2021 IEEE Smart Information Systems and Technologies (SIST)
28-30 April, 2021, Nur-Sultan, Kazakhstan

2021 IEEE International Conference on Smart Information Systems and Technologies (SIST) | 978-1-7281-7470-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/SIST50301.2021.9465943

Statistical and Network Analysis of Shared Bikes – In the Case of Almaty Bike

Almasbek Maulit[1*], Yerzhan Baiburin[2], Madiyar Rakhymbek[3], Gulnara Sadykova[4], Aliya Nugumanova[5]
[1*]S. Amanzholov East Kazakhstan University, Kazakhstan, maulit.almas@yandex.ru,
[2]S. Amanzholov East Kazakhstan University, Kazakhstan, ebaiburin@gmail.com
[3]S. Amanzholov East Kazakhstan University, Kazakhstan, madiyarrakhymbek@gmail.com
[4]University of International Business, Kazakhstan, gulnara.muratpek@mail.ru
[5]S. Amanzholov East Kazakhstan University, Kazakhstan, yalisha@yandex.kz

***Abstract* — In this paper we analyze bike riding patterns in Almaty by employing the annual data from Almaty Bike system. The main contribution of the paper is explaining the factors effecting bike sharing usage in Almaty. To our knowledge this is the first (albeit preliminary) study to investigate behaviors of bike-sharing users in Kazakhstan using statistical and network analysis methods. The analysis shows that mostly bikes are used no more than 20 minutes, for the most part during the daytime, although the peak use is between 7 and 9 pm, and there is no consistent difference between bike use on weekends and weekdays.**

Keywords* — *bike sharing systems; smart city; network analysis; visualization, temporal networks.

I. INTRODUCTION

Bike sharing systems (BSSs) are now becoming one of the most popular modes of transportation in cities and megacities all over the world. This is due at first to ever increasing air and environmental pollution issues, which are driving governments and international organizations to impose stringent regulations limiting carbon emissions from vehicles. At second, rapid depletion of fossil fuels threatens the sustainability of future generations. For these reasons, green and economical cycling makes bikes attractive for daily short city trips.

As of May 2018, there are over 1,600 bike sharing programs worldwide, providing over 18 million bikes for public use [1]. The largest BSSs are located in China, USA, France, Germany, Canada, England etc. In Asia, the number of bike sharing programs is increasing exponentially, 13 of the world's 15 largest systems are located in China [2]. The majority of these systems provide unprecedented amount of data on city trips, including date, time, location and other relevant information which can be used not only for identifying factors effecting BSS demand, but also for urban movement analysis.

In this paper, we investigate Almaty Bike, a bike-sharing system operating in the city of Almaty. Our goal is to identify patterns and characteristics of Almaty Bike ridership and to establish which factors responsible for these patterns. To our best knowledge this is the first (albeit preliminary) study to investigate behaviors of bike-sharing users in Almaty and in Kazakhstan in general, using statistical and network analysis methods.

II. RELATED WORK

There are a huge number of articles devoted to the analysis of BSSs. A detailed review of researches concerning various aspects of BSSs is presented in [3]. Authors of this work collect a total of 208 relevant articles published between 2010 and 2018, to conduct scientometric analysis. They note that the focus of BSS research is shifting toward to "more complex external impacts, system optimization, design and integration with public transit." [3].

For example, in [4], the authors used complex network methods to analyze the structure of BSS. They calculate degrees, radiation distance, and community structure of the networks to understand the internal relations of the public BSS. In [5], the authors study the problem of predicting individual trips for BSSs, and primarily analyze individuals' bike usage behaviors. Based on the analysis, they propose a new model for predicting the trip destination and duration.

In works [6]-[8], the large-scale Parisian BSS called Velib is analyzed. The system is investigated in several aspects such as usage patterns, communication between public transport and BSS, and imbalances in flows between stations. Another large BSS Mobike located in China, is analyzed in works [9]-[11]. These works focus on the spatial and temporal characteristics of the Mobike system, as well as investigate the most loaded stations and pay a lot of attention to the problem of rebalancing the system. Another major system CitiBike New York is analyzed in [12] - [15]. The authors of these works also carefully investigate the problems of rebalancing the system and at the same time consider what factors affect the operation of the system: weather conditions, social division, station locations, traffic flows, road quality, etc.

