Graph Database Schema for Multimodal Transportation in Semarang

Panji Wisnu Wirawan[1] *, **Djalal Er Riyanto [2]**, **Dinar Mutiara Kusumo Nugraheni [3]** , **Yasmin[4]**

[1)2)3)4)]*Universitas Diponegoro, Indonesia*

Jl. Prof Soedarto SH, Semarang

[1)]panji@if.undip.ac.id, [2)]errykom@gmail.com, [3)]dinar.mutiara@gmail.com, [4)]yasmin9658@gmail.com

Abstract

Background: Semarang has broad area that cannot be covered entirely by single transportation mode. To reach a specific location, people often use more than one public transportation mode. Apart from Bus Rapid Transit, another exist namely angkot or city transportation. Multimodal traveler information is then required to help passenger searching for a route. Several studies of multimodal traveler information system has been conducted, however the data model for multimodal transportation did not conceived in detail.

Objective: Proposes a database of multimodal transportation design using graph data model by taking Semarang as a case study.

Method: We create our model in oriented entity-relationship diagram (O-ERD) and map this O-ERD to the graph database schema.

Result: We develop our data model in graph database schema and we implement the model using Neo4J graph database for validation purpose. Our model consist of three graph node label namely Shelter, Angkot Stopper, and Closer Place. To validate our model, we execute a search query using the Cypher query to look for location with closer place to it.

Conclusion: Our data model was successfully developed and implemented. Searching transportation route in the implementation of our model has been conducted using cypher query. It can successfully display all possible paths and routes. Our query can distinguish between one mode of transportation with another.

I. INTRODUCTION

Semarang is the capital of Central Java Province with an area of 373,78 km^2. In 2015 the population of Semarang was 1,622,520 people. With a population of more than one million, Semarang is categorized as a metropolitan city [1]. As one of the big cities, Semarang has characteristics that are not different from some other major cities in Indonesia in terms of the city's traffic movements. The aspects of the city's traffic movements are busy and even tend to jam in peak hours both morning and evening. The high growth of vehicles operating on roads is generally dominated by the increase in private vehicles both cars and motorbikes [2].

The government has made various efforts to regulate traffic congestion, one of which is providing BRT (Bus Rapid Transit) [3]. Apart from BRT, there are other modes of transportation, including City Transportation or Angkot. Angkot is special transportation modes in Semarang, because it can be dropped the passanger off, anywhere along the street.

One transportation mode can not cover all of Semarang area. People have to use more than one mode of public transportation to reach a location known as multimodal. The use of several modes of transportation makes people have to choose the right method to get to the destination. Information regarding the transfer of transportation modes is needed so that people can drive public transportation to their intended destination. However, currently, there is a lack of information about multimodal transportation in Semarang. Therefore, an application or information system is needed to make it easier for people to search what public transportation modes can be used to reach certain location.

Previous studies proposed several multimodal transportation information system. Zhang et.al.[4] and Dib et.al. [5] proposed network model for Advance Traveler Information System (ATIS). Zhang et.al proposed model for public

* Corresponding author

and private modes and algorithm for multimodal transport model while Dib et.al proposed optmization algorithm for multicriteria routing algorithm. TIMIPlan, was proposed by Garcia et.al to solve uni-modal and multimodal transportation problem in Spain [6]. From the application development perspective, Garcia et.al. illustrated TIMIPlan architecture. Another work by Borole et al. also shows application architecture only [7]. It is an architecture for Real Time Transit Trip. However, none of the works shows the requiremnt for multimodal transportation data model.

A multimodal transportation information system requires storage media points that form a route. There are at least two options for storing the data, that is using a relational database and a non-relational or NoSQL database. NoSQL itself consists of several choices such as key-value, document, families and graph [8]. Several studies of the NoSQL database related to transportation have been conducted. Vela et al. use a document-oriented database to store and retrieve public transport routes [9]. Later, they designed graph database to store path which is generated from the mobile application [10]. Both of these studies were using the entity-relationship diagram for the basis of database modeling.

Previous studies show that graph database have better performance when searching for transportation routes compared to relational databases [11]. Thus, the graph database can be an alternative solution for storing multimodal transportation network data. Graph database is useful for storing relationship between data such as traversing in social network and deciding recommendation [12]. The discovery of transportation networks can take advantage of graph traversing operations such as finding neighborhoods, traversal transportation routes, and finding the shortest path.

