

УДК 519.8:004

## Компьютерное моделирование темпоральных сетей для анализа паттернов в функционировании динамических систем

*А. Маулит, А.С. Тлебалдинова, А.Б. Нугуманова, Е.М. Байбурун*

Восточно-Казахстанский государственный университет им. С. Аманжолова  
(Усть-Каменогорск, Казахстан)

## Computer Modelling of Temporal Networks for Bike Sharing Usage Patterns Analysis

*A. Maulit, A.S. Tlebaldinova, A.B. Nugumanova, Ye.M. Baiburin*

Sarsen Amanzholov East Kazakhstan State University (Ust-Kamenogorsk, Kazakhstan)

Представлены результаты анализа временной загрузки систем совместного использования велосипедов, выполненного с помощью темпоральных сетей. Темпоральные сети имеют множество применений в исследовании поведения сложных динамических систем, обладающих сетевой структурой. Например, их можно использовать для анализа и прогнозирования таких динамических показателей транспортных сетей, как интенсивность транспортных и пассажирских потоков, загруженность транспортных путей, пропускная способность транспортных узлов, оборот транспортных средств и т.д. В данной работе с помощью темпоральных сетей оцениваются показатели центральности станций и кластеров сети велопроката. Центральность станций оценивается на основе меры посредничества, а центральность кластеров — на основе меры централизации Фримена. На основании полученных оценок строятся визуальные модели «Тепловая карта» и «Временной ряд», в наглядной и компактной форме демонстрирующие пространственно-временные особенности велосети. Эксперименты, подтверждающие применимость построенных моделей, проводятся с использованием открытых данных системы CitiBike New York за апрель 2019 г. Они демонстрируют наличие суточных и месячных паттернов как среди отдельных станций, так и среди более крупных кластеров.

**Ключевые слова:** темпоральная сеть, CitiBike, байкшеринг, визуальная аналитика, центральность в графе, транспортная сеть.

DOI 10.14258/izvasu(2020)4-16

**Введение.** Системы совместного использования велосипедов (системы байкшеринга) — это системы, позволяющие арендовать велосипед на одной из авто-

This paper presents the results of analyzing the time load of stations in bike-sharing systems using temporal networks. Temporal networks have many applications in the study of the behavior of complex dynamic systems that have a network structure. In particular, they can be used to analyze and predict many dynamic indicators of transport networks, for example, such as the intensity of transport and passenger flows, traffic congestion, capacity of transport nodes, turnover of vehicles, etc. In this work, the indicators of the centrality of stations and clusters of a bike-sharing network are estimated using temporal networks. Based on the obtained estimates, visual models (Heat maps and Time Series) are constructed to demonstrate the spatial and temporal features of the bike network in a clear and compact form. The station centralities are estimated on the basis of the betweenness measure, and the cluster centralities are estimated on the basis of the Freeman centralization. Experiments confirming the applicability of the built models are conducted using open data from the CitiBike New York system for April 2019. They demonstrated the presence of daily and monthly patterns among both individual stations and more large station clusters.

**Key words:** temporal network, CitiBike, bike sharing, visual analytics, network centrality, transportation network.

матизированных станций, совершить поездку и вернуть велосипед на любую другую станцию, установленную в этом городе [1]. Такие системы становятся

все более популярными благодаря их экологичности и потенциалу в развитии «умных» городов, а математические аспекты их моделирования порождают множество интересных задач для исследователей.

В данной работе применен аппарат темпоральных сетей для анализа загруженности станций систем байкшеринга. Мы представляем систему байкшеринга как темпоральную сеть, рассматривая станции как узлы сети, а поездки между станциями — как связи.

## 1. Связанные работы

**1.1. Системы байкшеринга и основные проблемы их моделирования.** Эволюция систем совместного использования велосипедов насчитывает четыре поколения, последнее поколение — это цифровые системы, оборудованные датчиками, фиксирующими все действия пользователей в системе [2]. Проблемы проектирования и эксплуатации этих систем можно разбить на 3 класса [1].

