Journal of the Korean Physical Society, Vol. 60, No. 4, February 2012, pp. 595∼598

Effect of a Heterogeneous Site Size Distribution on the Branch Size Distribution of Tree Networks

Yup Kim, Huiseung Chae and Soon-Hyung Yook*
Department of Physics and Research Institute for Basic Sciences, Kyung Hee University, Seoul 130-701, Korea

(Received 7 October 2011, in final form 24 November 2011)

In this study, we investigate how the heterogeneous site size distribution affects the behavior of the branch size distribution of critical trees. By assuming a power-law, $P(s) \sim s^{-\alpha}$, for the site size distribution, we find that the exponent τ for the branch size distribution is smaller than that of a critical tree when $\alpha < 3$. When $\alpha \geq 3$, we find $\tau = 1.5$, which is the value of the critical tree for $P(s) = \delta(s-1)$. We also find $\tau = 1.5$ for other $P(s)$'s with finite $\langle s^2 \rangle$. Based on our measurement, the physical origin of the discrepancy between the measured branch size distribution in real river networks and that of the critical tree is suggested. As an extension of our study, we also measure the directory size distribution of computer operating systems and discuss the effect of the heterogeneity of the site size distribution τ of a supercritical tree.

PACS numbers: 89.75.Hc, 89.75.Da, 89.75.Fb, 05.90.+m
Keywords: Critical tree, River network
DOI: 10.3938/jkps.60.595

I. INTRODUCTION

Since many important natural phenomena can be easily described by mapping into phenomena on branched tree networks, studies on the topological properties of branched tree networks play great roles in understanding various phenomena in nature. Examples include branching processes such as nuclear chain reactions and directed percolation [1], specification in biology [2], decision making in sociology [3], and branched flows in rivers [4] and blood vessels [5]. Interestingly, many of those phenomena share a universal feature that the branch size distribution satisfies a power-law [4]

$$P(a) \sim a^{-\tau}. \tag{1}$$

Here, a is the size of branch, which is defined by the number of sites making up the branch in the tree. An important question involved in studies on the topological properties of branched network is how to decide the universality class of different natural and artificial branched networks.

Analytical derivations have shown that there can be two universality classes depending on the average branching ratio. If the average branching ratio is unity, then the resulting tree becomes critical [6]. For a critical branching process, it is well known that $\tau = 3/2$. When the average branching ratio is larger than unity, the resulting tree becomes supercritical and $\tau = 2$ [7]. The Internet [8], the communities in scientific collaboration networks [9], the distribution of taxa in biological specification [2], and the directory size distributions in computers [10] are well described by the branch size distribution of a supercritical tree. However, the drainage area distribution of river networks [4,11–13], communities in organizations [14], and jazz musician networks [15] are believed to be well described by the critical branching process. However, the empirical value of τ for the size distribution of the drainage area in river network is known to be $\tau \approx 1.41 \sim 1.45$ [11–13]. Here, the drainage area corresponds to the branch size of tree networks. the values of τ of other critical trees such as the Internet are also known to be $\tau \approx 1.45$ [7]. These values of τ are close to 3/2, but still noticeably smaller than 3/2. There have been several attempts to explain the deviations from the critical value. Scheidegger showed that the value of τ decreases to $\tau = 4/3$ by using a model based on the random walk system [16,17]. This value of τ obtained from Scheidegger's model was still far from the empirical values. More recently, when a slope-slope correlation of the morphology on various landscape exists, τ is known to be $\tau \approx 1.42$ [18], which agrees only with the lower bound of τ in river networks.

