

RESEARCH

Open Access



# A road network simplification algorithm that preserves topological properties

Jinyoung Pung<sup>1\*</sup>, Raissa M. D'Souza<sup>1,2,3†</sup>, Dipak Ghosal<sup>1†</sup> and Michael Zhang<sup>4†</sup>

<sup>†</sup>Raissa M. D'Souza, Dipak Ghosal and Michael Zhang assert joint second authorship

\*Correspondence:  
jpung@ucdavis.edu

<sup>1</sup> Department of Computer Science, University of California, Davis, CA 95616, USA

<sup>2</sup> Department of Mechanical and Aerospace Engineering, University of California, Davis, CA 95616, USA

<sup>3</sup> Santa Fe Institute, Santa Fe, NM 87501, USA

<sup>4</sup> Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA

## Abstract

A road network can be represented as a weighted directed graph with the nodes being the traffic intersections, the edges being the road segments, and the weights being some attribute of a road segment. Such a representation enables researchers to analyze road networks in consistent and automatable ways from the perspectives of graph theory. For example, analysis of the graph along with the traffic demand pattern can identify critical road segments based on centrality measures. However, due to the complexity of real-world road networks and the computationally expensive algorithms, it is challenging to extend the such methods to a large-scale road network. In this paper, we present a simple yet efficient network simplification framework based on graph theory that sub-samples and simplifies the graph while preserving key topological characteristics in the original network. Our method iteratively identifies and removes network elements that do not contribute to transportation functionality, such as self-loops, dead-ends, and interstitial nodes that lies on the same road line. We applied this method to three small cities with distinct street patterns and one large city, and showed that topological characteristics in the original networks are preserved by comparing two distinct kinds of centrality distributions in the original and simplified networks.

**Keywords:** Road network, Graph theory, Simplification, Network pruning, Betweenness centrality, Information centrality

## Introduction

In an urban road network, many roads and streets have complex inter-connections. Due to their complexity, researchers often represented road networks as spatial graphs for systematic analysis. This enables graph-theoretic algorithms to be applied to road networks. However, even a small city contains more than thousands of road segments and a detailed graph representation makes it challenging to use original road networks as input to analysis models (Bazzi et al. 2010).

To deal with this issue, researchers used a subsampled version of the city-level road network. Porta et al. (2006) investigated spatial graphs of four different cities by *Multiple Centrality Assessment* which assesses four different *centrality* measures to capture topological and geometric characteristics with different perspectives. They limited the area of study to 1-square-mile sub-regions of the original networks. Park and Yilmaz (2010)

investigated centrality measures and their entropy in road networks. In their study, small graphs with a maximum of 104 nodes are used to represent sub-regions of cities. Youn et al. (2008) assessed the *price of anarchy* by computing the difference in travel time between an *origin* and *destination* pair in social optima and Nash equilibria. Instead of the original network, they used the skeleton network of each city which consists of only principal arterial roads. Although these subsampling approaches make the studies computationally feasible, the results obtained from only a small portion of a city are biased to that selected part and may not be representative of the city. Also, selecting specific types of roads omits roads of other types regardless of their topological importance which can result in unexpected disconnections.

Similarly, traffic simulation studies of city-wide networks using microscopic traffic simulators such as VISSIM (Fellendorf and Vortisch 2010) also suffer from expensive computational costs. Many different user groups such as researchers and transportation system authorities have used traffic simulators to test new ideas and to easily collect data without interfering with real-world networks. However, they have considerable execution runtimes, even for moderate size cities. Furthermore, since traffic simulators must account for the stochastic nature of the traffic demand, they need to be run multiple times for reliable results. This further exacerbates the poor scalability with the size of the road network (Antoniou et al. 2014).

Motivated by these issues, we propose a novel road network simplification framework that preserves topological and geometric characteristics of the original network. Our method iteratively removes *cul-de-sacs* and gridiron patterns made of low-level roads from a road network represented as a directed graph. Porta et al. found that betweenness centrality and information centrality well represent the backbone and collective behaviors of road networks (Porta et al. 2006) so we assessed these two centrality distributions of the original and simplified networks to examine differences in topological characteristics and found that this simple proposed process significantly lightened the volume of the original networks while the simplified networks have very similar centrality distribution to the original networks.

Given that our framework preserves key topological characteristics of the original network after simplification, it suggests that a simplified network can be used for road network analysis instead of the large and detailed original network. This makes road network analysis more scalable with the size of the road network. Also, unlike most previous approaches, our method takes input as a *directed* graph that preserves the directionality of roads, which allows the simplification framework to be used for road network analysis of more diverse purposes such as routing of vehicles.

## Related works

In practice, reduced representations of road networks are commonly used for road network *generalization* in cartographic maps. Since the real-world space has enormous information, maps with limited space need to deliver information efficiently by highlighting relatively more important information. In this context, road network generalization methods usually work as a selective omission process, which sorts network elements in the order of given criteria, and selectively omits less important network elements with low importance ranks.

There are various techniques for road network generalization. Following early work by Mackaness and Beard (1993), many researchers (Mackaness 1995; Thomson and Richardson 1995; Jiang and Claramunt 2004; Jiang and Harrie 2004) proposed graph-based approaches for road network generalization, in which road segments and intersections are represented as nodes and edges respectively (or vice versa). These studies identified important network elements to be selected in generalized maps by incorporating graph-theoretic algorithms such as minimum-spanning-tree, shortest path, and centrality. However, since they consider mostly on topological relationships, traditional graph-based approaches ignored semantic and geometric properties of roads.

Thomson and Richardson (1999), based on the principle of 'Good Continuity', defined a *stroke* as a group of road segments that have the same road type and intersect at a small angle. Instead of individual road segments, they considered strokes for their generalization model, then computed the importance rank of strokes, and selected important strokes. In this approach, output networks have continuous network elements after generalization. Inspired by the stroke-based method, several studies (Thomson and Brooks 2001; Liu et al. 2010; Yu et al. 2020) developed their methods with strokes as selection units. Chen et al. (2009) used a *mesh*, a closed region by road segments, as a network element unit. Their method starts with the identification of meshes that have a density beyond a given threshold which are then merged with their adjacent meshes. Compared to the stroke-based approach, the mesh-based approach achieves an uniform distribution of roads in the generalized network.

