Modification of the gravity model and application to the metropolitan Seoul subway system

Segun Goh,[1] Keumsook Lee,[2,*] Jong Soo Park,[3] and M. Y. Choi[1,†]

[1]*Department of Physics and Astronomy and Center for Theoretical Physics, Seoul National University, Seoul 151-747, Korea*
[2]*Department of Geography, Sungshin Women's University, Seoul 136-742, Korea*
[3]*School of Information Technology, Sungshin Women's University, Seoul 136-742, Korea*
(Received 20 April 2012; published 6 August 2012)

The Metropolitan Seoul Subway system is examined through the use of the gravity model. Exponents describing the power-law dependence on the time distance between stations are obtained, which reveals a universality for subway lines of the same topology. In the short (time) distance regime the number of passengers between stations does not grow with the decrease in the distance, thus deviating from the power-law behavior. It is found that such reduction in passengers is well described by the Hill function. Further, temporal fluctuations in the passenger flow data, fitted to the gravity model modified by the Hill function, are analyzed to reveal the Yule-type nature inherent in the structure of Seoul.

I. INTRODUCTION

Social networks are complex systems and have attracted much interest [1–4]. For the maintenance as well as complexity of a social system, interactions between people are crucial. Human beings interact with each other, functioning as the fundamental source of couplings in the social system, by moving to meet or to avoid each other, as if massive particles attract each other.

Among social network systems, of particular interest is the subway system, where interactions are manifested directly by passenger flows, i.e., the numbers of passengers per day and the complex structure of the urban network can be verified. Whereas movements via airplanes [4] or highways [5] are not routine behaviors, daily movements of human beings in the modern society are accomplished in urban networks including the subway system. To probe the interactions in the subway system, one may wish to look into specific trips of individuals. Note, however, that there in general exist too many trips to handle individually. Further, we are not concerned with detailed information on individual trips, which is neither useful nor relevant in coarse-grained modeling [6]. Instead of such a "microscopic description," we thus resort to the "macroscopic description" based on the states of subway stations. The state of each station is described by various measures of the population of its influence area, which we call *mass* in this paper, by analogy with the mass in gravity. Namely, in terms of statistical mechanics, macrostates are specified by $\{M_1, M_2, \dots, M_N\}$, where M_i is the mass of the ith station with N the total number of stations. Since a passenger rides on the subway to go to work, to visit their friends, to take other intercity transportation, or to go shopping, the mass of a station should also reflect regional characteristics of the station.

Employing the macroscopic description with the mass of each station, one may apply the gravity model to the interactions, i.e., the passenger flows, between stations in the subway system. The gravity model, devised after Newton's law of gravity, has been applied widely to various social

systems, such as the Korean highway [5], worldwide commuters [7], cargo-ship movements [8], and intercity phone communications [9]. There has also been an attempt to build a universal model [10]. As discussed in the study of the London subway [11], however, direct human mobility measures exhibit results different from those of indirect measures. Therefore, the application of the gravity model to a direct measure of human movements is required to verify the validity of the model.

In this paper, we consider the Metropolitan Seoul Subway (MSS) system, the network of which is shown in Fig. 1, and examine passenger flows in the system. Specifically, the passenger flows on 24 June 2005 are analyzed. The smart card system, adopted in the MSS system, records the departure and arrival stations and times for each trip, thus allowing one to trace the trip of each individual passenger and providing the direct measure of the urban transportation. In this way, we collected the passenger flow data, which included $N = 357$ subway stations in the network and consisted of 4 907 541 completed trips (or transactions) on the given day [13].

We then apply the gravity model in the MSS system. To probe the scale invariant power-law behavior over the distance, we carry out a detailed analysis of the passenger flow. It is observed that in the long-distance regime, the gravity model indeed applies well. On the other hand, at short distances, there appears distinctive behavior, reflecting complications associated with the network structure, efficiency, convenience, etc.; this urges modification of the model. In addition, fluctuations in the modified gravity model are investigated.