In this paper we suggest a physical origin for the discrepancy between the empirically obtained value of τ in river networks and that for the critical tree. In general, the drainage area or branch size of the river network is defined as the number of sites connected through drainage directions. This implies that each site has a drainage area of uniform size. In other words, it has

*E-mail: syook@khu.ac.kr

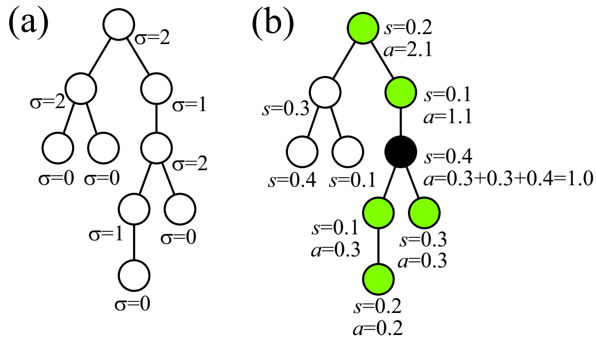

Fig. 1. (Color online) (a) Schematic diagram for generating a critical tree. σ represents the number of new branches. (b) Assignment of site size s to each site. s is randomly drawn from a given distribution $P(s)$. For each filled circle, a is the branch size of the site.

drainage area of unit size and does not have any internal structure. However, it is more reasonable to assume that each site has its own internal structure and that the area covered by a site fluctuates from site to site. Therefore, we introduce a fluctuation in the size of a site and show that the heterogeneity of the site size can change the value of τ. We also discuss the relationship between the rainfall distribution and the site size distribution as a possible origin of the heterogeneity in the site size of river networks. For a comparison, we also investigate the directory size distributions in computer systems.

This paper is organized as follows. In Sec. II, we define the model. The simulation results are shown in Sec. III. The relationship between the rainfall distribution and the site size distribution is given in Sec. IV. In Sec. V, we show the measurement of the directory size distributions for computer operating systems for a comparison. The summary and discussion are given in Sec. VI.

II. MODEL

In order to incorporate the fluctuation in site size into the branch size distribution of the critical tree, we first generate the critical tree. The tree becomes critical when the average branching ratio is unity; *i.e.*, $b = \sum_\sigma \sigma p_\sigma = 1$. Here, p_σ is the probability to trigger σ new branches (see Fig. 1(a)). Then, we assign a size of site i, s_i, for each i. The s_i is drawn from a give distribution $P(s)$. We let $\mathcal{S}$ be the set of sites that belong to the sub-tree rooted from i and i itself. Then, the size of the branch rooted from site i is defined as $a_i = \sum_{j\in\mathcal{S}} s_j$ (see Fig. 1(b)). In Fig. 1(b), we present an example to determine a_i at each site for filled circles. For example, a_i for the black site is determined as $a_i = 0.2 + 0.1 + 0.3 + 0.4$ as shown in Fig. 1(b). In the following simulations we show the results for $\sigma \in \{0, 1, 2\}$, but we check that the results are not affected by the upper bound on σ. Then, we measure $P(a)$ over at least 10^4 realizations of tree.

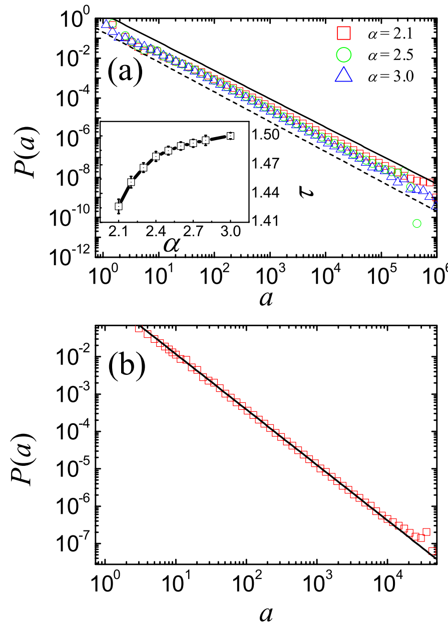

Fig. 2. (Color online) (a) Plot of $P(a)$ when $P(s) \sim s^{-\alpha}$. The dashed line denotes the relation $P(a) \sim a^{-\tau}$ with $\tau = 1.5$. The solid line corresponds to $\tau = 1.43$. Inset: Plot of τ against α. (b) Plot of $P(a)$ when $P(s) \sim \exp[-(s-1)^2/2]$. The solid line corresponds to $\tau = 1.5$.

