Raport Badawczy

Research Report

RB/55/2016

Identifying main center access hubs in a city using capacity and time criteria. The evolutionary approach

J. W. Owsiński, J. Stańczak,
A. Barski, K Sęp

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

Control and Cybernetics

vol. **45** (2016) No. 2

Identifying main center access hubs in a city using capacity and time criteria. The evolutionary approach*

by

Jan W. Owsiński[1,2], **Jarosław Stańczak**[1,2], **Aleksy Barski**[2], **Krzysztof Sęp**[1,2]

[1]Faculty of Computer Science
Warsaw School of Information Technology
ul. Newelska 6, 01-447 Warsaw, Poland
{owsinski, stanczak, alex, sep} @ibspan.waw.pl
[2]Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland

Abstract: In this paper we consider the well known Hub and Spoke problem, analyzed in the context of Warsaw Public Transport System. Our method was designed for data preprocessing so as to allow using a timetable obtained from the public transport web site after conversion into the required data format.

A dedicated evolutionary algorithm method that detects the hubs of almost all available transport means was also developed. The hubs identified are well connected to the center of the city and to other identified hubs (characterized by high capacity or short travel time). These hubs may become the skeleton of the public transport system and, in particular, good points for locating Park and Ride facilities.

Keywords: transport, urban transport system, Hub and Spoke, Park and Ride, evolutionary algorithm

1. Introduction

The incredibly fast pace of civilization takes mostly place in the big cities, involving virtually all of their inhabitants. As a result, there are traffic jams and pollution increase and several other problems, relative to urban systems. Fortunately, this is not the case, in which there is nothing that can be done. There are several ideas as to how the large cities can cope with pollution resulting from the exhaust gases. It is a well known fact that public transport produces much less pollution per a single passenger or passenger-mile than the private one and, in general, alleviates traffic congestion. The reason for analyzing and optimizing the city transport system in the perspective that we take

---

*This article is an extended version of the paper "The main quickest center access hubs in a big city. The evolutionary approach." presented at the BOS'2016 conference in October 2016 in Warsaw. Accepted for publication in November 2016.

in the present paper is quite simple: to convince the commuters to rather use the public transport than the private car. One of the ideas is The Park and Ride (P&R) solution - it allows the commuters to get to the city, but not closer to the center than some place in the suburbs or in surroundings, by their cars, and from there to allow them go further by public transport. It is important to ensure that the P&R facilities be located near to the well-connected points, which feature fast and high capacity connections to the different parts of the city (see Fig. 1). It seems to be good to locate these facilities (if this is possible) near transport hubs (Cambpell, O'Kelly, 2012). Such hubs or candidate locations for hubs can be identified by the method, which is proposed in this paper. The idea of Hub and Spoke (H&S) structure, which is widely used in designing transport systems, can be also helpful in the case of the urban transport system, where the fast transport means like the underground or urban trains may constitute a skeleton of the fast and high capacity connections, with slower transport means, like trams or buses, supporting local connections. In order to reorganize the transport system into the instance of the H&S structure it is necessary to detect appropriate hub locations in the graph of the city transport system. This is quite a heavy computational task, and so the evolutionary method, which is often used for solving this type of problems, was selected.

Figure 1. Warsaw and Park and Ride facilities in some main, well-connected points of the city (source: pl.wikipedia.orgwikiParkuj_i_Jedz_w_ Warszawie#mediaFile:Warszawa__Parkingi_P%2BR.svg)

2. Basic concepts

Since we shall be modelling the entirety of the Hub and Spoke problem through the graph representation and then solving it as a problem defined on graphs, we start with the basic notions from graph theory, here given following Wilson (1996).

A ***graph*** is a pair G=(V, E), where V is a non-empty set of *vertices* and E is a set of *edges*. Each edge is a pair of vertices $\{v_1, v_2\}$ with $v_1 \neq v_2$.

Two vertices, indexed i and j, in a graph G=(V, E) are called ***incident*** if for v_i, $v_j \in V$ there is $\{v_i, v_j\} \in E$ or $v_i = v_j$. Each vertex is, therefore, ***incident*** to itself.

A *subgraph* of graph G=(V, E) is a graph G'=(V', E'), where $V' \subseteq V$, and $E' \subseteq E$ such that for all $e \in E$ and $e = \{v_1, v_2\}$ if $v_1, v_2 \in V'$ then $e \in E$.

A ***degree*** of a vertex is the number of edges, to which this vertex belongs.

A ***clique*** (a complete subgraph) $Q = (V_q, E_q)$ in a graph $G = (V, E)$ is a graph such that $V_q \subseteq V$ and $E_q \subseteq E$ and $Card(V_q) = 1$ or each pair of vertices v_1, $v_2 \in V_q$ fulfils the condition $v_1, v_2 \in E_q$, see Cormen et al. (2009). Each subgraph of a clique is a clique.

An α***-clique*** (see Mażbic-Kulma et al., 2008; Potrzebowski, Stańczak and Sęp, 2006, 2007; Stańczak, Potrzebowski and Sęp, 2011) can be defined as follows: let $A = (V, E)$ be a subgraph of graph $G = (V, E), V \subseteq V, E \subseteq E, k = Card(V)$ and let k_i be a number of vertices $v_j \in V$ such that $v_i, v_j \in E$,

1. For $k = 1$ the subgraph A of graph G is an $\alpha - clique(\alpha)$. 2. For $k > 1$ the subgraph A of graph G is an $\alpha - clique(\alpha)$ if for all vertices $v_i \in V$ the condition $\alpha = (k_i + 1)/k$ is fulfilled, where $\alpha \in (0, 1]$.