II. GRAVITY MODEL AND STATION MASS

The gravity model, first applied to human movements [14] or trade [15], originates from the Newton's law of gravitational force. Namely, the exchange of personnel or the trade between two cities is modeled to be proportional to the product of the populations of the two cities and inversely proportional to the square of the distance between them. A generalized form of the gravity model for the flow F_{ij} between cities i and j is given by

$$F_{ij} = G \frac{M_i M_j}{r_{ij}^\alpha}, \tag{1}$$
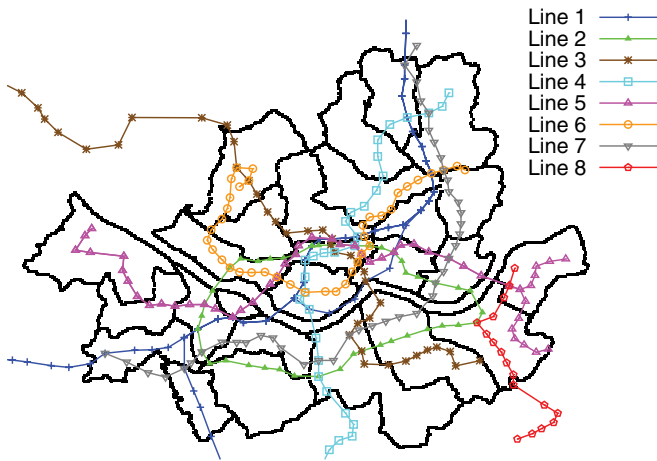
*kslee@sungshin.ac.kr
†mychoi@snu.ac.kr

FIG. 1. (Color online) Metropolitan Seoul subway network, consisting of eight lines (represented by lines with different symbols). Black lines without symbols represent the boundaries of districts in Seoul City. Detailed information is available on such web sites as [12].

where M_i is the population of city i, r_{ij} is the distance between cities i and j, and G is a constant. Here the distance dependence is described by the exponent α, the value of which is not necessarily 2. On the other hand, the exponent for populations M_i and M_j is still taken to be unity, as justified by the entropy maximization [16]. Defining the reduced flow $f_{ij} \equiv F_{ij}/M_i M_j$ and taking the logarithm of the above equation, we obtain the linear form

$$\ln f_{ij} = -\alpha \ln r_{ij} + \ln G, \tag{2}$$

which can be analyzed easily by the method of least squares. However, one should be careful in dealing with the linear form [17]. Note also that links of zero flow may not be handled and should be discarded in processing the data.

In applying the gravity model to a transportation network like the MSS system, it is usually more appropriate to take r_{ij} as the time distance, i.e., the average time taken in a trip between stations i and j. Then Eq. (1) or (2) gives the passenger flow F_{ij} from origin station i to destination j, in terms of their masses M_i and M_j and the time distance r_{ij} between them. (In the case of the nondirected flow between stations i and j, F_{ij} in fact stands for the sum $F_{ij} + F_{ji}$.)

As the mass of a station, we use three parameters: population, number of employees, or strength. The population stands for the number of residents of the *dong* where the subway station is located. (A *dong* is the administrative district of Korean cities which usually consists of tens of thousands of residents and the area reachable on foot.) The number of employees simply represents the number of people employed in the *dong* while the strength is defined to be the total number of passengers using the station: For example, the morning strength corresponds to the total number of passengers using the station in the morning, and the total or departure strength S_i is related to the nondirected or directed flow F_{ij} via [18]: $S_i = \sum_j F_{ij}$. Note that each person makes a constituent of the station mass. Specifically, for example, passengers riding the subway to go to work make up employees of the *dong* where the arrival station is located whereas the population of a departure station is expected to be a measure of the passengers

from that station. We use the data for populations and numbers of employees in the year 2005, compiled by the Government of Korea [19].

If those mass parameters are proportional to each other, there is no need to consider the cases separately. Indeed the number of employees correlates with the morning arrival strength, and with the evening departure strength (with adjusted $R^2 \approx 0.4$). Contrary to our expectation, however, there appears to be no linear correlation between the population and the strength of the morning departure or the evening arrival (adjusted $R^2 < 0.3$). It is also observed that the population is relatively uniform independently of other parameters. This reflects mainly the fact that the population of a region is absorbed into other regions where no subway lines are reached. Due to such absorption of the people, mass parameters tend to be independent of each other and measure rather different quantities.