Based on these descriptions, this paper proposes a database design for multimodal transportation using graph data model by taking Semarang as a case study. The model is then implemented on a Neo4j graph database, [13]. This model will be beneficial for helping multimodal transport implementations into the graph database, so that it can search multimodal transportation quickly, particulary for multimodal traveler information systems. From the application development perspective, this paper tries to fill the gap of the absence of data model in multimodal traveler information systems earlier studies.

II.   LITERATURE REVIEW

A.   Transportation

The definition of transportation comes from the Latin word, namely transportare , where trans means the other or the other side and portare means transporting or carrying. So, transportation means transporting or carrying (something) to the other side or somewhere to another place. Transportation can be defined as a business and activities transporting or carrying goods and / or passengers from one place to another . Transportation is an important element and serves as the lifeblood and economic, social, political and mobility development of the population that grows together and follows developments in various fields and sectors [14] . Transportation development is very important in supporting and driving the dynamics of development. Transportation has a strategic function in attaching territorial integrity and functioning as a catalyst in supporting economic growth and regional development [1] .

Multimodal transportation is the use of several modes of transportation to move passengers or goods from one place of origin to the destination. Passengers or goods are moved from one point to another using two or more modes of transportation. In addition to the use of different modes of transportation, walking and cycling can also be combined in multimodal transportation [15].

Several studies of multimodal transportation in Computer Science have been conducted including the shortest path problem [15] and multimodal transportation data models [16] [17][18]. Bieli et.al [16] models multimodal transportation networks using objects. Bieli proposes a framework that uses an algorithmic approach to solve shortest path problems and provide multimodal network modeling. Unlike Bieli, Booth et.al. [17] use the model graph for multimodal and relational transportation and several operations needed to implement the graph model that has been formed. In particular, Lopez et al. reviewed several network model for multimodal transportation network [19]. In his work, Lopez identfy several network models for multimodal shortest path transport network such as time dependent graph and hierarchical structure graph. This paper focuses on the data graph model as graph schema that will be implemented in the graph database.

B.   Graph Database and property graph

Graph database is a database that uses a graph structure to store data [20]. Graph databases provide facilities for creating, retrieving, updating and deleting (CRUD) plus other facilities such as indexing, searching, query, and path traversal. This chosen traversal path is the advantage of the graph database. Examples of graph databases are Neo4J,

InfoGrid, and Infinite Graph . To help implement the database graph, a graph data model is needed as a bridge between data needs analysis and implementation.

The graph data model is a model where the data structure for the schema and instance is modeled as a directed graph, or generalization of the data graph structure, where data manipulation is expressed by graph-oriented operations. Data graph modeling is applied in areas where information about the interconnectivity of data or topology is more important, or as important as the data itself. Graph is a modeling tool that has several advantages for this data type [21] . Data graph modeling allows more natural data modeling. Graph structure can be seen and allows natural ways to handle application data, such as hypertext data or geographic data. Graphs have the advantage of being able to store all information about an entity on one node and show information related to the edge connected to it.

Graph data models exists in common usage, including property graph, hypergraphs, dan triplets [22]. Our proposed model use property graph as a graph data model for modelling eficiency purpose. Property graph is a ' multigraph' where nodes and edges are labeled with data, which is a key-value pair. An example of a graph database can be seen in Fig. 1. The image represents information storage about blogs that have users as admin and followers. Node n 1 and n 2 represents a user and a blog. Both have id and name as properties . Edges between n 1 and n 2 representing relationships followers and admin with properties that specify these relationships [23].
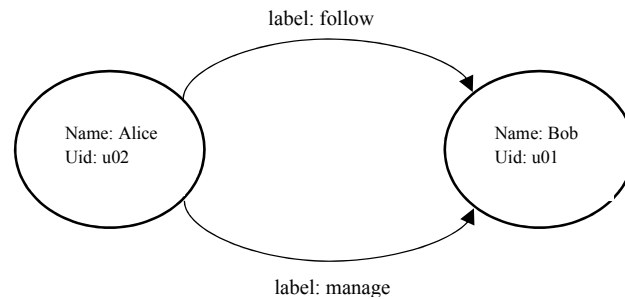

Fig. 1 Property graph example

III. METHOD

A model is a depiction of an actuality [24]. This paper proposes a graph model, in a schema form, for multimodal transportation in Semarang. We follow the method proposed by De Virgilio [23] to develop our graph database schema. This method begins with forming an Oriented Entity Relationship Diagram (O-ERD). Once arranged, OERD is weighted in the following rules:

1. One-to-one relations are transformed into edges that have two directions (double directed) $\omega (e) = 0$.
2. One-to-many relationships are transformed into one-way edges, from low to high-marginalized entities $\omega (e) = 1$.
3. Many-to-many relations are transformed into edges that have two directions (double directed) with different weights with one-to-one relations, namely $\omega (e) = 2$.