Further on, we will use the term of $\alpha - clique$ in the sense of $\alpha - clique(\alpha)$ for an earlier established α. A subgraph of an $\alpha - clique$ may not necessarily be an $\alpha - clique$ for the established α.

A graph G=(V, E) is a ***complete graph***, if for each pair of vertices there is an edge $e \in E$ between them.

An ***adjacency matrix*** of a graph G=(E, V) with $Card(V)$=n is a square binary matrix $n \times n$ with rows and columns corresponding to vertices. There is 1 in the a_{ij} cell of the adjacency matrix if vertices v_i and v_j are adjacent, and 0 in the opposite case.

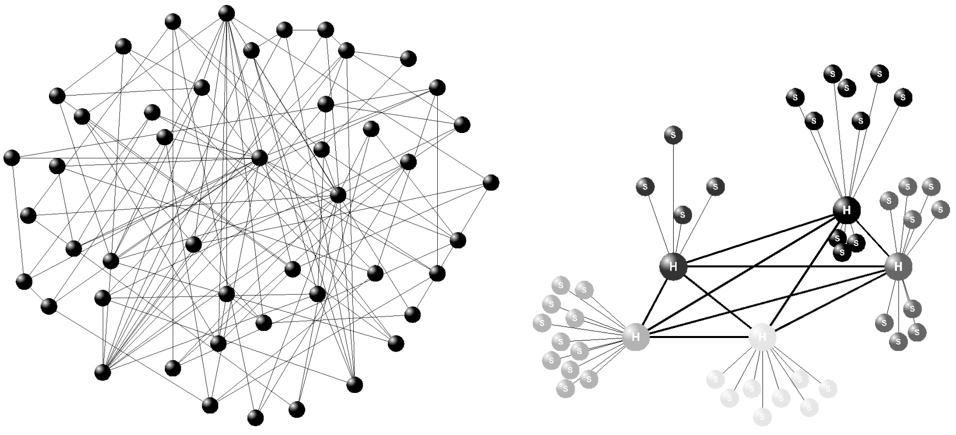
We propose an evolutionary algorithm that transforms the connection graph into an instance of the *hub and spoke* structure according to problem-specific conditions and requirements.

A ***hub and spoke*** structure (Fig. 2b) is a graph H_s=($G_h \cup G_s$, E) where the subset G_h is at least a connected graph[*] with the relevant subset of the set E, each vertex of the subset G_s has degree 1 and is connected exactly with one vertex from the subset G_h (thus forming a spoke) (see Horner and O'Kelly, 2001; O'Kelly and Bryan, 2002; Mażbic-Kulma et al., 2009).

---

[*] The subgraph of hubs should be as close to a clique as possible or should be an instance of the $\alpha - clique$ with α as close to 1 as possible, but in the case of sparse input graph it should just be a connected graph, so as to preserve its basic functionality.

This structure can be used in transport and logistic models, where direct connections between the spoke nodes, attached to respective hubs, are not very important and, thus, are not really necessary (O'Kelly, 1987). The hub and spoke structure can be derived using a predetermined, for instance - by some expert, number of communication hubs with the possibility of directly determining which nodes should become hubs or of selecting them with the use of the solving method.



a) b)

Figure 2. a) a source graph, b) the hub and spoke structure obtained from the source graph.

3. The problem - identification of hubs

We took under consideration an urban transport system, modeled as a graph of connections. It is quite a convenient method for analyzing such system (Coyle, Bardi and Novack, 1994, p.402).

We assume that there is a direct connection between two nodes (stops) if there is a public transport line they both belong to. Thus, the graph, constituting the model of the transportation network, used in our approach, is not a simple graph representation of city transport routes, but a denser structure, with interconnections among stops, belonging to the same transport line. Different lines may have common stops, so the transport graph consists of overlapping blocks of lines. In our approach, the graph of connections is a weighted graph, where all edges may have different values assigned. These values may represent not only the very fact of existence of a connection, but also some parameters of the connection:

- the number of scheduled courses per one connection (frequency)

- the travel time
- the potential capacity of the given means of transport (within a given period of time)
- the current/estimated number of passengers (in the vehicle: descending/boarding).

In this paper we concentrate on weights representing the connection capacity and the average travel times between stops.

Our evolutionary method is meant for picking some points, which can be treated as candidates for transport hubs. All of them should not necessarily become the hubs - the final decision should be left, according to our perspective, to an expert.

3.1. Problem definition

The problem is to find the points or hubs that are relatively close to the center in the sense of:

- the averaged and weighted trip time
  The quality function forces the evolutionary computations towards obtaining the desired graph structure. In the here considered problem, the quality function is formulated so as to direct the EA towards the desired H&S structure, accounting for the weights among the graph nodes being averaged trip times, so we try to minimize them between the identified potential hubs:

  $$min(Q) = \sum_{i=1}^{n}(\sum_{j=1}^{m} w_{ij}^{C} + \sum_{l=1}^{n} w_{il} + \sum_{p=1}^{k} w_{ip}) \tag{1}$$

  where: n is the imposed number of possible hubs, k is the number of nodes (stops) in the considered graph, w_{ij}^{C} is the weight of connection between the i^{th} hub and the j^{th} stop of the set of m important stops in the center of the city, w_{il} is the weight of connection between the i^{th} hub and the l^{th} hub, w_{ip} is the weight of connection between the i^{th} hub and the p^{th} node of the graph.
- the capacity of connections
  In this case the quality function plays the same role as in the preceding case, only the weights among the graph nodes have the meaning of the capacity values, which are maximized:

  $$max(Q) = \sum_{i=1}^{n}(\sum_{j=1}^{m} w_{ij}^{C} + \sum_{l=1}^{n} w_{il} + \sum_{p=1}^{k} w_{ip}) \tag{2}$$

  where: n is the imposed number of possible hubs, k is the number of nodes (stops) in the considered graph, w_{ij}^{C} is the weight of connection between the i^{th} hub and the j^{th} stop of the set of m important stops in the center of the city, w_{il} is the weight of connection between the i^{th} hub and the l^{th} hub, w_{ip} is the weight of connection between the i^{th} hub and the p^{th} node of the graph.