III. BRANCH SIZE DISTRIBUTION WITH A HETEROGENEOUS SITE SIZE DISTRIBUTION

In order to study the effect of heterogeneity in the site size on the branch size distribution, we first assume that $P(s)$ follows a power-law

$$P(s) \sim s^{-\alpha}. \tag{2}$$

In critical or supercritical branching, the branch size distribution can be easily calculated by using the generating function [7]. However, when $P(s)$ is not a trivial distribution as Eq. (2), the initial condition becomes nontrivial. Thus, the analytic derivation of $P(a)$ also becomes nontrivial. Therefore, we use a Monte Carlo simulation to obtain $P(a)$. In Fig. 2(a) we show the measured $P(a)$ from simulations for various values of $\alpha \in [2.1, 3.0]$. Using the least-squares fit of the data to Eq. (1), we obtain $\tau = 1.43 \pm 0.01$ for $\alpha = 2.1$, and $\tau = 1.50 \pm 0.01$ for $\alpha \geq 3.0$. This shows that τ is a function of α. In the inset of Fig. 2(a) we display the estimated τ against α. The data clearly shows that the value of τ increases as α increases. When α is close to $\alpha = 3$, τ approaches to $\tau = 1.5$, which corresponds to the value for the critical tree. Moreover, $\alpha \in [2.1, 2.4]$ gives the value of τ measured from the drainage area distribution in real river networks. In Fig. 2(b) we display the measured $P(a)$ when $P(s)$ follows the Gaussian distribution, $P(s) \sim \exp[-(s-1)^2/2]$. From the best fit of the data to Eq. (1), we obtain $\tau = 1.50 \pm 0.01$. For other

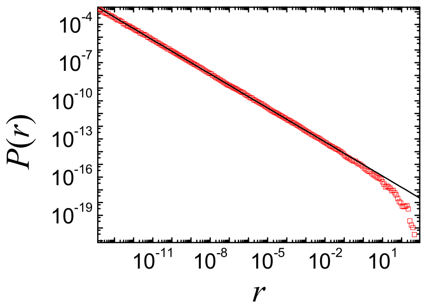

Fig. 3. (Color online) Measured $P(r)$ in Australia. The solid line represents the relation $P(r) \sim r^{-\beta}$ with $\beta = 0.9$.

distributions for $P(s)$ in which $\langle s^2 \rangle$ is finite, we obtain $\tau = 1.5$ (which is not shown). This result indicates that when the site size distribution is heterogeneous enough, or $\langle s^2 \rangle \to \infty$, then the value of τ significantly deviates from that for the critical tree. However, if the site size distribution becomes rather homogeneous, or $\langle s^2 \rangle$ is finite, the value of τ is equal to that for the critical tree.

IV. RAINFALL DISTRIBUTION

In real river networks, the area that is covered by each site would be closely related to the rainfall at each site in the following way. The flow at each site should be proportional to the rainfall, and the area covered by a site would be proportional to the flow at the site. However, the exact relation between the rainfall (flow) distribution and the distribution of the area covered by the site in a river network is not yet fully understood. In order to investigate the relationship between the rainfall and the area covered by each site, we measured the rainfall distribution, $P(r)$, by using the Australian daily rainfall gridded data [19]. The data in Ref. 19 have been recorded for 30 years (1961-1990). The area of each gridded region in the dataset covers around 25 km^2. From this dataset, we measure $P(r)$ for each gridded area in Australia. The result is shown in Fig. 3. From the data in Fig. 3, we find that $P(r)$ satisfies another power-law

$$P(r) \sim r^{-\beta}, \tag{3}$$

with $\beta = 0.9 \pm 0.1$. The value of β seems to be slightly different from country to country. For example, in Korea we also find that $P(r)$ follows Eq. (2) with $\beta \simeq 1.2$ [20]. The obtained value of β is much smaller than the value of $\alpha = 2.1 \sim 2.7$, which corresponds to the value of τ measured from the drainage area distribution in real river networks. This indicates that rainfall is not linearly proportional to the size of the site. The discrepancy between $P(s)$ and $P(r)$ can be understood from the definition of the drainage area. Since the drainage area is defined by the water flow on the surface, the part of the rainfall that permeates into the soil must be excluded when the