From the weights that has been formed, then nodes and edge graph databases are determined.

IV. RESULT

The formation of graphs in this study begins by forming nodes and edges of the data obtained. To form nodes and edges, first formed oriented entity relationship diagram (O- ERD) first. There are four entities in this study, namely Shelter, Angkot Stopper, Closer Place and Road. The Shelter Entity and Angkot Stopper have a relationship called " connect". While both entities are the same - they have a relationship with the entity Closer Place with the name "close_to" relation. Stopper Shelter entities and public transportation is also the same - each has a relationship with the entity Road called "it's_in". The O-ERD of this study can be seen in Fig. 2.

The O-ERD that has been formed is then calculated for the weight of each entity. There are seven nodes (entities) in the O-ERD. The seven nodes will be counted outcoming and incoming weights . The calculation of node weights is described below:

1) Closer Place

$$w^{+}(\text{Closer Place}) = w(\text{close-to}) + w(\text{close\_to}) \tag{1}$$

$$w^{-}(\text{Closer Place}) = w(\text{close\_to}) + w(\text{close\_to}) \tag{2}$$

Because the weight of the outcoming of the node Closer Place is worth 4 and the incoming weight is from the node Closer Place is worth 4 so that the node has w - (n)> 1 and w + (n) ≥ 1, then the Closer Place node becomes a stand-alone node .

2) Shelter

$$w^+(\text{Shelter}) = w(\text{connect}) + w(\text{close\_to}) \quad (3)$$
$$w^-(\text{Shelter}) = w(\text{connect}) + w(\text{close\_to}) + w(\text{it's\_in}) \quad (4)$$

Due to the weight of the node Shelter outcoming value of 4 and the weight of the node Shelter incoming worth 5 so that these nodes have w - (n)> 1 and w + (n) ≥ 1, then the node shelter into a stand-alone node.

3) Angkot Stopper

$$w^+(\text{ Angkot Stopper}) = w(\text{it's\_in}) + w(\text{connect}) + w(\text{close\_to}) \quad (5)$$
$$w^-(\text{ Angkot Stopper}) = w(\text{it's\_in}) + w(\text{connect}) + w(\text{close\_to}) \quad (6)$$
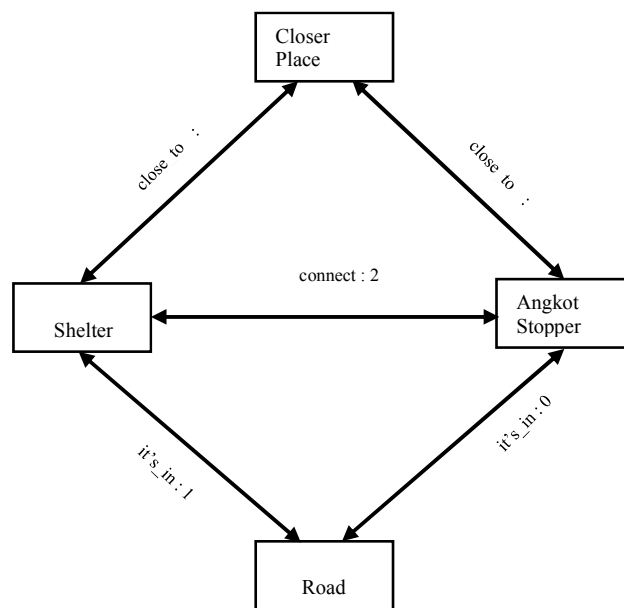

Fig. 2 O-ERD for multimodal transportation

Because the weight of the outcoming of the node Stopper Angkot is 4 and the incoming weight is from the node Stopper Angkot is 4 so that the node has w - (n)> 1 and w + (n) ≥ 1, then the node Angkot Stopper becomes a stand-alone node.

