Imperial College London

On Geodesics in Directed Acyclic Graphs

S. Moon and S. Sanei

Supervisor: Dr Tim Evans, Theoretical Physics Group

Aims

1. Investigate the relationship between different paths & the geodesic for a Directed Acyclic Graph.
2. Devise measures for quantifying how different paths may be taken as a geodesic.

Directed Acyclic Graphs (DAGs)

Graph – A group of points or **Nodes**, with some pairs of them being connected with **Edges.**

Directed – Edges point in one direction only.

Acyclic – There are no **Cycles**, paths that back onto themselves.

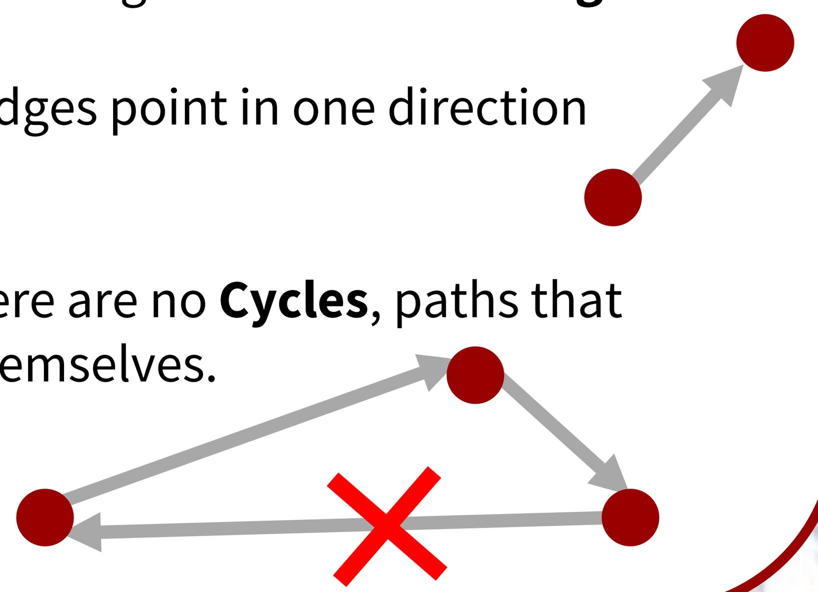
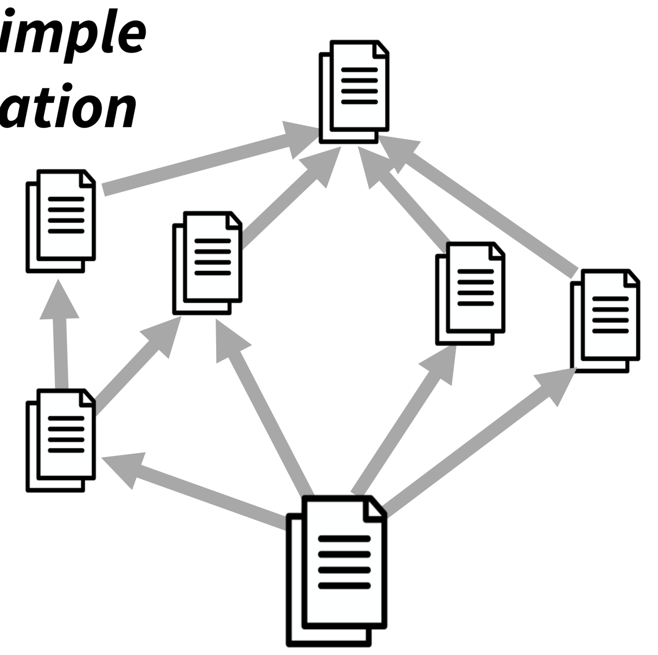


Figure 1: A simple DAG as a Citation Network



Geodesic

A **Geodesic** is a curve that something would move along without any forces to push it off course.

It takes the **shortest step** (or sometimes the **longest step**) along every part of the path.