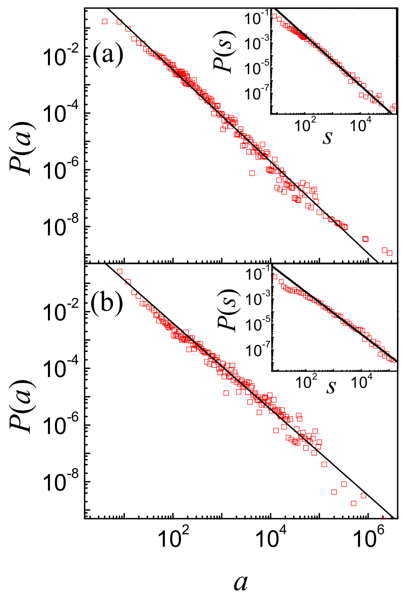

Fig. 4. (Color online) (a) Plot of the directory size distribution for Linux. The solid line represents $\tau = 1.6$, which is significantly smaller than $\tau = 2.0$ for supercritical branching. Inset: Plot of the data size distribution stored in each directory. This data size distribution corresponds to $P(r)$ of the rainfall distribution. The solid line denotes $\alpha = 2.0$. (b) Plot of the directory size distribution for Microsoft Windows 7. The line represents $\tau = 1.5$. Inset: Plot of the data size distribution. The line denotes $\alpha = 1.7$.

distribution of the drainage area or $P(s)$ is estimated. Form our measurement, we expect that the amount of water on the surface, w, would satisfy a power-law relation to the rainfall as $w \sim r^{\delta}$. If $w \sim s$ as the simplest assumption, then δ is in the range $\delta \in [1.8, 2.3]$.

V. SIZE DISTRIBUTION OF THE DIRECTORY IN COMPUTER OPERATING SYSTEMS

For a comparison, we also measured the directory size distribution for operating systems (OS) of a computer. The branch size distribution of the directories in a computer is also known to follow the power-law in Eq. (1) with $\tau \approx 2$ [10] under the assumption that the sizes of all sites (or directories) are the same. This value of τ corresponds to that for supercritical branching [7]. To investigate the effect of the site size fluctuation on the branch size distribution of the supercritical tree, we consider a directory as a site in Sec. II, and we define s_i for each site i as the sum of file sizes over all files in the directory i. In the sum, the files in subdirectories of i are of course excluded. Then, the branch size a_i of a directory i becomes the sum of the file sizes over all files in the directory and its subdirectories.

In Fig. 4(a) we show the measured $P(a)$ for the Linux system. From the best fit of the data to Eq. (1), we obtain $\tau = 1.6 \pm 0.1$, which is significantly smaller than $\tau = 2.0$ for the supercritical tree. In the inset of Fig. 4(a), we show the the measured $P(s)$ for the Linux system. As shown in the inset of Fig. 4(a), we find that $P(s)$ for the Linux system also satisfies Eq. (2). From the data we obtain $\alpha = 2.0 \pm 0.1$. We also obtain a similar distribution for other OS's. For example, in Fig. 4(b), we display the measured $P(a)$ for the Microsoft Windows 7 system. The solid line in Fig. 4(b) represents the power-law relation in Eq. (1) with $\tau = 1.5 \pm 0.1$. The inset of Fig. 4(b) shows that the $P(s)$ for the Microsoft Windows 7 system also satisfies Eq. (2) with $\alpha = 1.7 \pm 0.1$. These results show that not only for the critical tree but also for the supercritical tree, the exponent τ for the branch size distribution is strongly affected by the heterogeneity of the site size distribution.