III. METHODOLOGY

Our methodology includes 3 stages. At the first stage, we create some histograms to visualize main distribution

characteristics of Almaty Bike system. At the second stage, we identify the so-called imbalanced stations, which are characterized by difference in the amount of incoming and outgoing traffic [16]. We can plot the difference between the incoming and outgoing traffic for each station on a map as a circle to see which stations are the most imbalanced. The size of the circle represents the absolute difference and the color represents in which direction this station is imbalanced [16]. Then we analyze the map in order to determine factors effecting imbalance. At the third stage we apply network analysis to identify stations-authorities and stations-hubs.

IV. EXPERIMENTS AND RESULTS

A. *Data*

In this work, we use data from the Almaty Bike system, which has been operating in the city of Almaty since 2016 [17]. The system includes 190 stations, of which 187 are active. The data covers the period from January 16 to December 31, 2020 and consists of two datasets: trip-data and stations. The trip-data dataset consists of 866854 rows and 7 columns (i.e. bikeid, starttime, start.station.id, stoptime, end.station.id, covered.distance, trip.duration). The stations dataset contains information about stations' locations, i.e. longitude and latitude (see Figure 1) and capacities (see Figure 2).

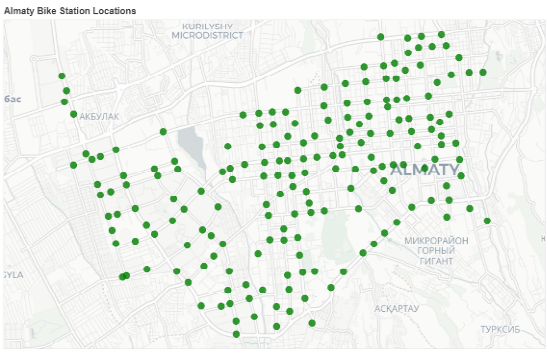

Figure 1 – Almaty Bike stations on the map

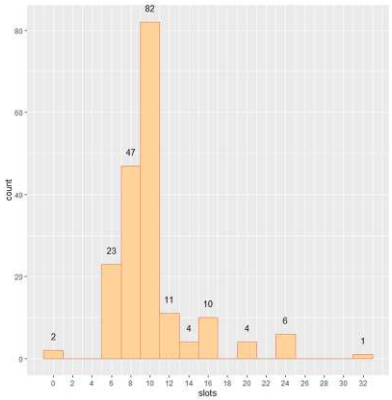

Figure 2 – Distribution of stations by the number of slots

B. *First stage: base distribution plots*

The distribution of Almaty Bike trips by month is subject to the normal law, the peak falls in the summer months and the trough in the winter (see Figure 3), with the exception of the months of April and May, when a tough lockdown was announced in the Kazakhstan due to the coronavirus pandemic.

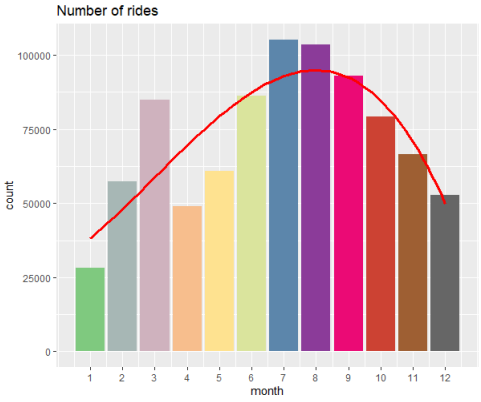

Figure 3 – Total number of rides by months

In contrast to the monthly distribution, the distribution of Almaty Bike trips by day of the week does not show a significant difference between weekdays and weekends (see Figure 4).

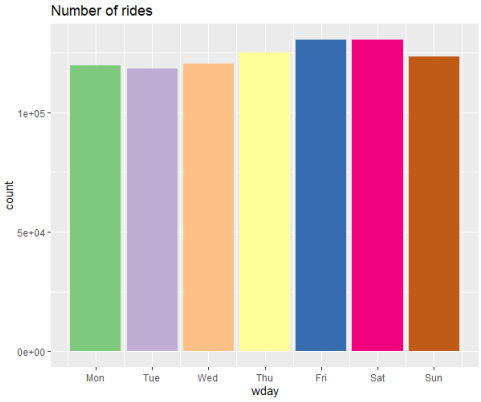

Figure 4 – Total number of rides by months

Then we analyze the distribution of Almaty Bike trips by duration and time of day. It shows that the most popular duration is above 20 minutes (see Figure 6) and the most popular time of day is from 12 noon to 6 pm, however, peak hours are from 7pm to 9pm (see Figure 7).

