Long Short-Term Memory and Graph Convolution Network for Forecasting the Crude Oil Traffic Flow

QI OUYANG, TENGDA SUN, YUANYUAN XUE, AND ZHEHUI LIU
China Transport Telecommunications and Information Center, Beijing 100011, China
Corresponding author: Tengda Sun (suntengda@cttic.cn)

ABSTRACT
Understanding maritime network structure and traffic flow changes is a challenging task that must incorporate economic, energy, geopolitics, maritime transportation, and network sciences. Crude oil is the most imported energy in the world. Investigating the crude oil maritime network status and predicting the crude oil traffic flow changes has great significance for the global trade, especially for key crude oil importing/exporting regions and countries. To address this, a system-based approach using long short-term memory and graph convolution network for the crude oil traffic flow forecasting named LGCOTFF is introduced. The LGCOTFF approach constructs a maritime transportation network firstly, and then calculates and predicts the node traffic flow based on trajectory data and crude oil berth geographical position. Firstly, we construct a maritime crude oil transportation network based on supply-demand relationship, ship trajectory and route information. Then, we design an approach to calculate how many crude oil ships finished up-load/offtake tasks in a single week for each port, and gather this data to countries and regions. Finally, we design a deep learning neural network named long short-term memory and graph convolution network (L-GCN) to extract the temporal and spatial characteristics of crude oil transportation, and predict the node traffic flow. We evaluate the proposed model on China, Russia, Middle East and America respectively and observe consistent improvement of more than 10% over state-of-the-art baselines.

INDEX TERMS
Crude oil transportation network, traffic flow, graph convolution network.

I. INTRODUCTION
Crude oil is one of the most important energies in the world. As one of the main transportation ways of crude oil transportation, maritime transportation undertakes almost all transportation tasks of the global intercontinental crude oil trade. As Clarkson’s statistics shows, the global crude oil maritime transportation volume for three consecutive years in 2017, 2018, and 2019 reached more than 2 billion tons. Namely, the crude oil maritime transportation network also plays one of the busiest international trade networks in the world.

With the high frequency and changeability characteristics of international crude oil trade, the crude oil maritime transportation network always changes correspondingly. For instance, affected by the outbreak of the COVID19, Clarkson’s statistics shows that the global crude oil maritime transportation volume in 2020 drop down to 1.8 billion tons, a year-on-year decrease of 6.6% compared with 2019. This event also triggered big changes in the structure of global crude oil import and export (I/E). The import volume in major crude oil importing regions and countries has experienced negative growth with different degrees, and the export structure and volume in major crude oil exporting regions and countries has also changed significantly. Therefore, dynamically monitoring the status of crude oil transportation in major I/E regions and countries and predicting the crude oil demand for a certain period of time in the future is very meaningful, which can help us optimize the maritime capacity structure in advance. Furthermore, crude oil belongs to The Materials and Methods should be described with sufficient details to allow others to dangerous goods, so mastering how many crude oil ships are arriving at certain region in the future will also do a favor to ensure the safety of transportation operations.

So far, there are a lot of related research works in an area of maritime traffic flow calculation and prediction, but they all have two problems. The first one is that while calculating traffic flow volume through a node, common method is...
taking the total number of ships in a certain fenced area into account. Considering that in the process of crude oil maritime transportation, ships may only refuel or supply when passing by some regions and countries such as Singapore, and counting the total number of ships in the fenced area will increase the weight of this node and affect the accuracy of the prediction results. The second one is that as predicting traffic flow, recent researches mainly focused on time series data of single ports. Through signal decomposition algorithms, time domain information for predicting port traffic flow volume or throughput is extracted. However, in the global crude oil maritime transportation network, all of that crude oil I/E regions and countries have a great impact on each other. So, only considering a single port for prediction will also affect the accuracy of the results.

To solve the problems above and improve the prediction accuracy, our work makes the following contributions:

(i) We propose an algorithm to calculate how many crude oil ships actually finished upload/offtake tasks in a period of time, and use this data as the node features to correct errors caused by refuel or supply operations.

(ii) We build a graph of maritime crude oil transportation. This graph takes supply and demand relationship into consideration, and gathers the traffic flow volume of each port that belongs to a certain country, as well as forms the spatial structure characteristics of the crude oil maritime transportation network.

(iii) We propose a L-GCN model to extract temporal and spatial features in real automatic identification system (AIS) and geographic information data. Then we take China, Russia, Middle East and America as crude oil importing and exporting country examples respectively, and experiment results show that the proposed approach achieves more than 10% relative error reduction over state-of-the-art baseline methods for crude oil traffic flow volume predicting. In our knowledge, it is the first time that a graph convolutional network (GCN) is used in maritime traffic predicting. In the following part of this paper, the traffic flow volume represents the number of ships that finished upload/offtake tasks.

The rest parts of this paper are organized in the following ways. In section 2, the related researches are introduced. The experimental data and methods are explained in detail afterwards. In section 4, we evaluate our experimental results comparatively. In section 5, we give a conclusion to our work.

