A Topological Interpretation for Mass Transit Network Connectivity

Chulmin Jun, Seungjae Lee, Hyeyoung Kim and Seungil Lee
The University of Seoul,
Dongdaemoon-Ku, Seoul, Korea
E-mail: {cmjun, sjlee, mhw3n, silee}@uos.ac.kr

ABSTRACT

The City of Seoul recently reformed the entire bus system as part of an intensified effort to relieve traffic congestion by encouraging mass transit means. Although the reformed system receives relatively positive reactions, it is still viewed to need improvements in route organization. Evaluation and organization of mass transit system becomes complex as the number of transportation modes increases. This paper presents an alternative method to assess the connectivity of mass transit system based on the topological structure of the routes network. It shows how to compute the connectivity of current transport routes configuration of multi-modal network and describe the results using the GIS. This study employs the hierarchical computation process of space syntax theory. The main methodological issue starts from the fact that the more transfers take place, the deeper the connectivity becomes making that area evaluated as less advantageous as for mass transit accessibility. By computing the connectivity of each bus or subway station with all others in a city, we can quantify the difference in the serviceability of city areas based on the mass transit. In computing the paths of origin-destination of routes, it employs the genetic algorithm. The process is illustrated using a network data of Seoul City built in a GIS.

Keywords: mass transit, transfer, space syntax, genetic algorithm, GIS

1. INTRODUCTION

Traffic congestion in the City of Seoul has lead to public-oriented transportation policy that encourages use of mass transit instead of privately-owned cars and recently the city government reformed the bus system on a large scale. However, there are some criticisms that the reformed system still partly shows over- or under- supply throughout the city area. Differences in accessibility to other areas from bus stops or subway stations cause differences in time, expenses and metal burden of users who travels the same distances. Current limitations in public transport planning call for robust methodology to assess the accessibility or serviceability of the transport routes.

Space syntax is the technique that has been used to derive the connectivity of urban or architectural spaces (Hillier 1996). The theory has primarily been applied in the research areas that seek to find the movement of human beings among indoor spaces or pedestrian paths and it has helped to compute the connectivity of the network of built environment quantitatively based on the topological structure of spatial links. The theory sees that spatial structure or layout has great impact on human social activities and displacement (Hillier 1984, Jiang 1999). Its primary principle is to model a spatial structure as a set of axial lines and compute spatial indices to derive the relationships between different parts of urban or indoor spaces. The resulting index is expressed as the integration of that space which is the degree to which that space is integrated and connected with other spaces in the defined area. Although transportation network problems are not included in typical applications of space syntax, we see the analogy between space syntax's spatial integration and mass transit links in that both are based on the hierarchical transitions between spaces.

This paper proposes an alternative method to evaluate connectivity of mass transport network based on its topological structure and the computational principle from space syntax. The primary idea of the paper is that we consider a transfer of vehicles as a connection node that links two different routes, and the more transfers take place, the deeper the connectivity becomes making that area evaluated as less advantageous as for public transport accessibility. We suggest an algorithm to show how geometric accessibility based on the connectivity of transport routes, rather than their physical distance, can be computed. Also, in order to calculate the optimal path from an origin to a destination, we used genetic algorithm. Transport network data including routes and stops were built in a GIS and relational database and the algorithm was programmed in C# language. The bus and subway network data set in Seoul are used to illustrate the proposed algorithm.

2. AXIAL LINE-BASED NETWORK CONFIGURATION

Human movement is frequently described in an abstracted form using its topology. Topological description allows researchers to focus on the structural relationship among units of movement while disregarding the details of phenomena. For example, pedestrian movement can be described using network of simple lines without considering the details such as sizes of forms, number of people and speed of movement. Such network configuration is also referred to as graph, which is a way to represent a network by a set of vertices and a set of edges that connects pairs of vertices. Figure 1 illustrates how meandering streets can be mapped to a graph. Following space syntax principle, spaces are first broadly perceived as discrete components, for instance, linear lines, and then are combined forming a continuous network. These lines are called 'axial lines' in space syntax.

