



Article The Design of a Metro Network Using a Genetic Algorithm

Aleksander Król^{1,*} and Małgorzata Król²

- ¹ Faculty of Transport, Silesian University of Technology, Krasińskiego 8, 40-019 Katowice, Poland
- ² Faculty of Energy and Environmental Engineering, Silesian University of Technology, Konarskiego 18, 44-100 Gliwice, Poland; malgorzata.krol@polsl.pl
- * Correspondence: aleksander.krol@polsl.pl; Tel.: +48-326-034-120

Received: 30 December 2018; Accepted: 26 January 2019; Published: 28 January 2019



Featured Application: A quick method that allows one to obtain a draft of a metro network.

Abstract: The process of planning the metro system is very complex. This is due to the large amount of required input data originating from many domains. Additionally, the required extraordinary budget makes the design phase difficult and forces some trade-offs to meet most expectations. Designing a metro system requires often a multi-variant analysis to examine possible future ways of urban agglomeration development. In such circumstances there is a need to apply an efficient method, which allows for obtaining a draft design of a metro system. This paper presents an application of a genetic algorithm for two stages of a metro system design. The first stage determines the layout of metro stations, which optimally meets the transportations demands of the inhabitants. The second stage generates the optimal topology of lines connecting the stations, taking into account the minimization of construction costs and the total travel time of all passengers. The parameters, which control the operation of the optimization procedures, were tuned with the use of the Taguchi method.

Keywords: metro system; station location; lines layout; genetic algorithm; Taguchi method

1. Introduction

Metro systems is the best way to improve the quality of public transport in large cities due to their large capacity, rapidity, time reliability, and almost total separation from the urban space [1]. The construction of a metro network entails extraordinary costs; therefore, the design phase is of great importance. Since a metro network should serve the inhabitants living in a given area, their transportation demands should be the main criterion while designing such a system [2]. Obviously, there are always some budget constraints that could significantly limit the range of the investment. However, a feedback relationship also exists—if a new metro station is built in a relatively underdeveloped place, it could stimulate the development of such an area [3–6]. Furthermore, the cost of such a metro infrastructure location could be significantly lower. In such a situation, there are two approaches when considering the planned location of a new infrastructure: the first oriented towards the current transportation demands and the second oriented toward future land development [7]. After the above-mentioned conceptual choice has been decided, a detailed study on station location can take place. The main issue of this stage is accessibility, which is twofold:

- Attraction accessibility, which means the ease of reaching the station using any mode of transport;
- Radiation accessibility, which means the ease of reaching other stations from the given station [8].

Accessibility is the main criterion of assessing the degree of satisfaction of passengers' transportation demands [9–11]. In general, when considering the optimization of the structure of the

public transportation systems in cities, the streets topology, the origin-destination demands, and the existing and available transportation infrastructure must be taken into account [12].

The above-mentioned issues involve the diverse domains of economy, engineering, science, anthropology, and management [13], and therefore lead to difficult theoretical problems [14]. Although, there are known few exact algorithms that allow for design such complex systems, the computational complexity of such a class of problems induces the use of some heuristic algorithms [12,15–17]. The application of heuristic methods allows one to find an almost optimal solution in a reasonable time. Additionally, there is no need to fully understand or even know about all internal dependencies of the problem when applying such methods. The only demand is the existence of an objective function that describes the quality of each solution variant. This is why heuristic methods, especially artificial intelligence methods, are widely used in various fields of science and medicine [18–20].

The previously described case study of the central part of Upper Silesia region in Southern Poland can be regarded as a model example, because it involves designing an entire metro network. The considered area is a highly populated region with many cities of small and medium size. The region entered the post-industrial phase almost 20 years ago. Heavy industry has ceased to prevail, and now the industry of new technologies has started to dominate. There is a very dense road network, and a railway system, over the whole region. The system of public communication is also highly developed, but it suffers severely from the old infrastructure. Generally, public communication systems in Poland are less efficient than in other European countries (there is only one metro network in the whole country). For some time, an idea of building a metro network in the central part of Upper Silesia region has been considered. However, no budget constrains have been determined, and no guidelines concerning the overall scope of the investment have been established.

The study presented in this paper covers two stages of metro system design: first, the optimal locations of metro stations were found, based mainly on the distribution of population density; next, the optimal layout of metro lines was designed. Therefore, the approach oriented towards current transportation demands was adopted in the first stage. In turn, the second stage involved optimization mainly regarding radiation accessibility. The optimization method used in both stages was a genetic algorithm. As was mentioned above, such a choice resulted from the high computational complexity of the undertaken tasks. The main contribution of the presented work is introducing a quick method that allows one to obtain a draft of a metro system. The strength of the presented method is the capability to provide a reasonable draft of a metro system basing on general data, which characterizes the considered region.

The structure of this paper is as follows: first, the criteria determining the stations location and the lines layout are discussed; then, the optimization method is described, which contains some important detailed solutions. The next section contains a case study: the region being considered is described; the control parameters of the proposed method are tuned; and, finally, the obtained results are presented. In the last section, a short summary and the conclusions are provided.

2. Model Description

The design process is divided into two stages: in the first stage, the optimal station location is determined, and in the second stage the optimal layout of lines connecting the stations is found. Such an approach comes from quite different and independent objectives of both sub-issues. Therefore, the problem can be split since decreasing the number of the degrees of freedom significantly reduces the computational complexity. Additionally, if the task is not split, it might happen that much computational time will be spent on optimizing the lines topology to the as yet not optimal station layout.