II. RELATED WORKS

Crude oil is one of the most widely used energy sources and the most important commodities in the global economy, which plays a pivotal role in global economic development, social stability, and national security in almost all countries. The crude oil maritime transportation network is also one of the most important networks in the world, which is directly related to the global economy, financial market, transportation and so on.

Many scholars have made great contributions to the construction of crude oil maritime transportation network. Their research is very meaningful and does a lot of favor to our study. In their research [11]–[4], countries are divided into two groups. One is the oil import-dependent country group and the other is the oil export-dependent countries. Each group contains several countries. The routes from oil exporting countries to oil importing countries constitute the crude oil transportation network. In this way, only the major crude oil producing or consuming countries are included in the network. Some key nodes are ignored. Besides, America becomes a net exporter of crude oil in 2019, but it also imports a lot of crude oil at the same time. It is not enough to simply treat these nodes as import or export dependent countries. In order to correct the problem and make the network structure more accurate, we define these nodes as crude oil import and export equilibrium countries in transportation network.

In maritime traffic flow volume statistics, the method of setting door lines or fences is widely used. L. Kang et al. and L. Zhang et al. [5], [6] divided the waters of the Singapore port into multiple parts according to actual conditions, then they calculated how many ships passed the set door line or fence during a period of time for each part and analyzed the traffic behavior such as speed, destination and so on of each ship. B. Gunnarsson et al., E. Merico et al. and L. Qi et al. [7]–[9] also used this method to obtain the ship behavior of Northern Sea, Adriatic Sea and Changshan Channel. Based on the obtained ship behavior, researchers studied the atmospheric impact and policy impact on maritime transportation. They have achieved a lot of exciting results in maritime traffic behavior field. However, if we use the number of ships passing through an area in global crude oil transportation network, the key nodes ships refueled or supplied such as Singapore will be given a greater weight which may increase the impact of this node on other nodes. To fix this, we propose an algorithm to calculate the number of ships that finished upload or takeoff tasks.

Traffic flow or transportation volume prediction has always been the research focus of scholars. There is a rich amount of works on this topic, including predicting bike flows, the traffic demand, the port throughput and so on [10]–[13]. Compared with land transportation, maritime transportation information is more difficult to obtain. So, in the maritime transportation prediction, traditional machine learning methods are widely used. M. Intihar et al. [14] used auto regressive integrated moving-average model with exogenous inputs (ARIMAX) model to forecast container throughput. In ARIMAX model, more factors such as GDP and surrounding port throughput are considered within the time series. With the help of more features, the model achieves better results than the time series model. M. Eskafi et al. [15] applied a Bayesian statistical method to forecast the annual throughput of the multipurpose port of Isafjordur in Iceland. In this model, the national GDP (NGDP), the average yearly CPI (ACPI), the world GDP (WGDP), the volume of national export trade (VNET), the volume of national import trade (VNIT), and the national population (NPOP) are concerned. L. Mo et al. [16] proposed a group method of data
handling (GMDH) neural network to forecast container throughput. In GMDH neural network, the seasonal auto-regressive integrated moving average (SARIMA) approach is used to predict the linear trend and support vector regression (SVR), back-propagation (BP) neural network was used to predict the nonlinear sub-series. GMDH can overcome the short-comings of the time series model partly. Other combined approaches [17]–[19] are also widely used by researchers to achieve a better performance in forecasting. In the road traffic area, because more data is available, graph convolutional network (GCN) is widely used in traffic flow prediction. L. Bai et al., B. Yu et al. and L. Zhao et al. [20]–[22] used GCN to forecast the traffic flow, the traffic speed and the taxi demand, respectively. They all achieved better results than baseline methods. The maritime transportation network is quite different from the traditional road transportation. The maritime transportation network has no certain routes, so we have to extract a graph from the AIS data. In this paper, we also use the LSTM structure obtain the time characteristics after using GCN. These are the main differences between our work and the related work.

III. MATERIALS AND METHODS

To achieve a better crude oil prediction result, we propose a crude oil traffic flow forecast model containing three sub-models in this paper. Our model mainly uses L-GCN as our key deep learning method to do crude oil forecast tasks, so we named our model LGCOTFF. The L-GCN model uses a long short-term memory to process the time series dependence in data and uses the graph convolutional network to extract the spatial relationship between data. Compared with the research introduced above, taking the temporal and spatial characteristics in maritime traffic flow forecasting is our main innovation point in this research.

Besides the L-GCN neural network, our model also contains the crude oil transportation network establishment and the traffic flow calculation. These two parts helps to improve the accuracy of crude oil transportation network structure and the feeding quality of forecast tasks.

Our dataset is proved by the national water information service platform. The dataset contains 2000 crude oil ship, 200 million crude oil ship location point data from January 2020 to July 2021 and 10 thousand berth location and information data which is a real big dataset.

In the LGCOTFF model, the first sub-model aggregates berth data to countries or regions, then constructs the maritime crude oil transportation network based on three different attributes of region. The second sub-model uses AIS data and the berth information provided by myships.com to calculate the crude oil ship number finished upload or take-off tasks in a period of time. The third sub-model feeds the results calculated by the models above into the L-GCN model to forecast the node features. In our case, the node features represent the crude oil ship number finished upload or takeoff tasks. Figure 1 shows the structure of our model.

