

Article Identifying Complex Junctions in a Road Network

Jianting Yang ¹, Kongyang Zhao ¹, Muzi Li ², Zhu Xu ^{1,3,*} and Zhilin Li ^{1,3}

- ¹ Faculty of Geosciences and Environment Engineering, Southwest Jiaotong University, Chengdu 611756, China; jantty_y@my.swjtu.edu.cn (J.Y.); zhaoky@my.swjtu.edu.cn (K.Z.); lszlli@polyu.edu.hk (Z.L.)
- ² Faculty of Geography and Resource Sciences, Sichuan Normal University, Chengdu 610066, China; lmz@sicnu.edu.cn
- ³ National Collaborative Innovation Center for Rail Transport Safety, Southwest Jiaotong University, Chengdu 611756, China
- * Correspondence: xuzhu@swjtu.edu.cn

Abstract: Automated generalization of road network data is of great concern to the map generalization community because of the importance of road data and the difficulty involved. Complex junctions are where roads meet and join in a complicated way and identifying them is a key issue in road network generalization. In addition to their structural complexity, complex junctions don't have regular geometric boundary and their representation in spatial data is scale-dependent. All these together make them hard to identify. Existing methods use geometric and topological statistics to characterize and identify them, and are thus error-prone, scale-dependent and lack generality. More significantly, they cannot ensure the integrity of complex junctions. This study overcomes the obstacles by clarifying the topological boundary of a complex junction, which provides the basis for straightforward identification of them. Test results show the proposed method can find and isolate complex junctions in a road network with their integrity and is able to handle different road representations. The integral identification achieved can help to guarantee connectivity among roads when simplifying complex junctions, and greatly facilitate the geometric and semantic simplification of them.

Keywords: generalization; scale; topological relationship; road network; pattern recognition

1. Introduction

Map generalization is a special variant of spatial modelling, which involves deriving a less detailed spatial representation from a detailed one and was traditionally performed manually by cartographers [1]. In digital environment, generalization of spatial data is the basis of constructing multiscale representations in spatial databases and is highly desired to be automated [2,3].

Generalization demands highly intelligent processing of spatial data and is very difficult to automate in general [1,4]. Great effort has been made to its automation in the past half century [5]. In the last three decades, significant progress has been made on transforming data representation [6,7], enriching data semantics [5,8], and developing intelligent algorithms [9–11]. Some of the achievements have been applied in data production [12,13].

Generalization of road network data has been among the greatest concern of generalization researchers and practitioners due to the importance of road data and the difficulty involved [6,14,15]. Lots of research have been carried out on road selection [16–18], analysis of patterns in road networks [19–21], and techniques for enriching road data semantics [22–24].

This study is particularly concerned with the identification of complex junctions in a road network. Complex junctions are where roads meet and join in a complicated way, and properly simplify them is a key issue in road network generalization [25–27]. Unlike simple planar junctions, such as crossings and T-junctions where major roads directly intersect by themselves, complex junctions can be planar ones with slip roads for smooth turning and can be grade-separated interchanges with ramps for bridging gaps in height.



Citation: Yang, J.; Zhao, K.; Li, M.; Xu, Z.; Li, Z. Identifying Complex Junctions in a Road Network. *ISPRS Int. J. Geo-Inf.* 2021, *10*, 4. https:// dx.doi.org/10.3390/ijgi10010004

Received: 13 October 2020 Accepted: 21 December 2020 Published: 24 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/ licenses/by/4.0/). What is more, they can be very complex in structure and twisted in shape. To simplify the representation of such junctions without breaking the connectivity among the relevant roads, it is necessary to first identify them precisely in the network of roads [25,26].

Identifying complex junctions in a road network has been shown to be quite challenging by previous studies due to the difficulties in definitely characterizing them and ascertaining their boundary in the network [21,27]. Mackaness and Mackechnie [25] characterized complex junctions as where nodes of road intersection are dense and proposed a density analysis approach to detecting them. By carefully tuning the density threshold, complex junctions as a collection of densely clustered nodes could be identified, although street blocks might be wrongly included as well. Savino et al. [23,26] noticed node density did not work in some cases. They used statistics of redundant circles (or loops in topological terms) to characterize different forms of complex junctions and tried to detect them by searching for such circles in a road network. Touya [19] applied the method of Grosso [28] to detect highway interchanges. Their method is similar to that of Mackaness and Mackechnie [25]. But, instead of ordinary intersection nodes, they used fork and y-nodes which were considered more characteristic of highway interchanges. Similarly, Zhou and Li [27] took a complex junction as a collection of simple junctions, including T-, fork and crossing junctions etc. Like Savino et al. [23,26], they attempted to characterize and detect different forms of complex junctions with their compositional statistics of the simple junctions.