Проблемы первого класса связаны с проектированием систем байкшеринга [3, 4]. Их решение направлено на оптимальное планирование и размещение станций велопроката и учитывает множество факторов, начиная с топографических особенностей города, прогнозирования спроса и заканчивая принципами социальной справедливости [2].

Проблемы второго класса связаны со стимулированием клиентов [5–7]. Стимулирование клиентов является необходимой частью сервиса велопроката в условиях перегруженности станций (например, когда на станциях отсутствуют велосипеды или свободные доки) [8]. Оно предполагает гибкую ценовую политику, зависящую от текущей ситуации (времени суток, погодных или сезонных явлений, календарных событий). Данные поступают от внешних сервисов, например, сервисов погоды, сервисов пробок и т.д.

Проблемы третьего класса связаны с ребалансировкой станций (перераспределением велосипедов между станциями) [9–12]. Они вызваны коммутационными паттернами, присущими мегаполисам, например, регулярными поездками горожан на работу, в результате которых наблюдается нехватка велосипедов в утренние часы в спальнях районах, а в вечерние часы, наоборот, — в деловых центрах [13–15]. Ребалансировка станций должна, с одной стороны, отвечать ожидаемым потребностям станций в велосипедах, а с другой — минимизировать стоимость управления системой, в том числе расходы на переброску велосипедов [8].

**1.2. Открытые данные системы совместного использования велосипедов CitiBike New York.** Не все системы байкшеринга предоставляют накопленные данные в открытом доступе, хотя такие данные очень ценны как самостоятельный ресурс для апробации моделей и методов. Мы используем открытые данные системы байкшеринга CitiBike NYC. Система совместного использования велосипедов в Нью-Йорке CitiBike NYC открылась в мае 2013 г., по состоянию на январь 2020 г. насчи-

тывает 13 000 велосипедов и 850 станций [16]. Данные, генерируемые системой, публикуются на облачном сервере компании Amazon [17], они представляют ежемесячные отчеты из 15 полей: продолжительность поездки; дата и время начала и окончания поездки; код, название и геокоординаты начальной и конечной станций поездки; код велосипеда; тип пользователя, год рождения и пол.

## 2. Предлагаемый подход

**2.1. Темпоральные оценки центральности станций.** Измеряем важность узлов (станций) велосипедной сети на основе оценки по посредничеству [18]. Чтобы ее вычислить, нужно знать количество кратчайших путей в сети, которые проходят через данный узел. Сначала идентифицируются все кратчайшие пути в сети, а затем для каждого узла подсчитывается, сколько раз он оказался на кратчайших путях:

$$C_i^B = \sum_{j \in V} \sum_{\substack{k \in V \\ k \neq j}} \frac{\sigma_{jk}(i)}{\sigma_{jk}}, \quad (1)$$

где  $\sigma_{jk}$  — количество кратчайших путей от узла  $j$  до узла  $k$ , а  $\sigma_{jk}(i)$  — количество таких кратчайших путей, которые проходят через узел  $i$ . Суммирование ведется по всем узлам. Таким образом определяются узлы, которые являются «мостами» между остальными узлами, т.е. ускоряют прохождение потоков внутри сети. В социальных сетях центральность по посредничеству применяют для определения людей, которые являются посредниками между слабо связанными сообществами.

Чтобы распространить понятие центральности на случай темпоральной сети, можно вычислять ее в каждый промежуток времени [19]. Тогда формула (1) останется неизменной, однако кратчайшие пути будут рассчитываться на основе только тех связей, которые существуют в сети в указанный промежуток. Таким образом, темпоральная центральность по посредничеству указывает, насколько интенсивно данная станция участвовала в обороте велосипедов между станциями в заданный промежуток времени.

**2.2. Темпоральные оценки центральности кластеров.** Существует большое количество работ, в которых анализ или прогнозирование трафика велосипедной сети предваряется кластеризацией [20, 21]. Необходимость кластеризации объясняется тем, что под влиянием большого количества сложных факторов трафик одной отдельно взятой станции выглядит слишком хаотичным, чтобы можно было делать на его основе какие-то выводы [22]. После группировки станций в кластер периодичность и регулярность трафика становятся более очевидными и прогнозируемыми, что позволяет сформировать более сглаженную картину трафика.