A. THE CONSTRUCTION OF CRUDE OIL TRANSPORTATION NETWORK

In order to calculate these data efficiently, we use our knowledge related to crude oil supply-demand to form a maritime crude oil transportation network. Due to the addition of prior knowledge, the graph convolutional neural network can extract spatial features faster and increase the calculation speed.

Most scholars divide all the countries into crude oil importers or crude oil exporters. However, there are two countries that both import and export oil. One is United Kingdom (GB), the other is United States (US). (https://www.worldstopexports.com/). Singapore (SG) is the main country for crude oil spot trading. These three countries play a different role in the crude oil transportation network. So, we divide them into an oil-balance group. The other two groups are oil-import-dependent group and oil-export-dependent group, respectively.

Figure 2 shows the fifteen oil-import-dependent countries, including sixteen European countries [Sweden (SE), Spain (ES), Slovakia (SL), Romania (RO), Portugal (PT), Poland (PL), Netherlands (NL), Lithuania (LT), Italy (IT), Ireland (IE), Greece (GR), Germany (DE), Belgium (BE), Bulgaria (BG), Croatia (HR) and France (FR)], and six Asian countries [Thailand (TH), Philippines (PH), South Korea (KR), Japan (JP), India (IN) and China (CN)].
Q. Ouyang et al.: Long Short-Term Memory and Graph Convolution Network for Forecasting Crude Oil Traffic Flow

Figure 2 also shows the twenty oil-export-dependent countries in the study, including two European countries [Russia (RU) and Norway (NO)], seven Asian countries [Saudi Arabia (SA), Iraq (IQ), Iran (IR), Kuwait (KW), United Arab Emirates (AE), Qatar (QA) and Oman (OM)], six African countries [Libya (LY), Guinea (GN), Nigeria (NG), Algeria (DZ), Angola (AO) and Gabon (GA)], three South American countries [Ecuador (EC), Venezuela (VE) and Brazil (BR)], and two North American countries [Canada (CA) and Mexico (MX)] [4]. These countries include the fourteen members of the Organization of Petroleum Exporting Countries (OPEC) along with the other top eight countries export oil.

According to the relationship between countries and ports, we aggregate ports to the country these ports belong to. Then, based on the crude oil import and export characteristics and geographical distribution around the world, we will gather some countries into regions to reduce the size of the transportation network. There is total eleven major nodes in the network, including China, other Asian crude oil import countries, European crude oil import countries, the Middle East, the West Africa, the South America, the Central and North America, Russia, the United States, the United Kingdom, and Singapore.

Since the graph convolution depends on the degree matrix, the nodes in network must have in-degree and out-degree at the same time. Combining the import and export characteristics of the nodes, when a node only has in-degree or out-degree (a node only imports or exports crude oil), the node needs a self-connection operation. Based on the above relevant information, the process of constructing a global crude oil maritime transportation network is shown in figure 3. Through the country aggregation operation (a country contains serval berths, then we add the ship number together if these berths belong to the same country), we reduce the number of nodes in the graph from 6000 by 50. Through the region aggregation operation (for example, the middle east contains serval countries, then we add the ship number together if these countries belong to the same region), we reduce the scale of the graph structure to 11 nodes once again. By setting the oil-balance group, we have obtained a more accurate crude oil transportation network. This structure improves the accuracy for some nodes. We will discuss this in detail in the section 4.

B. THE ALGORITHM OF PORT CRUDE OIL SHIP OPERATION TIMES OBTAINED

In this part, we will try to distribute node traffic flow to the crude oil transportation network. As we mentioned before, many current researches are based on fence data, which will magnify the value of some nodes. Our task in this part is to extract the node traffic flow through 200 million ship position points and 10 thousand berth locations and reduce errors caused by fence data efficiently.

This algorithm takes the basic information of crude oil ships, the longitude and latitude information of berths and the AIS data of ships as the input. Using the relationship between the temporal and spatial information hidden in ship trajectories and geographic information of berths, we calculate the time when the ship enters the berth and the time when the ship leaves the berth. Then, the operation times of crude oil ships in the port in a certain period of time are obtained. For the import-dependent countries and the export-dependent countries, this value is equal to the traffic flow of crude oil ships. To complete the calculations, this algorithm mainly consists of the following three parts. Figure 4 shows the three parts of the algorithm and the connection between the three parts.

STEP1: For a single ship, the ais dataset can be described as \[\{(u_1,lon_1, lat_1, sog_1), \ldots, (u_n,lon_n, lat_n, sog_n)\}\], where \(u\) is
FIGURE 4. Total structure of GCOTFF and the roles each part played in the model.

the universal time coordinated (UTC) timestamp, lon is the longitude the ship reported at time \(u\), lat is the latitude the ship reported at time \(u\), and sog is the speed over ground (SOG) of the ship at time \(u\). For a single berth, the berth dataset contains two attributes (blon, blat). blon represents the berth longitude, blat is for the berth latitude and \(i\) is the berth number.

