Changes in properties of the transport network graph when using simplification algorithms

Cite as: AIP Conference Proceedings **2172**, 080008 (2019); https://doi.org/10.1063/1.5133566
Published Online: 13 November 2019

Maxim Sabanov, and Galina Timofeeva

Changes in Properties of the Transport Network Graph when Using Simplification Algorithms

Maxim Sabanov[1, b)] and Galina Timofeeva[1,2, a)]

[1]*Ural State University of Railway Transport, Yekaterinburg, Russia*
[2]*Ural Federal University, Yekaterinburg, Russia*

[a)]Corresponding author: gtimofeeva@usurt.ru
[b)] sabanovus@gmail.com

Abstract. In the course of solving problems of managing and planning the development of large transport networks, a big amount of heterogeneous data from various sources is accumulated. The methods of graph theory are used to systematize information, process it, and calculate network characteristics. However, the problem arises of the enormous complexity and duration of computations over graphs of large networks. One of the approaches to solve this problem is the simplification of the original graph. In our report we study the influence of processing and basic graph simplification on the initial properties of the network.

INTRODUCTION

The development of transport infrastructure is one of the key factors affecting the quality of life and the possibilities for the economic development of regions.

At the same time, in Russia over the past 10–15 years, rail transportation occupies a leading position in the field of cargo transportation, whose share in the transportation market is more than 65% in terms of data for 2016 [1]. In connection with this, the object of the study was precisely the railway network.

Property analysis and transport network management is a complex task. One of the methods for solving this problem is the analysis of network properties using graph theory. Often, the initial information about the transport network is contained in maps and geographic information systems.

Such a representation may include tens of thousands of lines and hundreds of thousands of points. This amount of information is redundant and creates difficulties in processing and calculating the properties of the system [2]. In addition, the source data that can be obtained from public sources contain errors that impede the automated processing of information.

Algorithms for processing and simplification of graphs help solve the problem [3-4]. Initially, the task of enlarging arose because of the need to manage complex systems, described by a huge number of interconnections, which creates the need for more and more computing power needed to make operational control actions. However, the influence of the simplification process on the properties of the original network is often ignored. This article discusses the influence of the algorithm of simplification [5] on the properties of the transport network of railways of the Sverdlovsk region.

Proceedings of the 45th International Conference on Application of Mathematics in Engineering and Economics (AMEE'19)
AIP Conf. Proc. 2172, 080008-1–080008-6; https://doi.org/10.1063/1.5133566
Published by AIP Publishing. 978-0-7354-1919-3/$30.00

INITIAL DATA

As a source of information about the transport network, were used materials published by the open cartographic project OpenStreetMap [6], namely, a vector layer containing geographical information (points and lines with coordinates, meta-information) about the railway network of the Sverdlovsk region.

The original representation of the vector layer, after exclusion from it based on the meta-information of all inactive lines, as well as tram networks and the metro, includes 8810 line-objects and 97378 point-objects. The visualization of the original vector layer is shown in Fig. 1.