Чтобы от отдельных оценок центральности станций перейти к общей оценке центральности кластера, используется мера централизации Фримена [23]. Она

отражает степень, в которой подсеть (кластер) состоит из одного узла с высокой централизацией, окруженного периферийными узлами [23]. Мера представляет собой сумму разностей между центральностью самого центрального узла подсети и центральностями всех остальных узлов, деленную на максимально возможную сумму разностей, которая может существовать в подсети с таким набором узлов:

$$F = \frac{\sum_{i \in V} (c_* - c_i)}{\max \sum_{i \in V} (c_* - c_i)}, \quad (2)$$

где  $c_*$  — центральность самого центрального узла в подсети (кластере);  $c_i$  — центральность очередного узла  $i$  в подсети (кластере).

### 3. Экспериментальная работа

**3.1. Данные.** Экспериментальный набор данных был взят из открытых данных системы CitiBike NYC за апрель 2019 г. Набор состоит из 1 766 094 записей, описывающих поездки на велосипедах между 791 станцией. Алгоритм K-means был применен для группировки станций в кластеры по их географическому расположению (по широте и долготе) (рис. 1).

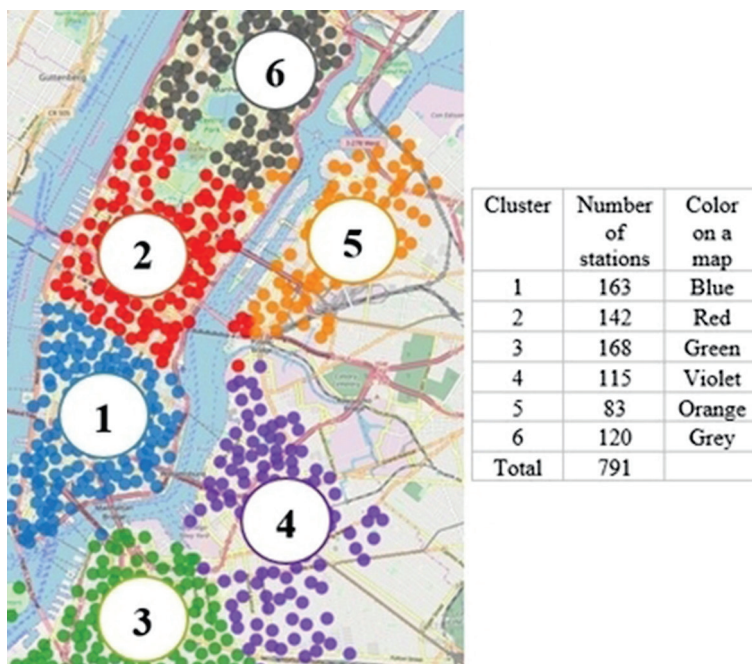


Рис. 1. Кластеризация станций по широте и долготе

После выполнения кластеризации каждый кластер рассматривается как темпоральная сеть со станциями в качестве узлов и поездками в качестве связей. Значения темпоральной центральности каждой станции внутри соответствующего кластера рассчитываются по формуле (1).

На рисунке 2 приведены графики центральности для двух станций кластера 2, которые имеют самые высокие суточные итоги. Измерения проводились в утренние часы в воскресенье и понедельник, и, как можно видеть, воскресная активность велосипедистов началась намного позже, чем в понедельник.

