Urban sensing as a random search process

Kevin, Brandon, Carlo ...

Here comes abstract

I. INTRODUCTION

Random search processes [1–4] are a well studied topic with a bounty of practical applications. Examples include the spreading of diseases and rumours [5], gene transcription [2], animal foraging [6–9], immune systems chasing pathogens [10], robotic exploration [11], and transport in disordered media [12, 13]. Early research on random searches focused on the simple, symmetric random walk in Euclidean spaces. Over the years many variants have been considered, such as persistent random walks [14–16], intermittent random walks [7, 17, 18], and Levy flights [19–22], which have advantageous spreading properties in certain contexts. Topologies other than Euclidean spaces, such as random graphs or real-world networks, have also been studied [23–25].

In this work, we explore a new random search process: *the taxi drive*. As the name suggests this process models the movement of taxis. The motivation for studying such a process comes from an recent (theoretical) work in urban sensing [26], in which sensors are deployed on taxis, thereby allowing air pollution, road congestion, and other urban phenomena to be monitored 'parasitically'. As such, this drive-by approach [27–30] to urban sensing can be viewed as a random search problem, in which a city's environment must be 'sensed' (i.e. searched) opportunistically by sensor-bearing taxis driving around serving passengers.

Similar to Levy walks [19, 20] or the run-and-tumble motion of bacteria [31, 32], the motion of taxis is part-random, part-regular: passenger destinations are chosen randomly, but the routes taken to those destinations are (approximately) deterministic. This mix of ballistic and random motion makes the spreading properties of taxis unusual; as shown in [26], the stationary distribution of the taxi-drive process on real-world street networks follow Zipf's law, in agreement with large, real-world taxi data from nine cities worldwide. The obeying of Zipf's law was not trivial. For example the stationary distributions of a standard random walk on the same street networks, are qualitatively different, being skewed and unimodal (see Figure X in []).

The purpose of the present work is to examine if other aspects of the taxi drive process are also atypical. Our goal is theoretical: we wish to study the taxi-drive as a stochastic processes. We compute X and find Y.

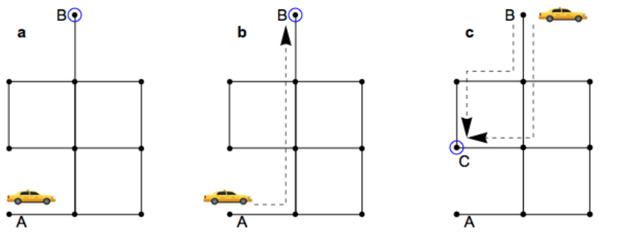

FIG. 1: **The taxi drive**. (a) A taxi picks up a passenger at node A and a destination node B, circled blue, is randomly chosen. (b) The shortest path between A and B is taken as indicated by the dashed arrow. (c) Now at B, the taxi's next pickup is at C. There are two shortest paths connecting B and C, so one is chosen at random. This process then repeats.

II. MODEL

A schematic of the taxi drive process is shown in Figure 1. As can be seen, the taxi moves on a street network S whose edges represent street segments, and whose nodes represent possible passenger pickup and dropoff locations. Having picked up a passenger at node A, the taxis move via shortest path to the destination node B, with ties between multiple shortest paths are broken at random. In order to quantitatively capture the behavior of real taxis, the destination node B is *not* chosen uniformly at random. Instead, previously visited nodes are chosen preferentially: the probability of selecting the n'th node is $q_n \propto 1 + v_n^{\beta}$, where v_n is the number of times node n has been previously visited and β is free parameter. This 'preferential return' mechanism captures the statistical properties of human mobility [33], shown in [26] also capture those of taxis.

In this work, however, we are interested in the taxi drive process from a theoretical stand point. So for simplicity's sake, we set $\beta = 0$, so that destination nodes are chosen uniformly at random.

Before showing our results, we discuss how the taxi drive is related to other models used in random searches. As stated in the introduction, the blend of random and ballistic motion makes the taxi drive similar to the levy walk. In the levy walk, agents also travel to the randomly chosen directions at constant speed, but the choosing of the destinations is different to the taxi drive: a radial distance is chosen from a heavy tailed distribution $P(r) \propto r^{-1-\alpha}$ (the direction of motion is chosen uniformly at random) as opposed to picking a random node as destination. Notice that by choosing the distance of a trip, and not the destination directly, the graph on which the levy walk is run must be embedded in a metric space (note, levy walks need

