

# Asymptotic behavior of the Kleinberg model

Shai Carmi,<sup>1,2</sup> Stephen Carter,<sup>3</sup> Jie Sun,<sup>3</sup> and Daniel ben-Avraham<sup>4,3,\*</sup>

<sup>1</sup>*Minerva Center & Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel*

<sup>2</sup>*Center for Polymer Studies, Boston University, Boston, MA 02215, USA*

<sup>3</sup>*Department of Mathematics & Computer Science, Clarkson University, Potsdam, NY 13699-5815*

<sup>4</sup>*Physics Department, Clarkson University, Potsdam, New York 13699-5820, USA*

We study Kleinberg navigation (the search of a target in a  $d$ -dimensional lattice, where each site is connected to one other random site at distance  $r$ , with probability  $\sim r^{-\alpha}$ ) by means of an exact master equation for the process. We show that the asymptotic scaling behavior for the delivery time  $T$  to a target at distance  $L$  scales as  $T \sim \ln^2 L$  when  $\alpha = d$ , and otherwise as  $T \sim L^x$ , with  $x = (d - \alpha)/(d + 1 - \alpha)$  for  $\alpha < d$ ,  $x = \alpha - d$  for  $d < \alpha < d + 1$ , and  $x = 1$  for  $\alpha > d + 1$ . These values of  $x$  exceed the rigorous lower-bounds established by Kleinberg. We also address the situation where there is a finite probability for the message to get lost along its way and find short delivery times (conditioned upon arrival) for a wide range of  $\alpha$ 's.

PACS numbers: 05.40.Fb, 02.50.-r, 89.75.Hc, 05.60.-k

In a now famous study [1] the social psychologist Stanley Milgram asked randomly chosen people to send a postcard to a disclosed target in the USA. The participants were to send the card to a person they knew on a first-name basis, who will then send it on to another acquaintance, and so on, until it reached the desired destination. 20% of the cards successfully reached the target using, on average, chains of 6.5 acquaintances, confirming the notion that the network of social contacts has the *small-world* property [2]: a very short path, typically logarithmic in the size of the system, connects between any two nodes. How does the message find its way, let alone so efficiently? Searching and navigation problems such as this [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] are relevant to several disciplines, from sociology, to efficient algorithms in computer science, to the understanding of the foraging behavior of insects and animals.

The problem has been elegantly addressed in the seminal work of Jon Kleinberg [3, 4]. Kleinberg considers an  $(L \times L)$ -square lattice, where in addition to the links between nearest neighbors each node  $i$  is connected to a random node  $j$  with a probability  $p_{ij} = r_{ij}^{-\alpha} / \sum_k r_{ik}^{-\alpha}$  ( $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$  is the Euclidean distance between nodes  $i$  and  $j$ , and the sum in the denominator excludes  $k = i$ ). Suppose that a message is to be passed from a “source” node  $s$  to a “target” node  $t$ , along the links of the network, by a *decentralized* algorithm — an algorithm that relies solely on *local* information. Kleinberg shows that when the exponent  $\alpha$  equals  $d$ , the lattice dimensionality, an algorithm exists that requires less than  $\ln^2 L$  steps to complete the task. If  $\alpha \neq d$ , the delivery time,  $T$ , grows as  $L^x$ , with rigorous lower bounds,  $x \geq x_K$  [3, 4, 14];

$$T \sim L^x, \quad x \geq x_K = \begin{cases} \frac{d-\alpha}{d+1} & 0 \leq \alpha < d, \\ \frac{\alpha-d}{\alpha-d+1} & \alpha > d. \end{cases} \quad (1)$$

Moreover, no local algorithm performs better, functionally, than the simple-minded *greedy* algorithm: *Pass the message forward to the neighbor node that is closest to the target* (geographically).

In this letter we study the asymptotic long-time behavior of the Kleinberg search process. We find the exact form of the exponent  $x(\alpha)$ , and we show that  $T \sim \ln^2 L$  is the actual scaling (not just a bound) for the special case of  $\alpha = d$ . Our approach is based on a master equation for the full probability distribution for completing a search within a given time. This formalism also enables one to consider the possibility of the message getting lost along its way, and we discuss briefly some surprising outcomes of that scenario.

