Network Analysis of World Subway Systems Using Updated Graph Theory

Sybil Derrible and Christopher Kennedy

This paper demonstrates that network topologies play a key role in attracting people to use public transit; ridership is not solely determined by cultural characteristics (North American versus European versus Asian) or city design (transit oriented versus automobile oriented). The analysis considers 19 subway systems worldwide: those in Toronto, Ontario, Canada; Montreal, Quebec, Canada; Chicago, Illinois; New York City; Washington, D.C.; San Francisco, California; Mexico City, Mexico; London; Paris; Lyon, France; Madrid, Spain; Berlin; Athens, Greece; Stockholm, Sweden; Moscow; Tokyo; Osaka, Japan; Seoul, South Korea; and Singapore. The relationship between ridership and network design was studied by using updated graph theory concepts. Ridership was computed as the annual number of boardings per capita. Network design was measured according to three major indicators. The first is a measure of transit coverage and is based on the total number of stations and land area. The second relates to the maximum number of transfers necessary to go from one station to another and is called directness. The third attempts to get an overall view of transfer possibilities to travel in the network to appreciate a sense of mobility; it is termed connectivity. Multiple-regression analysis showed a strong relationship between these three indicators and ridership, achieving a goodness of fit (adjusted R^2 value) of .725. The importance of network design is significant and should be considered in future public transportation projects.

Subway systems contribute greatly to metropolitan regions around the world. They are part of the identities of cities and are an essential economic component (*1*). Their extent and use are likely to increase when the present challenges that the world is facing in terms of sustainable development (the growth of urban areas) and global warming (lower greenhouse gas emissions from the use of mass transit and electricity-powered technologies) are considered. The analysis of subway systems is often associated solely with demand characteristics, that is, geographic features (density) and cultural features (Europe versus Asia versus North America). The supply side is habitually considered only briefly (e.g., by comparing the number of lines of transit systems), leaving only a general appreciation of the complexities of networks.

This research focuses instead on the topologies of subway systems, bringing attention to the network design as a whole. This is done specifically by incorporating some updated graph theory concepts by the use of such characteristics as the numbers of lines, stations, and transfers and other characteristics. The efficiency of a subway system is then compared with ridership data to reflect its net attractiveness rather than only network characteristics. In this paper, subway systems include all urban rail rapid transit systems that have an exclusive right-of-way, whether they run underground or not (e.g., the El in Chicago, Illinois, does not); however, data for lighter mass transit technologies, such as light rail and streetcars, are not included. Data were collected for the subway systems of 19 cities on three continents around the world: those in Toronto, Ontario, Canada; Montreal, Quebec, Canada; Chicago; New York City; Washington, D.C.; San Francisco, California; Mexico City, Mexico; London; Paris; Lyon, France; Madrid, Spain; Berlin; Athens, Greece; Stockholm, Sweden; Moscow; Tokyo; Osaka, Japan; Seoul, South Korea; and Singapore.

Graph theory is inherently linked to transportation. Euler first introduced graph theory in the 18th century after he looked at the Seven Bridges of Konigsberg, Russia. In the 1960s and 1970s, it was extensively applied to road networks (*2*). Kansky explains the main graph theory concepts and enlists various indicators, such as the cyclomatic number, the network diameter, and other dimensionless ratios (α, β, γ, etc.) (*3*). He also made an attempt to link the topology of road networks with economic development. Bon looked at the effect of network growth on the actual and potential number of edges in a system using empirical data (*4*). Nevertheless, the application of graph theory seemed to lose its stamina in the mid-1970s, being replaced by further, more adequate travel demand models.

The literature on the application of graph theory to public transit networks is sparse. Taaffe (*5*) recalls the indicators presented by Kansky (*3*). Lam and Schuler also enlisted them and introduced a new one related to travel time (*6, 7*). Gattuso and Miriello used them only for network analysis purposes and therefore did not include ridership data (*8*).

Musso and Vuchic established a comprehensive series of new indicators, the line overlapping index, the circle availability, and network complexity, which more adequately grasp the public transportation networks issue (*9*). However, they also did not incorporate ridership data. Vuchic later recalled these indicators in his book *Urban Transit: Operations, Planning, and Economics* (*10*). On the other hand, Levinson produced interesting results by incorporating ridership levels but did not link them to graph theory (*11*). Instead, he contemplates the future of rail transit and considers network data (e.g., route length) at the country level.

The primary goal of the present research was to investigate the relationship between network design and ridership by using transit-appropriate indicators. In the process, extensive data were collected

---

Department of Civil Engineering, University of Toronto, 35 St. George Street, Toronto, Ontario M5S 1A4, Canada. Corresponding author: S. Derrible, sybil.derrible@utoronto.ca.

Transportation Research Record: Journal of the Transportation Research Board, No. 2112, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 17–25.
DOI: 10.3141/2112-03

at the city and the network levels. The work presented here follows three steps:

1. Collect new and adequate data on subway network characteristics;
2. Develop new measures of urban transit networks, notably, by using graph theory; and
3. Establish an empirical relationship between ridership and network characteristics.