VI. SUMMARY AND DISCUSSION

In summary, we study the effect of the fluctuation in the site size on the branch size distribution of the critical tree. For several distributions of $P(s)$, we show that the branch size distribution exponent τ is closely related to the second moment of the site size distribution $\langle s^2 \rangle$. When $\langle s^2 \rangle$ diverges, the value of τ significantly deviates from $\tau = 1.5$ for the critical tree and is smaller than $\tau = 1.5$. On the other hand, for finite $\langle s^2 \rangle$, τ is not affected by the site size fluctuation, and $\tau = 1.5$. Therefore, our results provide a clue to understand the origin of the deviation of τ for real trees from that for a critical tree. For example, in river networks, the heterogeneous site size distribution caused by the heterogeneity in rainfall distribution strongly affects the value of τ. Moreover, from measurements of the directory size distribution for Linux systems and Microsoft Windows 7 systems, we find a similar behavior in τ. The measured $P(s)$'s both for Linux directories and Microsoft Windows directories show that $\langle s^2 \rangle \to \infty$. Consequently, τ's for computer systems are much smaller than $\tau = 2$ for the supercritical tree as for the critical tree. Therefore, the effect of the heterogeneous site size distribution on the branch size distribution is physically crucial.

ACKNOWLEDGMENTS

We thank Dr. J. Park for useful discussions. This work was supported by National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MEST) (No. 2009-0073939 and No. 2011-0015257). One of the authors (Yup Kim) acknowledges the support from Kyung Hee University through the requested research year 2011.

REFERENCES

[1] J. Cardy and R. L. Sugar, J. Phys. A: Math. Gen. **13**, L423 (1980).
[2] B. Burlando, J. Theor. Biol. **163**, 161 (1993).
[3] B. Blasius and R. Tönjes, Phys. Rev. Lett. **103**, 218701 (2009).
[4] I. Rodriguez-Iturbe and A. Rinaldo, *Fractal River Basins: Chance and Self-Organization* (Cambridge University Press, New York, 1997).
[5] G. B. West, J. H. Brown and B. J. Enquist, Science **276**, 122 (1997).
[6] T. E. Harris, *The Theory of Branching Processes* (Dover, New York, 1989).
[7] P. de Los Rios, Europhys. Lett. **56**, 898 (2001).
[8] G. Caldarelli, R. Marchetti and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[9] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto and D. Parisi, Proc. Nat. Acad. Sci. **101**, 2658 (2004).
[10] K. Klemm, V. M. Eguíluz and M. S. Miguel, Phys. Rev. Lett. **95**, 128701 (2005).
[11] A. Maritan, A. Rinaldo, R. Rigon, A. Giacometti and I. Rodriguez-Iturbe, Phys. Rev. E **53**, 1510 (1996).
[12] I. Rodriguez-Iturbe, R. L. Bras, E. Ijjasz-Vasquez and D. G. Tarboton, Water Resour. Res. **28**, 988 (1992).
[13] S. P. Breyer and R. S. Snow, Geomorphology **5**, 143 (1992).
[14] R. Guimera, L. Danon, A. Díaz-Guilera, F. Giralt and A. Arenas, Phys. Rev. E **68**, 065103(R) (2003).
[15] A. Arenas, L. Danon, A. Díaz-Guilera, P. Gleiser and R. Guimera, Eur. Phys. J. B **38**, 373 (2004).
[16] A. E. Scheidegger, Geol. Soc. Am. Bull. **72**, 37 (1961).
[17] A. E. Scheidegger, Water Resour Res. **4**, 167 (1968).
[18] H-J. Kim, I-M. Kim and J. M. Kim, Phys. Rev. E **62**, 3121 (2000).
[19] http://www.water.gov.au and http://www.bom.gov.au.
[20] http://www.kma.go.kr.