Article

An Area Partitioning and Subgraph Growing (APSG) Approach to the Conflation of Road Networks

Hoa-Hung Nguyen [1] and Han-You Jeong [2,*]

[1] Department of Electrical and Electronics Engineering, Pusan National University, Republic of Korea; nguyenhoahungit@gmail.com
[2] Department of Electrical Engineering, Pusan National University, Republic of Korea; hyjeong@pusan.ac.kr
* Correspondence: hyjeong@pusan.ac.kr

Abstract: A road network represents a set of road objects in a geographic area and their interconnections, and is an essential component of intelligent transportation systems (ITS) enabling emerging new applications such as dynamic route guidance, driving assistance systems, and autonomous driving. As the digitization of geospatial information becomes prevalent, a number of road networks with a wide variety of characteristics may coexist. In this paper, we present an *area partitioning and subgraph growing (APSG)* approach to the conflation of two road networks with a large difference in the level of details and representation rules. Our area partitioning (AP) scheme partitions the geographic area using the Network Voronoi Area Diagram (NVAD) of the low-detailed road network. Next, a subgraph of the high-detailed road network corresponding to a complex intersection is extracted and aggregated into a supernode so that high precision can be achieved via 1:1 road object matching. For the unmatched road objects due to missing road objects and different representation rules, we also propose a subgraph growing (SG) scheme that sequentially inserts a new road object while keeping the consistency of its connectivity to the matched road objects by the AP scheme. From the numerical results at Yeouido, Seoul, Korea, we show that our APSG scheme can achieve an outstanding matching performance in terms of the precision, recall, and F1-score.

Keywords: Road network conflation; area partitioning; subgraph growing; intelligent transportation systems.

Citation: Nguyen, H.-H.; Jeong, H.-Y. Title. *Sensors* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Geographic information systems (GIS) provide the solutions for capturing, manipulating, analyzing and visualizing the geospatial data for many application fields, such as transportation, agriculture, commerce, etc. [1,2]. Initially, government agencies have built authoritative GIS because the construction of geospatial information requires extensive and accurate surveys of the land [3,4]. Recently, as the digitization of geospatial information has recently become prevalent, some portal sites or mobile service providers have constructed proprietary GIS that combines authoritative GIS, aerial photos, mobile-mapping service (MMS), and crowdsourcing data, etc. [5,6]. On the other hand, voluntary GIS, such as the openstreetmap (OSM), has been constructed by the participation of voluntary users carrying a GPS-enabled mobile terminal [7]. Currently, more than 7.8 million registered users all around the world contribute to the OSM [8].

A *road network* is a subset of GIS that focuses on road objects, attributes, and their interconnectivity. It is usually represented by a graph, where a node represents an intersection, an endpoint of a road, or a point of attribute change, whereas an edge represents a road segment connecting two nodes. The road network is an important component of many Intelligent Transportation System (ITS) applications. For example, turn-by-turn navigation establishes the shortest route connecting the origin and destination in the road network. In addition, the current traffic situation on the road segment

Table 1. Characteristics of authoritative, proprietary, and voluntary road networks

Characteristics	**Authoritative** [4]	**Proprietary**[5,6]	**Voluntary** [7]
Raw dataset	Accessible	Inaccessible	Accessible
Quality	Intermediate	High	Low
Level of detail	Low	High	High
Real-time data	Available	Available	Not available
Software packages	None	Limited	Abundant

is indexed by the corresponding identifier in the road network, and then broadcast as public transportation data (PTD), which enables novel ITS applications, such as dynamic route guidance [9–12] and dynamic traffic management [13–15]. In a high-precision map for autonomous driving, each lane of a road can be represented in connection with the corresponding road segment of the road network [16].

Table 1 shows the characteristics of authoritative, proprietary, and voluntary road networks. First, the authoritative road network called *node-link map (NLM)* is designed to support ITS applications in Korean major roads [4]. It provides the representation of a road object associated with its PTD attributes, such as average speed, road incidents, variable-message signs, and CCTV streamings [17]. Two major limitations of the NLM are the lack of software packages for ITS applications and the low-detailed representation of the road network. Second, the proprietary road network has good characteristics to support ITS services, but the access to its raw dataset and the ITS software packages is either very limited or impossible. The voluntary road network called the *OSM road network (ORN)* provides a detailed view of the road network as well as a variety of open-source software packages: map editing tools (Potlatch 2 [18] and JOSM [19]), rendering tools, (Mapnik [20] and the Tirex [21]), geocoding tools (Nominatim [22]), and especially routing tools (the open-source routing machine [23] and the Valhalla [24]). However, it has been reported that the quality of OSM objects obtained from crowdsourcing can be diverse in terms of accuracy, completeness, and consistency [25].

Taking into account the characteristics of road networks, we consider the *road network conflation (RNC)* between the authoritative and voluntary road networks, i.e. NLM and ORN, for emerging new ITS services. The RNC can be seen as a generalization of the road network matching (RNM) in [26–40]: Given two road networks, the RNM finds the association between a set of objects in one road network and another set in the other, where both sets represent the same road entity. Since the RNM is done without any modifications of input road networks, it cannot address the problem of missing road objects that can be found in the voluntary road networks [25]. The RNC relaxes this restriction by allowing to add road objects to one input road network. Since each road network has its own strengths and weaknesses, a successful RNC solution can enhance the strengths and compensate for the weaknesses. In particular, it can suggest a new direction to the emerging new ITS applications through the integration of NLM-indexed real-time transportation data with ORN software packages. The challenge of RNC is how to address the difference between two road networks, including level of details (LoD) [30,35,40], missing road objects [30,31,35], and representation rules.

In this paper, we present an *area partitioning and subgraph growing (APSG)* approach to the RNC that consists of two schemes: the area partitioning (AP) scheme for the RNM and the subgraph growing (SG) scheme for the unmatched NLM objects by the AP scheme. Our AP scheme exploits the network Voronoi area diagram (NVAD) in [41] to partition the map area into a set of regions centered on each node in the NLM graph. For each partitioned region, it extracts the ORN subgraph of a complex intersection and then aggregates it into an ORN supernode so that it can be associated with NLM node via 1:1 node matching. For the unmatched NLM subgraph due to missing road objects and different representation rules, we also propose the SG scheme that sequentially inserts an ORN road object corresponding to the unmatched NLM subgraph while keeping the consistency of its connectivity to the matched NLM subgraph by the AP scheme. The numerical results at Yeoui-do, Korea's autonomous vehicle testing site, show that our APSG approach can achieve an outstanding RNC performance in terms of precision and recall. The contributions of this paper are summarized as follows:

- As far as we are aware of, this is the first work to provide a formal definition of RNC that allows to insert new road objects into one road network only, and to present a novel APSG approach that achieves an outstanding matching performance;
- The proposed AP scheme can accurately cluster the nodes at a complex intersection not only by partitioning the map area using the NVAD but also extracting the precise subgraph that yields the maximum number of paths across the NVAD; and
- To address the problem of missing road objects and different representation rules, the proposed SG scheme inserts a new road object into the ORN subgraph so that it is as consistent as possible with the existing matchings by the AP scheme.

The remainder of this paper is organized as follows. Section 2 introduces the related works of the RNC. Section 3 describes the characteristics of two road networks and formulate the RNC problem. In section 4, our AP scheme for the RNM is presented in detail. Section 5 presents the SG scheme for the unmatched NLM objects. The numerical results are discussed in section 6, and finally the conclusion of this paper is given in section 7.

2. Related Work

Given two input road networks, RNM is the process of associating road objects and combining their attributes that represent the same road entity without any modifications of the input road networks. In the literature, numerous research efforts have focused on the RNM [27–40]. For a complete solution to RNM, it is necessary to comprehensively take into account the geometric and topological characteristics of all road objects in both input road networks. However, since it is difficult to reflect their global information, most of the existing approaches sequentially match a road object with its counterpart based on its local information. Depending on the type of matching road objects, the existing RNM approaches are classified into the node (or point) [26–30], path (or line) [31–34,36,37], and subgraph matching [38,39].

First, the node matching focuses on the matching between the points in the input road networks, such as intersections, traffic monitoring points, and the endpoints of overpass/underpass, bridges, and tunnels. The basic idea of node matching is to assess the proximity of the points to be matched, as well as the similarity of their geometric and topological properties of incident edges. The seminal work in [26] presents an iterative scheme for RNM between the United States Geological Survey (USGS) and the Bureau of the Census: At each iteration, given a part of nodes already matched with their counterparts, the remaining nodes are relocated by the rubber-sheet transformation and then a new set of 1:1 node matchings is obtained again. In [27], the 1:1 node matching between two road networks with an order of scale difference exploits a few geometric dissimilarity measures, such as the Euclidean distance, nodal degree, and average orientation difference of incident edges. Given a node matching of a node and its all neighbor nodes, paper [28] presents a round-trip walk scheme for evaluating the local topological consistency along the round-trip path across the two road networks and node matching. Although this paper also identifies the difficulties of 1:n and m:n node matchings, they are left as an open problem. By replacing the Euclidean distance of the DBSCAN clustering in [42] with the graph distance of road network, the authors in [30,40] presents a node clustering scheme that aggregates the multiple nodes at a complex intersection into a single node. However, their clustering approach aggregating all intermediates nodes with an empirically determined stroke-length threshold may include too many nodes that do not belong to the complex intersection (as shown in Figure 15(a)), which significantly degrades the overall matching performance. On the contrary, the proposed AP scheme can accurately cluster the nodes at the complex intersection not only by partitioning the whole map area based on the NVAD, but also by extracting the precise subgraph that yields the maximum number of paths across the NVAD, which will be shown in section 6.3.

Second, the path matching associates a path in one road network with another path in the other: Depending on the number of edges in each path, the path matching can be classified into 1:1, 1:n, m:1, and m:n edge matchings. A buffer-growing approach is proposed to address the most general m:n edge matching, where the merit function of potential matching pairs are computed by the mutual information of positions, angles, lengths, and forms within two-hop distance, and the one with the highest mutual information is eventually selected as the matching pair [31]. An adaptive algorithm is proposed to determine the appropriate buffer size of buffer-growing algorithm [32]: If the buffer size is too small, no candidate path can be found, and if the buffer size is too large, the computation complexity becomes high. However, the buffer-growing algorithm has two limitations: 1) To reduce the global errors between two input road networks, it requires an initial affine transformation using manually selected control points at the preprocessing step; and 2) To compute the mutual information, it also needs the statistical distribution of previously matched data from the same pair of input road networks, which is not usually available. A probabilistic relaxation scheme is also presented in [33], where it initializes the probability matrix based on the geometric dissimilarity of paths, iteratively updates the matching probabilities by evaluating the compatibility of neighbor candidate pairs, and selects the final 1:1 and 1:n matching pairs from the probability matrix. The probabilistic relaxation scheme in [34] improves the matching performance not only by considering both geometric and topological characteristics in the computation of probability matrix but also by inserting a virtual node in order to address m:n matching pattern. To mitigate the user errors in the OSM crowdsourcing process, our APSG approach to the RNC problem also inserts a new node and edge into the ORN subgraph so that it can better match with the NLM.

Finally, the subgraph matching starts from an initial matching between the seed nodes, and the matched subgraph grows through a sequence of path and node matchings at each iteration. The semi-automated RNM in [38] consists of automated and interactive matching algorithms: The former includes the establishment of an initial matching for seed nodes and the expansion of the matching via cluster-based node/path matching algorithms, while the latter allows a human operator to manually correct the incorrect and improper initial matchings. On the other hand, the iterative matching algorithm in [39] initially performs the rubber-sheet transformation and topologically splits a path to maximize the number of 1:1 edge matchings. Then, starting from a subset of seed nodes, its combined edge and node matching algorithm gradually adds 1:1 matchings at the boundary of the existing matching set. Since the subgraph matching associates two *existing* road objects that represent the same road entity, its subgraph growing is determined by the similarity measure of their geometric and topological characteristics. The prime difference of our SG scheme is that *new road objects* are sequentially inserted into the subgraph of one road network to address the problem of missing road objects and different representation rules. In this process, the order of inserted road objects is carefully determined so that the resulting subgraph is as consistent as possible with the existing matchings by the AP scheme.