This research is part of a broader study of the effects that public transit networks have on cities by looking closely at ridership and economics. Earlier work was already established at the microscale (*12*) by looking at the local effects of adding one individual transit line in Toronto. The present work relates to the macroscale (cities around the world) and is planned to be extended to grasp the economic effects. Future work will then consider the mesoscale by radically changing the transit network of a single city.

DATA

Extensive data were collected throughout the study. Information on the designs of the 19 networks was first gathered. Accurate data on ridership were then collected for each network, as was information on city characteristics (population, density, and area).

Network Design

Network design is a substantial part of this research. The book *Transit Maps of the World,* by Ovenden (*13*), was useful for identifying major subway systems as well as getting the number of stations (n_S), the number of lines (n_L), and the route length (L; in kilometers) for each network. The respective transit authority websites for each system were also examined to validate the data. These websites typically show key figures of networks, and in most cases the values found in the book were consistent with the data from the websites. Moreover, the maps included in Ovenden's book (*13*) were used to create data for this study.

The maps of the transit systems of all cities studied are redrawn as graphs by consideration of only the subway mode. Figure 1 offers a typical example of a network along with the graph theory components used in this research. The values in Figure 1 are related to an example network; they are not generic. The network has four transit lines.

The nodes in the graph are called vertices. Not all stations are vertices; only those stations that host at least two lines or those that are line terminals are vertices; for instance, the network in Figure 1 could have any number of stations. Two types of vertices are distinguished: transfer vertices (transfer stations) and end vertices (line terminals). The example network has nine vertices (six end vertices and three transfer vertices); one end vertex that is also a transfer vertex is considered to be only a transfer vertex.

$$v = v^t + v^e \tag{1}$$

where

v = total number of vertices,
v^t = total number of transfer vertices, and
v^e = total number of end vertices and where

$$v = \sum_i v_i \tag{2}$$

$$v^t = \sum_{i,l\neq 1} v_{i,l} \tag{3}$$

$$v^e = \sum_{i,l=1} v_{i,l} \tag{4}$$

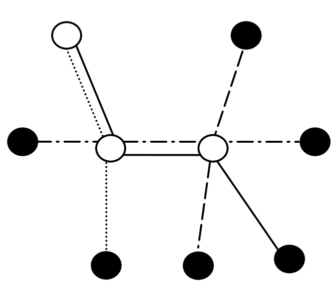
In Equations 2, 3 and 4, the subscripts i label the vertices and the subscripts l attribute the number of lines crossing the vertex. These are not shown in Figure 1; but each vertex could be labeled from 1 to 9 as $v_1, v_2, v_3, \ldots, v_9$, respectively.

Links in the network, which are called edges in graph theory, connect terminal vertices to transfer vertices or transfer vertices to other transfer vertices. These were divided into two groups: single use and multiple use. If two consecutive transfer vertices are connected by two parallel edges, then one single-use edge and one multiple-use edge are considered. It is therefore possible to avoid the travel path redundancies generated by overlapping lines, whether the lines share the rail track or not. The example network has 10 edges (eight single-use edges and two multiple-use edges).

$$e = e^s + e^m \tag{5}$$

where

e = total number of edges,
e^s = total number of single-use edges, and
e^m = total number of multiple-use edges and where



Name	Symbol	Value	Image
Route length	L	NA	
Stations	n_S	NA	
Lines	n_L	4	
Vertices	v	9	
end	v^e	6	
transfer	v^t	3	
Edges	e	10	
single	e^s	8	
multiple	e^m	2	
Max number of transfers	δ	2	

FIGURE 1 Graph theory components and example of network.

$$e = \sum_{j} e_j \tag{6}$$

$$e^s = \sum_{j,k=1} e_{j,k} \tag{7}$$

$$e^m = \sum_{j,k\neq 1} e_{j,k} \tag{8}$$

In Equations 6, 7, and 8, each edge is labeled by a subscript j, similar to the method used for the vertices (Figure 1 would have edges one to 10), and by a subscript k that represents the multiple-use characteristic ($k = 1$ normally, $k = 2$ if it is the second edge to link two vertices, etc.).

Finally, the maximum number of transfers (δ) needed to go from any vertex to another vertex but taking the shortest path was also measured. It is analogous to the concept of the diameter of a graph but focuses instead on the least number of transfers. This number is used to appreciate the transfer properties of a system and is a significant value because transfers are often seen as being negative. In this instance, data were collected from the actual transit maps and not from the graphs. The values remained relatively low, from one transfer on the San Francisco Bay Area Rapid Transit (BART) in California to three transfers on more complex networks (e.g., the Paris Metro). The maximum number of transfers of the example network is two, which would be required to go from the station at the bottom left to the station at the upper right.

It should also be noted that in the cases of Paris and Berlin, appropriate portions of the Réseau Express Régional (RER) and S-Bahn (Stadtschnellbahn, or fast city train) networks, respectively, were included in the analysis. Indeed, both cities have regional rail services that act as part of the subway system in the city core (and for which the fare is the same as that for the subway and on which intermodal transfers are easy). It was therefore believed to be preferable to include segments of these two extra systems.

Ridership

As mentioned in the introduction, the various components of network designs are analyzed relative to ridership levels. Nevertheless, a few choices were available in terms of the ridership indicator, and it is worth going through them here.

