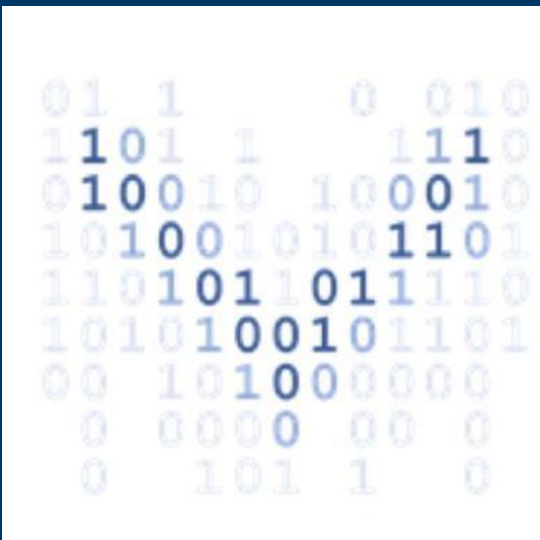


Pseudonymising livestock movement data: why it matters how you do it

Carlijn Bogaardt¹, Matthew Denwood² and Jessica Enright¹

¹School of Computing Science, University of Glasgow, Glasgow, United Kingdom

²Department of Veterinary and Animal Sciences, University of Copenhagen, Copenhagen, Denmark



Commercial livestock movements are an important factor in the spread of infectious diseases of livestock. Such movements can be analysed by social network analysis to provide valuable public health intelligence: network properties reflecting effectiveness at transmitting infection, or the relative importance of different holdings in the network. But livestock movement data are commercially sensitive – and often need to be made non-identifiable before they can be shared with data analysts and modellers. How this is done may affect the outcomes of analyses and ultimately the value of the public health intelligence.

Q: What are the effects of various pseudonymisation methods on network measures of relevance to disease transmission?

Methods

Example: Effect of *temporal* pseudonymisation on *max. reachability*

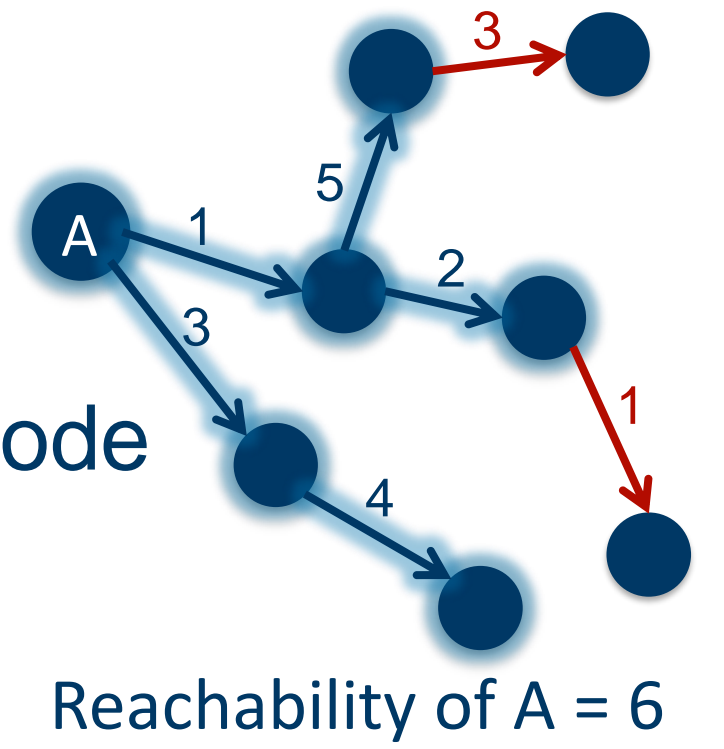
Data: Pig movements 2020, Danish Central Husbandry Register

Pseudonymisation methods for movement dates:

- **Jitter (n days):** Adds random noise (range -n to n days) to date
- **Rounding (unit):** Rounds date down to first date of time unit