III. APPLICATION OF THE GRAVITY MODEL TO THE SUBWAY SYSTEM

We first apply the gravity model to the whole MSS system and plot the result in Fig. 2, which shows that the passenger flow decreases indeed linearly (in the logarithmic scale) with the time distance. This allows one to compute the exponent α, depending on the origin-destination mass parameters and directionality of passenger flows, for the given time zone, e.g., morning (4 a.m. to 10 a.m.), afternoon (11 a.m. to 4 p.m.), and evening (5 p.m. to 1 a.m.). Here the reduced χ^2 is chosen as the fitting error because the time distance is fixed and exactly determined when the departure and arrival stations are given. These results are summarized in Table I.

Before proceeding further, we remark on the origin of errors in the gravity model. In fact, if we ignore the real noise effects which are ubiquitous in complex systems, there remain mainly three reasons for the data to stray from the gravity model. The first one is the overall fluctuations, which originate mostly from the mis-selection of the mass parameter. It does not mean that the perfect fit is possible, but it would be possible to minimize fluctuations by selecting a proper mass parameter. Second, there is a curvature around $r_{ij} \approx 6.5$, which corresponds to about 15 min in time. The existence of such a curvature indicates that if the destination is not far enough, people would hesitate to ride the subway and rather take the bus or walk on foot. In fitting, therefore, we use only the data in the range $6.5 \lesssim r_{ij} \lesssim 9.0$, where the data follow faithfully the power law. The curvature in this range is reexamined in the next section. Finally, transfers can complicate the situation, possibly affecting the exponent in the gravity model.

It is of interest that exponents for the population are always larger than those for the number of employees. Further, exponents in the evening are always larger than those in the morning. As demonstrated, the population is more uniformly distributed, and passengers tend to diffuse in the evening rather than to return to their origins [20]. Naively, these behaviors can be explained with the concept of the coupling range. Passenger flows of the population or in the evening are distributed rather uniformly, giving rise to interactions with nearby stations. As for flows of employees
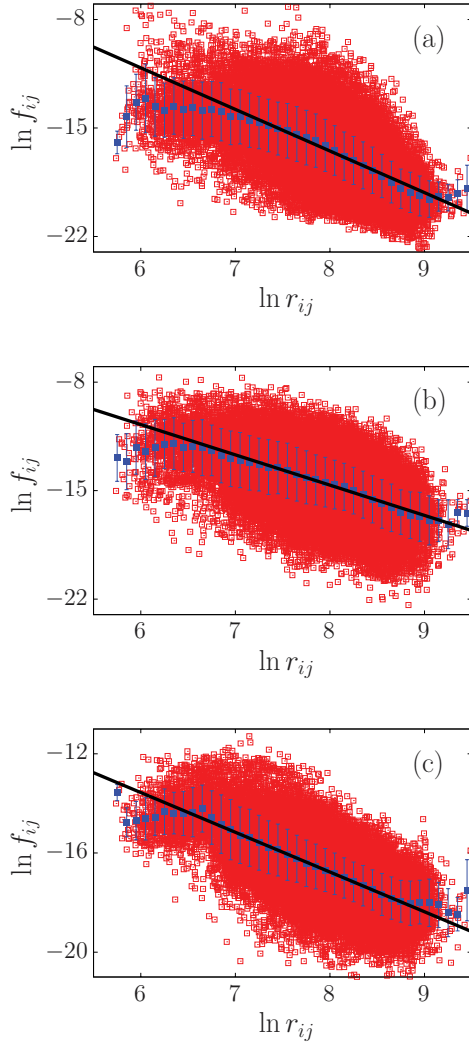
FIG. 2. (Color online) Reduced flow f_{ij} versus time distance r_{ij} in the logarithmic scale. The mass parameter is taken to be (a) the population, (b) the number of employees, and (c) the strength. Nondirected passenger flows for all day are used in all cases. Empty squares ($\square$) represent data, the mean values of which are plotted by filled squares ($\blacksquare$) with error bars estimated by standard deviations. Power-law behaviors are observed in the range $6.5 \lesssim \ln r_{ij} \lesssim 9$; power-law fits over this range are plotted by thick straight lines.

or in the morning, the city appears biased and passengers have to move relatively further. Qualitatively speaking, shorter coupling ranges correspond to more rapid reductions of the distribution with the distance and thus larger values of the exponent.

In addition, reduced χ^2 for flows from the population to the number of employees in the morning is found smaller than that for nondirected flows of the population or the number of employees. It is also true for flows from the number of employees to the population in the evening. These results indicate that the subway is used largely for commuting. Meanwhile, when the strength is used as the mass, the fitting error becomes minimized, reflecting that passenger flows actually constitute the strength.