As can be seen, existing methods are all based on statistical analysis of geometric and topological properties of complex junctions. They therefore suffer from the problem of threshold setting. Real world complex junctions vary greatly in node density, enclosed shape and size, and structural form. Thresholds of them inevitably vary from area to area and city to city due to spatial heterogeneity. What is more, there was no definition of the boundary of a complex junction included in those methods. Therefore, they cannot identify complex junctions with their integrity, as the boundary would change with the thresholds used. Further, junction representation is scale-dependent. A change from single-line representation of roads to dual-line one can make a complex junction very different in geometry and topology. Currently, no known methods deal with such scale-dependence explicitly.

We took a structural pattern approach to identifying complex junctions. According to the domain knowledge of junction design [29,30], complex junctions are those having secondary roads, such as slip roads and ramps, to help primary roads join smoothly. In that sense, a complex junction can be structurally defined in topological terms no matter how complicated it is. That definition opens a gate to a straightforward way to identify complex junctions.

The rest of this paper is organized as follows. From the perspective of road and junction design, Section 2 formalizes a structural pattern of complex junctions, which characterizes them in a scale-independent way and defines the exact topological boundary of them. A simple algorithm based on that pattern is then proposed for identifying complex junctions in Section 3. In Section 4, experimental test of applying the method to a challenging case of real-world road network and a comparison between this method and a conventional one are presented. The influences of errors in input data on the identification are discussed in Section 5. In Section 6, scale-dependence during identifying complex junctions and relation between primary-secondary roads and road network generalization are discussed. Finally, conclusions are drawn in Section 7.

2. Topological Boundary of a Complex Junction

As analyzed above, more essential characteristics of complex junctions need to be revealed to provide more effective basis for identifying them. We turn to the domain knowledge of road and junction design and formulate a topological definition of the complex junction. According to the principles of road and junction design [29–32], two types of roads can be differentiated in a road network system, the primary ones and the secondary ones. Primary roads (or major roads in the common sense) serve to link places distributed over geographic space. In contrast, secondary roads such as slip roads and ramps serve to link primary roads and help them join more smoothly than in the direct intersection way to increase throughput at their joints. In other words, secondary roads are dependent on primary roads. These two kinds of roads are illustrated in Figure 1 with a schematic road network T. Roads in T are labeled for easy reference. Those in light or dark grey, i.e., a–j, are primary ones, and those in black, i.e., k–p, are secondary ones. With the roads so classified, the composition of a complex junction becomes conceptually clear. Two dashed circles are drawn around the two complex junctions in Figure 1 to make their composition more explicit. It should be clear that a complex junction consists of a group of secondary roads together with the relevant primary roads that make the former connected to the rest roads.



Figure 1. A schematic road network T for illustrating primary roads, secondary roads, and complex junctions.

The above conception of complex junctions needs to be formalized to be rigorous and made computable. To do so, we transform the ordinary segment-based representation of a road network to a stroke-based representation, and further on to a dual topological representation (or dual network for brevity). The concept of road stroke was proposed by Thomson and Richardson [33], which refers to a concatenation of road segments according to the principle of good continuity. Intuitively, instead of road segment it is road stroke that naturally corresponds to the concept of road in everyday use. The road stroke has been widely accepted and used in road network analysis since its proposed [18,19,32]. In fact, the labels in Figure 1 are those of road strokes instead of road segments. Therefore, the road network T in Figure 1 used the stroke-based representation.

The dual network of a stroke-based road network is widely used in topological analysis of road networks [34–36]. In the dual network, a road is turned into a node and there is an edge between two nodes if the two roads they represent intersect. To illustrate, the dual network of road network T shown in Figure 1 is presented in Figure 2. As can be seen, each road in Figure 1 is now a node in Figure 2. In addition, there is an edge between two roads (i.e., nodes) if they intersect. In Figure 1, the intersection relationship among roads is indicated by the line colors. Generally, if two lines intersect, the two roads they represent intersect. Exceptions are the cases in which a light grey line intersects with a dark grey line. In other words, road a does not intersect either road f or g, and neither does b. Therefore, the dual network of T is like that shown in Figure 2, where there are not edges between a and f, a and g, b and f, and b and g.



Figure 2. The dual network of the schematic road network T and the topological boundary of complex junctions depicted in Figure 1.

In the dual network of the stroke-based representation of a road network, a complex junction can be defined as a group of nodes of secondary road isolated from the rest of the road network by the relevant nodes of primary road, which the former serves to connect or depend on. That is clear in Figure 2, where nodes of secondary road are purposely greyed to indicate they are 'bounded' by the relevant primary roads. Two dashed ellipses are drawn to make explicit the topological boundaries of the two complex junctions, which otherwise might not be apparent. As can be seen, nodes of secondary road, say (k, l, m, n), are blocked from the rest of the network by the relevant nodes of primary road, i.e., (a, b, f, g). In other words, the rest of the road network cannot be reached from the secondary roads without passing through the primary roads they serve, and the latter constitute the topological boundary of a complex junction. Reversely, the interior secondary roads of a complex junction cannot be reached from the roads external to the junction without passing through its boundary primary roads.

