Proceedings of the ASME 2020
International Design Engineering Technical Conferences
and Computers and Information in Engineering Conference
IDETC/CIE2020
August 17-19, 2020, Virtual, Online

DETC2020-22631

TOWARDS ENGINEERING COMPLEX SOCIO-TECHNICAL SYSTEMS USING NETWORK MOTIFS: A CASE STUDY ON BIKE-SHARING SYSTEMS

Yinshuang Xiao, Zhenghui Sha[1]
Department of Mechanical Engineering
University of Arkansas
Fayetteville, AR

ABSTRACT

The socio-technical system (STS) is an important topic in Systems Engineering and Design Science. Its performance is not only affected by social aspects but also influenced by various technical factors. To understand the relationships and interactions among different components and subsystems in STS, many studies have been done either at individual agent level or at the system level, yet few studies were conducted at the local structural level in such systems. Motivated by this research gap, we developed an approach to analyzing STS based on the network motif theory. In this study, we apply this approach to three bike-sharing systems (BSS) to validate its feasibility. We focus on studying the size-3 motif, the most basic building block of complex networks, and its correlations to a BSS's rebalancing performance in three different cities, i.e., NYC, Chicago, and Los Angeles. This paper reaches three conclusions. First, both seasonal and city effects play a significant role in affecting BSS's network structure and its local motif characteristics. Second, the rebalancing issue, i.e., the imbalance between bike returns and rentals, happened at the local transit level can be different from that observed at the system level, and vice versa. Third, the average geographical distance of size-3 trip motifs follows strong patters correlated to the motif structures as well as the number of directed links in a motif. Compared with previous studies, these insights would be beneficial to guiding system designers in engineering STS, particularly from a bottom-up manner (e.g., using mechanisms or incentives), to achieve desired system-level performance. This study also provides an in-depth understanding of the relations between local system structures and system-level performance in an STS, therefore contributes to both the design theory of complex systems and the BSS research community from a new network motif-based perspective.

Keywords: Socio-technical System, Bike-sharing System, Network Motif, Complex Network, System Rebalance.

1. INTRODUCTION

1.1 STS and Network-based Research on STS

The socio-technical system (STS) is one type of complex system that emerged from the interaction between social behavior and technological development. It is important because people, as one unpredictable factor in many complex systems, have a critical impact on the performance of a system. Thus, considering the social aspects can make system modeling more robust, and better guide engineering design.

The term STS is originally proposed by Trist et al. in their paper discussing the interrelation between humans, machines, and the context in English coal mines[1]. In 1986, Cherns established a set of general socio-technical design principles to guide abstract STS development [2]. In 2000, Clegg extended those principles to three interrelated types: meta, content, and process, by considering new information technologies and management practices [3]. The authors also discussed the role of these principles in system design. To gain a better understanding of STS, more in-depth analyses are conducted. For example, human factors were analyzed by Carayon to evaluate their influence on complex systems in the areas of health care as well as computer security [4]. Gordon Baxter analyzed the obstacles of applying the socio-technical design method to complex system development and presents a newly STS engineering framework based on information systems, the group investigating work design, computer-based collaborative work, and cognitive engineering. [5].

In recent years, the complex network has been wildly used in STS studies because of its capability in capturing the inter-connections between social entities and technological parts [6,7]. For example, Sha and Panchal [8] developed three different models based on complex networks to estimate the node-level behaviors and their impacts on the Internet, one of the most important STS in human society. The authors also analyzed

[1] Contact author: zsha@uark.edu

airlines' decisions on city-pair route addition and deletion behaviors in the US domestic air transpiration systems by integrating discrete choice models and complex networks. Based on random utility theory, Sha and Panchal [9] developed a degree-based decision-centric (DBDC) network model that can produce a variety of network topologies representing different STS structures. A series of studies [10–15] based on stochastic network models (e.g., the Exponential Random Graph Model) were conducted in support of customer preferences modeling, choice behavior modeling, and customer-product interactions in complex vehicle-product market systems.

1.2 BSS and Literature Review of BSS

In this paper, we study a particular STS, called bike-sharing systems (BSS) – one emerging transportation system arising with the sharing economy. The technical aspect of BSS refers to the bikes and the station (or dock) system, while the social aspect relates to customers' transit patterns as well as the social networks. The perspective of viewing BSS as a socio-technical complex network is unique and important, for it can provide network-based approaches to better understanding BSS. The history of bike-sharing project can be traced back to 1965 when the first generation of BSS emerged in Amsterdam. Although BSS experienced several updates, it did not gain much attention from academia until the explosion of data science and the rising sharing economy. According to the Scientometric Review [16], earlier articles mainly focus on some basic concepts, such as the policy, system safety, and system benefits, etc. ([17–20]). Since then BSS becomes a more interdisciplinary research subject, which covers environmental factors, population distribution, spatial characteristics, and social elements etc.

Regarding to recent studies, Yi et al. [21] divided them into three categories: system rebalancing, system prediction, and system design and traffic pattern analysis, among which many mathematical models, algorithms, and optimization methods were developed. For example, a novel deep learning model was established by Lin et al. to predict station-level hourly demand in BSS [22]. The main contribution of this study is that their models take the potential correlations between stations into consideration. Besides demand prediction, many studies also aim to provide rebalancing strategies using both customer-oriented and truck-based methods [21][23,24]. For example, Yi et al. [21] developed a customer-oriented rebalancing strategy based on one-dimensional Random Walk with Two Absorption Walls (RWTAW) and The Largest The First (TLTF) algorithms. Thirdly, for system design and traffic patterns analysis, an integrated approach, which jointly considers station location and capacity, was proposed by Çelebi et al. [25] to guide the BSS design.

Despite of extensive studies on BSS, most of them apply statistical models to analyze the performance of individual stations or local regional stations. No network-based approaches have yet been developed to analyze the performance of BSS from a systems engineering and design point of view. Lin et al. [22] analyzed the BSS network, but their research purpose was to verify the performance of the Data-driven Graph Filter (DDGF) model instead of studying the local-global structural interactions and performance correlations in BSS. The research gap also exists in the lack of knowledge in understanding the relationship between the local trip patterns and the system performance of BSS.

The **objective of this paper** is to fill the research gap by developing an approach based on the theory of network motif to analyze the local-global relations in STS. It is expected the new knowledge obtained can help designers gain a better understanding of the local behaviors in an STS and more insights into how it can be engineered. We apply our approach on three BSSs, i.e. Metro Bike (Los Angeles and California metropolitan areas), Divvy Bike (Chicago), Citi Bike (New York City), each of which represents a particular developmental stage of BSS from underdeveloped, developing, and to developed, respectively. The **main contributions of this study** are:

1) A network motif-based method for measuring BSS rebalancing performance is presented using nodes' (bike stations') number of in- and out-degrees. And a new understanding for the local rebalance performance based on the correlation between the size-3 motifs and their average geographical distance is achieved.
2) An understanding of the relationships between the local network motifs and the system-level performance in BSS, and how BSS trip behaviors quantified by network motifs vary with seasons.

To the best of our knowledge, this is the first study using a network motif-based approach to analyzing STS. The new motif-based interpretations can facilitate the development of new strategies and incentives to engineering BSS structures towards a desired system performance.