IV. UNIVERSALITY IN THE SUBWAY SYSTEM

Heretofore we have considered three candidates for the mass parameters in the gravity model. The analysis of the goodness-of-fit measures in Sec. III discloses that the strength serves as the best mass parameter. Unlike intercity transportation, such as the highway [5], it is not obvious to determine the station influence area in the complex urban system, and the gravity model appears to be contaminated by additional spatial fluctuations due to the selection of the "dong" as an influence area. Henceforth, we therefore present results only for the strength as the mass parameter. Furthermore, there are eight lines in the MSS system, and transfers are needed for trips between stations on different lines. To probe the bare interactions without complications due to transfers, we consider each line separately in applying the gravity model. We thus extract data (r_{ij}, F_{ij}) with stations i and j on the same line. Here it should be pointed out that line 2 is a circle line, thus having a different topology from other lines.

Figure 3 demonstrates the gravity model applied to single lines. Note that line 2, which is the longest circle line in the world, connects the most important regions and encircles through downtown areas in Seoul. It is observed that Bundang and Incheon lines as well as line 2 with the unique topology display somewhat different behaviors as expected. Except those, plots of other (ordinary) lines coincide reasonably well with each other. It is remarkable for the ordinary lines to display the power-law behavior with essentially the same exponent (ranging from 1.91 to 1.97). This manifests the emergence of a universality even when individual subway lines pass through quite different regions and some have seriously detouring paths due to the topography and geographic considerations. In addition, it appears that relatively many people use the Incheon and Bundang lines. Since the station mass is the strength, this corresponds to the fact that people taking the line tend to get off at stations on the same line. In other words, these lines are not strongly coupled to subway lines in Seoul, and stations on these lines exchange relatively many passengers with stations on the same lines. On the contrary, data for line 2 are located below those of other lines: Because of the characteristic circular topology, line 2 hosts more transfer stations, thus it is coupled strongly to other lines.

V. MODIFICATION BY THE HILL FUNCTION

As noted in the plot of f_{ij} versus r_{ij}, there is curvature around $r_{ij} \approx 6.5$. Factoring out this, we write the generalized gravity model in the form

$$f_{ij} = \frac{G}{r_{ij}^{\alpha}} g(r_{ij}). \tag{3}$$

Suppose that the curvature can be described by the Hill function:

$$g(r_{ij}) = \frac{r_{ij}^n}{r_{ij}^n + K^n} \tag{4}$$

with the Hill coefficient n. The Hill function, first introduced to describe the equilibrium relationship between the oxygen tension and the saturation of hemoglobin [21], is widely used for chemical reactions, such as binding problems of ligands and

TABLE I. Obtained values of exponent α and reduced χ^2, depending on the time zone, directionality, and origin-destination mass parameters.

			Population-population	Employees-employees	Strength-strength	Population-employees	Employees-population
Morning	Directed	α	1.90(1)	1.26(2)	1.24(1)	1.63(1)	1.53(1)
		Reduced χ^2	2.28	2.81	0.965	1.83	3.36
	Non-directed	α	2.29(1)	1.59(2)	1.39(1)		
		Reduced χ^2	2.33	2.57	0.971		
Afternoon	Directed	α	1.93(1)	1.41(2)	1.22(1)	1.68(1)	1.66(1)
		Reduced χ^2	2.15	2.24	0.860	2.22	2.36
	Non-directed	α	2.24(2)	1.65(2)	1.39(1)		
		Reduced χ^2	2.30	2.28	0.803		
Evening	Directed	α	2.17(1)	1.50(1)	1.37(1)	1.78(1)	1.89(1)
		Reduced χ^2	2.35	2.60	0.847	3.05	2.04
	Non-directed	α	2.48(2)	1.78(2)	1.52(1)		
		Reduced χ^2	2.46	2.52	0.811		
All day	Directed	α	2.50(1)	1.80(1)	1.50(1)	2.12(1)	2.17(1)
		Reduced χ^2	2.46	2.49	0.806	2.50	2.58
	Non-directed	α	2.68(2)	1.95(2)	1.60(1)		
		Reduced χ^2	2.61	2.58	0.810		