Each hub should be close to the other hubs and, if it is possible, to the rest of stops. By the center we mean a subset of stops, chosen a priori, situated in the center of the city. Getting to the city is the here assumed aim of all passengers.

3.2. Preprocessing of data

The input data were obtained from the ZTM (Warsaw Transport Board) site and preprocessed by our application. As a result, we obtained an example of the weighted adjacency matrix. The matrix fulfills the following conditions:

1. Two vertices are adjacent if and only if there exists a line to which they both belong.
2. In the case of findig hubs with the shortest connection time, the weight of the edge is an average trip time divided by the number of segments between stops belonging to the trip.
3. In the case of looking for the highest capacity hubs, the weight of the edge is the minimum of capacities of all the consecutive components of the route.
4. In the case of many lines that operate the path between two stops, the average value (the shortest time problem) or the sum of capacities (the highest capacity problem) of all courses is taken.

In view of the fact that there may be one way segments on some routes and lines, loops or different routes in both directions, the obtained matrix is, in general, not symmetric.

3.3. Complexity of the problem

As we mentioned before, the problem is not easy to solve. With the assumption that the trip time between each pair of stops is 1 we deal, in fact, with the k-vertex cover problem, see Cormen et al. (2009). So, the k-hub problem is at least NP hard.

3.4. The evolutionary method of hub identification

The scheme of the standard evolutionary algorithm (EA) is presented in Algorithm 1.

It is, naturally, possible to use the standard EA formulation from Algorithm 1 to solve the problem here stated, but in order to obtain a truly efficient method it is necessary to develop a specialized version of EA (Michalewicz, 1996). This is quite difficult, because there are no precise rules for doing it. First, it is necessary to invent the proper encoding of solutions, then to develop specialized genetic operators, tailored for the solved problem, and finally, to formulate the fitness function to be optimized by the algorithm. In this case, the solution can be encoded as a set of vectors with the index number of the

Algorithm 1 The standard evolutionary algorithm.

```
Input: Input data
Output: Output data

begin

Random initialization of the population of solutions (or individuals)

while (stop condition is not satisfied )
        begin
                Reproduction and modification of solutions using genetic operators
                Valuation of obtained solutions
                Selection of individuals for the next generation
        end;
end;
```

hub node and with attached index numbers of spoke nodes (the number of hub nodes being imposed). The set of developed genetic operators consists of:

- mutation - exchange of randomly chosen nodes from different sets of spokes,
- relocation of a randomly chosen spoke of one hub node to a different, randomly chosen set of spokes,
- exchange of a randomly selected hub for a randomly selected spoke.

Application of more than two specialized genetic operators requires a method of selecting them for execution in the consecutive iterations of the algorithm. In the EA here used (Stanczak, 2003), every individual is characterized by an additional vector of floating point numbers, besides the encoded solution. Each number corresponds to one genetic operator and is a measure of its quality and after normalization becomes the value of probability of its execution. The higher is the factor, the higher is the probability of executing the corresponding operator. The method of updating quality factors is based on reinforcement learning (Cichosz, 2000; Sutton and Barto, 1998). When the operator, selected by the individual i is applied, this can be regarded as an agent performing action a_i leading him to a new state s_i, which, in this case, is a new solution. The agent receives a reward or penalty, depending on the quality of the new state. The aim of the agent is to perform the actions that secure obtaining the highest long term discounted cumulative reward V^*, maximizing the chances of its offspring to evolve in next generations.

The following formula is used for evaluation purposes:

$$V(s_{t+1}) = V(s_t) + \alpha * (r_{t+1} + \gamma * V^*(s_{t+1}) - V(s_t)) \tag{3}$$

where: t - current time, $V(s_t)$ - quality factor or discounted cumulative reward, $V^*(s_{t+1})$ - estimated value of the best quality factor (in our experiments we take the value attained by the best operator), α - the learning factor, γ - the discount factor, r_{t+1} - reward for the best action, equal to the improvement of the quality of solution after execution of the genetic operator.

4. Real data

As the methods of EA, and especially their more complex modifications, like the one applied in this problem, cannot be formally proved for their properties, especially as concerns the optimum or the divergence from the optimum, we provide here an evaluation of the method, based on the real set of data.

For this purpose we use the data, concerning Warsaw public transport, taken from the official site of the ZTM (ftp://rozklady.ztm.waw.pl/). The whole list of stops contains more than 10 000 items, but for evaluation purposes we reduced it to more than 2 500, combining the stops, bearing the same names, but oriented in opposite directions, or divided into sub-stops at bigger terminals.

The data contain, in particular, information about:

- the type of the day (a holiday or an ordinary day)
- schedules for an ordinary day
- schedules for a holiday
- the list of the transport lines
- the list of all stops, containing:
    - a unique stop number
    - name of the stop group
    - the town token
    - the name of the town
- the list of lines assigned to each stop.

