Analysis of Driving Safety and Cellphone Use Based on Social Media

Chao Qian[1,2], Yueqing Li[1,1], Wenchao Zuo[1,2] and Yuhong Wang[2]

[1] Department of Industrial Engineering, Lamar University, Beaumont, Texas 77710, USA
[2] School of Business, Jiangnan Universtiy, Wuxi, Jiangsu 214122, China
{cqian, yueqing.li}@lamar.edu, wenchaozuo@gmail.com, yuhongwang@jiangnan.ed.cn

Abstract. The use of mobile phones while driving is a major driver distraction leading to traffic accidents. Using Twitter Archiving Google Sheet, this paper collected 5,208 tweets containing the hashtags: #distracteddriving, #textanddrive, #textinganddriving. By using R studio and Tableau, the tweets data was visualized and aggregated. The virtual cellphone distracted driving community was also mapped through Gephi. The results could visualize the snapshot of the attitudes and opinions surrounding driving safety and cellphone use.

Keywords: Text Mining · Driving Safety · Cellphone Use · Twitter

1 Introduction

The usage of mobile phones while driving is a primary driver distraction which makes traffic accidents and fatal crashes increased. Observational survey conducted in the United Sates exposed that 32.7% of drivers have distracting activities, which include talking on the phone (31.4%), and texting or dialing a phone (16.6%) [1]. It reported that 3.4% of drivers use handheld cellphone while driving in the United Kingdom [2], and 141.1% in Spain [3]. The National Highway Traffic Safety Administration (NHTSA) reported that approximately 481,000 drivers have used cell phones while driving during daylight hours [4]. This cellphone distracted driving will remain stable or even increase due to the high degree of integration of this technology into society [5].

Social media emerges as a revolutionary way for the connection among individuals through the development of Internet-based communication services. There are different social media platforms, such as Twitter, Facebook, LinkedIn and YouTube. Twitter, as a significant platform for public communication, has 313 million active users per month and 500 million tweets per day around the world [6]. It has become a realtime data network for information sharing as well as a discussion about the political and social issues. Users do so by creating short message under 140 characters, known as tweet, which can be shared by other users. Besides pictures, videos and websites, tweets also contain hashtags (words begin with #) which can turn into links to find certain terms

[1] Corresponding author: Yueqing Li, yueqing.li@lamar.edu

easily. With the Application Programming Interface (API) of Twitter developer, Twitter data is relatively straightforward to collect and analyze, which can be useful in understanding the attitudes and opinions about driving safety and cellphone use.

Compared to typical quantitative analysis applied to numerical data, text data is "unstructured, amorphous and difficult to deal with" [7]. Different from the driving simulator and questionnaire survey, text mining can better retrieve people's attitudes on cellphone distraction of driving directly through finding patterns in text and extracting useful information. Text mining also make it possible to measure the quantitative data of tweets and create a more historical vision of distracted driving study. This paper has collected 5,208 tweets containing the hashtags: #distracteddriving, #textanddrive, #textinganddriving, from December 24th, 2018 to January 14th, 2019 through Twitter Archiving Google Sheet (TAGS version 6.1). By using R studio (R studio Software Inc, Boston, MA) and Tableau (Tableau Software Inc, Sattle, WA), the Twitter data become visualization and aggregation. Gephi (Gephi version 0.9.1) is used to map the virtual cellphone use distracted driving community. These three methodological tools are provided to visualize the snapshot of the attitudes and opinions surrounding the driving safety and cellphone use.

2 Literature Review

Researchers have proposed different perspectives and methods to estimate this significant study filed recently. The three common studies of cellphone distracted driving are reported data, questionnaire survey and driving simulator experiment or naturalistic driving study.

William et. al (2016) showed that the handheld electronic devices usage is detrimental to driver safety via a National Academy of Sciences-sponsored naturalistic driving dataset [8]. Claudio et. al (2018) collected the data on road crashes with fatalities from seven Italian metropolitan areas from National Institute of Statistics (ISTAT) to analyze the association between mobile phone traffic volume and road crash fatalities [9].

The 2015 National Survey on Distracted Driving Attitudes and Behavior (NSDDAB) conducted by NHTSA made a national survey of the distracted driving activities including the cellphone use among 6,001 interviews over 16 who had driven a motor vehicle [10]. Oscar et. al (2017) made a survey of 484 drivers using an online questionnaire to analyze the risk factors of mobile phone use while driving in Queensland [11]. Amanda et al. (2018) made an online survey of 18 young adult drivers to explore the mobile phone functions used while driving [12]. Motao et. al (2018) enrolled 22 experts and 7 young drivers to propose behavioral indicators and consequential indicators through a Delphi Survey [13].