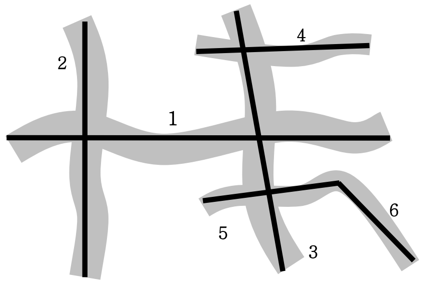

Figure 1. Axial lines of a street network

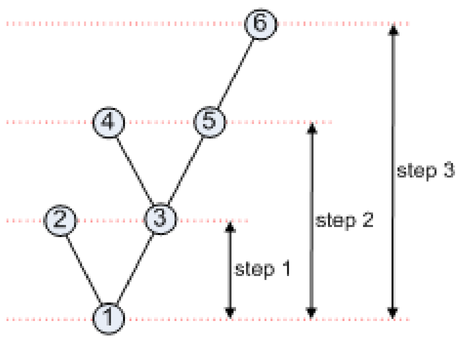

Figure 2. Hierarchical representation of street '1' from Figure 1

When spaces are mapped to a graph, the hierarchical relationship of component units is obviously captured. All this is best illustrated using Figure 1. Line 2 is accessible from line 1 by one turn, whereas line 4 is accessed by two turns. In other words, the relationship of 2 and 3 is called symmetrical with respect to 1 whereas the relationship of 4 to 1 is asymmetrical. In the literature of space syntax theory, this relationship is described thought a variable called depth (Bafna 2003). If one were to represent each component with a node and a turn with a link connecting their respective nodes, one could then describe the hierarchy from each node as shown in Figure 2. Figure 2 shows the hierarchy from node (or street) 1.

Depth of one node from another can be directly measured by counting the number of steps (or turns) between two nodes. The greater the depth of two nodes, the greater the hierarchical difference between them. The depth of a node (or a street) is defined by the number of nodes distant from a given number of steps to that node. If we take the example of Figure 2, the depth of node 1 for immediate neighbors (eg. step 1 nodes) is 2 since there are two nodes that can be accessed by one turn. On the other hand, the depth of node 1 in 2 steps distance is 4 since there are two nodes that can be reached by two turns, that is, 2 (nodes) × 2 (steps). Thus, the total depth from a node to all other nodes can be measured by summing the product of the level of step and the number of nodes in that step as given by:

$$TD_i = \sum_{s=1}^{m} s \times N_s \qquad (1)$$

, where

TD_i : the total depth of node *i*

s : the step from node *i*

m : the maximum number of steps extended from node *i*

N_s : the number of nodes at step *s*

The mean depth then is given by the total depth divided by k – 1, where k is the total number of nodes in the graph (Hillier 1996). This means the average depth of a particular node. Figure 3 shows extreme cases from node 1 in a network of same number of nodes. One case contains a node that extend to the maximum number of steps, which is k – 1 with the rest of nodes, one in each of intervening steps (case b in Figure 3), and the other contains only neighboring nodes to node a (case a in Figure 3).

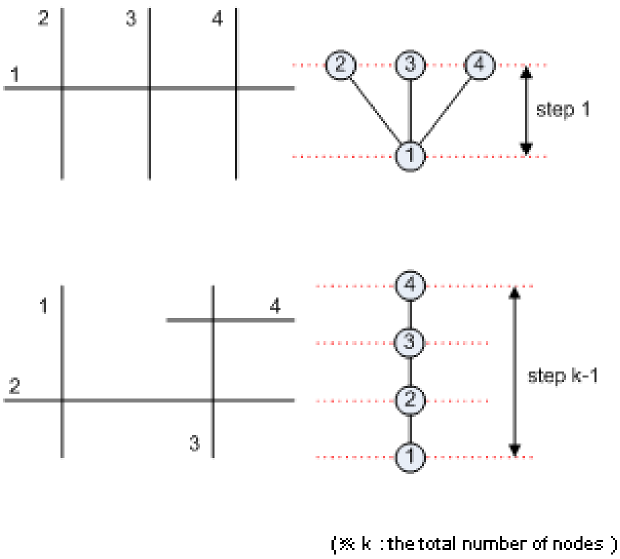

Figure 3. Symmetrical(a) and asymmetrical(b) layouts of streets

In the case (a), completely symmetrical structure, in Figure 3, MD is computed as follows,