More recently, data-driven approaches have been proposed to reflect traffic flow patterns in addition to the geometric and topological properties of networks. Yu et al. (2020) modified the stroke-based method by including the relationship between road segments and traffic flows in computing the importance score of strokes. Van De Kerkhof et al. (2020) sorted a set of car trajectories that consist of consecutive road segments, and selected the road segments that belong to high-rank trajectories. Those approaches can have better connectivity for routes that are frequently used by drivers.

The above-mentioned road network generalization methods can reduce the volume of road networks by any given threshold in continuous ways. However, the primary objective of those approaches is in making the map in different scales more "readable" and the threshold is set regardless of the structure of network, causing unexpected changes in connectivity and topological characteristics after generalization. Although recent data-driven approaches attempt to prevent unexpected disconnection of functionally important road segments, they rely on past data and thus potentially useful road segments can be removed causing undesired disconnection.

There are several road network *simplification* methods that work without any pre-determined threshold for determining the size of simplified networks. Boeing (2017) simplified road networks by removing interstitial nodes that lie on the same road line. Since these nodes are just extending the roads which do not branch, they are treated as redundant and replaced with a single edge that concatenates all the road segments into one. Huynh and Selvakumar (2020) proposed a simplification method that iteratively cut short dangling paths, identify clusters of road network components, and then collapsed each cluster into a single node.

However, there are several drawbacks in these approaches. First, the method proposed by Boeing (2017) only focuses on sequential road chunks on the same road line. Second, the method proposed by Huynh and Selvakumar (2020) is applicable only to undirected graphs and thus cannot be used for analysis of the dynamic nature of road networks. Also, their method makes the assumption that road networks are strictly *planar* meaning that they can be well represented by a simplified two-dimensional model without any underpass or overpass which is not always true. In this paper, we propose a novel road network simplification method that preserves key topological characteristics similar to the original, in which a directed graph converges to a simplified directed graph without any pre-determined threshold.

## Method

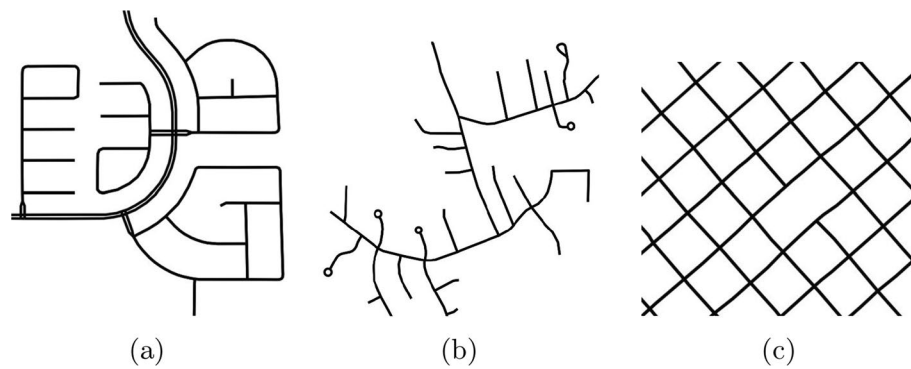
### Road network data

OSMnx (Boeing 2017) is a Python package that downloads road networks from OpenStreetMap (Haklay and Weber 2008) and constructs them into *primal*, *non-planar* and *weighted* multidigraphs. This means that nodes and directed edges represent intersections and roads respectively (primal), grade-separated roads such as overpass and underpass do not have an intersection (non-planar), and geographic and spatial information of roads such as road length is included in the edge attributes that can be used as weights. We utilized OSMnx for our study since the above distinctive features of the package takes the dynamic nature of the road networks into account.

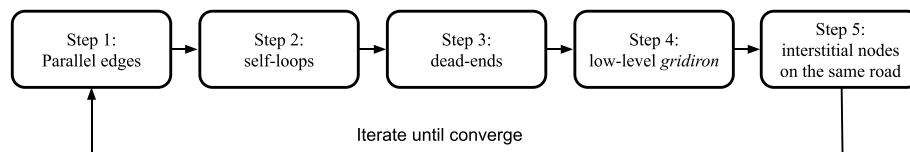
### Framework

Urban road networks have hierarchical structures. High-level roads (e.g., motorways and arterial roads) transport a large number of vehicles at fast speeds while low-level roads (e.g., residential roads) have lower speed limits and are used to provide access between high-level roads and local areas. Low-level roads have little impact on the vehicular dynamics from the perspective of global transportation. However, some low-level roads may provide detours and shortcuts between sub-regions or distribute traffic avoiding congestion on high-level roads. Thus, arbitrary loss of low-level roads may have non-trivial impact on traffic flow and topological context should be considered in any road network simplification process. In our method, we distinguish trivial low-level roads which are superfluous from topologically important roads, then selectively omit such roads so that the topological characteristics of the original network can be preserved after simplification. To identify the redundant roads, we utilize three patterns in residential street network suggested by Southworth and Ben-Joseph (2013): *loops and lollipops*, *lollipops on a stick*, and *gridiron*. Figure 1 shows an example network for each pattern.