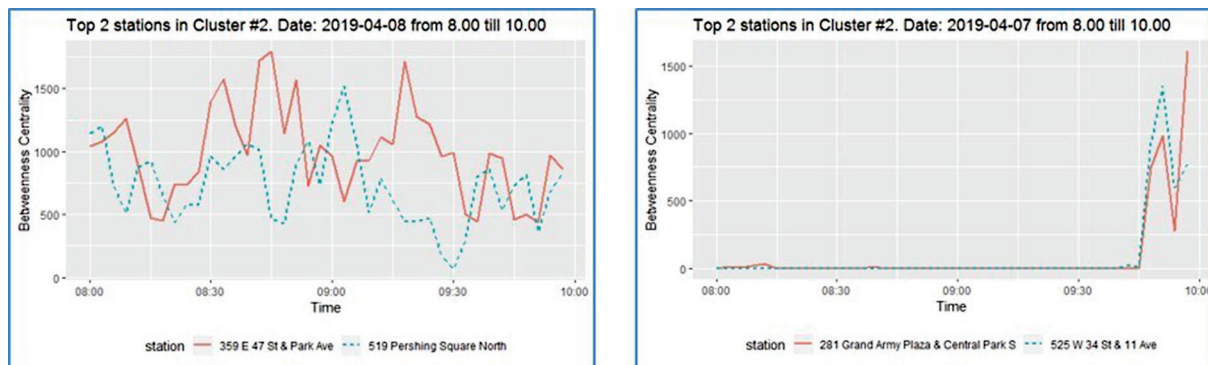


Рис. 2. Данные по двум самым активным станциям в воскресенье утром

**3.2. Тепловые карты.** После того как временные показатели центральности были рассчитаны на уровне отдельных станций, их можно агрегировать на уровне кластеров в соответствии с формулой (2). Затем, выбрав максимальные значения централизации для каждого кластера, можно визуализировать нагрузки на кластер. Тепловые карты на рисунке 3 представляют средние значения централизаций кластера по 100 пиковым значениям.

Как следует из рисунка, на тепловых картах кластеров присутствуют белые пятна в левом нижнем углу, это означает, что интенсивность совместного использования велосипедов в субботу и воскресенье утром является низкой, независимо от кластера. Тепловая карта кластера 1 содержит гораздо меньше белых пятен, чем тепловая карта кластера 2, таким образом, нагрузка на кластер 1 является более равномерной.

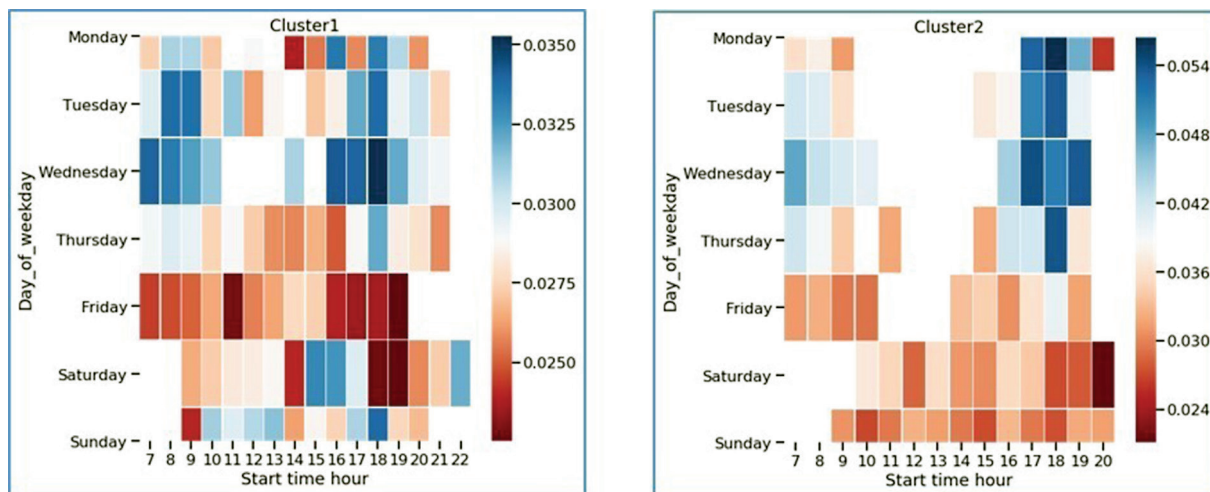


Рис. 3. Тепловая карта централизаций кластеров 1 и 2

В целом, анализируя все тепловые карты, можно сделать вывод, что самый интенсивный в плане оборота велосипедов — это кластер 1. Высокая централизация кластера означает наличие в нем такой центральной станции-посредника, которая «замыкает» на себе все обороты велосипедов, т.е. большинство станций обмениваются своими велосипедами именно через нее. Низкая централизация наблюдается в кластере 1 во все дни после 20:00.