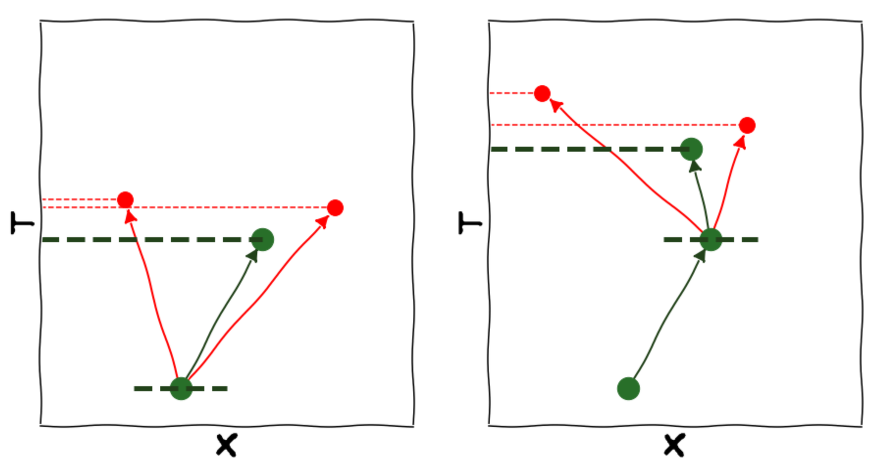
This definition works in continuous spaces.

But there is **NO** good definition of geodesics in graphs *yet* [1].

Paths

- **Random –** Path that picks the next node randomly out of what it can reach. This is our control.
- **Longest –** Path that goes through the most nodes. This scales as the $\sqrt[D]{N}$, for a D-dimensional and N-node graph [2]. ***In continuous Minkowski space, the geodesic is the longest path.***
- **Greedy Paths –** Paths that maximise or minimise some "decision" at each node as they move up the graph. ***This looks similar to the formal definition of a geodesic.***

Figure 2: Process for a Greedy Path



Our Graphs

For **N-node graphs:**

1. **Sprinkle nodes randomly** in a unit square.
2. **Add** a **Source** node at the origin, and a **Sink** at (1,1).
3. **Rotate** the square into a diamond. The new co-ordinate system is **time** on the vertical, and **space** on the horizontal.

This gives us a graph in a unit diamond of **1+1 Minkowski Spacetime.**

Method

1. Make the 1+1 Minkowski spacetime Graph.
2. Add in the edges. There is an edge from nodes *a* to *b* only if *b* is in the **Causal Future** of *a*.
3. Find the **Longest**, **Random**, and **Time-Greedy Path**. Save these.
4. Repeat 1000 times.

For every path, find and plot the *Protrusion.*

Protrusion Δx

Protrusion is the maximum spatial deviation from the geodesic and **has not been studied before.**

This tells us how far the path gets from the straight-line geodesic in the continuous limit.

If a path is a good approximate to a geodesic, we'd expect this value to tend towards 0 as the number of nodes N increases.

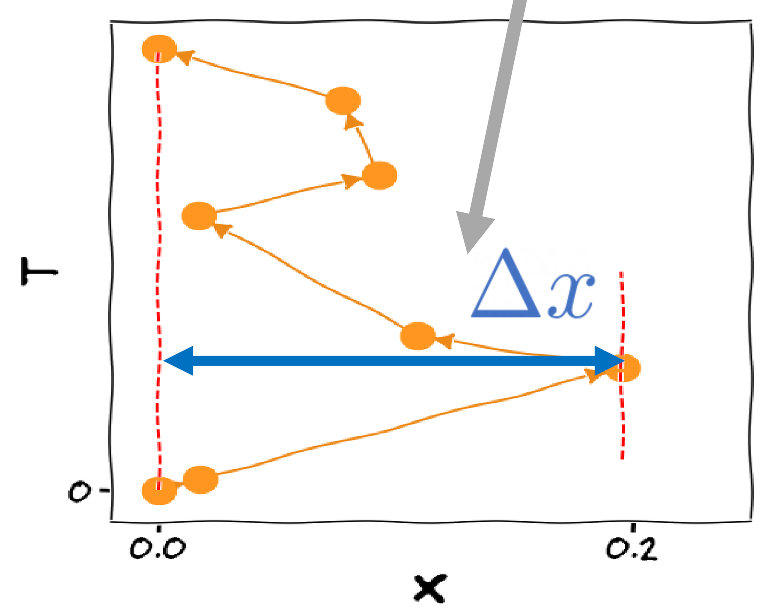
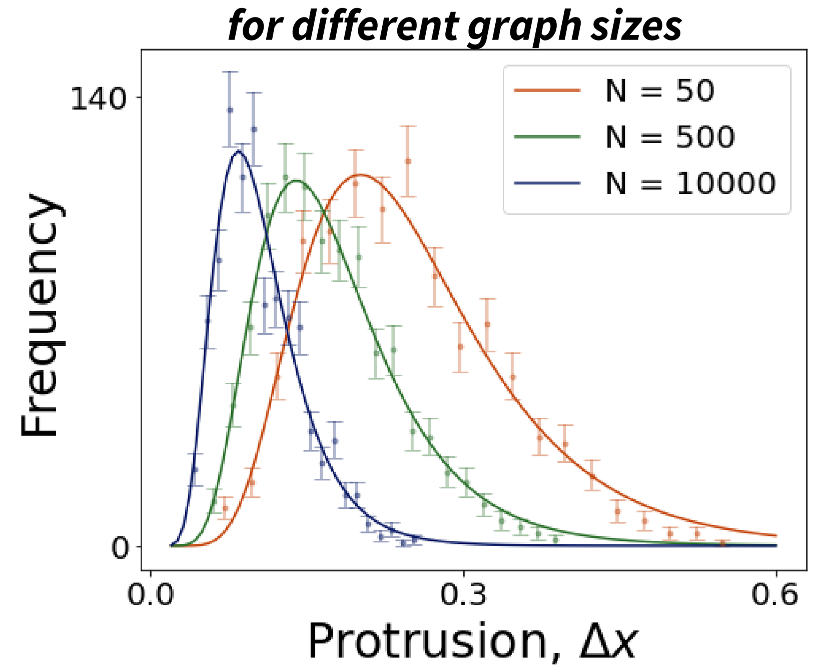