Typically, transit systems are assessed in comparison with other transport modes by using mode split figures (the numbers of individuals driving an automobile, taking transit, walking, or cycling). However, such figures typically emphasize the home–work trips, for which regional rail is often a determinant but is not part of the analysis. Moreover, users of large transit networks are expected to use public transportation not only to go to work but also for different purposes. Therefore, a better indicator may be the numbers of transit trips. Again, only the subway mode is considered. Moreover, these indicators are relatively hard to acquire and validate on an international scale. In addition, the main sources that include this information are the Millennium Cities Database for Sustainable Transport (*14*) and the Mobility Cities Database (*15*). Although it is comprehensive, the Millennium Cities Database contains data from 1995 (which is not appropriate for recent networks, e.g., the network in Madrid), and the Mobility Cities Database does not contain information on all the network systems evaluated in the present study. However, they have been useful for comparison and validation of the data collected.

Consequently, it became logical to use the number of boardings (subway only) as the principal indicator. The average daily ridership of subway systems is most often cited on transit authority websites; however, subway use on the weekend is relevant, again so that the home–work trip is not overemphasized and to account for subway trips for shopping or entertainment. As a result, the use of annual figures was preferred. Although they were not always readily available, the authors managed to collect reasonable values by looking up transit financial reports and data from statistical agencies or by communicating directly with transit authorities. All the data on annual boardings collected date between 2005 and 2007. They are therefore reasonably recent. These values were then validated by using the two sources mentioned above. However, it is not possible to tell whether the boardings are linked or unlinked trips, as that information was not specified.

Finally, to standardize the data, the annual number of boardings was further divided by the population. This action was critical because larger cities typically have higher numbers of boardings. By using the population, it is therefore possible to appreciate the net attractiveness of a network design as opposed to only its ridership. The values chosen for populations are discussed in the next section.

City Characteristics

The three city characteristics that were considered were population, density, and area. Population was collected for the reason mentioned above. Density was then collected as a means of acquiring the area served by public transit.

These types of data are difficult to collect or measure (*16*). Historically, most cities had rather small area boundaries that have not been updated with urban growth at the administrative level (e.g., Lyon). Naturally, these values cannot be considered, given that they do not reflect the real effects of public transportation networks. However, the next level offered is the urban area. The concept of urban area is not similar in every country, and it usually encompasses an area much larger than that served by the transit networks. Moreover, some urban areas actually have separate transit systems (e.g., Tokyo–Yokohama). This information was even more important because the study was dealing with international cities; hence, there was a need to acquire balanced standardized values. For instance, London has 7.5 million inhabitants and an urban area population of 8.3 million, whereas Osaka has 2.6 million inhabitants and an urban area population of 15.5 million. It is clear that the use of the city population or the urban area population individually is not sensible. Figure 2 shows an example of a subway network in a city within its urban area. The area shaded light gray is the area served and is the area desired for use in the analysis.

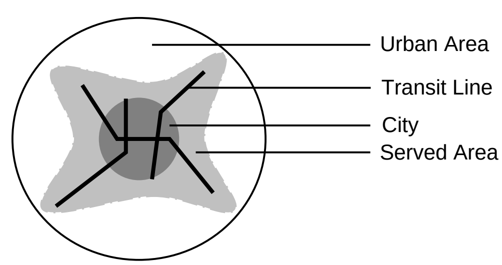


FIGURE 2 Typical city, served area, and urban area with subway network.

Of the 19 cities, only Stockholm, Moscow, and Tokyo had readily available information on the areas served. Data for Stockholm and Tokyo were available from their transit agencies. For Moscow, city official data that encompass true population levels were used. Paris was tackled differently because of the unique geographic feature of the population. Data were collected for the City of Paris, its three surrounding administrative regional municipalities, plus two further regional municipalities that had most of their inhabitants located in the same overall region.

For the remaining networks, it was found that the mean of the city population and the urban area population gave the most practical measure of the effective population served. Not only does this approach seem realistic in terms of the served area by transit, but also it offsets other uncertainties concomitant with the fact that the study was dealing with international cities (different administrative systems). Data on population and density were collected mostly from www.citypopulation.de (*17*) at the city level and from Demographia (*18*) at the urban area level.

Similar to the problem with population and density, administrative figures of area often incorporate data for uninhabited lands, whereas the area served is required. To go around the problem, it was decided that area would be calculated by dividing the population by the density. This is a rather pragmatic approach to handling the issue of defining the area served by subway networks.

NETWORK INDICATORS

At first, the initial indicators presented by Kansky (*3*) were applied to the data. However, it rapidly became clear that they were not well adapted to explaining boardings on public transit networks. Consequently, the indicators were reevaluated and new concepts were put forth. Finally, three ratios seemed to meet the study objectives: coverage, directness, and connectivity.

Coverage

The concept of coverage is relatively simple. The purpose is to measure the coverage area accessible by subway networks. Typically, this is done by considering the sum of coverage areas of each station of the network divided by the area served (A). By taking a threshold value of a 500-m radius for a coverage area, coverage (σ) is

$$\sigma = \frac{n_S \cdot \pi \cdot 0.5^2}{A} \tag{9}$$

where n_S is the number of stations.