$$MD = \frac{k-1}{k-1} = 1$$

whereas MD of the case (b) is computed by

$$MD = \frac{1+2+...+(k-1)}{k-1} = \frac{(k-1)k/2}{k-1} = \frac{k}{2}$$

Here, $1 \le MD \le \frac{k}{2}$ is derived. Therefore, MD is normalized as follows:

$$0 \le \frac{2(MD-1)}{k-2} \le 1 \qquad (2)$$

Now, using the normalized depth (ND), the depth from a node in a graph can be represented by a number ranging from 0 to 1. Using ND values makes it possible to compare depths of nodes from graphs with different number of nodes.

3. APPLYING TO MASS TRANSIT PROBLEM

The hierarchical description in space syntax summarized in the previous section primarily targets the abstraction of free movement of people mostly in built spaces rather than the movements in the mass transit which take place along fixed routes. However, we can derive the similarity of these two problems. Hierarchical framework of space syntax focuses on turns of spaces which are the basis for computing the depth of a certain space to others, while the mass transit problem generally entails transfers between vehicles. Thus, we can map turns of spaces to transfers between transportation means. In the hierarchical network description, the deeper the depth from a space to others, the more relatively difficult it is to move from that space to others. On the other hand, in mass transit, cost generally increases as the number of transfers between different modes increases. In this case, the cost can be either total fares or time taken in transfers, or it can even be seen as the mental burden that a traveler feels when he or she moves to or waits for the next vehicle in transfer areas. If we map the components using nodes and links in the previous section to a mass transit network, a node (or a street in Figure 1) can be seen as a stop, regardless of bus or subway, whereas a link between two nodes can be mapped to a transfer between two vehicles. This relationship is illustrated in Figure 4.

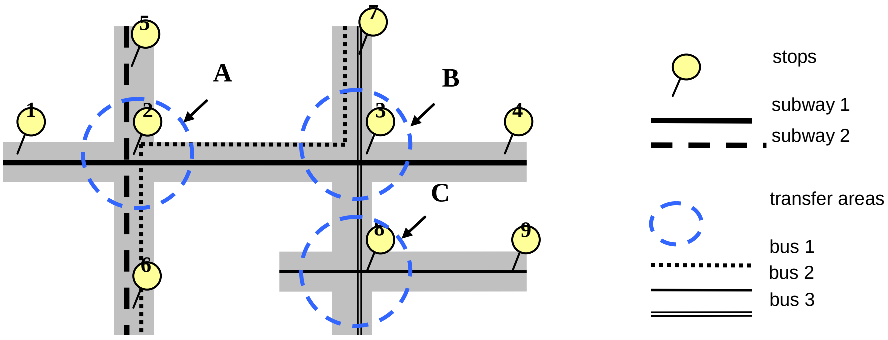

Figure 4. Mass transit network

If a person moves from stop 1 to stop 2, 3 or 4, he or she does not need to transfer because these stops are on the same route, subway line 1. However, if the traveler wants to go to stop 5 or stop 6 from stop 1, he or she has to transfer in the area A since stop 1 and stop 5 are not on the same route. Similarly, if the origin is stop 1 and the destination is stop 7 or 8, one transfer is needed in area B and, if the destination is stop 9, he or she has to transfer two times, one in area B and then in area C. One transfer from a transportation mode to another is the 'spatial transfer' which becomes one depth between spaces. If this network were that of pedestrian streets, when a person were to move from point 6 to point 7, he or she needs two turns, which means these points are two depths away from each other. However, in case of using pre-laid transport routes, the existence of a route that connects two points is first taken into account in computing the depths. Therefore, no transfer is needed in case of moving from stop 7 to 6 because there is a direct line connecting these two points.