3. Input Road Networks and Problem Specification

In this section, we describe the characteristics of two road networks, i.e. NLM and ORN, and then formulate the RNC problem.

3.1. Node-Link Map

The Korean government has initiated the national GIS project in 1995, and completed the construction of the geospatial database in 2009 [43]. The NLM is the road network of this database that represents major road objects in Korea [4]. It also provides

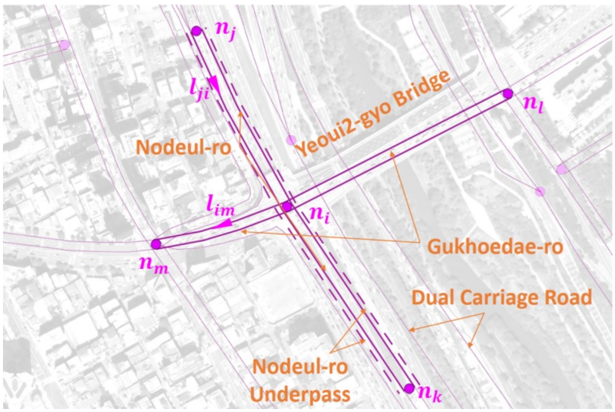


Figure 1. NLM graph representation around Yeoui2-gyo intersection

Table 2. The *road_rank* attribute in the NLM.

road_rank	**Explanation**
101	Highway
102	Urban expressway
103	National road
104	Metropolitan city road
105	Aerial or inter-province road
106	Intra-province road
107	Intra-city or island road
108	Other roads

a unified identifier (ID) hierarchy to its road entity. In order to efficiently exchange the ITS information, the Korean law enforces that all ITS applications must use the NLM ID hierarchy to exchange road and traffic information [17].

Figure 1 shows NLM graph representation of Yeoui2-gyo intersection, Yeoui-do, Seoul, Korea overlaid on top of the aerial view, where Gukhoe-daero (east-west road) and the access ramps of Nodeul-ro (north-south underpass) are interconnected. The NLM graph is a *directed* graph $\mathcal{G}_N = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of nodes representing the points at which the road characteristics are changed, such as intersection (n_i, n_l, and n_m), traffic monitoring point, administrative boundary, and the endpoints of road, overpass, and underpass (n_j and n_k). A single NLM node $n_i \in \mathcal{N}$ is used to represent a complex intersection (Yeoui2-gyo) without a detailed view of the internal road network. We define subgraph $\mathcal{G}_N(n_i) = (\mathcal{N}(n_i), \mathcal{L}(n_i))$ consists of NLM node n_i, its directly connected links (pink solid links in $\mathcal{G}_N$), and the neighbor NLM nodes (n_j, n_k, n_l, and n_m). An NLM node is placed at the crosspoint of two roads, where a road consists of two parallel links each of which represents a unidirectional road segment. In a dual carriage road, it is placed at the endpoint of two NLM links.

In the NLM, the geometric shape of a link is approximated by a sequence of concatenated line segments. For example, unidirectional links l_{ji} and l_{im} are shown by pink solid lines with triangular marks for their directions. The underpass and overpass links are also placed in parallel with the main road segment with additional spacing between them. In this paper, we represent each NLM underpass/overpass by the pink dashed line, as shown in Figure 1. Each link has a set of attributes, such as *link_id, f_node, t_node, road_rank, road_type, connect, road_use*, etc., where the *road_rank* attribute represents the class of road segment as shown in Table 2, *road_type* specifies the type of road, such as overpass, underpass, bridge, tunnel, etc., *connect* specifies the type of ramps depending on *road_rank* attribute, and *f_node* and *t_node* represent the start and end node indexes of NLM link, respectively.

Table 3. Major *highway* tag of an OSM way.

highway tag group	*highway* tag value
Roads	*motorway, trunk, primary, secondary, tertiary, unclassified, residential, service*
Link roads	*motorway_link, trunk_link, primary_link, secondary_link, tertiary_link*
Special roads	*living_street, pedestrian, track, bus_guideway, escape, raceway, road*
Paths	*footway, bridleway, steps, path, cycleway*
Sidewalks	*sidewalk*
Cycleways	*cycleway*

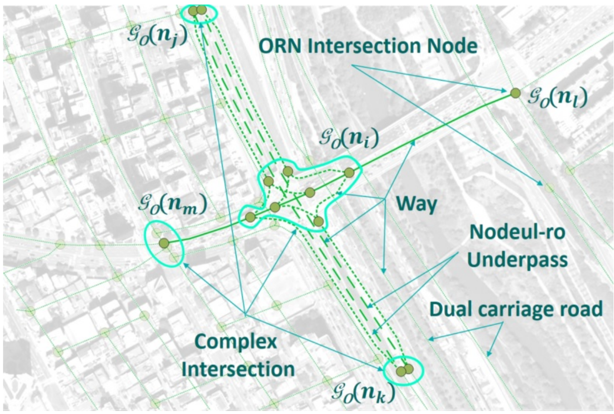


Figure 2. ORN graph representation around Yeoui2-gyo intersection

3.2. OpenStreetMap Road Network (ORN)

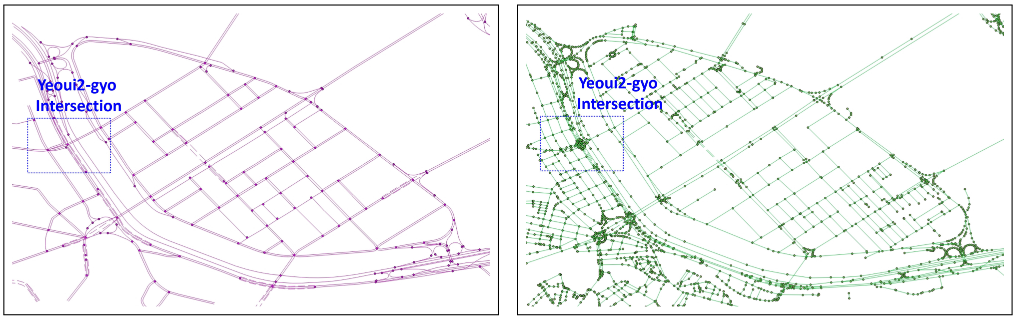
The ORN is a subset of OSM objects with *highway* tag, where a tag is an ordered pair of (key, value) identifying the attribute of a road object. Table 3 shows the *highway* tag of way which is classified into a few groups. In each group, the tag values are ordered from the most important to the least important. The main focus of this paper is on the *road* and *link road* groups, where the former is a way for representing a road while the latter is a way for connecting two roads in a complex intersection. We initially prune all ORN objects in *special roads*, *paths*, *sidewalks*, and *cycleways* groups that do not correspond to the NLM objects. This pruning process removes approximately 20 % of unnecessary road objects from the original ORN. Furthermore, we also remove the subgraphs for underpass/overpass in both road networks because they can be easily matched via their attribute, such as NLM *road_type*, and ORN *tunnel* and *bridge* tags[1].

Figure 2 shows the ORN graph representation which can be modeled by *undirected* graph $\mathcal{G}_O = (\mathcal{V}, \mathcal{E})$. Contrary to NLM graph $\mathcal{G}_N$, ORN graph $\mathcal{G}_O$ is designed to reflect the detailed road network at a complex intersection. This feature makes the ORN more suitable for ITS applications, such as navigation and autonomous driving.

In $\mathcal{G}_O$, an ORN node $v \in \mathcal{V}$ is connected to at least three neighbor ORN nodes. In the RNC, NLM node n is associated with ORN subgraph $\mathcal{G}_O(n)$, where the ORN subgraph can be a single ORN intersection node , e.g. $\mathcal{G}_O(n_l)$, disconnected subgraphs, e.g. $\mathcal{G}_O(n_j)$ and $\mathcal{G}_O(n_k)$, or a connected subgraph, e.g. $\mathcal{G}_O(n_i)$ and $\mathcal{G}_O(n_m)$, in Figure 2. If an intersection consists of a single ORN intersection node, it is called a simple intersection; otherwise, a complex intersection.

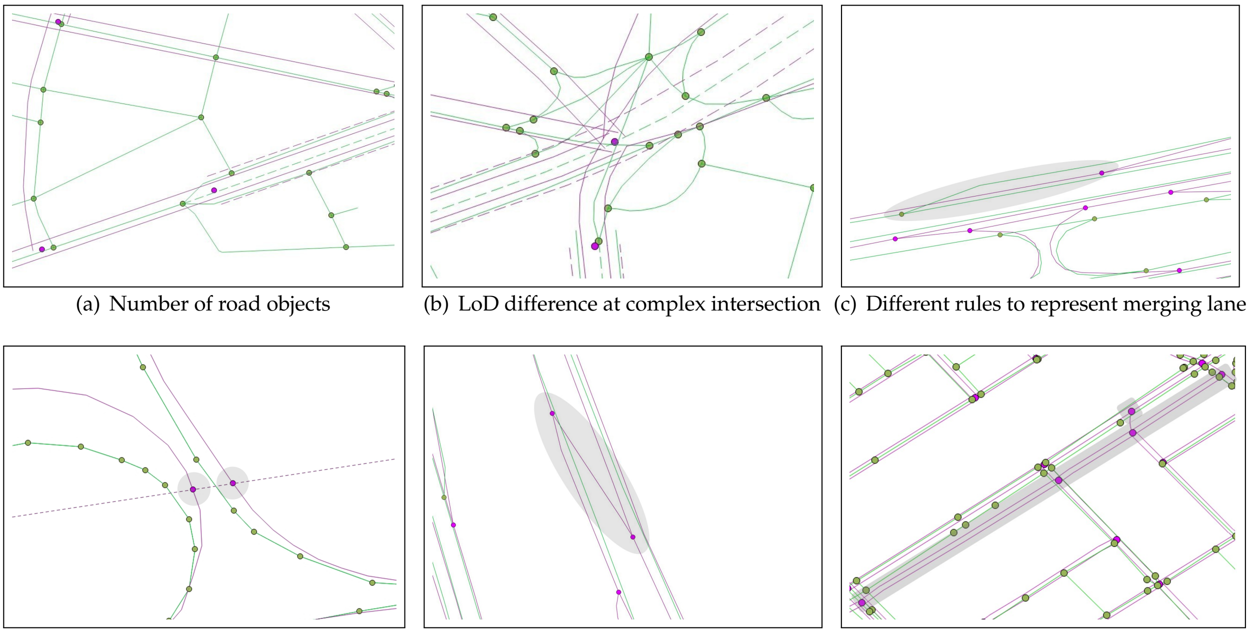
The atomic unit for representing an ORN road is a way $w \in \mathcal{W}$ which may span multiple ORN nodes [7]. If way w includes more than two ORN nodes, it is decomposed into consecutive ORN edges $e \in \mathcal{E}$ so that each edge connects two ORN nodes only. In Figure 2, the Gukhoe-daero in the ORN subgraph $\mathcal{G}_O(n_i)$ consists of edges with *road* tag group only, shown in solid green lines, whereas all remaining edges in $\mathcal{G}_O(n_i)$ belong to *link road* tag group, represented by dotted green lines. On the other hand, all intersecting

[1] In some figures, we still illustrate ORN underpass/overpass with green dashed line for the clarity of expression.



(a) NLM graph of Yeoui-do roads (b) (Pruned) ORN graph of Yeoui-do roads

Figure 3. NLM and ORN graph representation of Yeoui-do roads



(a) Number of road objects (b) LoD difference at complex intersection (c) Different rules to represent merging lane

(d) No corresponding ORN nodes (e) No corresponding ORN edge (f) No corresponding ORN subgraph

Figure 4. Examples of the representational dissimilarities between NLM and ORN

edges at a simple intersection, such as $\mathcal{G}_O(n_l)$, belong to *road* tag group. In the case of dual carriage road, a distinct edge is used for each ORN edge whose direction is specified to the *direction* tag.