One comment about the radius used should be made at this point. The value of 500 m was chosen arbitrarily and could be anywhere from 400 to 800 m or even 1 km. In fact, however, it is of low relevance to the analysis. As the concern here is more with the relationship of each network with one another, the constant $\pi \cdot 0.5^2$ actually gets absorbed in the slope of the equation during the regression analysis. Nevertheless, it was left in the original equation so that each system could be looked at individually.

Directness

The concept of directness is related to the concept of the indicator π, as introduced by Kansky (*3*). He explains π as being "a number expressing the relationship between the circumference of a circle and its diameter. . . . We can apply this notion to transportation networks. . . . [L]et us assume that the total mileage of a transportation system is analogous to the circumference of a circle and the total mileage of all edges of the diameter of a network is analogous to the diameter of the circle." In other words, π does not deal with line type (circle, radial, spinal, etc.), but instead, it measures the ratio of the total track length to the length of the longest route.

To apply the concept to public transportation, the numerator can be the total route length (L). As the denominator is related to the diameter of the system, from a transit network point of view, it is concomitant with the maximum number of transfers needed to go from any vertex to another, introduced above as δ. In fact, this is one measure of the longest route possible in the network and can be associated with a diameter. As a matter of consistency, δ should be multiplied by some sort of length to keep the ratio dimensionless. One suggested solution is to multiply it by a segment (κ) of the average line length. If n_L is the total number of lines of a network, L/n_L is the average line length; hence, $L/(\kappa \cdot n_L)$ is a segment of a line; directness (τ) is

$$\tau = \frac{L}{\frac{L \cdot \delta}{\kappa \cdot n_L}} = \frac{\kappa \cdot n_L}{\delta} \tag{10}$$

Therefore, a switch from an overall network-size feature to a maximum transfer appreciation has been made. The value of κ is again irrelevant, as it is absorbed by the slope during a regression analysis. However, one interesting case happens when κ equals δ. In this particular instance, the directness ratio equals the total number of lines. Although a relationship between the number of lines and the number of boardings per capita exists, it is not as strong. Moreover, it undermines the purpose of measuring the transfer properties of a given subway network. This study simply kept the directness indicator (τ) to be equal to

$$\tau = \frac{n_L}{\delta} \tag{11}$$

Connectivity

Connectivity in this case relates to the ability to travel freely within the network. It could also be associated with the degree of mobility or the density of transfer possibilities. In other words, it shows the degree by which a network is connected, allowing more travel path choices. It may be determined by summing all the transfer opportunities, v_c^t, which is the sum of the number of subway lines going through a transfer station minus 1 and where c indicates connectivity (that is, the number of transfer vertices v^t becomes an indicator of transfer possibilities, v_c^t). It differs from the traditional definition of degree of a node; the total number of edges at a vertex was not counted; instead, the number of possibilities for the switching of transit lines was simply counted. For instance, a transfer station sharing two transit lines offers one transfer possibility, another sharing three lines offers two possibilities, and so on.

$$v_c^t = \sum_{i,l \neq 1} (l - 1) \cdot v_{i,l} \tag{12}$$

To account for the net sum of transfer possibilities (avoid redundancies), the number of multiple-use edges (e^m) needs to be subtracted from v_c^t. This accounts for the net number of transfer possibilities but is not yet an index. It still needs to satisfy the international context criteria, which is the wide range of network topologies that exist. Consequently, to standardize connectivity (ρ), the equation needs to be further divided by an appropriate variable, which in this case is the total number of transfers.

$$\rho = \frac{v_c^t - e^m}{v^t} = \frac{\sum_{i,l \neq 1} (l-1) \cdot v_{i,l} - \sum_{j,k \neq 1} e_{j,k}}{\sum_{i,l \neq 1} v_{i,l}} \tag{13}$$

The denominator allows the ratio to become an index. A dispersed network can have a high number of transfer stations and still have a relatively low connectivity feature. For instance, the London subway system has a higher net sum of transfer possibilities than the Madrid system; nevertheless, Madrid's network is more highly connected and has a higher number of boardings per capita.

RESULTS

This section illustrates the results for the three indicators (Table 1). The values of coverage were always much less than 1; they ranged from 0.0269 in Athens to 0.2538 in Seoul (i.e., 25% of the served area calculated is located within 500 m of a station). On the other hand, directness was always greater than 1 and had values that ranged from 1.33 to 6.5. For connectivity, values less than 1 can be generated, for example, 0.6 for San Francisco and 1.6212 for Paris, because of the presence of multiple-use edges.

Each indicator is first analyzed separately and is then grouped in a multiple linear regression. The *t*-test method was used to prove the validity of the parameters calculated. Because the subway systems in 19 cities were evaluated, the *t*-test (with right-tail probabilities) value should be at least 1.328 for 90% confidence and 1.729 for 95% confidence.