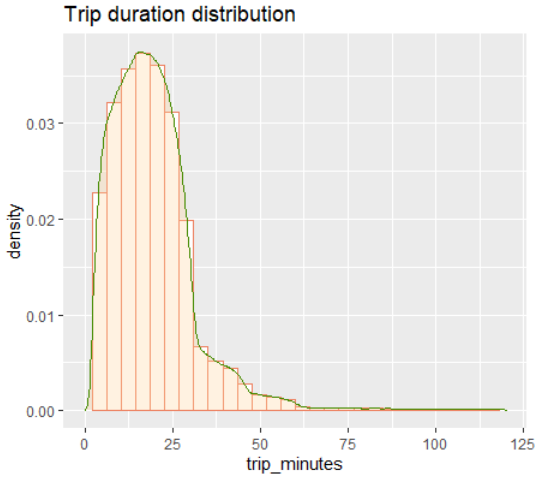

Figure 5 – Distribution of trip durations

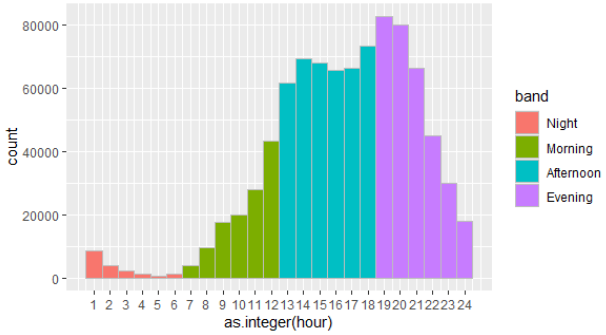

Figure 6 – Distribution of trips by time of day

C. *Second stage: Determining imbalanced stations*

By grouping the trips by the station id and counting the number of trips, we find how many trips have started from (outbound) or ended (inbound) at each station (see Figures 7-8). Then we can calculate the difference between outbound and inbound bikes for each station (see Figure 9). As Figure 9 shows, most stations have about as many trips ending there as starting from that station. However, there are a certain number of unbalanced stations where the difference between inbound and outbound trips is negative or positive.