Noam et. al (2013) performed two experiments to study whether decisions to engage in cell phone conversation while driving and the consequences of such decisions [14]. Oscar et. al (2019) investigated the influence of driving demands, secondary task characteristics, and personal characteristics on behavioral adaptation of mobile phone distracted drivers through driving simulator experiment [15]. Puspha et al. (2019) did a comparative analysis of young drivers (N=25) and professional drivers (N=24) on their driving performance with the help of driving simulator experiments [16]. A data

acquisition system has been used to continuously record video of driver's and vehicle's message including cell phone use from 204 drivers who took part in the study for 31 days (on average) from February 2011 to November 2011 by NHTSA [17].

Database from Security Department often suffers from the neglect of recording the exposure to mobile phone use, which may be inaccurate in estimation. Questionnaire and online survey can collect the subjective results but can't reflect the drivers' actual behaviors. Driving simulator experiment is not suitable for estimating natural crash risk as the virtual scenario. Hence, the use of social media data of cellphone distracted driving shall be an exploratory vision, to better understand the attitudes and opinions about cellphone distracted driving. Although the social media has been a popular data source in other research areas, it has seldomly been used in the research of driving safety.

The rest of this paper is organized as follows. Section 3 clarifies the methods of acquisition and analysis of the social media data. Section 4 presents the empirical results. Section 5 summarizes the study and make a discussion.

3 Materials and Methods

3.1 Social Media Data Acquisition

To obtain the tweets of interest, this paper used Twitter Archiving Google Sheet (TAGS version 6.1). TAGS sampled tweet data in real time using user-defined hashtag search terms over a defined period. These hashtags were chosen to collect a broad scope of tweets, while not bias the sample by targeting specific tags. The following information was stored into a Google spreadsheet: Twitter username, userprovided geographic identifier, tweet content, tweet generated time, user follower count, and user friend count.

3.2 Social Media Network Data Analysis

The goal of the social media analysis was to characterize the interconnection between driving safety and cellphone use. R studio (R studio Software Inc, Boston, MA) and Tableau (Tableau Software Inc, Seattle, WA) were used in the data analysis. The statistics for term frequencies were displayed through Word Cloud by R studio. Other basic statistics for users, user activities, tweet contents, temporal trends were tabulated by Tableau.

3.3 Mapping the Distracted Driving Community Network

Open source network visualization software (Gephi version 0.9.1) was used to map the virtual cellphone use distracted community. It uses specialized algorithms to map and measure the distance between users and followers based on the number of interactions between users. Under the operation of the Forceatlas2 algorithm, the linkages between Twitter users were generated as a spatial map by the network software [18].

4 Results

This paper has collected 5,208 tweets from December 24th, 2018 to January 14th, 2019, through the hashtags related to driving safety and cellphone use: #distracteddriving, #textanddrive, and #textinganddriving.

4.1 Social Media Network Data Analysis

Temporal patterns analysis generated the number of hashtagged tweets over the study period time. Three types of tweets (@mention, original tweet, retweet) are shown in different colors. A tweet wrote primitively by a user is an original one. While shared publicly by other users with their followers, it known as a retweet. When you want to mention someone, use @ before their username, it is an @mention.

In this study, a spike in retweets identified one marked tweet that corresponded with a widely publicized alarming. The alarming declared that "distracted driving (mainly refers to the cellphone usage while driving) kills & maims more people than drunk drivers". The tweet has received relatively extensive attention with retweet by other Twitter users. (Fig. 1; the content of spike shown inset).

