The Nearest Unvisited Vertex Walk on Random Graphs

David Aldous

2 February 2020

My theme phrase is `hide in plain sight`.

- Two 1970s results, not-quite-obviously related, in a deterministic setting.
- Not at all difficult – one page proofs – so people moved on.
- From the proofs one can extract a "key observation" that is useful in the random setting,
- The content in this talk is mostly quite elementary, but it suggests intriguing questions in the random setting that no-one seems to have considered before.

Consider a connected undirected graph G on n vertices, where the edges e have positive real lengths $\ell(e)$. Imagine a robot that can move at speed 1 along edges. We need a rule for how the robot chooses which edge to take after reaching a vertex. Most familiar is the "random walk" rule, choose edge e with probability proportional to $\ell(e)$ or $1/\ell(e)$. One well-studied aspect of the random walk is the *cover time*, the time until every vertex has been visited.

Instead of the usual random walk model, let us consider the *nearest unvisited vertex* (NUV) walk

> *after arriving at a vertex, next move at speed* 1 *along the path to the closest unvisited vertex*

and continue until every vertex has been visited. Note this is deterministic and has some length (= time) $L_{NUV}(G, v_0)$.

Of course *distance* $d(v, v')$ is shortest path length. In informal discussion we imagine lengths are scaled so that distance to closest neighbor is order 1, so L_{NUV} must be at least order n.

Natural first question: when is it $O(n)$ rather than larger order?

The old algorithms literature discussed the NUV walk as heuristics for TSP or robot algorithms, but then quickly moved on to better algorithms, and the key starting math observation below is rather obscured.

My preprint on arXiv gives citations.

Consider **ball-covering**: for $r > 0$ define $N(r) = N(G, r)$ to be the minimal size of a set S of vertices such that every vertex is within distance r from some element of S. In other words, such that the union over $s \in S$ of $\mathrm{Ball}(s, r)$ covers the entire graph.

Proposition

(i) $N(r) \le 1 + L_{NUV}/r, \ 0 < r < \infty$*.*
(ii) $L_{NUV} \le 2 \int_0^{\Delta/2} N(r) \, dr$ *where* $\Delta = \max_{v,w} d(v, w)$ *is the diameter of the graph.*

Note that for continuous spaces, *metric entropy* implies a notion of *dimension* via $N(r) \approx r^{-\mathrm{dim}}$ as $r \downarrow 0$. In our discrete context, if we have *dimension* in the sense

$$N(r) \approx n r^{-\mathrm{dim}}, \quad 1 \ll r \ll \Delta$$

then the Proposition has informal interpretation that L_{NUV} is always $O(n)$ when $\mathrm{dim} > 1$.

Proposition

(i) $N(r) \leq 1 + L_{NUV}/r, \ 0 < r < \infty.$
(ii) $L_{NUV} \leq 2 \int_0^{\Delta/2} N(r) \, dr$ *where* $\Delta = \max_{v,w} d(v,w)$ *is the diameter of the graph.*

Part (i) actually holds for the length L of **any** walk that visits every vertex. Just pick vertices $s_0, s_1, s_2, \ldots$ such that s_{i+1} is the first vertex visited at time later than r after the visit to s_i.

Part (ii) has a proof by picture. What does the NUV walk look like inside a ball?