The *loops and lollipops* pattern is characterized by the presence of loops and *cul-de-sacs*. Both loops and *cul-de-sacs* are not likely to contribute to transportation functionality since a loop tends to get back to its starting point and a *cul-de-sac* does not provide any through pass to the rest of the network. The *lollipops on a stick* pattern consists of a few through streets with branching off *cul-de-sacs* from those streets, where *cul-de-sacs* are considered as redundant as explained above. Some studies highlight the effect of *cul-de-sacs* on road networks. The studies in Batac and Cirunay (2022) and Distel (2015) point out that travel from a dead-end node to another is



**Fig. 1** Example networks of residential street patterns: **a** loops and lollipops, **b** lollipops on a stick, and **c** gridiron



**Fig. 2** The five-step simplification framework and components being removed in each step

sinuous, especially if the length of the path is very short, which may translate into a degradation in the quality of travel. Also, in the study in Li et al. (2022), cul-de-sacs may provide access points where traffic may flow into a traffic analysis zone (TAZ) of interest from outside the TAZ. As some features of a TAZ are computed using the number of access points to the TAZ, cul-de-sacs are not trivial in their study. Consequently, studies in Batac and Cirunay (2022), Li et al. (2022), and Distel (2015) show that cul-de-sacs have local impacts on road networks. Since in this study, we perform a city-level analysis of road networks and focus on the network-wide properties, we remove *cul-de-sacs* in the proposed simplification framework ignoring their potential local impacts.

The *gridiron* pattern is a simple system of two series of parallel streets crossing at right angles to form a pattern of rectangular blocks, which provides a lot of route choices. Also, the studies in Distel (2015) and Daganzo et al. (2011) argue that *gridiron* pattern may have a critical impact on the road network as it may cause gridlock traffic congestion especially when traffic demand is very high. In our study, however, we simplify *gridiron* pattern that only consists of low-level roads. These low-level roads will remain unused since they have the same direction and length as their nearest high-level roads and will be used only when the high-level roads are highly degraded or disrupted. Thus we consider them as trivial elements from a network-wide traffic perspective, and decided to remove them through simplification. We used the road type information that is tagged by OSM to identify low-level roads and roads with *residential* tag are considered as low-level roads.

Our method identifies the target patterns using the topological, geometric, and semantic information of a road network. The framework is depicted in Fig. 2 and consists of five steps:

1. Parallel edges are removed from the input graph leaving only the shortest edge between two adjacent nodes. Those edges are relatively long and generally used to provide access from main roads to residential areas.
2. Self-loop edges are removed from the graph. Circular ends for easy turning at the end of roads, which are represented as self-loops in a graph, not only add unnecessary overhead, but also make a dead-end node have at least two adjacent nodes (itself and its neighbor). Removing self-loops ensures dead-end nodes have a single adjacent node for the next step.
3. The graph is simplified by removing dead-ends, which are the nodes that have only one adjacent node and incident edges of the nodes. These components are only used to provide access to the end node and can be collapsed to the entrance of each *cul-de-sac*.
4. Areas with *gridiron* pattern are simplified by removing low-level components. The nodes that satisfy all the following conditions are removed along with their incident edges: (1) have exactly 4 adjacent neighbors, (2) the maximum length of the incident edges is less than 300 m, (3) the road type of all the incident edges is residential, and (4) at least two nodes under the conditions above are adjacent.
5. The interstitial nodes on a single road line are removed by replacing the sub-edges with a single unified edge. We used the method proposed by Boeing (2017) for this step. These five steps are iterated until the input graph converges to the final graph upon which no further simplification can be made.

#### Tracking regional node density of the original network

The simplified network has lower node density than the original, especially in residential areas of the network, as the framework removes nodes in target areas. This can lead to different topological characteristics (e.g. centrality measures) after simplification. To circumvent this, we set a node attribute `aggr_node_number` to keep track of the regional node density in the original network. The value of the attribute is initialized to 1 for each node in the original graph. When a node or a group of nodes is removed for simplification, the `aggr_node_number` value of the node, or the summation of `aggr_node_number` values of the group of nodes being removed, is distributed equally to its neighbors. Intuitively, this attribute of a node in the simplified network represents the number of other nodes that were collapsed into it, which can be used to approximate the node density in the original network. In our study, we used it to better estimate the centrality measure of original networks from simplified networks as explained in detail later. There are other potential benefits of this attribute such as using it for generating origin and destination pairs in traffic simulations.

**Algorithm 1:** Simplification algorithm

---

```

input: Directed graph  $G = (V, E)$ 
while  $G$  is not converged do
  foreach edge  $\{u, v\} \in E$  do
    if edge  $\{u, v\}$  is not the shortest edge from  $u$  to  $v$  then
      | remove  $\{u, v\}$  from  $E$  ▷ parallel edges simplified
    end
  end
  foreach edge  $\{u, v\} \in E$  do
    if  $u == v$  then
      | remove edge  $\{u, v\}$  from  $E$  ▷ self-loops simplified
    end
  end
  foreach node  $v \in V$  do
    if  $v$  has only a single neighbor then
      | remove incident edges of  $v$  from  $E$ 
      | remove node  $v$  from  $V$  ▷ dead-ends simplified
    end
  end
   $U \leftarrow \{\}$ 
   $V_{grid} \leftarrow \{\}$ 
  foreach node  $v \in V$  do
     $N_v \leftarrow$  the number of neighbors of  $v$ 
    if  $N_v == 4$  then
      |  $E_v \leftarrow$  incident edges of  $v$ 
      | if all edges in  $E_v$  are residential road and have length of less than 300m then
        | | add  $v$  to  $U$ 
      | end
    end
  end
  foreach node  $u \in U$  do
    if any neighbor of  $u$  is in  $U$  then
      | add  $u$  to  $V_{grid}$ 
    end
  end
  remove incident edges of nodes in  $V_{grid}$  from  $E$ 
  remove  $V_{grid}$  from  $V$  ▷ gridiron pattern simplified
  foreach sequence of nodes on the same road line  $P$  do
     $s \leftarrow$  start node of  $P$ 
     $t \leftarrow$  end node of  $P$ 
     $V_{int} \leftarrow \{v : v \in P, v \neq s, v \neq t\}$ 
    if  $V_{int} \neq \{\}$  and there is not any branch between  $s$  and  $t$  then
      | remove incident edges of  $V_{int}$  from  $E$ 
      | remove  $V_{int}$  from  $V$ 
      | add a new edge  $\{s, t\}$  to  $E$  ▷ Interstitial nodes simplified
    end
  end
end

```

---

**Centrality measures and estimation**

In graph theory and network analysis, the basic idea of *centrality* is that there are relatively more *central* or important nodes and edges in a network. Since the first set of centrality indices were defined for social network analysis by Freeman (1978), various centrality indices have been suggested and widely applied to many other fields of study including road network analysis (Porta et al. 2006; Park and Yilmaz 2010; Zhang et al. 2011; Huynh and Selvakumar 2020). In our study, we utilized two centrality indices: *betweenness centrality* and *information centrality*. Porta et al. (2006) showed those centrality indices nicely capture the backbone structure of a road network and collective behaviors. Observing the difference in the distribution of centrality measurements



before and after simplification provides a measure of how much a simplification method distorts the topological characteristic of a road network.