**3.3. Тренды.** Рассчитанные временные значения центральностей могут быть представлены в виде временных рядов, сравнение которых может быть полезным с точки зрения выделения трен-

дов. С этой целью в каждом кластере мы выбрали топ-10 станций с самыми высокими средними значениями центральности и построили графики сезонности для каждой из топ-10 станций. Во всех кластерах, кроме первого, тренды станций имели одинаковый вид.

На рисунках 4–7 показаны тренды четырех станций из топ-10 в кластере 6 и кластере 3. Разница видна невооруженным глазом: в то время как тренды станций кластера 6 резко снижаются после 17 апреля, а затем имеют пик в районе 22 апреля, все тренды кластера 3 после такого же спада имеют слабый пик в районе 25 апреля.

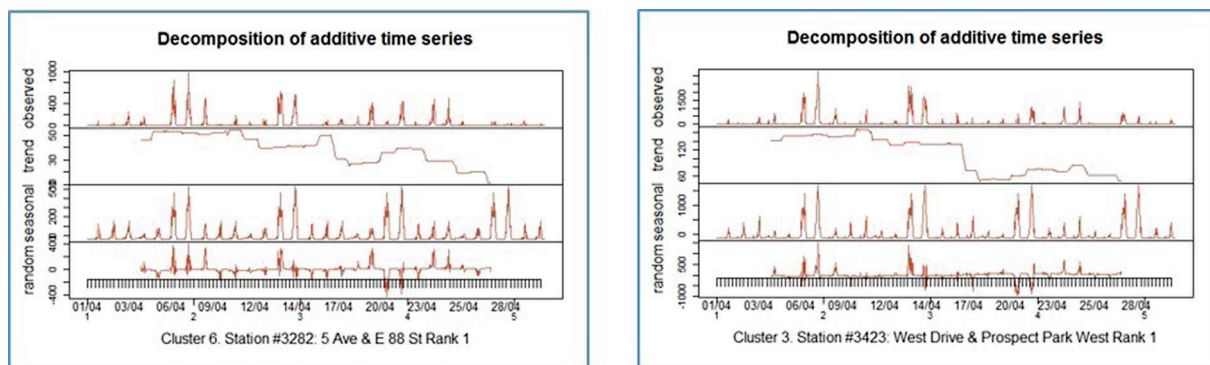