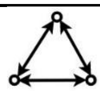
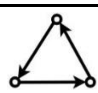
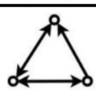
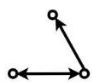
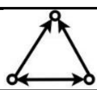
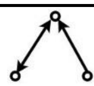
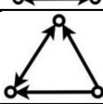
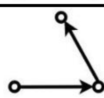
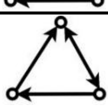
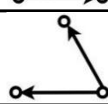
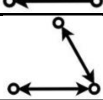
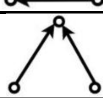
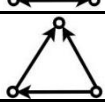
The remainder of this paper is organized as follows: the technical background of network motifs is introduced in Section 2. The proposed STS analysis approach is presented in Section 3, and the BSS case study is presented in Section 4. In this section, a summary of the three BSSs is provided. Then in Section 4.1, the trip networks are constructed, modeled, and their motifs are extracted by the motif mining tools. Based on the established BSS trip network, the global-level analysis is conducted in Section 4.2, and Section 4.3 focuses on the motifs and local-level analysis. We discuss the results obtained in Section 4.5; and finally, the paper is concluded in Section 5 with closing thoughts and our vision of future research directions.

2. TECHNICAL BACKGROUND

2.1 Network Motifs

Network motifs are interconnecting patterns emerging in specific networks with higher numbers than those in the random network [26]. Before introduced by Milo in 2002 [26], network motif had been researched for a long period, which was considered as certain sub-graphs that appear in a network statically different from that in the corresponding random network [27,28]. Since then, the research focus is more on two topics: motif structure interpretation ([29–31]) and motif mining ([32–34]). Based on the number of nodes that a motif has, it can be categorized into different types. The most common types are size-2 motifs (dyads), size-3 motifs (triads), and size-4 motifs (tetrads) [30]. Dyads are the simplest motifs but have a significant impact on network formation and higher-level motifs. Triads are “transitivity” motifs that can powerfully influence the growth of the social network, and they are also often applied to analyze the clustering characteristics of a complex network. Tetrads are motifs that have caught attention in recent years, and certain patterns have been identified in various real networks such as biological networks, electronic networks, and social networks [30]. Network motifs can also be classified as undirected motifs and directed motifs. Given that the triad is the basic “transitivity” motif and has been wildly studied in social network [30] and is regarded as the most basic building block for any other complex graphs, applying size-3 motif to analyze the STS will facilitate the understanding of its social aspects. Accordingly, in this study we mainly focus on size-3 directed motifs. The thirteen motif structures and corresponding IDs are listed in Table 1 [35].

TABLE 1: SIZE-3 DIRECTED MOTIF LIST

Structure	Adjacent Matrix	ID	Structure	Adjacent Matrix	ID
	0 1 1 1 0 1 1 1 0	238		0 1 0 0 0 1 1 0 0	140
	0 1 1 1 0 1 1 1 0	174		0 0 0 0 0 1 1 1 0	14
	0 0 0 1 0 1 1 1 0	46		0 1 0 1 0 0 1 0 0	164
	0 1 0 1 0 0 1 1 0	166		0 0 0 0 0 1 1 0 0	12
	0 0 1 1 0 0 1 1 0	102		0 0 0 0 0 0 1 1 0	6
	0 0 1 0 0 1 1 1 0	78		0 0 0 1 0 0 1 0 0	36
	0 0 0 1 0 0 1 1 0	38			

A network motif has three statistical properties:

1) Motif Concentration

Given a graph G and an n-size motif G′, the frequency of G′ in G is defined as the times that G′ occurred in G which is represented by $F_G(G')$. Then the motif concentration is defined as follows:

$$C_G(G') = \frac{F_G(G')}{\sum_i F_G(G_i)} \tag{1}$$

, where i represents the total numbers of non-isomorphic n-size motifs [32].

2) Motif Z-score

Considering the mean frequency of G′ in a random network R be $u_R(G')$ and the corresponding standard variance be $\sigma_R(G')$, then the Z-score is defined as:

$$Z_G(G') = \frac{F_G(G') - u_R(G')}{\sigma_R(G')} \tag{2}$$

Z-score indicates the significance of sub-graph G′ in G. The higher the Z-score, the more important the G′ as a motif in G. Additionally, it is worth noting that motifs in a larger real networks (with more nodes and links) may have higher z-score than that in a smaller network [26].

3) P-value

P-value indicates the probability of $F_R(G') > F_G(G')$, where $F_R(G')$ is the frequency of G′ in random network R. P-value can be calculated by

$$P_G(G') = \frac{1}{N}\sum_{j=1}^{N} \delta\left(F_R(G') > F_G(G')\right) \tag{3}$$

, where N represents the total number of considered random networks, and j is the index of each random network. δ equals to 1 when $F_R(G') > F_G(G')$; and 0 otherwise. In general, one motif pattern is significant if the P-value is smaller than a typical threshold, commonly 0.01 or 0.05.

3. THE RESEARCH APPROACH

In this section, we introduce the proposed approach to analyzing STS based on network motif theory, as shown in Fig. 1. The first step of this approach is to define and construct the complex network that best captures the STS structures and dynamics based on the research interests and the problems to be investigated. In this step, we need to verify if a system can be interpreted as an STS, and then define the node, link, whether links are directed or undirected, and whether links carry weigh or not. With the established network, in step two, motif-mining tools such as FANMOD and MFINDER can be applied to search for significant motif patterns. Then according to the motif-mining results, local-level network analysis will be conducted. This step is important because the recurring sub-graphs are the potential building block of a complex network, thus may have significant impact on the system-level performance. On the other hand, the global-level network analyses, including both network structure analysis and performance analysis, are conducted. Finally, critical correlations or variations between the local structures and behaviors and the global network performance are studied. The insights obtained from the last step can inspire new design ideas and guidelines for engineering STS.

We apply this approach in BSS and demonstrate how this approach can help reveal the local-global relations in BSS for engineering and design purpose. The BSS is chosen for the case study because of the richness of data, thus provides a good testbed for validation and verification.

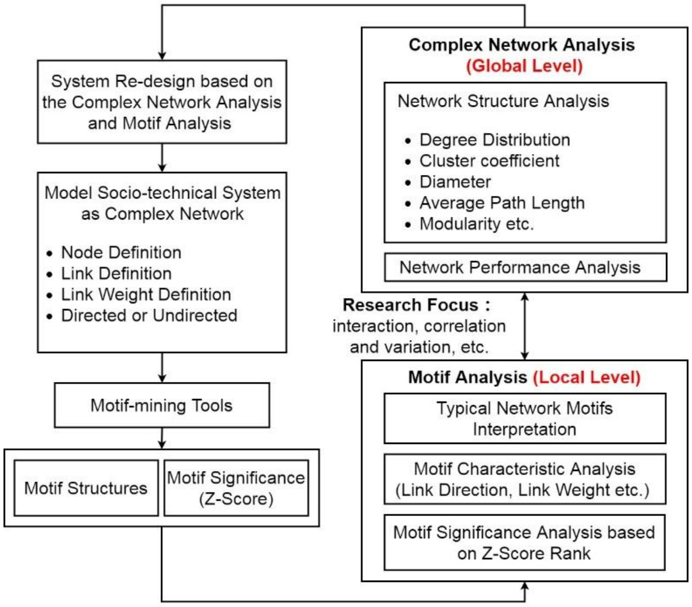


FIGURE 1: THE PROPOSED APPROACH TO ANALYZING SOCIOTECHNICAL SYSTEMS

4. CASE STUDY

Viewing BSS as an STS, the technical part refers to the station system while the social part relates to the transit patterns naturally formed by customers. The BSSs selected are Citi Bike in New York City, Divvy Bike in Chicago, and Metro Bike in the Los Angeles and California metropolitan area, respectively. All historical data are publicly available online [36–38], and we use the data of the year 2018 in this paper. Figure 2 illustrates the station distributions of three BSSs, and Table 2 shows the number of stations in each region and the date when the system was initiated. According to Fig. 2, the spatial distributions of these BSS stations are quite different. Citi Bike stations are evenly distributed in NYC, Divvy Bike stations show a radial structure in Chicago, while Metro Bike stations are separately located in four areas of the Greater LA area.