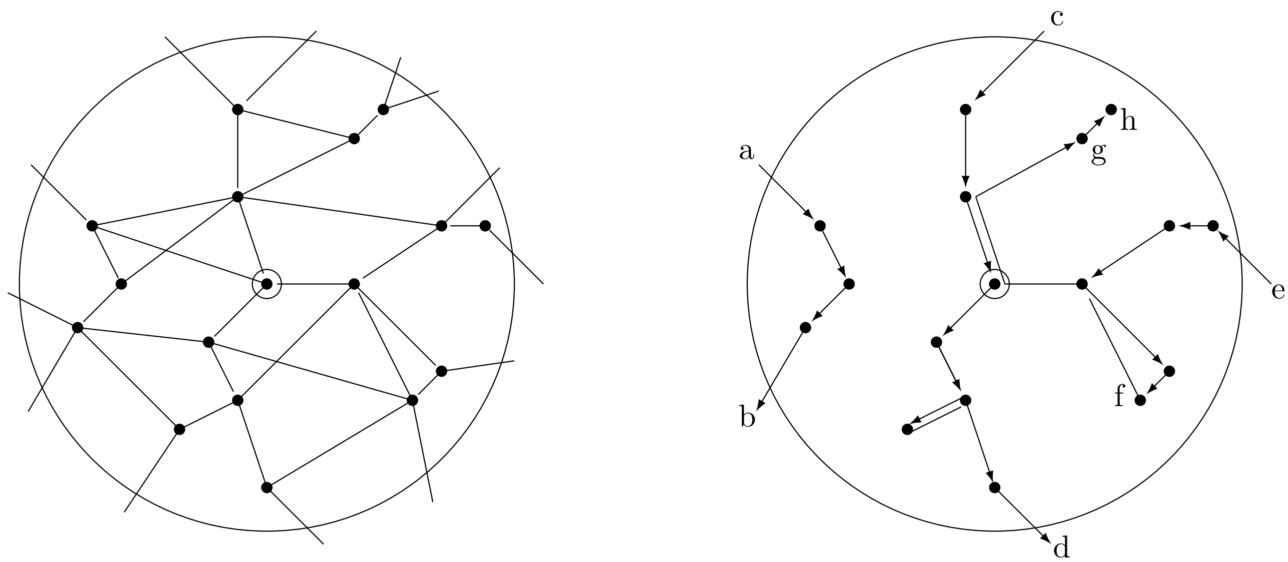


Figure 1: Illustration of Lemma 2. The left panel shows the subgraph within a radius-r ball. The NUV walk must consist of one or several excursions within the ball. These excursions depend on the configuration outside the ball, and the right side shows one possibility. The first excursion enters via edge a and exits via edge b. The second excursion enters via edge c and exits via edge d, en route backtracking across one edge. The third excursion enters via edge e and proceeds to vertex f; at that time only vertices g, h within the ball are unvisited, and the next step of the walk is a path going via three previously-visited vertices to reach g and then h. The next step from h, not shown, might be very long, depending on whether nearby vertices outside the ball have all been visited.

The picture shows

Lemma

Fix a vertex v^* and a real $r > 0$. For all $v \in \mathrm{Ball}(v^*, r)$ except perhaps the last-visited, the NUV step away from the first visit to v has length $\leq 2r$.

So the number of steps of length $\geq 2r$ is at most $N(r)$, and the bound (ii) easily follows:

$$L_{NUV} \leq 2 \int_0^{\Delta/2} N(r) \, dr.$$

The two "classical" results about NUV walks are simple Corollaries of the Proposition. First, the argument for (i) shows

$$N(r) \le 1 + L_{TSP}/r, \ 0 < r < \infty$$

where L_{TSP} is length of shortest covering walk. So do the integral in (ii)

$$L_{NUV} \le 2 \int_0^{\Delta/2} N(r) \, dr$$

and note $N(r) \le n$ and $\Delta \le L_{TSP}$.

Corollary

Let $a(n)$ be the maximum, over all connected n-vertex graphs with edge lengths and all initial vertices, of the ratio L_{NUV}/L_{TSP}. Then $a(n) = O(\log n)$.

Several authors have given examples where ratio is order $\log n$, but very artificial.

Here is the second classical result.