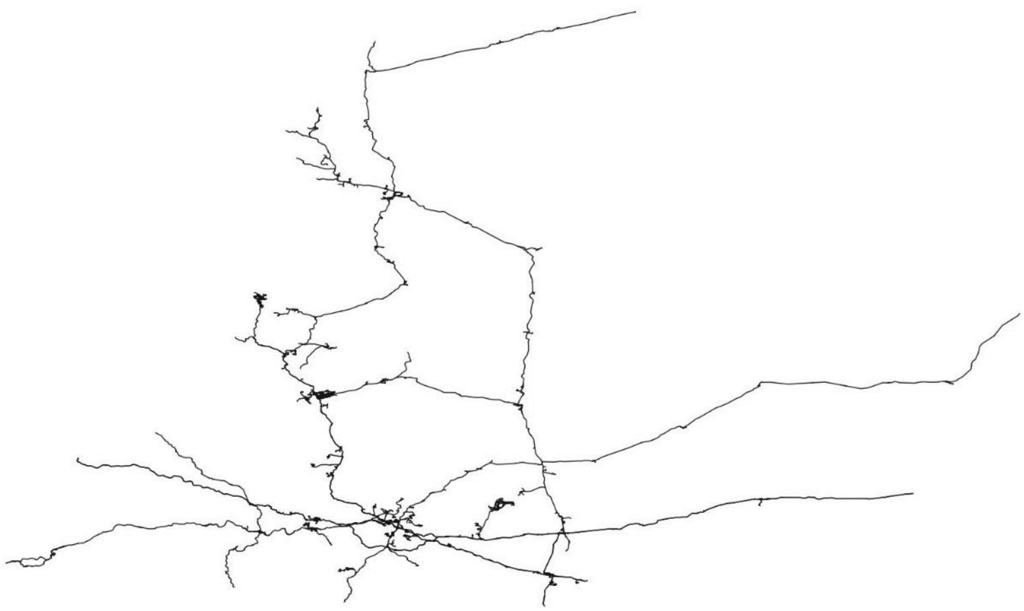

FIGURE 1. Initial representation of the transport network

The algorithm of preprocessing and simplifying was applied to this vector layer. The types of objects of the layer were corrected, the accuracy of the coordinates for all objects was aligned. In addition, for all settlements, relative to which, in the future, the properties of the transport network connecting them will be calculated, the internal transport systems were simplified to the center of the settlement polygon. Also, points that are vertices of a graph of degree 2 have been removed from the vector layer (they do not reflect the interrelationships of the graph elements, affecting only the curvature of the lines).

After simplifying by the algorithm, a new vector layer of railways began to contain 2670 lines and 7669 points, which gives us a tenfold simplification of the graph in terms of the number of elements.

SIMPLIFICATION ALGORITHM EFFECT

Simplification algorithms have the greatest impact on the original network trace. It would be logical to assume that the influence will be greater, the more extensive the route in the network we consider. Denote L_{orgign} as the original length of the route, L_{simpl} as the length of the route after simplification and delta = ($L_{origin} - L_{simpl}$).

Consider the results of network processing by the algorithm presented in the correlation field (Fig. 2). The lengths of the shortest paths between settlements are displayed on the abscissa axis, and the delta values obtained for these paths are displayed on the ordinate axis.

We can observe that for the vast majority of measurements (80% of the sample fall inside the contour in Fig. 2) there is a weak linear dependence of the resulting error on the path length, described by a correlation equation of the form delta = 5.885 – 0.00117 L_{origin}. However, for some of the values, the delta values are much higher than the rest of the sample. Let us try to compare instances with approximately equal values of L_{origin}, but different values of delta.

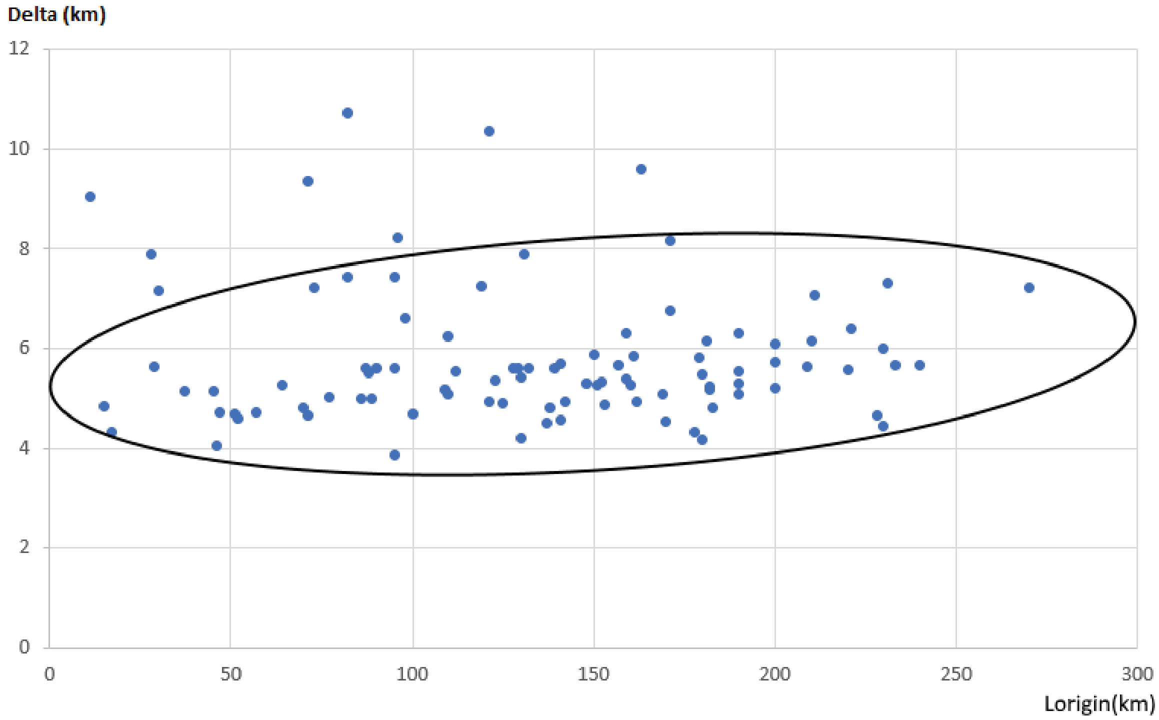

FIGURE 2. Impact on shortest path length

Path A (numbered as 5 on the vertical axis of Fig. 2) is the shortest path between the settlements of Bogdanovich and Kamishlov. The initial length of the route is almost 44 kilometers, after simplification - 41 km, which gives us a total deviation of 3 km. This path is shown on Fig. 3.