The total number of lines and number of stations are assumed in advance. These numbers can be estimated after the review of the structures of some metro networks existing in the world.

2.1. Criteria of Station Locations

A station is regarded as an object whose location must assure meeting the transportation demands of inhabitants. It was assumed that the transportation demands are proportional to the number of people living in an area. Therefore, the main input here is a map of population density covering the considered region. In the rough approach, which involves neglecting the street layout and existing possibilities of inhabitants' mobility, the main criterion of passenger satisfaction is the Euclidean distance to the nearest metro station—e.g., circular shape of the catchment areas was assumed [21]. The present work expresses the number of inhabitants being satisfied with the location of *i*-th metro station as [22,23]

$$n_i = \int\limits_A e^{-\frac{r_i^2}{\sigma^2}} \rho dA \tag{1}$$

where r_i —distance to *i*-th station; ρ —local population density; σ —standard deviation, determining the distance regarded as satisfactory; and *A*—considered area.

The model also takes into account the existence of point generators of transportation demands like stadiums, multi-arenas, railway stations, airports, or big factories. In this case, the number of people related to *j*-th generator whose transportation demands are satisfied by *i*-th station can be expressed as

$$n_{ij}^{g} = e^{-\frac{r_{ij}^{2}}{\sigma^{2}}} g_{j}$$
(2)

where r_{ij} —distance between *j*-th generator and *i*-th station, and g_j —number of people related to *j*-th generator.

The important issue is to establish the proper value of σ parameter. This depends on the way the passengers reach the station. Schlossberg and Brown considered walking accessibility and adopted the values in range 400–800 m [24]. Brinckerhoff used questionnaire research to determine the acceptable distance to the nearest metro station: about a half of respondents claimed it to be less than a 1500 m, and one fifth of respondents claimed more than 3000 m [25]. The distance up to 1000 is commonly accepted as comfortable [26]. The accessibility of a metro station can be significantly enhanced by the introduction of Park & Ride systems [27]; it concerns especially areas of a low population density.

In the optimization process, the locations of all stations are varied in such a way as to maximize the objective function, which is simply the number of people whose transportation demands are satisfied, as follows:

$$f_{obj}^{1} = \sum_{i} n_{i} + \sum_{i} \sum_{j} n_{ij}^{g} \stackrel{opt}{\to} \max.$$
(3)

After optimization is complete, the best locations of all stations are known and the number of potentially serviced people is known for each station as well. Let us denote this value for *i*-th station as s_i and then it can be expressed as

$$s_i = n_i + \sum_j n_{ij}^g. \tag{4}$$

2.2. Criteria of Lines Layout

A metro line is a series of stations. There are no preliminary assumptions determining the network topology, except for one requirement: the network must be coherent. In other words, it must be possible to travel from any origin station to any destination station. The aim of the optimization is to find such a layout of metro lines that the total cost of the network is minimized and the total travel time of all the passengers is minimized too. The model does not take into account the time needed for transfers; what is more, the travel time is assumed to be dependent only on journey length, neglecting the contribution of transfers entirely. Such an assumption could be partially justified when the frequency of trains is relatively high. The above criteria are contradictory: if the travel time is expected to be minimized, the topology of the network has to be somewhat like a grid. This, in turn, means that some connections are duplicated, which causes the increase of the costs. Generally, some basic patterns of

metro network topology can be listed: a triangle, a star, and a cartwheel [28], but having in mind the irregular shape of the populated areas and the irregular distribution of the population density in the considered case, no such assumption was made.

The construction costs of a line segment depend on its length; additionally, it is possible to take into account the cost dependency on some local conditions like rivers, lakes, built up areas, or industry zones. For this purpose, a map describing the cost distribution is introduced into the model, and the building cost of *i*-th segment of *j*-th line between two stations can be expressed as

$$c_{ij} = \int_{L} k dL \tag{5}$$

where *k*—local unit building cost and *L*—segment length.

The second criterion taken into account is the overall network efficiency. There are many different measures developed to asses it [29–31], but from the passengers' point of view the expected travel time is the most important. The total travel time of all potential passengers calculated below is only an estimation, but it reflects the passengers' expectations. In the optimization process, the shortest distance between each pair of stations is calculated (d_{ij}); it could include transfers, but the transfer time is not taken into account. Then, the total travel time between a pair of origin and destination stations for all their potential passengers can be estimated in the following way:

$$t_{ij} \sim d_{ij}(s_i + s_j). \tag{6}$$

If the total travel time is taken into account, the constructed objective should result in a highly interconnected network topology. It will lead to the construction's cost increase, but the passengers' demands will be satisfied in a higher degree [32]. This criterion will remain valid if a lesser number of passengers is taken into account, as long as the population distribution is the same. Just the proportional factor will change. Of course, considering the explicit costs, the investment will appear less beneficial because the costs per a single passenger will increase.