Due to the SOG of ships is always small when berthing, we extract the ship ais data when the SOG is less than 0.5. According to the trajectory information of the ship and the geographical information of the berth, we obtain the distance \(dis_{i,j}\) between a certain ship and a certain berth at time \(u_{i}\). If \(dis_{i,j}\) is less than a super parameter \(d\), \((u_{i}, lon_{i}, lat_{i}, sog_{i}, blon_{i}, blat_{i}, dis_{i,j})\) is a possible berthing point.

For multiple ships and multiple berths, we combine the data belonging to different groups through the Cartesian product firstly, and then we calculate the distance between ships and berths. This way, by increasing the calculation space, a part of the calculation time can be reduced.

STEP2: In the possible berthing point dataset, a ship trajectory may belong to two different berths at the same time. We consider a ship trajectory only belongs to the nearest berth at a certain time. Namely, if \(dis_{i,j} < dis_{i,k}\), we will drop \((u_{i}, lon_{i}, lat_{i}, sog_{i}, blon_{i}, blat_{i}, dis_{i,j})\). Thus, we establish a one-to-one relationship between the ship trajectories and the berth data.

STEP3: After two steps above, we have a new dataset for a certain ship and berth \([u_{i}, lon_{i}, lat_{i}, sog_{i}, blon_{i}, blat_{i}, dis_{i,j}], \cdots, (u_{n}, lon_{n}, lat_{n}, sog_{n}, blon_{n}, blat_{n}, dis_{n,j})\] if \(u_{i+1} - u_{i} < t_{1}\), these two records belong to the same berthing operation. If \(u_{i+1} - u_{i} > t_{2}\), these two records belong to different berthing operations respectively. \(t_{1}, t_{2}\) are two super parameters. The dataset becomes \([u_{i}, \alpha_{1}], \cdots, (u_{n}, \alpha_{k})\], where \(k\) is the \(k\)-th berthing operation. For operation \(k\), the earliest time is the arrival time and the latest time is the departure time.

Due to the particularity of crude oil transportation, some countries do not necessarily have ships berthing every day, so a very sparse traffic flow matrix will be obtained in the unit of day. In order to avoid the occurrence of a sparse matrix, we count the traffic of each node by week. This operation also reduces the size of input data.

By using the data of berth points, we obtain more accurate node flow at certain nodes. The relevant results will be discussed in the section 4. At the same time, due to changes in the flow calculation methods of some key nodes, the nodes connected to them in the graph will also change. The relevant situation will also be discussed in section 4.

C. THE L-GCN FOR NODE FEATURE FORECAST

Through the above two steps, we build a global maritime crude oil transportation network with three different node attributes. In this part, we will introduce how to get the traffic flow changes of each node in the graph through a period time data.

According to our marine crude oil transport network above, each key country or region is defined as a node \(v \in V\) in the graph. \(V\) denotes the set of all regions in the global crude oil maritime transportation network. Let \(X_{t}(x_{1}, \ldots x_{v})\) represent the number of finished operation ships in all regions at the \(t\)-th interval. Then the forecasting problem is formulated as a single step spatiotemporal prediction given input with a fixed temporal length in \(\mathbb{R}^{V \times 1}\). Namely, learning a function maps historical values of all regions to the demand in the next timestep

\[
[X_{t-T}, \ldots X_{t}] \xrightarrow{f(\cdot)} X_{t+1}
\]

(1)

In a certain graph, each element in \(X_{t}\) has a spatial connection with other adjacent nodes. This relationship can be expressed by the adjacency matrix.

\[
A_{N,ij} = \begin{cases} 
1, & v_i and v_j are adjacent \\
0, & otherwise 
\end{cases}
\]

(2)

Different nodes have different number of edges and different weight. If a node has many edges, the node feature will be a great value. So, it is necessary to normalize the information (including self-loop information) from neighbors before the node updates. The normalization process can be expressed by Equation 3.

\[
X^* = \tilde{D}^{-1/2}A\tilde{D}^{-1/2}X
\]

(3)

where \(\tilde{D} = \sum_j \tilde{A}_{ij}\).
Thus, in a one-layer graph convolution at time $t$, the following layer-wise propagation rule\cite{23}:

$$X^1_t = \sigma(D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}} X^0_1 W^{-1}_1)$$  \hspace{1cm} (4)

where $X^0_1 \in \mathbb{R}^{V \times 1}$ is the matrix of activations in the 0th layer, $\sigma(\cdot)$ denotes an activation function.

Through the above operations, we get the spatial information of all the nodes in the network. Next, we will use the LSTM network to extract the temporal features of all the nodes. After graph convolution at all time step, the total dataset can be described as $[X_1(t_1,x_1,1, \cdots, x_t,v), \cdots, X_l(t_l,v_1, \cdots, x_l,v)]$. Then, we transpose the data according to the index of the node. The new dataset is $[N_1(x_1,1, \cdots, x_1), \cdots, N_l(x_1,v_1, \cdots, x_l,v)]$. For each node, we feed the sequence with a total length $t$ into the LSTM structure. The LSTM structure \cite{24} can be described by the following equations:

$$i_t = \sigma(W_{ix} x_t + W_{ih} h_{t-1} + b_i)$$  \hspace{1cm} (5)

$$f_t = \sigma(W_{fx} x_t + W_{fh} h_{t-1} + b_f)$$  \hspace{1cm} (6)