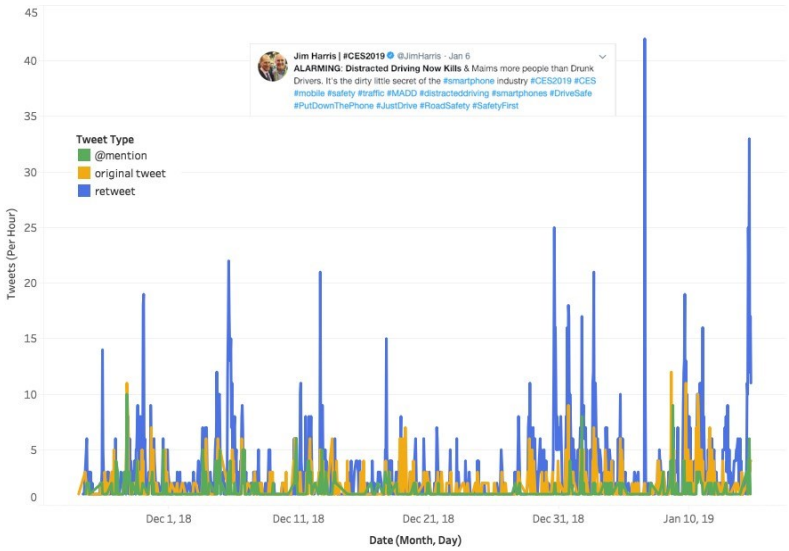

Fig. 1. Temporal patterns of tweets per hour over the study period. Insert is characterization of the alarming on January 6th, 2019.

The word cloud represents the frequency of words appeared top 100 in the collected tweets dataset (Fig. 2). The words "driving" and "distracted" are associated to the hashtags (#distracteddriving, #textanddrive, #textinganddriving) directly. The words "drivers" and "phone" reflect the causes of distracted driving, in which the "drivers" is the crash-maker and the "phone" is the main distraction.

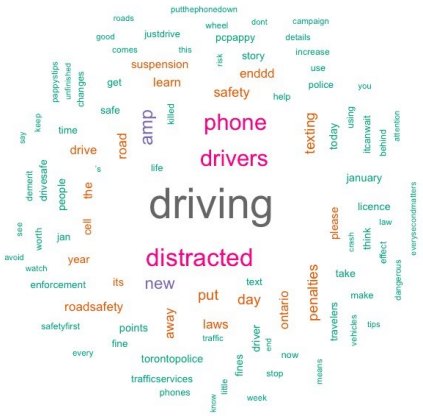

Fig. 2. The word cloud displaying frequently tweeted stemmed terms. Words that larger represent more frequent terms.

The visible and active user is the statistic of the number of tweets from different users. In this study period, the most visible user was “pcpappy”, and the most active user was “Cellcontrol” (Fig. 3). The “pcpappy” user account as an individual user is a policer of Traffic Enforcement (Community Complaints) in Toronto Police Service. The “Cellcontrol” user account represents a company (https://www.cellcontrol.com/) which focus on the technology to eliminate distracted driving.

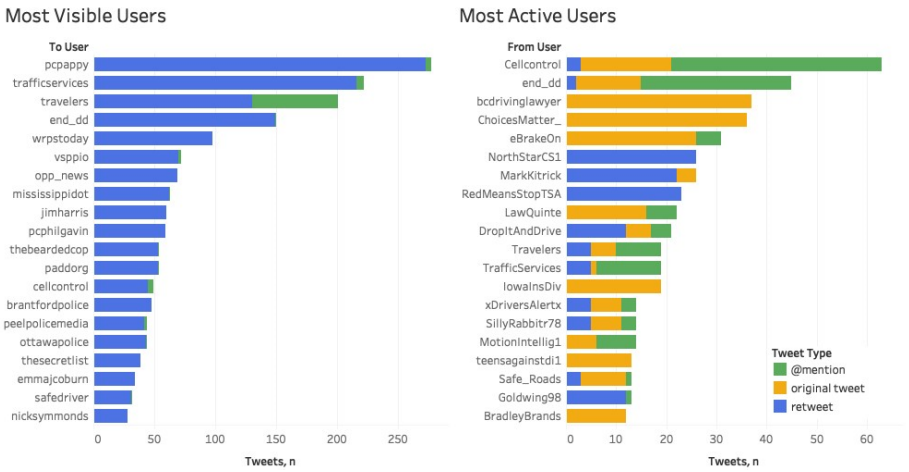

Fig. 3. The most visible Twitter users (Top 20) and most active Twitter users (Top 20) in the study period.

Twitter users use multiple hashtags in their tweets. The hashtags including “pappystips”, “unfinishstories”, “itcanwait” and “smartphone” were the most frequent words (>60 tweets) coupled with the hashtags used for collecting tweets (Fig. 4). The distribution of the secondary hashtags by hours indicated the users’ different preference

and habits on time. Most tweets with secondary hashtags (>80 tweets) occurred from 2 pm to 3 pm, and fewest tweets occurred from 6 am to 8 am (Fig. 5).