The applied optimization method requires an increasing objective function; thus, both components of the objective function are expressed as inverts

$$f_{cost}^2 = \frac{1}{\sum_j \sum_i c_{ij}}$$
(7)

$$f_{time}^2 = \frac{1}{\sum_j \sum_i t_{ij}}.$$
(8)

Finally, the objective function is a product of both factors

$$f_{obj}^2 = f_{cost}^2 f_{time}^2 \xrightarrow{opt} \max.$$
(9)

Such formulation of the objective function allows for simultaneous optimization in relation to both criteria without the use of arbitrarily adopted weights.

3. Optimization Method

The idea of genetic algorithms is based on the evolution processes in the world of living nature [33]. It was first described by Holland [34]. Due to their advantages, genetic algorithms have quickly become an efficient tool in science and technology [35]. The optimization process is based on the following assumptions:

- Individuals representing different versions of solutions compete with each other;
- The structure of each individual is coded by an entity called a genotype;

- A genotype is randomly varied by mutations;
- Across-over operation causes random pairs of individuals to partially exchange their genotypes;
- The quality of each individual can be evaluated as a single value called fit function;
- Fit function determines the probability of transition to the next generation, which corresponds to the selective pressure;
- The combination of random mutation and cross-over with aimed selective pressure leads the whole population towards the optimal solution.

The main advantage of the genetic algorithms is they are able to find a solution close to the optimal one investigating only a scant part of the solution's space. Additionally, they are able to avoid getting stuck in the local optimum. A specific cross-over operation gives the generally faster convergence in comparison with other methods based on the solutions space exploration.

The selective pressure means that individuals with a higher value of the fit function are preferred to pass to the next generation. In the presented work, the method of the roulette wheel was used, in which the probability of passing to the next generation is directly proportional to the value of the fit function. Additionally, a few of the best individuals pass unconditionally (elite selection). The size of the elite selection should be relevant to avoid ousting from the population individuals that are poor now but could be improved.

Additionally, the current leader of each generation is compared with the stored general leader. The better of both becomes a new general leader. This stored solution does not take part in the optimization process, but it contains the result.

The described issue involved two optimization problems—optimization of stations locations and optimization of lines layout.

3.1. Optimization of Stations Locations

The genotype for the first sub-problem is a simple table containing the following coordinates of stations (x_i and y_i denote coordinates of *i*-th station):

$$\{(x_1, y_1), \dots, (x_i, y_i), \dots (x_n, y_n)\}.$$
(10)

A mutation is a shift of a randomly selected station. There are random numbers added to its coordinates. The distribution of these shifts is normal with the mean equal to zero, and the standard deviation is a program parameter.

$$\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\} \xrightarrow{m} \{(x_1, y_1), \dots, (x_i + \Delta x, y_i + \Delta y), \dots, (x_n, y_n)\}.$$
(11)

A cross-over between a pair of individuals is an exchange of random subset of stations

$$\{ (x_1, y_1), \dots, (x_i^1, y_i^1), \dots, (x_n, y_n) \} \xrightarrow{X} \{ (x_1, y_1), \dots, (x_i^2, y_i^2), \dots, (x_n, y_n) \} \\ \{ (x_1, y_1), \dots, (x_i^2, y_i^2), \dots, (x_n, y_n) \} \xrightarrow{X} \{ (x_1, y_1), \dots, (x_i^1, y_i^1), \dots, (x_n, y_n) \}.$$

$$(12)$$

The probability of mutation and the probability of cross-over are program parameters.

3.2. Optimization of Lines Layout

The genotype for the second sub-problem is a set of stations series. Each series corresponds to a single metro line (L_i , lowercase letters correspond to successive stations) as follows:

$$\{L_1: (e, f, \dots, a); L_2: (p, a, \dots, q); \dots; L_n: (x, q, \dots, y)\}.$$
(13)

The number of the lines is the same for all individuals, and a station can appear at most once in a line (no loops). However, there are no other limitations apart the above: Any station can belong to many lines, so lines can go parallel and can cross each other.

Because of the specific structure of this genotype, the special mutation operators were designed. There are six types of mutations, which allow for completely free changing of a genotype.

1. Exchange of the position of two randomly selected station for a randomly selected line, as follows:

$$\{\ldots; L_i: (p,\ldots,\boldsymbol{x},\ldots,\boldsymbol{y},\ldots,q);\ldots\} \stackrel{m_1}{\to} \{\ldots; L_i: (p,\ldots,\boldsymbol{y},\ldots,\boldsymbol{x},\ldots,q);\ldots\}$$
(14)

2. The order reversion of the randomly selected sub-series of stations for a randomly selected line, as follows:

$$\{\ldots; L_i: (p,\ldots,\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\ldots,q);\ldots\} \xrightarrow{m_2^2} \{\ldots; L_i: (p,\ldots,\boldsymbol{z},\boldsymbol{y},\boldsymbol{x},\ldots,q);\ldots\}$$
(15)

3. Exchange of a pair of randomly selected stations between randomly selected lines, as follows:

$$\{\ldots; L_i: (\ldots, \mathbf{x}, \ldots); \ldots L_j: (\ldots, \mathbf{y}, \ldots); \ldots\} \xrightarrow{m_3} \{\ldots; L_i: (\ldots, \mathbf{y}, \ldots); \ldots L_j: (\ldots, \mathbf{x}, \ldots); \ldots\}$$
(16)