Because the message gets closer to the target with each successive step, nodes are never revisited and one can view each long-range link as being created only as the message arrives at its site of origin. The message closes in on the target in a peculiar kind of directed Lévy walk, consisting of a mix of “short” (one lattice spacing) and “long” (power-law distributed) steps. In this search the dimensionality of the lattice enters into consideration mainly as a Jacobian in the various sums (or integrals) of the equations describing the process. We therefore limit the following derivations to one dimension and generalize the results for higher dimensions, having made the necessary adjustments.

For convenience, to render the  $p_{ij}$  independent of  $i$ , we adopt periodic boundary conditions. Specifically, consider a ring of length  $4L$ , with the source at 0 and the target at  $L$ , and the range of the long-contact links limited to  $2L - 1$  [15]. In that case,

$$p_{i,i+k} = Ak^{-\alpha}; \quad A = \left( 2 \sum_{k=1}^{2L-1} k^{-\alpha} \right)^{-1}, \quad (2)$$

where  $A$  is a normalizing factor. Let  $P(n;l)$  denote the probability that a message, at distance  $l$  from the target, takes  $n$  additional steps to reach the target. Once the message is at the target it takes no additional time

\*Electronic address: benavraham@clarkson.edu

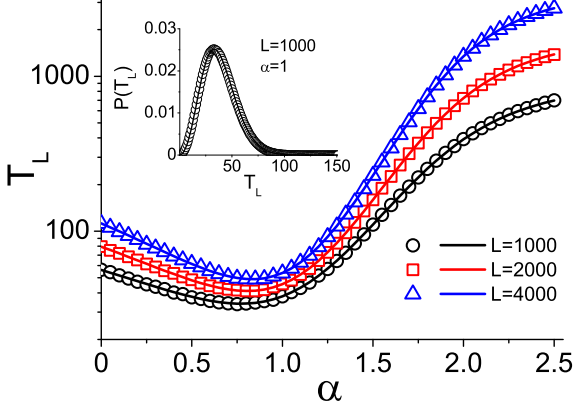


FIG. 1: (Color online) Mean delivery time,  $T_L$ , as a function of the long-contact exponent,  $\alpha$ . Note the perfect agreement between simulations (solid line) and the results from Eq. (4) (symbols). Shown are results for three values of  $L$ . Inset: Distribution of the delivery time for the case of  $\alpha = 1$  and  $L = 1000$  as computed from (3) (solid line) is compared to simulations (symbols).

to reach it, so  $P(n; 0) = \delta_{n,0}$ . Likewise,  $P(0; l) = \delta_{0,l}$ , since the only way to reach the target in 0 steps is if the message is already there to begin with.

$P(n; l)$  satisfies the equation

$$P(n; l) = A \sum_{k=1}^{2l-1} k^{-\alpha} P(n-1; |l-k|) + \left(1 - A \sum_{k=1}^{2l-1} k^{-\alpha}\right) P(n-1; l-1). \quad (3)$$

The first term on the rhs represents the events that the first of the additional  $n$  steps is a long step of length  $k$ , in which case the message would come to within distance  $|l-k|$  from the target. The second term represents a short step, that advances the message a single lattice spacing.

For our main purpose here it is sufficient to consider just the first moment  $\langle n \rangle \equiv T_l$ , that is, the mean delivery time from a site a distance  $l$  away from the target. Multiplying Eq. (3) by  $n$  and summing over  $n$ , we get

$$T_l = A \sum_{k=1}^{2l-1} k^{-\alpha} (1 + T_{|l-k|}) + \left(1 - A \sum_{k=1}^{2l-1} k^{-\alpha}\right) (1 + T_{l-1}), \quad (4)$$

for  $l = 1, 2, \dots, L$ . Numerical integration of Eqs. (3) and (4) yields perfect agreement with the results from direct simulation of the Kleinberg navigation process on a ring (Fig. 1).