### Edge betweenness centrality

Edge betweenness centrality is a concept that generalize Freeman's betweenness centrality to edges (Girvan and Newman 2002), which shows how frequently an edge lies on the shortest paths connecting a pair of nodes in a graph. In a road network, the higher betweenness an edge has, the more it provides shortest routes and is likely to contribute to the transportation in a city. The betweenness centrality  $C^B$  of an edge  $e$  is defined by

$$C^B(e) = \frac{1}{N(N-1)} \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)} \quad (1)$$

where  $N$  is the number of nodes in a graph,  $V$  is the set of nodes,  $\sigma(s,t)$  is the number of shortest paths between an origin and destination pair  $(s,t)$ , and  $\sigma(s,t|e)$  is the number of those paths that passing through edge  $e$ .

### Edge information centrality

Based on the concept of efficient propagation of information over a social network, Latora and Marchiori defined *information centrality* as the relative drop in the *network efficiency* caused by the removal of a node (Latora and Marchiori 2004) where network efficiency represents how efficiently information is exchanged over the network (Latora and Marchiori 2001). Fortunato et al. (2004) generalized information centrality to edges and defined *edge information centrality*. Applying edge information centrality to a road network, we recognize network efficiency as the summation of the ratio between the length of the straight line and the shortest path between each origin and destination pair  $(s,t)$ . The normalized network efficiency  $E$  for a weighted graph  $G$  as proposed in Vragović et al. (2005) is given by:

$$E(G) = \frac{1}{N(N-1)} \sum_{s,t \in V; s \neq t} \varepsilon_{st} = \frac{1}{N(N-1)} \sum_{s,t \in V; s \neq t} \frac{d_{st}^{Eucl}}{d_{st}} \quad (2)$$

where  $\varepsilon_{st}$  is the efficiency of travel from node  $s$  to  $t$ ,  $d_{st}^{Eucl}$  is the Euclidean distance between a pair of nodes  $s$  and  $t$ , and  $d_{st}$  is the length of the shortest path from node  $s$  to  $t$ . In case there is no path from  $s$  to  $t$ ,  $d_{st} = \infty$  and, consequently,  $\varepsilon_{st} = 0$ . Thus edge information centrality is well defined for either weakly connected or disconnected graphs. The removal of an edge forces origin and destination pairs to choose alternative paths or there may not be any available alternatives, in either case the network suffers from a high drop in efficiency. For example, the removal of bridges or the removal of the only connection to a large sub-graph is likely to cause high drop in efficiency. Thus such edges have high information centrality. The information centrality  $C^I$  of an edge  $e$  is defined by

$$C^I(e) = \frac{E(G_{org}) - E(G_{cut}^e)}{E(G_{org})} \quad (3)$$

where  $G_{org}$  is an original graph and  $G_{cut}^e$  is the graph that removed edge  $e$  from  $G_{org}$ .



### Estimating centrality from a simplified network

Roads and intersections in residential areas are significantly simplified by our method. To preserve the original density of nodes we use the method of preserving overall `aggr_node_number` discussed above. Although the removed nodes have minimal effect on the transportation functionality, the decreased node density in those areas would result in a difference in topological characteristics and centrality distribution between the simplified and original networks.

Specifically, we estimate centrality in the original network from a simplified network by using the value of `aggr_node_number` attribute which represents the number of nodes that are removed in the vicinity of a node. For computing centrality in a simplified network, we can use the attribute as a weight to estimate centrality of the same element in the original network.

The estimated betweenness centrality  $\hat{C}^B$  of an edge  $e$  is defined by

$$\hat{C}^B(e) = \frac{1}{N(N-1)} \sum_{s,t \in V'} \frac{\sigma(s,t|e)}{\sigma(s,t)} \times aggr_s \times aggr_t \quad (4)$$

where  $V'$  is the set of nodes in the simplified graph, and  $aggr_v$  is the `aggr_node_number` value of a node  $v$ . Similarly to  $\hat{C}^B$ , we estimate the network efficiency  $E$  of the original graph  $G$  from a simplified graph  $G_{simple}$ . The estimated network efficiency  $\hat{E}$  is defined by

$$\hat{E}(G_{simple}) = \frac{1}{N(N-1)} \sum_{s,t \in V'; s \neq t} \frac{d_{st}^{Eucl}}{d_{st}} \times aggr_s \times aggr_t \quad (5)$$

By plugging in  $\hat{E}$  into Eq. (3), the estimated information centrality  $\hat{C}^I$  can be derived.