The topological boundary of a complex junction provides an approach to generalize the complex junction by properly discarding its interior secondary roads. Figure 3 shows an example to illustrate this generalization process by simplifying the complex junctions in the schematic road network T in Figure 1. The complex junctions are generalized by discarding the nodes of secondary road, i.e., (k, l, m, n, o, p), but preserving the connectivity between the boundary primary roads served by the discarding ones.



Figure 3. (**a**–**c**) The process of generalization of complex junction by discarding the interior secondary roads.

It is worth emphasizing that the above definition is a purely topological one and is thus inherently scale-independent. That means it works no matter how complicated or simple the complex junction is. That also means it works no matter a road is represented by a single line, dual lines or even multiple lines. The case of dual-line representation is illustrated in Figure 2. For the pairing lines a and b of a dual carriageway in Figure 1, we can logically take them as a composite node A in the dual network as shown in Figure 2. Similarly, f and g make another composite node F. In addition, as the definition includes an exact topological boundary of a complex junction, it can thus be applied to identify complex junctions with their integrity, as explained below.

3. Algorithm of Identifying Complex Junctions

The process of finding and isolating a complex junction in the dual network starts with a seed node of secondary road (sd) and then continues by expanding the found part till the boundary is reached. More specifically, the algorithm proceeds as follows:

(1) Global initialization: concatenate road segments into strokes. Derive the primaryor-secondary type of roads and pair counterparts. By transform to the dual network, take the set of nodes of secondary road (SG). Prepare an empty set (J) to collect complex junctions to be found. If SG is empty, return J and quit. Otherwise, mark all nodes in SG as unprocessed. The global initialization of the schematic road network T shown in Figure 4. SG corresponds to the set of secondary roads, i.e., (k, l, m, n, o, p).



Figure 4. (a,b) The global initialization of the schematic road network T.

(2) Seeding: pick an unprocessed secondary road sd from SG. If it is an appropriate one (i.e., it connects two different roads instead of the counterparts of one road), proceed to the next step. Otherwise, mark it as processed in SG. Try others till an appropriate one is picked. Or otherwise, return J and quit, as there are no more unprocessed nodes in SG.

(3) Local initialization: prepare two empty sets, one for primary roads (P) and the other for secondary ones (S), to collect a complex junction's boundary nodes and interior nodes respectively. Put sd into S. By picking node k as sd as shown in Figure 5a, S contains a secondary road in its collection as (k) after the local initialization.



Figure 5. The seed node of secondary road k (a) and its S expanding (b).

(4) S expanding: for each unexpanded appropriate node (x) in S, fetch all of its neighboring nodes from the dual network and put each of them into either P or S according to its road type, and then mark x as expanded. If new nodes are added to S, recursively expand the new ones till some primary roads are reached. In the process of recursive expanding, mark those in S as expanded but leave those in P as unexpanded. Figure 5b shows an example of the process of S expanding by expanding node k in the dual network. The nodes involved in the process of S expanding are indicated by their color as shown in Figure 2, and the black edges correspond to links between the processing nodes, so are those below in Figures 6–8. By S expending of node k, P collect the node a, and f.



Figure 6. (a,b) The multi-line processing of node of primary road f and a.



Figure 7. The P expanding of node of primary road f (a), g (b), a (c), b (d).



Figure 8. A complex junction in the dual network for illustrating its collections of nodes. P contains the nodes of the complex junction's boundary as (a, b, f, g), and S contains the interior nodes of secondary road as (k, l, m, n).

(5) Multi-line processing: for each node x in P that is a counterpart of a dual-line or multi-line road r, put r's other counterparts into P if they are not in yet, and leave them as unexpanded. Figure 6 shows an example of multi-line processing by collecting the

counterparts, g and b, of node f and a in the dual network. By the multi-line processing, P contains the nodes of complex junction's boundary as (a, b, f, g).

(6) P expanding: prepare three sets for temporary use: first for the unexpanded appropriate nodes of secondary road neighboring with the processing one (T), second for collecting primary roads involved in the process of expanding (TP) and third for collecting secondary roads during expanding (TS). For each unexpanded node x in P, clear T, fetch all of x's neighboring unexpanded appropriate secondary roads, and put them into T. For each unexpanded node (y) in T, clear TS and TP, recursively expand y till some primary roads are reached, and put the roads obtained in the process of expanding (including y) into TS or TP respectively. If any node in TP is one in P other than x, put all those in TS and TP into S and P respectively if they are not there yet, and mark those added into S as expanded and leave those added into P as unexpanded. If x is connected to any node in S, mark it as expanded. Otherwise, remove x from P. The P expanding of node f, g, a, b in the dual network is shown in Figure 7. By P expanding, S contains its interior nodes as (k, l, m, n).

(7) Local recursion: repeat step 5–6 till all nodes in both P and S are expanded;

(8) Collecting results: now a complex junction j has just been identified, for which P contains its boundary primary roads and S contains its interior secondary roads. Mark the nodes in S as processed in SG, put j into J, and clear sd. Figure 8 shows an example of a complex junction with its node collections.