Using the fact that  $T_{-k} = T_k$ , and defining  $D_k = T_k -$

$T_{k-1}$ , we obtain, after some rearranging,

$$D_{l+1} - D_l = A \left\{ \sum_{k=1}^l [(l+1-k)^{-\alpha} - (l+k)^{-\alpha}] D_k - \sum_{k=1}^{2l-1} k^{-\alpha} D_l \right\}, \quad (5)$$

for  $l = 1, 2, \dots, L-1$ , while for  $l = 0$  we have  $D_1 = T_1 - T_0 = 1$ .

As a quick check, consider the limit of  $\alpha \rightarrow \infty$ , when all the long-range contacts are restricted to length 1, and therefore one expects  $T_l = l$ . Indeed, in this case all the  $k^{-\alpha}$  terms in the equation tend to zero, unless  $k = 1$ , and we get  $D_{l+1} - D_l = 0$ , which along with  $D_1 = 1$  yields  $D_k = 1$ , and  $T_l = \sum_{k=1}^l D_k = l$ , just as expected.

Next, consider the opposite limit, of  $\alpha = 0$ , where the distribution of long-range contacts is homogeneous. In this case  $A = [2(2L-1)]^{-1}$  and we obtain from (5),

$$D_{l+1} - D_l = -\frac{2l-1}{2(2L-1)} D_l.$$

Although this equation can be solved exactly, a continuous approximation,

$$\frac{d}{dl} D(l) = -\frac{l}{2L} D(l),$$

assuming  $L \geq l \gg 1$ , works just as well. In view of the boundary condition  $D(0) = 1$ , this has the solution  $D(l) = \exp(-l^2/4L)$ . Then,  $T(L) = \int_0^L D(l) dl$ . The upper integration limit may be safely replaced with  $\infty$ , due to the rapid decay of the gaussian, and a simple change of variables yields  $T(L) \sim L^{1/2}$ , in perfect agreement with the Kleinberg bound for  $\alpha = 0$ .

For larger values of  $\alpha$  we are not as fortunate as to find a full analytic solution, but we can still obtain the asymptotic behavior. For  $0 \leq \alpha < 1$  we take a hint from the solution for  $\alpha = 0$  and make the *ansatz*  $D(l) = f(l^\beta/L)$ , where  $f(x)$  is a smoothly decreasing function;  $f(x) = \mathcal{O}(1)$  for  $x \lesssim 1$ , and decays very rapidly (e.g., exponentially) for  $x \gtrsim 1$ . Consistent with this behavior, the derivative at the crossover point  $x_* = 1$  is  $f'(x_*) = -\mathcal{O}(1)$ . This ansatz is nicely confirmed by numerical integration of Eq. (5).

Apply now Eq. (5) to the crossover length  $l_* = L^{1/\beta}$ . The lhs is

$$D_{l_*+1} - D_{l_*} \approx \frac{d}{dl} D(l)|_{l=l_*} \sim -L^{-1/\beta},$$

while the sums on the rhs can be estimated by replacing  $D_l$  with a constant for  $l < l_*$ , and zero for  $l > l_*$ , yielding  $-A l_*^{1-\alpha}$ . But  $A \sim 1/L^{1-\alpha}$ , leading to  $-1/\beta = (1-\alpha)/\beta - (1-\alpha)$ , or  $\beta = (2-\alpha)/(1-\alpha)$ . Finally,

$$\begin{aligned} T(l) &= \int_0^L D(l) dl \approx \int_0^\infty D(l) dl \\ &= \frac{L^{1/\beta}}{\beta} \int_0^\infty f(x) x^{1/\beta-1} dx \sim L^{1/\beta}, \end{aligned}$$

so

$$T(L) \sim L^{(1-\alpha)/(2-\alpha)}, \quad 0 \leq \alpha < 1. \quad (6)$$

For  $\alpha > 1$ , we sum Eq. (5) over  $l$ , taking into account that  $D_1 = 1$ , and rearrange:

$$D_{l+1} - 1 = -A \sum_{k=1}^l \sum_{m=-k+2}^k (l+m)^{-\alpha} D_k. \quad (7)$$

This can be obtained more directly also by rearranging Eq (4). The inner sum over  $m$  can be approximated by an integral, yielding

$$D_{l+1} - 1 \approx \frac{A}{\alpha - 1} \sum_{k=1}^l [(l+k)^{1-\alpha} - (l-k+2)^{1-\alpha}] D_k. \quad (8)$$

