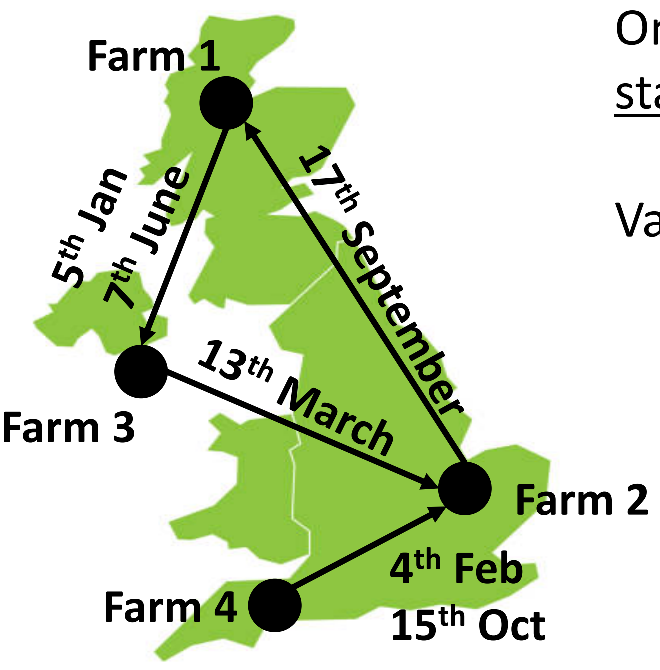
Sampling temporal networks inspired by livestock movement

Ciara Losel and Dr. Jessica Enright, University of Glasgow

Movements like cattle trades can be represented by temporal networks, which can give us key information about important holdings. However, these networks are often extremely large, making them hard to analyse.



On the left is an example of a temporal network. It has nodes (farms) and edges (cattle movements) which have time stamps showing the date of specific movements. There are approximately **216,000 farms in the UK**.

Various metrics allow us to characterise temporal networks e.g., temporal degree centrality and reachability ratio.

Temporal degree centrality	Reachability ratio
The total number of incoming and outgoing edges a node is adjacent to over all time points.	The proportion of other nodes a node can reach within a certain time frame.

We want to come up with sampling methods that make calculating these metrics more computationally feasible.

QUESTION: How many nodes/edges do we need to sample to accurately estimate these parameters and what sampling method should we use?

Consider the following different topologies and time distributions we can use to generate different models of temporal networks.

Topology	Time distribution
Erdos-Renyi network: Each contact between two nodes is present with equal specified probability	**Random time stamps:** time stamps across network edges have uniform distribution
Geometric random network: nodes and edges placed geometrically e.g. cattle trade network	**Bursty time stamps:** Edge activity tends to happen in concentrated time periods

Combining these gives 4 different temporal networks to sample from

+

Random walk sampling | Depth-first-search sampling | Breadth-first search-sampling | Forest fire sampling

4 different sampling methods

=

16 different samples

To find out **how big** the different samples need to be get accurate representations of the networks…

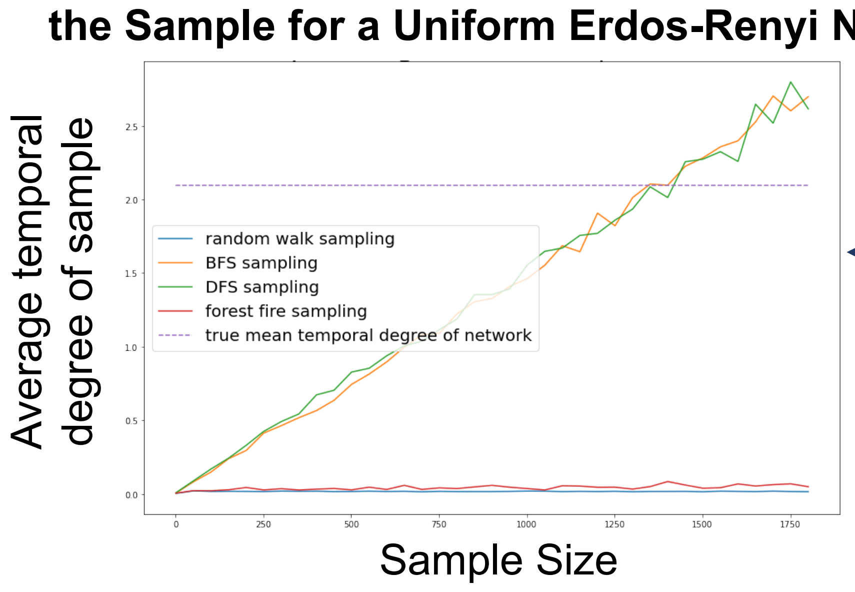
Get **different sized samples** of each temporal network and **compare metrics** of samples with actual values