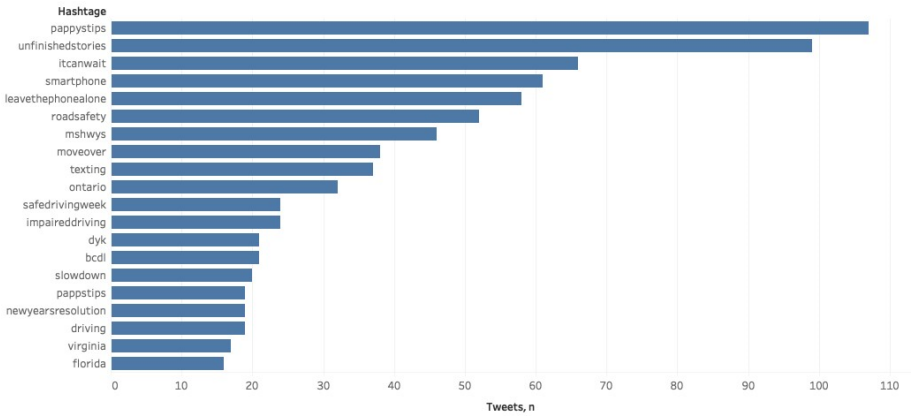

Fig. 4. The secondary hashtags (Top 20 of 488 members) in the study period. The view is filtered on hashtags which excluded: #distracteddriving, #textanddrive, #textinganddriving.

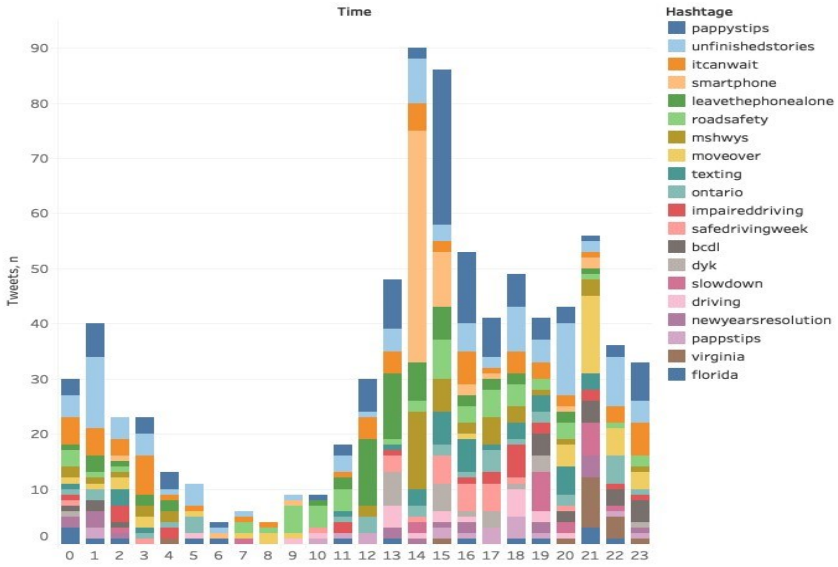

Fig. 5. The distribution of the secondary hashtags (Top 20 of 488 members) by time (Hour) in the study period. The color shows detail about the secondary hashtags.

4.2 Mapping the Distracted Driving Community Network

Based on the visualization of the cellphone distracted driving community, some major trends appeared in Fig. 6. Two major and distinct subcommunities can be recognized, which including different types of accounts.

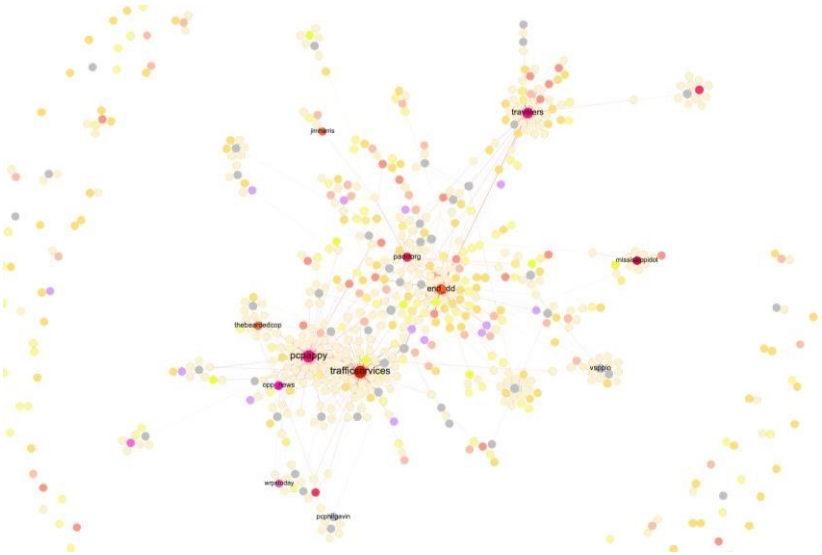

Fig. 6. Twitter social media network map. Nodes mean user accounts. Edges mean Twitter follower connections. Colors represent the algorithm-derived clusters of similarity.