3.3. Problem Specification

Figure 3 shows the NLM and ORN graph representation of Yeoui-do roads which are given as the input of our RNC problem. The NLM in Figure 3(a) is a low-detailed road network consisting of major public roads only, while the ORN in Figure 3(b) has a much more detailed representation of the road network. Given NLM $\mathcal{G}_N$ and ORN $\mathcal{G}_O$, the RNC problem is an association problem that finds the ORN subgraph corresponding to each NLM object while allowing to add new road objects to the ORN.

Since each road network has its own rules for representing its road objects, there are several differences in representing road objects between two road networks as shown in Figure 4: Figure 4(a) shows different numbers of road objects, where the ORN shows both major and minor roads in a geographical area while the NLM displays major public roads only. Figure 4(b) illustrates different LoDs at a complex intersection, where the ORN illustrates all detailed connectivity at the intersection whereas the NLM aggregates them into a single NLM node. Figure 4(c) reveals two different rules to represent a merging lane, where it is a part of the mainline road in the NLM, while it is a part of on-ramp with *trunk_link* tag in the ORN. Figure 4(d) shows two NLM nodes that do not have the corresponding ORN subgraphs at the crosspoints of the administrative boundary. Figure 4(e) illustrates an NLM link without the corresponding ORN object due to its omission during the crowdsourcing process of OSM. Figure 4(f) also shows an NLM subgraph that does not have the corresponding ORN subgraph due to the OSM crowdsourcing errors.

To summarize, a comprehensive solution to the RNC problem needs to address the fundamental issues of these representational differences, as follows:

1. To identify the ORN subgraph of a complex intersection in order to alleviate the LoD difference between two road networks,
2. To find a reliable methodology to cope with the differences in the representation of merging lane and administrative boundary, and
3. To create a new ORN subgraph corresponding to the unmatched NLM subgraph while keeping the consistency of its connectivity to the matched NLM subgaph.

4. Area Partitioning for LoD Difference at a Complex Intersection

For a given NLM node n_i with NLM subgraph $\mathcal{G}_N(n_i)$ and ORN graph $\mathcal{G}_O = (\mathcal{V}, \mathcal{E})$, the challenging task is to accurately extract ORN subgraph $\mathcal{G}_T(n_i)$ against a wide variety of intersection topology as shown in Figure 4(b). An inaccurate ORN subgraph incurs an incorrect matching which in turn influences the accuracy of another matching. This propagation eventually results in severe degradation of RNC performance.

Our AP scheme first computes the region of the map dedicated to NLM node n_i in which the corresponding ORN subgraph may exist. Then, it extracts ORN subgraph $\mathcal{G}^*_O(n_i)$ along the path connecting each pair of entering and exiting points across the region boundary, taking into account the turning information and geometry of intersection. Finally, ORN subgraph $\mathcal{G}^*_O(n_i)$ is replaced by an ORN supernode v^*_i so that it can be matched with NLM node n_i via 1:1 node matching.

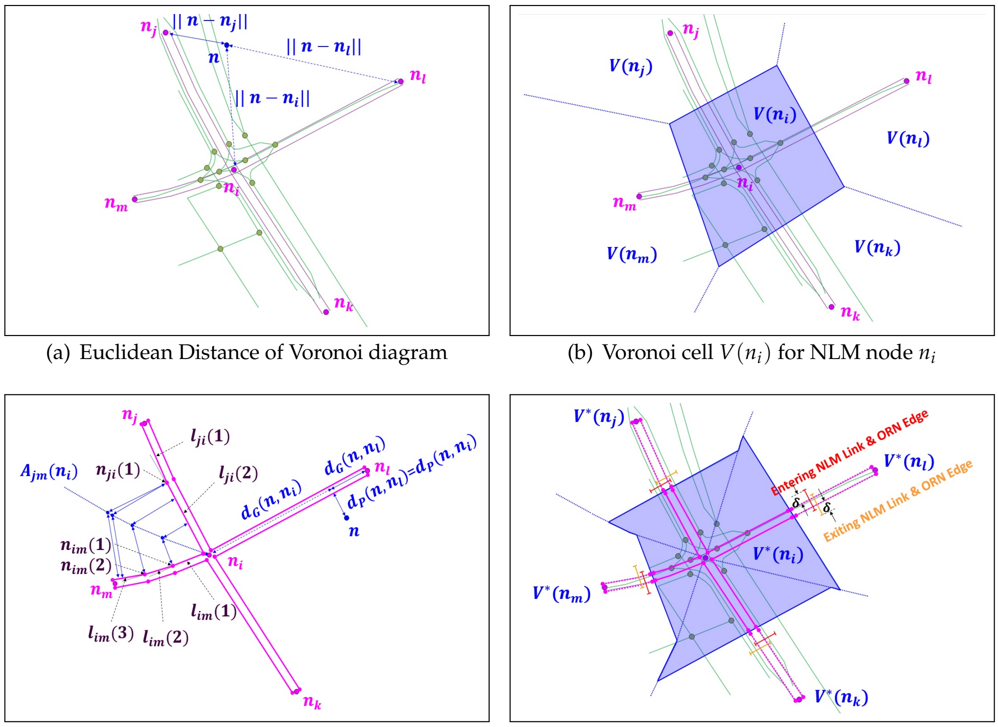
4.1. Network Voronoi Area Diagram (NVAD) for Partitioning Map Area

Given NLM subgraph $\mathcal{G}_N(n_i)$ and the corresponding map area $\mathcal{A}(n_i)$ around n_i, the first task of our AP scheme is to partition this area into regions, where each region is centered at an intersection in $\mathcal{N}(n_i)$. A simple method called the *Voronoi diagram (VD)* partitions the map area $\mathcal{A}(n_i)$ based on the Euclidean distance [41]. The basic idea is to associate a point $n \in \mathcal{A}(n_i)$ with the region of the closest intersection n_x, called the Voronoi cell $V(n_x)$, in terms of the Euclidean distance metric:

$$V(n_x) = \{n|\ \|n - n_x\| \leq \|n - n_y\| \quad \forall y \neq x,\ n_x, n_y \in \mathcal{N}(n_i)\}, \tag{1}$$

where $\mathcal{N}(n_i) = \{n_i, n_j, n_k, n_l, n_m\}$ for NLM graph $\mathcal{G}_N(n_i)$ in Figure 5(a). Given an ORN node $n \in V(n_j)$ in map area $\mathcal{A}(n_i)$, the Euclidean distances from three closest NLM nodes are shown in Figure 5(a). For two NLM nodes n_x and n_y ($n_y \in \mathcal{N}(n_i) \backslash \{n_x\}$), the boundary of Voronoi cells becomes a hyperplane that is equidistant from both NLM nodes. Finally, Voronoi cell $V(n_x)$ is constructed by intersecting all half-spaces in which NLM node n_x is located. For example, Voronoi cell $V(n_i)$ is illustrated with the blue transparent quadrilateral in Figure 5(b).

However, given NLM subgraph $\mathcal{G}_N(n_i)$, the Euclidean norm is no longer a fair measure to evaluate the distance between point $n \in \mathcal{A}(n_i)$ and the set of NLM nodes in $\mathcal{N}(n_i)$. This is because the Euclidean distance metric does not account for the distance from the curved roads in $\mathcal{G}_N(n_i)$. To address this problem, our AP scheme adopts the *network Voronoi area diagram (NVAD)* whose measure reflects two distance factors [41]: First, if point n is on subgraph $\mathcal{G}_N(n_i)$, the distance should be the length of shortest path to NLM node $n_x \in \mathcal{N}(n_i)$ in $\mathcal{G}_N(n_i)$, called the graph distance $d_G(n, n_x)$. If point n lies in $\mathcal{A}(n_i) \backslash \mathcal{G}_N(n_i)$, the measure should also consider the projection distance $d_P(n, n_x)$ to the closest NLM link of subgraph $\mathcal{G}_N(n_i)$. Figure 5(c) shows these distances between



(a) Euclidean Distance of Voronoi diagram

(b) Voronoi cell $V(n_i)$ for NLM node n_i

(c) Distance metrics and projection boundary

(d) NVAD $V^*(n_i)$ for NLM node n_i

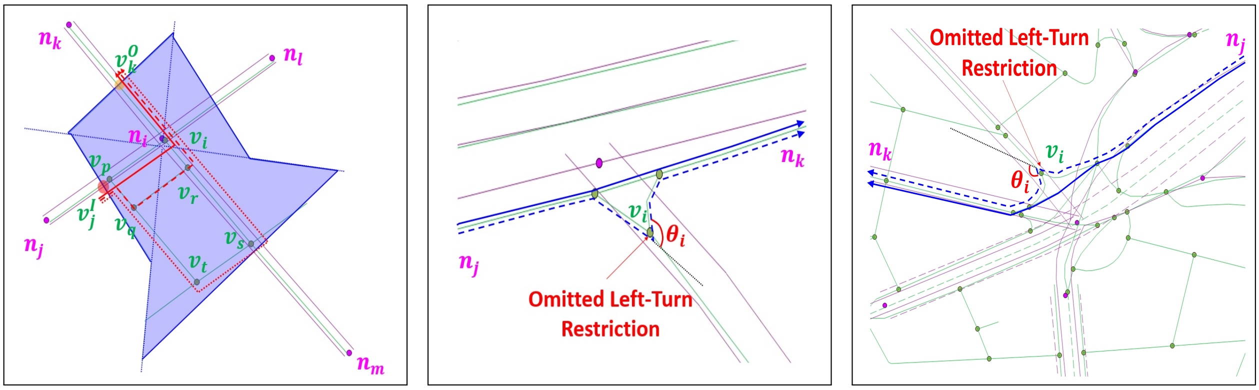
Figure 5. VD and NVAD to partition map area $\mathcal{A}(n_i)$ around NLM node n_i.

point n and two closest intersections n_i and n_l. Consequently, the distance metric of NVAD is defined as the sum of these two distance components, i.e.,

$$\|n - n_x\| = d_G(n, n_x) + d_P(n, n_x). \tag{2}$$

To determine the NLM link onto which a given point n is projected, we choose an example of map area $\mathcal{A}_{jm}(n_i)$ surrounded by unidirectional NLM links l_{ji} and l_{im} in Figure 5(c), where the former (latter) consists of two (three) line segments. The k-th line segment and vertex of NLM link l_{ji} are denoted by $l_{ji}(k)$ and $n_{ji}(k)$, respectively, where $n_{ji}(0) = n_j$ and $n_{ji}(2) = n_i$. Our approach draws the equiangle boundary starting from the center of intersection n_i until its projection approaches the endpoint of shorter line segment $n_{im}(1)$. Notice that any points on this projection boundary are equidistant from both NLM links l_{ji} and l_{im}. In Appendix A, we demonstrate that the projection boundary curve becomes a concatenation of linear or parabolic segments. Figure 5(c) shows the resulting blue dotted projection boundary of map area $\mathcal{A}_{jm}(n_i)$.

Figure 5(d) shows all projection boundaries that partition map area $\mathcal{A}(n_i)$ into four projection areas each of which has a pair of NLM links between n_i and its neighbor NLM node. At the middle point of these links, we draw a perpendicular line that bisects the projection area. Then, NVAD cell $V^*(n_i)$ is determined by the union of the bisected map area in which NLM node n_i is located, as shown by the blue transparent polygon in Figure 5(d). For each NLM link l passing through the NVAD cell boundary, we finally build a list of candidate ORN edges $\mathcal{E}_l = \{e_l(1), e_l(2), \cdots\}$ of the same direction whose distance along the boundary line is less than threshold δ. For example, in Figure 5(d), NLM links l_{ji} and l_{ik} have two ORN edges in their lists, while all remaining NLM links have only one ORN edge. In the next section, the candidate ORN edges will be examined to be the correspondent of an NLM link.