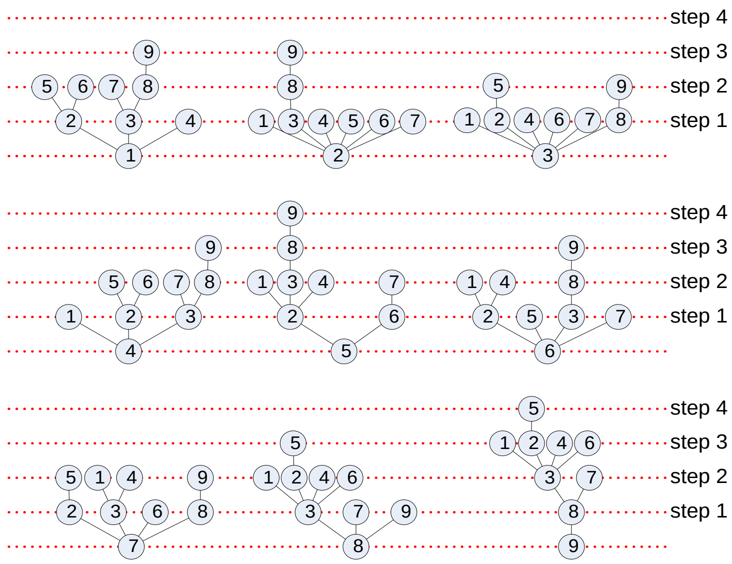

Figure 5. Hierarchical representation of network connectivity

The relationship of stops via routes is described in Figure 5. The connectivity from each stop to all others is hierarchically mapped to a graph. The procedure for generating a graph is iterative, starting with a stop and then progressively identifying the next neighboring stops until the entire stops are covered. The procedure first identifies the stops that are directly accessible from an origin using one route (step 1), then among these stops finds those stops that are belonged to transfer areas. Then, it looks for the routes that share these stops in the transfer areas. Next, it finds those stops that can be reached using the identified routes (step 2). It again looks for the stops belonged to the transfer areas. It continues iteratively in this manner.

Here, we can assume symmetrical and asymmetrical cases as we did in the previous section (Figure 6). The first case is when all stops are accessed from an origin via only one route. On the contrary, the other case is when all stops and routes are laid such that every stop is accessed by different means from the previous means. The former one is the completely symmetrical case, and the latter is completely asymmetrical one. Eqs. (1) and (2), which were defined for space connectivity, hold true for public transport network. We can then compute TD, MD and ND as shown in Table 1.

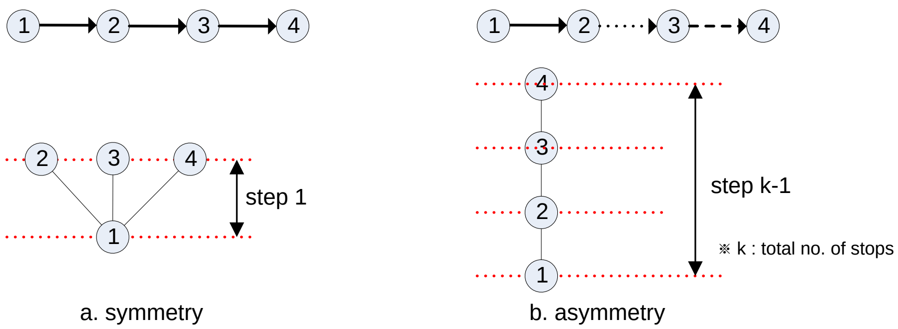

Figure 6. Symmetry and asymmetry of the route connectivity

Table 1. Depths from each stop in Figure 4

Stop No.	TD	MD	ND	ND^{-1}
1	14	1.750	0.214	4.67
2	11	1.375	0.107	9.33
3	10	1.250	0.071	14.00
4	14	1.750	0.214	4.67
5	17	2.125	0.321	3.11
6	13	1.625	0.179	5.60
7	12	1.500	0.143	7.00
8	14	1.750	0.214	4.67
9	21	2.625	0.464	2.15

Note that the reciprocal of ND is also calculated. These values help intuitive interpretation about the relationship between the graph and the accessibility. That is, higher values of nodes indicate that the stop is less deep on an average from all other stops and, in other words, shows better accessibility to other stops on an average. As one may easily expect, stop 3 shows the highest accessibility, 14.0, followed by 9.33 of stop 2. Note that stop 7 ranks in the third. It's because stop 7 has the routes that pass transfer areas, which makes it possible for a traveler to go to any other stops from stop 7 by only one transfer.