Fig. 7 reflects the subcommunities which combined with different police service institutions. The “pcappy” as an individual account, and “trafficservices” as an official account are both from Toronto Police Service. The “opp_news” and “wrpstoday” are different police agencies from Ontario and Waterloo.

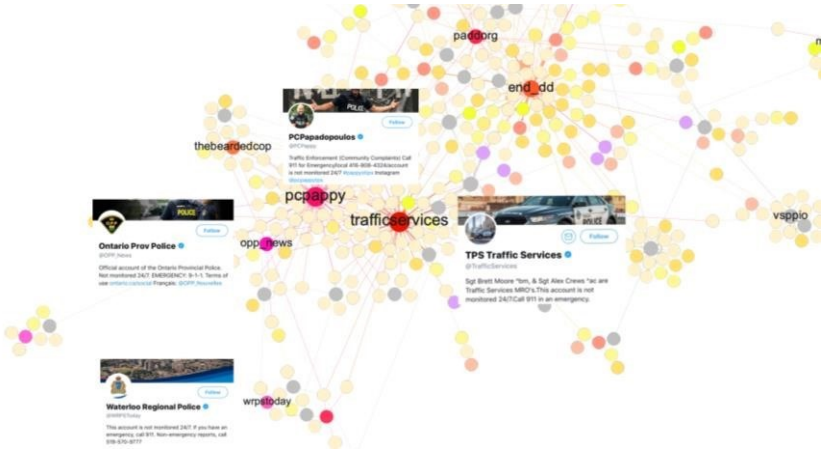

Fig. 7. Twitter cellphone distracted driving network subcommunity (Police service institutions). Insets are representative user account profiles.

Fig. 8 indicates several social justice organizations including “end_dd” (a website against distracted driving), “travelers” (an insurance provider for auto), “paddorg” (an

organization against distracted driving), “mississippidot” (a website for up-to-date roadway information).

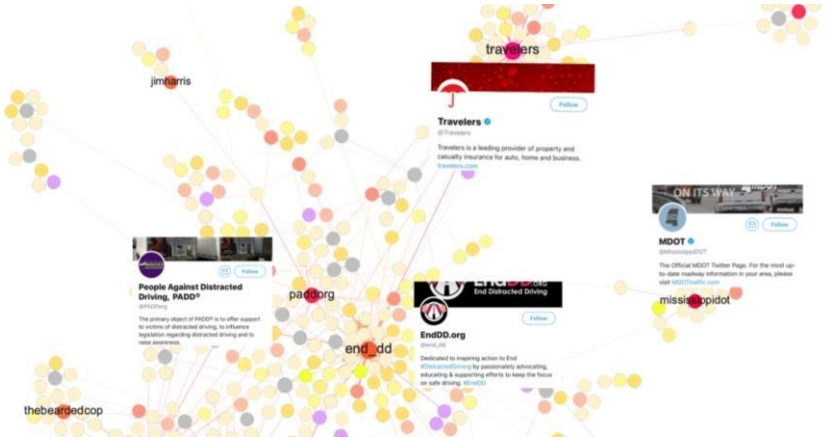

Fig. 8. Twitter cellphone distracted driving network subcommunity (Social justice organizations). Insets are representative user account profiles.

5 Discussion & Conclusion

The purpose of this research was to use text mining tools to explore social media about driving safety and cellphone use. Different text mining software made it possible to analyze the trends, frequencies and associations of Twitter data. The results above indicated that:

(1) The volume of retweets is far more than other two types (original and @mention tweet). Many of the statistical facts, media reports, and driving alarming about distracted driving were retweeted. For example, the alarming tweet that the distracted driving, especially the cellphone use kills more people than drunk driving, received most retweets in the study period. The retweeting of these tweets makes sure the increasing exposure of the valuable information.

(2) The majority of the top 20 active Twitter users are commercial/for-profit institutions. Most of the top 20 visible Twitter users come from organizations or agencies rather than individual users. The similar phenomenon can be discovered in the Twitter social media network map. Considering that the social media can enlarge the influence scale, increase the awareness degree and attract more audiences, it is not surprised to find out these results.