4. Transfer of a randomly selected station between a pair of randomly selected lines, as follows:

$$\{\ldots; L_i: (\ldots, \mathbf{x}, \ldots); \ldots L_j: (\ldots, \mathbf{y}, \ldots); \ldots\} \xrightarrow{m4} \{\ldots; L_i: (\ldots, \ldots); \ldots L_j: (\ldots, \mathbf{x}, \mathbf{y}, \ldots); \ldots\}$$
(17)

5. Removing of a randomly selected station from a randomly selected line, as follows:

$$\{\ldots; L_i: (\ldots, \mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots); \ldots\} \xrightarrow{m_5} \{\ldots; L_i: (\ldots, \mathbf{x}, \mathbf{z}, \ldots); \ldots\}$$
(18)

6. Addition of a randomly selected station to a randomly selected line, as follows:

$$\{\ldots; L_i: (\ldots, \mathbf{x}, \mathbf{z}, \ldots); \ldots\} \xrightarrow{m_6} \{\ldots; L_i: (\ldots, \mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots); \ldots\}$$
(19)

The cross-over operation involves an exchange of a random single metro line between a pair of individuals, as follows:

$$\{ L_1^1, L_2^1, \dots, L_i^1, \dots, L_n^1 \} \xrightarrow{X} \{ L_1^1, L_2^1, \dots, L_i^2, \dots, L_n^1 \}$$

$$\{ L_1^2, L_2^2, \dots, L_i^2, \dots, L_n^2 \} \xrightarrow{X} \{ L_1^2, L_2^2, \dots, L_i^1, \dots, L_n^2 \}.$$

$$(20)$$

After each operation of mutation or cross-over the network structure is checked, special attention is paid to its coherency. If the test fails, the operation is canceled.

The shortest distance between a pair of stations (6) is calculated using fast Dijkstra's algorithm with a priority queue implemented by the heap.

4. Results—A Case Study

The presented method was applied to obtain a preliminary design of a metro system for Upper Silesia conurbation in Southern Poland.

4.1. Case Study Description

Commonly, the development of a metro network is a process lasting many years. The considered case differs here, because the center of Upper Silesia region is an agglomeration of a dozen of cities, which practically form a uniform urban structure, but due to the lack of funding there was no attempt to build a common metro communication system in the past. There are lots of bus and trams lines, which serve over 1 million passengers daily. There is also a network of railways, but the train frequency is much too low to serve as typical urban trains. Very dense road network including two motorways and an urban motorway crossing the centers of most of cities cause many citizens to use their own

cars; this results in congestions, city centers overcrowding and the environment pollution. Nowadays, in such situation an idea has arisen to consider a possibility to build a metro network. Figure 1 presents a map of considered area, and Table 1 lists the biggest cities forming the center of Upper Silesia region. The region extends from E18.6150 to E19.3030 and from N50.1118 to N50.4575 (65 km \times 41 km).



Figure 1. Map of considered area (OpenStreetMap).

No	City	Population	No	City	Population
1	Będzin	69,000),000 10 Piekary Śląskie		58,000
2	Bytom	175,000	11	Ruda Śląska	142,000
3	Chorzów	111,000	12	Siemianowice	70,000
4	Dąbrowa Górnicza	125,000	13	Sosnowiec	214,000
5	Gliwice	186,000	14	Świętochłowice	52,000
6	Katowice	307,000	15	Tarnowskie Góry	61,000
7	Knurów	39,000	16	Tychy	129,000
8	Mikołów	40,000	17	Zabrze	179,000
9	Mysłowice	75,000			

Table 1. The main cities of Upper Silesia region.

4.2. Determination of the Main Metro System Parameters

The first issue to be discussed is the number of metro lines and stations. As there are no strict guidelines, research on a few of the world's metro systems was done. Some of this research is gathered in Table 2. The data are commonly available, but sometimes it is difficult to interpret them

because metro networks often cross the city boundaries and the size of serviced population cannot be determined exactly.

City	No of Lines	No of Stations	Total Length [km]	Population [10 ⁶]	Area [km ²]
London	11	270	402	8.20	1579
Berlin	10	173	146	3.40	892
Paris (agglomeration)	16	302	218	11.90	12,000
Moscow	12	200	333	12.20	2511
Beijing	19	344	574	18.59	16,800
Prague	3	61	65.5	1.28	496
Shanghai	14	364	588	24.15	6341
New York	27	468	368	9.00	1213
Los Angeles	6	80	141	3.93	1302
Rome	3	60	74	2.87	1287
Istanbul	5	73	95.3	14.80	1539

Table 2. Selected metro networks in the world.

The dependence between the population and the number of metro lines is not strict, but some regularity can be observed (Figure 2).



Figure 2. Dependency between the population size and the number of metro lines.

As is can be noted except the case of New York (off-standing red square), all other points are more or less aligned along the regression line. The conclusion is as follows: one metro line should be for about 800,000 citizens. In the same way, the overall number of stations can be determined (Figure 3).



Figure 3. Dependency between the number of metro lines and the number of the stations.

In contrast to the previous issue, there is no exception here—the regression line fits the data points very well, and the average value of 19 stations per line can be stated as a conclusion.

There is a population of about 2,700,000 inhabitants living in considered region; therefore, three metro lines were assumed with, respectively, 54 stations. Additionally, some calculations were done for four lines, keeping the same number of stations.