Coverage

The number of boardings per capita versus coverage showed a clear relationship. Higher coverage tends to render higher ridership per capita. This is expected, as greater accessibility attracts more people either for the origin or for the destination ends of trips. One well-known trade-off exists between speed and coverage; that trade-off stipulates that higher coverage induces lower speed and, hence, longer travel times. That was not studied here, however. First, speed cannot be compared alone because it is also considerably dependent on technology; therefore, a small network could still have fairly high speeds but not carry many passengers. Second, it was assumed that there is no universal station spacing for the variable to be included in the analysis effectively. Third, at this level, it was believed that the measurement of coverage (accessibility) was more sensible than, for instance, measurement of the travel time to the city center. It could, nevertheless, be included in further studies, notably, by considering the indicator introduced by Lam and Schuler (*6, 7*) that relates time and connectivity.

The existing relationship (not shown), however, did not appear to be linear. It seemed that an increase in coverage did not necessarily imply a linear increase in ridership. Replotting of the natural logarithm of coverage showed a much better fit (Figure 3). In fact, this means that as systems increase in size, the marginal boardings per capita decrease. In other words, an already large network will not generate as many additional boardings per capita as a smaller network by building new stations. This is quite interesting. One may think that by adding new stations to a large network more people are prone to use it, as the choice of destinations is already extensive, whereas, in fact, the opposite is true. This seems to follow the theory of diminishing marginal returns.

All values calculated for coverage were therefore negative because the logarithmic function was being used and all coverage values were less than 1.

TABLE 1 Boardings and Three Indicators for the 19 Subway Networks

Subway Network	Pop. (million)	Boardings (million per year)	Boardings per Capita	Coverage (σ)	Directness (τ)	Connectivity (ρ)
Toronto	3.435	265.3	77.23	0.0522	1.3333	0.8000
Montreal	2.418	278.2	115.05	0.0663	2.0000	1.2500
Chicago	5.571	186.8	33.53	0.0674	2.3333	0.7692
New York City	13.007	1,804	138.69	0.1594	3.6667	1.0426
Washington, D.C.	2.258	259.4	114.88	0.0743	2.5000	0.8750
San Francisco	1.987	99.3	49.97	0.0724	4.0000	0.6000
Mexico City	12.932	1,417	109.57	0.0446	3.6667	1.3913
London	7.898	1,078	136.49	0.1661	6.5000	1.0000
Paris	8.403	1,860.9	221.46	0.2354	6.3333	1.6212
Lyon	0.908	96.5	106.28	0.1885	2.0000	1.0000
Berlin	3.642	475.0	130.42	0.1444	4.0000	1.1228
Madrid	4.016	690	171.81	0.1937	4.3333	1.2500
Athens	3.379	92	27.23	0.0269	3.0000	1.0000
Stockholm	1.9	297	156.32	0.0748	3.0000	0.8889
Moscow	10.521	2,475.6	235.30	0.2079	6.0000	1.2500
Tokyo	15.07	2,974	197.35	0.1036	6.5000	1.2889
Osaka	9.043	912	100.85	0.0924	2.6667	1.0833
Seoul	13.71	2,264	165.13	0.2538	3.6667	1.0000
Singapore	4.277	434.9	101.68	0.0956	1.3333	0.8333

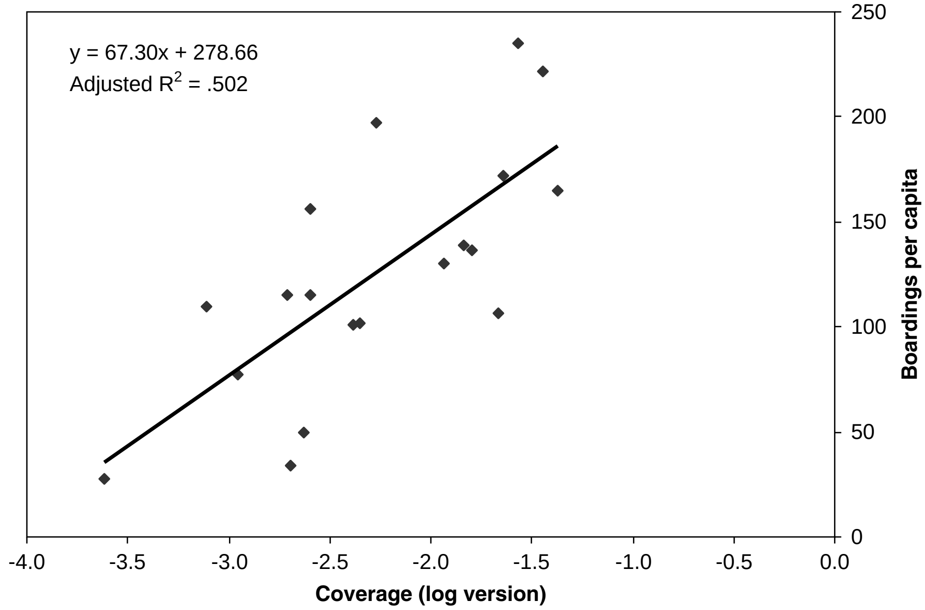


FIGURE 3 Boardings per capita versus natural log of coverage.

A linear regression analysis (least-squares method) resulted in a slope value of 67.30 and an intercept of 278.66, with the *t*-test statistics being 4.37 and 7.71, respectively, and the goodness of fit (adjusted R^2 value) being .502. The model parameters were therefore statistically significant, as they met the 95% criteria.