These three BSSs are geographically distributed from West Coast to Mid-West and to East Coast, and the system sizes rank from small to large, i.e., 131 stations, 618 stations, and 846 stations, respectively. So, the selection of these three systems is beneficial to learning the spatial effects and the system performance in different developing stages of BSS.

TABLE 2: BASIC INFORMATION OF CASE BIKE-SHARING SYSTEMS

System Name	Metro Bike	Divvy Bike	Citi Bike
Locale	Los Angeles and California metropolitan area	Chicago	New York City
Date of Operation Began	July 7, 2016	June 28, 2013	May 27, 2013
Number of Stations in 2018	131	618	846

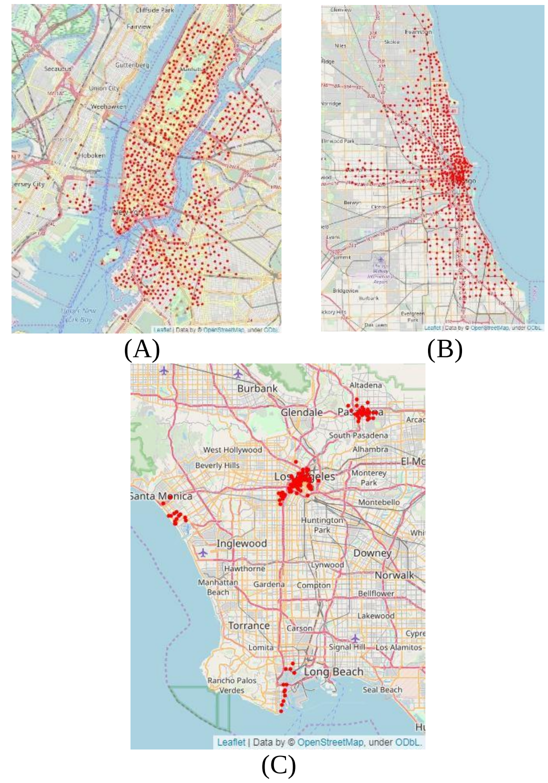


FIGURE 2: BSS STATION DISTRIBUTIONS OF THREE CITIES. (A) CITI BIKE IN NEW YORK CITY; (B) DIVVY BIKE IN CHICAGO; (C) METRO BIKE IN LOS ANGELES AREAS

The trip data packages include information like the trip duration, trip date, start, and end station ID, station geographic coordinates, and customer information. We followed four steps to process the raw data. The first step is to extract the essential information of a trip or transit which includes trip start and end times, start station ID and geographic latitude and longitude, and end station ID and geographic location. Based on the extracted data, in the second step, the following treatments are conducted: 1) deleted trips with incomplete information; 2) deleted stations marked as test stations and the associated trip data; 3) categorize trips by month based on their starting time; 4) count the number of times that one trip from one station to another station. After step two, the data frame only has seven columns (including start station ID, start station latitude, start station longitude, end station ID, end station latitude, end station longitude, and the number of trips) that will be used for establishing the directed and weighted trip network construction (see Section 4.1). Finally, we apply the same process to all the three BSSs for all 12 months in 2018, thus we have 36 networks in total. In each network, the stations are defined as nodes, and links represent the transit between two stations. The direction of link points from the start station to the end station. The weight on a link represents the total number of trips that happened between two stations in a month.

4.1 Trip Network Refinement and Motif Mining

Given that there are a large number of one-time trips occurring in each month, we decided to binarize the links in this study and only focus on the links having more frequent trips occurred. The threshold for the network binarization is set as the minimum mean value (μ) of the weights in the 36 networks:

$$\mu = Min\left(\mu_{1,1}, \dots, \mu_{i,j}, \dots, \mu_{3,12}\right), (i = 1,2,3; j = 1,2, \dots, 12) \quad (4)$$

, where j represents the index of twelve months, i represents the index of three BSSs, $\mu_{i,j}$ is the mean value of link weight in month j of BSS i. With this threshold, if the number of trips between two stations is smaller than μ, it is considered as an underused route, and we will assume there is no link between the two station nodes. The reason for using the minimum mean as the threshold is to compare the three BSSs on the same scale and only keep the trip links with high frequency in the BSS networks. With Equation (4), we get $\mu = 4.21$. Therefore, only the links with weights greater than 4.21 are kept in the networks. Figure 3 illustrates the link weight distributions of three BSS networks in July 2018. It is interesting that even if the three BSSs have stations distributed quite differently (see Fig. 2), their link weight distributions look very similar. All show that most pairs of stations are connected with few trips and very few stations are connected with a large number of trips. For example, one pair of stations in Chicago has over 1000 trips taken place in July 2018 while over 10,000 pairs of stations only have one trip in-between in that month.

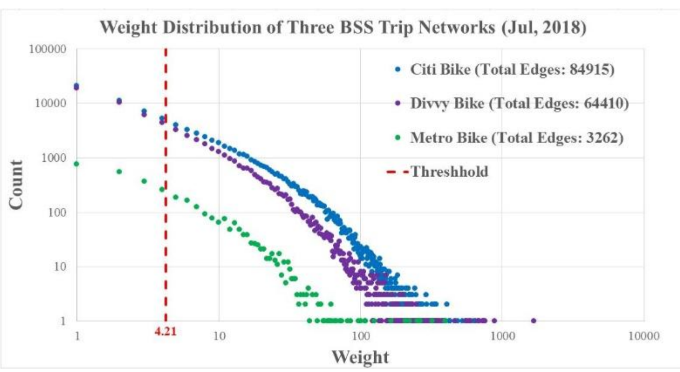


FIGURE 3: WEIGHT DISTRIBUTION EXAMPLES OF THREE TRIP NETWORKS (JUL, 2018)

After getting the binary trip networks, the motif mining tool, FANMOD [35], is used to detect significant motifs. The results from FANMOD report the significant motif structures, z-scores, p-values, frequencies, and the adjacent matrix list of all existing motifs. The detailed analysis of the identified motifs is presented in Section 4.3. In the following section, we first analyze the networks at a global level.

4.2 Global-level Trip Network Analysis

This section firstly presents the analysis of the trip network including its visualizations and key network measurements. Then a network-based method for evaluating rebalancing performance is introduced, and the rebalancing problems in three BSSs are analyzed. The network visualization and analysis are performed using GEPHI [39].

a. Network Structure Analysis

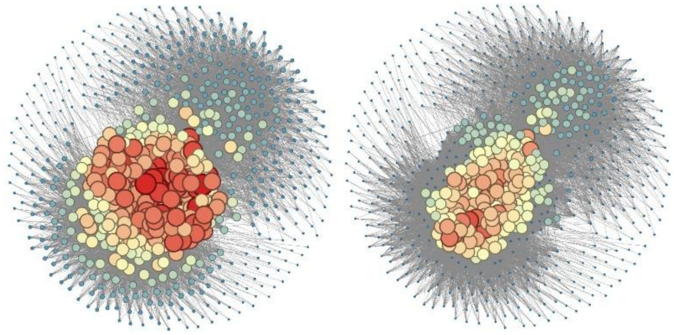
Table 3 presents the basic network structural metrics of Citi Bike in July 2018. The average clustering coefficient is 0.592 meaning that the possibility of the nodes in Citi Bike network grouping to form communities is high. The clustering coefficient is consistent with the observations in many social networks [7]. For example, if customers frequently travel between station A and station B, and between Station A and station C, the likelihood that a trip exists between station B and station C is high. This may be due to the nature of the NYC road transportation system (e.g., the grid structure) which makes every pair of stations have no distinct difference in terms of travel conditions (e.g., similar trip distance, similar road condition, etc.). This helps the formation of triangle structures in the network which is proportion to the clustering coefficients. The average path length is 1.947 implying that any pairs of stations in Citi Bike can be connected by two trips on average, and the diameter of 4 indicates at most 4 trips are needed to connect any pairs of stations in Citi Bike. Comparing to a complete network in which the graph density is 1, the graph density of 0.082 indicates the trip network is quite sparse and the system itself is still not fully exploited. Lastly, the network modularity is 0.288. It means that the possibility of a network being divided into different modules is less than 30%. This is true from the observation on the station distribution of Citi Bike on the map in Fig. 2.