(a) Simple intersection (b) Exceptional case I (c) Exceptional case II

Figure 6. Candidate ORN paths between the ORN edges closest to l_{ji} and closest to l_{ik} at the boundary of NVAD cell $V^*(n_i)$

4.2. Extraction of Candidate ORN Subgraph

Given NVAD cell $V^*(n_i)$, allowable turn information at NLM node n_i, and set $\mathcal{E}_l$ of candidate ORN edges for NLM link l, the objective of this section is to extract the corresponding ORN subgraph $\mathcal{G}_O^*(n_i)$ in $V^*(n_i)$ that corresponds to NLM node n_i. Our key observation is that *an intersection allows at most one path for each pair of roads, where one enters to and the other exits from NVAD cell $V^*(n_i)$*. Starting with null ORN subgraph having no ORN node and edge, the basic idea of our approach is to sequentially insert an ORN path passing through the intersection along which the turn restriction is satisfied at each pair of consecutive ORN edges. Without loss of generality, we focus on the construction of ORN path p_{jk} as shown in Figure 6.

Figure 6(a) shows an example of simple intersection, where NLM node n_i connects a two-way road (l_{ki} and l_{ik}) and three one-way roads (l_{ji}, l_{il}, and l_{mi}). Due to the LoD difference, the ORN subgraph in $V^*(n_i)$ consists of two components: 1) the true ORN subgraph almost overlapped with NLM subgraph $\mathcal{G}_N(n_i)$, and 2) the remaining ORN subgraph representing a minor road network around intersection n_i. Denoting by v_j^I and v_k^O the crosspoints of the entering and exiting ORN edges at the boundary of NVAD cell $V^*(n_i)$, respectively, there are three candidate paths in Figure 6(a): $p_{jk}(1) = v_j^I \rightarrow v_i \rightarrow v_k^O$ (red solid path), $p_{jk}(2) = v_j^I \rightarrow v_p \rightarrow v_q \rightarrow v_r \rightarrow v_k^O$ (red dashed path), and $p_{jk}(3) = v_j^I \rightarrow v_p \rightarrow v_t \rightarrow v_s \rightarrow v_k^O$ (red dotted path). Among these paths, our *candidate ORN subgraph extraction (COSE)* scheme chooses the path that has smallest sum of turning angles regardless of its direction. For example, path $p_{jk}(1)$ has the smallest total turning angle since it makes only one left-turn at v_i compared to three turns in the other two paths.

Although ORN subgraph $\mathcal{G}_O^*(n_i)$ is much more complex than NLM subgraph $\mathcal{G}_N(n_i)$ around a complex intersection, it is surprising that our key observation is valid for all complex intersections in Yeoui-do except for the blue dashed paths in Figures 6(b) and 6(c). They are evidently originated from the crowdsourcing error that omits a left-turn restriction at ORN node v_i by the participating users, and eventually, turn out to be invalid paths. Unfortunately, these human errors are inevitable in the crowdsourcing-based ORN. To exclude these exceptional paths from ORN subgraph $\mathcal{G}_O^*(n_i)$, we exploit the second key observation that *the geometry of connecting roads in a complex intersection is designed in a way that the curvature changes linearly with the curve length*, which is known as the *clothoids*. Based on this observation, the COSE scheme discards a path, if it has two consecutive edges and the angles between them abruptly change, e.g. the blue dashed path at node v_i in Figures 6(b) and 6(c).

The final step of the COSE scheme is the derivation of ORN subgraph $\mathcal{G}_O^*(n_i)$ for NVAD node n_i. It first calculates the number of allowable ORN paths for each ingress-

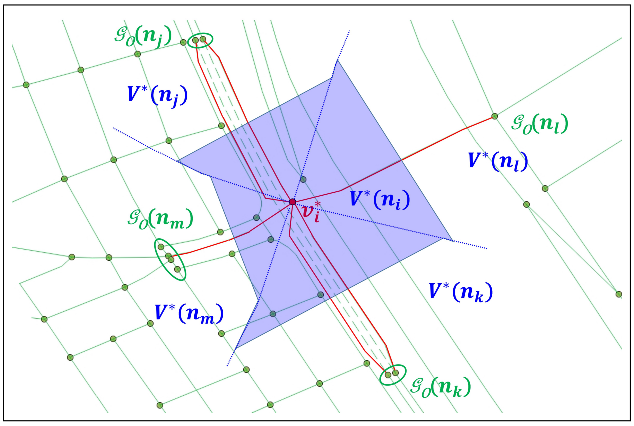


Figure 7. ORN supernode v_i^* replacing the final ORN subgraph $\mathcal{G}_O^*(n_i)$

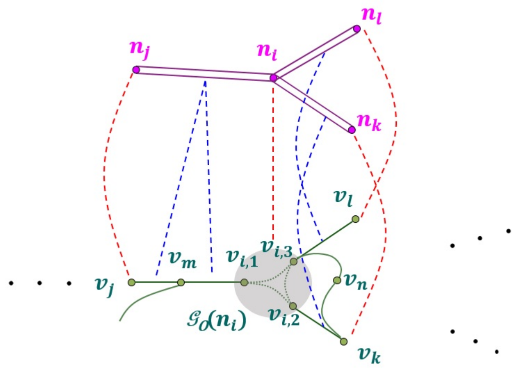


Figure 8. Example of RNM results

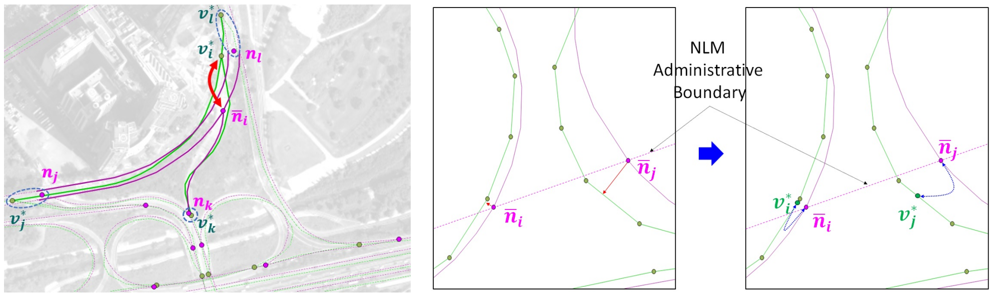
egress pair of ORN edges at the boundary of NVAD cell $V^*(n_i)$. Then, it chooses the optimal ORN subgraph $\mathcal{G}_O^*(n_i) = (\mathcal{V}^*(n_i),\ \mathcal{E}^*(n_i))$ that yields the largest number of allowable ORN paths. Finally, it extracts all ORN nodes from $\mathcal{V}^*(n_i)$. If there are more than one ORN node in $\mathcal{V}^*(n_i)$, our AP approach replaces them with ORN supernode v_i^* located at the center of them, as shown in Figure 7. By this replacement, the node matching becomes a simple 1:1 matching between NLM node n_i and ORN supernode v_i^*.

4.3. Classification of RNM Result

Figure 8 shows the matching results between NLM subgraph $\mathcal{G}_N(n_i)$ and ORN graph $\mathcal{G}_O$. Depending on which road object belongs to ORN subgraph $\mathcal{G}_O^*(n_i)$, both node and edge matching results can be one of the following four matching types: *correct match (CM)*, *incorrect match (IM)*, *partial match (PM)*, and *missing match (MM)*. For each NLM node, the node matching result can be determined as follows:

- The red dashed lines in Figure 8 represent the CM between NLM and ORN nodes, where the sets of true ORN nodes for NLM nodes n_i, n_j, n_k, and n_l are denoted by $\mathcal{V}_T(n_i) = \{v_{i,1}, v_{i,2}, v_{i,3}\}$, $\mathcal{V}_T(n_j) = v_j$, $\mathcal{V}_T(n_k) = v_k$, and $\mathcal{V}_T(n_l) = v_l$, respectively;
- A node matching becomes MM, if its set of ORN nodes is empty, i.e. $\mathcal{V}^*(\cdot) = \phi$;
- A node matching becomes IM, if its set of ORN nodes is *disjoint* with the set of true ORN nodes, i.e. $\mathcal{V}^*(\cdot) \cap \mathcal{V}_T(\cdot) = \phi$; and
- A node matching becomes PM, if its set of ORN nodes satisfies two conditions $\mathcal{V}^*(\cdot) \cap \mathcal{V}_T(\cdot) \neq \phi$ and $\mathcal{V}^*(\cdot) \neq \mathcal{V}_T(\cdot)$.

At the boundary of two adjacent NVAD cells $V^*(n_i)$ and $V^*(n_j)$, the COSE scheme also yields a solution to the edge matching between NLM link l and two ORN edges:



(a) Unmatched NLM node in merging lane (b) Unmatched NLM node at administrative boundary

Figure 9. Examples of unmatched single NLM node $\overline{n}_i$

one from set $\mathcal{E}_l \cap \mathcal{E}^*(n_i)$ in area $\mathcal{A}(n_i)$ and the other from $\mathcal{E}_l \cap \mathcal{E}^*(n_j)$ in area $\mathcal{A}(n_j)$, respectively. Similarly, the type of edge matching result is determined as follows:

- The blue dashed lines in Figure 8 represent the CM between NLM link and ORN edges, where the sets of true ORN edges for NLM link l_{ij}, l_{ik}, and l_{il} are denoted by $\mathcal{E}_T(l_{ij}) = \{(v_{i,1}, v_m), (v_m, v_j)\}$, $\mathcal{E}_T(l_{ik}) = (v_{i,2}, v_k)$, and $\mathcal{E}_T(l_{il}) = (v_{i,3}, v_l)$, respectively;
- An edge matching becomes MM, if its set of ORN edges is empty, i.e. $\mathcal{E}^*(\cdot) = \phi$;
- An edge matching becomes IM, if its set of ORN edges is *disjoint* with the set of true ORN edges, i.e. $\mathcal{E}^*(\cdot) \cap \mathcal{E}_T(\cdot) = \phi$; and
- An edge matching becomes PM, if its set of ORN edges satisfies two conditions $\mathcal{E}^*(\cdot) \cap \mathcal{E}_T(\cdot) \neq \phi$ and $\mathcal{E}^*(\cdot) \neq \mathcal{E}_T(\cdot)$.

Finally, we partition NLM graph $\mathcal{G}_N$ into the matched and unmatched NLM subgraphs $\mathcal{G}_N^* = (\mathcal{N}^*, \mathcal{L}^*)$ and $\overline{\mathcal{G}}_N = (\overline{\mathcal{N}}, \overline{\mathcal{L}})$, where the former includes all NLM road objects of CM, IM, and PM types, while the latter has those in MM type only.

5. ORN Subgraph Growing for Unmatched NLM Subgraph

The unmatched NLM subgraph is mainly originated from missing ORN objects in the OSM crowdsourcing process or the differences in representation rule. In general, a connected subgraph of unmatched NLM graph $\overline{\mathcal{G}}_N$ can be either NLM node $\overline{n}_i$, NLM link $\overline{l}_{ij}$, or NLM component $\overline{\mathcal{C}}_N$ consisting of at least two NLM road objects. First, we present two schemes for unmatched single NLM node due to the differences in representation rule: the NVAD cell expansion (NCE) scheme for a merging lane in Figure 4(c) and the NLM node projection onto ORN edge (NPOE) scheme for administrative boundaries in Figure 4(d). Second, we present the ORN edge insertion (OEI) scheme for unmatched single NLM link in Figure 4(e). Finally, we present the sequential ORN subgraph growing (SOSG) scheme for unmatched NLM component in Figure 4(f). Finally, we also address the internal structure design of new ORN nodes by the SG scheme.

5.1. Schemes for Unmatched Single NLM Node

We present two schemes to address the difference in representation rule: the NCE scheme for merging lane and the NPOE scheme for the administrative boundary. This difference results in an isolated NLM node as shown in Figure 9.

5.1.1. NCE Scheme for Merging Lane

Figure 9(a) shows a typical example of different rules for representing a merging lane, where it is a part of the mainline road in NLM while it is a part of the on-ramp in ORN. As a result, the ORN edge connecting v_k^* and v_i^* is longer than the corresponding NLM link l_{ki}. This rule difference results in unmatched single NLM node $\overline{n}_i$ with $|\mathcal{L}(\overline{n}_i)| \geq 3$ because its corresponding ORN node v_i^* is located outside its NVAD cell $\mathcal{V}^*(\overline{n}_i)$.