4. INTEGRATING INTO GIS

1) Building GIS Data

In order to apply the procedure proposed here to the mass transit network problem, we should consider utilizing GIS capabilities due to complexity and size of the network. However, current GIS data of the City of Seoul available to us as of now have limitations in reflecting the information necessary for the computation process suggested in this study. Mass transit networks of Seoul are composed of different modes such as metro bus, short-distance connection bus, airport limousine and subway and each of these modes is built in separate GIS data set. Also, a typical GIS data set is composed of geographical feature data and the table data, each record of which reflects its corresponding geographic feature. Thus, currently used GIS data structure alone can not capture the complex relationship of characteristics in mass transit.

To make the GIS network data of different modes usable, we first needed to organize them into a spatially integrated data set, where coordinate system and topology are matched against each other. Then, we added some attribute data tables that contain information about relationships between GIS network data features. The relationship among streets, routes, stops and transfer areas can be abstracted into an entity-relationship model in a relational database as shown in Figure 7. The procedure begins by defining a relation for each entity. One street section may contain more than one bus or subway route. Therefore the relationship between streets and routes is one to many (1:N). On the other hand, the relationship between routes and stops is many to many (M:N) because a route may include many stops and a stop may be shared by more than one route. We also need information about transfer areas for the computation process. Since one transfer area may contain more than one stop, the relationship of these two entities is one to many (1:N). The entities included in the E-R model are classified into two types; one is those that are built as the attribute tables of GIS data and the other is separately built tables. The former case includes Street, Route, and Stop entities and the latter Transfer Area. Also, the intersection entity generated from M:N relationship is also has to be created as a separate table. Since the data are constructed using such different formats, we cannot use macro languages provided as a subsidiary function of a proprietary GIS package. Thus, we used C# language to implement the proposed algorithm. We first created a test version using artificial networks similar to the one illustrated in Figure 4 and then extended it to the real networks of the Seoul City as illustrated in the following section.

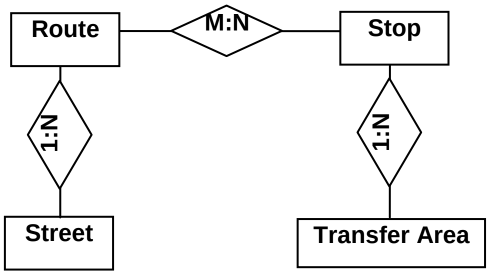

Figure 7. E-R diagram for mass transit network

2) Generating a Path using GA

Computing the depth of a stop or a station in a target area entails finding paths from that stop to all others in the area, each of which being the minimum-cost path. In this study, a minimum-cost path is defined as the one that has the minimum number of transfers between the origin and the destination.

Finding an optimal path in a mass transit network is much more complex than in the network of privately-owned cars. It is because the network is composed of multi-modes of vehicles and also has time-constraints at transfer areas. If a vehicle has a list of pre-specified departure times and transfers to other modes take place, the departure time is constrained for each available mode and its departure schedule, and comparison among these different departure times needs to be performed in order to explore the minimum time path (Desrochers *et al.* 1988).

Genetic algorithms aim at such complex problems. The study used the GA-based approach in finding the minimum-cost paths. Although we avoid detailed description about GA because there are many books or papers that introduce GA principles, we summarize the major part that shows how GA was modified to fit the mass transit problem.

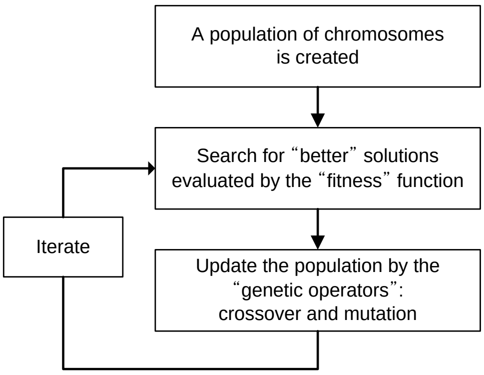

Figure 8. Generating a solution in GA

As shown in Figure 8, in GA, a global search process is performed on a certain population of chromosomes by gradually updating the population. Search processes are conditioned by two objectives: exploiting the best solutions and exploring the search space (Michalewicz 1994). The process for creating the first population is called the initialization. The updating processes of the population, the creation of successive generations, are done using so-called the genetic-operators: crossover and mutation. These genetic operators alter the composition of children of parent chromosomes. The search process is continued until it reaches the maximum number of generations while searching for "better" solutions that are evaluated by the "fitness" function. Therefore, the fitness function along with some parameters such as population size and probabilities of applying genetic operators are required in advance. Figure 9 shows a simple example of a network where different types of vehicles are present. In nodes such as 3 or 7, transfer does not happen. But the rest nodes allow the traveler to transfer to another mode.