To find out **which sampling methods work best** for each of the networks…

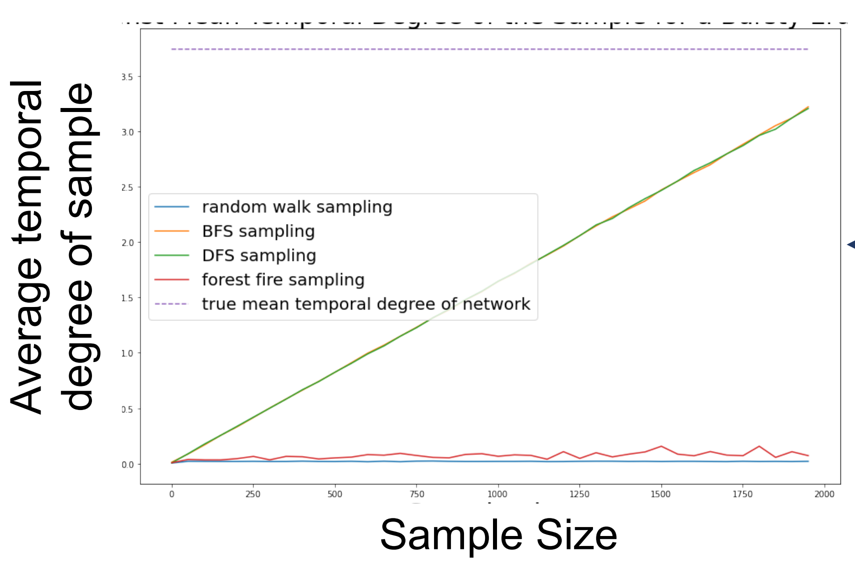
Get **lots of equal-sized samples** of each temporal network and **compare metrics** of samples with actual values

ANSWER:

Sample Size Against Average Temporal Degree of the Sample for a Uniform Erdos-Renyi Network



Sample Size Against Average Temporal Degree of the Sample for a Bursty Erdos-Renyi Network



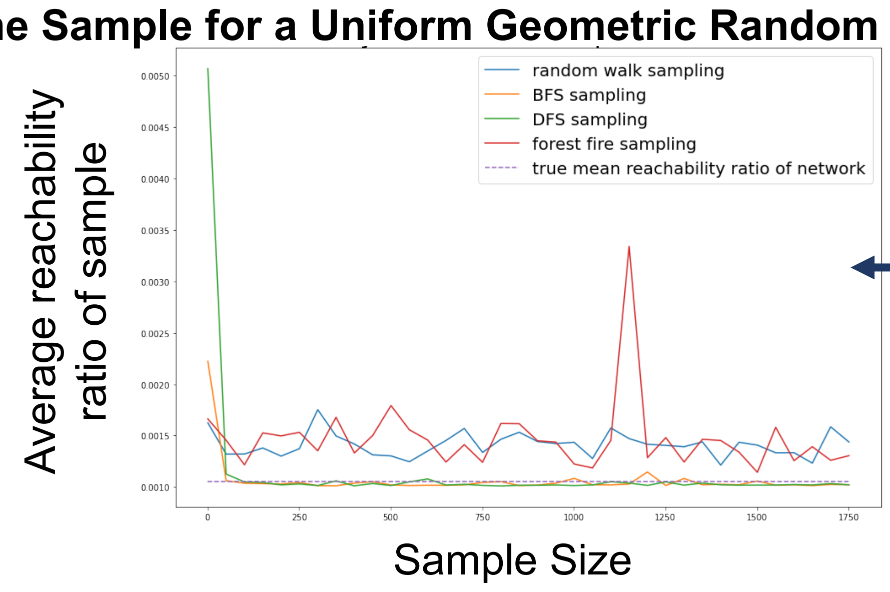
For an Erdös-Renyi network with uniform time stamps…