Corollary

There is a constant A such that, for the complete graph on n arbitrary points in the area-n square, with Euclidean lengths,

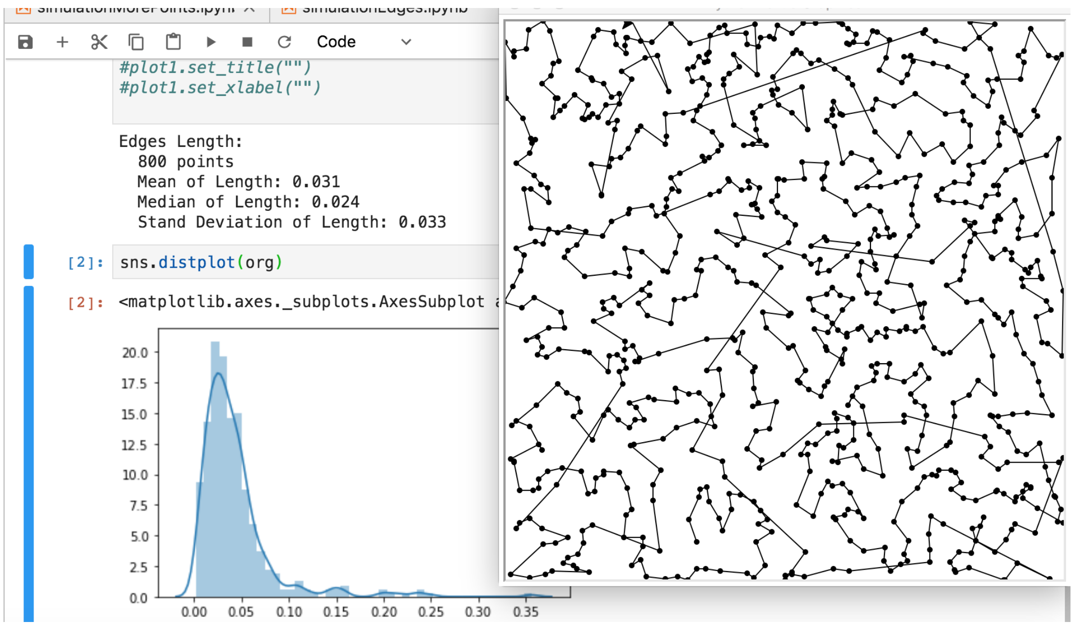
$$L_{NUV} \leq An.$$

Note this implies the well known corresponding result $L_{TSP} \leq An$.

Proof. By continuum ball-covering there is a numerical constant C such that $N(r) \leq Cn/r^2$, and so (ii) gives