Directness

Figure 4 illustrates the number of boardings per capita as a function of directness. Figure 4 shows a comparatively good relationship between directness and the number of boardings as well. Moreover, note that in this case the relationship is linear, unlike the relationship for coverage.

Three networks (the bottom three datum points in Figure 4) did not seem to follow the general pattern. These networks were, from left to right, Chicago, Athens, and San Francisco, respectively. All three have the particularity of having a low number of boardings per capita. The Chicago subway system has a low value of δ because of the presence of the Loop (almost all the lines interconnect downtown). The Athens Metro, although its lines are well connected, is fairly small in comparison with the size of the population (three lines for almost 3.4 million inhabitants). The San Francisco BART system is often associated with a regional rail system (average station spacing of

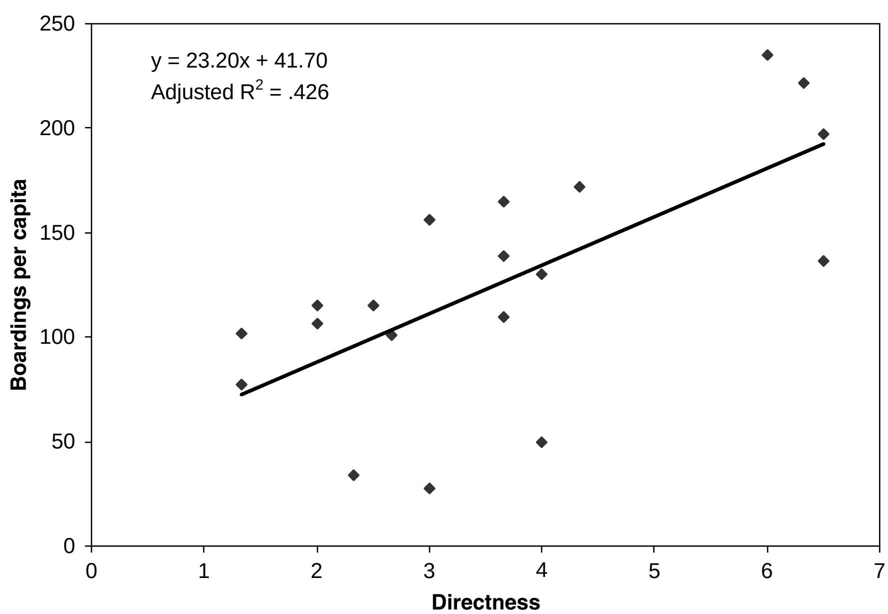


FIGURE 4 Boardings per capita versus directness.

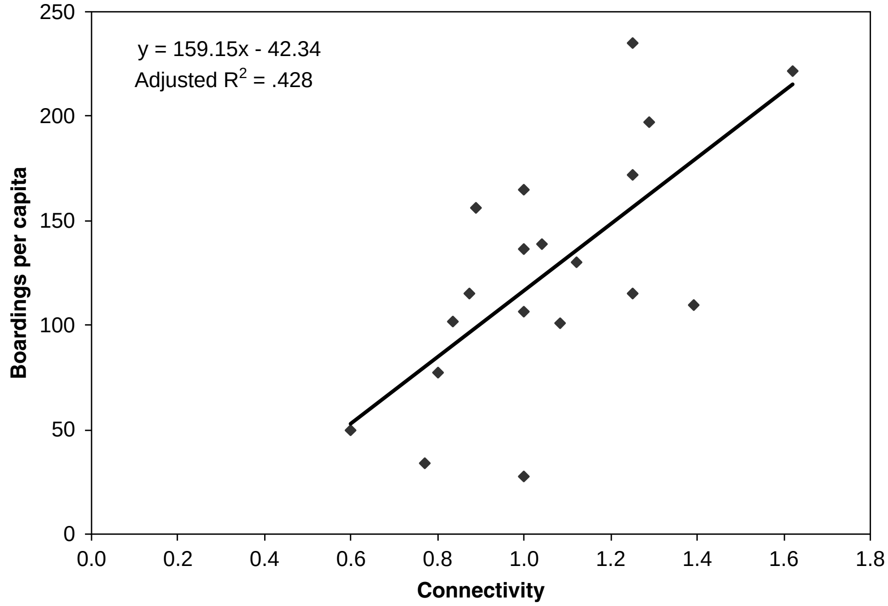


FIGURE 5 Boardings per capita versus connectivity.

more than 3 km); its design is very linear and contains many multiple-use edges. The last two networks seem to lack accessibility, which translates into low ridership. Nevertheless, the lack of accessibility is already accounted for in the coverage indicator.

When a linear regression analysis was carried out, the slope had a value of 23.20 and the intercept was 41.70, with *t*-test statistics of 3.79 and 1.71, respectively. The goodness of fit (adjusted R^2 value) was calculated to be .426. The model parameters were therefore statistically significant as well. By omitting the data for Chicago, Athens, and San Francisco, the slope becomes 20.66 and the intercept becomes 65.58, with the goodness of fit (adjusted R^2 value) becoming .637. This further demonstrates the strong relationship between directness and the number of boardings per capita and shows the reasonable consistency of the slope.