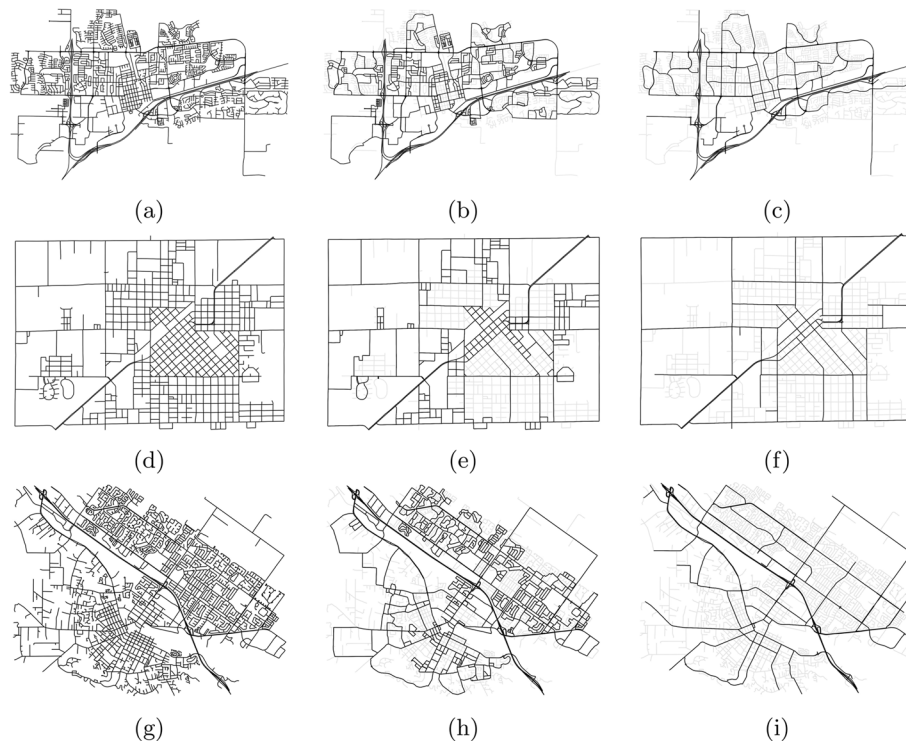
The time complexity of computation of centrality is  $O(|V||E|)$  (Girvan and Newman 2002) for betweenness centrality and  $O(|V||E|^3)$  (Fortunato et al. 2004) for information centrality where  $|V|$  is the number of nodes and  $|E|$  is the number of edges in a network. The attribute `aggr_node_number` is simply a multiplicative factor in the time complexity which is dominated by the size of graph defined by the number of nodes and edges in the graph. Thus, estimating the centrality measures in the original network from a simplified network has substantially lower computational cost than directly computing centrality measures of the original network.

## Experiment and results

We simplified the road networks of three cities in the United States to evaluate our method. In the first experiment, we investigate three small cities where each city has very distinct features. The first selected city is Davis, a small college town in California. The city has plenty of residential areas with *cul-de-sacs*. On the contrary, the next selected city, Portales in New Mexico mostly consists of *gridiron* pattern such as two series of parallel streets crossing at right angles. Finally, we selected Petaluma in California which has multiple street patterns and a geographical constraint which is a river that crosses the city. The latter has more general and combined street patterns than the two other cities. Also, as the separated regions by the river are interconnected with a

**Table 1** Four cities with distinct features and the description of street pattern in the cities

City	Distinct feature	Description
Davis, CA	Cul-de-Sacs	Street pattern most common in residential areas
Portales, NM	Gridiron pattern	Street pattern most common in residential areas
Petaluma, CA	Geographical constraint	Sensitive to distortion due to geographical constraint More general and combined street patterns
Columbia, SC	Complex road network	Various patterns are interconnected Most general road network

**Fig. 3** Visualization of original and simplified road networks. Left column (a, d, g): input network, middle column (b, e, h): simplified by the proposed method, right column (c, f, i): simplified by omitting all residential road. Top row (a, b, c): Davis, middle row (d, e, f): Portales, bottom row (g, h, i): Petaluma

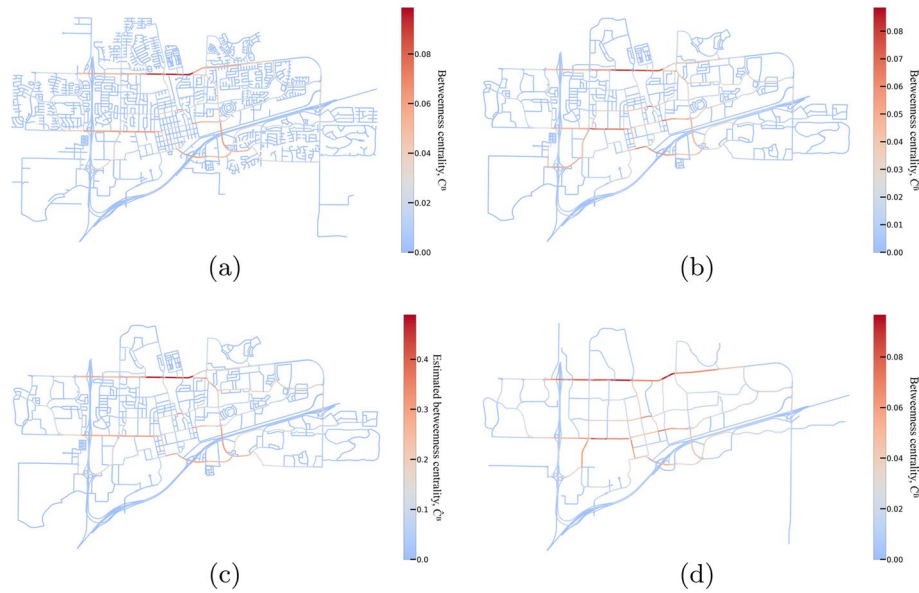
few bridges, it is likely to be sensitive to possible distortion in topological characteristics after simplification. In the second experiment, we extend the study to a big city, Columbia in South Carolina, where streets and roads in various patterns are interconnected in a complex manner. Since this city has multiple patterns in its road network, the analysis of this city demonstrates that our simplification algorithm works for general cities. The distinctive features of the road network in the selected cities are summarized in Table 1.

In addition to our proposed simplification method, we considered a benchmark which is a naive but frequently used method that omits all the residential roads in a network for comparison.