$$L_{NUV} \leq 2 \int_0^{\sqrt{n/2}} \min(n, Cn/r^2) \ dr \leq 4C^{1/2} n.$$

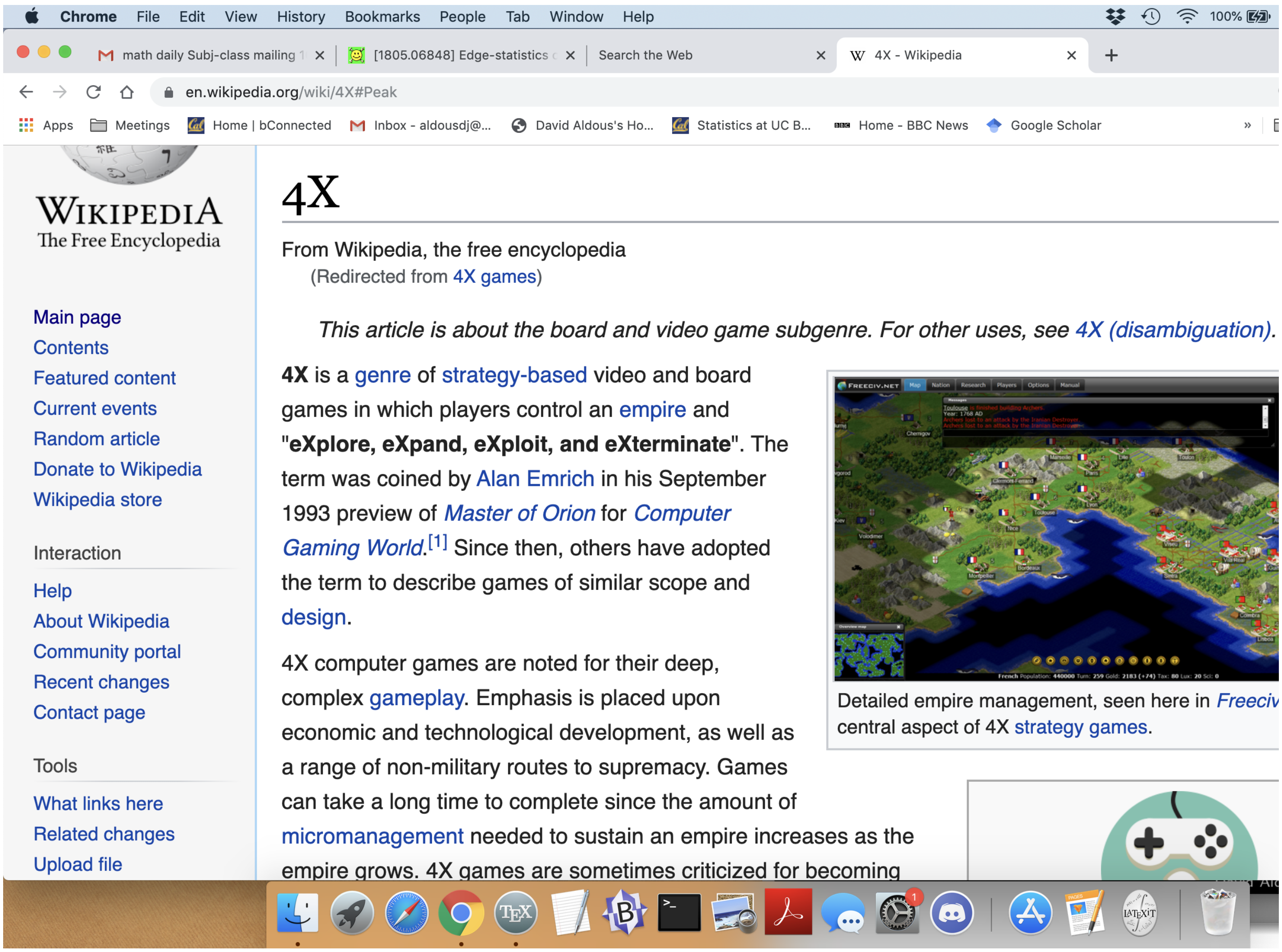
NUV walk on 800 random points in the square.
Simulation by Yechen Wang.



I have told you everything that's known in the deterministic setting.
Here's an open problem.

Is $\frac{\max_v L_{NUV}(G,v)}{\min_v L_{NUV}(G,v)}$ bounded over all finite graphs G?

So why did I myself get interested in this model, recently?



4X

From Wikipedia, the free encyclopedia

(Redirected from 4X games)