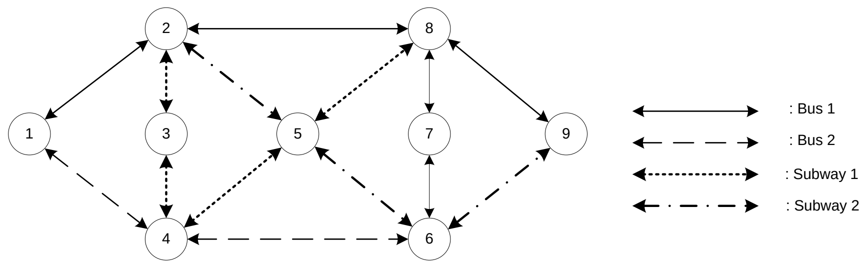

Figure 9. An example of a multi-modal network

Representation: In this study, a chromosome is represented by linking the stops from the source to the destination. If the source is Node 1 and the destination is the Node 9, a chromosome is an array of nodes that include Node 1 at the first position and Node 9 at the last.

Initialization: The initial population of chromosomes is created according to the preset population size. All nodes for each chromosome are initialized randomly as the following manner;

C_1 = (1, 2, 8, 9)

C_2 = (1, 4, 5, 6, 9)

C_3 = (1, 2, 5, 6, 7, 8, 9)

…

Evaluation: The evaluation function or the fitness function plays the role of the environment, rating potential solutions in terms of their fitness. Evaluation function *eval* for node vectors C can be set as the total time taken in the path. For this study that considers the connectivity of paths, we used the number of transfers taken from the origin to the destination instead of the total time as follows;

$$eval(C) = gene\_transfers(x),$$

Selection: Selection is a preparatory process that is needed for updating the current population. In order to preserve good chromosomes, some of them are reproduced in the next generation instead of participating in the mutation or crossover. This way, we can prevent those elite chromosomes from being deleted in the process. Selection process also includes the process that selects the parent chromosomes for crossover or mutation, which is described in the following section.

Genetic Operators: Some members in the initial population undergo alteration by means of two genetic operators: crossover and mutation. Crossover combines the features of two parent chromosomes to form two similar children by swapping corresponding segment of the parents. For example, if the parents are C_2 and C_3, then a common node (e.g. Node 5) can be selected and the portions of chromosomes after this node are crossed generating new children:

C_2 = (1, 4, **5**, 6, 9) → C_2’ = (1, 4, **5**, 6, 7, 8, 9)
C_3 = (1, 2, **5**, 6, 7, 8, 9) C_3’ = (1, 2, **5**, 6, 9)

Mutation arbitrarily alters the positions of one or more genes. In the transportation example, just exchanging a certain gene can generate a chromosome having disconnected link of nodes. Thus, we can modify the mutation process to fit this problem. If a certain gene is selected as the target of mutation, it can be thought of the temporary origin and then a portion of chromosome is created that reaches the destination. Assume C_2 has been selected and third gene, Node 5 has been selected as the mutation. Then, Node 5 becomes the temporary origin yielding a chromosome from this node to Node 9. After the mutation, new C_2 can be created as

C_2 = (1, 4, **5**, 2, 8, 7, 6, 9) → C_2’= (1, 4, **5**, 2, 8, 7, 6, 9)

As seen from this, mutation can either increase or decrease the value of selected chromosomes.

3) Case study

In order to show the feasibility of the method proposed here, we chose a CBD area called 'Kangnam District' of Seoul that shows high complexity in the network structure. Since the area contains a variety of modes such as subway and different types of bus lines, it was viewed suitable for the study.

The bus stops that are located near each other are grouped into a transfer area as shown in Figure 10, and the Integration value(ND^{-1}) was computed for each of these transfer areas.