Рис. 4. Тренды станций ранга 1 из топ-10 в кластерах 3 и 5



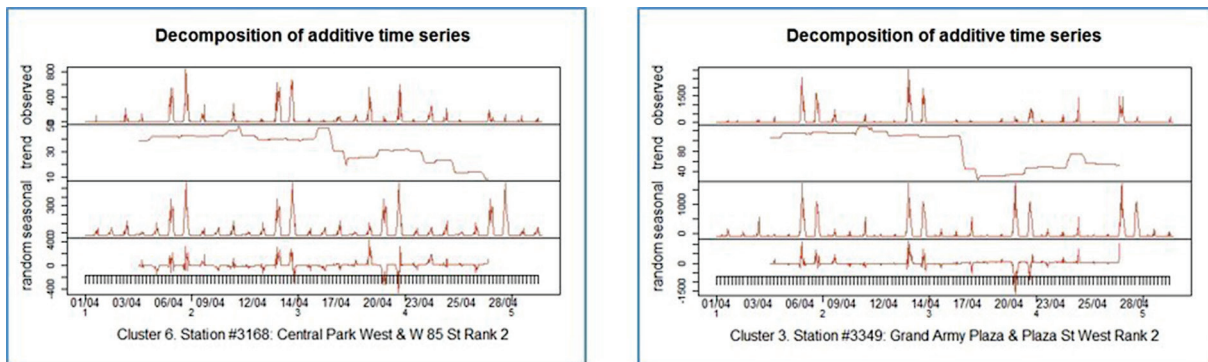


Рис. 5. Тренды станций ранга 2 из топ-10 в кластерах 3 и 5

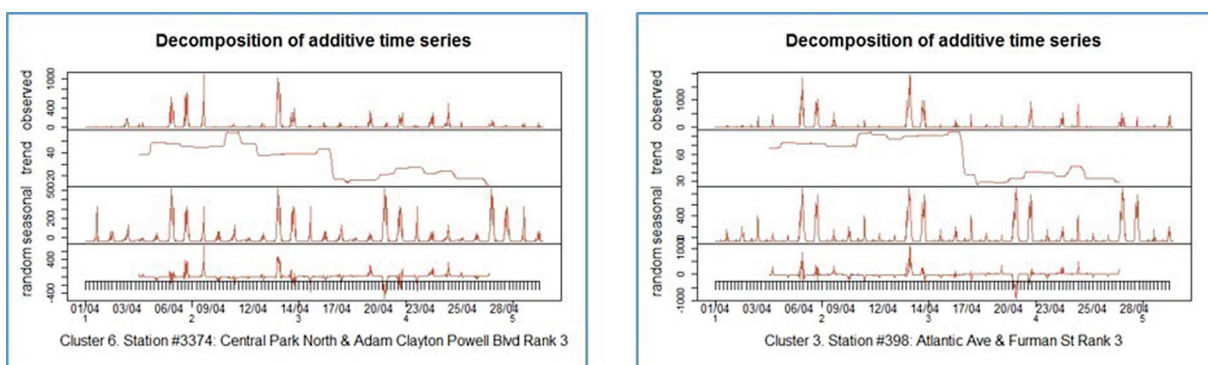


Рис. 6. Тренды станций ранга 3 из топ-10 в кластерах 3 и 5

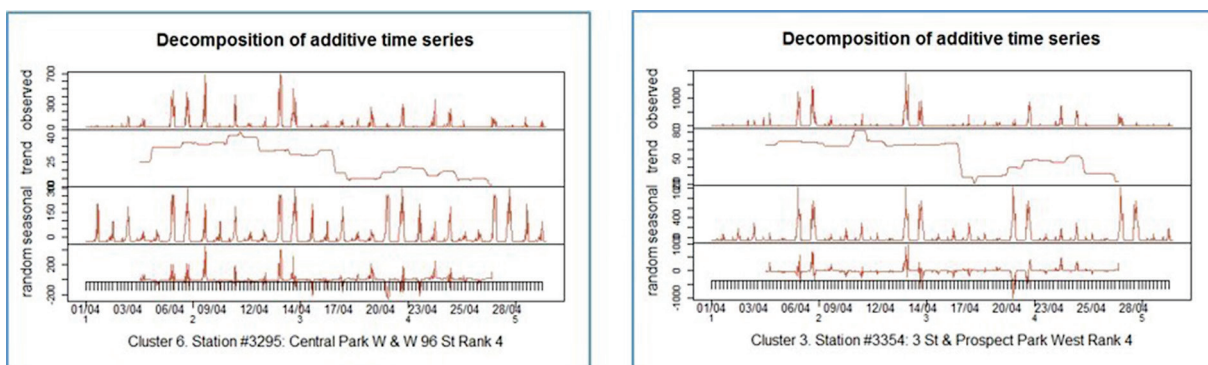


Рис. 7. Тренды станций ранга 4 из топ-10 в кластерах 3 и 5

**Заключение.** Несмотря на то что в этой работе использован небольшой набор данных, ограниченный только одним месяцем, и выполнена простейшая кластеризация, считаем, что цель нашей работы достигнута. Мы доказали применимость темпораль-

ных центральностей для определения закономерностей и трендов в работе системы байкшеринга. Наша будущая работа будет заключаться в расширении наборов данных, улучшении методов кластеризации и в сравнении показателей центральности.