receptors or enzyme molecules [22,23]. While the Hill function describes the saturation of the receptors, the individual may be regarded as a receptor and the subway as a ligand. In this manner, the transportation system corresponds to the chemical system with receptors (urban populations) and various kinds of ligands (transportation means including subway, bus, or taxi). If the (time) distance is short and the pressure to take the subway is not high, namely, if the concentration of subway ligands is low, the individual would prefer to take a bus or walk rather than take the subway. In contrast, if the distance is long enough, people would take the subway as receptors are bound by ligands. By analogy with the dissociation constant of a chemical reaction, K in Eq. (4) measures the time distance at which half of the traveling people would use the subway for transportation; we thus call K the time constant. In short, total movements of people tend to grow algebraically as the distance
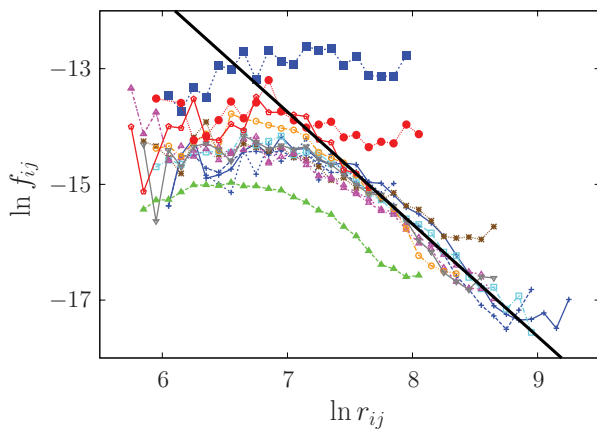
decreases. However, there are other transportation means, e.g., bus, taxi, and walking, in addition to the subway; this leads to the modification described by the Hill function.

In Fig. 4, data for nontransfer flows on all lines but line 2 with the strength as the mass are fitted to Eq. (3). As revealed in the inset, $g(r_{ij})$ is indeed given by the Hill function in Eq. (4). The gravity exponent α, Hill coefficient n, time constant K, and reduced χ^2, obtained via least-squares fitting, are summarized in Table II. Note that the modified model covers the whole range of the time distance, with the reduced χ^2 given by 0.495 while the traditional gravity model is applicable only to the long (time) distance regime. It is also pleasing that the Hill coefficient turns out to be integers ($n = 3$) within the error bars. As is well known, the Hill coefficient larger than unity implies cooperativity on binding of ligands and gives the number of binding sites of the receptor. This is interpreted as cooperativity among transportation means: When a passenger
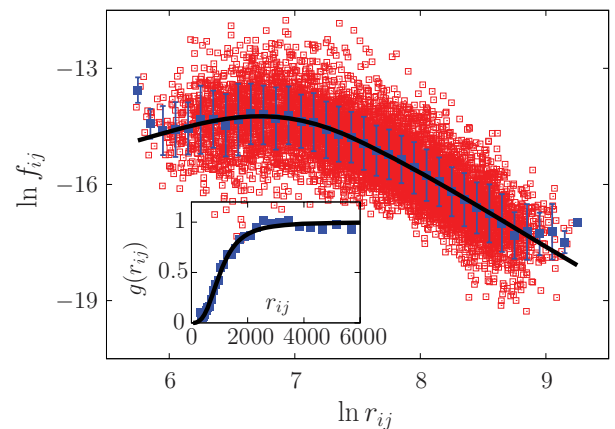

FIG. 3. (Color online) Reduced flow f_{ij} versus time distance r_{ij} on individual lines with the strength as the mass. Data for the Bundang line (●) and the Incheon line (■), as well as for lines 1 to 8 (represented by the same symbols as those in Fig. 1), are plotted together, with lines between the data points being guides to the eye. Linear lines in Seoul (i.e., except line 2) apparently display a universal behavior with the exponent $\alpha = 1.94$ represented by the thick straight line.

FIG. 4. (Color online) Reduced flow f_{ij} versus time distance r_{ij} on all lines but line 2 in Seoul, with the strength as the mass. The thick curved line describes the gravity model modulated by the Hill function, with the results of a least-squares fitting summarized in Table II. The inset shows $g(r_{ij})$ versus r_{ij}, confirming that $g(r_{ij})$ fits well to the Hill function (thick line).