For the purpose of generating the adjacency matrix, a dedicated application was designed and implemented.

The matrix contained the weights of the edges in the transport graph. Each weight corresponds to the distance between stops. By distance we understand here the time needed to go from one stop to another.

5. Results obtained

In this section we present some of the obtained results for the imposed number of 20, 50 and 80 potential communication hubs for both problems considered. The big dots in the respective figures represent the stops, which ensure fast or high capacity (depending on the solved problem) connections to the virtual center of the city and to other hubs. As it was expected, the chosen points (stops) are connected with fast rail communication means, like underground, fast city trains, suburban trains and trams. This demonstrates that development of such means of transport, even though expensive, should be a priority for modern big cities. Buses, it appears, should play a secondary role and ought to be used for local transport from hubs to less important destinations.

Table 1. Selected fragments of the data file

```
300305  DWORKOWA,                       --    Y= 52.204978     X= 21.022889
   *TD  3
      DP  DZIEŃ POWSZEDNI
         *WG  19
            G  3   5:  [18]   38    58
            G  4   6:   18   [27]^  39   [58]
            G  5   7:   11    24   [36]   44   [52]
            G  8   8:   00    08    16   [24]   32    40    48   [56]
            G  6   9:   04    15   [25]   29m^  36    48
            G  5  10:  [00]   12   [24]   36    48
            G  5  11:   00   [12]   24    36   [48]
            G  5  12:   00   [12]   24    36   [48]
            G  5  13:   00   [12]   24    36    48
            G  8  14:  [00]   07^   12    24    29^  [36]   48    54^
            G  7  15:  [00]   12    20   [28]   36   [44]   52
            G  8  16:   00    08    16   [24]   32    40    48    56
            G  7  17:  [04]   12    20   [28]   36    44   [52]
            G  7  18:   00   [08]   16    24    32    40   [48]
            G  7  19:   00    12    19m^  24   [36]   42m^  48
            G  6  20:  [00]   11   [22]   34    44    58
            G  5  21:   05m^ [18]   38    48m^  58
            G  3  22:  [18]  [38]  [59]
            G  3  23:   08m^  19   [39]m^
         #WG
         *OD 107
             5.18  TP-WYS/DP/04.35___
             5.38  TP-WYS/DP/04.55___
             5.58  TP-WYS/DP/05.15___
```

5.1. The problem of minimizing the travel time

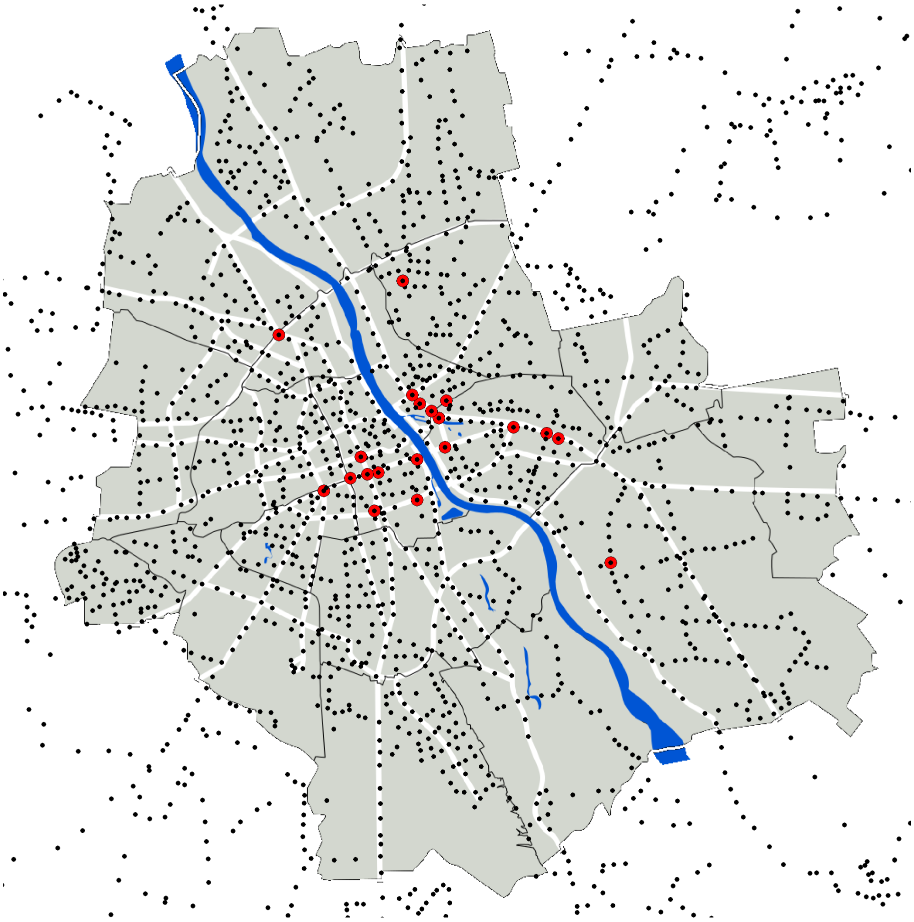


Figure 3. Results obtained for the assumed number of 20 candidates for hubs (the shortest time problem)

5.2. The problem of maximizing the connection capacity

Even a superficial observation of the series of results provided in the figures, presenting the maps of Warsaw with the candidate stops indicated as bigger dots, makes apparent several conclusions of practical character, namely:

== for smaller number of required hubs, in both problems, the selected stops concentrate close to the center and along the primary transport routes of the city;

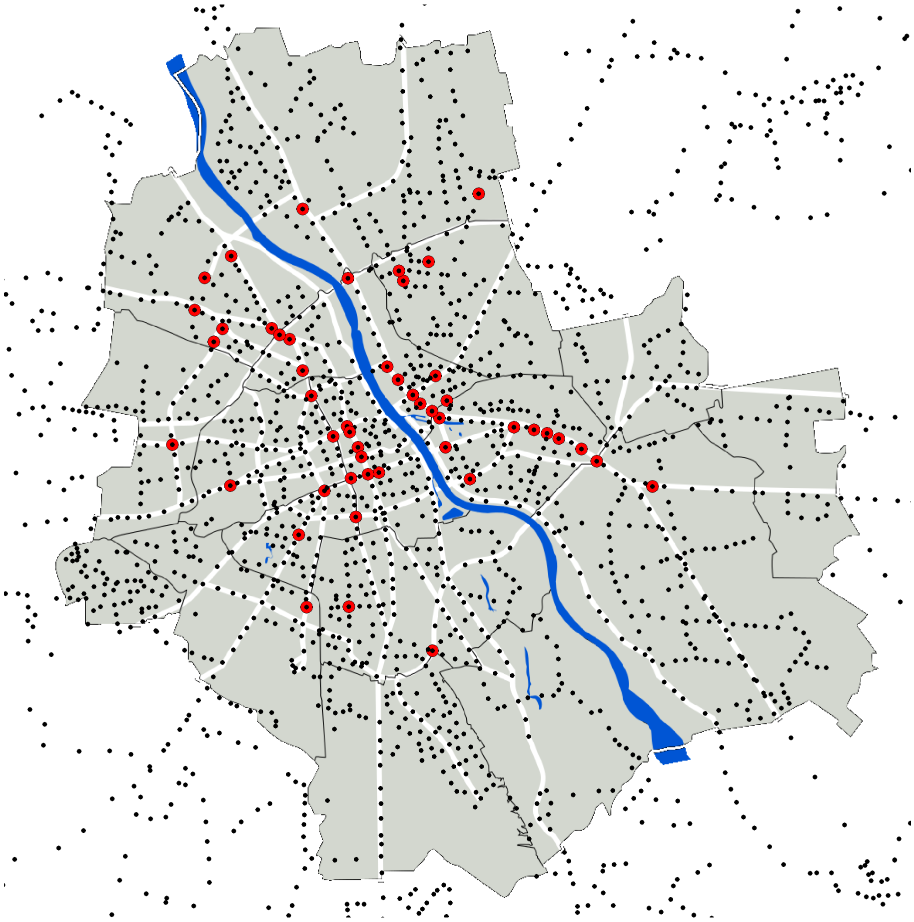


Figure 4. Results obtained for the assumed number of 50 candidates for hubs (the shortest time problem)

== the maximum capacity problem appears to be more conducive to bigger dispersion of the locations, especially for the smaller number of selected stops;

== there are very clear directions and lines, along which the candidate stops appear first, when the required number of hubs is increased (especially for the shift from 20 to 50 hub candidates).

All this, along with more detailed considerations, concerning the concrete areas and lines, should constitute the basis for further analyses of the transport system development for the agglomeration of Warsaw.

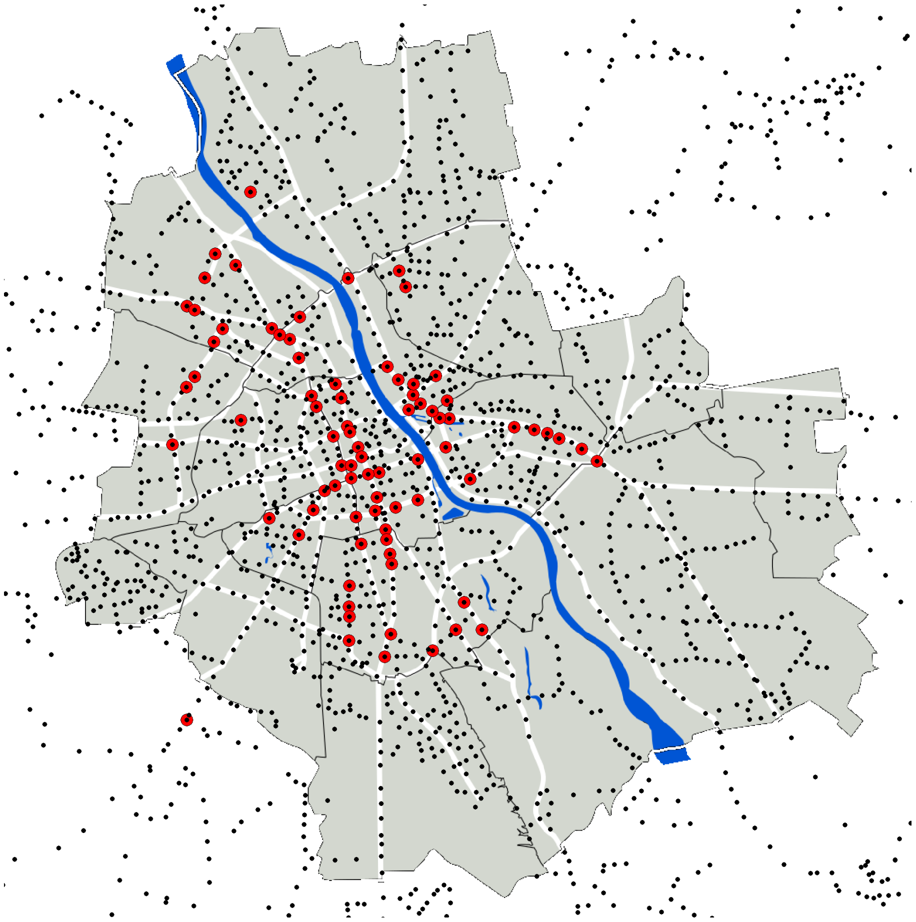


Figure 5. Results obtained for the assumed number of 80 candidates for hubs (the shortest time problem)