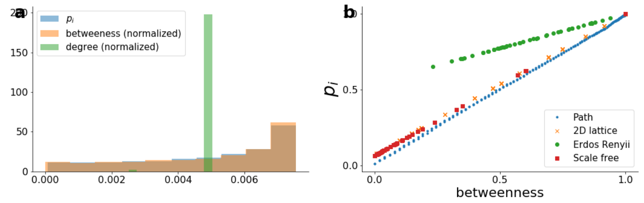

FIG. 2: **Stationary densities** (a) Distributions of p_i, betweenness and degree for a path graph. We have normalized the betweenness and degree do that their sum is 1. As can be seen, p_i agree with the betweenness, and not the degree. In panel (b) we show $p_i \propto b_i$ trend holds true for other graphs, both regular and random. Note we have rescaled by $p_i \to p_i/p_{i,max}$ and $b_i \to b_i/b_{i,max}$.

not be confined to a graph; a lot of studies consider them in continuous spaces). Since the taxi drive chooses nodes as destinations directly – that is, doesn't consider distances – it can be applied on networks with arbitrary topology.

- Are there other models of levy walk on graphs? Must check this

III. RESULTS

A. Stationary densities

We first look at the stationary densities of the taxi drive on different graph topologies. With the start with the simple path graph, in which N nodes are connected in a straight line. We ran the taxi drive until equilibration, and counted the relative number of times each node was visited p_i. Figure 2(a) shows a histogram of the p_i along with histograms of the betweenness b_i and degree d_i of the nodes in the graph. Interestingly, for the taxi drive, the stationary densities of a given nodes is proportional its betweenness $p_i \propto b_i$. This contrasts with the stationary densities of the random walk, which are proportional to the nodes degree $p_i \propto d_i$. In Figure 2(b) we show the $p_i \propto b_i$ relation holds for other graphs. The $p_i proptob_i$ relation follows from the similarities between the taxi drive and the definition of a nodes betweenness. A node's betweenness is defined by the fraction of paths that go through the node, where the total number of paths is found by considering all combinations of start and end nodes. Since in the long run, every start-end node combination will be sampled by the taxi-drive, we anticipate $p_i = b_i$.

- Note: some definitions of betweenness do NOT include the start and end node. We need to to get the $b_i \propto p_i$ relation.
- Question: is selecting the betweenness good for anything? Community detection? Foraging?.

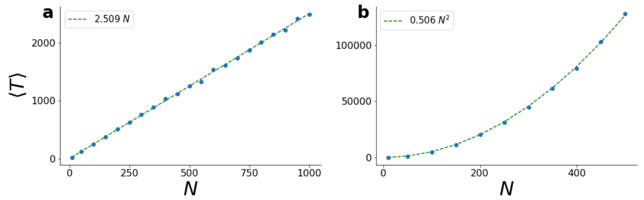

FIG. 3: **Cover times**. (a) Mean cover time on ring graph versus graph size N. Error bars shows variances. The line of best fit suggests $\langle T \rangle \sim \frac{5}{2}N$ (b) Mean cover time for path graph, showing $\langle T \rangle \sim \frac{1}{2}N^2$.

- The $p_i \propto b_i$ relation is intuitive, but a rigorously proof is lacking.

B. Cover times

A quantity commonly considered in random searches the cover time T, the time taken for each node to be visited at least once. Here we explore the cover times of the taxi process, and pose *the curious tourist problem*: A curious tourist arrives in a city. He decides to explore the city by taking taxis to randomly chosen locations. How long does it take him to cover every road at least once?

Due to the non-markovian nature of the taxi drive (the markovian property is violated since taxis move deterministically when serving passengers), we were unable to solve the curious tourist problem analytically. That said, a clean solution might be possible on simple topologies. In Figure 3 we plot the mean cover time $\langle T \rangle$ on the ring graph, and the path graph (the path graph is a straight line of nodes) versus graph size N. Lines of best fit indicate that $\langle T \rangle_{ring} = 2.509N$ and that $\langle T \rangle_{path} = 0.506N^2$. Since the coefficients are close to simple fractions, we conjecture

$$\langle T \rangle_{ring} = \frac{5}{2}N \tag{1}$$

$$\langle T \rangle_{path} = \frac{1}{2}N^2. \tag{2}$$

- Am I sure of this claim? I must collect more numerical data, and do more stats on the parameters of best fit.
- Maybe I should say $\langle T \rangle \sim 0.5N^2$, i.e. make a claim about asymptotics.