TABLE II. Obtained values of the gravity exponent α, the Hill coefficient n, the time constant K, and reduced χ^2.

α	n	K (min)	Reduced χ^2
1.94(1)	3.0(1)	17(1)	0.495

wishes to make a trip and decides the trip path, he or she may use the subway together with some other transportation means. In other words, subway ligands tend to bind to receptors together with other transportation ligands. The Hill coefficient then corresponds to the number of transportation means used including the subway. It is quite common in Metropolitan Seoul for passengers to transfer from bus to subway and once again from subway to bus to reach their destinations and the Hill coefficient with the value 3 appears reasonable.

The gravity model manifests scale invariance, i.e., power-law behaviors of complex systems. Under modification by the Hill function, there are two scale invariant regimes with different exponents, depending on r_{ij}: In the short (time) distance regime ($r_{ij} \ll K$), we have

$$f_{ij} \sim r_{ij}^{n-\alpha}, \tag{5}$$

while for $r_{ij} \gg K$, the usual behavior

$$f_{ij} \sim r_{ij}^{-\alpha} \tag{6}$$

is recovered. Such cut-off behavior of the power-law distribution was reported in various spatial networks [24,25]. Our model thus provides an explanation of the cut-off behavior in terms of the Hill function.

VI. TEMPORAL FLUCTUATIONS

There exist fluctuations in the data of flows described by Eq. (3), independent of the time distance r_{ij}. To accommodate them, we generalize further the modified gravity model in Eq. (3) as follows:

$$f_{ij} = G \frac{g(r_{ij})}{r_{ij}^{\alpha}} \eta_{ij}. \tag{7}$$

For the strength as the mass, η_{ij} accounts for temporal fluctuations in the strength. To probe them, we divide a time zone into n equal intervals, $\{t_1, t_2, \ldots, t_n\}$, and introduce (fluctuating) strength $S_i(t_k) \equiv S_i e^{\epsilon_i(t_k)}$ at time t_k, the (time) average of which gives the strength S_i: $\langle S_i \rangle \equiv n^{-1} \sum_{k=1}^{n} S_i(t_k) = S_i$. Similarly, the flow in the time zone is given by the average: $F_{ij} = (G/r_{ij}^{\alpha}) g(r_{ij}) \langle S_i(t_k) S_j(t_k) \rangle$, where it is straightforward to identify that $\eta_{ij} = \langle e^{\epsilon_i + \epsilon_j} \rangle$. Taking the logarithm, we have the cumulant expansion: $\ln \eta_{ij} = \langle e^{\epsilon_i + \epsilon_j} - 1 \rangle_c = \langle \epsilon_i \epsilon_j \rangle - \langle \epsilon_i \rangle \langle \epsilon_j \rangle$, where higher-order cumulants have been discarded on the assumption of the average taken over a normal distribution and the constraint $\langle e^{\epsilon_i} \rangle = 1$ has been noted.

Therefore $\ln \eta_{ij}$ is given by an average of random variables, which follows a normal distribution according to the central limit theorem. This implies that η_{ij} displays a log-normal distribution, which emerges from a multiplication of many independent random variables; such a multiplicative nature of fluctuations is inherent in the form $S_i e^{\epsilon_i(t_k)}$. Note that the time evolution of the mass parameters, appearing as fluctuations on short-time scales, reflects the growth of the city on long-time
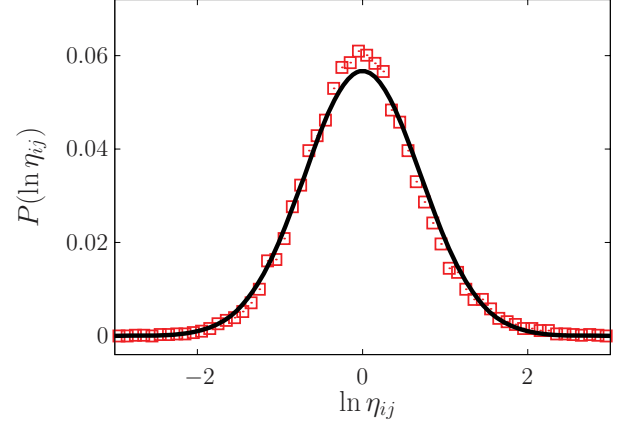

FIG. 5. (Color online) Distribution of fluctuations in the modified gravity model. Data for the strength as the mass ($\square$) follow the Gaussian distribution with a standard deviation of 0.703 (thick line), which is the square root of the reduced χ^2 in Table II.

scales, via the Yule process [20]. These fluctuations can be computed from Eq. (7) with the flow and strength data. Figure 5 shows the result for the nontransfer flows on all lines but line 2 in Seoul; indeed the data fit perfectly to the normal distribution.