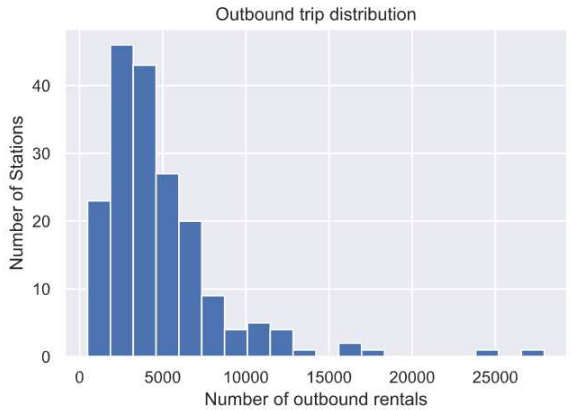

Figure 7 – Number of outbound bikes

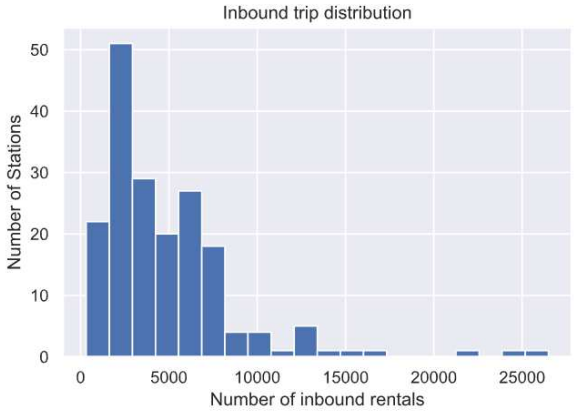

Figure 8 – Number of inbound bikes

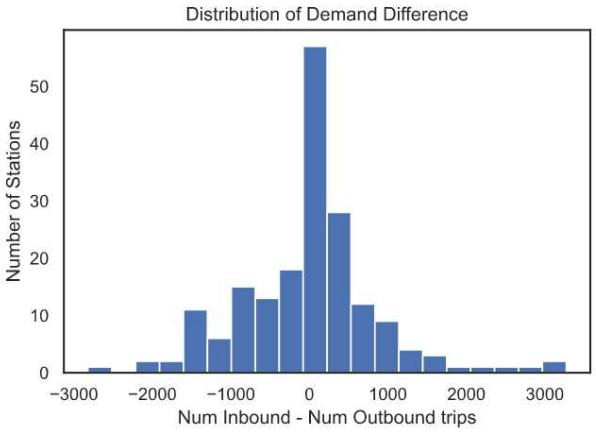

Figure 9 – Difference between inbound and outbound trips

By mapping these differences, we can immediately identify some problematic stations (see Figure 10). There are stations where many more people end their trips than they start (for example, Atakent station, 2 stations on Esetntai mall). There are also stations where many more people start their trips than they end (for example, stations on Tole-bi street - Gaidar street, Tole-bi street - Utegen batyr street). The map of the differences also reveals an interesting flow pattern: more trips start from stations that are nearby on Tole-bi street. The darker circles on the map show that more trips are outbound than inbound. Thus, we see a natural picture associated with the mountainous relief of Almaty – from the south, i.e. from the upper part of the city, more people descend than ascend from the north, i.e. from the bottom part. Obviously, this is due to the reluctance of people to ride bikes uphill.

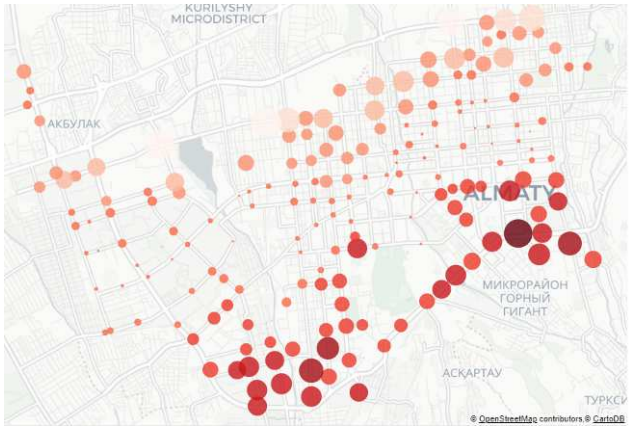

Figure 9 – Map of imbalance

D. Third stage: Determining hubs and authorities

At this stage, we use the HITS algorithm to calculate authority stations and hub stations. The algorithm is based on the concept of Kleinberg centrality of a node. An authority is a node that is referenced by a large number of other nodes. A hub is a node that gives many links to other nodes. As it shown from Figure 10, hubs and authorities are basically the same; however, neither the hubs nor the authorities are the same as the major stations.

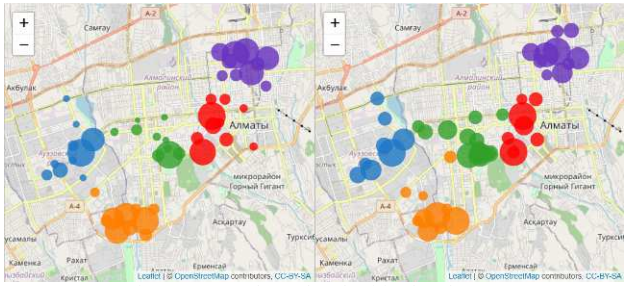

Figure 10 – Hubs and authorities by clusters

The largest hubs (they are also authorities) are located mainly along Timiryazev street and along Abai Avenue (see Figure 11). For example, station 2035 is located near the arch of KazNU named after al-Farabi. we can explain this by the large number of educational institutions located in the area.

Figure 11 – The largest hubs and authorities

When calculating hubs, trips that start and end at the same station are not taken into account. However, returning to the same station is the most popular ride, although trips to another, second most popular station are not too inferior in quantity. The question arises, are there groups of stations between which there is an intensive exchange? On the heat map in Figure 11, the intersections of these stations are colored red, which corresponds to the number of trips between these stations. The heat map shows that the stations form groups.

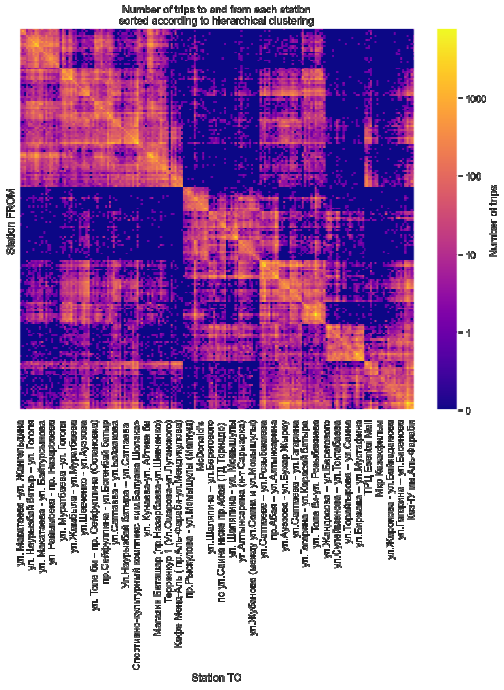

Figure 12 – Number of trips to and from each station

V. CONCLUSION

We employed an exploratory approach that focused on spatial, network, and statistical analysis for identifying the patterns of Almaty Bike System usage. Our focus was specifically on the factors effecting on these patterns. In this context, we borrowed methods from the BSSs researches and were able to identify some valuable factors.