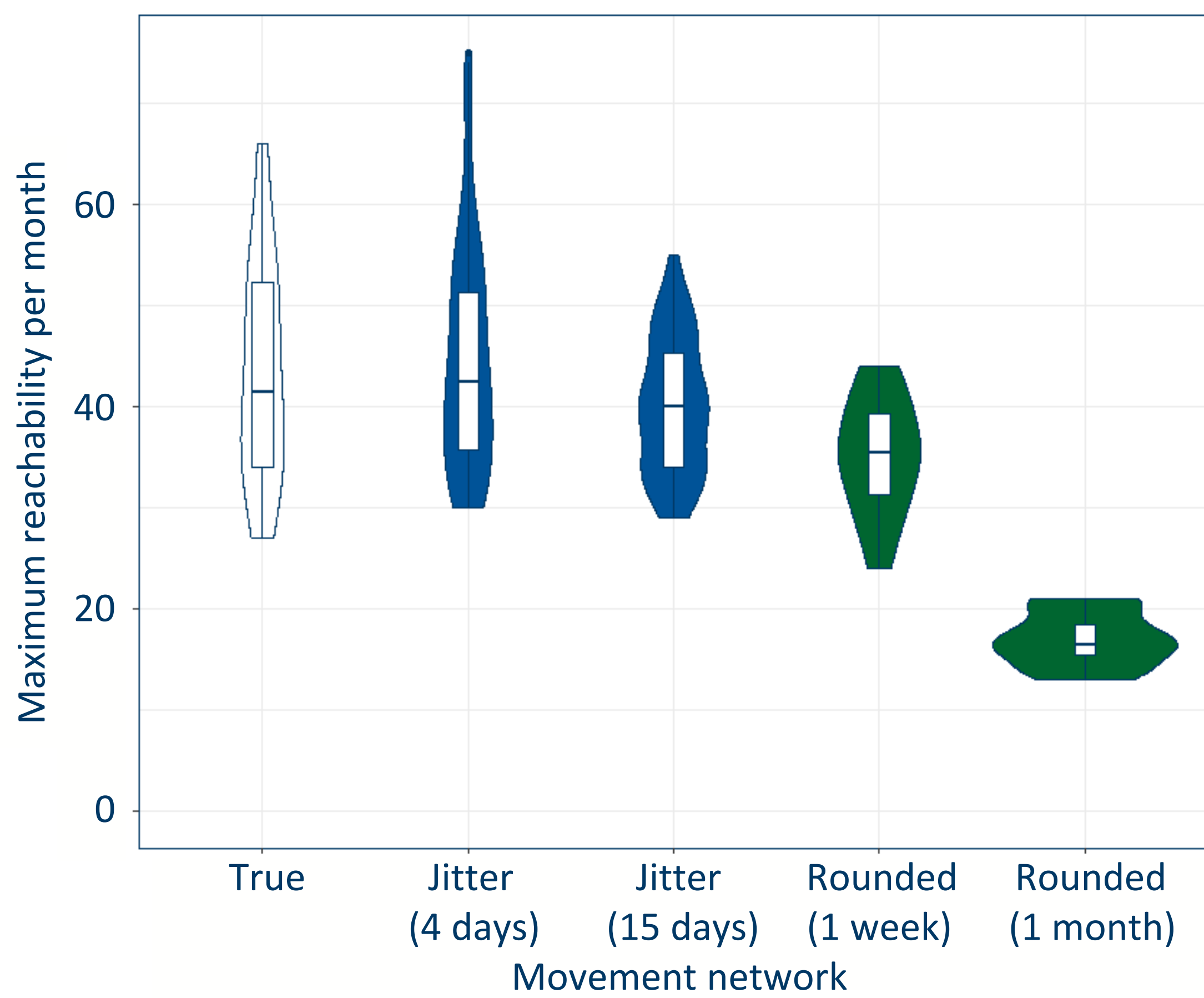
Maximum reachability:

- Reachability: How many nodes can be reached from a node, following a temporal path through the network
- Max. possible outbreak size, if starting with 1 infected node
- Effectiveness of the network at transmitting infection
- We assume each movement takes 1 full day



Results

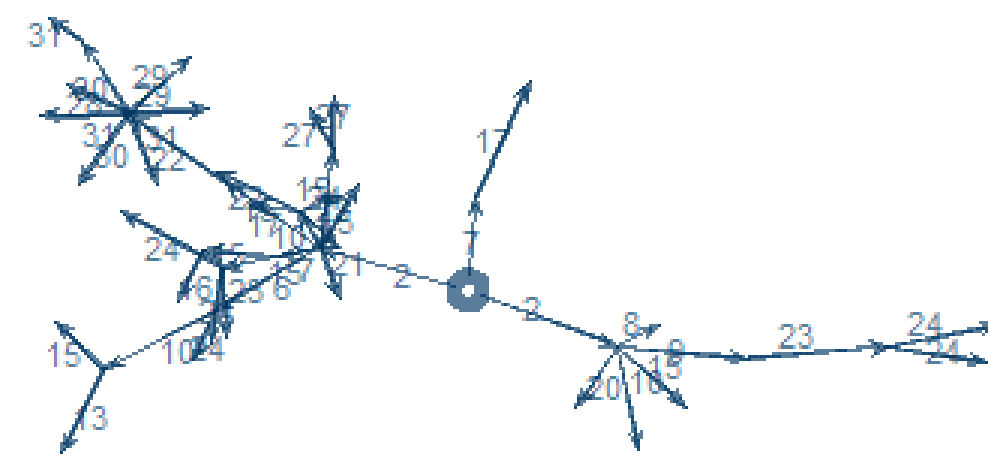
Jitter → little effect on monthly max. reachability;
Rounding → underestimates monthly max. reachability



Greater rounding unit → underestimates max. possible path length → max. reachability excessively reflects direct connections

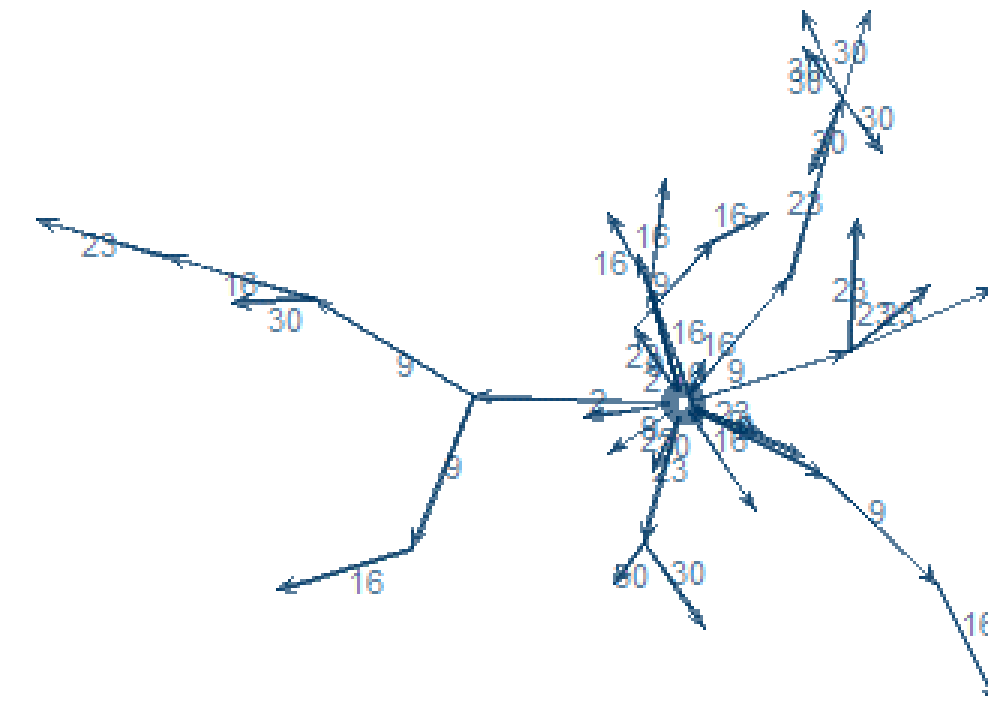
Example paths (over 1 month) from maximally reaching node:

True
Max. possible monthly path length = 31

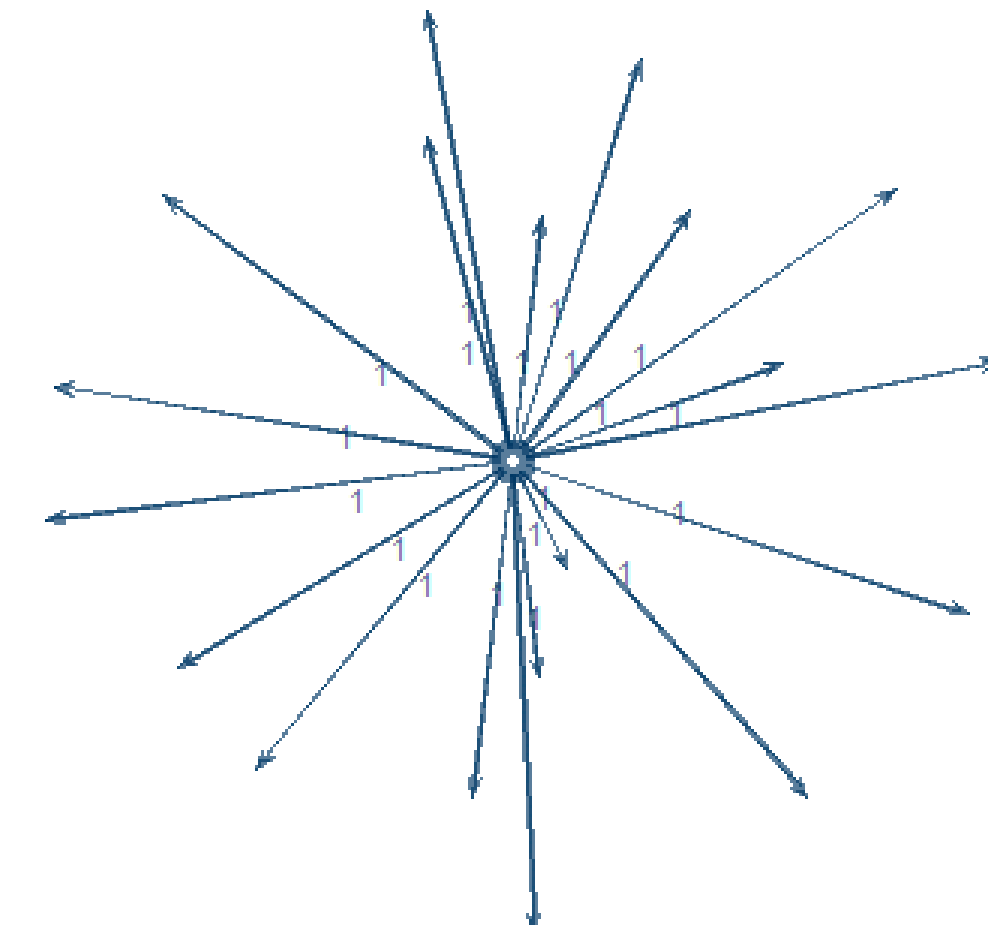


Jitter (4 days or 15 days)
Max. possible monthly path length = 31 [not shown]