4.3. Selection of the Relevant Values of Optimization Parameters

An optimization process is controlled by a number of parameters. In the case of genetic algorithm, the most important are the population size, the number of generations, the probability of mutation, and the probability of cross-over. Some additional parameters are possible depending on the details of the problem. The increase of the population size leads to better convergence and allows for avoiding local optima, but causes the calculation time to grow up. The size of population was preliminary determined by trial and error; finally, it was equal to 250 individuals at the first stage (stations locations) and 1000 at the second stage (lines layout). To select the relevant number of generations, the optimization process was monitored, and then it was stopped when the progress of increasing the objective function value significantly slowed down.

The most important of the other parameters for the first stage are the probability of mutation (PM), the probability of cross-over (PX), the elite size (ES), and the standard deviation for the station shift (ND, expressed as a fraction of the region diameter). The parameters taken into account in second stage are PM, PX, and ES. A genetic algorithm is a non-deterministic procedure, so there is a need to perform a series of algorithm runs for each combination of parameters. If each parameter took only two values (from a typical range), the full analysis of all possible parameters combinations would require 16 series at the first stage and 8 series at the second stage. Therefore, the values of the parameters were determined by the use of Taguchi method [36]. The L8 and L4 orthogonal tables were used to design these preliminary tests. Taguchi proposed a procedure of determining the relative importance of the examined parameters. This idea involves an application of a quantity called signal to noise ratio (S/N). This quantity is taken from signal theory, and it expresses the relation between the signal—a desired feature and the noise—undesired disturbing factors. In the considered case, as high value of the objective function as possible is desired, so according to Taguchi the S/N value should be expressed as follows:

$$SN_i = -10\log_{10}\frac{1}{n}\sum_{j=1}^n \frac{1}{X_{ij}^2}$$
(21)

Tables 3 and 4 show the selected parameters values and the optimization process outputs (scaled values of the objective function and S/N) for both stages. The input data described a simplified metro system with 30 stations and three lines to make these preliminary calculations quick. Four procedure runs were carried out in each series. All the values of the objective function were referred to the highest one (normalized to unity).

Series	Parameters				Scaled Objective Function				S/N
	PM	РХ	NE	NS	1	2	3	4	0/11
1	0.1	0.5	1	5	0.9900	0.9991	0.9923	0.9702	-0.1075
2	0.1	0.5	1	8	0.9646	0.9920	0.9560	0.9905	-0.2164
3	0.1	0.8	5	5	0.9672	0.9555	0.9527	0.9559	-0.3746
4	0.1	0.8	5	8	0.9825	0.9702	0.9899	0.9668	-0.2001
5	0.2	0.5	5	5	0.9308	0.9590	0.9705	0.9598	-0.4028
6	0.2	0.5	5	8	0.9879	0.9372	0.9753	0.9619	-0.3092
7	0.2	0.8	1	5	0.9744	0.9586	0.9410	0.9635	-0.3624
8	0.2	0.8	1	8	0.9333	0.9613	1.0000	0.9667	-0.3144

Table 3. Scaled outputs for the first stage according to L8 orthogonal table.

Series	Parameters			Scaled objective Function				S/N
	PM	РХ	NE	1	2	3	4	- 5/1
1	0.1	0.5	2	0.9190	0.9868	0.9037	0.9506	-0.5521
2	0.1	0.8	8	0.9733	0.9868	0.9162	0.9162	-0.4780
3	0.2	0.5	8	0.9394	0.9715	1.0000	0.9506	-0.3134
4	0.2	0.8	2	0.9318	0.9411	0.9162	0.9557	-0.5757

Table 4. Scaled outputs for the second stage according to L4 orthogonal table.

In the next step, the average values of S/N for each parameter and for each value level must be calculated, and then charts showing the relative importance of each parameter can be constructed. Figure 4a,b show this and then allow for determining the optimal parameters for both stages of thegenetic optimization.



Figure 4. Relative importance of control parameters: (**a**) for stations locations optimization and(**b**) for line layout optimization.

Such adjusted parameters were further applied for the final optimization. The described above procedure allowed for determining the best conditions for the optimization process. What is worth noting, parameters PM and NE were adjusted in a different way for both optimization stages (lower values of PM and NE are preferred in the first stage; meanwhile, higher are more appropriate for the second).

4.4. Obtained Structures of the Metro Network

Figure 5 presents the adopted map of population density [37]. Cities are marked according to Table 1. Due to avoiding the uncertainty resulting from non-deterministic nature of genetic algorithm, a series of 10 runs of optimization procedure was carried out. The obtained results were similar but not the same. Figure 6 shows as examples two of the obtained stations layouts. The shaded circles mark the degree of satisfying the transportation demands, and bold black circles mark the additional generators of the transportation demands. These generators and relevant number of people demanding the transport have been added after analyzing the list of such facilities at the examined area. As it can be seen, most of the stations are located mainly on highly populated areas, which was expected. However, the detailed locations of stations differ significantly. Some of the transportation demands generators are not served as well (however, it concerns those of lesser significance). The main reason is probably that the number of stations was underestimated, and it was too low to sate all the highly populated areas. Additionally, as was stated, the objective function did not take into account the details of local topography.



Figure 5. Map of population density (white—below 250 pers./km², black—above 4000 pers./km²).



Figure 6. Examples of obtained stations layouts.