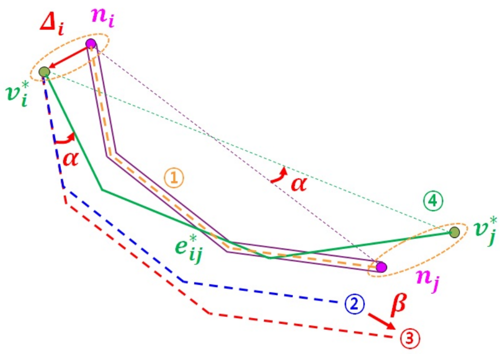


Figure 10. Example of OEI scheme for correspondent-missing NLM link $\bar{l}_{ij}$

To address this problem, we present the NCE scheme as follows: It first expands its NVAD cell $\mathcal{V}^*(\bar{n}_i)$ through the union of all NVAD cells in map area $\mathcal{A}(\bar{n}_i)$, i.e. $\cup V^*(n_x)$ for each $n_x \in \mathcal{N}(n_i)$. Next, the COSE scheme in section 4.2 is used to extract the corresponding ORN node v_i^* from all possible ORN paths, e.g. paths $v_j^* \rightarrow v_l^*$ and $v_k^* \rightarrow v_l^*$ in Figure 9(a).

5.1.2. NPOE Scheme for Administrative Boundary

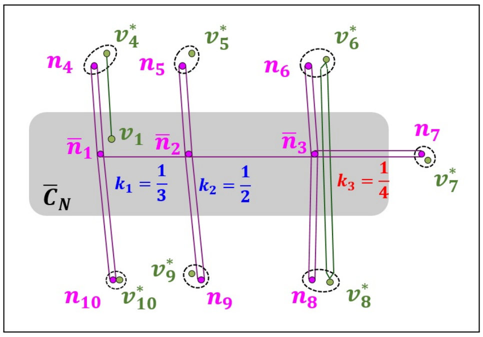
Figure 9(b) shows an example of different rules for indicating a road crossing an administrative boundary: Two nodes $\bar{n}_i$ and $\bar{n}_j$ are created to represent the administrative boundary in the NLM links, while no corresponding ORN node exists in NVAD cells $V^*(\bar{n}_i)$ and $V^*(\bar{n}_j)$, respectively. To address this problem, we propose the NPOE scheme that projects the unmatched NLM nodes $\bar{n}_i$ and $\bar{n}_j$ onto the ORN subgraphs $\mathcal{G}_O(\bar{n}_i)$ and $\mathcal{G}_O(\bar{n}_j)$ obtained from the COSE scheme, respectively. For example, Figure 9(b) shows two ORN nodes v_i^* and v_j^* that are matched with unmatched NLM nodes $\bar{n}_i$ and $\bar{n}_j$, respectively. If the unmatched NLM node is on dual carriage roads, the NPOE scheme collapses the projected ORN nodes into an ORN node located at the middle of them (See ORN node v_3^* in Figure 11(b)).

5.2. OEI Scheme for Missing ORN edge

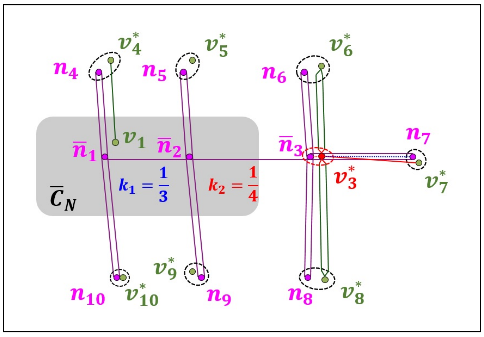
Figure 10 shows an example of OEI scheme to address the problem that there is no ORN edge corresponding to NLM link $\bar{l}_{ij}$. In this example, both endpoints n_i and n_j of NLM link $\bar{l}_{ij}$ are matched with ORN nodes v_i^* and v_j^* via the AP scheme, respectively. However, the ORN edge connecting these ORN nodes is missing possibly due to user errors in the OSM crowdsourcing process. The goal of this section is to insert an ORN edge e_{ij}^* that corresponds to NLM link $\bar{l}_{ij}$. To aim this, our OEI scheme considers three factors: 1) the displacement Δ_i between NLM node n_i and ORN node v_i^*, 2) the angle difference α between NLM line segment $(n_i,\, n_j)$ and ORN line segment $(v_i^*,\, v_j^*)$, and 3) the length ratio β of ORN line segment $(v_i^*,\, v_j^*)$ to NLM line segment $(n_i,\, n_j)$, where

$$\beta = \frac{\left\| v_i^* - v_j^* \right\|}{\left\| n_i - n_j \right\|}. \tag{3}$$

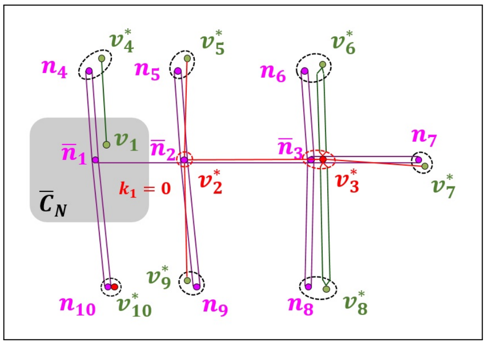
The OEI scheme first computes an orange dashed link between NLM nodes n_i and n_j which is equally distant from both NLM links $\bar{l}_{ij}$ and $\bar{l}_{ji}$. Next, it obtains a blue dashed link by shifting the orange dashed link by Δ_i so that it can start from ORN node v_i^*. Then, it computes a red dashed link by multiplying the scaling factor β to the blue dashed line. Finally, ORN edge e_{ij}^* in Figure 10 is obtained by rotating the red dashed link by angle α around ORN node v_i^*.



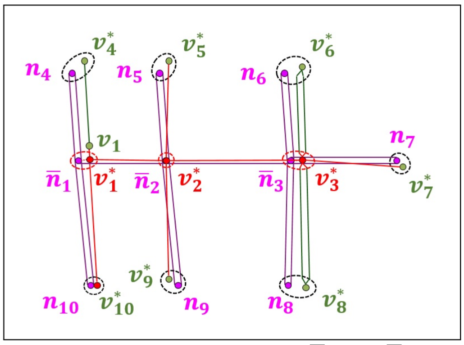
(a) Initial unmatched NLM component $\overline{C}_N$



(b) After the extraction of $\overline{n}_3$ from Q



(c) After the extraction of $\overline{n}_2$ from Q



(d) The ORN subgraph $\mathcal{G}_O(\overline{C}_N)$ for $\overline{C}_N$

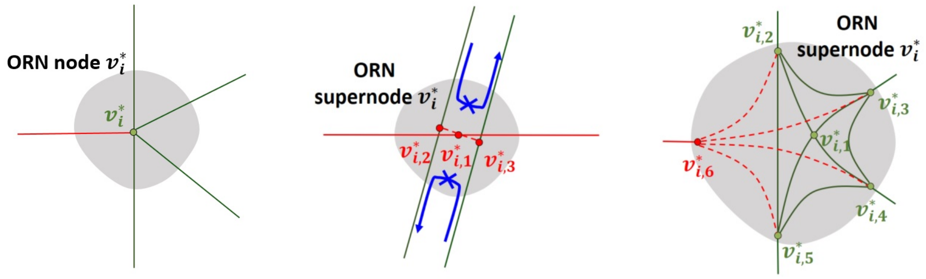
Figure 11. Example of SOSC scheme for unmatched NLM component $\overline{C}_N$

5.3. SOSG Scheme for Unmatched NLM Component

During the OSM crowdsourcing process, the ORN subgraph $\mathcal{G}_O(\overline{C}_N)$ corresponding to unmatched NLM component $\overline{C}_N$ may not exist due to the misinterpretation of the road network (See the example in Figure 4(f)). Figure 11(a) shows an example of NLM component $\overline{C}_N$ consisting of three unmatched NLM nodes ($\overline{n}_1$, $\overline{n}_2$, and $\overline{n}_3$), and 16 unmatched NLM links: Two unmatched NLM links connect two unmatched NLM nodes in $\overline{C}_N$ while 14 unmatched NLM links pass through the boundary of $\overline{C}_N$. The objective of our SOSC scheme is to construct a simple ORN subgraph $\mathcal{G}_O(\overline{C}_N)$ that corresponds to NLM component $\overline{C}_N$.

The basic idea of the SOSC scheme is to sequentially examine an unmatched NLM node in $\overline{C}_N$, and for each unmatched NLM node, to construct the corresponding ORN subgraph using both OEI and NPOE schemes. It maintains priority queue Q that determines the order of unmatched NLM nodes sequentially extracted from $\overline{C}_N$. To better associate with the neighbor NLM nodes in matched NLM subgraph $\mathcal{G}_N^*$, the key k_i of unmatched NLM node $\overline{n}_i$ in priority queue Q is defined as the ratio of unmatched neighbor NLM nodes to all neighbor NLM nodes $\mathcal{N}(\overline{n}_i)\backslash\{\overline{n}_i\}$. Since the three key values are $k_1 = \frac{1}{3}$, $k_2 = \frac{1}{2}$, and $k_3 = \frac{1}{4}$ in Figure 11(a), NLM node $\overline{n}_3$ is first extracted from Q.

For unmatched NLM node $\overline{n}_3$ extracted from Q, it first investigates the existence of an ORN edge corresponding to NLM links l_{3j} and/or l_{j3}, where NLM node n_j belongs to the set of matched neighbor NLM nodes $\mathcal{N}(\overline{n}_3) \cap \mathcal{N}^*$. Figure 11(a) shows a dual carriage edge between two neighbor ORN nodes v_6^* and v_8^*. In this case, it uses the NPOE scheme to insert ORN node v_3^* to the center of two projection points onto ORN edges (v_6^*, v_8^*) and (v_8^*, v_6^*) in Figure 11(b). Once ORN node v_3^* is created, the OEI scheme is used to insert a new ORN edge e_{37}^*. Then, unmatched NLM node $\overline{n}_3$ and new NLM links $l_{36}, l_{63}, l_{37}, l_{38}$, and l_{83} that have their corresponding ORN edges are removed from NLM component $\overline{C}_N$, and then inserted to matched NLM subgraph $\mathcal{G}_N^*$, which reduces the key value of unmatched NLM node $\overline{n}_2$ to $k_2 = \frac{1}{4}$ as shown in Figure 11(b).



(a) Simple intersection (b) Dual carriage road (c) Complex intersection

Figure 12. Addition of ORN subgraph to the existing ORN (super)node v_i^*

Next, unmatched NLM node $\overline{n}_2$ extracted from Q is examined to find an existing ORN edges corresponding to NLM links $l_{23}, l_{25}, l_{52}, l_{29}$, and l_{92}. Since there is no such ORN edge, the SOSG scheme overlays ORN node v_2^* on top of $\overline{n}_2$, and uses the OEI scheme to insert these ORN edges e_{23}^*, e_{25}^*, and e_{29}^* as shown in Figure 11(c). The newly matched NLM objects are removed from $\overline{C}_N$ and inserted to matched NLM subgraph $\mathcal{G}_N^*$. Finally, the key value of the last NLM node $\overline{n}_1$ is updated to zero ($k_1 = 0$).

Similarly, the last unmatched NLM node $\overline{n}_1$ in $\overline{C}_N$ has one ORN edge between ORN nodes v_1 and v_4^*. The SOSG scheme creates an ORN node v_1^* at the projection point onto the extended ORN edge, and inserts an ORN edge $\overline{v_1 v_1^*}$ in Figure 11(d). Finally, it also uses the OEI scheme to add the ORN edges e_{12}^*, e_{14}^*, and e_{110}^*, which completely covers the unmatched NLM component $\overline{C}_N$.