$$o_t = \sigma(W_{ox} x_t + W_{oh} h_{t-1} + b_o)$$  \hspace{1cm} (7)

$$\tilde{C}_t = \tanh(W_{cx} x_t + W_{ch} h_{t-1} + b_c)$$  \hspace{1cm} (8)

$$C_t = i_t \otimes \tilde{C}_t + f_t \otimes C_{t-1}$$  \hspace{1cm} (9)

$$h_t = o_t \otimes \tanh(C_t)$$  \hspace{1cm} (10)

where $t$ stands for $t$-th timestamp. $i_t$ refers to the output of the input gate. $f_t$ refers to the output of forget gate, and $o_t$ is the output of output gate. $x_t$ is the input vector. $C_t$ is the state vector, and $h_t$ is the hidden vector. $h_{t-1}$ is the former output of $h_t$, and $C_t$ is the input state and output state of the memory cell. $\sigma$ and $\tanh$ are activation functions between different layers. $\otimes$ is the Hadamard product. Other variables like $W_{ix}, W_{ih}, b_i$ are needed to be solved by minimizing the loss function.

Researches on the forecasting of maritime freight are always based on temporal features and lack the consideration of spatial information. Therefore, this paper proposes the L-GCN neural network to carry out related prediction tasks from the temporal dimension and the spatial dimension. First, the relevant features of each country or region are normalized through the normalization layer. Then the relevant data is feed into a two-layer GCN structure to complete the spatial feature extraction. After that, the spatial features of each node are transformed according to the node index and the time step. Each node uses the LSTM structure to calculate the relevant weights and share the weights. The final out-puts are the predicted value of the node at time $t+n$. Detailed information on the L-GCN model is shown in figure 5.

**IV. RESULTS AND DISCUSSION**

In this section, we will discuss the accuracy of the ship berthing time calculation algorithm, the accuracy of the node traffic flow calculation algorithm, the superiority of the L-GCN algorithm in maritime and the total performance of the LGCOTFF model.

**A. THE ACCURACY OF SHIP BERTHING CALCULATION ALGORITHM**

In order to verify the accuracy of the calculation results of the berthing time calculation algorithm, we manually check the berthing time and berthing ports of 200 ships through satellite maps. Compared with the calculated berthing time, the results are exactly the same. The calculation results and the check results of 50 ships are shown in Appendix A. In order to see the accuracy of the calculation results intuitively, we intercepted relevant satellite images of 5 ships.

Table 1 shows the real trajectory and the calculated arrival berthing time and the calculated departure berthing time of 5 ships (real trajectories obtained from hifleet.com). We can see that the calculation result is accurate.

**B. THE ACCURACY OF THE NODE TRAFFIC FLOW CALCULATION ALGORITHM**

Singapore is an extremely important and special node in international maritime transportation, as is the case in the crude oil transportation network. Singapore is not a crude oil producing country, but almost all crude oil exported from the Middle East or West Africa East Asia and Southeast Asian countries has to pass through Singapore. Since our goal is to predict the flow of all the nodes in the graph more accurately, if we feed all the ships passing through Singapore into the transportation network, the overall traffic flow of the network will be increased. Therefore, it is very necessary in our model to only calculate how many ships finishing upload or offtake operations. Otherwise, the increased traffic flow will spread to other nodes through this network.

In order to show the difference between the results calculated based on fence data and berth position more clearly, we calculate the number of three main crude oil carriers (VLCC, SUEZMAX and AFRAMAX) in Singapore from the 14th week to the 18th week of 2021 using the port fence data and the berth data. The detailed results are shown in table 2.
From table 2, we can see that the calculated amount of crude oil carriers using the port fence data is far greater than that using the berth data. If we use the port fence data results, the node importance of Singapore will be increased significantly. At the same time, because Singapore is not an oil-producing country, this increased data cannot be eliminated and will cause the overall data in the network to become larger. Through the analysis above, our redesigned traffic flow calculation algorithm and the addition of node types in the crude oil transportation network are very meaningful.
C. THE PERFORMANCE OF L-GCN NEURAL NETWORK

In this part, we will introduce the performance of different methods and discuss the advantages and disadvantages of different methods. In the experiment part, we use the AIS data of all crude oil carriers from January 2020 to July 2021 and collect the node traffic flow volume weekly. We choose China, Russia, Middle and America as our test nodes. We use History Average, Xgboost, LSTM, Time Series, GCN and L-GCN doing a forecasting task respectively. We evaluate the performance based on two popular metrics, i.e., Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) [25].