(9) Global recursion: repeat step 2–8 to find more complex junctions in the road network. Regarding the algorithm, some explanation may be helpful. It is worth noting that step 4 expands outward from the seed node of a complex junction, while step 6 first expands inward from its boundary nodes of primary road, and then outward to other boundary nodes. Additionally, note the simple processing in step 5 makes the algorithm applicable to the dual-line and multi-line representations of roads in addition to the single-line one. What is more, the result obtained in step 8 is stroke-based and can be easily transformed to a segment-based representation which specifies the boundary and interior of a complex junction in terms of road segment and segment node. As can be seen, the overall process is basically a traversal of the dual network and is therefore highly efficient. A demonstration of the algorithm will be given in the next section.

4. Experimental Tests and Result Analysis

A challenging road network of real world was used to test the proposed method. The test dataset covers a 6 km by 4 km area of city San Diego, California, US, which is shown in Figure 3. As can be seen, the roads were represented as segments in the dataset as indicated by their ends, the light grey dots on the map. The dataset used the dual-line representation for dual carriageways. It contained nineteen complex junctions, which were labeled from one to nineteen for easy reference. The rough extents of the complex junctions are circled with the black dotted ellipses on the map. As shown, the complex junctions vary greatly in size and complexity and are embedded in thousands of roads. The dataset is therefore considered challenging and suitable for testing if the proposed method can integrally identify complex junctions embedded in the surroundings, and also if it can deal with both the single-line and dual-line representations of roads.

The data was preprocessed with ESRI ArcGIS Desktop 10.2.2. Stroke assembling, counterpart pairing, and the derivation of the primary-or-secondary type of roads were based on road attributes, mainly road names and height grades, which is an attribute indicate whether two intersected road segments of 2-D plane in database connect with each other. Thanks to the quality of road attribute data, all of the software-based preprocessing turned out to be of high accuracy in terms of percentage of correct inferences, but there were errors. Manual check and error correction were performed after each main step of software-based preprocessing to ensure clean input to the next step.

The results of each main preprocessing step are all shown on the map in Figure 9. About 1000 strokes were obtained from near 3000 segments in the test data. Each of the strokes was uniquely colored on the map to be distinguishable from the surroundings,

although adjacent strokes might be in close colors due to the use of a random coloring scheme. The paired dual lines were indicated by their light grey background on the map. The secondary roads were indicated by their yellow background on the map.



Figure 9. The test dataset and its preprocessing results after manual correction. The end points of the road segments in the raw data are represented by the light grey dots. The nineteen complex junctions to be identified are labeled from one to nineteen. The rough extents of the complex junctions are circled with the black dotted ellipses. The strokes assembled are uniquely colored using a random coloring scheme. The paired dual lines are indicated by their light grey background. The secondary roads are indicated by their yellow background.

Finally, the dual network of the stroke-based representation of the dataset was generated and used for identification. We take complex junction No. 9 as shown in Figure 10 as an instance to demonstrate the transformation and identification processes. Although No. 9's internal structure is not very complicated, its surrounding setting is more complicated than most of the others in the test area. It is next to No. 3 and No. 10, and connects five strokes of primary roads, i.e., (a, b, c, d, e), among which a and b are dual and the other three are single-line.

To simply the demonstration, we have complex junction No.9 and its surroundings clipped to concentrate on. The resulting road network N is shown in Figure 10, where the roads of concern are labeled for reference. Its dual network is shown in Figure 11. As can be interpreted from Figure 10, roads a–n are primary and o–v are secondary. In addition, a and b are dual. So, o–v were greyed, and a and b were grouped by a dotted ellipse in Figure 11. The dashed ellipse doesn't represent information known up to now and is drawn to help visualize the composition of complex junction No.9, which is to be identified.



Figure 10. Road network N around complex junction No.9 clipped from the test data.



Figure 11. Dual network D_N of road network N.

With the dual network D_N , complex junction No.9 can be identified by the algorithm proposed. Table 1 lists the steps of the identification process, in which the states of each step are given.

The result of identification in the test network is shown on the map in Figure 12. For an identified complex junction, its secondary roads are in the same bright color while the participant primary roads are in black as one primary road can participate in more than one complex junctions. The rest roads that are not part of the complex junctions are in light grey.

From the test result, some properties of the algorithm can be observed. Firstly, the algorithm identified all the nineteen complex junctions with their respective integrity. In other words, no complex junctions to be identified were missed; no non-complex-junction parts of the road network were mistakenly included in the identification result; no component roads of any junction were missed; and no non-component roads were mistakenly included in any junction. The integral identification results come from the deterministic nature of the algorithm, which is based on the definite topological boundary of the complex junction. With that said, it is not to claim the algorithm is error-tolerant, but rather it is integrity-maintaining when the input is clean. The influence of errors in input data will be discussed in Section 5.