Connectivity

Figure 5 shows the number of boardings per capita versus connectivity. It is possible to observe that despite a steeper slope, a relationship is present. The lowest point away from the regression line is for the Athens Metro, for the reasons mentioned above. Again, the relationship is linear, unlike that for coverage. This means that the addition of connectivity has the same results for small or large networks. The regression analysis resulted in a value for the slope of 159.15 and a value of –42.34 for the intercept, with *t*-test statistics of 3.81 and –0.93, respectively. The goodness of fit (adjusted R^2 value) was calculated to be .428. Although the *t*-test statistic of the intercept is low, the overall relationship is still clear. Moreover, this value is not relevant, insofar as more emphasis is put on the multiple regression analysis, which is described in the next section.

Multiple Regression Analysis

Up to this point, the three indicators were considered separately. None of them seemed to be more influential than the others. Indeed, they all showed similar goodness-of-fit values. The next step was to perform a multiple linear regression analysis. The results are shown in Table 2. The equation relating the network characteristics to boardings per capita (Bpc) is

$$\text{Bpc} = 44.963 \cdot \sigma + 7.579 \cdot \tau + 92.316 \cdot \rho + 102.947 \tag{14}$$

The intercept of 102.947 is not the lowest number of boardings per capita because the log version of the coverage is negative. Moreover, by considering the values for the parameters in Equation 14 with their respective variables, it seems that the three indicators share relatively equal weights.

The goodness of fit (adjusted R^2 value) of .725 is reasonably high. The R^2 value was .771. Three *t*-test calculations reached the 95% confidence level; the remaining value (directness) fell in the 90% confidence range. The four parameters are therefore significant.

A separate analysis further showed little correlation between the three variables. The strongest correlation occurred with directness and connectivity, for which the adjusted R^2 value was .26, but this is understandable. When the number of transfer stations is increased, directness should logically increase, as there are more paths to go from one vertex to another one. Figure 6 shows the predicted number of boardings per capita versus the actual number, and a 45° line shows the desired results. All the points are scattered evenly and are close to the line.

TABLE 2 Linear Multiple Regression Statistics

	Coefficients	Standard Error	*t*-Stat.
Intercept	102.947	51.787	1.988
Coverage	44.963	13.196	3.407
Directness	7.579	5.508	1.376
Connectivity	92.316	34.813	2.652

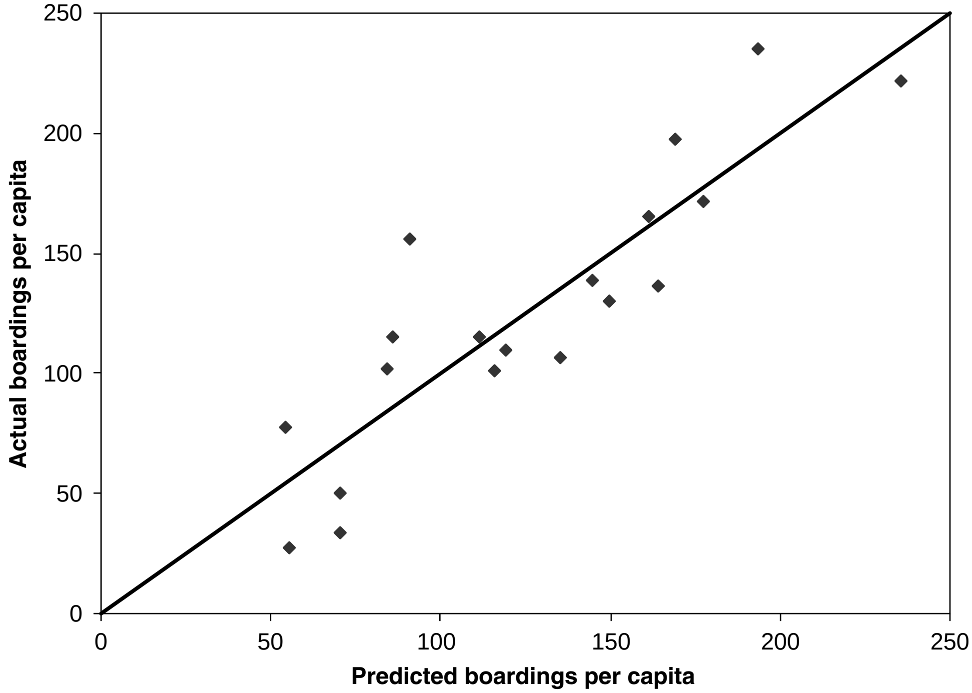


FIGURE 6 Predicted versus actual number of boardings per capita.

DISCUSSION OF RESULTS

The statistical analysis showed a high degree of correlation between ridership per capita and coverage, directness, and connectivity. Other characteristics, such as line type, could also be included in future studies. Vuchic identified seven types: radial, diametrical, tangential, circumferential, trunk with branches, trunk with a feeder, and loops (*10*). A problem may arise with the way in which line type could be modeled. The use of dummy variables (e.g., 1 if the line is radial and 0 otherwise) could be introduced, but the low number of networks studied makes it difficult to account for.