In the light of such variability of the obtained results, a conclusion arose that the found stations layouts should be regarded just as the preliminary. The determination of the exact location of each station would require further work and far more detailed data. Anyway, such a matter could be expected due to simplicity and rawness of the model. On the other side, the values of fit function were quite close to each other. This is shown in Figure 7, which presents the progress of the optimization process for five runs. The value of fit function was normalized to unity.

As it can be noted, all the solutions were very close to the final value of fit function about 2000th generation. All of them converged in the same manner: the optimization progress was very quick during the first thousand generations; next, it slowed down; and; finally; it came up to the same value.

In the next step, the obtained stations layout with the highest value of fit function was the input data for lines topology optimization. The cost of new-build metro line per unit length depends on local conditions, and this amount can vary even by a factor of 12 [38]. This variety comes mainly from difficult urban infrastructure, geology land acquisition, and type of construction. Table 5 shows the indicative costs of the construction of a metro line per one kilometer. Since the considered region includes also some less urbanized areas, a big part of the metro network could be constructed at the ground level. The underground or elevated facilities should be built in the centers of the cities.

Because the presented idea of building of a metro system is just in the initial phase, no cost analyses, even superficial ones, have been carried out on this issue.



Figure 7. Progress of the process of the station layout optimization.

Table 5. Indicative costs of a metro line per one kilometer [38]

Construction Type	Costs [US \$ million]	Ratio
Ground level	15–30	1
Elevated	30-75	2-2.5
Underground	60–180	4–6

In such a situation, no explicit costs were entered into the model and the costs were treated relatively. Therefore, just the simplified map of the building cost coefficient (k) was introduced (Figure 8). Light gray areas mark the normal cost for less urbanized areas (ratio = 1); dark gray areas mark the increased cost due to dense built-up terrains (ratio = 2), and black areas mark the costs for the city centers (ratio = 4).



Figure 8. Map of the construction cost coefficient.

The optimization of lines topology was carried out for different assumptions concerning the number of lines and the formulation of the objective function according to the below scheme (Table 6). Disabling Equation (8) means the coherency of the metro network is assured, but the average travel time is not minimized.

		Objective Function					
		Full (Equation (9))	Equation (8) Disabled				
No of	3	А	В				
lines	4	С	D				

Table 6. Scheme of the optimization of lines topology.

The results of optimization of lines topology are shown in Figure 9. As previously, a series of 10 calculations was carried out for each case, and the best solutions are shown. The same regularity was observed too: despite almost the same values of fit function, the obtained topologies of metro network differed significantly. Because the fit function was determined strictly here, a different explanation is suggested. The possible reason for such an effect is the equifinality of the problem. This means the same desired result can be achieved in different ways.



Figure 9. The best solutions of lines topology for all examined cases (according to Table 6).

There is a conspicuous difference between pairs of cases A, C and B, D: the latter pair assures the coherency of the metro network, but sometimes it is needed for a long journey between stations despite their physical proximity. In contradiction, very dense network for cases A and C allows for fast travelling between any pair of stations, although sometimes with more than one transfer. This difference of the radiation accessibility is the obvious effect of the different formulations of the objective function. The short summary of the obtained results is shown in Table 7.

Case	Relative Cost	Longest Trip [km]	Average Trip [km]	Relative Total Distance Travelled (Equation (6))
А	0.98	59.6	21.7	0.61
В	0.76	98.1	35.2	1.00
С	1.00	60.3	22.0	0.62
D	0.76	91.3	33.6	0.95

Two quantities (building cost and total distance travelled) are expressed relatively (referring to the maximal values) due to avoid large numbers without meaningful units. Having in mind the non-deterministic nature of applied optimization method, one can easily conclude that the metro network performance does not depend strongly on the number of the lines. In the light of data in Table 6, cases A and C seem to be almost the same; this concerns cases B and D as well. Therefore, a conclusion has appeared that the total construction cost does not depend really on the number of lines. The proposed model does not provide the premises to make a decision on the number of lines: one could expect the lower number of transfers for three lines, but a system with four lines seems to be less prone to disturbances. The choice between coherency (variants B and D) and highly interconnected network (variants A and D) is dependent obviously on the available budget.

Figure 10 shows the progress of optimization process for 5 runs of case A.



Figure 10. Progress of the process of the optimization of lines topology.

The progress of the optimization slowed down at about the 5000th generation. Therefore, the 10,000 generations can be regarded as the number sufficient to obtain the converged solution. As previously, the optimization process ran in a similar manner: quick and significant improvements of the solution were made in the beginning phase; afterwards, only rare and small corrections were made.

5. Conclusions

The paper presents a method allowing one to obtain a quick draft of a metro network. Since the task is computationally complex, the two-stage approach with the use of the genetic algorithms was proposed. The method was applied to prepare a preliminary design of the metro network for the central part of the Upper Silesia region in Southern Poland. In the first stage, the optimal locations of stations accounting for population density were found. In the second stage, the optimal lines topology was established; here, two alternative assumptions were considered: just the network coherency was required, or the high radiation accessibility was expected. Two variants of the metro network with three or four metro lines were checked too.

The obtained results were the effect of a few hours of calculations using a standard PC with i7 CPU 2.9 GHz and 8GB RAM. Some effort was required to tune the optimization parameters. The input data were in the form of simple bitmaps, which allowed one to examine different variants of the assumed inputs.