Secondly, the algorithm turned out to be adaptable as anticipated. It worked for the simply structured complex junctions like No. 12–19, and the complicatedly structured ones like No. 1–11. It worked for the small ones like No. 12–19, the medium ones like No. 1, 2, 4, 5, 8, 9, 10, and 11, and the large ones like No. 3, 6 and 7. It worked for the single-line ones like No. 12–19, the dual-line ones like No. 1, 3 and 7, and the mixture ones like No. 2, 4, 5, 6, 8, 9, 10 and 11.

	S	Р	sd	Т	TS	ТР	SG	J
Global initialization	/	/	/	/	/	/	(o, p, q, r, s, t, u, v)	()
<i>Notes:</i> (1) s, t, u, and v are not eligible for seeding in D_N as each of them connects to one road only;								
(2) without loss of generality, suppose s, t, u, and v are tried first and then p is picked as the seed.								
Seeding	/	/	р	/	/	/	(o, p, q, r, s', t', u', v')	()
Local initialization	(p)	0	р	/	/	/	(o, p, q, r, s', t', u', v')	0
S expanding, p	(p ')	(a, c, d)	р	/	/	/	(o, p, q, r, s', t', u', v')	0
Multi-line processing	(p')	(a, b , c, d)	р	/	/	/	(o, p, q, r, s', t', u', v')	0
P expanding, a	(o , p')	(a ′, b, c, d)	р	(0)	0	0	(o, p, q, r, s', t', u', v')	0
expanding, o	(o' , p')	(a', b, c, d)	р	(o)	(o)	(a, c, d)	(o, p, q, r, s', t', u', v')	()
P expanding, b	(o', p', q, r)	(a', b ', c, d)	р	(q , r)	0	0	(o, p, q, r, s', t', u', v')	0
expanding, q	(o', p', q' , r)	(a', b', c, d, e)	p	(q, r)	(q)	(b, c, e)	(o, p, q, r, s', t', u', v')	0
expanding, r	(o', p', q', r')	(a', b', c, d, e)	p	(q, r)	(r)	(b, c, e)	(o, p, q, r, s', t', u', v')	0
P expanding, c	(o', p', q', r')	(a', b', c' , d, e)	p	0	0	0	(o, p, q, r, s', t', u', v')	0
P expanding, d	(o', p', q', r')	(a', b', c', d' , e)	p	()	0	0	(o, p, q, r, s', t', u', v')	()
Local recursion: proceed to Multi-line processing as there may be new nodes in P .								
Multi-line processing	(o', p', q', r')	(a', b', c', d', e)	р	/	/	/	(o, p, q, r, s', t', u', v')	0
P expanding, e	(o', p', q', r')	(a', b', c', d', e')	p	0	0	0	(o, p, q, r, s', t', u', v')	()
Local recursion: proceed to the next step as all node in S and P are <i>expanded</i> .								
Collecting results	(o', p', q', r')	(a', b', c', d', e')	/	/	/	/	(o' , p' , q' , r' , s', t', u', v')	(No. 9)
Global recursion: proceed to Seeding								
Seeding: return J and quit. As all nodes in SG are processed.								

Table 1. The steps and states of the algorithm in identifying complex junction No.9 in **D**_N (the changed states are in **bold** in each step).



Figure 12. The result of applying the proposed method to the test dataset. For an identified complex junction, its secondary roads are in the same bright color while the participant primary roads are in black as one primary road can participate in more than one complex junctions. The rest roads that are not part of the complex junctions are in light grey.

Thirdly, the algorithm was not affected by geometric complexity of junctions as it is based on the topological relationship among roads. For No. 4 and 11, which are not connected but look very much like one complex junction instead of two, the algorithm did not combine them. For No. 5 and 6, of which the geometric extents are almost overlapping, the algorithm did not mix them up. For No. 6, it correctly gathered the four secondary roads that are quite apart but integrally connect the four primary roads (or six primary road strokes), among which two are dual and the other two are single-line.

A comparison between this method and the conventional one, which defines complex junctions as dense regions of vertices, proposed by Mackaness and Mackechnie in 1999 is conducted in this study. Figure 13 shows the result of complex junction recognitions generated from a hierarchical clustering method [27], which was applied to detect dense regions of vertices in this comparison. By testing the thresholds involved in the clustering approach rigorously, we finally set the Distance, which means the maximum distance between points in a cluster, at 130 m, and the least number of points in a cluster at 4.

It can be seen most of the clusters are outside the targets through comparison with the complex junctions highlighted in Figure 12. In addition to the error detected regions, the complex junctions were identified with several defects. The missed complex junctions like No. 8 and 13, the overfitting ones like No. 4, 12 and 14–19, the other ones with various degree of inexactness, including missing intersections, overfitting or both, like No. 1, 2, 3, 5, 6, 7, 9, 10 and 11. More detail were discussed in the previous studies [25–27].



Figure 13. Complex junction detections using the approach proposed by Mackaness and Mackechnie (1999). The nineteen labels of the complex junctions to be identified are in two grey levels. Dark grey for various degree of inexactness, and light grey for overfitting. For those completely missing identified complex junctions, red dashed ellipses are used to circle the extents of concern.