This study also did not include the fact that other transit modes were present in each of the cities. For instance, the bus mode often acts as a feeder to the subway. This is also true for light rail transit and segregated streetcars. The performance of the other transit modes is most likely significant and is related to the performance of the subway. An attempt was made to overcome this by considering the boarding levels for the subway mode only. Moreover, this is why cities with extensive and predominant subway systems were mainly chosen. This minimizes the impacts that these other modes could have on the analysis. In fact, the Rome subway system was not included for that reason. Nevertheless, future work could include the other transit modes, but that work will likely be time-consuming; the different transit technologies would have to be carefully modeled as well. Another option could be to include an intermodality indicator. Furthermore, other factors that have an impact on transit ridership, such as car ownership rates and the price of gasoline, were not accounted for. First, the focus of the study was on network design and not mode split. Second, this research also aimed to prove that the attractiveness of transit does not depend solely on such characteristics (gas prices in North America versus those in Europe) and city design (transit-oriented versus automobile-oriented cities) but depends greatly on network design.

CONCLUSION

This study analyzed 19 subway networks located around the world. The purpose was to investigate the role of network design on ridership. To do so, updated graph theory concepts were used. Moreover, 19 subway network designs were compared by using the annual numbers of boardings per capita as the performance indicator. First, coverage was introduced as a way to measure the relevance of accessibility to transit. This was done by considering the number of stations present in a system over the area served. Second, directness evolved from the concept of the indicator π in graph theory, but its meaning is different. It deals with the maximum number of transfers to go from one station to another; a high number of transfers is often seen as a negative. The third indicator was connectivity. It is an index of the number of transfer possibilities in a network. The three indicators showed a relatively strong relationship with the number of boardings per capita.

A multiple regression analysis showed an even stronger relationship, with a goodness of fit (adjusted R^2 value) of .725. In addition, all *t*-test statistics were statistically significant. It is apparent that the three components introduced play key and equal roles in the design of a network. They therefore have an important impact on ridership. If the three indicators are considered in future public transportation projects, the aim should be to maximize coverage adequately (without impeding speed) and maximize connectivity while making trips as direct as possible.

Finally, more work could be included, for instance, adding line type characteristics or travel time. In addition, interesting results could supposedly emerge by including all transit modes available in a city.

REFERENCES

1. Banister, D. *Transport Investment and Economic Development.* UCL Press, London, 2000.
2. Kozyrev, V. P. Graph Theory. *Journal of Mathematical Sciences*, Vol. 2, No. 5, 1974, pp. 489–519.
3. Kansky, K. J. *Structure of Transportation Networks: Relationships Between Network Geometry and Regional Characteristics.* University of Chicago Press, Chicago, Ill., 1963.
4. Bon, R. Allometry in Topologic Structure of Transportation Networks. *Quality and Quantity*, Vol. 13, No. 4, 1979, pp. 307–326.
5. Taaffe, E. J. *Geography of Transportation.* Prentice-Hall, Upper Saddle River, N.J., 1996.
6. Lam, T. N., and H. J. Schuler. *Public Transit Connectivity,* Vol. 1. Report DMT-084. Institute of Transportation Studies, University of California, Irvine, 1981.
7. Lam, T. N., and H. J. Schuler. Connectivity Index for Systemwide Transit Route and Schedule Performance. In *Transportation Research Record 854*, TRB, National Research Council, Washington, D.C., 1982, pp. 17–23.
8. Gattuso, D., and E. Miriello. Compared Analysis of Metro Networks Supported by Graph Theory. *Networks and Spatial Economics,* Vol. 5, No. 4, 2005, pp. 395–414.
9. Musso, A., and V. R. Vuchic. Characteristics of Metro Networks and Methodology for Their Evaluation. In *Transportation Research Record 1162,* TRB, National Research Council, Washington, D.C., 1988, pp. 22–33.
10. Vuchic, V. R. *Urban Transit: Operations, Planning, and Economics.* John Wiley and Sons, Inc., Hoboken, N.J., 2005.
11. Levinson, H. S. Rail Transit in the Next Millennium: Some Global Perspectives. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1704,* TRB, National Research Council, Washington, D.C., 2000, pp. 3–9.
12. Derrible, S., C. A. Kennedy, E. J. Miller, and M. Roorda. Light Rail Transit in Toronto: Underground or Above-Ground. Presented at 37th CSCE Annual Conference 2008, Canadian Society for Civil Engineering, Quebec City, Quebec, Canada, 2008.
13. Ovenden, M. *Transit Maps of the World.* Penguin Books, London, 2007.
14. Millennium Cities Database for Sustainable Transport. International Association of Public Transport, Brussels, Belgium, 2001.
15. Mobility in Cities Database. International Association of Public Transport, Brussels, Belgium, 2006.
16. Miron, J. R. *Urban Sprawl in Canada and America: Just How Dissimilar?* Cities Lab, University of Toronto at Scarborough, Toronto, Ontario, Canada, 2003.
17. Brinkhoff, T. CityPopulation. www.citypopulation.de. Accessed May 2008.
18. *World Urban Areas (500,000+): Population, Density*. Demographia, Belleville, Ill., 2007.

The Rail Transit Systems Committee sponsored publication of this paper.