Figure 3 and Table 2 show the visualizations and statistical results of the network simplification process for the selected cities. We observed that the proposed method efficiently simplified road networks in all three cities by about the factor of two. The naive

**Table 2** Statistic results of simplification

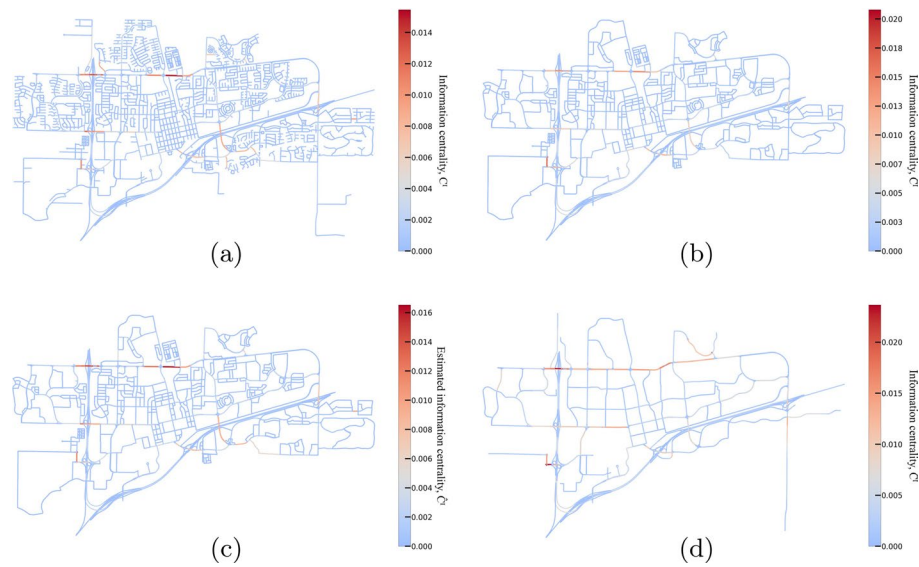
Simplification	Davis		Portales		Petaluma	
	Node	Edge	Node	Edge	Node	Edge
Original	2110	4968	794	2428	2374	5957
Proposed	960 (− 55%)	2397 (− 52%)	373 (− 53%)	1091 (− 55%)	892 (− 62%)	2425 (− 59%)
W/o residential	782 (− 63%)	1448 (− 71%)	402 (− 49%)	821 (− 66%)	633 (− 73%)	1164 (− 80%)


**Fig. 4**  $C^B$  distribution in Davis road network: **a** original, **b** proposed method, **c** proposed method with centrality estimation, and **d** simplified by omitting all residential roads

heuristic that omits all the residential roads can simplify the networks to more reduced representations than our method. However, it oversimplifies networks and makes arbitrary disconnections regardless of the functionality of transportation, distorting the topological characteristics of the original networks.

To identify the difference in topological characteristics that resulted from simplification, we measured betweenness centrality  $C^B$  and information centrality  $C^I$  in each network before and after simplification. For the networks simplified by our proposed method, each centrality is measured in two different ways: the standard centrality  $C$  and estimated centrality  $\hat{C}$  using the `aggr_node_number` approach. Figures 4 and 5 visualize the distributions of  $C^B$  and  $C^I$  in the city of Davis respectively. In the visualizations, the simplified maps resulting from the proposed method share key features and details with the original network especially with respect to centrality measures. However, the map which simply omits residential roads shows a clearly different distribution, having more widespread centrality or high values in erroneous components. Centrality distribution in Portales and Petaluma also showed the same tendency as Davis.

For quantitative evaluation, we used correlation coefficient to compare the centrality of the edges of an original network and the corresponding edges of the simplified network. If the correlation coefficient has a high value, there is a strong association between



**Fig. 5**  $C^I$  distribution in Davis road network: **a** original, **b** proposed method, **c** proposed method with centrality estimation, and **d** simplified by omitting all residential roads

**Table 3** Pearson ( $\rho$ ) and Spearman ( $R$ ) correlation coefficient between centrality measurement of original and simplified networks

Simplification	Davis		Portales		Petaluma	
	$C^B$	$C^I$	$C^B$	$C^I$	$C^B$	$C^I$
Proposed, with estimation	$\rho = 0.947$ $R = 0.939$	$\rho = 0.978$ $R = 0.933$	$\rho = 0.761$ $R = 0.845$	$\rho = 0.998$ $R = 0.997$	$\rho = 0.931$ $R = 0.907$	$\rho = 0.947$ $R = 0.891$
Proposed	$\rho = 0.916$ $R = 0.895$	$\rho = 0.928$ $R = 0.893$	$\rho = 0.735$ $R = 0.831$	$\rho = 0.980$ $R = 0.964$	$\rho = 0.869$ $R = 0.837$	$\rho = 0.897$ $R = 0.833$
W/o res. roads	$\rho = 0.817$ $R = 0.798$	$\rho = 0.666$ $R = 0.739$	$\rho = 0.593$ $R = 0.687$	$\rho = 0.560$ $R = 0.523$	$\rho = 0.857$ $R = 0.853$	$\rho = 0.382$ $R = 0.513$

the centrality of the original network and the simplified network, which supports that the key topological properties in the original network are preserved after simplification.

Specifically, we used Pearson ( $\rho$ ) and Spearman ( $R$ ) correlation coefficient. Pearson correlation coefficient represents the linear relationship between two variables, that is, a change in the centrality of the simplified network is proportional to the centrality of the original network. Spearman correlation coefficient can evaluate the monotonic relationship between two variables, i.e., how the rank of centrality measurements is similar between the original and simplified network.

Since the edge segments that lie on the same road line are replaced with a single unified edge throughout our proposed method, the maximum centrality measurement of such edge segments in the original network is matched to the centrality measurement of the unified edge in a simplified network to compute the correlation coefficient.