5. Influence of Errors in Input Data

As can be seen from the above experiment, preprocessing is needed to semantically enrich the raw data and transform the representation of data before the identification, which includes stroke assembling, counterpart pairing, road classification and dual network generation. As we know, this is not special but rather common in map generalization, as spatial data in their ordinary form are generally not suitable to be generalized directly, and data enrichment and representation transformation are regularly required [2].

While the proposed algorithm is integrity-maintaining, dirty input can influence and even spoil the identification. To have some idea of how significant the influence is, we made a trial by letting the identification process run upon the uncorrected preprocessing result. The uncorrected preprocessing result and the corresponding identification result are shown in Figures 14 and 15 respectively. In Figure 15, the nineteen labels of the complex junctions to be identified are in three grey levels. White stands for exact identification, light grey for various degree of inexactness, and dark grey for misses. As can be seen, ten were identified exactly, and eight inexactly, and one is missed. In addition, four artefact ones were mistakenly identified, which were highlighted and labeled with red numbers on the map. By comparing Figure 14 with Figures 9 and 15 with Figure 12, it is not difficult to figure out that all of the inexact identification was due to imperfect stroke assembling. Additionally, all of the missed and artefact complex junctions were due to misclassification of roads, although some of the classification errors were related to the mistakes in stroke assembling. There were also dual-line pairing errors, but they did not manifest themselves in the result of this trial of identification.



Figure 14. The test dataset and its preprocessing result without manual correction. The map used the same symbolization as that of Figure 9 except for the random stroke colors. The end points of the road segments of the raw data are represented by the light grey dots. The nineteen complex junctions to be identified are labeled from one to nineteen. The rough extents of the complex junctions are circled with the dashed ellipses. The strokes assembled are uniquely colored using a random coloring scheme. The paired dual lines are indicated by their light grey background. The secondary roads were indicated by their yellow background.



Figure 15. The identification result obtained from the input with errors. The nineteen labels of the complex junctions to be identified are in three grey levels. White stands for exact identification, light grey for various degree of inexactness, and dark grey for misses. For those inexactly identified or missed, red dashed ellipses are used to circle the extents of concern. The four artefact complex junctions that were mistakenly identified are highlighted and labeled with the red numbers.

The influence exposed above should be not surprising. To make it worse, there are no known error-free algorithms for the three widely used preprocessing operations of data enrichment, i.e., stroke assembling, counterpart pairing and road classification. That means the result of purely software-based preprocessing would normally contain errors, which if not corrected would significantly influence the identification. Manual check and correction after each operation is therefore a necessity. In addition, the above trial indicates that it is still so even though each step produces output of high accuracy, as the errors will propagate and accumulate through the preprocessing flow. That also means user-friendly efficient tools for interactive check and correction are invaluable in production environment, and reliable estimation of the uncertainty of processing results is highly desired to be provided by any inexact algorithm in addition to the result itself.

6. Discussion

In this article, we consider the scale-dependence of identifying complex junctions as two aspects: the first is spatial heterogeneity among complex junctions, and the second is different representations of a complex junction at its relative scales (multiple representations). For dealing with the scale-dependence, this article proposed a novel approach to identify complex junctions based on their topological boundaries. No matter how complicated or simple a complex junction is, it contains its boundary primary roads and interior secondary ones. Therefore, the topological boundary of a complex junction is inherently scale-independent that can be free from the influence of spatial heterogeneity and multiple representations.

From the comparison, it can be found that the topological boundary of a complex junction is a practical characteristic to integrally identify complex junctions from avoiding overfitting problems. Therefore, it seems that representing complex junction as the primarysecondary road structure is cleaner and clearer than point-based (or intersection-based) cluster in a road network. This approach can also be appropriately integrated into the process of road network generalization based on stroke representation. By simplifying complex junctions without breaking connectivity among relative roads, the generalization of road network can be purely dealt with primary ones. Figure 16 shows an example to explain this process by transforming dual-line representation to single-line one for the road network in Figure 3c. The pairing lines a and b of a dual carriageway are combined with the primary road c. Similarly, f and g compose another road F. Therefore, classifying roads into the primary and secondary categories can be considered as a preprocessing operations of data enrichment for road network generalization.



Figure 16. A result of the road network generalization by combining the pairing primary roads.

Thereby the criteria to differentiate between primary roads and secondary ones is a necessity. In this article, strokes are assembled and classified based on road attributes. In the domain knowledge of junction design [29–31], the principle of secondary road (auxiliary lanes, slip roads or ramps) design is rigorously considered, such as number of lanes, speed restriction and driving direction. Primary roads and secondary roads are with significantly difference on their design principles for such as road capacity, safety, location and space [29]. Therefore, the standard of interchange design may be applied as criteria to classify roads of primary-or-secondary type in the future.

7. Conclusions

This paper presented a neat solution to the problem of identifying complex junctions. In addition to preserving the integrity of complex junctions, the proposed method is free from the troublesome threshold-setting, able to deal with different representations of roads, and efficient. All these together make it favorable for practical use. With the boundary and interior clearly identified, complex junctions in a road network can be properly simplified without breaking the connectivity among roads, thereby maintaining the topological validity of the simplified road network. What is more, such topological clarity can greatly facilitate the geometric and semantic simplification of complex junctions [23,25,26].