TABLE 3: CITI BIKE TRIP NETWORK STRUCTURE METRICS (JULY, 2018)

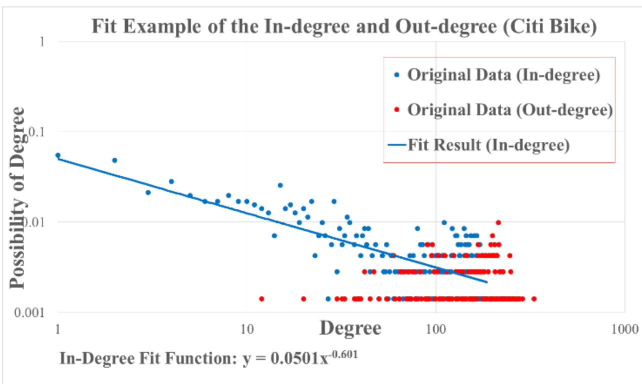
Item	Value	Item	Value
Nodes	707	**Diameter**	4
Links	41053	**Average Path Length**	1.947
Average In/Out Degree	58.066	**Graph Density**	0.082
Average Clustering Coefficient	0.592	**Modularity**	0.288

Figure 4 further illustrates the network topologies of the Citi Bike in July 2018 and the corresponding in- and out-degree distributions (Fig. 4 (C)). In Fig. 4 (A) and (B), the size of nodes is proportional to the node degree, and the color from blue to red corresponds to the number of the in- or out-degree from low to high. The in-degree and out-degree distributions of the network are plotted in Fig. 4(C).

The in-degree distribution follows a power law, as shown in Fig. 4(C) meaning that there are a few stations serving as the hubs (i.e., those big red nodes) that receive bikes returned from other stations, while most of the stations (i.e., those small blue nodes) only receive bikes from one or two stations. However, the out-degree distribution follows no apparent laws, and over half of the nodes have no out-degree (a few trips were started from those stations) and most of the remaining nodes have high out-degree values. The top five start and end station hubs are shown in Table 4. There are no stations found being both end and start station hubs in the Citi Bike.



(A) (B)



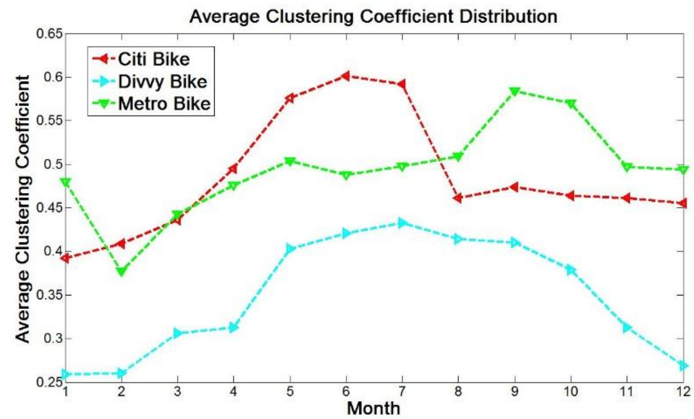
(C)

FIGURE 4: CITI BIKE TRIP NETWORK (JULY, 2018) VISUALIZATION AND DEGREE DISTRIBUTION FITTING. (A) TRIP NETWORK (IN-DEGREE); (B) TRIP NETWORK (OUT-DEGREE); (C) IN-DEGREE AND OUT-DEGREE DISTRIBUTION.

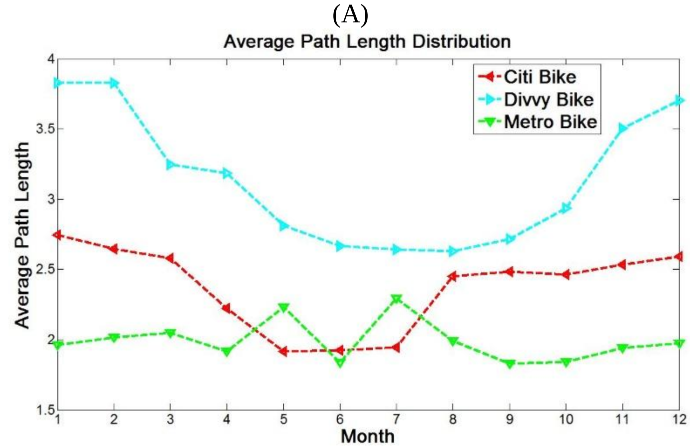
TABLE 4: CITI BIKE TOP FIVE START AND END STATION HUBS (JULY, 2018)

Station ID	Station Name	Out-degree
Top Five Start Station Hubs		
519	Pershing Square North	331
477	W 41 St & 8 Ave	288
402	Broadway & E 22 St	277
514	12 Ave & W 40 St	275
426	West St & Chambers St	270
Top Five End Station Hubs		
Station ID	**Station Name**	**In-degree**
497	E 17 St & Broadway	186
151	Cleveland Pl & Spring St	184
3435	Grand St & Elizabeth St	184
403	E 2 St & 2 Ave	184
368	Carmine St & 6 Ave	183

In Fig. 5, the network metrics, i.e., the average clustering coefficient and the average path length, of different cities are compared. It is found that in general spatially different cities' BSSs behave quite differently in terms of these two metrics; and temporally, the networks' structure evolves with obvious patterns. For example, as shown in Fig. 5 (A), the average clustering coefficient of the Divvy Bike network is lower than those of Citi Bike and Metro Bike networks. It indicates that the connectivity of stations in Divvy Bike is low comparing to that in Citi Bike and Metro Bike, which also can be verified by the average path length distributions (Fig. 5 (B)). Besides, the average path length distributions also illustrate that any pairs of stations in Divvy Bike are averagely linked by more trips than that in Citi Bike and Metro Bike.