5.3. The comparison of detected hubs

Although different criteria were used for the two considered problems, several identical hubs were chosen for both cases. In our opinion, these stops seem to be very important nodes in the city transportation system. Here in Fig. 9 we present the results of such comparison for 20 hubs.

6. Conclusion

It is obvious that the necessity and the actual design of the ways to improve city transport systems is nowadays a big challenge for scientists, engineers and urban authorities. However, such factors as:

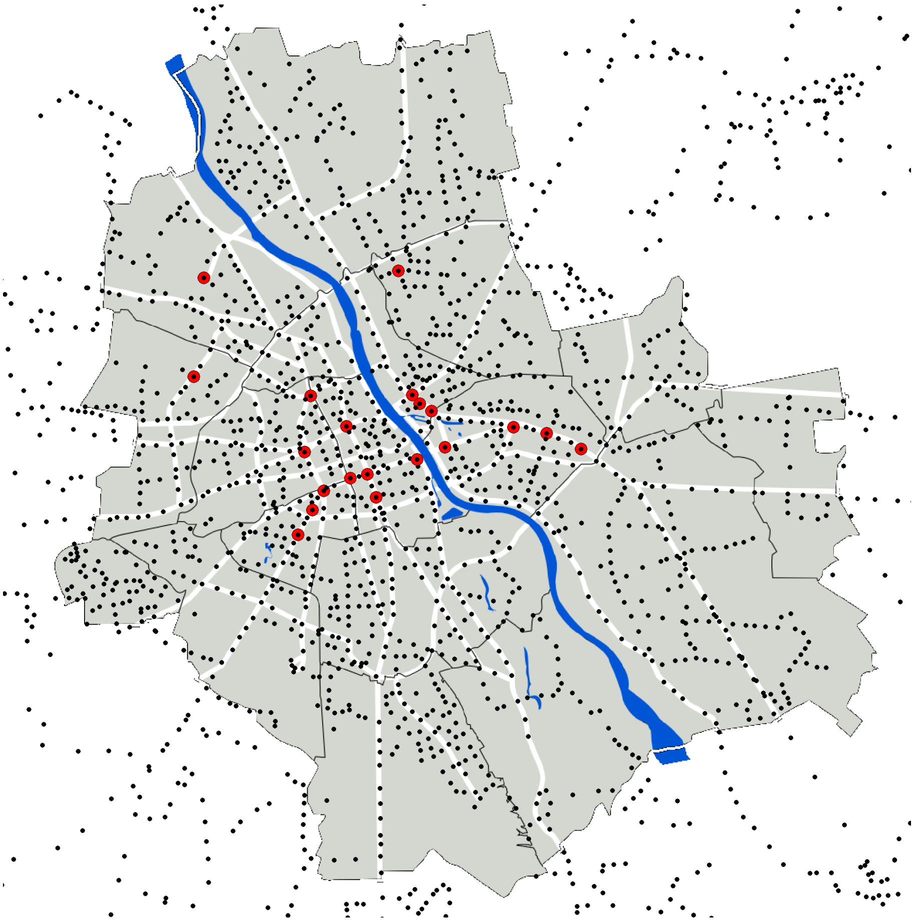


Figure 6. Results obtained for the assumed number of 20 candidates for hubs (the highest capacity problem)

- decrease of pollution
- alleviation of traffic jams
- lowering of total costs
- increase in trip comfort
- shortening of travel times

should constitute a sufficient reason for embarking on respective undertakings. Our method, presented here, provides a tool for indicating the most important points in the city transport network that should be taken under consideration when analyzing the potential directions in the city development planning. Our method can be useful, as well, in the other branches of broadly understood logistics (Cetiner, Sepil and Süreal, 2010; Lin and Chen, 2003; Min and Gou,