This study consolidates the effectiveness of the intelligent approach to spatial data generalization that is characterized by an appropriate combination of data enrichment, representation transformation and pattern recognition [8,19]. Although this study is mainly concerned with the composite pattern of complex junction, it is based on the components that are progressively richer in semantics and more complex in structure, which in turn are roads strokes, classified roads and the bundled counterparts of a road. Therefore, we think all of the three operations of data enrichment (i.e., stroke assembling, road classification and counterparts pairing) are crucial and desired to be better addressed so as to improve the accuracy of algorithms, make reliable uncertainty estimation for effectively locating potential errors in results, and provide interactive tools for efficiently checking results and correcting errors.

In that regard too, the two new conceptions adopted in this study can be of use wider than defining the composition of a complex junction. Firstly, we believe the categorization of roads into two basic types, the primary and the secondary, is not only appropriate but also useful in other processing and analysis of road networks. As is shown in this study, the topological boundary of complex junctions instantly becomes clear once roads are classified this way. In fact, we found it also very useful for better pairing the dual lines of a dual carriageway and analyzing other patterns in road networks in our related studies ongoing. Secondly, the concept of topological boundary may also apply to other structural patterns in road networks besides complex junctions. For instance, street blocks may be better delineated topologically than geometrically. We will examine such possibility in our future studies.

Author Contributions: Zhu Xu and Zhilin Li designed this research. Jianting Yang implemented the methodology and performed the experimental validation. Kongyang Zhao and Muzi Li performed the data preprocessing and conducted some of the experiments. All authors have contributed to this article with discussion and writing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (No. 40971209 and 41971330) and the Key Laboratory of Digital Mapping and Land Information Application of National Administration of Surveying, Mapping and Geoinformation (No. DM2016SC07).

Acknowledgments: This work was supported by the New Century Talent Supporting Project of education ministry.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Brassel, K.E.; Weibel, R. A review and conceptual framework of automated map generalization. *Int. J. Geogr. Inf. Syst.* **1988**, 2, 229–244. [CrossRef]
- 2. Weibel, R. Map generalization in the context of digital systems. Cartogr. Geogr. Inf. Syst. 1995, 22, 259–263. [CrossRef]
- 3. Jones, C.B.; Mark Ware, J. Map generalization in the web age. Int. J. Geogr. Inf. Sci. 2005, 19, 859–870. [CrossRef]
- 4. Buttenfield, B.P.; McMaster, R.B. *Map Generalization: Making Rules for Knowledge Representation;* Longman Scientific and Technical: Harlow, UK, 1991; ISBN 978-047-021-803-7.
- 5. Li, Z. Digital map generalization at the age of enlightenment: A review of the first forty years. *Cartogr. J.* **2007**, *44*, 80–93. [CrossRef]