5.4. Internal Structure Design of New ORN Node

Figure 12 shows a few examples of adding a set of new ORN edges to an existing ORN (super)node v_i^*, where green road objects represent the existing ORN subgraph, and red objects represent new ORN subgraph by the SG scheme. There are three possible cases in the addition of a new ORN subgraph: 1) simple intersection, 2) dual carriage road, and 3) complex intersection.

To make the resulting ORN subgraph simple for the first two cases, our SG scheme restricts that *all ORN paths through the intersection must intersect at the same ORN node.* In addition, a new relation must be inserted into the ORN in order to reflect a turn restriction between a new ORN edge and an existing ORN edge. Since there is only one ORN node at a simple intersection, the new ORN edge is directly connected to ORN node v_i^* as shown in Figure 12(a). On the other hand, the ORN supernode v_i^* for dual carriage road is placed in the middle of two parallel ORN edges. In Figure 12(b), our SG scheme overlays an ORN node $v_{i,1}^*$ to this supernode, and then requires that all additional ORN edges must intersect at this point. To interconnect the dual carriage edges to ORN node $v_{i,1}^*$, it also inserts two internal (red dashed) ORN edges which connect this node and its projection onto two opposite ORN edges, i.e. ORN nodes $v_{i,2}^*$ and $v_{i,3}^*$. To avoid the u-turns via new internal ORN edges, it is also required to add an additional ORN relation that restricts the u-turns between two dual carriage edges.

However, it is not easy to define a single ORN node for connecting all ORN edges in a complex intersection due to the wide diversity of its internal structure. Figure 12(c) shows an example of ORN subgraph for complex intersection, where the set of ORN nodes are partitioned into two subsets: 1) the subset $\mathcal{V}_{i,C}^*$ of *core* ORN nodes where each ORN edge is connected to another ORN node in the complex intersection, and 2) the subset $\mathcal{V}_{i,B}^*$ of *boundary* ORN nodes having at least one ORN edge that connects to an ORN node outside the complex intersection. For example, $\mathcal{V}_{i,C}^* = \{v_{i,1}^*\}$ and $\mathcal{V}_{i,B}^* = \{v_{i,2}^*, v_{i,3}^*, v_{i,4}^*, v_{i,5}^*\}$ in Figure 12(c). In order to add a new ORN edge regardless of the internal structure, the SG scheme first adds a new boundary ORN node $v_{i,6}^*$, and then add a new (red dashed) ORN edge that directly connects this new node with every

Table 4. Statistical description of NLM and ORN in Yeouido.

Road network	Spatial extent	Number of nodes	Number of edges	Total road length
NLM	3.5 km x 2.8 km	177	434	124.74 km
ORN (Pruned)	3.5 km x 2.8 km	590	1005	140.67 km

other boundary ORN node. To reflect a turn restriction between a new ORN edge and an existing ORN edge, a new relation should be inserted into the ORN similarly to the previous two cases.

6. Numerical Results

In this section, we present the numerical results of the RNC between ORN and NLM at Yeoui-do island, Seoul, Korea: The former is extracted from the XML file at the official OSM website[44] and the latter is a shape file downloaded from the Korean ITS website[4]. Both road networks are imported to PostgreSQL database for the RNC [45]. Table 4 shows the statistical information on the area, the number of nodes, road segments, and the total length of road networks.

6.1. The Existing RNM Schemes

In this paper, the proposed AP scheme is compared with three existing node matching schemes, as follows:

- **Nearest first matching (NFM)**: In the NFM, the Euclidean distance between each NLM and ORN node pair that is within a distance threshold (100 m) is initially stored in a priority queue. At each step, the matching (n_i^*, v_j^*) with the smallest Euclidean distance in the priority queue is chosen, and then all remaining matchings with either NLM node n_i^* or ORN node v_j^* are removed from the priority queue.
- **Round-trip walk matching (RWM)** [28]: Given an initial matching, the RWM check the topological consistency of the matching through the following three steps: First, it extracts the corresponding ORN node v_j of each neighbor NLM node $n_j \in \mathcal{N}(n_i) \backslash n_i$. Second, for each corresponding ORN node v_j, it examines the topological consistency by checking whether the corresponding ORN node v_i of NLM node n_i is also its neighbor ORN node or not. Finally, the ratio of the topologically inconsistent neighbor node is stored in a priority queue so that an NLM node with the highest topological consistency is extracted first for the final matching.
- **RWM with DBSCAN clustering (RWM-DC)**: Since both NFM and RWM are 1:1 node matching, they do not account for the LoD difference at a complex intersection. To mitigate this problem, the RWM-DC scheme combines the RWM with a clustering algorithm called the DBSCAN [40,42].

Given all pairs of matched NLM and ORN nodes, we use the score-based matching (SM) for the edge matching of the above three schemes [37]. The SM first computes a discrete similarity score based on multiple independent measures, i.e. the Hausdorff distance[39], orientation[31,39], mean perpendicular distance, and the nodal degree of endpoint nodes [28], and then chooses a pair with the highest score.

In our AP scheme, the threshold δ in section 4.1 is chosen to the maximum width of the general highway and local road in Korea (34 m) [3].

6.2. RNM Results

In this section, we compare the RNM results of our AP scheme with those of three other RNM schemes. In section 4.3, the matching result can be either CM, IM, PM, or MM. If we think of the RNM result as a binary classification, the CM can be interpreted as true positive, and the IM and PM as false positive. On the other hand, if we look at how a true

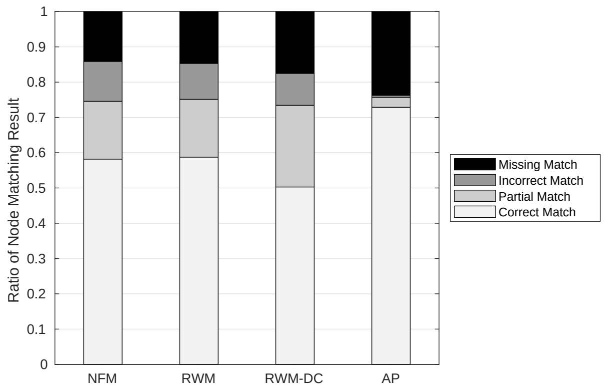


Figure 13. Ratio of node matching results

ORN subgraph is matched to which NLM object, we can classify the matching result into three different cases, as follows: First, a matching scheme successfully finds the NLM object that corresponds to the true ORN subgraph, the matching result becomes CM. Second, if it fails to find the right NLM object corresponding to the true ORN subgraph, the matching result is classified into *failed match (FM)*, which can be interpreted as false negative: The FM can be further partitioned into PM, IM, and MM. Third, there is an exceptional case of binary classification, where the true ORN subgraph does not exist due to the errors in the OSM crowdsourcing process. Denoting the cardinality of type-m matching result by $|\mathcal{M}(m)|$, the precision, recall, and F1-score of matching result can be defined as follows:

$$Precision = \frac{|\mathcal{M}(CM)|}{|\mathcal{M}(CM)| + |\mathcal{M}(IM)| + |\mathcal{M}(PM)|}, \tag{4}$$

$$Recall = \frac{|\mathcal{M}(CM)|}{|\mathcal{M}(CM)| + |\mathcal{M}(FM)|}, \tag{5}$$

and

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{6}$$

respectively.

6.2.1. Node Matching Results

Figure 13 shows the ratio of node matching results against the RNM schemes. We first observe that the proposed AP scheme can achieve an outstanding CM ratio of 0.73 at least 14.1 percent higher than the other RNM schemes: Its (CM, PM, IM, MM) ratio is (0.73, 0.028, 0.006, 0.237). The NFM and RWM schemes that do not support node clustering show almost similar RNM performance: The (CM, PM, IM, MM) ratios of NFM and RWM schemes are (0.582, 0.164, 0.113, 0.141) and (0.588, 0.164, 0.102, 0.147), respectively. The inaccurate node clustering of RWM-DC degrades the CM ratio to 0.503 while increasing the PM and MM ratios to 0.232 and 0.175, respectively. The excellent node clustering performance of AP scheme originates from its low false positive ratio of 0.028, which is at least 8.29 times smaller than those of the other RNM schemes. Furthermore, the AP scheme has the lowest IM ratio of 0.006 while those of the other RNM schemes are at least 0.09. Since the node matching is performed sequentially for each NLM node, an IM of the previous NLM node may block the CM of a subsequent NLM node, which can significantly reduce the CM ratios of the other RNM schemes. The only problem with the AP scheme is its relatively high MM ratio, which will be addressed in section 6.3.

Figure 14 shows the precision, recall, and F1-score of node matching in the NFM, RWM, RWM-DC, and AP schemes. It can be seen that the precision, recall, and F1-score of the AP scheme are at least 26.7, 17.1, and 21.7 percent higher than the other

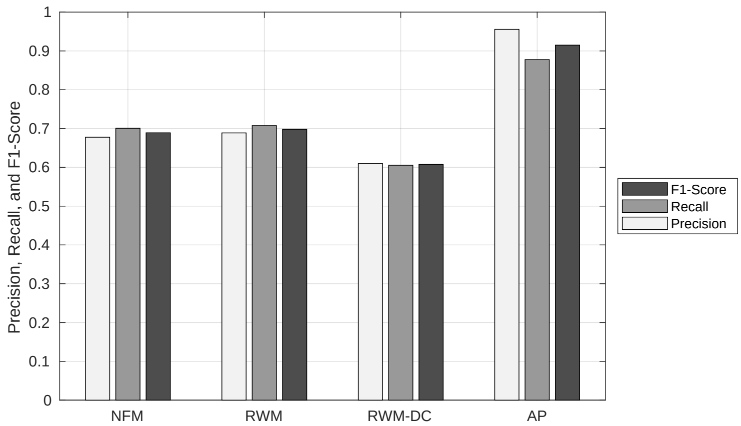


Figure 14. Precision, recall, and F1-score of node matching

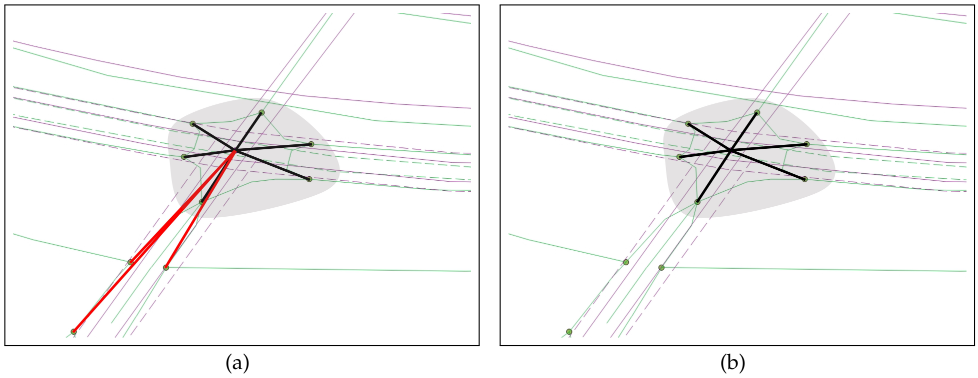


Figure 15. Example of node clustering in the (a) RWM-DG and (b) AP schemes

RNM schemes, respectively. Similar to the ratio of node matching result, NFM and RWM schemes show a similar precision, recall, and F1-score: The difference in their performance is within 1.1 percent. The RWM scheme shows the lowest precision, recall, and F1-score due to its inaccurate node clustering. For example, the node clustering results of RWM-DC and AP schemes are shown in Figures 15(a) and 15(b), respectively, for the same complex intersection in the shaded region. While the AP scheme extracts the exact ORN nodes for the complex intersection, the RWM-DC scheme cannot distinguish three red ORN nodes belonging to minor intersections.

To summarize, the proposed AP scheme achieves an excellent node matching performance, in terms of precision, recall, and F1-score, compared with the existing three RNM schemes.

6.2.2. Edge Matching Results

In this section, the edge matching performance of the AP scheme is compared with those of three existing RNM schemes in section 6.1.