The History Average and the Time Series only take the time index and ship number as input. For each country we need train a new model. Then totally, we have four different sub models in these two models. The Xgboost and the LSTM take the time index, ship number and country one-hot code as input. The predicted model for all countries can be trained from one model. The GCN takes ship number and country one-hot code as input and it also need a graph structure before training. The L-GCN model takes all the elements needed in other models. Table 3 shows different input and sub-model number for each method.

| Method   | Input Data                          | Sub Model Number |
|----------|-------------------------------------|------------------|
| History Average | [Time index, Ship number]      | 4                |
| Xgboost  | [Time index, Ship number, Country one-hot code] | 1                |
| Time Series         | [Time Index, Ship Number]        | 4                |
| LSTM     | [Time index, Ship number, Country one-hot code] | 1                |
| GCN      | [Graph structure, Ship number, Country one-hot code] | 1                |
| L-GCN    | [Graph structure, Time index, Ship number, Country one-hot code] | 1                |

Table 4 shows the test error comparison of different approaches for forecasting over ten runs. L-GCN achieves the best performance regarding all the metrics on both two nodes, out-performing the second-best baseline by at least 10% in terms of relative error reduction, which suggests the effectiveness of proposed approaches for spatiotemporal correlations modeling.

From table 4, the proposed L-GCN model is the best model for node China, Russia and Middle East. For these three countries, the performance of GCN is similar to that of the time series and LSTM, and these three methods rank in the second place. History Average and Xgboost are not good chose for forecasting crude oil traffic flow, because all the features in the data are homogeneous and these features are not friendly for the Xgboost or other tree methods. This result is similar to that mentioned in Gorishniy’s paper [26].

For America, the weekly crude oil traffic flow is very stable, so the performances of LSTM and L-GCN are almost the same. The History Average also is a good chose for American crude oil traffic flow forecasting.

Figure 6 shows the relationship between the true value of each test node and the predicted value of different methods. Figure 6 (a) shows the predicted trend of different methods for China crude oil ship number. Figure 6 (b), figure 6 (c) and figure 6 (d) shows the predicted trend of different methods for Russia, Middle Est and America, respectively. In each sub-figure the red line is the true value. The orange line is the predicted value of History Average. The yellow line, green line, watchet line and blue line represents Time Series, XGBOOST, LSTM and GCN, respectively. It can be seen from the figure that the predicted values of GCN and L-GCN have similar trends and fit the true curve better. For Russia and the Middle East, the trends obtained by forecasting methods except HA are almost the same. For China and America, the trends predicted by Time Series, XGBOOST, and LSTM are similar, while the trends predicted by GCN and L-GCN are similar. From the experimental results, we also can see that the predicted trend of the LSTM method and the GCN method are always different. This means the temporal features extracted by LSTM and spatial features extracted by GCN have different effects on the prediction.
### D. THE PERFORMANCE OF GCOTFF MODEL

In this part, we will evaluate the overall performance of the GCOTFF model. We will also discuss different results caused by the node value and the network structure changing.

Besides the L-GCN, the ship berthing calculation algorithm and the node traffic flow calculation algorithm also do a lot of favor in the GCOTFF model. The ship berthing calculation algorithm extracts the number of ships finish upload/offtake.
tasks accurately. If we use the fence data in Singapore, the node value and the total ship number in the network will be increased. This increased value will be propagated through the network to the neighboring nodes. The predicted value of China will become larger in our experiment. Detailed information shows in Figure 7(a).

According to data from worldstopexports.com, the United States and the United Kingdom are both the top 15 crude oil importers and the top 15 crude oil exporters in the world in 2020. The United States crude oil imports reached 81.6 billion dollars and exports reach 50.3 billion dollars in 2020. The United Kingdom crude oil imports reach 15.6 billion dollars and exports reach 16.1 billion dollars in 2020. The detailed data of top crude oil import and export countries are shown in Appendix B and Appendix C. So, it is unfair to treat these two countries as dependent on oil import-dependent country or oil export-dependent country in the transportation network as researchers did. If we treat the United States as an import-dependent country, the role the United States plays in the crude oil transportation network will be the same as Europe plays. The node value of the United States itself won’t change a lot because of the existence of self-loop. However, the connected node such as China will be decreased. Detailed information shows in Figure 7(b). From Figure 7, we can see that the RMSE of predicted value will become larger after using fence data or treating oil-balance countries as import-dependent countries. So, node status calculation and the graph structure establishment are two important parts in the LGCOTFF model. These two parts make a great contribution to reduce the overall error of the model.

V. CONCLUSION

The main goal of this paper is to propose a more accurate global maritime crude oil transportation forecasting model. In the process of trying to build the model, we find that there are several points that can be improved based on the previous researches of scholars. One is the calculation of the node traffic flow and the construction of the global crude oil transportation network. The other is a prediction model considering both the temporal dimension and the spatial dimension information. Our research results can be summarized as follows.

(i) According to the AIS data and the berth data, we propose a new method to calculate the number of ship operations. Compared with the existing methods, this method is more accurate in the calculation of the traffic flow value at some important transportation nodes. It can effectively reduce the error of traffic flow caused by supplying or refueling.

(ii) In the crude oil maritime transportation network, according to the actual situation, we divide the major nodes with both import and export or spot trading into one category besides the import-dependent country and the export-dependent country. As a result, the constructed crude oil maritime transportation network is closer to the actual situation.

(iii) We propose a crude oil network node traffic prediction model named L-GCN. When evaluated on two major nodes, the proposed approach achieved significantly better results than state-of-the-art baselines.

However, there are avenues for further research and investigation:

(i) Our prediction approach is a model based on a static graph, which means that the structure change of the graph is not taken into account in the prediction process. In the future, we can apply muti-graph in our model to improve this problem.