- 6. Mackaness, W.A.; Beard, K.M. Use of graph theory to support map generalization. *Cartogr. Geogr. Inf. Syst.* **1993**, *20*, 210–221. [CrossRef]
- Jones, C.B.; Bundy, G.L.; Ware, M.J. Map generalization with a triangulated data structure. *Cartogr. Geogr. Inf. Syst.* 1995, 22, 317–331. [CrossRef]
- 8. Neun, M.; Weibel, R.; Burghardt, D. Data enrichment for adaptive generalisation. In Proceedings of the 8th ICA Workshop on Generalisation and Multiple representation, Leicester, UK, 20–21 August 2004.
- Barrault, M.; Regnauld, N.; Duchène, C.; Haire, K.; Baeijs, C.; Demazeau, Y.; Hardy, P.; Mackaness, W.; Ruas, A.; Weibel, R. Integrating multi-agent, object-oriented and algorithmic techniques for improved automated map generalization. In Proceedings of the 20th International Cartographic Conference, Beijing, China, 6–10 August 2001; pp. 2110–2116.
- 10. Lee, J.; Jang, H.; Yang, J.; Yu, K. Machine learning classification of buildings for map generalization. *Isprs Int. J. Geo-Inf.* **2017**, *6*, 309. [CrossRef]
- 11. Feng, Y.; Thiemann, F.; Sester, M. Learning cartographic building generalization with deep convolutional neural networks. *Isprs Int. J. Geo Inf.* **2019**, *8*, 258. [CrossRef]
- 12. Stoter, J.; Post, M.; van Altena, V.; Nijhuis, R.; Bruns, B. Fully automated generalization of a 1: 50k map from 1: 10k data. *Cartogr. Geogr. Inf. Sci.* 2014, 41, 1–13. [CrossRef]
- 13. Šuba, R. Design and development of a system for vario-scale maps. Archit. Built Environ. 2017, 13–14. [CrossRef]
- Edwardes, A.J.; Mackaness, W.A. Intelligent generalisation of urban road networks. In Proceedings of the GIS Research UK 2000 Conference, York, UK, 5–7 April 2000; pp. 81–85.
- 15. Touya, G.; Lokhat, I.; Duchêne, C. CartAGen: An Open Source Research Platform for Map Generalization. In Proceedings of the 29th International Cartographic Association (ICC 2019), Tokyo, Japan, 15–20 July 2019. [CrossRef]
- 16. Jiang, B.; Claramunt, C. A structural approach to the model generalization of an urban street network. *GeoInformatica* 2004, *8*, 157–171. [CrossRef]
- 17. Benz, S.A.; Weibel, R. Road network selection for medium scales using an extended stroke-mesh combination algorithm. *Cartogr. Geogr. Inf. Sci.* 2014, *41*, 323–339. [CrossRef]
- 18. Yu, W.; Zhang, Y.; Ai, T.; Guan, Q.; Chen, Z.; Li, H. Road network generalization considering traffic flow patterns. *Int. J. Geogr. Inf. Sci.* 2020, 34, 119–149. [CrossRef]
- 19. Touya, G. A road network selection process based on data enrichment and structure detection. *Trans. Gis* **2010**, *14*, 595–614. [CrossRef]
- Yang, B.; Luan, X.; Li, Q. An adaptive method for identifying the spatial patterns in road networks. *Comput. Environ. Urban. Syst.* 2010, 34, 40–48. [CrossRef]
- Shao, L.; Mittelstädt, S.; Goldblatt, R.; Omer, I.; Bak, P.; Schreck, T. Analysis and Comparison of Feature-Based Patterns in Urban Street Networks. In Proceedings of the 11th International Joint Conference on Computer Vision, Imaging and Computer Graphic, Rome, Italy, 27–29 February 2016; pp. 287–309. [CrossRef]
- 22. Balboa, J.L.G.; López, F.J.A. Generalization-oriented road line classification by means of an artificial neural network. *GeoInformatica* 2008, 12, 289–312. [CrossRef]
- Savino, S.; Rumor, M.; Zanon, M.; Lissandron, I. Data enrichment for road generalization through analysis of morphology in the CARGEN project. In Proceedings of the 13th ICA Workshop on Generalisation and Multiple Representation, Zürich, Switzerland, 12–13 September 2010.
- 24. Wenzel, S.; Bulatov, D. Simultaneous Chain-Forming and Generalization of Road Networks. *Photogramm. Eng. Remote Sens.* 2019, 85, 19–28. [CrossRef]
- 25. Mackaness, W.A.; Mackechnie, G.A. Automating the detection and simplification of junctions in road networks. *GeoInformatica* **1999**, *3*, 185–200. [CrossRef]
- 26. Savino, S.; Rumor, M.; Lissandron, I. Road Junction Generalization in Large Scale Geodatabases; CRC Press/Balkema: London, UK, 2009; ISBN 978-042-920-671-9.
- 27. Zhou, Q.; Li, Z. Experimental analysis of various types of road intersections for interchange detection. *Trans. GIS* **2015**, *19*, 19–41. [CrossRef]
- 28. Grosso, E. Etude des carrefours d'un réseau routier–Première approche: Modélisation et implémentation d'outils de détection des carrefours (plate-forme Oxygène). Master's Thesis, University of Paris 1 Pantheon-Sorbonne, Paris, France, February 2004.
- 29. Lamm, R.; Psarianos, B.; Choueiri, E.M.; Mailaender, T. Interchange Planning and Design-An International Perspective. *Transp. Res. Rec.* **1993**, 1385, 69–83.
- 30. Lunenfeld, H. Human factors associated with interchange design features. Transp. Res. Rec. 1993, 1385, 84-89.
- 31. Chlewicki, G. New interchange and intersection designs: The synchronized split-phasing intersection and the diverging diamond interchange. In Proceedings of the 2nd Urban Street Symposium, Anaheim, CA, USA, 28–30 July 2003.
- 32. Thomson, R.C. The 'stroke' Concept in Geographic Network Generalization and Analysis. In *Progress in Spatial Data Handling*; Riedl, A., Kainz, W., Elmes, G.A., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 681–697. [CrossRef]
- 33. Thomson, R.C.; Richardson, D.E. The 'good continuation'principle of perceptual organization applied to the generalization of road networks. In Proceedings of the 19th International Cartographic Conference, Ottawa, ON, Canada, 14–21 August 1999.
- 34. Lämmer, S.; Gehlsen, B.; Helbing, D. Scaling laws in the spatial structure of urban road networks. *Phys. A Stat. Mech. Its Appl.* **2006**, *363*, 89–95. [CrossRef]

- 35. Newman, M.E.J. *Networks: An Introduction;* Oxford University Press: New York, NY, USA, 2010; pp. 35, 121–124, ISBN 978-019-920-665-0.
- 36. Corcoran, P.; Mooney, P. Characterising the metric and topological evolution of OpenStreetMap network representations. *Eur. Phys. J. Spec. Top.* **2013**, *215*, 109–122. [CrossRef]