(A)



(B)

FIGURE 5: TYPICAL NETWORK PARAMETER DISTRIBUTIONS OF THREE BIKE-SHARING SYSTEMS IN 2018. (A) AVERAGE CLUSTERING COEFFICIENT; (B) AVERAGE PATH LENGTH

On the time scale, Citi Bike has a higher average clustering coefficient and correspondingly lower average path length from April to July than the other two networks. For Divvy Bike, this period starts from May to October. However, the change in the average clustering coefficient in the Metro Bike network is minimal. These temporal distributions imply that in summer, stations in both the Citi Bike and the Divvy Bike have higher connectivity, and any pairs of stations in BSS are averagely connected by fewer trips. The possible factors that lead to these variations can be from both spatial and temporal aspects. From the spatial point of view, the geographic variation between stations in different BSS networks is potentially the key factor. According to Fig. 2, the station distribution of the Divvy Bike is expanding from the downtown area to the suburbs, and the stations are distributed much sparser in suburbs than those in downtown areas. As a result, the connectivity between suburb stations or between suburb and urban stations may be influenced by the distance between stations. On the contrary, a more even distribution of the Citi Bike stations makes the whole network have better transitivity. For the Metro Bike, stations are geographically divided into four main areas, as shown in Fig. 2 (C). The connectivity within each area determines the performance of the global network. From the temporal point of view, the seasonal effect should be the key factor because it is intuitive that more rides and outdoor activities would happen in warmer seasons. As shown in Fig. 5, there exist strong patterns in terms of the change of metrics during summer months in both Citi Bike and Divvy Bike networks. For Metro Bike, since there is not much weather fluctuation all year round in the Greater LA area, no significant changes of the trip network structure are observed.

b. Network Performance Analysis

Within a trip network, the in-degree and out-degree of a node correspond to how many trips originate from and end up to that node (station). Supposing there are a stations connected from station i, and b stations connected to station i, rebalance issues emerge when the difference between a and b is significant. Therefore, in this study, the rental problem is defined as that the connecting-from stations are less than the connecting-to stations, the return problem is defined as that the connecting-from stations are more than the connecting-to stations. At the system level, the binary network-based rental metric α and return metric β are defined to characterize a BSS's rebalancing performance. As shown in Fig. 6, assuming the number of in-degrees of the station i is a_i, and the out-degrees are b_i; We define $c_i = a_i - b_i$ as the metric to quantify the rebalancing performance at station i. Ideally, we expect $c_i = 0$ so that station is perfectly rebalanced in a month. If $c_i < 0$, that station will be more frequently serving as a trip starting point instead of ending point, thus would cause rental problems, and if $c_i > 0$, that station will be more frequently serving as a trip destination instead of origination, thus would cause a return problem.

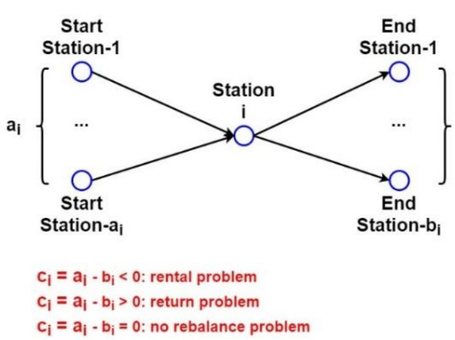


FIGURE 6: BSS TRIP NETWORK REBALANCE PROBLEM EVALUATION

If there are M stations having rental problems and N stations having return problems, α and β can be represented as:

$$\alpha = \frac{1}{M}\sum_{i=1}^{M}|c_i| \quad (5)$$

$$\beta = \frac{1}{N}\sum_{i=1}^{N}|c_i| \quad (6)$$

Figure 7 shows the results of α and β the 12 trip networks in three cities. Since larger values of α and β imply more serious rebalancing issues, it is clearly seen from Fig. 7 that Citi Bike has the worst rebalancing performance among the three BSSs, especially from April to July when the system is active and rides are frequent. Both return and rental curves reach a peak in June at 57.196 and 32.278, which means the rental problem in NYC is worse than the return problem. It implies that bikes from many stations come to one particular station, but not many stations will these bikes go to. And this problem becomes more significant in summer (tour season). The imbalance issue of Divvy Bike and Metro Bike networks is not as significant as Citi Bike, but rental problems remain worse in these two systems.

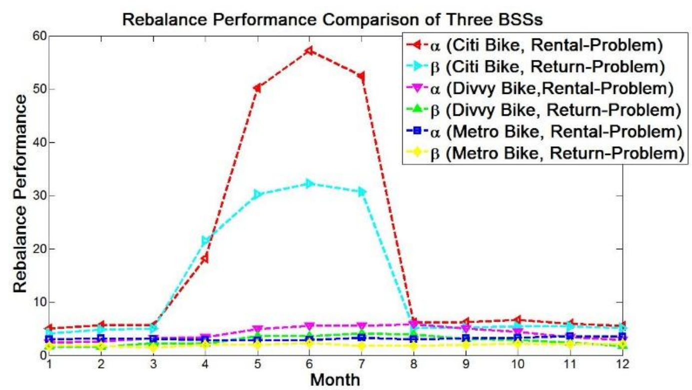


FIGURE 7: TRIP NETWORK REBALANCING PERFORMANCE OF CITI BIKE, DIVVY BIKE, AND METRO BIKE

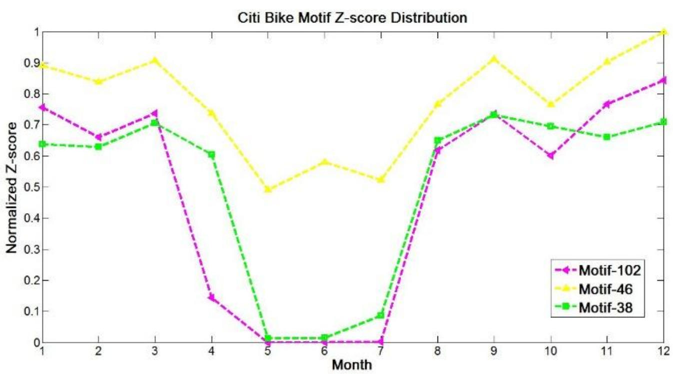
It should be noted that the rebalancing evaluation conducted in this study is different from those studies in operational research where probabilistic models and statistical methods are often used [23,24,40]. Our method focus more on the network structure representing the trip patterns. One limitation of the current method arises from the neglect of the link weight. That means when we define the rebalance performance of a station, it only reflects the number of stations that this station connects from (return) or connects to (rental), but not considers how many trips (i.e., the number of bikes) returned or rented to that station. Since we binarize the network based on the minimum mean of link weight, this operation would not result in significant bias (because most links only have low weights in the network as shown in Fig. 3). But further assessment is still needed in future research by using weighted networks. So, when a return problem is claimed in this study, it means *the number of satiations* where the bikes are received from is more than *the number of stations* where the bikes are running to. The same interpretation applies to the rental problem.

4.3 Local-level Trip Motifs Analysis

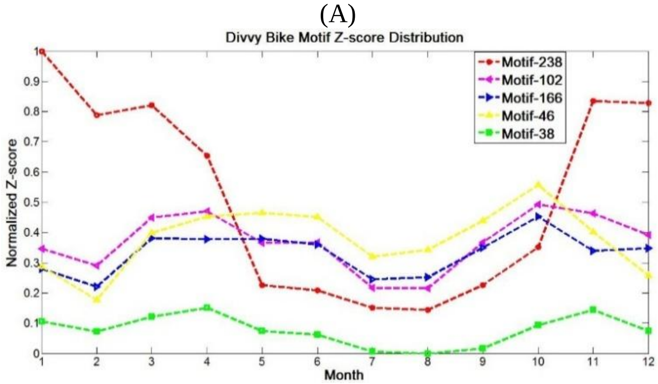
This section analyzes the size-3 motif characteristics of the trip networks, i.e., the local network structures, in three cities. Particularly, we present three analyses, including motif significance, interpretations of typical trip motifs, and motif structure characteristics.

a. Trip Motifs Significance Analysis

As aforementioned, network with larger size tends to have a higher z-score, the motif z-scores of Citi Bike are generally higher than those in the other two networks, and Metro Bike has the lowest z-scores. For the sake of offsetting the size effect, the z-scores are normalized. Figure 8 shows the z-scores of the top five most significant motifs in each network overtime in three cities, and Table 5 gives their corresponding interpretations. The rebalance metrics α and β are applied to study the rebalance performance of every local structure.