VII. CONCLUSION

We have applied the gravity model to the passenger flows in the Seoul Metropolitan Subway system. The exponent in the model has been obtained for different mass parameters, time zones, and trip directionality, and analysis of the coupling on each line reveals a universality present in the system. It has also been demonstrated that the model should be modified by the Hill function, which reflects cooperation in transportation. In addition, we have examined temporal fluctuations; disclosed are fluctuations of the mass parameters following log-normal distributions, the analysis of which helps to probe the evolution of the urban structure.

The main subject of this paper is to established the relevant model for the metropolitan subway passenger flow. Accordingly, we have presented the modified model only with the strength as the mass parameter even though the population or the number of employees should be more useful for infrastructure planning. The analysis based on these mass parameters would reveal a more detailed spatial structure of the city and its subway system.

Finally, we remark that other transportation means have been neglected due to lack of relevant data. In reality, transfer to and from another means would affect the behavior of passengers, altering the exponent. For example, the bus or taxi may be used as an alternative, leading to the exponent different from that for the subway. Further, the analysis including other transportation data would reveal the detailed origin of the Hill coefficient. These are left for further study.

[1] M. Barthélemy, Phys. Rep. **499**, 1 (2011).

[2] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[3] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley, Proc. Natl. Acad. Sci. USA **97**, 11149 (2000).

[4] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004).

[5] W.-S. Jung, F. Wang, and H. E. Stanley, Europhys. Lett. **81**, 48005 (2008).

[6] For this point of view in modeling for macroscopic variables, see, e.g., H. G. Schuster and W. Just, *Deterministic Chaos* (Wiley-VCH, Weinheim, 2005), Chap. 2.

[7] D. Balcan, V. Colizza, B. Gonalves, H. Hu, J. J. Ramasco, and A. Vespignani, Proc. Natl. Acad. Sci. USA **106**, 21484 (2009).

[8] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, J. R. Soc. Interface **7**, 1093 (2010).

[9] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel, J. Stat. Mech. Theor. Exp. (2009) L07003.

[10] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, Nature (London) **484**, 96 (2012).

[11] C. Roth, S. M. Kang, M. Batty, and M. Barthélemy, PLoS ONE **6**, e15923 (2011).

[12] Official web site, www.seoulmetro.co.kr/eng

[13] K. Lee, J. S. Park, H. Choi, M. Y. Choi, and W.-S. Jung, J. Korean Phys. Soc. **57**, 823 (2010).

[14] J. Q. Stewart and W. Warntz, J. Regional Sci. **1**, 99 (1958).

[15] J. Tinbergen, *Shaping the World Economy; Suggestions for an International Economic Policy* (Twentieth Century Fund, New York, 1962).

[16] A. G. Wilson, J. Transp. Econ. Pol. **3**, 108 (1969).

[17] J. M. C. S. Silva and S. Tenreyro, Rev. Econ. Stat. **88**, 641 (2006).

[18] K. Lee, W.-S. Jung, J. S. Park, and M. Y. Choi, Physica A **387**, 6231 (2008).

[19] The National Statistical Office of the Republic of Korea, http://www.index.go.kr

[20] K. Lee, S. Goh, J. S. Park, W.-S. Jung, and M. Y. Choi, J. Phys. A **44**, 115007 (2011).

[21] A. V. Hill, J. Physiol. **40**, 4 (1910).

[22] S. Goutelle, M. Maurin, F. Rougier, X. Barbaut, L. Bourguignon, M. Ducher, and P. Maire, Fundam. Clin. Pharm. **22**, 633 (2008).

[23] J. N. Weiss, FASEB J. **11**, 835 (1997).

[24] M. Barthélemy, Europhys. Lett. **63**, 915 (2003).

[25] R. Guimerà and L. Amaral, Eur. Phys. J. B **38**, 381 (2004).