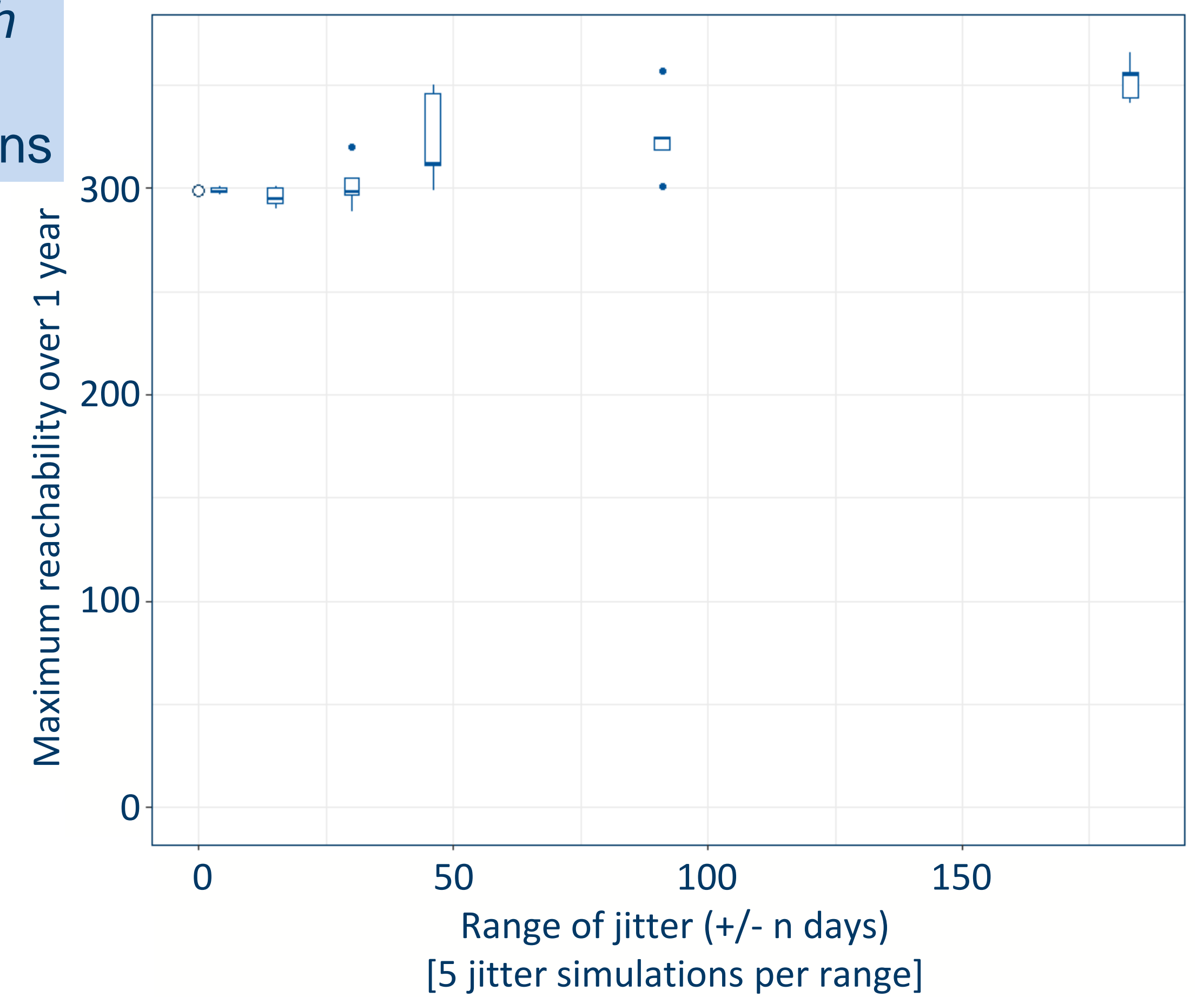
Rounded (1 week)
Max. possible monthly path length = 5