Since  $\alpha > 1$ ,  $A$  converges, as  $L \rightarrow \infty$ , and we may simply follow powers of  $l$ . Assume first that  $D_l \sim l^{-\beta}$ . The dominant  $-1$  on the lhs of the equation must be balanced by the dominant term on the rhs, which scales as  $-l^{2-\alpha-\beta}$ , so  $\beta = 2 - \alpha$ . This leads to

$$T \sim L^{\alpha-1}, \quad 1 < \alpha < 2. \quad (9)$$

The upper limit on  $\alpha$  follows from the fact that  $T_L$  cannot increase faster than linearly in  $L$ .

For the special case of  $\alpha = 1$ , we have

$$D_{l+1} - 1 \approx A \sum_{k=1}^l \ln \frac{l+k}{l-k+2} D_k,$$

and  $A \sim (\ln L)^{-1}$ . To counter the  $-1$  on the lhs one must then allow that  $D_l \sim (\ln L)/l$ , for large  $l$ , which leads to

$$T \sim \ln^2 L, \quad \alpha = 1, \quad (10)$$

exactly as the Kleinberg *upper limit* for this case.

The dominant  $-1$  term on the lhs of (8) might also be cancelled if  $D(l) = 1 - g(l)$ , where  $g(l)$  vanishes as  $l \rightarrow \infty$ . Substituting this ansatz in (8) we find  $g(l) \sim l^{2-\alpha}$ , and integration of  $D(l)$  yields

$$T \sim L + \mathcal{O}(L^{3-\alpha}), \quad \alpha > 2. \quad (11)$$

In this case, the condition  $\alpha > 2$  prevents the correction term from growing faster than linearly.

In Fig. 2a we compare our various results for  $x(\alpha)$  in one dimension to computer simulations [16] and to the Kleinberg bounds,  $x_K(\alpha)$ . For  $\alpha \rightarrow 2$  we expect logarithmic terms, as the main contribution and the correction term in (11) approach then the same power-law. The logarithmic behavior makes it very difficult to extract reliable estimates from simulations for the exponent  $x(\alpha)$  near  $\alpha = 1$  and 2.

The foregoing results, argued for one-dimensional lattices, are easily generalized to higher dimensions: the space dimensionality,  $d$ , enters in the various integrals

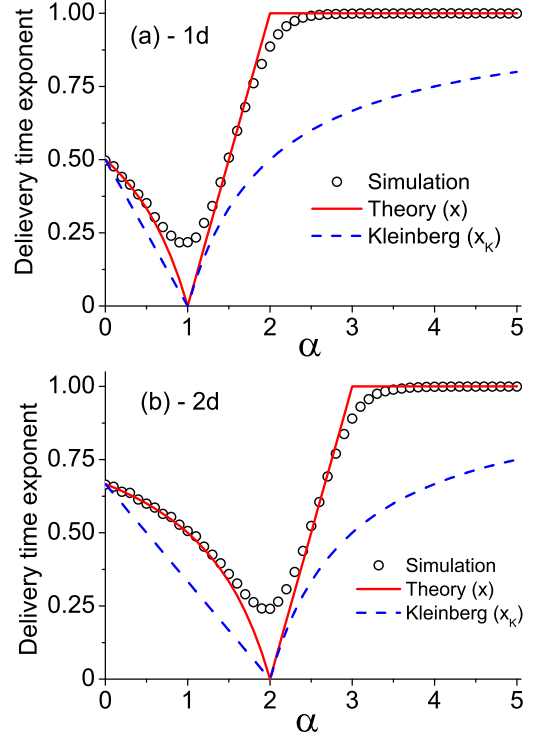


FIG. 2: (Color online) Delivery time exponent,  $x$ , as a function of  $\alpha$  in (a)  $d = 1$  and (b)  $d = 2$  dimensions. Results from Eq. (12) (solid lines) are compared to the Kleinberg bounds (broken lines) and to simulations (symbols).