Figure 16 shows the ratio of edge matching results against the RNM schemes. We observe that the proposed AP scheme shows an excellent edge matching performance compared with the other RNM schemes: It has the highest CM ratio of 0.873 (at least 32 percent higher than the others), the lowest false positive ratio of 0.05 (at least 12.7 percent lower than the others), and the lowest MM ratio of 0.076 (at least 19.4 percent lower than the others). This outstanding performance of AP scheme comes from its highly accurate node clustering at a complex intersection that minimizes both PM and IM ratios, which restricts the propagation of false positive in the subsequent edge matching. On the other hand, an inaccurate node matching of three RNM schemes results in a high MM ratio of edge matching. This is because, in a generic road network with limited nodal degree,

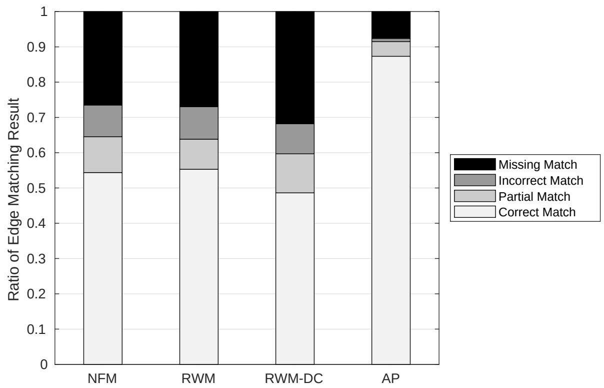


Figure 16. Ratio of edge matching results

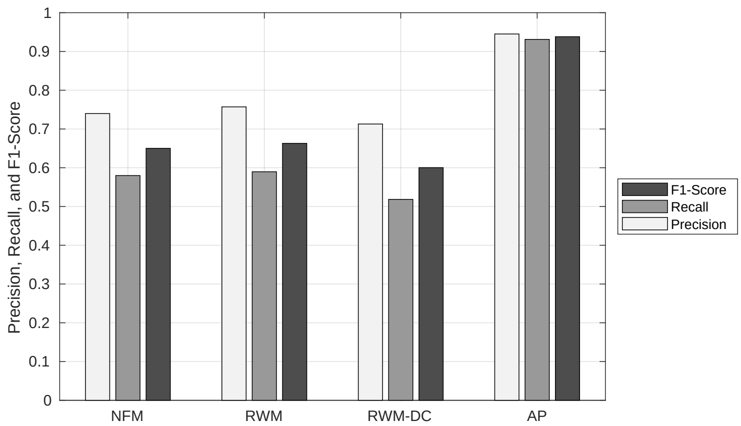


Figure 17. Precision, recall, and F1-score of edge matching.

a change in the endpoints of ORN edge leads to a non-existent ORN edge with high probability.

Figure 17 shows the precision, recall, and F1-score of the edge matching against the RNM schemes. The AP scheme achieves superior edge matching performance with at least 18.8, 34.1, and 27.5 percent higher precision, recall, and F1-score, respectively, than the other RNM schemes. We also observe that the high MM ratio of three existing RNM schemes significantly degrades their recall performance.

From these results, we demonstrate that the proposed AP scheme can also achieve an outstanding edge matching performance compared with the existing RNM schemes.

6.3. RNC Results

In this section, we investigate how our APSG scheme can further improve the matching performance of AP scheme. Table 5 lists the number of CM, PM, IM, and MM results of AP and APSG schemes at Yeoui-do island consisting of 177 NLM nodes and 434 NLM links. By adding ORN objects, the APSG scheme further improves the node matching performance of AP scheme: The number of CM results is increased by 41, while the number of MM results is reduced by 42. As a result, it can increase the recall by 8.29 percent while slightly improving the precision by 0.49 percent. The APSG scheme also improves the edge matching performance compared with AP scheme: It improves both the precision and recall of AP scheme by 1.8 and 3.19 percent, respectively.

In Table 5, we also found the limitation of our APSG scheme in an exceptional node matching where an MM result of AP scheme becomes an IM result by the APSG scheme. The shaded region in Figure 18 shows the complex intersection consisting of two nodes in both road networks. The NLM interprets this complex intersection as the combination of two intersections: n_i connects a road with an underpass and n_j connects three NLM links. On the other hand, the ORN interprets it as a single intersection with

Table 5. Number of matching results $|\mathcal{M}(\cdot)|$ in AP and APSG schemes

Number of Matches	Node Matching					Edge Matching				
	CM	PM	IM	MM	FM	CM	PM	IM	MM	FM
AP	129	5	1	42	18	379	18	4	33	28
APSG	170	5	2	0	7	418	12	4	0	16

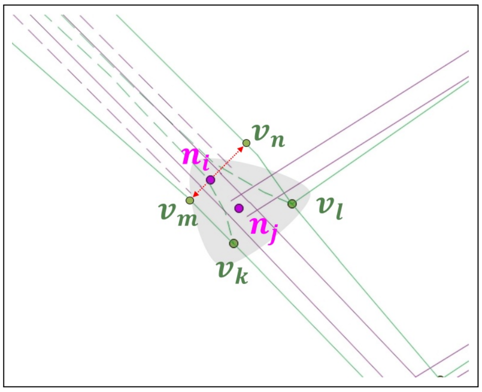


Figure 18. IM case in the APSG.

ORN nodes v_k and v_l interconnecting a dual carriage road, a road, and an underpass. This difference in the interpretation of road objects leads to 2:2 node matching which cannot be addressed by our APSG scheme: In the AP scheme, the matching results for NLM nodes n_i and n_j are MM and PM, respectively. The APSG scheme projects NLM node n_i onto the ORN nodes v_m and v_n in dual carriage road, which changes the matching result to IM.

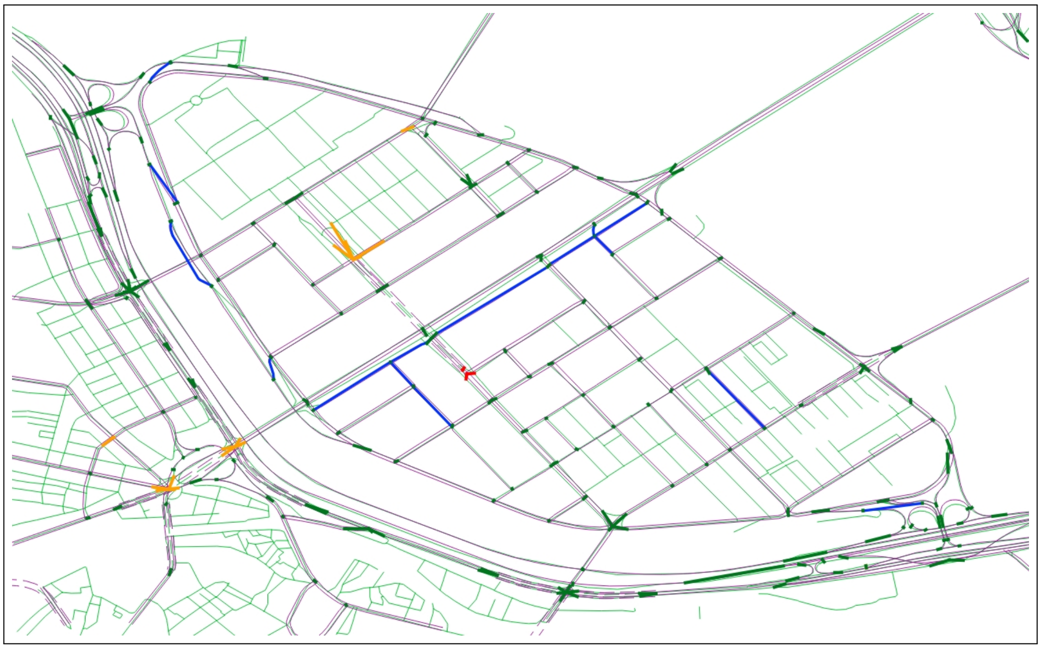
Finally, the matching results of our APSG scheme at Yeoui-do island are shown in Figure 19, where Figures 19(a) and 19(b) illustrate the node and edge matching results, respectively. The blue subgraph represents the new subgraph added to the ORN by the APSG scheme. In addition, the thick dark green, orange, and red lines indicate the CM, PM, and IM results, respectively, between NLM and ORN objects. We can see that the proposed APSG scheme achieves outstanding node and edge matching performance.

7. Conclusions

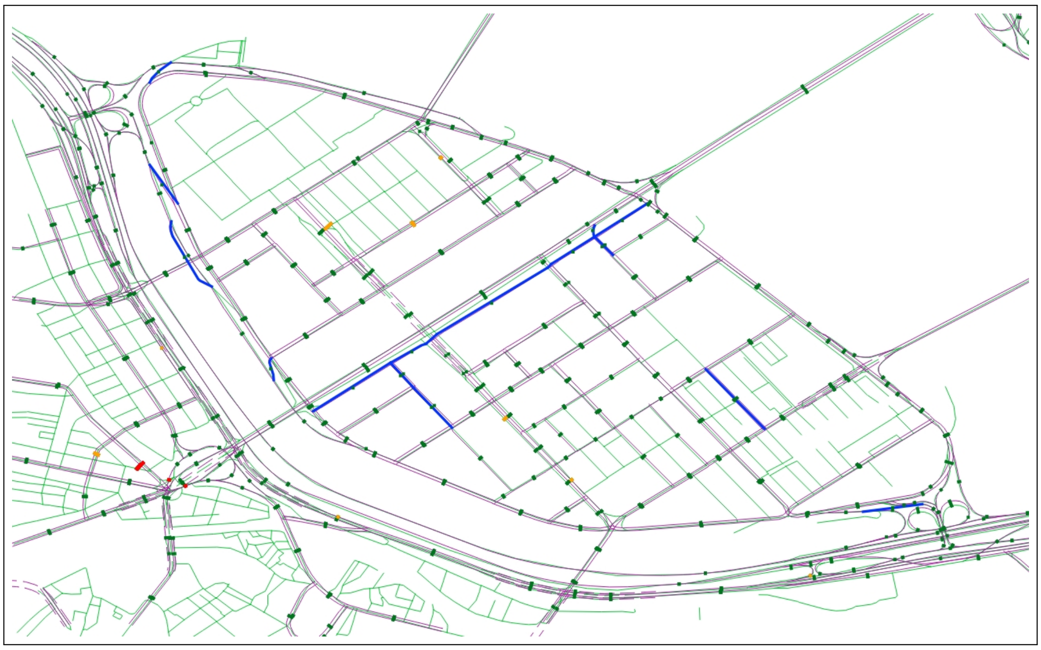
This paper presents the APSG approach to the conflation between administrative and voluntary road networks. The AP scheme addresses the LoD problem of complex intersection through the partition of map area, extraction of candidate ORN subgraph, and aggregation to a supernode. For the unmatched NLM subgraph, the SG scheme sequentially inserts an ORN object while satisfying the connectivity with the matched NLM subgraph by AP scheme. The numerical results show that our APSG scheme achieves an outstanding node and edge matching performance in terms of the precision, recall, and F1-score, compared with the existing RNM schemes.