4) Road

$$w^+(\text{ Road}) = w(\text{it's\_in}) + w(\text{it's\_in}) \quad (7)$$
$$w^-(\text{ Road}) = w(\text{it's\_in}) \quad (8)$$

Because the weight of the outcoming of the node Road is worth 1 and the incoming weight is from the node Road is 0 so that the node has w - (n) ≤ 1 and w + (n) ≤ 1, then the node Road joins the node connected to it. Node Road joins the Shelter node and the Stopper Angkot node .

After all nodes and group nodes were formed, we build a graph database schema. After the weighting is complete, it can be seen that the O-ERD of the Semarang multimodal transportation model has 3 node from the previous four nodes. This happens because the road nodes must join the Shelter nodes and nodes Angkot Stopper . This means all the properties of the node Road also attached to or merged with nodes and node Shelter Angkot Stopper . The final result, graph database schema, is shown in Fig. 3.
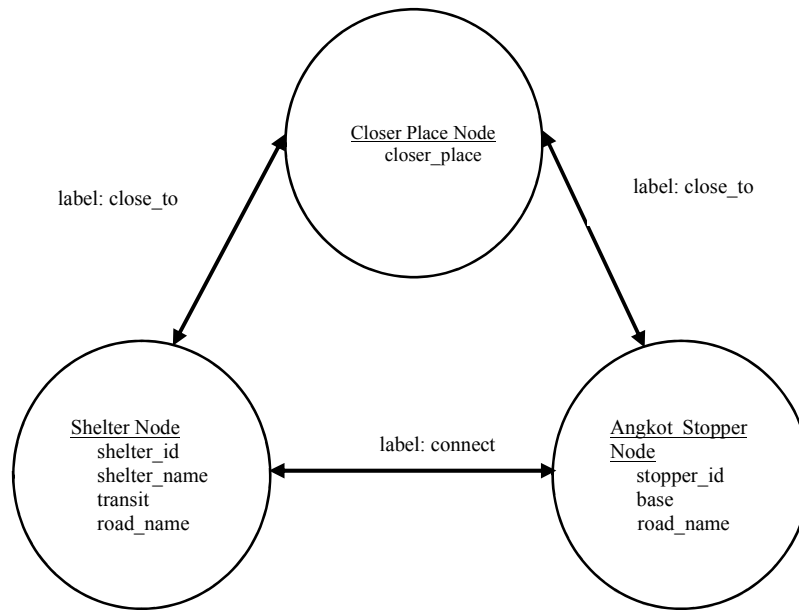
Fig. 3 Graph database schema for multimodal transportation.

Graph databases are implemented in Neo4j version 3.2.3 and we use Cypher query to implement our model in Fig 3. Fig. 4 is a query used to create a Closer Place node. Closer place node is formed first, because Closer Place can only exist once. So when adding a shelter or the previous Public Stopper will be checked whether the entered Closer Place has ever existed before or not. If there has been a node will be called with the MATCH operation but if there is no node Closer Place will be created first with the CREATE operation.

```
CREATE (a:closer_place{closer_place:'name of place'})
```
Fig. 4. Cypher query to create closer place node.

Fig. 5 is a query to build relationships with the label "connected" to the BRT corridor property and its path. When building this relationship has also been built shelters node connected to the corridor together at once connected with the corresponding Closer Place.

```
MATCH (a:closer_place)
WHERE a.closer.place='name of place'
CREATE   (b:shelter{shelter_id:'shelter   ID',   shelter_name:'shelter
name',road_name:'road name', transit: 'transition status'}), (c:shelter{
id_shelter:'shelter   ID',shelter_name:'shelter   name',road_name:'road
name',  transit:  'transition  status'}),  (b)-[:close_to]->(a),  (b)-
[:connect{koridor:'number of route', route:'name of route'}]->(c)
```
Fig. 5. Cypher query to form a relationship connected with the corridor property.

Fig. 6 is a query to build a relationship with the label "connected" with id_angkot property along its path. When building this relationship, built as well as public transportation nodes connected to id_angkot Stopper simultaneously.

```
MATCH (a:closer_place)
WHERE a.closer_place='name of place'
CREATE  (b:angkot{stopper_id:'stopper  ID',road_name:'name  of  route',
transit:'transition    status'}),    (c:angkot{    stopper_id:'stopper
ID',road_name:'name  of  road',  transit:  'transition  status'}),  (b)-
[:close_to]->(a),  (b)-[:connect{angkot_id:'angkot  ID',  route:'name  of
route'}]->(c)
```
Fig. 6. Cypher query to form relationship.