(A)



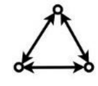
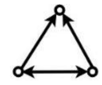
(B)

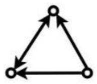
(C)

FIGURE 8: SIGNIFICANT MOTIF Z-SCORE DISTRIBUTION. (A) CITI BIKE; (B) DIVVY BIKE; (C) METRO BIKE

There are only three motifs (motif-102, motif-46, motif-38) in the Citi Bike trip network that are statistically significant in the whole year, so Fig. 8(A) has three curves only. The changes in their z-scores show consistent patterns, i.e., a sharp decrease is observed from April to July. In Divvy Bike and Metro Bike networks, in addition to the same three significant motifs, two more motifs (motif-238 and motif-166) are found significant. But, the trends of the z-scores in these two networks are quite different. For example, motif-238 in the Divvy Bike network has an obvious decline from May to October, yet in Metro Bike, the decline happens from July to December. One common characteristic of the significant motifs in these three cities is that trip motifs with closed topology (as shown in Table 5) are popular in all these three BSSs. No open triangles (those shown in Table 1) are found significant. It implies if both two stations have the same end station or start station, trips are very likely to happen between these two stations. So, in BSS, the trips in local network structures are highly correlated. This could be the reason that any two stations in these closed size-3 motifs can be easily reached physically (e.g., not too many blocks between them and the distance is short), therefore, customers can pick any route among the three to achieve their destinations. To test this hypothesis, we analyze the relationship between motif patterns and the average distance between stations in the flowing section.

TABLE 5: SIGNIFICANT TRIP MOTIFS INTERPRETATION

ID	Configuration	Rebalance evaluating metrics α, β	Interpretation
238		α = 0 β = 0	Completely connected trip motifs. Because of the shortest average distance, it has the highest transportation efficiency.
46		α = 1 β = 2	When one station is the end station of two different trips, then the start stations of these two trips are easy to be connected. This trip

			pattern has a potential return problem.
166		$\alpha = 2$ $\beta = 1$	When one station is the start station of two different trips, then the end stations of these two trips are easy to be connected. This trip pattern has a potential rental problem.
102		$\alpha = 1$ $\beta = 1$	When one station is the transitional station of two different stations, then direct connections exist between these two stations with high possibility.
38		$\alpha = 2$ $\beta = 2$	Two trip routes exist between two stations. This trip pattern has potential both return and rental problems.

Lastly, with our network motif-based method, the rebalance performances of all the motifs in all three BSSs can be evaluated. The size-3 motifs with serious rental problem (α=2) are: motif-166, motif-38, and motif-6; the size-3 motifs with serious return problem (β=2) are: motif-46, motif-38, and motif-36; the size-3 motifs with no rebalance problem (α=0, β=0) are: motif-238, motif-78, motif-140. Note that the remaining motifs (only α=1 or β=1, or both α=1 and β=1) are considered having the rental or return issues but not serious.

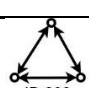
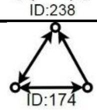
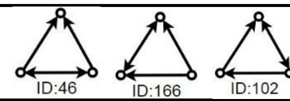
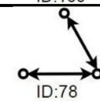
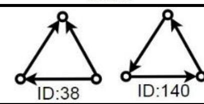
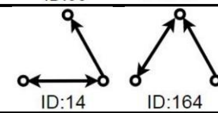
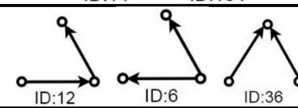
b. The Relationship between Motif Patterns and Average Distance

Given that the trip distance is one of the most important factors in customers' travel preference, relationships between motif patterns and the average distance between stations are analyzed. Firstly, assuming there is a size-3 motif *k* with nodes A, B, C, emerging in a trip network *h* times, and the latitudes and longitudes of A, B, C are known, the geographic distances between A, B, C (d_{AB}, d_{AC}, d_{BC}) can be calculated based on the GeoPy client [41]. Then, the average distance of motif *k* is:

$$D_{avg} = \frac{1}{h}\sum_{i=1}^{h}\left(d_{AB}^{i} + d_{AC}^{i} + d_{BC}^{i}\right) \tag{7}$$

Secondly, we categorize the size-3 motifs in different sets based on whether they are opened or closed and the number of directed links (i.e., the number of arrows). See Table 6.

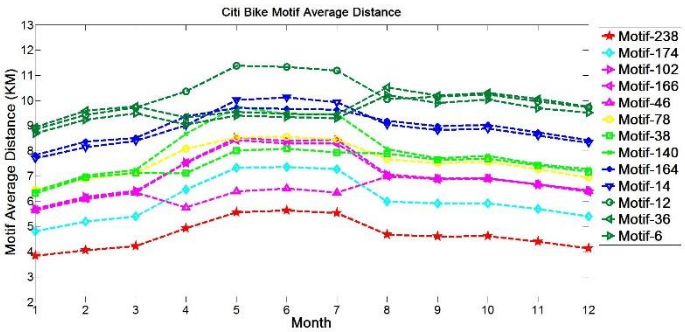
TABLE 6: MOTIF SET DEFINITION

Set Number	Structure	Structure Form	Arrow Number
Set 1		Closed	6
Set 2		Closed	5
Set 3-1		Closed	4
Set 3-2		Open	4
Set 4-1		Closed	3
Set 4-2		Open	3
Set 5		Open	2

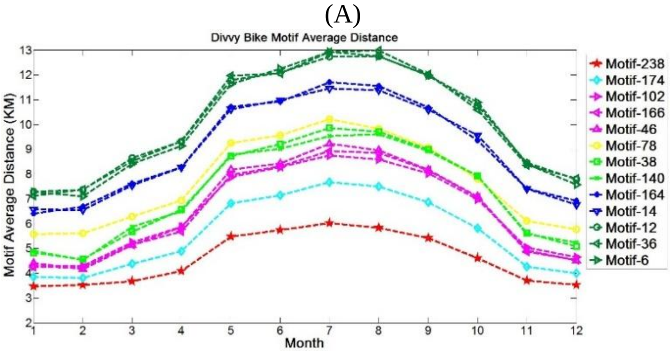
The average geographic distances of the motifs in the 36 trip networks for all three cities all around the year are calculated based on Equation (7). The results are plotted in Fig. 9, and there are two interesting findings: 1) a motif's average distance is larger in the summer time than that in the winter time, and 2) the average distance well corresponds to the motif patterns along the y direction. For example, especially the Divvy Bike, the seven groups of curves are vertically distributed, well corresponding to the seven sets of motifs from top to bottom in Table 6. With further analysis, the second finding is concluded as follows:

1) **Observation 1:** The average distance of motifs in closed-form is shorter than those in open-form.
2) **Observation 2:** Motifs with less number of arrows have a longer distance.
3) **Observation 3:** Motif form has a higher priority than the number of arrows in determining a motif's average distance. For example, considering motif-78, motif-38, and motif-140. Even if motif-78 has a greater number of arrows than motif-38 and motif-140, it has a longer average distance because it is an open-form motif.