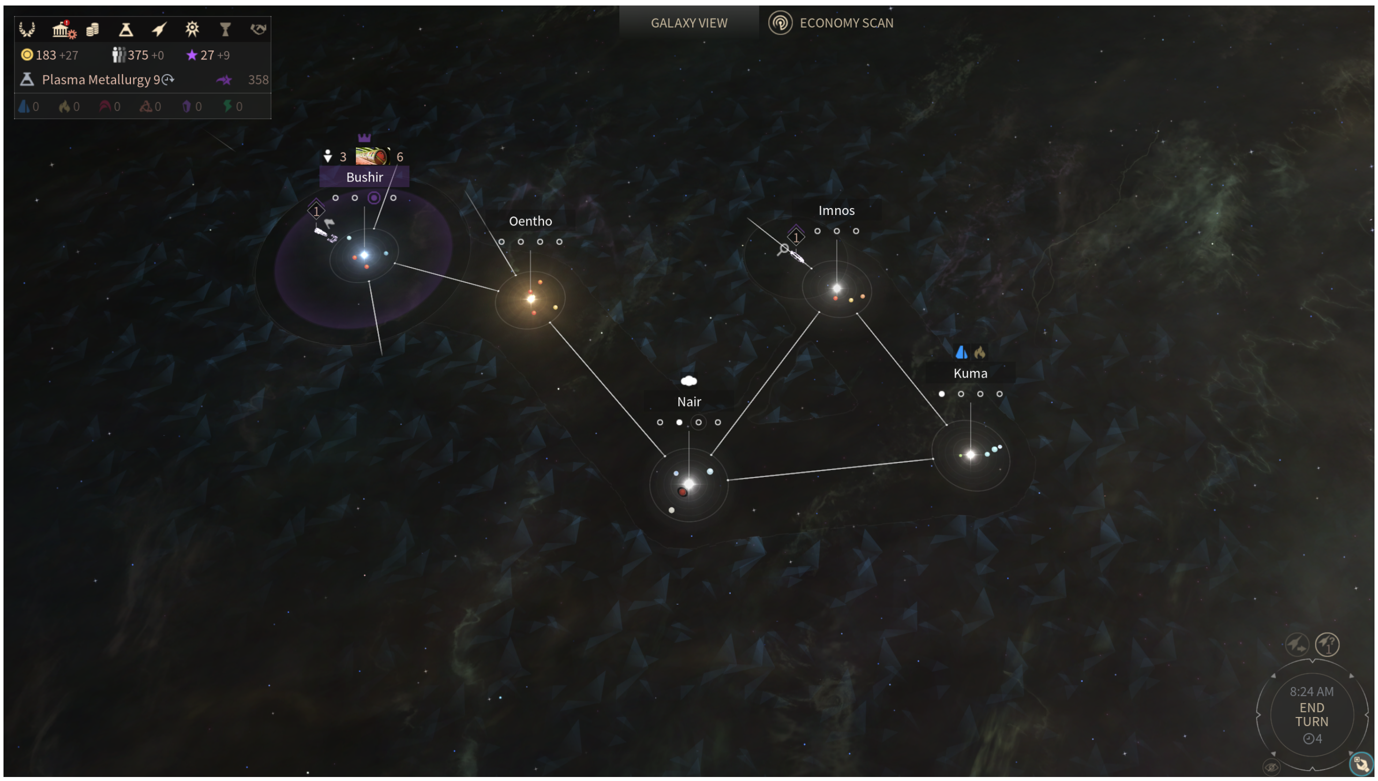
This article is about the board and video game subgenre. For other uses, see 4X (disambiguation).

4X is a genre of strategy-based video and board games in which players control an empire and "**eXplore, eXpand, eXploit, and eXterminate**". The term was coined by Alan Emrich in his September 1993 preview of *Master of Orion* for *Computer Gaming World*.[1] Since then, others have adopted the term to describe games of similar scope and design.

4X computer games are noted for their deep, complex gameplay. Emphasis is placed upon economic and technological development, as well as a range of non-military routes to supremacy. Games can take a long time to complete since the amount of micromanagement needed to sustain an empire increases as the empire grows. 4X games are sometimes criticized for becoming

Detailed empire management, seen here in *Freeciv*, central aspect of 4X strategy games.