Table 3 is the results of quantitative analysis. In every case, the proposed method has the highest level of correlation with the two measures of centrality, having very high values from 0.761 to 0.998. It implies that the topological characteristics in the simplified networks are very similar to the original networks, and the centrality distribution in the

original network can be accurately approximated in an efficient way using a simplified network. On the other hand, omitting all the residential roads for simplification resulted in low values of correlation factors from 0.382 to 0.857, and turned out to change the topological characteristics of the original network considerably.

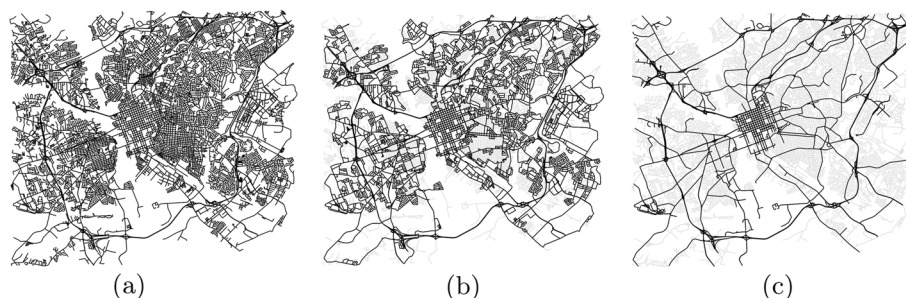
We extended our experiment to a big city where streets and roads in various patterns are interconnected in a more complicated manner. We simplified the road network in Columbia which is the capital city of South Carolina in the United States. The original network has 11,113 nodes and 29,781 edges. Our method reduced the size by 5457 nodes (– 51%) and 15,275 edges (– 49%), whereas simply omitting all residential roads reduced the size by 3433 nodes (– 69%) and 6851 edges (– 77%) (Fig. 6).

Similar to the first experiment, the number of nodes and edges are simplified by the factor of two by our proposed method. Although the naive approach that omits all residential roads can further reduce the size of the road network, it ends up having undesired disconnections between sub-regions. Figure 7 and Table 4 clearly shows our method has a very similar  $C^B$  distribution to the original network and high correlation coefficient while the other naive method has a different distribution and low correlation coefficient.

Considering the heavy computational cost to compute the information centrality for Columbia, we do not obtain this result. However, here we provide an estimate of the time it would take to compute information centrality in the original and simplified network with our machine which has a 2.6 GHz CPU. In the original network, it would take about 240 days, but in the simplified network by our method, it is estimated to take less than 42 days which is about six times faster than the original network.

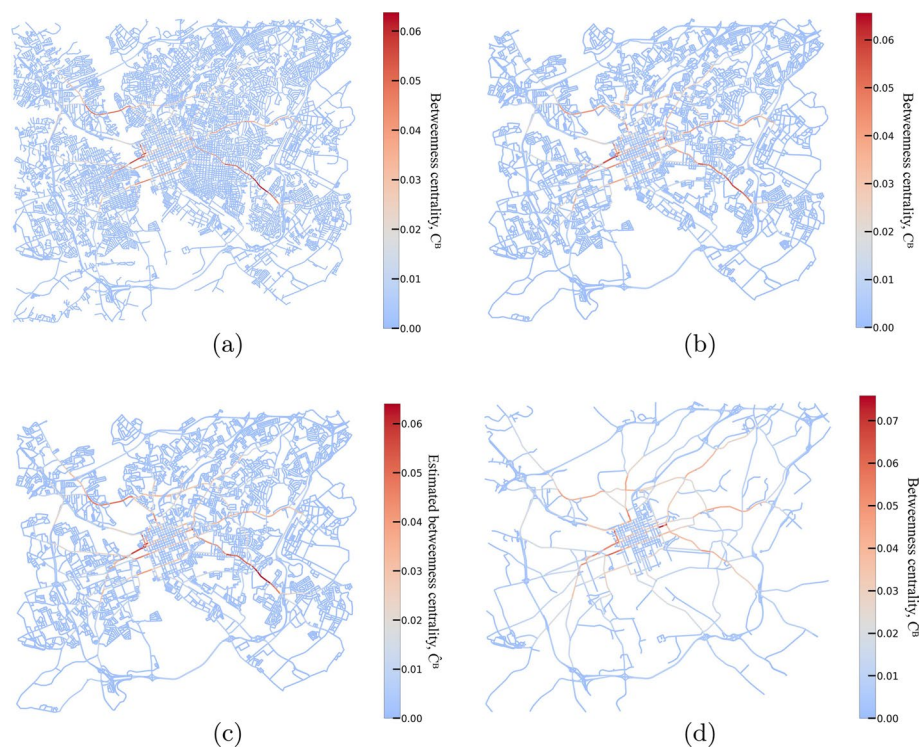
## Conclusion

This article proposes a road network simplification framework that selectively removes network components that have little impact on transportation functionality: *cul-de-sacs* and gridiron patterns consisting of low-level roads. In this way, the topological characteristics of a road network are preserved while efficiently de-densifying the network. The method keeps track of the regional node density of the original network, which can be used to more precisely estimate topological characteristics of the original network such as centrality measurement. We applied our method to three small cities with distinct street patterns and a big complex city, then computed



**Fig. 6** Visualization of the road network in the city of Columbia: **a** input network, **b** simplified by the proposed method, and **c** simplified by omitting all residential road





**Fig. 7**  $C^B$  distribution in Columbia road network: **a** original, **b** proposed method, **c** proposed method with centrality estimation, and **d** simplified by omitting all residential roads

**Table 4** Pearson ( $\rho$ ) and Spearman (R) correlation coefficient between centrality measurement of original and simplified networks in Columbia

Simplification method	Correlation coefficient
Proposed, centrality estimation	$\rho = 0.964$ , $R = 0.911$
Proposed	$\rho = 0.952$ , $R = 0.883$
w/o residential roads	$\rho = 0.805$ , $R = 0.834$

centrality distributions in the road networks of the cities. By measuring the correlation coefficient between the centrality measurements in the original and simplified networks, we quantitatively showed that the topological characteristics in the original network are successfully preserved after simplification. We argue that this is an important step to enable large-scale road network analysis. Having similar topological characteristics to the original network, the simplified network can be used for the analysis of vehicular dynamic behaviors instead of a massive, thus computationally expensive, original network.