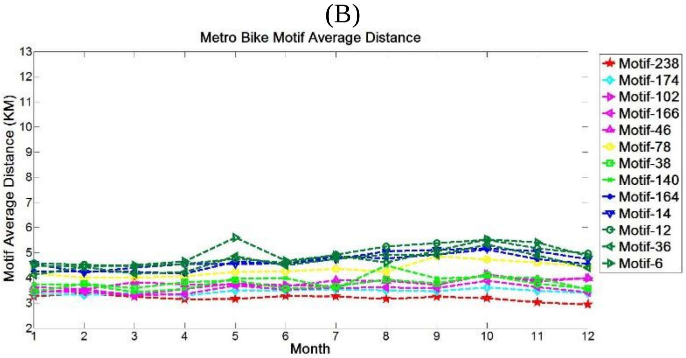
In Fig. 9(B), all the results follow these observations perfectly; in Fig. 9(A), the average distance of all the motifs follows those observations except from April to July; in Fig. 9 (C), the distance of motif-238 is distinctly shorter than the motifs in set 5. The possible explanations for these two inconsistent observations are as follows. First, the geographic distance of two stations is calculated instead of the route distance for simplification, so the distance data might not be accurate. Second, for Metro Bike, since the differences between any two motifs' average distance are very small, the impact of such a distance inaccuracy can be amplified which makes inconsistency happen. For example, according to Fig. 9 (C), the average distance of motif-174 is smaller than that of motif-238 in February, but based on our observation, the average distance of motif-174 should be larger. For Citi Bike, motif-46, motif-38, motif-36, and motif-6 are also inconsistent with our observation but share a very similar trend in the moths from April to July. The possible reason is that in the tour season of NYC (from April to July), Citi Bike is frequently used. Consequently, the size of its trip network size grows, thus more uncertainties can be introduced. To completely understand these inconsistencies, more detailed analyses need to be conducted in our future study.



(A)



(B)



(C)

FIGURE 9: MOTIF AVERAGE DISTANCE DISTRIBUTION. (A) CITI BIKE; (B) DIVVY BIKE; (C) METRO BIKE.

4.4 Discussion

In this section, we discuss the potential correlation between the local trip motif patterns and the global-level performance of the BSS networks. Firstly, based on all the results presented, it is found that the seasonal effect has a significant impact on both of the global trip networks and the local trip motifs. For example, during the summer time (from April to August) in Chicago, there is a significant increase in terms of trip motifs' average distance (Fig. 9(B)) and a decrease in the average path length of the trip networks (Fig. 5(B)). Secondly, according to Table 5, motif-46, i.e., one of the most significant motifs in all three networks indicated by the z-score in Fig. 8(A), indicates the return problem. However, Fig. 7 shows that the rental problem is more serious than the return problem at the global level (from April to July). Merely evaluating the system-level performance does not help acquire a complete picture of a BSS's rebalance performance. The entire system exhibits a rebalance problem is actually because unbalance station connections exist which means that the number of stations being treated as start points is different from that of the station being treated as end points. Solving the rebalancing issue of the entire system requires an in-depth understanding of the local-level rebalance problems, and vice versa.

According to the distributions of the average distance shown in Fig. 9, strategies for adding new stations or customer-oriented rebalancing strategies can be developed to promote the formation of certain motifs from the bottom up to solve the system-level rebalancing issues. For instance, as shown in Fig. 10, if motif-46 is identified in the trip network with stations A, B, C, the system designer could establish a new station D with shorter distances to B and C to promote the formation of trip motif-238 which has better rebalance performance. In addition, if a closer station D has existed, certain incentives (e.g., discounts) or more bike allocation can be made to encourage customers to travel between stations B, C, and D instead of stations A, B, and C to promote the formation trip motif-238.

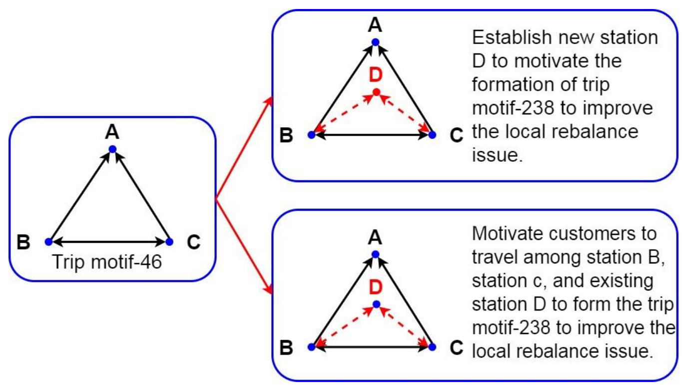


FIGURE 10: EXAMPLE OF OPTIMIZING STRATEGIES OF LOCAL REBALANCING ISSUES.

5. CONCLUSION

In this paper, to fill the research gap of gaining a better understanding of the local-global relations of the STS, we propose an STS analysis approach based on the network motif theory. This approach takes both global-level network topology and local-level network motifs into consideration and can help designers and systems engineers more accurately identify the potential causes of inadequate system performance. In this study, BSS is chosen as an example for demonstrating and validating the proposed approach. To the best of our knowledge, this is the first time BSS is treated from an STS point of view, and the network motif theory is applied to analyze its system structure and rebalance performance.

Using the BSSs from three cities, we have analyzed the global trip network structures and the rebalance performance as well as the characteristics of the local trip motifs. The conclusions are: 1) season effect is a critical factor to influence both the global and the local network structures and performance; 2) global trip networks could exhibit performance that is different from what is illustrated in the local trip motifs; 3) obvious hierarchical distributions are found by calculating the average distances of the trip motifs, and the distance closely relates with the motif topologies: a. motifs with closed-form have shorter distance than those with open form; b. motifs with more directed arrows have shorter distance than those with less directed arrows; c. motif form has a higher priority than the directed arrow numbers. Based on these conclusions, rebalancing strategies can be developed by considering factors such as season, geographical distance, and motif patterns. Besides the mentioned customer-oriented strategies, truck-based rebalance strategies also can be established by considering motif structures with serious rebalance problems. Tests of the design strategies developed based on network motif-based insights will be considered in our future study.

Additionally, compared to the BSS dynamic rebalancing analyses in operation research that predict station pick-up and drop-off demands based on the data in a shorter time period (e.g., hourly) [42,43], our network-based analyses study travel paterns in a longer period (i.e., monthly) aiming to understand the relations between local travel patterns (represented by size-3 motifs) and the system rebalancing issues. In our future study, the weighted BSS network (taking trip frequency as the link weight) will be analyzed to inform the design of strategies for the improvement of system performance. In such a study, those approaches to demand prediction from the operational research can be integrated into the network-based framework to test if certain mechanisms and/or interventions could be indeed effective in improving a BSS's rebalance performance.

ACKNOWLEDGMENTS

We would like to thank Khoinguyen Trinh for sharing the research ideas of motif analysis. We would also like to thank Xingang Li and Jared Poe for helping the proofreading and the revision of the manuscript.

REFERENCES

[1] Trist, E. L., and Bamforth, K. W., 1951, "Some Social and Psychological Consequences of the Longwall Method of Coal-Getting," Hum. Relations, **4**(1), pp. 3–38.

[2] Cherns, A., 1976, "The Principles of Sociotechnical Design," Hum. Relations, **29**(8), pp. 783–792.

[3] Clegg, C. W., 2000, "Sociotechnical Principles for System Design," Appl. Ergon., **31**(5), pp. 463–477.

[4] Carayon, P., 2006, "Human Factors of Complex Sociotechnical Systems," Appl. Ergon., **37**(4 SPEC. ISS.), pp. 525–535.

[5] Baxter, G., and Sommerville, I., 2011, "Socio-Technical Systems: From Design Methods to Systems Engineering," Interact. Comput., **23**(1), pp. 4–17.

[6] Mitchell, M., 2006, "Complex Systems: Network Thinking," Artif. Intell., **170**(18), pp. 1194–1212.

[7] Costa, L. D. F., Rodrigues, F. A., Travieso, G., and Boas, P. R. V., 2007, "Characterization of Complex Networks: A Survey of Measurements," Adv. Phys., **56**(1), pp. 167–242.

[8] Sha, Z., and Panchal, J. H., 2013, "Estimating the Node-Level Behaviors in Complex Networks from Structural Datasets," *Proceedings of the ASME Design Engineering Technical Conference*, American Society of Mechanical Engineers, Portland, Oregon, USA, pp. 1–14.