- BFS and DFS sampling get accurate temporal degree when sample size is around 1500
- Random walk and forest fire sampling not good for estimating degree

For an Erdös-Renyi network with bursty time stamps…

- BFS and DFS need larger sample sizes to predict temporal degree
- Random walk and forest fire sampling not good for estimating degree

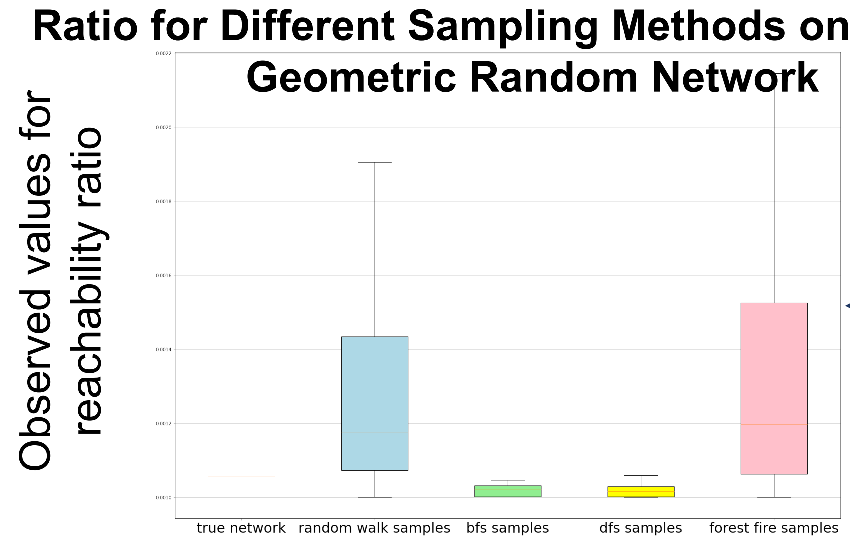
Sample Size Against Average Reachability Ratio of the Sample for a Uniform Geometric Random Network



For a geometric random network with uniform time stamps…

- BFS and DFS predict reachability ratio **extremely well** even for small sample sizes
- Random walk and forest fire overestimate

Box Plot to show the Distribution of Reachability Ratio for Different Sampling Methods on a Bursty Geometric Random Network



Network of 1000 nodes sampled at the same size 500 times for each of the four sampling methods

For a geometric random network with bursty time stamps…

- RW and FF overestimate rratio
- RW and FF predictions vary
- BFS and DFS consistently predict rratio accurately

Conclusions

- **Breadth-first and depth-first search sampling work best for calculating both average temporal degree centrality and reachability ratio**
- **BFS and DFS sample sizes need to be slightly larger for bursty networks in order to get accurate average temporal degree centrality**
- **Random walk and forest fire sampling can give close overestimates of reachability ratio but can't predict temporal degree centrality**
- **BFS and DFS sampling predictions vary for average temporal degree centrality so should be sampled multiple times and then averaged**
- **RW and FF sampling predictions vary for average reachability ratio so should be sampled multiple times and then averaged**