Figure 4: Protrusion illustrated as Δx

Results

Figure 5: Longest Path Protrusions for different graph sizes



As graph size increases, the peak tends towards zero.

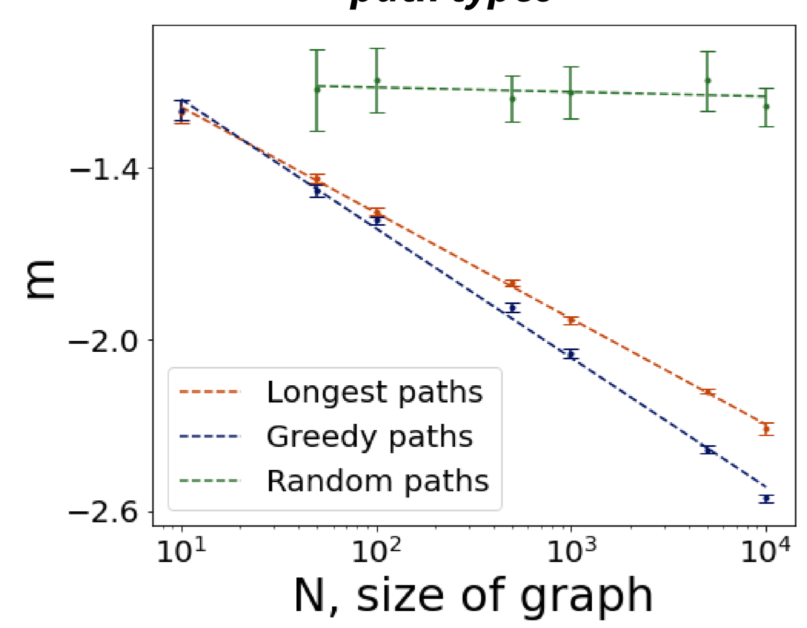
By fitting a lognormal distribution, we can quantify how the peak tends towards zero. *m* is $\ln(median)$.

Lognormal fitting:

$$\frac{A}{\Delta x s\sqrt{2\pi}} exp\left(-\frac{(ln(\Delta x) - m)^2}{2s^2}\right)$$

***M*-fitting:** $m = \alpha log(N) + \beta$

Figure 6: m-scaling for different path types



Longest and greedy path protrusion scales but random doesn't.

Path	α
Longest	-0.371(1)
Greedy	-0.45(1)
Random	-0.02(2)

Greedy path protrusion tends to zero **faster** than longest path!

Conclusions

- The Greedy and the Longest Paths both have protrusions that tend to 0 as N increases. **BOTH are good candidates for geodesics.**
- Time-Greedy paths tend towards the geodesic **faster** than Longest paths.

Future Research

- **More Graph Measures –** Jaggedness, Wonkiness, and Deviation.
- **Different Networks –** Higher Dimensions & Different Spaces.
- **Applications –** Causal Sets, Network Analysis, and more…

References

[1] M. van Steen, "*Graph Theory and Complex Networks*". self-published, 2010.

[2] B. Bollobás & G. Brightwel (1991) "*Box-Spaces and Random Partial Orders*". Transactions of the American Mathematical Society 324, Vol. 1, pp 59-72.