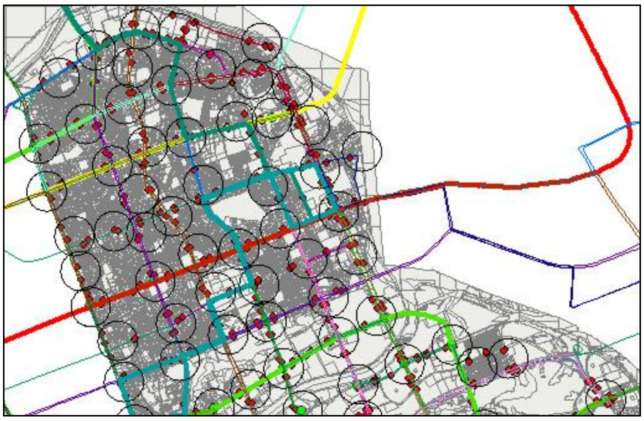

Figure 10. Transfer areas in the test area

Figure 11 shows the computed Integration(ND^{-1}) for the study area. The stops grouped in a transfer have the same I-values. For the purpose of visual recognition, we divided the resulting I-values into 5 classes in the same color scheme. The darker the color, the higher the I-value becomes meaning the stop is located in more connected area with other locations.

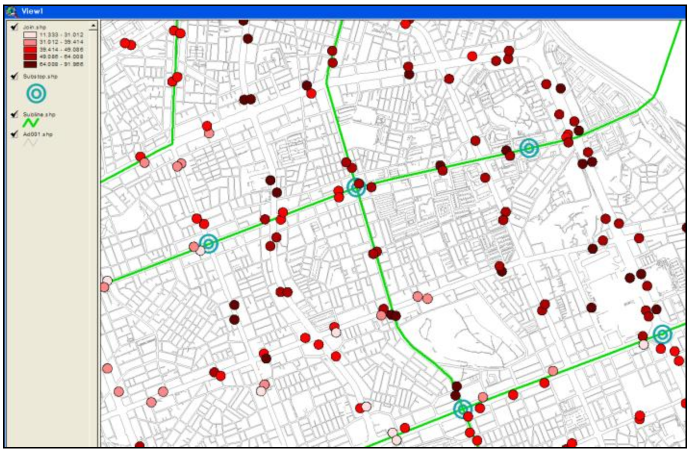

Figure 11. Integration(ND^{-1}) values for stops in the test area

5. CONCLUDING REMARKS

This paper presented an alternative method to assess the connectivity of mass transit network by defining the network relationship onto a graph of hierarchical structure. We derived an analogy between the concept of depths in space syntax and the degree of connectivity of network of transport routes. We interpreted the spatial transitions between spaces in space syntax as the transfers between transportation modes. We developed an algorithm to automate the computing process using GIS data. In order to generate optimal paths during the process, we employed the Genetic Algorithm. We constructed a relational database model to capture the relationship between GIS data features and added attribute tables. By applying the proposed method to a CBD area, we could quantify the differences in the connectivity of city areas using the space syntax values.

We plan to expand this algorithm to more real situations such that it can incorporate origin-destination movements and real distances in the paths. In that case, the term 'the topological connectivity' used here can be modified to more general term such as 'the accessibility'.

ACKNOWLEDGEMENT

This study was supported by the 2005 University Research Fund of the University of Seoul.

REFERENCES

[1] Bafna, S., 2003. Space syntax: a brief introduction to its logic and analytical techniques, *Environment and Behavior*, 35(1): 17-29.
[2] Desrochers, M. and F. Soumis, 1988. A Reoptimization Algorithm for the Shortest Path Problem with Time Windows, *European Journal of Operational Research*, ***35*:** 242-254.
[3] Hillier, B., 1984. *The Social Logic of Space*, Cambridge University Press.
[4] Hillier, B., 1996. *Space is the Machine*, Cambridge University Press.
[5] Jiang, B., C. Claramunt and M. Batty, 1999. Geometric accessibility and geographic information: extending desktop GIS to space syntax, *Computers, Environment and Urban Systems*, 23: 127-146.
[6] Michalewicz, Z., 1994. *Genetic Algorithm + Data Structure = Evolution Programs*, Springer-Verlag.