The Cypher query used for finding multimodal route shown in Fig. 7. We use the road as a reference to find routes. These routes are then aggregated into a single list using the collect function [25], so it can be processed later. This function will save effort to process data because it can minimize loops.

```
MATCH (a)-[:close_to]->(c:closer_place), (b)-[:close_to]->(d:closer_place),
p=(a)-[r:connect*]->(b)
WHERE a.road_name='origin_road' AND b.road_name='destination_road'  AND
c.closer_place='closer_place_in_origin' AND
d.closer_place='closer_place_in_destination'
RETURN collect(p) as path,(a),(b)
```
Fig. 7. Cypher query to search multimodal transportation.

The results of the search query can be seen in Fig. 8. This figure is taken from Neo4j browser. The blue nodes represent nodes shelter while red represents stopper public transportation. As an alternative, the query result can be represented as JavaScript Object Notation (JSON) to be consumed in application [26].
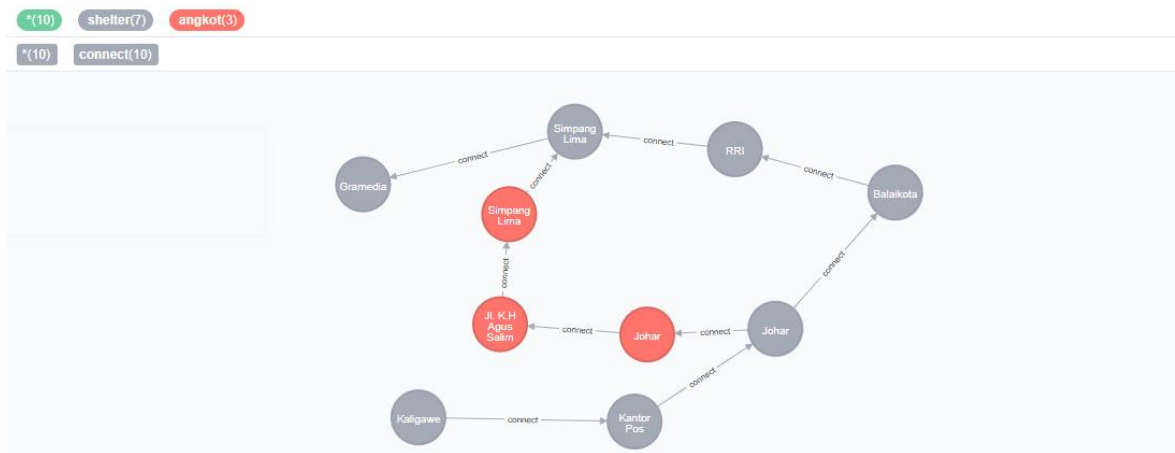

Fig. 8. Query result in Neo4j browser.

V. DISCUSSION

We have created a data model using a graph database schema for multimodal transportation in Semarang. The modes of transportation in the model are all public transportation. The formation of nodes always starts with a closer place node because the node is an independent node compared to BRT shelter or angkot stopper. Angkot stopper is a unique characteristic of this model because it models where prospective passengers can stop angkot along the road being passed, because on Semarang angkot can stop anywhere on the road it passes.

```
MATCH (a)-[:close_to]->(c:closer_place), (b)-[:close_to]->(d:closer_place),
p=(a)-[r:connect*]->(b)
WHERE a.road_name='Jl. Kaligawe' AND b.road_name='Jl. Pandanaran'  AND
c.closer_place='Kaligawe' AND d.closer_place='Gramedia'
RETURN collect(p) as path,(a),(b)
```
Fig. 9. Example of cypher query to search multimodal transportation.

The search process begins by determining the nearest place and street name, where the passenger is and the final destination of the passenger. The final destination is not only the name of the road, but is accompanied by the closest place to the final destination. The query for that purpose is shown in Fig. 7. The query in Fig. 7 still uses the collect function because route search is done for all possible routes, indicated by using connect * on kalusa MATCH.

An example of a query result is shown in Fig. 8, it is the result of a query to search a route from Kaligawe road to Pandanaran road where Gramedia is the closest place to the destination. The query representation is shown in Fig. 9.

