, M., Gupta, A., Almeida, J., Kumaraguru, P., & Almeida, V. (2012, December 10). Beware of What You Share: Inferring Home Location in Social Networks. *Proceedings of the 12th International Conference on Data Mining Workshops*, Brussels, Belgium (pp. 571–578). doi:10.1109/ICDMW.2012.106

Ramírez, R. (2012). Twitter, datos y estadísticas. Retrieved from http://isopixel.net/archivo/2012/02/twitter-datos-y-estadisticas/

Romero, D. M., Galuba, W., Asur, S., Huberman B.A. (2011). Influence and passivity in social media. In *Machine learning and knowledge discovery in databases* (pp. 18-33). Springer Berlin Heidelberg.

Rui, F., Jichang, Z., Chen, Y., & Xu, K. (2013). Anger is more influential than joy: Sentiment correlation in Weibo. *PLoS ONE*, *9*(10). PMID:25333778

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010, April 26-30). Earthquake shakes Twitter users: real-time event detection by social sensors.*Proceedings of the 19th international conference on World wide web*, Raleigh, North Carolina, USA (pp. 851-860). doi:10.1145/1772690.1772777

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLoS ONE*, *6*(5), e19467. doi:10.1371/journal.pone.0019467 PMID:21573238

Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. *Science*, *327*(5968), 1018–1021. doi:10.1126/science.1177170 PMID:20167789

Takahashi, T., Abe, S., & Igata, N. (2011). Can twitter be an alternative of real-world sensors? Human-Computer Interaction: Towards Mobile and Intelligent Interaction Environments, LNCS (Vol. 6763, pp. 240-249). Springer Berlin Heidelberg,

Tremayne, M. (2014). Anatomy of Protest in the Digital Era: A Network Analysis of Twitter and Occupy Wall Street. *Social Movement Studies*, *13*(1), 110–126. doi:10.1080/14742837.2013.830969

Twitter Inc. (n. d.). About Twitter, Inc. Retrieved from https://about.twitter.com/company

Victor R. Martinez is a first-year PhD student in University of Southern California Viterbi School of Engineering. Currently, he is part of the Signal Analysis and Interpretation Lab led by Professor Shrikanth Narayanan. He received an MSc degree in Computer Science, BSc in Computer Engineering, and BSc Applied Mathematics from Instituto Tecnológico Autónomo de México (ITAM). His research interests lie somewhere in between Social Computing and Natural Language Processing. Miguel A. Escalante works as a Data Scientist. He holds a BSc in Applied Mathematics from Instituto Tecnológico Autónomo de México (ITAM), where he is also about to complete his MSc in Data Science. Over the past 5 years he has worked with both private and public institutions transforming and analyzing data in order to make better decisions. He currently is a Semi-Senior Data Scientist in Globant, a software consulting company from Argentina. Miguel specializes in database management, machine learning, big data infrastructure design, data analysis statistics, optimization, numerical calculus, and differential equations. Mariano Beguerisse-Díaz is a Research Fellow at Department of Mathematics, Imperial College London. His research interest is the mathematical representation of biological and human made systems with a special focus on complexity and networks. He received his PhD in Mathematical Biology from Imperial College London in 2012, a MSc in Mathematical Modeling and Scientific Computing from University of Oxford, and a BSc in Applied Mathematics from Instituto Tecnologico Autonomo de Mexico.

Victor M. González is Professor and Head of the Academic Department of Computer Engineering at Instituto Tecnologico Autonomo de Mexico. He works on research projects investigating the Interaction of People with Computers and Information Systems, with a particular emphasis on phenomena related to work fragmentation, information overload and attention deficit. He received a PhD and MSc degrees in Information and Computer Science from the University of California at Irvine, an MSc in Telecommunication and Information Systems from the University of Essex, and a BSc in Electronic Communication Engineering from Tech of Monterrey. He is a Level 1 member of the National Systems of Research (CONACyT) and past president of the Computer Science Mexican Society.

APPENDIX

Table 1. Mean traveled distance and standard deviation for every zone in Mexico City

zone	Mean traveled distance (km)	Standard deviation (km)
0	2.77	6.83
1	2.73	3.62
2	3.06	8.38
3	2.62	11.99
4	3.00	13.41
5	2.83	12.32
6	2.73	6.90
7	3.78	7.41
8	2.34	4.40
9	2.56	4.49
10	3.06	7.99
11	2.58	4.40
12	2.97	18.42
13	2.48	3.54
14	2.97	12.52
15	2.54	3.91
16	2.88	3.52
17	3.15	3.69
18	3.01	5.61