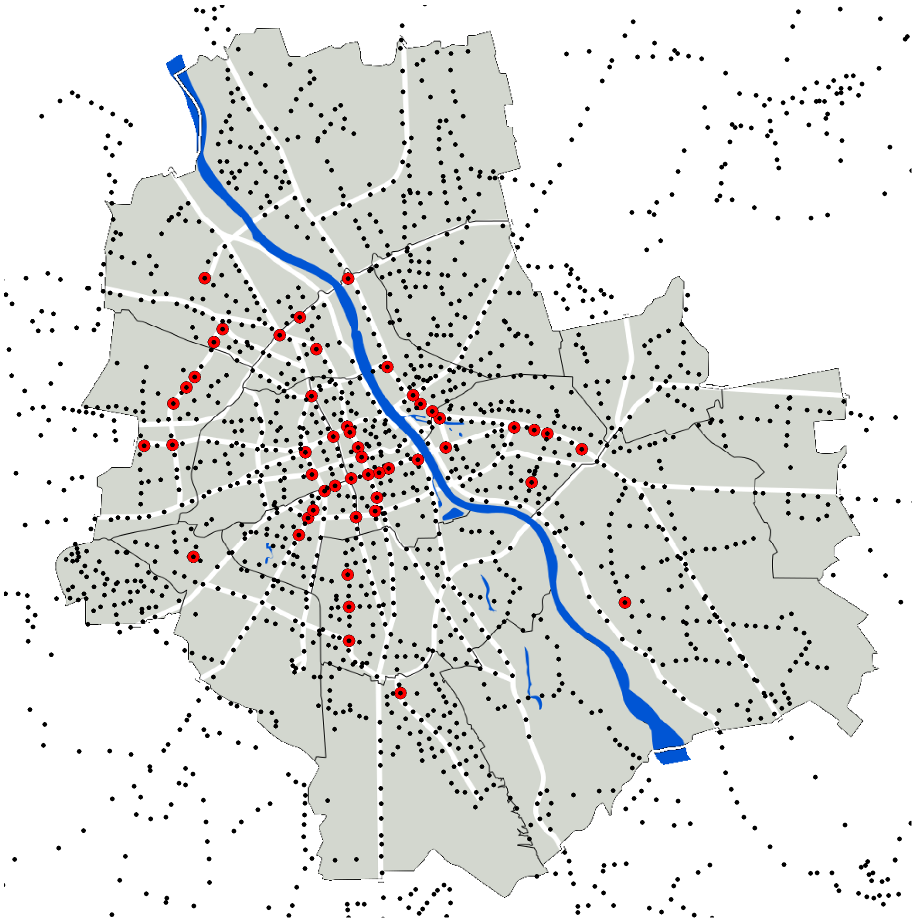


Figure 7. Results obtained for the assumed number of 50 candidates for hubs (the highest capacity problem)

2004). Hubs may become not only P&R facilities, but also main transport change points, and be used for designing new tube and other fast transport means lines. Our method can, likewise, be useful in the analysis of the cell phone networks or social networks. This method may be applied in every model, which can be improved by using the hub and spoke structure.

References

Campbell, J.F., O'Kelly, M.E. (2012) Twenty-Five Years of Hub Location Research. *Transportation Science*, **46** (2), 153–169.

Çetiner, S., Sepil, C., Süral, H. (2010) Hubbing and routing in postal delivery systems. *Annals of Operations Research*, 181, 109–124.

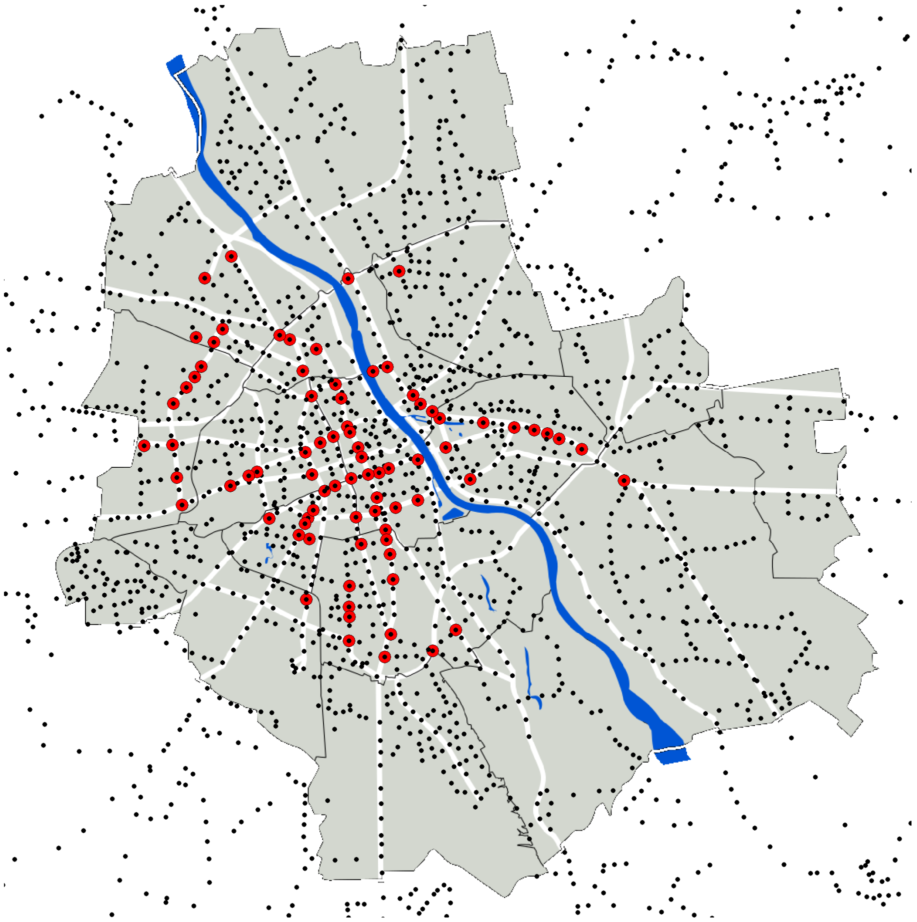


Figure 8. Results obtained for the assumed number of 80 candidates for hubs (the highest capacity problem)

Liu, Ch.-Ch. and Chen, Y.-Ch. (2003) The integration of Taiwanese and Chinese air networks for direct air cargo services. *Transportation Research Part A: Policy and Practice* **37**, 7, August 2003, 629–647.

Cichosz, P. (2000) *Systemy uczące się (Self-learning systems; in Polish).* WNT, Warszawa.