The results of query execution in Fig. 9 are shown in Fig. 8, where there are two choices of multimodal transportation uses, namely BRT and angkot. The first route (gray node) is the route using BRT that passes through the Kaligawe shelter → Kantor Pos → Johar → Balaikota → RRI → Simpang Lima → Gramedia. The second route is a combination of the use of BRT and angkot which are indicated by the red color on the node. The route taken is longer, namely: Kaligawe → Kantor Pos → Johar -> (move away with angkot in Johar-> Jl KH Agus Salim -> Simpang Lima -> (move away with BRT at Simpang Lima -> Gramedia. Thus, the query in Fig. 8 can determine the mode of transportation and the route that is passed for each mode involved. The execution of query in Fig 8 verifies that our model can adopt two modes of transportation with different characteristics.

The model we have developed can adopt entity for the closest place that public transport passengers want to go to. The use of the nearest place can help passengers who want to go to a certain place while he only knows the name of the road. Cypher query in Fig. 9 exclude the nearest place node. The closer path in a road can be requested by adding the text "(c)" and "(d)" in the end of query, as the representation of closer_place label in both origin and destination. The query for requesting closest place depicted in Fig. 10.

```
MATCH  (a)-[:close_to]->(c:closer_place),  (b)-[:close_to]->(d:closer_place),
p=(a)-[r:connect*]->(b)
WHERE  a.road_name='Jl.  Kaligawe'  AND  b.road_name='Jl.  Pandanaran'  AND
c.closer_place='Kaligawe' AND d.closer_place='Gramedia'
RETURN collect(p) as path,(a),(b),(c),(d)
```

Fig. 10. Query for searching multimodal transportation and closer path in origin and destination .

Thus, our proposed model is differ from the model developed by Vela et.al.[10]. In addition to developing a model for the transportation route, the model developed by Vela et.al. adopts a 'special need', which is additional information about the route such as photos, comments, and so on. Information about the closest place to destination is not presented in the model.

Previous work shows that building a multimodal transport model is accompanied by the creation of an algorithm for route search especially for the shortest path route [19]. The model that we have created, can be implemented directly to the graph database, therefore we can rely on the searching algorithm to graph the database management system. This is very beneficial in terms of application development because application developers can only think about how to arrange search queries and retrieve the results, rather than developing searching algorithm.

VI. CONCLUSION

This paper has successfully developed graph data model for multimodal transportation. Our model consist of three labels namely shelter, angkot stopper, and closer place. They are the representation of BRT shelter, city transportation and closer place near them. Our angkot stopper node is a unique node because the node models an angkot that can stop anywhere on the road it passes. Our model has been implemented in Neo4j graph database, where nodes and relationships are created using Cypher query.

We also have validated our model by querying our implementation to search transportation route by the guidance of the closer place near passanger. One of the important functions of the search query is the collect function, which can make search results better in forming the path. Unlike previous studies, the searching for routing algorithm depend on graph database management system itself. Consequently, it is fairly easy for building multimodal transportation with the help of our model because all route generated from graph database system.

REFERENCES

[1]   I. Y. Hidayati and S. Febriharjati, "Pengembangan Transportasi Berkelanjutan Di Kota Semarang," *J. Riptek*, vol. 10, no. 1, pp. 43–56, 2016.

[2]   D. Purwanto and E. E. Yulipriyono, "Efektifitas Pemberlakuan Sistem Satu Arah pada Jalan Indraprasta Kota Semarang," *J. MKTS*, vol. 21, no. 1, pp. 47–55, 2015.

[3]   Pemerintah Kota Semarang, "Kadishub Khadik Himbau Masyarakat Manfaatkan Transportasi Massal Yang Murah Dan Nyaman," 2017. [Online]. Available: http://semarangkota.go.id/berita/read/7/berita-kota/1972/kadishub-khadik-himbau-masyarakat-manfaatkan-transportasi-massal-yang-murah-dan-nyaman. [Accessed: 23-Apr-2018].

[4]   J. Zhang, F. Liao, T. Arentze, and H. Timmermansa, "A multimodal transport network model for advanced traveler information systems," in *Procedia Computer Science*, 2011, vol. 5, pp. 912–919.