(3) The attitudes and opinions of Twitter users can be noticed from the word cloud (Fig. 2) and secondary hashtags (Fig. 4 and Fig. 5). Except the words associated to the hashtags directly, other words like “avoid” and “stop” can reflect the attitudes of the Twitter users. The other hashtags combined with the three used for collecting tweets, like “itcanstop”, “leavethephonealone” can see the opinions of the Twitter users.

This exploratory research still has some limitation which can be improved in the future work. It’s significant to select suitable and associated hashtags to make sure the quality of the tweets. Another necessary work is to lengthen the study period to enrich

the amount of the Twitter dataset. Further analysis of these data includes the relationship and association between terms. The specific classification of large scales of users, the definitions of the abbreviations of hashtags and the geographic location are other interesting paths. In addition, how to combine the social media with the real traffic accidents is a research perspective with realistic meaning.

Acknowledgements

This study was partially supported by the Natural Science Foundation (1726500). The findings and conclusions of this paper are those of the authors and do not necessarily represent the official position of NSF.

References

1. Huisingh, C., Griffin, R., & McGwin, G.: The Prevalence of Distraction Among Passenger Vehicle Drivers: a Roadside Observational Approach. Traffic Injury Prevention. 16(2), 140-146 (2015)
2. Sullman, M. J., Prat, F., & Tasci, D. K.: A Roadside Study of Observable Driver Distractions. Traffic Injury Prevention. 16(6), 552--557 (2015)
3. Prat, F., Planes, M., Gras, M. E., & Sullman, M. J. M.: An Observational Study of Driving Distractions on Urban Roads in Spain. Accident Analysis & Prevention. 74, 8--16 (2015)
4. Distracted Driving, National Highway Traffic Safety Administration, https://www.nhtsa.gov/risky-driving/distracted-driving
5. Young, K., & Regan, M.: Driver Distraction: a Review of the Literature. Distracted driving, 379--405 (2007)
6. Twitter Statistics Directory, Socialbakers, https://www.socialbakers.com/statistics/twitter/?v=1243
7. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Cambridge (2016)
8. Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., King, M. B., & Hankey: Driver Crash Risk Factors and Prevalence Evaluation Using Naturalistic Driving Data. Proceedings of the National Academy of Sciences. 113(10), 2636--2641 (2016)
9. Gariazzo, C., Stafoggia, M., Bruzzone, S., Pelliccioni, A., & Forastiere, F.: Association between Mobile Phone Traffic Volume and Road Crash Fatalities: A Population-based Casecrossover Study. Accident Analysis & Prevention. 115, 25--33 (2018)
10. National Survey on Distracted Driving Attitudes and Behaviors 2015, https://www.nhtsa.gov
11. Trespalacios, O. O., King, M., Haque, M. M., & Washington, S.: Risk Factors of Mobile Phone Use While Driving in Queensland: Prevalence, Attitudes, Crash Risk Perception, and Task-management Strategies. PLoS One. 12(9), e0183361 (2017)
12. George, A. M., Brown, P. M., Scholz, B., Parker, S. B., & Rickwood, D.: "I Need to Skip a Song Because It Sucks": Exploring Mobile Phone Use While Driving Among Young Adults. Transportation Research Part F: Traffic Psychology and Behaviour. 58, 382--391 (2018)
13. Zhu, M., Rudisill, T. M., Rauscher, K. J., Davidov, D. M., & Feng, J.: Risk Perceptions of Cellphone Use While Driving: Results from a Delphi Survey. International Journal of Environmental Research and Public Health. 15(6), 1074 (2018)
14. Tractinsky, N., Ram, E. S., & Shinar, D.: To Call or not to Call—That is the Question (While Driving). Accident Analysis & Prevention. 56, 59--70 (2013)
15. Trespalacios, O. O., Haque, M. M., King, M., & Washington, S.: "Mate! I'm Running 10 min Late": An Investigation into the Self-regulation of Mobile Phone Tasks While Driving. Accident Analysis & Prevention. 122, 134--142 (2019)
16. Choudhary, P., & Velaga, N. R.: Effects of Phone Use on Driving Performance: a Comparative Analysis of Young and Professional Drivers. Safety Science. 111, 179--187 (2019)
17. Fitch, G. Ms., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L. & Dingus, T. A.: The Impact of Hand-held and Hands-free Cell Phone Use on Driving Performance and Safety-critical Event Risk (No. DOT HS 811 757) (2013)
18. Jacomy, M., Venturini, T., Heymann, S., & Bastian, M.: ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PloS One. 9(6), e98679 (2014)