## Библиографический список

1. Shaheen S. A., Guzman S., Zhang H. Bikesharing in Europe, the Americas, and Asia: past, present, and future // *Transportation Research Record*. 2010. T. 2143. №. 1. DOI: 10.3141/2143-20.
2. Lozano Á. et al. Multi-agent system for demand prediction and trip visualization in bike sharing systems // *Applied Sciences*. 2018. T. 8. DOI: 10.3390/app8010067.
3. Frade I., Ribeiro A. Bike-sharing stations: A maximal covering location approach // *Transportation Research Part A: Policy and Practice*. 2015. T. 82. DOI: 10.1016/j.tra.2015.09.014.
4. Çelebi D., Yörüsün A., Işık H. Bicycle sharing system design with capacity allocations // *Transportation research part B: methodological*. 2018. T. 114. DOI: 10.1016/j.trb.2018.05.018.
5. Singla A. et al. Incentivizing users for balancing bike sharing systems // *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015.
6. Pan L. et al. A deep reinforcement learning framework for rebalancing dockless bike sharing systems // *Proceedings of the AAAI conference on artificial intelligence*. 2019. T. 33. DOI: 10.1609/aaai.v33i01.33011393.
7. Angelopoulos A. et al. Incentivization schemes for vehicle allocation in one-way vehicle sharing systems // *2016 IEEE International Smart Cities Conference (ISC2)*. IEEE, 2016. DOI: 10.1109/ISC2.2016.7580857.
8. Raviv T., Tzur M., Forma I. A. Static repositioning in a bike-sharing system: models and solution approaches // *EURO Journal on Transportation and Logistics*. 2013. T. 2, №. 3. DOI: 10.1007/s13676-012-0017-6.
9. Alvarez-Valdes R. et al. Optimizing the level of service quality of a bike-sharing system. *Omega* (2015), pp. 1-13. h DOI: 10.1016/j.omega.2015.09.007.
10. Liu J. et al. Rebalancing bike sharing systems: A multi-source data smart optimization // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. DOI: 10.1145/2939672.2939776.
11. Xu F., Chen F., Liu Y. Bike Sharing Data Analytics for Smart Traffic Management // *2019 5th International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE, 2019. DOI: 10.1109/BIGCOM.2019.00020.
12. Zheng Z., Zhou Y., Sun L. A Multiple Factor Bike Usage Prediction Model in Bike-Sharing System // *International Conference on Green, Pervasive, and Cloud Computing*. Springer, Cham, 2018. DOI: 10.1007/978-3-030-15093-8\_28.
13. Oppermann M., Möller T., Sedlmair M. Bike sharing Atlas: visual analysis of bike-sharing networks // *International Journal of Transportation*. 2018. T. 6, №. 1. DOI:10.14257/ijt.2018.6.1.01.
14. Zhou X. Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago // *PloS one*. 2015. T. 10. №. 10. DOI: 10.1371/journal.pone.0137922.
15. Papazek P. et al. Balancing bicycle sharing systems: an analysis of path relinking and recombination within a GRASP hybrid // *International Conference on Parallel Problem Solving from Nature*. Springer, Cham, 2014. DOI: 10.1007/978-3-319-10762-2\_78.
16. Kaufman S. M. et al. Citi Bike: the first two years // *The Rudin Center for Transportation Policy and Management*. 2015.
17. CitBike System Data. Available online: <https://www.citibikenyc.com/system-data> (accessed on 17 January 2020).
18. Nicosia V. et al. Graph metrics for temporal networks // *Temporal networks*. Springer, Berlin, Heidelberg, 2013. DOI:10.1007/978-3-642-36461-7\_2.
19. Li Y. et al. Traffic prediction in a bike-sharing system // *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2015. DOI: 10.1145/2820783.2820837.
20. Dai P. et al. Cluster-Based Destination Prediction in Bike Sharing System // *Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference*. 2018. DOI: 10.1145/3299819.3299826.
21. Caggiani L. et al. Spatio-temporal clustering and forecasting method for free-floating bike sharing systems // *International Conference on Systems Science*. Springer, Cham, 2016. DOI: 10.1007/978-3-319-48944-5\_23.
22. Feng S. et al. A hierarchical demand prediction method with station clustering for bike sharing system // *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*. IEEE, 2018. DOI: 10.1109/DSC.2018.00133.
23. Borgatti S.P., Everett M.G. Extending centrality // *Models and Methods in Social Network Analysis*. 2005. T. 28. DOI:10.1017/CBO9780511811395.