[9] Sha, Z., and Panchal, J. H., 2016, "A Degree-Based Decision-Centric Model for Complex Networked Systems," *Volume 1B: 36th Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, Charlotte, New Carolina, USA, pp. 1–11.

[10] Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y., and Chen, W., 2016, "Forecasting Technological Impacts on Customers' Co-Consideration Behaviors: A Data-Driven Network Analysis Approach," *Volume 2A: 42nd Design Automation Conference*, American Society of Mechanical Engineers.

[11] Wang, M., Chen, W., Huang, Y., Contractor, N. S., and Fu, Y., 2016, "Modeling Customer Preferences Using Multidimensional Network Analysis in Engineering Design," Des. Sci., **2**, p. e11.

[12] Fu, Jiawei, Zhenghui Sha, Yun Huang, Mingxian Wang, Yan Fu, and W. C., 2017, "Two-Stage Modeling of Customer Choice Preferences in Engineering Design Using Bipartite Network Analysis," *ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Cleveland, Ohio, USA, pp. 1–11.

[13] Sha, Z., Huang, Y., Fu, J. S., Wang, M., Fu, Y., Contractor, N., and Chen, W., 2018, "A Network-Based Approach to Modeling and Predicting Product Coconsideration Relations," Complexity, **2018**, pp. 1–14.

[14] Bi, Y., and Chen, W., 2018, "Modeling Spatiotemporal Heterogeneity of Customer Preferences in Engineering Design," *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Quebec City, Quebec, Canada, pp. 1–12.

[15] Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y., and Chen, W., 2018, "Predicting Product Co-Consideration and Market Competitions for Technology-Driven Product Design: A Network-Based Approach," Des. Sci., **4**, p. e9.

[16] Si, H., Shi, J., Wu, G., Chen, J., and Zhao, X., 2019, "Mapping the Bike Sharing Research Published from 2010 to 2018: A Scientometric Review," J. Clean. Prod., **213**, pp. 415–427.

[17] Midgley, P., 2009, "The Role of Smart Bike-Sharing Systems in Urban Mobility," Journeys, **2**(1), pp. 23–31.

[18] Parkes, S. D., Marsden, G., Shaheen, S. A., and Cohen, A. P., 2013, “Understanding the Diffusion of Public Bikesharing Systems: Evidence from Europe and North America,” J. Transp. Geogr., **31**, pp. 94–103.
[19] DeMaio, P., 2009, “Bike-Sharing: History, Impacts, Models of Provision, and Future,” J. Public Transp., **12**(4), pp. 41–56.
[20] Bachand-Marleau, J., Lee, B. H. Y., and El-Geneidy, A. M., 2012, “Better Understanding of Factors Influencing Likelihood of Using Shared Bicycle Systems and Frequency of Use,” Transp. Res. Rec. J. Transp. Res. Board, **2314**(1), pp. 66–71.
[21] Yi, P., Huang, F., and Peng, J., 2019, “A Rebalancing Strategy for the Imbalance Problem in Bike-Sharing Systems,” Energies, **12**(13), p. 2578.
[22] Lin, L., He, Z., and Peeta, S., 2018, “Predicting Station-Level Hourly Demand in a Large-Scale Bike-Sharing Network: A Graph Convolutional Neural Network Approach,” Transp. Res. Part C Emerg. Technol., **97**(October), pp. 258–276.
[23] Chemla, D., Meunier, F., and Wolfler Calvo, R., 2013, “Bike Sharing Systems: Solving the Static Rebalancing Problem,” Discret. Optim., **10**(2), pp. 120–146.
[24] Dell’Amico, M., Hadjicostantinou, E., Iori, M., and Novellani, S., 2014, “The Bike Sharing Rebalancing Problem: Mathematical Formulations and Benchmark Instances,” Omega (United Kingdom), **45**(January), pp. 7–19.
[25] Çelebi, D., Yörüsün, A., and Işık, H., 2018, “Bicycle Sharing System Design with Capacity Allocations,” Transp. Res. Part B Methodol., **114**, pp. 86–98.
[26] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U., 2002, “Network Motifs: Simple Building Blocks of Complex Networks,” Science (80-.)., **298**(5594), pp. 824–827.
[27] Holland, P. W., and Leinhardt, S., 1975, “The Statistical Analysis of Local Structure in Social Networks,” Sociol. Methodol., pp. 1–45.
[28] Stone, L., Simberloff, D., and Artzy-Randrup, Y., 2019, “Network Motifs and Their Origins,” PLoS Comput. Biol., **15**(4), pp. 1–7.
[29] Alon, U., 2007, “Network Motifs: Theory and Experimental Approaches,” Nat. Rev. Genet., **8**(6), pp. 450–461.
[30] Felmlee, D., Mcmillan, C., and Towsley, D., 2018, “Social Network Motifs : A Comparison of Building Blocks across Multiple Social Networks,” *Annual Meetings of the American Sociological Association*, Philadelphia, US., pp. 1–10.
[31] Paranjape, A., Benson, A. R., and Leskovec, J., 2017, “Motifs in Temporal Networks,” *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining - WSDM ’17*, ACM Press, New York, USA, pp. 601–610.
[32] Kashtan, N., Itzkovitz, S., Milo, R., and Alon, U., 2004, “Efficient Sampling Algorithm for Estimating Subgraph Concentrations and Detecting Network Motifs,” Bioinformatics, **20**(11), pp. 1746–1758.
[33] Wernicke, S., and Rasche, F., 2006, “FANMOD: A Tool for Fast Network Motif Detection.,” Bioinformatics, **22**(9), pp. 1152–3.
[34] Choobdar, S., Ribeiro, P., and Silva, F., 2012, “Motif Mining in Weighted Networks,” *Proceedings - 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012*, pp. 210–217.
[35] Florian, R., Sebastian, W., “FANMOD Fast Network Motif Detection MANUAL” [Online]. Available: http://theinf1.informatik.uni-jena.de/motifs/fanmod-manual.pdf.
[36] “Citi Bike System Data | Citi Bike NYC” [Online]. Available: https://www.citibikenyc.com/system-data. [Accessed: 21-Feb-2020].
[37] “Divvy System Data | Divvy Bikes” [Online]. Available: https://www.divvybikes.com/system-data. [Accessed: 21-Feb-2020].
[38] “Data – Metro Bike Share” [Online]. Available: https://bikeshare.metro.net/about/data/. [Accessed: 21-Feb-2020].
[39] Bastian, M., Heymann, S., and Jacomy, M., 2009, “Gephi: An Open Source Software for Exploring and Manipulating Networks Visualization and Exploration of Large Graphs,” *Third International AAAI Conference on Weblogs and Social Media*.
[40] Cruz, F., Subramanian, A., Bruck, B. P., and Iori, M., 2017, “A Heuristic Algorithm for a Single Vehicle Static Bike Sharing Rebalancing Problem,” Comput. Oper. Res., **79**(2009), pp. 19–33.
[41] “Welcome to GeoPy’s Documentation! — GeoPy 1.21.0 Documentation” [Online]. Available: https://geopy.readthedocs.io/en/stable/#. [Accessed: 22-Feb-2020].
[42] Chiariotti, F., Pielli, C., Zanella, A., and Zorzi, M., 2018, “A Dynamic Approach to Rebalancing Bike-Sharing Systems,” Sensors (Switzerland), **18**(2), pp. 1–22.
[43] Liu, J., Sun, L., Chen, W., and Xiong, H., 2016, “Rebalancing Bike Sharing Systems: A Multi-Source Data Smart Optimization,” Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., **13-17-August-2016**, pp. 1005–1014.