ACKNOWLEDGEMENT

The data was kindly provided by the "Arman Aimagy" public foundation, which is responsible for the operational management of the Almaty Bike project.

REFERENCES

[1] https://www.statista.com/chart/14542/bike-sharing-programs-worldwide/.
[2] The Bike Share Map URL https://bikesharemap.com
[3] Si H. et al. Mapping the bike sharing research published from 2010 to 2018: A scientometric review ///Journal of cleaner production. – 2019. – Vol. 213. – P. 415-427.
[4] Yao Y, Zhang Y, Tian L, Zhou N, Li Z, Wang M. Analysis of Network Structure of Urban Bike-Sharing System: A Case Study Based on Real-Time Data of a Public Bicycle System.

Sustainability.2019;11(19):5425.https://doi.org/10.3390/su11195425.

[5] J. Zhang, X. Pan, M. Li and P. S. Yu, "Bicycle-Sharing System Analysis and Trip Prediction," 2016 17th IEEE International Conference on Mobile Data Management (MDM), Porto, 2016, pp. 174-179, doi: 10.1109/MDM.2016.35.

[6] Rahul Nair, Elise Miller-Hooks, Robert C. Hampshire & Ana Bušić (2013) Large-Scale Vehicle Sharing Systems: Analysis of Vélib', International Journal of Sustainable Transportation, 7:1, 85-106, DOI: 10.1080/15568318.2012.660115

[7] Come E. et al. Spatio-temporal analysis of dynamic origin-destination data using latent dirichlet allocation: Application to Vélib' bike sharing system of Paris //TRB 93rd Annual meeting. – TRANSPORTATION RESEARCH BOARD, 2014. – 19p.

[8] Y. Feng, R. Costa Affonso, M. Zolghadri, Analysis of bike sharing system by clustering: the Vélib' case, IFAC-PapersOnLine, Volume 50, Issue 1, 2017,pp 12422-12427,ISSN 2405-8963,https://doi.org/10.1016/j.ifacol.2017.08.2430.

[9] Cao, Min, and Mengxue Huang. "Analysis on the Riding Characteristics of Mobike Sharing Bicycle in Beijing City, China." Abstracts of the ICA, vol. 1, 2019, p. NA. Accessed 28 Jan. 2021.

[10] E. Gorgul, Chaoran Chen, A Visualization Based Analysis to Assist Rebalancing Issues Related to Last Mile Problem for Bike Sharing Programs in China: A Big-Data Case Study on Mobike, The International Conference on Computational Design and Robotic Fabrication, CDRF 2019: Proceedings of the 2019 DigitalFUTURES pp 145-153.

[11] Oppermann, M., Möller, T. and Sedlmair, M. "Bike sharing Atlas: visual analysis of bike sharing networks". International Journal of Transportation, 6(1), 2018. pp.1-14.

[12] O'Mahony, E., & Shmoys, D. (2015). Data Analysis and Optimization for (Citi)Bike Sharing. Proceedings of the AAAI Conference on Artificial Intelligence, 29(1). Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/9245.

[13] A. Faghih-Imani, N. Eluru,Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system, Journal of Transport Geography,Volume 54,2016, p.p. 218-227,ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2016.06.008.

[14] C. McBain, B. Caulfield, An analysis of the factors influencing journey time variation in the cork public bike system, Sustainable Cities and Society, Volume 42, 2018, p.p. 641-649, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2017.09.030.

[15] Masih A. Babagoli, Tanya K. Kaufman, Philip Noyes, Perry E. Sheffield, Exploring the health and spatial equity implications of the New York City Bike share system, Journal of Transport & Health, Volume 13, 2019, Pages 200-209, ISSN 2214-1405, https://doi.org/10.1016/j.jth.2019.04.003.

[16] https://brendanhasz.github.io/2018/08/02/nice-ride-eda.html

[17] https://almatybike.kz/ . Дата обращения: 1 февраля 2021

[18] Punithavelan N., Jaganathan B. New web page rank method using HITS Centrality //Global Journal of Pure and Applied Mathematics. – 2017. – Т. 13. – №. 10. – С. 7229-7235.