Cormen, T. H., Leiserson C. E., Rivest R. L., Stein. C. (2009) *Introduction to Algorithms.* MIT.

Coyle, J.J., Bardi E.J, Novack R.A., (1994) *Transportation.* Fourth Edition, West Publishing Company, New York.

Horner M., O'Kelly M.E. (2001) Embedding economy of scale concepts for hub network design. *Journal of Transport Geography*, **9** (4), 255–265.

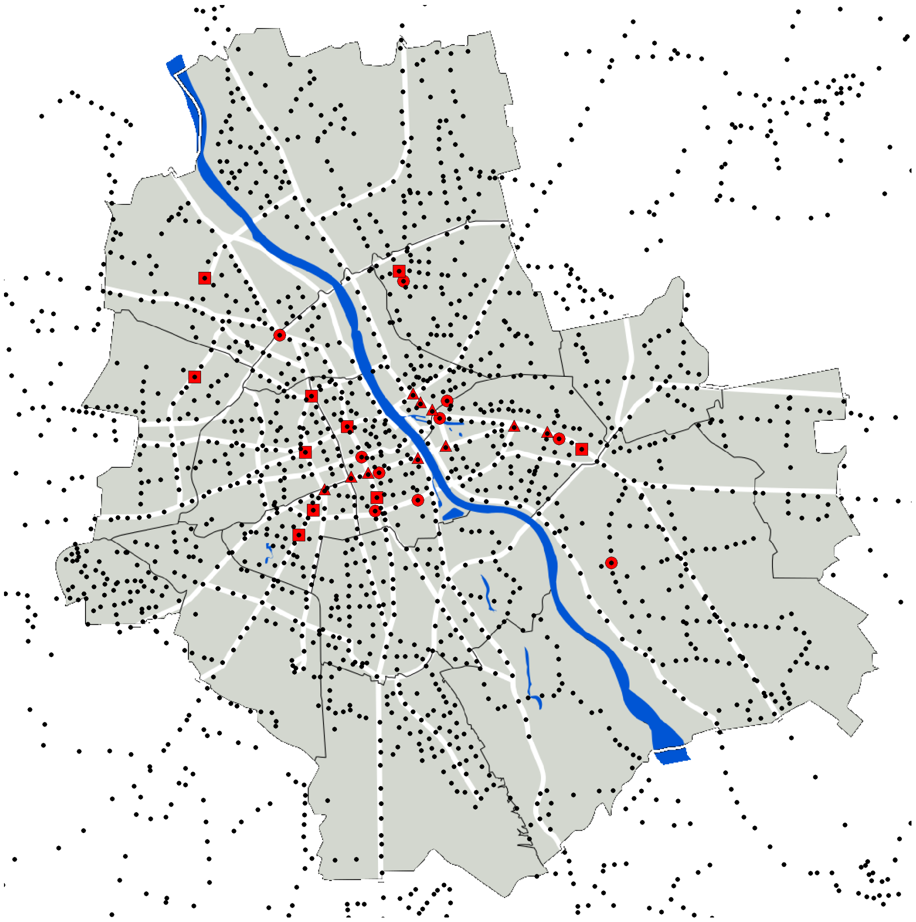


Figure 9. Results of comparison of 20 candidates for hubs for both problems, the results on the time crtiterion are indicated by circles, for the capacity criterion by squares, and 10 candidates common for both problems are indicated by triangles

O'Kelly, M.E, (1987) A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* 32, 392–404.

O'Kelly, M.E., Bryan D. (2002) Interfacility interaction in models of hubs and spoke networks. *Journal of Regional Science*, **42** (1), 145–165.

Mażbic-Kulma, B., Potrzebowski, H., Stańczak, J., Sęp, K. (2008) Evolutionary approach to solve hub-and-spoke problem using $\alpha-$ cliques *Evolutionary Computation and Global Optimization, Prace naukowe PW*, Warszawa, 121–130.

Mażbic-Kulma, B., Potrzebowski, H., Stańczak, J., Sęp, K. (2009) Evolutionary approach to find kernel and shell structure of a connection graph *Total Logistic Management*, AGH, Cracow, 37–50.

Min, H., Gou, Z. (2004) The location of hub-seaports in the global supply chain network using a cooperative competition strategy. *Integrated Supply Management*, **1** (1), 51–63.

Michalewicz, Z. (1996) *Genetic Algorithms + Data Structures = Evolution Programs.* Springer Verlag, Berlin–Heidelberg.

Potrzebowski, H., Stańczak, J., Sęp, K. (2006) Evolutionary Algorithm to find graph covering subsets using α - cliques. *Evolutionary Computation and Global Optimization, Prace naukowe PW*, Warsaw, 351–358.

Potrzebowski H., Stańczak J., Sęp K. (2007) Separable decomposition of graph using α-cliques. In: M. Kurzyński, E. Puchała, M. Woźniak, A. Żołnierek. eds., *Computer recognition systems 2*, *Advances in soft computing.* Springer-Verlag, Berlin-Heidelberg, 386–393.

Stańczak, J. (2003) Biologically inspired methods for control of evolutionary algorithms. *Control and Cybernetics*, **32** (2), 411–433.

Stańczak, J., Potrzebowski, H., Sęp, K. (2011) Evolutionary approach to obtain graph covering by densely connected subgraphs. *Control and Cybernetics*, **41**, 3, 80–107.

Sutton, R.S., Barto, A.G. (1998) *Reinforcement Learning: An Introduction.* MIT Press.

Wilson, R.J. (1996) *Introduction to graph theory.* Addison Wesley Longman.