The presented approach can be helpful regarding preliminary considerations of a metro system structure. As no precise data concerning transportation demands and construction costs were introduced into the model, the obtained results cannot be regarded as the final design of the metro system. This concerns mainly the problem of stations locations—this issue requires additional considerations on street topology and the location of other urban facilities. The presented results can be the basis for discussing some variants of stations layout or lines topology. They can be subjected to the public discussion. They can also be used when estimating the required budget.

An important finding is that the requirement of the high radiation accessibility will result in a significant increase of building costs (of about 30%). Otherwise, the increase of the average trip length should be expected (of about 60%). The finally chosen option depends on the budget available. Obviously, a trade-off solution is reasonable as well, but it cannot be found using the presented model as it is currently. At present, the model just minimizes the relative construction costs. Thus, it should be extended to take into account the explicit budget limitations and the real construction costs. It was also shown that the level of service and the costs do not depend significantly on the number of the metro lines.

Despite the mentioned limitations, the obtained results give an insight into the designed metro system structure. The main advantage of the presented method is the possibility to quickly make a draft of topology of the metro network by introducing only the basic data describing the analyzed region. The simplicity of the presented model makes it possible to easily change the introduced data to analyze different scenarios of the future development of the region.

All the above remarks do not mean that the applied method—the genetic algorithm—is unable to provide a detailed design of a metro system or any of its components. Artificial intelligence methods, including genetic algorithms, can be applied at any level of the design process. The accuracy of the obtained results is limited only by the model structure, not by the applied optimization method.

Author Contributions: A.K. participated in the preparation of the research. He built the numerical models, tuned the controlling parameters, and took part in the optimization analyses. He participated in the writing of the article. M.K. took part in the data acquisition and in the optimization analyses. She analyzed the obtained results. She participated in the writing of the article as well.

Funding: This research received no external funding.

Acknowledgments: Publication supported as a part of the Rector's habilitation grant. Silesian University of Technology, grant number 12/050/RGH18/0071.

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

References

- Li, Z.; Luo, J. The evolution analysis of Guangzhou subway network by complex network theory. *Procedia Eng.* 2016, 37, 186–195. [CrossRef]
- Samanta, S.; Jha, M.K. Identifying feasible locations for rail transit stations two-stage analytical model. *Transp. Res. Rec. J. Transp. Res. Board* 2008, 2063, 81–88. [CrossRef]
- 3. Lee, E.S.W.; Leung, C.K. Transport planning for new town development: Experience of Hong Kong. *J. Chin. Geogr.* **1994**, *4*, 1–13.
- 4. Hess, D.B.; Almeida, T.M. Impact of proximity to light rail rapid transit on station-area property values in Buffalo. New York. *Urban Stud.* **2007**, *44*, 1041–1068. [CrossRef]
- 5. Calvo, F.; de Ona, J.; Arán, F. Impact of the Madrid subway on population settlement and land use. *Land Use Policy* **2013**, *31*, 627–639. [CrossRef]
- 6. Yu, H.; Zhang, M.; Pang, H. Evaluation of transit proximity effects on residential land prices: An empirical study in Austin, Texas. *Transp. Plan. Technol.* **2017**, *40*, 841–854. [CrossRef]
- 7. Yang, J.; Chen, J.; Le, X.; Zhang, Q. Density-oriented versus development-oriented transit investment: Decoding metro station location selection in Shenzhen. *Transp. Policy* **2016**, *51*, 93–102. [CrossRef]
- 8. Li, L.; Ren, H.; Zhao, S.; Duan, Z.; Zhang, Y.; Zhang, A. Two dimensional accessibility analysis of metro stations in Xi'an, China. *Transp. Res. Part A* **2017**, *106*, 414–426. [CrossRef]
- 9. Cheng, J.; Bertolini, L. Measuring urban job accessibility with distance decay, competition and diversity. *J. Transp. Geogr.* **2013**, *30*, 100–109. [CrossRef]
- 10. El-Geneidy, A.; Levinson, D.; Diab, E.; Boisjoly, G.; Verbich, D.; Loong, C. The cost of equity: Assessing transit accessibility and social disparity using total travel cost. *Transp. Res. Part A* **2016**, *91*, 302–316. [CrossRef]
- Tasic, I.; Bozic, C.; Hanss, E. Towards True Multimodal Transportation Accessibility: Data, Measures, and Methods. In Proceedings of the 96th Annual Meeting of the Transportation Research Board 2017, Washington, DC, USA, 8–12 January 2017.