Beyond simple topologies however, it is unlikely an analytic solution to the curious tourist problem exists. Indeed, even for the simple, symmetric random walk, exact results for cover times are rare (see [34] for a review)

Does the taxi-drive offer any advantages over the random walk? The answers depends on topology, as shown

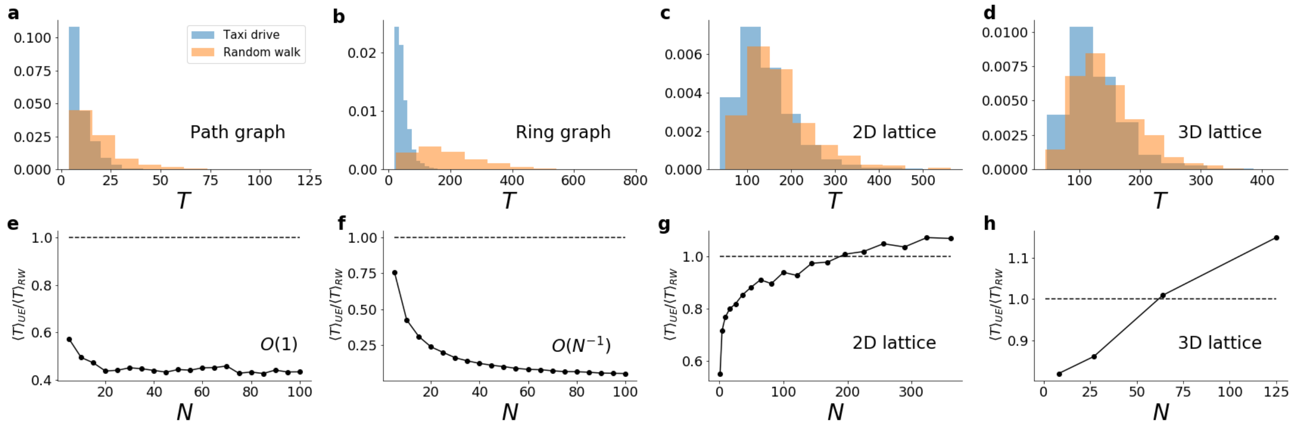

FIG. 4: **Cover times of regular graphs**. Top row: distribution of cover times of urban explorers and random walkers on graphs with regular topology. Comment. Bottom row: ratio of mean cover time of urban explorer to that of random walkers. The dashed line indicated the value 1, which indicates the point at which urban explorers cover the graphs more efficiently (on average) that random walker. Comment.

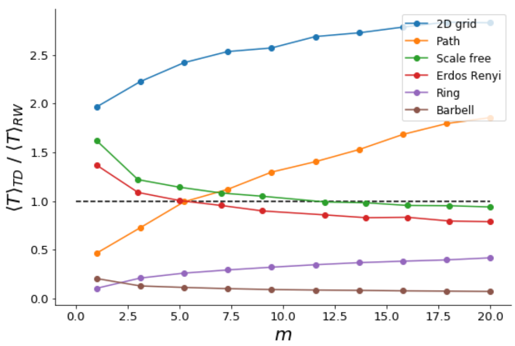

FIG. 5: ***m*-cover times**. Some text

in Figure 4. Distributions of the cover times are in the top row, while in the bottom the ratio $\langle T\rangle_{TD}/\langle T\rangle_{RW}$ is shown. For the ring and graph, the taxi drive is more efficient for all graph sizes N. But for lattices in 2D and 3D, there is a critical N^* beyond which $\langle T\rangle_{RW} < \langle T\rangle_{TD}$.

m-cover times. In some settings, it is necessary to cover each node more than once. For example, in urban sensing, more than one sample per node is necessary to capture the temporal fluctuations of certain quantities. This leads us to consider the $m-$ cover time, the time it takes to cover every node at least m times. We show the plots of this in Figure 5. Interesting trends are evident.

When $m = 1$, the random walker is more efficient for all graphs but the barbell. This is to be expected, since it is known that random walk cover times are maximized by the barbell topology (the rationale here is that the walker gets stuck in the bells; the ballistic aspect of the taxi drive movements insulates it from this trapping). When $m > 1$ however there is a crossover in the behavior on the Erdos-Renyi and scale-free graphs, not yet understood theoretically.

- Open question: what graph maximizes the cover-time of the taxi drive?