GALAXY VIEW
THE MEANS OF PRODUCTION ONGOING
In 30 turns, own the System producing the most Industry in the galaxy.
END TURN
Idrus
Esh
HD 44179
Aeri
Hydrus
Caelum
Orion
Lyra
RNGC 43278
Bundus
Nair
Giazar
Wordan
Lupus
Heracles
Saiph
Alioth
Amosa
Centaurus
Phaeton
Oshyi
Ingris
Zuben
Nusakan
Jyem
Kuma
Ying
IRAS 48796+25.
3C 75
Baten
Octans
Cran
Serte
Lonica
Oiad
Artemis
RFGC 88-43
Imex
Cetus
Asellus
Qorum
Cygnus
Arkab
Veran
Echius
Phad
Rasam
GNY 8
Olvaldi
Sathrus
Grus
Traitor's Reach
NGC 205
Delphinus
Wyrix
IRAS 48838+23.
Vorta
Volans
Temos
Ita
Libra


GALAXY VIEW
ECONOMY SCAN
183 +27
375 +0
27 +9
Plasma Metallurgy 9
358
Bushir
Oentho
Imnos
Kuma
Nair
8:24 AM
END
TURN

The ball-covering relation is not helpful from the algorithms viewpoint. But it is useful for some random graph models. In particular, in a model where we take a unweighted graph and then assign random edge-lengths, understanding "balls" is precisely the basic issue in **first passage percolation**.

Consider the random graph G_m that is the $m \times m$ grid, that is the subgraph of the Euclidean lattice $\mathbb{Z}^2$, assigned i.i.d. edge-lengths $\ell(e) > 0$. Because the shortest edge-length at a given vertex is $\Omega(1)$, clearly L_{NUV} is $\Omega(m^2)$.

Corollary

For the 2-dimensional grid model G_m above, the sequence $(m^{-2} L_{NUV}(G_m),\ m \geq 2)$ is tight.

We strongly believe that in fact $m^{-2} L_{NUV}(G_m)$ converges in probability to a constant, but we do not see any simple argument.

Outline proof.

For a vertex v of G_m write $B(v, r)$ for the random set of vertices v' with graph distance $d(v, v') \le r$, and write $D(v, r)$ for the non-random set of vertices v' with Euclidean distance $||v - v'|| \le r$. Standard results for FPP on $\mathbb{Z}^2$ imply that there exist constants c_1, c_2, c_3 (depending on the distribution of $\ell(e)$) such that

$$\mathbb{P}(D(v, r) \not\subseteq B(v, c_1 r)) \le c_2 \exp(-c_3 r), \ 0 < r < \infty.$$

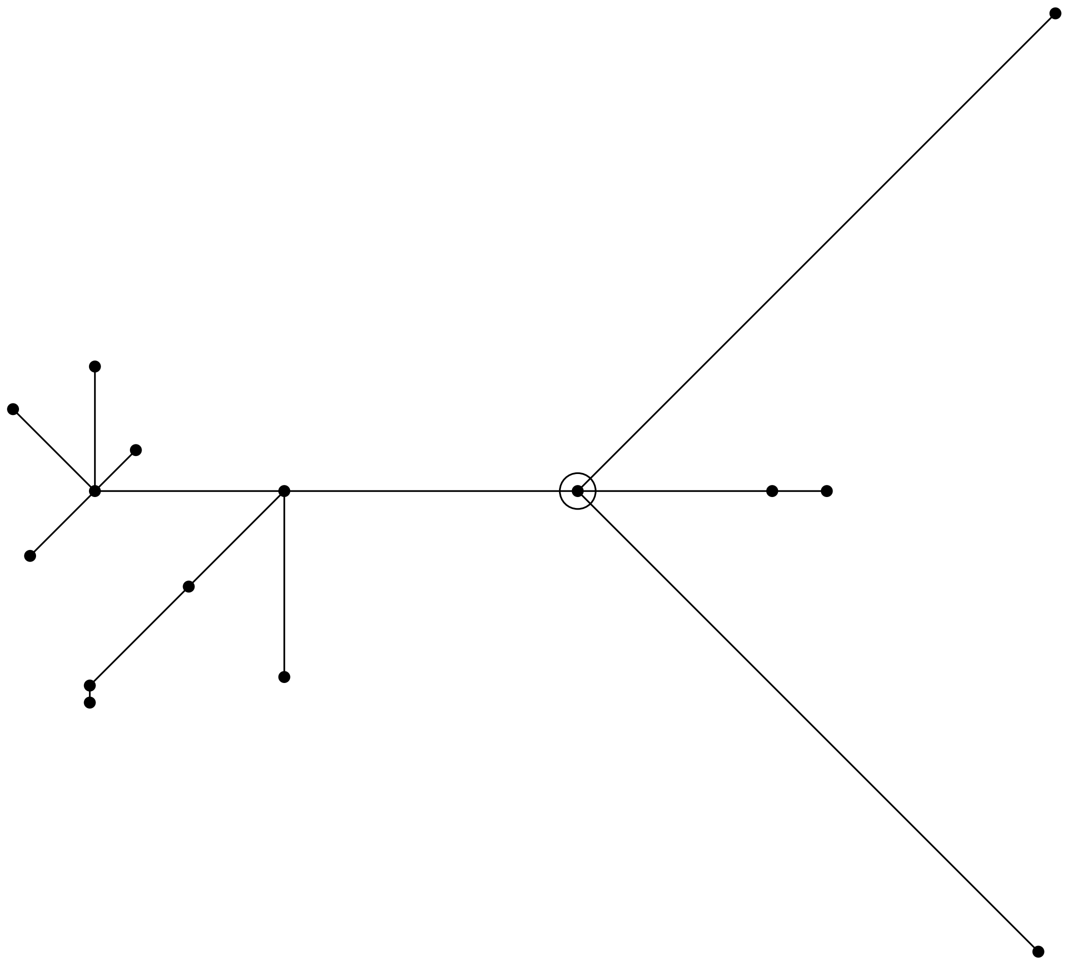
Now we simply use Euclidean ball-covering.