(or sums) through the Jacobian for  $d$ -dimensional integration. Making the necessary adjustments, we find

$$T \sim L^x, \quad x = \begin{cases} \frac{d-\alpha}{d+1-\alpha} & 0 \leq \alpha < d, \\ \alpha - d & d < \alpha < d+1, \\ 1 & \alpha > d+1. \end{cases} \quad (12)$$

Once again, these results agree with the Kleinberg bounds  $x \geq x_K$ . Eq. (12) fits simulations results for  $d = 2, 3, 4$  nicely (we did not test higher dimensions). Results for  $d = 2$  are shown in Fig. 2b.

Returning now to the probability distribution,  $P(n; l)$ , a standard way to tackle Eq. (3) is through the generating function

$$\hat{P}(z; l) \equiv \sum_{n=0}^{\infty} P(n; l) z^n.$$

The “fugacity”  $z$  may be interpreted as the probability to complete a single step successfully. In that case  $P(z; l)$  is the total probability to complete the delivery successfully (to a target at distance  $l$ ). We defer a more detailed study of  $P(n; l)$  for future work and focus for the moment on the conditional average of the delivery time,  $T_{z,l}$ , that is, the average delivery time conditioned upon successful

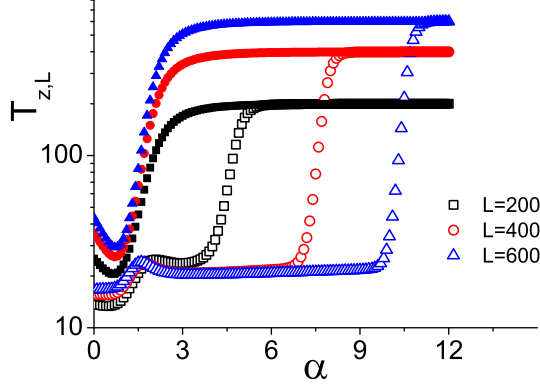


FIG. 3: (Color online) Conditional delivery time,  $T_{z,l}$ , as a function of  $\alpha$ , for  $L = 200$  (squares), 400 (circles), and 600 (triangles). Results for  $z = 0.9$  (empty symbols) are compared to perfect transmission (solid symbols).

arrival at the target:

$$T_{z,l} = \frac{\sum_{n=0}^{\infty} n P(n;l) z^n}{\sum_{n=0}^{\infty} P(n;l) z^n} = z \frac{\partial \hat{P}(z;l)/\partial z}{\hat{P}(z;l)}. \quad (13)$$

Multiplying (3) by  $z^n$  and summing over  $n$ , minding the boundary conditions, we obtain

$$\begin{aligned} \hat{P}(z;l) &= A \sum_{k=1}^{2l-1} k^{-\alpha} z \hat{P}(z;l-k) \\ &+ \left(1 - A \sum_{k=1}^{2l-1} k^{-\alpha}\right) z \hat{P}(z;l-1), \end{aligned} \quad (14)$$

Taking the derivative of this equation with respect to  $z$ , and writing  $\partial \hat{P}(z;l)/\partial z \equiv \hat{F}(z;l)$ , we have

$$\begin{aligned} \hat{F}(z;l) &= A \sum_{k=1}^{2l-1} k^{-\alpha} [z \hat{F}(z;l-k) + \hat{P}(z;l-k)] \\ &+ \left(1 - A \sum_{k=1}^{2l-1} k^{-\alpha}\right) [z \hat{F}(z;l-1) + \hat{P}(z;l-1)]. \end{aligned} \quad (15)$$

We are now ready to address the conditional average, at least numerically. Fixing the value of  $z$ , we solve Eqs. (14) and (15) recursively, up to  $l = L$ , and use Eq. (13) to compute the average. Typical results are shown in Fig. 3. For a fixed distance  $L$  the delivery time is remarkably small throughout a wide range of long-contact exponents,  $\alpha \lesssim \alpha_*(z;L)$ , and saturates rapidly,  $T_{z,L} \sim L$ , as soon as  $\alpha$  exceeds  $\alpha_*$ . Thus, it seems that in the presence of losses small-world behavior does *not* imply a particular value of the long-contact exponent. If anything, there appears to be a small local *maximum* (barely perceptible in the figure) around a value of  $\alpha$  that creeps towards  $d$  as  $L$  increases. These results seem relevant to the Milgram experiment, where it was estimated that  $z \approx 0.75$  [17].