- 12. Cancela, H.; Mauttone, A.; Urquhart, M.E. Mathematical programming formulations for transit network design. *Transp. Res. Part B* 2015, *77*, 17–37. [CrossRef]
- 13. Cai, J.; Zheng, C.; Yang, M. Research on Rail Transit Network System and its Connection Model in the Metropolitan Area. *Procedia Soc. Behav. Sci.* 2013, *96*, 1286–1292. [CrossRef]
- 14. Laporte, G.; Mesa, J.A.; Ortega, F.A. Optimization methods for the planning of rapid transit systems. *Eur. J. Oper. Res.* **2000**, *122*, 1–10. [CrossRef]
- 15. Laporte, G.; Mesa, J.A.; Ortega, F.A.; Sevillano, I. Maximizing trip coverage in the location of a single rapid transit alignment. *Ann. Oper. Res.* **2005**, *136*, 49–63. [CrossRef]
- 16. De Sá, E.M.; Contreras, I.; Cordeau, J.F. Exact and heuristic algorithms for the design of hub networks with multiple lines. *Eur. J. Oper. Res.* **2015**, *246*, 186–198. [CrossRef]
- Canca, D.; De-Los-Santos, A.; Laporte, G.; Mesa, J.A. The Railway Network Design, Line Planning and Capacity Problem: An Adaptive Large Neighborhood Search Metaheuristic. In *Advanced Concepts, Methodologies and Technologies for Transportation and Logistics. EWGT 2016, EURO 2016;* Zak, J., Hadas, Y., Rossi, R., Eds.; Advances in Intelligent Systems and Computing; Springer: Cham, Switzerland, 2018; Volume 572. [CrossRef]
- 18. Ferdyn-Grygierek, J.; Grygierek, K. Multi-variable optimization of building thermal design using genetic algorithms. *Energies* **2017**, *10*, 1570. [CrossRef]
- 19. Martínez, H.S.; Moyano, A.; Coronado, J.M.; Garmendia, M. Catchment areas of high-speed rail stations: A model based on spatial analysis using ridership surveys. *Eur. J. Transp. Infrastruct. Res.* **2016**, *16*, 364–384.
- 20. Król, M.; Białecki, R. Optimization of a window frame by BEM and genetic algorithm. *Int. J. Numer. Methods Heat Fluid Flow* **2003**, *13*, 565–580. [CrossRef]
- Nowakowski, P.; Król, A.; Mrówczyńska, B. Supporting mobile WEEE collection on demand: A method for multi-criteria vehicle routing, loading and cost optimization. *Waste Manag.* 2017, 69, 377–392. [CrossRef]
- 22. Ingram, D.R. The concept of accessibility: A search for an operational form. *Reg. Stud.* **1971**, *5*, 101–107. [CrossRef]
- 23. Kwan, M. Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geogr. Anal.* **2010**, *30*, 191–216. [CrossRef]
- 24. Schlossberg, M.; Brown, N. Comparing transit-oriented development sites by walkability indicators. *Transp. Res. Rec. J. Transp. Res. Board* 2004, 1887, 34–42. [CrossRef]
- 25. Brinckerhoff, P. *Metrorail Bicycle & Pedestrian Access Improvements Study;* Washington Metropolitan Area Transit Authority: Washington, DC, USA, 2010.
- 26. Charles, P.; Galiza, R. *Improving Rail Station Access in Australia*; Project Number: R1.133. Project Name: Station Access; CRC for Rail Innovation: Brisbane, Australia, 2013.
- 27. Mesa, J.A.; Ortega, F.A. Park-and-ride station catchment areas in metropolitan rapid transit systems. In *Mathematical Methods on Optimization in Transportation Systems*; Pursula, M., Niittymäki, J., Eds.; Applied Optimization; Springer: Boston, MA, USA, 2001; Volume 48. [CrossRef]
- Laporte, G.; Pascoal, M.B. Path based algorithms for metro network design. *Comput. Oper. Res.* 2015, 62, 78–94. [CrossRef]
- 29. Curtin, K.M.; Biba, S. The transit route arc-node service maximization problem. *Eur. J. Oper. Res.* **2011**, 208, 46–56. [CrossRef]
- 30. Wu, X.; Dong, H.; Tse, C.K.; Ho, I.W.H.; Lau, F.C.M. Analysis of metro network performance from a complex network perspective. *Phys. A* **2018**, *492*, 553–563. [CrossRef]
- 31. Fedeli, F.; Mancini, R.; Mannino, C.; Ofria, P.; Oriolo, G.; Pacifici, A.; Piccialli, V. Optimal design of a regional railway service in Italy. *J. Rail Transp. Plan. Manag.* **2017**, *7*, 308–319. [CrossRef]
- 32. Chiaradia, A.; Moreau, E.; Raford, N. Configurational exploration of public transport movement networks: A case study, the London underground. In Proceedings of the 5th International Space Syntax Symposium, Delft, The Netherlands, 13–17 June 2005; pp. 541–552.
- 33. Darwin, C. On the Origin of Species by Means of Natural Selection: Or the Preservation of Favoured Races in the Struggle for Life; D. Appleton: New York, NY, USA, 1869.
- 34. Holland, J.H. *Adaptation in Natural and Artificial Systems*; The University of Michigan Press: Ann Arbor, MI, USA, 1975.
- 35. Goldberg, D.E. *Genetic Algorithms in Search. Optimization and Machine Learning*; Addison-Wesley Longman Publishing Co. Inc.: Boston, MA, USA, 1989.

- 36. Roy, R.K. *Design of Experiment Using the Taguchi Approach;* John Wiley & Sons: New York, NY, USA, 2001; ISBN 978-0-471-36101-5.
- 37. Official Journal of Silesian Voivodeship. *Resolution no CXIX/12/2013*; Silesian Voivode: Katowice, Poland, 2013. (In Polish)
- 38. Flyvbjerg, B.; Bruzelius, N.; van Wee, B. Comparison of capital costs per route-kilometre in urban rail. *Eur. J. Transp. Infrastruct. Res.* **2008**, *8*, 17–30.



© 2019 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 (http://creativecommons.org/licenses/by/4.0/).