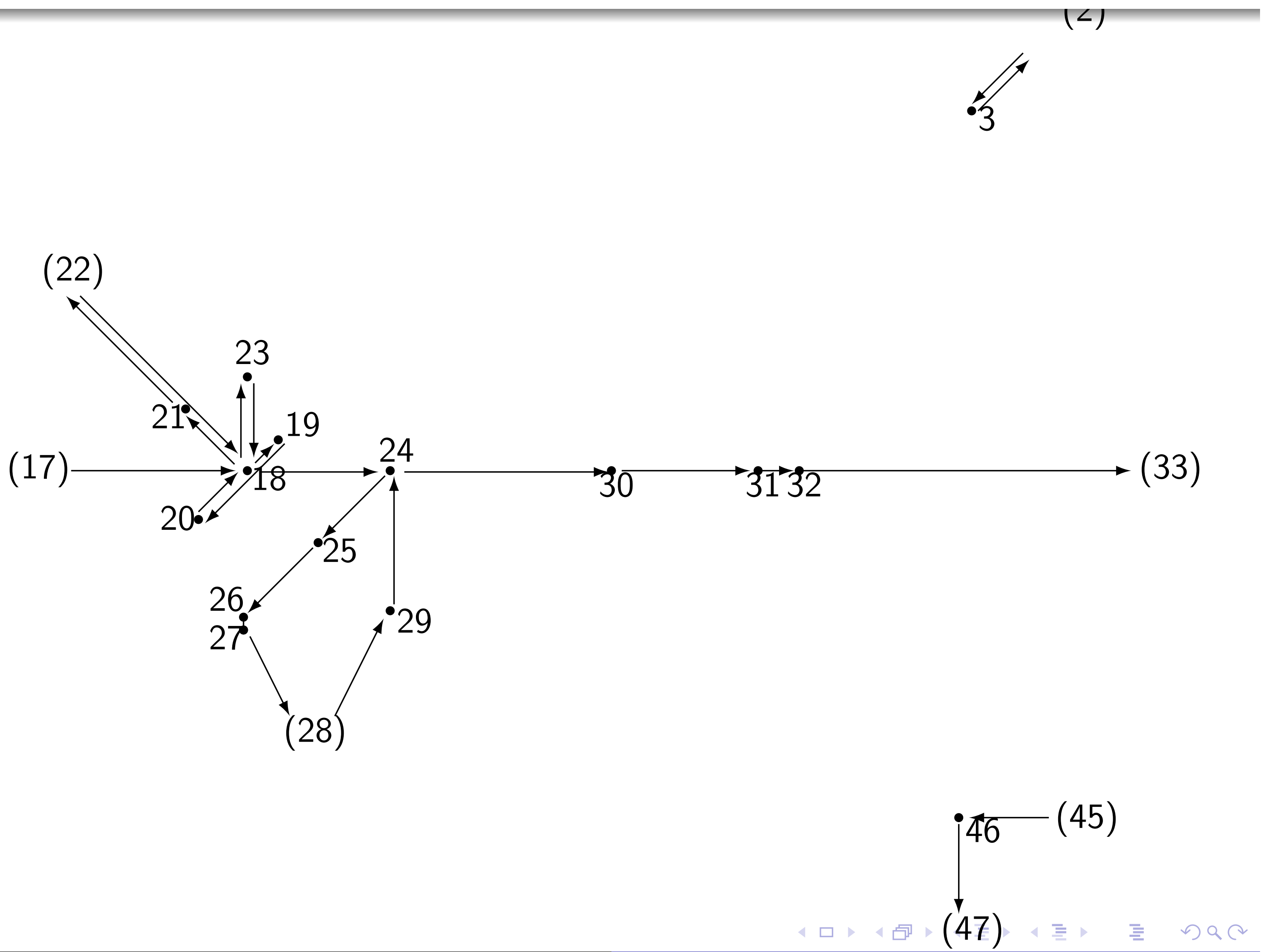
The mean-field model of distance
Take the complete graph on n vertices and assign to edges i.i.d. random weights with Exponential (mean n) lengths. This "mean-field model of distance" G_n turns out to be surprisingly tractable, because the smallest edge-lengths at a given vertex are distributed (in the $n \to \infty$ limit) as the points of a rate-1 Poisson point process on $(0, \infty)$, and as regards short edges the graph is locally tree-like. A now classical result of Frieze proves that the length of the MST is asymptotically $\zeta(3)n$, and a remarkable result of Wästlund formalizing ideas of Mézard - Parisi shows that the length of the TSP path is asymptotically cn for an explicit constant $c = 2.04....$.