In summary, we have found the actual delivery time exponent in the Kleinberg navigation model, for the non small-world cases of  $\alpha \neq d$ , and we have confirmed that  $T \sim \ln^2 L$  is the actual scaling (not just a bound) for the special case of  $\alpha = d$ . We have also introduced master equations that allow one to study the Kleinberg model analytically in greater generality, and have looked briefly into the delivery time in the case of imperfect transmission, when there is a finite probability for the message to get lost along its way. The distribution of delivery times and the analytical treatment of imperfect transmission are important problems left open to further inquiry.

## Acknowledgments

We thank L. M. Glasser and O. Gromenko for many useful discussions. We are grateful to the Israel Science Foundation, to the Adams Fellowship Program of the Israel Academy of Sciences and Humanities (SC), and to the NSF, award PHY-0555312 (DbA), for partial support of this work.

- 
- [1] S. Milgram, *Psych. Today*, **2**, 60–67 (1967).
  - [2] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998); D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton University Press, Princeton, New Jersey, 1999).
  - [3] J. Kleinberg, *Nature* **406**, 845 (2000).
  - [4] J. Kleinberg, In *Proc. 32nd ACM Symp. on Theory of Computing*, 163–170 (2000).
  - [5] M. Boguñá, D. Krioukov, and K. C. Claffy, *Nature Physics*, doi:10.1038/nphys1130 (2008).
  - [6] M. A. Lomholt, T. Koren, R. Metzler, and J. Klafter, *PNAS* **105**, 11055 (2008).
  - [7] G. Oshanin, H. S. Wio, K. Lindenberg, and S. F. Burlatsky, *J. Phys.: Condens. Matter* **19**, 065142 (2007).
  - [8] O. Bénichou, M. Coppey, M. Moreau, P.-H. Suet, and R. Voituriez, *Phys. Rev. Lett.* **94**, 198101 (2005); O. Bénichou, C. Loverdo, M. Moreau, and R. Voituriez, *Phys. Rev. E* **74**, 020102(R) (2006); *J. Phys.: Condens. Matter* **19**, 065141 (2007).
  - [9] A. M. Edwards et al., *Nature* **449**, 1044 (2007).
  - [10] J.-Z. Chen, W. Liu, and J.-Y. Zhu, *Phys. Rev. E* **73**, 056111 (2006).
  - [11] M. C. Santos, G. M. Viswanathan, E. P. Raposo, M. G. E. da Luz, *Phys. Rev. E* **72**, 046143 (2005).
  - [12] L. A. Adamic and E. Adar, *Social Networks* **27**, 187–203 (2005); L. A. Adamic, R. M. Lukose, A. R. Puniyani, and

- B. A. Huberman, *Phys. Rev. E* **64**, 046135 (2001).
- [13] D. J. Watts, P. S. Dodds, and M. E. J. Newman, *Science* **296**, 1302 (2001).
- [14] M. R. Roberson and D. ben-Avraham, *Phys. Rev. E* **74**, 017101 (2006).
- [15] The range of the long contacts must be limited for  $\alpha \leq d$ , for the contact probabilities  $p_{ij}$  to be normalizable. In other words, for  $\alpha \leq d$  the problem must be posed on a *finite* lattice (of linear size  $L$ ).
- [16] We actually used  $p_{i,i+k} \propto \int_k^{k+1} m^{-\alpha} dm$ , instead of (2), for the ease of simulations. This does not affect the asymptotic behavior of the search model.
- [17] H. C. White, *Social Forces* **49**, 259 (1970).