[5]   O. Dib, M. A. Manier, L. Moalic, and A. Caminada, "A multimodal transport network model and efficient algorithms for building advanced traveler information systems," *Transp. Res. Procedia*, vol. 22, no. 2016, pp. 134–143, 2017.

[6]   J. García, A. Torralba, Á. García-Olaya, J. E. Flórez, and D. Borrajo, "Solving multi-modal and uni-modal transportation problems through TIMIPlan," *IFAC Proc. Vol.*, vol. 45, no. 24, pp. 203–208, 2012.

[7]   N. Borole, D. Rout, N. Goel, P. Vedagiri, and T. V. Mathew, "Multimodal Public Transit Trip Planner with Real-time Transit Data," *Procedia - Soc. Behav. Sci.*, vol. 104, pp. 775–784, Dec. 2013.

[8]   P. J. Sadalage and M. Fowler, *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. Upper Saddle River, NJ: Addison-Wesley, 2013.

[9]   B. Vela, J. M. Cavero, P. Caceres, A. Sierra, and C. E. Cuesta, "Defining a NoSQL Document Database of Accessible Transport Routes," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2017, pp. 1125–1129.

[10]   B. Vela, J. M. Cavero, P. Cáceres, A. Sierra, and C. E. Cuesta, "Using a NoSQL graph oriented database to store accessible transport routes," in *Proceedings of the Workshops of the EDBT/ICDT 2018 Joint Conference*, 2018, pp. 62–66.

[11]   P. W. Wirawan, D. E. Riyanto, and K. Khadijah, "PEMODELAN GRAPH DATABASE UNTUK MODA TRANSPORTASI BUS RAPID TRANSIT," *J. Inform.*, vol. 10, no. 2, Jul. 2016.

[12]   A. Ribeiro, A. Silva, and A. R. da Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.

[13]   Neo4J, "Neo4J : The World's Leading Graph Database." [Online]. Available: http://neo4j.com. [Accessed: 01-Jan-2016].

[14]   A. Kadir, "Transportasi: Peran Dan Dampaknya Dalam Pertumbuhan Ekonomi Nasional," *J. Perenc. Pengemb. Wil. Wahana Hijau*, vol. 1, no. 3, pp. 121–131, 2006.

[15]   A. Idri, M. Oukarfi, A. Boulmakoul, K. Zeitouni, and A. Masri, "A new time-dependent shortest path algorithm for multimodal transportation network," *Procedia Comput. Sci.*, vol. 109, pp. 692–697, 2017.

[16]   M. Bielli, A. Boulmakoul, and H. Mouncif, "Object modeling and path computation for multimodal travel systems," *Eur. J. Oper. Res.*, vol. 175, no. 3, pp. 1705–1730, Dec. 2006.

[17]   J. Booth, P. Sistla, O. Wolfson, and I. F. Cruz, "A data model for trip planning in multimodal transportation systems," in *Proceedings of the 12th International Conference on Extending Database Technology Advances in Database Technology - EDBT '09*, 2009, p. 994.

[18]   I. Kabashkin, "Modelling of Regional Transit Multimodal Transport Accessibility with Petri Net Simulation," *Procedia Comput. Sci.*, vol. 77, pp. 151–157, 2015.

[19]   D. López and A. Lozano, "Techniques in multimodal shortest path in public transport systems," *Transp. Res. Procedia*, vol. 3, no. July, pp. 886–894, 2014.

[20]   J. Powell and M. Hopkins, "Graph databases and how to use them," in *A Librarian's Guide to Graphs, Data and the Semantic Web*, Elsevier, 2015, pp. 197–207.

[21]   R. Angles and C. Gutierrez, "Survey of graph database models," *ACM Comput. Surv.*, vol. 40, no. 1, pp. 1–39, 2008.

[22]   I. Robinson, J. Webber, and E. Eifrem, *Graph Databases*, 2nd ed. O'Reilly Media Inc., 2015.

[23]   R. De Virgilio, A. Maccioni, and R. Torlone, "Model-Driven Design of Graph Databases," 2014, pp. 172–185.

[24]   M. A. Rodriguez and P. Neubauer, "Constructions from dots and lines," *Bull. Am. Soc. Inf. Sci. Technol.*, vol. 36, no. 6, pp. 35–41, Aug. 2010.

[25]   C. Kemper, *Beginning Neo4j*. Apress, 2016.

[26]   A. Goel, *Neo4j Cookbook*. Packt Publishing, 2015.