FIGURE 3. Shortest path between Bogdanovich and Kamishlov

Path B - Sysert-Aramil (Fig. 4) has an initial length of 50.6 km; however, after simplification, the length was 43.5 km, which gives an error of more than 7 km.

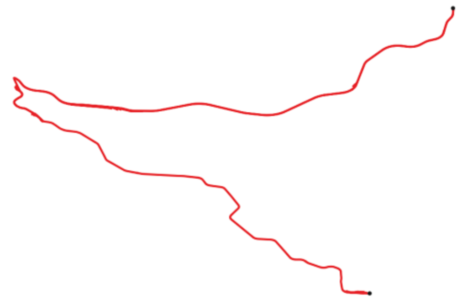

FIGURE 4. Shortest path between Sysert and Aramil

In this case, if we consider the original outline of the routes A and B (presented in Fig. 3), which have comparable length and a large difference in the delta index, we will find that the deviation is much greater for the section with a non-linear structure.

This fact suggests that the degree of influence of the integration algorithm depends on the complexity of the outline of the original graph. Thus, if we consider the railways of the Siberian region, which are characterized by a simpler and more linear structure, we can expect significantly less influence than, for example, when considering the networks of central regions that have a complex radial structure.

Transport accessibility and betweenness centrality were considered as the estimated characteristics of the transport network. This pair of parameters was chosen based on the fact that they use different characteristics of the transport network graph - one of them depends on the direct outline of the network, the second uses only the concept of the shortest path and the weight characteristics of the key nodes that do not participate in the simplification.

Transport accessibility

Transport accessibility is a characteristic of a certain point or territory, indicating the extent to which the selected methods can overcome the space that separates this point or territory from other points or territories under consideration. This indicator for the rail transport network will depend on the total traffic flow passing through the node and the length of the shortest route between each pair of nodes. Transport accessibility is adapted for use on large transport networks measure DivAct [7].

$$\varphi(\nu) = \frac{\sum_{u \in V} N(u) \cdot L_{sp}(\nu, u)}{\sum_{u \in V} N(u)}, \tag{1}$$

v – graph node;
V - set of nodes of the graph;
$N(u)$ - traffic flow through the node u;
L_{sp} (v,u) - length of the shortest path between nodes v and u.

Consider the impact of simplification on indicators of transport accessibility of large network nodes. Table 1 shows the calculated data of the indicator, calculated on the enlarged and original graph. Here φ_{simpl} is the value of transport accessibility according to the simplified graph, φ_{origin} is the value of transport accessibility according to the original graph, Sum of L_{sp} is the total length of the shortest paths for a node without weighting factors and delta = 1- (φ_{simpl} / φ_{origin}).

We can observe a picture similar to the effect on the length of the shortest path. The largest values of delta deviations do not correspond to the nodes with the highest indicators of the total length of the shortest paths and the influence of the network tracing is enhanced by the size of the traffic flow passing through the node. For example, the Sysert node, which has the largest deviation value in the entire network, is characterized by a large traffic flow, but is connected to the remaining nodes through a network section with a complex outline.

TABLE 1. Estimated indicators of transport accessibility

Node	φ_{simpl} (km)	φ_{origin} (km)	Sum of L_{sp} (km)	delta
Kamishlov	311	331	336359	6,1%
Yekaterinburg	308	330	335338	6,6%
Pervouralsk	222	233	236121	4,7%
Bogdanovich	215	228	231420	5,8%
Kamensk Uralsky	164	170	172866	3,7%
Aramil	107	113	114608	5,2%
N. Tagil	107	113	103597	5,2%
Nevyansk	78	82	83120	5,1%
Alapaevsk	76	81	81823	5,8%
Novouralsk	69	73	74276	5,8%
Artyomovsky	3	3	33855	6,4%
Polevskoy	23	25	25469	8,4%
V. Tagil	22	24	23971	6,5%
Kirovgrad	22	23	23325	5,5%
Sysert	11	12	12439	10,7%