Discussion

- We have found a new spreading process
- State why different to levy walk. (Non-euclidean)
- They have these properties.
- They are useful in X
- Future work could do X. Mean first passage times. Mixing times.
- Open question, what graph maximises the cover time of the urban explorer?
- Curious tourist problem

[1] S. Condamin, O. Bénichou, V. Tejedor, R. Voituriez, and J. Klafter, Nature **450**, 77 (2007).
[2] O. Bénichou and R. Voituriez, Physics Reports **539**, 225 (2014).
[3] S. Redner, *A guide to first-passage processes* (Cambridge University Press, 2001).
[4] V. Méndez, D. Campos, and F. Bartumeus, in *Stochastic foundations in movement ecology* (Springer, 2014), pp. 177–205.
[5] A. L. Lloyd and R. M. May, Science **292**, 1316 (2001).

[6] G. M. Viswanathan, M. G. Da Luz, E. P. Raposo, and H. E. Stanley, *The physics of foraging: an introduction to random searches and biological encounters* (Cambridge University Press, 2011).
[7] O. Bénichou, C. Loverdo, M. Moreau, and R. Voituriez, Reviews of Modern Physics **83**, 81 (2011).
[8] G. M. Viswanathan, V. Afanasyev, S. Buldyrev, E. Murphy, P. Prince, and H. E. Stanley, Nature **381**, 413 (1996).
[9] F. Bartumeus, M. G. E. da Luz, G. M. Viswanathan, and J. Catalan, Ecology **86**, 3078 (2005).
[10] M. L. Heuzé, P. Vargas, M. Chabaud, M. Le Berre, Y.-J. Liu, O. Collin, P. Solanes, R. Voituriez, M. Piel, and A.-M. Lennon-Duménil, Immunological reviews **256**, 240 (2013).
[11] M. Vergassola, E. Villermaux, and B. I. Shraiman, Nature **445**, 406 (2007).
[12] S. Havlin and D. Ben-Avraham, Advances in Physics **36**, 695 (1987).
[13] D. Ben-Avraham and S. Havlin, *Diffusion and reactions in fractals and disordered systems* (Cambridge university press, 2000).
[14] V. Tejedor, R. Voituriez, and O. Bénichou, Physical review letters **108**, 088103 (2012).
[15] P. Cénac, A. Le Ny, B. de Loynes, and Y. Offret, Journal of Theoretical Probability **31**, 232 (2018).
[16] L. Basnarkov, M. Mirchev, and L. Kocarev, in *International Conference on ICT Innovations* (Springer, 2017), pp. 102–111.
[17] G. Oshanin, H. Wio, K. Lindenberg, and S. Burlatsky, Journal of Physics: Condensed Matter **19**, 065142 (2007).
[18] M. Azaïs, S. Blanco, R. Bon, R. Fournier, M.-H. Pillot, and J. Gautrais, PLOS ONE **13**, e0206817 (2018).
[19] M. F. Shlesinger, J. Klafter, and B. J. West, Physica A: Statistical Mechanics and its Applications **140**, 212 (1986).
[20] A. Blumen, G. Zumofen, and J. Klafter, Physical Review A **40**, 3964 (1989).
[21] G. Viswanathan, V. Afanasyev, S. V. Buldyrev, S. Havlin, M. Da Luz, E. Raposo, and H. E. Stanley, Physica A: Statistical Mechanics and its Applications **282**, 1 (2000).
[22] M. A. Lomholt, K. Tal, R. Metzler, and K. Joseph, Proceedings of the National Academy of Sciences (2008).
[23] J. D. Noh and H. Rieger, Physical review letters **92**, 118701 (2004).
[24] T. Weng, J. Zhang, M. Small, B. Harandizadeh, and P. Hui, Physical Review E **97**, 032320 (2018).
[25] N. Masuda, M. A. Porter, and R. Lambiotte, Physics reports (2017).
[26] K. O'Keeffe, A. A, P. Santi, S. H. Strogatz, and C. Ratti, x **X**, X (2018).
[27] U. Lee and M. Gerla, Computer Networks **54**, 4 (2010).
[28] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden, in *Proceedings of the 4th international conference on Embedded networked sensor systems* (2006), p. 125138.
[29] P. Mohan, V. N. Padmanabhan, and R. Ramjee, in *Proceedings of the 6th ACM conference on Embedded network sensor systems* (2008), p. 323336.
[30] A. Anjomshoaa, F. Duarte, D. Rennings, T. Matarazzo, P. de Souza, and C. Ratti, IEEE Internet of Things Journal pp. 1–1 (2018).
[31] M. J. Schnitzer, Physical Review E **48**, 2553 (1993).
[32] H. C. Berg, *E. coli in Motion* (Springer Science & Business Media, 2008).
[33] C. Song, T. Koren, P. Wang, and A.-L. Barabási, Nature Physics **6**, 818 (2010).
[34] M. Abdullah, arXiv preprint arXiv:1202.5569 (2012).