(ii) For the node features in the transportation network, we only take the traffic volume into consideration. There are no macroeconomic features for each node. In the future, related factors can be added to the model. Thus, we can build a connection between the economy and crude oil transportation to make our research more meaningful.

APPENDIX

See Tables 5–7.

| Rank | Country (Region) | Dollar Value (billion) | Proportion in The World (%) |
|------|------------------|------------------------|----------------------------|
| 1    | China            | 176.3                  | 25.8                       |
| 2    | United States    | 81.6                   | 12.0                       |
| 3    | India            | 64.6                   | 9.5                        |
| 4    | South Korea      | 44.5                   | 6.5                        |
| 5    | Japan            | 43.5                   | 6.4                        |
| 6    | Germany          | 27.4                   | 4.0                        |
| 7    | Netherlands      | 22                     | 3.2                        |
| 8    | Spain            | 18.2                   | 2.7                        |
| 9    | Thailand         | 17.6                   | 2.6                        |
| 10   | Italy            | 16.2                   | 2.4                        |
| 11   | United Kingdom   | 15.6                   | 2.3                        |
| 12   | Singapore        | 14.4                   | 2.1                        |
| 13   | Taiwan (China)   | 12.6                   | 1.8                        |
| 14   | France           | 12                     | 1.8                        |
| 15   | Belgium          | 9.8                    | 1.4                        |

| Rank | Country (Region) | Dollar Value (billion) | Proportion in The World (%) |
|------|------------------|------------------------|----------------------------|
| 1    | Saudi Arabia     | 113.7                  | 17.2                       |
| 2    | Russia           | 72.6                   | 11.0                       |
| 3    | Iraq             | 50.8                   | 7.7                        |
| 4    | United States    | 50.3                   | 7.6                        |
| 5    | United Arab      | 47.9                   | 7.2                        |
| 6    | Emirates         | 47.6                   | 7.2                        |
| 7    | Kuwait           | 28.3                   | 4.3                        |
| 8    | Nigeria          | 25.2                   | 3.8                        |
| 9    | Kazakhstan       | 23.7                   | 3.6                        |
| 10   | Norway           | 22.7                   | 3.4                        |
| 11   | Angola           | 20.2                   | 3.1                        |
| 12   | Brazil           | 19.6                   | 3.0                        |
| 13   | United Kingdom   | 16.1                   | 2.4                        |
| 14   | Oman             | 15                     | 2.3                        |
| 15   | Mexico           | 14.9                   | 2.2                        |
REFERENCES

[1] Z. Song, Q. Zhu, and M. Han, “Tele-connection of global crude oil network: Comparisons between direct trade and embodied flows,” Energy, vol. 217, Feb. 2021, Art. no. 119359.

[2] M. Wang, L. Tian, and R. Du, “Research on the interaction patterns among the global crude oil import dependency countries: A complex network approach,” Appl. Energy, vol. 180, pp. 779–791, Oct. 2016.

[3] X. Xi, J. Zhou, X. Gao, D. Liu, H. Zheng, and Q. Sun, “Impact of changes in crude oil trade network patterns on national economy,” Energy Econ., vol. 84, Oct. 2019, Art. no. 104490.

[4] H. Yu, Z. Fang, F. Lu, A. T. Murray, H. Zhang, P. Peng, Q. Mei, and J. Chen, “Impact of oil price fluctuations on tanker maritime network structure and traffic flow changes,” Appl. Energy, vol. 237, pp. 390–403, Mar. 2019.

[5] L. Kang, Q. Meng, and Q. Liu, “Fundamental diagram of ship traffic in the Singapore strait,” Ocean Eng., vol. 147, pp. 340–354, Jan. 2018.

[6] L. Zhang, Q. Meng, and T. Fang Fwa, “Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters,” Transp. Res. E, Logistics Transp. Rev., vol. 129, pp. 287–304, Sep. 2019.

[7] B. Gunnarsson, “Recent ship traffic and developing shipping trends on the Northern sea route—Policy implications for future Arctic shipping,” Mar. Policy, vol. 124, Feb. 2021, Art. no. 104369.

[8] E. Merico, A. Gambaro, A. Argirizou, A. Alebic-Juretic, E. Barbaro, D. Cescar, L. Chasapidis, S. Dimopoulos, A. Dinoi, A. Donateo, and C. Giannaros, “Atmospheric impact of ship traffic in four adriatic-ionian port-cities: Comparison and harmonization of different approaches,” Transp. Res. D, Transp. Environ., vol. 50, pp. 431–445, Jan. 2017.

[9] L. Qi, Z. Zheng, and L. Gang, “A cellular automaton model for ship traffic flow in waterways,” Phys. A, Stat. Mech. Appl., vol. 471, pp. 705–717, Apr. 2017.

[10] V. Gosasang, W. Chandraprakaikul, and S. Kiattisin, “A comparison of traditional and neural networks forecasting techniques for container throughput at Bangkok port,” Asian J. Shipping Logistics Transp. Rev., vol. 27, no. 3, pp. 463–482, Dec. 2011.