Betweenness centrality

Betweenness centrality is an indicator defining the hierarchy of graph edges, based on how often the shortest routes between two nodes pass through a specific edge, as well as on the weights of these nodes. This definition is supplemented with the help of the weights of the nodes of centrality from the paper [8]

$$C(e) = \frac{\sum\limits_{u,v \in V} \sigma_{uv}(e) N(u) N(v)}{(\sum\limits_{u \in V} N(u))^2}, \tag{2}$$

e – the graph edge;
V - set of nodes of the graph;
σ_{uv} (e) – equals 1 if e is the shortest path between u and v, else equals 0;
L_{sp} (v, u) - length of the shortest path between nodes v and u;
$N(u)$ - traffic flow through the node u.

Consider the impact on the indicator of centrality. If we turn to formula (2), by which the indicator is calculated, you can see that the path length is not involved in it, and only the relative characteristic of the face of the graph plays a role - is it part of the shortest path between the two nodes, as well as data on the nodes themselves. Thus, the integration does not affect the numerical values, but it creates errors in determining the location of the central routes of the network.

However, if we superimpose the sections on each other, one can see that even on simple linear sections of the paths we get a deviation from the actual passage of the central route to more than 1 kilometer (Fig. 5).

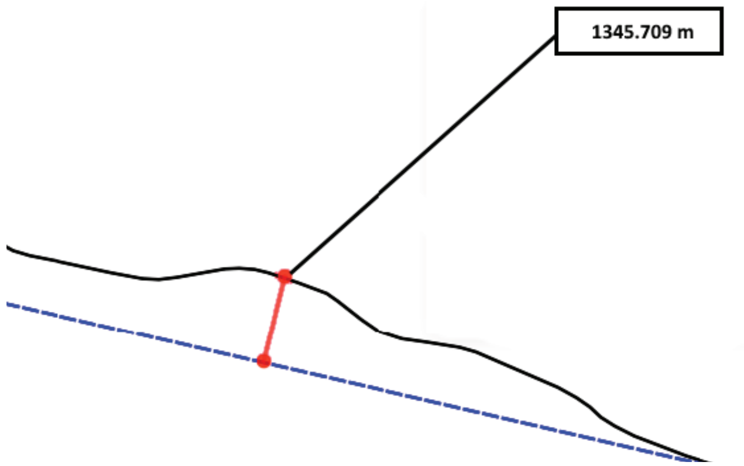

FIGURE 5. Effects on Betweenness centrality

Transport networks with high complexity of the organization (where many routes laid on a compact territory), such deviations can give is not quite correct representation of the passage of the most central roads.

CONCLUSION

The average impact of the considered simplification algorithm on calculated indicators ranges from 3.6% to 10%. On average, the deviation is 6%, which is a relatively good indicator for simplify a graph by 10 times in the number of elements. With further application of simplification algorithms, it is possible to reduce the effect of graph simplification on the network parameters, taking into account the obtained estimates.

ACKNOWLEDGMENTS

The investigation was funded by RFBR, project number17-08-01123-a.

REFERENCES

1. A.M. Butov, Rail Freight Market, Report, National Research University Higher School of Economics – Centre of Development Institute, Moscow, 2016.
2. M.A. Sabanov and G.A. Timofeeva, Modern Methods for Processing and Enlargement the Graphs of Transport Network, Transport of Urals **58**(3), 8–12 (2018).
3. E. Ko, M. Kang, H. J. Chang and D. Kim, "Graph-Theory Based Simplification Techniques for Efficient Biological Network Analysis," *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, San Francisco, CA, 2017, pp. 277-280.
4. F. Bonchi, G. De Francisci Morales, A. Gionis and A. Ukkonen, Activity Preserving Graph Simplification, Data Mining and Knowledge Discovery, **27**(3), 321–343 (2013).
5. M.A. Sabanov, Methods of Transformation of Geographical Data on Transport Network, Herald of the Ural State University of Railway Transport **39**(3), 84–90 (2018).
6. Official Web-site of OpenStreetMap Project, see https://www.openstreetmap.org/copyright/en.
7. X. Albacete, D. Olaru, V. Paül and S. Biermann, Appl. Spatial Analysis **10**(2), 161-188 (2017).
8. S.K.R. Unnithan, B. Kannan, M. Jathavedan, Betweenness Centrality in Some Classes of Graphs, International Journal of Combinatorics, **2014**, ID 241723 (2014).