Appendix A Transient Curve of Projection Boundary

In this appendix, we identify the transient curve of projection boundary around a vertex in an intersection area. Figure A1 shows two examples of projection boundary in area $A_{jm}(n_i)$, where vertex $n_{ji}(p)$ connects two NLM line segments $l_{ji}(p)$ and $l_{ji}(p+1)$ in one projection side, and NLM line segment $l_{im}(q)$ is common on the other projection side. Starting from NLM node n_i, the projection boundary is the bisector b_1 of the angle created by $l_{ji}(p+1)$ and $l_{im}(q)$, and is illustrated by the blue dotted line in both examples. It is clear that every point on this projection boundary should have the same projection distance to l_{ji} and l_{im}, e.g. $d_{P,1} = d_{P,2}$. Our goal is to determine the point where

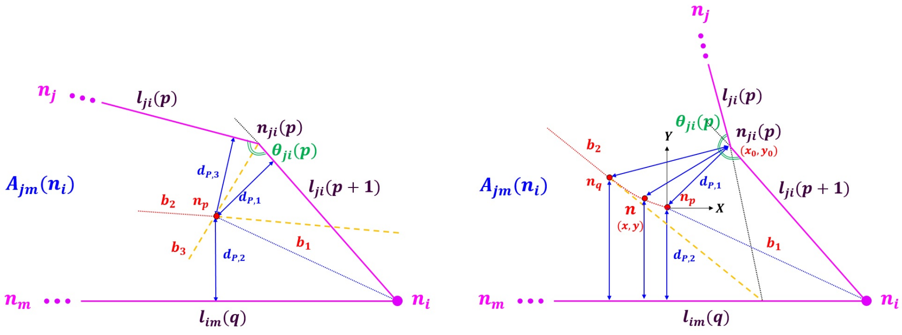


(a) Node matching results of APSG scheme



(b) Edge matching results of APSG scheme

Figure 19. Matching results of APSG scheme



(a) $\theta_{ji}(p) < 180°$ (b) $\theta_{ji}(p) > 180°$

Figure A1. Construction of projection boundary in map area $A_{jm}(n_i)$

the projection boundary deviates from b_1 and find the equidistant projection boundary between two NLM line segments $l_{ji}(p)$ and $l_{im}(q)$. Without loss of generality, we examine the projection boundary curve in two different cases: 1) The internal angle of vertex $n_{ji}(p)$ is less than 180° ($\theta_{ji}(p) < 180°$); and 2) It is greater than 180° ($\theta_{ji}(p) > 180°$).

Figure 1(a) shows an example where $\theta_{ji}(p) < 180°$. To find the point where projection boundary deviate from b_1, we draw two additional bisectors that intersect with bisector b_1 at point n_p: bisector b_2 of the angle between $l_{ji}(p+1)$ and $l_{im}(q)$ and bisector b_3 of angle $\theta_{ji}(p)$. At point n_p, the projection distance to NLM line segments $l_{ji}(p)$, $l_{ji}(p+1)$, and $l_{im}(q)$ becomes the same. After point n_p, the projection boundary deviates from b_1 and becomes the red dotted line segment b_2.

When $\theta_{ji}(p) > 180°$ as shown in Figure 1(b), bisector b_2 is similarly obtained from the crosspoint of $l_{im}(q)$ and the extended line of $l_{ji}(p)$. Next, we determine point n_q on bisector b_2 so that its distance to point $n_{ji}(p)$ is equal to the projection distance to $l_{im}(q)$. It is clear that, beyond point n_q, bisector b_2 becomes the projection boundary. The remaining problem is to determine the projection boundary between points n_p and n_q. To address this problem, we first define a Cartesian coordinate whose X-axis crossing at the origin point n_p is parallel to $l_{im}(q)$. We denote the Cartesian coordinate of point n on the transient boundary curve by (x, y). Similarly, the Cartesian coordiantes of point $n_{ji}(p)$ is denoted by (x_0, y_0). Since $y > 0$, the projection distance of point n to $l_{im}(q)$ becomes $y + d_{P,2}$ which must be equal to the distance between points n and $n_{ji}(p)$, i.e.,

$$\sqrt{(x - x_0)^2 + (y - y_0)^2} = y + d_{P,2}. \tag{A1}$$

Finally, the transient curve of projection boundary becomes a *parabola* satisfying the following equation:

$$y = \frac{(x - x_0)^2 + y_0^2 - d_{P,2}^2}{2(y_0 + d_{P,2})}. \tag{A2}$$

References

1. Geography, G. 1000 GIS Applications and Uses – How GIS Is Changing the World. https://gisgeography.com/gis-applications-uses/, accessed on 2021-12-02.
2. Clarke, K.C. Advances in Geographic Information Systems. *Computers, Environment and Urban Systems* **1986**, *10*, 175–184. doi:10.1016/0198-9715(86)90006-2.
3. Korean Ministry of Land Infrastructure and Transport. The Road Extension per Person in Korea is 2 Meters. Technical report, 2019.
4. Korean National Transport Information Center. Node Link Map. http://nodelink.its.go.kr/, accessed on 2021-12-02.
5. Google LLC. Google Map. http://maps.google.com/, accessed on 2021-12-02.
6. Daum Kakao. Daum map. http://map.daum.net/, accessed on 2021-12-02.
7. OpenStreetMap Contributors. OpenStreetMap Wiki. http://wiki.openstreetmap.org/, accessed on 2021-12-02.
8. OpenStreetMap Contributors. OpenStreetMap Wiki - Stats. https://wiki.openstreetmap.org/wiki/Stats, accessed on 2021-12-02.

9. Yuan, M.; Yang, Z.; Niu, S. Study on dynamic route guidance method of vehicle based on travel time reliability. *Proceedings - 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010* **2010**, *1*, 292–295. doi:10.1109/ICACC.2010.5487007.
10. Hajiahmadi, M.; Knoop, V.L.; De Schutter, B.; Hellendoorn, H. Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). IEEE, 2013, number Itsc, pp. 1022–1028. doi:10.1109/ITSC.2013.6728366.
11. Zuo, L.; Zhang, N.; He, Y.; Zhou, T. Research on Dynamic Route Guidance and Navigation System Based on Multi Information Feedback. *Proceedings - 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control, SDPC 2017* **2017**, *2017-Decem*, 421–424. doi:10.1109/SDPC.2017.86.
12. Lin, J.; Yu, W.; Yang, X.; Yang, Q.; Fu, X.; Zhao, W. A Real-Time En-Route Route Guidance Decision Scheme for Transportation-Based Cyberphysical Systems. *IEEE Transactions on Vehicular Technology* **2017**, *66*, 2551–2566. doi:10.1109/TVT.2016.2572123.
13. Yang, Q.; Koutsopoulos, H.N.; Ben-Akiva, M.E. Simulation laboratory for evaluating dynamic traffic management systems. *Transportation Research Record* **2000**, pp. 122–130. doi:10.3141/1710-14.
14. Van Hinsbergen, C.P.; Schreiter, T.; Zuurbier, F.S.; Van Lint, J.W.; Van Zuylen, H.J. Localized extended kalman filter for scalable real-time traffic state estimation. *IEEE Transactions on Intelligent Transportation Systems* **2012**, *13*, 385–394. doi:10.1109/TITS.2011.2175728.
15. Zygouras, N.; Zacheilas, N.; Kalogeraki, V.; Kinane, D.; Gunopulos, D. Insights on a scalable and dynamic traffic management system. EDBT 2015 - 18th International Conference on Extending Database Technology, Proceedings, 2015, number March, pp. 653–664. doi:10.5441/002/edbt.2015.65.
16. Korean Land Information Platform. Precise Road Map. http://map.ngii.go.kr/ms/pblictn/preciseRoadMap.do, accessed on 2021-12-02.
17. Korean National Transport Information Center. Public Traffic Information Service. http://openapi.its.go.kr/, accessed on 2021-12-02.
18. Fairhurst, R. Potlatch 2. https://wiki.openstreetmap.org/wiki/Potlatch_2, accessed on 2021-12-02.
19. Scholz, I.; Stöcker, D. JOSM. https://josm.openstreetmap.de/, accessed on 2021-12-02.
20. Pavlenko, A. Mapnik. http://mapnik.org/, accessed on 2021-12-02.
21. Ramm, F.; Topf, J. Tirex. https://wiki.openstreetmap.org/wiki/Tirex, accessed on 2021-12-02.
22. Quinion, B.; Hoffmann, S.; Metten, M.T. Nominatim. https://nominatim.org/, accessed on 2021-12-02.
23. Luxen, D.; Vetter, C. Real-time routing with OpenStreetMap data. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '11* **2011**, p. 513. doi:10.1145/2093973.2094062.
24. Gearhart, D.; Knisely, G.; Kreiser, K.; DiLuca, K.; Nesbitt, D. Valhalla. https://github.com/valhalla, accessed on 2021-12-02.
25. SinghSehra, S.; Singh, J.; Singh Rai, H. Assessment of OpenStreetMap Data - A Review. *International Journal of Computer Applications* **2013**, *76*, 17–20, [1309.6608]. doi:10.5120/13331-0888.
26. Rosen, B.; Saalfeld, A. Match criteria for automatic alignment. Proceedings of 7th international symposium on computer-assisted cartography (Auto-Carto 7), 1985, pp. 1–20.
27. Lüscher, P.; Burghardt, D.; Weibel, R. Matching road data of scales with an order of magnitude difference. *Proc. XXIII International Cartographic Conference, Moscow, Russia* **2007**, pp. 1–11.
28. Filin, S.; Doytsher, Y. Detection of corresponding objects in linear-based Map conflation. *Surveying and land information systems* **2000**, *60*, 117–128.
29. Krishnamurthy, G.; Devarajan, V.; Dragan, I. Nonrigid conflation for vector datasets using EM algorithm and mixture of Gaussian approach. *International Geoscience and Remote Sensing Symposium (IGARSS)* **2010**, pp. 3964–3967. doi:10.1109/IGARSS.2010.5654391.
30. Yang, B.; Luan, X.; Zhang, Y. A pattern-based approach for matching nodes in heterogeneous urban road networks. *Transactions in GIS* **2014**, *18*, 718–739. doi:10.1111/tgis.12057.
31. Walter, V.; Fritsch, D. Matching spatial data sets: a statistical approach. *International Journal of Geographical Information Science* **1999**, *13*, 445–473. doi:10.1080/136588199241157.
32. Zhang, M.; Shi, W.; Meng, L. A generic matching algorithm for line networks of different resolutions. *Proceedings of 8th ICA Workshop on Generalisation and Multiple Representation* **2005**, pp. 1–8.
33. Yang, B.; Zhang, Y.; Luan, X. A probabilistic relaxation approach for matching road networks. *International Journal of Geographical Information Science* **2013**, *27*, 319–338. doi:10.1080/13658816.2012.683486.
34. Zhang, J.; Wang, Y.; Zhao, W. An improved probabilistic relaxation method for matching multi-scale road networks. *International Journal of Digital Earth* **2017**, *11*, 635–655. doi:10.1080/17538947.2017.1341557.
35. Zhang, M. Methods and implementations of road-network matching. *PhD Dissertation* **2009**.
36. Fan, H.; Yang, B.; Zipf, A.; Rousell, A. A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data. *International Journal of Geographical Information Science* **2016**, *30*, 748–764. doi:10.1080/13658816.2015.1100732.
37. Hacar, M.; Gökgöz, T. A new, score-based multi-stage matching approach for road network conflation in different road patterns. *ISPRS International Journal of Geo-Information* **2019**, *8*. doi:10.3390/ijgi8020081.
38. Xiong, D.; Sperling, J. Semiautomated matching for network database integration. *ISPRS Journal of Photogrammetry and Remote Sensing* **2004**, *59*, 35–46. doi:10.1016/j.isprsjprs.2003.12.001.

39. Volz, S. An Iterative Approach for Matching Multiple Representations of Street Data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **2006**, *36*, 101–110.
40. Yang, B.; Luan, X.; Li, Q. Generating hierarchical strokes from urban street networks based on spatial pattern recognition. *International Journal of Geographical Information Science* **2011**, *25*, 2025–2050. doi:10.1080/13658816.2011.570270.
41. Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S.N.; Kendall, D. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*; Wiley Series in Probability and Statistics, John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2000. doi:10.1002/9780470317013.
42. Ester, M.; Kriegel, H.p.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. 2nd International Conference on Knowledge Discovery and Data Mining, 1996, pp. 226–231, [10.1.1.71.1980]. doi:10.1.1.71.1980.
43. Korean Ministry of Land Infrastructure and Transport. Spatial Information Portal. https://www.nsdi.go.kr, accessed on 2021-12-02.
44. OpenStreetMap Contributors. OpenStreetMap. https://www.openstreetmap.org/, accessed on 2021-12-02.
45. PostgreSQL Global Development Group. PostgreSQL. http://www.postgresql.org, accessed on 2021-12-02.