[11] J. Ke, H. Zheng, H. Yang, and X. M. Chen, “Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach,” J. Transp. Res. C, Emerg. Technol., vol. 85, pp. 591–608, Dec. 2017.

[12] Q. Ouyang, Y. Lv, J. Ma, and J. Li, “An LSTM-based method considering history and real-time data for passenger flow prediction,” Appl. Sci., vol. 10, no. 11, p. 3788, May 2020.

[13] J. Zhang, Y. Zheng, and D. Qi, “Deep spatio-temporal residual networks for citywide crowd flows prediction,” in Proc. 31st AAAI Conf. Artif. Intell., 2017, pp. 1655–1661.

[14] M. Intihar, T. Kramberger, and D. Dragan, “Container throughput forecasting using dynamic factor analysis and ARIMAX model,” PROMET Traffic Transp., vol. 29, no. 5, pp. 529–542, Nov. 2017.

[15] M. Eskafi, M. Kowsari, A. Dastgheib, G. F. Ulfarsson, G. Stefansson, P. Taneja, and R. I. Thorarinsdottir, “A model for port throughput forecasting using Bayesian estimation,” Maritime Econ. Logistics, vol. 23, no. 2, pp. 348–368, Jun. 2021.

[16] L. Mo, L. Xie, X. Jiang, G. Teng, L. Xu, and J. Xiao, “GMDH-based hybrid model for container throughput forecasting: Selective combination forecasting in nonlinear subspecies,” Appl. Soft Comput., vol. 62, pp. 478–490, Jan. 2018.

[17] P. Du, J. Wang, W. Yang, and T. Niu, “Container throughput forecasting using a novel hybrid learning method with error correction strategy,” Knowl.-Based Syst., vol. 182, Oct. 2019, Art. no. 104853.

[18] Y. Rashid, H. Meersman, C. Sys, E. Van De Voorde, and T. Vanneste, “A combined approach to forecast container throughput demand: Scenario analysis for the harbours of the destination ports,” Transp. Res. A, Policy Pract., vol. 117, pp. 127–141, Nov. 2018.

[19] G. Xie, N. Zhang, and S. Wang, “Data characteristic analysis and model selection for container throughput forecasting within a decomposition-ensemble methodology,” Transp. Res. E, Logistics Transp. Rev., vol. 108, pp. 160–178, Dec. 2017.

[20] B. Bai, L. Yao, C. Li, X. Wang, and C. Wang, “Adaptive graph convolutional recurrent network for traffic forecasting,” 2020, arXiv:2007.02842.

[21] B. Yu, Y. Lee, and K. Sohn, “Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN),” Transp. Res. C, Emerg. Technol., vol. 114, pp. 189–204, May 2020.

[22] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, “T-GCN: A temporal graph convolutional network for traffic prediction,” IEEE Trans. Intell. Transp. Syst., vol. 21, no. 9, pp. 3848–3858, Sep. 2019.

[23] T. N. F. Kip and M. Welling, “Semi-supervised classification with graph convolutional networks,” 2016, arXiv:1609.02907.

[24] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[25] J. S. Armstrong and F. Collopy, “Error measures for generalizing about forecasting methods: Empirical comparisons,” Int. J. Forecasting, vol. 8, no. 1, pp. 69–80, Jun. 1992.

[26] Y. Gorishniy, I. Rubacev, V. Khrulkov, and A. Babenko, “Revisiting deep learning models for tabular data,” 2021, arXiv:2106.11959.

QI OUYANG was born in Qiqihaer, Heilongjiang, China in 1990. He received the Ph.D. degree from Beijing Jiaotong University, in 2020. He continues his researches as a Postdoctoral Researcher with the China Transport Telecommunications and Information Center. He published two SCI papers, “Passenger Travel Regularity Analysis Based on A-Large-Scale Smart Card Data” and “An LSTM-Based Method Considering History and Real-Time Data for Passenger Flow Prediction,” as the first author. His main research interest includes maritime big data analysis.

TENGA SUN received the Ph.D. degree. He is currently a Professor-Level Senior Engineer, a General Manager of the Data Application Division, China Transport Telecommunications and Information Center, the Deputy Director of the Transportation Beidou Satellite Navigation System Application Development Center, and the Head of the National Transportation Logistics Information Platform. He has been engaged in the research, construction, and management of informatization in the transportation industry for a long time, and has rich experience in informatization construction and management. He has presided over and participated in a number of national and ministerial research projects, published more than 30 articles, two monographs, and three invention patents in important domestic and international core journals.

YUANYUAN XUE was born in Xining, Qinghai, China, in 1974. She received the Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University, in 2017. She is currently an Associate Professor at the China Transport Telecommunications and Information Center. Her research publications have appeared in several academic journals (SCI) and top conferences, such as Computer Systems Science and Engineering journal, Journal of Biomedical and Health Informatics, Conference on Pervasive & Ubiquitous Computing, and Conference on Extending Database Technology. Her research interests include data analysis on water transportation, data mining, and sentiment analysis in social networks.

ZHEHUI LIU received the master’s degree from the State Key Laboratory of Resource and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, in 2010. She currently works as a Senior Engineer with the China Transport Telecommunications and Information Center. Her research interests include waterborne transportation information technology and GIS application.