**Abbreviations**  
 $C^B$  Betweenness centrality  
 $C^B$  Estimated betweenness centrality  
 $C^I$  Information centrality  
 $C^I$  Estimated information centrality

### Author contributions

JP, RD, DG, and MZ conceived the work and coordinated the research. JP is responsible for the design, implementation, and evaluation of the simplification framework, and authored the manuscript. RD, DG and MZ contributed research expertise and edited the manuscript. All authors read and approved the final manuscript.

### Funding

JP acknowledges the support of the Republic of Korea Navy.

### Availability of data and materials

We have used publicly available road network data. More information of these data are provided in the corresponding references.

### Declarations

#### Competing interests

The authors declare that they have no competing interests.

Received: 27 August 2022 Accepted: 16 November 2022

Published online: 01 December 2022

### References

- Antoniou C, Barcelò J, Brackstone M, Celikoglu H, Ciuffo B, Punzo V, Sykes P, Toledo T, Vortisch P, Wagner P (2014) Traffic simulation: case for guidelines
- Batac RC, Cirunay MT (2022) Shortest paths along urban road network peripheries. *Phys A Stat Mech Appl* 597:127255
- Bazzi A, Masini BM, Pasolini G, Torreggiani P (2010) Telecommunication systems enabling real time navigation. In: 13th International IEEE conference on intelligent transportation systems. IEEE, pp 1057–1064
- Boeing G (2017) OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput Environ Urban Syst* 65:126–139
- Chen J, Hu Y, Li Z, Zhao R, Meng L (2009) Selective omission of road features based on mesh density for automatic map generalization. *Int J Geogr Inf Sci* 23(8):1013–1032
- Daganzo CF, Gayah VV, Gonzales EJ (2011) Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability. *Transp Res B Methodol* 45(1):278–288
- Distel MB (2015) Connectivity, sprawl, and the cul-de-sac: an analysis of cul-de-sacs and dead-end streets in Burlington and the surrounding suburbs
- Fellendorf M, Vortisch P (2010) Microscopic traffic flow simulator VISSIM. In: *Fundamentals of traffic simulation*. Springer, New York, pp 63–93
- Fortunato S, Latora V, Marchiori M (2004) Method to find community structures based on information centrality. *Phys Rev E* 70(5):056104
- Freeman LC (1978) Centrality in social networks conceptual clarification. *Soc Netw* 1(3):215–239
- Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc Natl Acad Sci* 99(12):7821–7826
- Haklay M, Weber P (2008) Openstreetmap: user-generated street maps. *IEEE Pervasive Comput* 7(4):12–18
- Huynh HN, Selvakumar R (2020) Extracting backbone structure of a road network from raw data. In: *International conference on computational science*. Springer, pp 582–594
- Jiang B, Claramunt C (2004) A structural approach to the model generalization of an urban street network. *Geoinformatica* 8(2):157–171
- Jiang B, Harrie L (2004) Selection of streets from a network using self-organizing maps. *Trans GIS* 8(3):335–350
- Latora V, Marchiori M (2001) Efficient behavior of small-world networks. *Phys Rev Lett* 87(19):198701
- Latora V, Marchiori M (2004) A measure of centrality based on the network efficiency. *Open-Access J Phys*. <https://arxiv.org/pdf/math/0402050.pdf>
- Li H, Wu D, Zhang Z, Zhang Y (2022) Safety impacts of the discrepancies and accesses between adjacent traffic analysis zones. *J Transp Saf Secur* 14(3):359–381
- Liu X, Zhan FB, Ai T (2010) Road selection based on voronoi diagrams and “strokes” in map generalization. *Int J Appl Earth Observ Geoinf* 12:194–202
- Mackaness W (1995) Analysis of urban road networks to support cartographic generalization. *Cartogr Geogr Inf Syst* 22(4):306–316
- Mackaness WA, Beard KM (1993) Use of graph theory to support map generalization. *Cartogr Geogr Inf Syst* 20(4):210–221
- Park K, Yilmaz A (2010) A social network analysis approach to analyze road networks. In: *ASPRS annual conference*, San Diego, pp 1–6
- Porta S, Crucitti P, Latora V (2006) The network analysis of urban streets: a primal approach. *Environ Plann B Plann Des* 33(5):705–725
- Southworth M, Ben-Joseph E (2013) Streets and the shaping of towns and cities
- Thomson RC, Brooks R (2001) Exploiting perceptual grouping for map analysis, understanding and generalization: the case of road and river networks. In: *International workshop on graphics recognition*. Springer, pp 148–157
- Thomson RC, Richardson DE (1995) A graph theory approach to road network generalisation. In: *Proceeding of the 17th international cartographic conference*, pp 1871–1880



- Thomson RC, Richardson DE (1999) The 'good continuation' principle of perceptual organization applied to the generalization of road networks
- Van De Kerkhof M, Kostitsyna I, Van Kreveld M, Löffler M, Ophelders T (2020) Route-preserving road network generalization. In: Proceedings of the 28th international conference on advances in geographic information systems, pp 381–384
- Vragović I, Louis E, Díaz-Guilera A (2005) Efficiency of informational transfer in regular and complex networks. *Phys Rev E* 71(3):036122
- Youn H, Gastner MT, Jeong H (2008) Price of anarchy in transportation networks: efficiency and optimality control. *Phys Rev Lett* 101(12):128701
- Yu W, Zhang Y, Ai T, Guan Q, Chen Z, Li H (2020) Road network generalization considering traffic flow patterns. *Int J Geogr Inf Sci* 34(1):119–149
- Zhang Y, Wang X, Zeng P, Chen X (2011) Centrality characteristics of road network patterns of traffic analysis zones. *Transp Res Rec* 2256(1):16–24

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---