Might it be possible to get a similar explicit result for the NUV length? We get the correct order of magnitude by essentially the same method as above.

Corollary

For the mean-field model of distance G_n , the sequence $(n^{-1}L_{NUV}(G_n),\ n \geq 2)$ is tight.




(2)
3
(22)
23
21
19
(17)
18
24
30
31
32
(33)
20
25
26
27
29
(28)
(45)
46
(47)

Outline proof.

In this model the $n \to \infty$ limit of the sizes of balls, that is for fixed v the process

$$(|B_n(v,r)|, \ 0 < r < \infty)$$

is the Yule process, and for $r < \frac{1}{2} \log n$ this size distribution is approximately Exponential, mean e^r (birthday problem). So we can cover the graph using around ne^{-r} balls (centered randomly).

Summary

- We have merely observed that one can apply the key Proposition to get the correct $\Theta(n)$ order of magnitude of L_{NUV} in familiar models of random connected graphs.
- A natural next question: under what conditions can one show that the variance is also $\Theta(n)$?
- One "structural" property of the NUV walk is that, if two vertices are each other's nearest neighbor, then every (over starting vertex) NUV walk uses the linking edge. This suggests a very general $\Omega(n)$ lower bound for variance, from randomness of such edge-lengths. But our small-scale simulations suggest slightly sub-linear.
- In other contexts the "locally tree-like" property of models like the mean-field model of distance allows calculations, but not so clear for NUV walks.