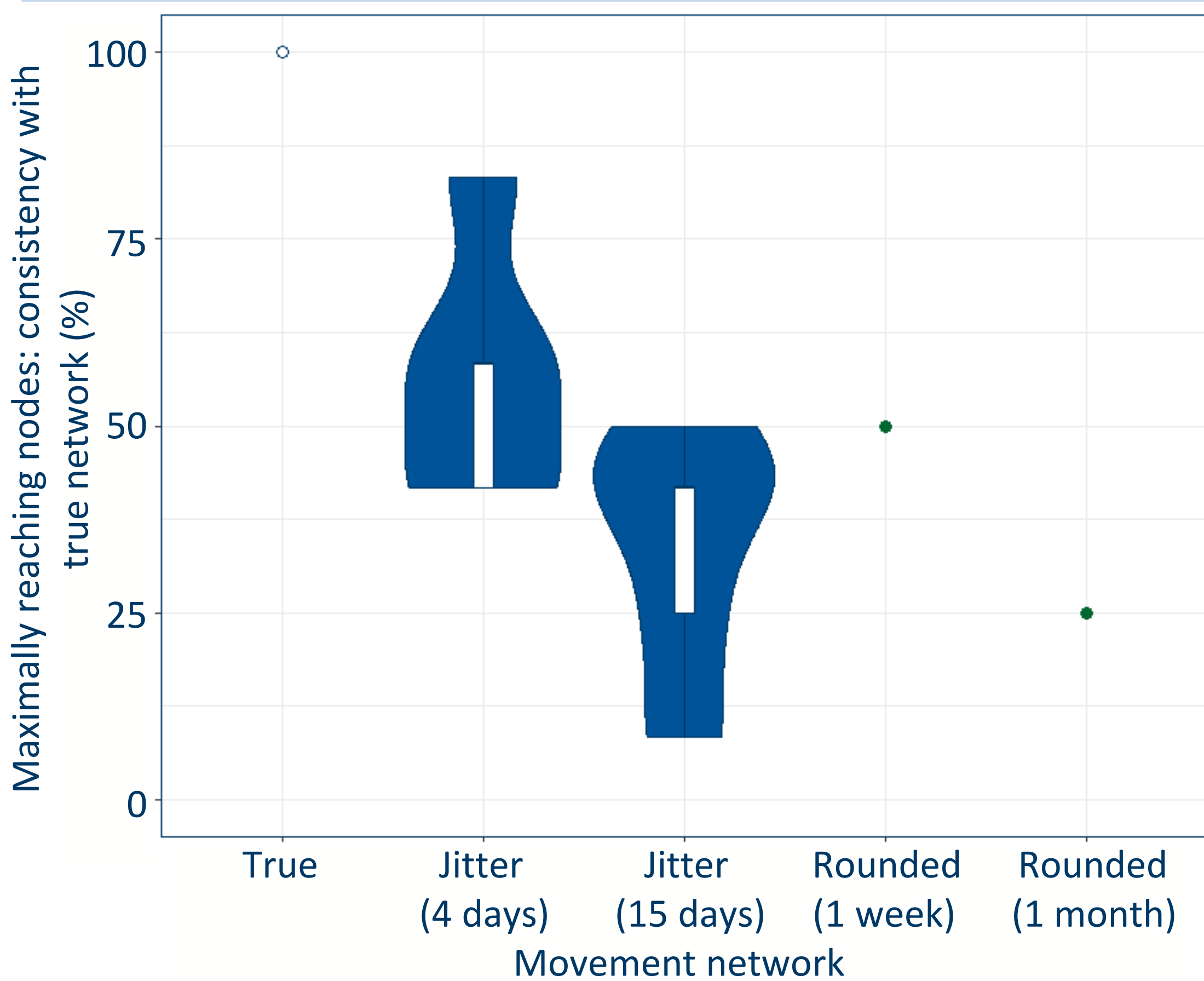
Rounded (1 month)
Max. possible monthly path length = 1



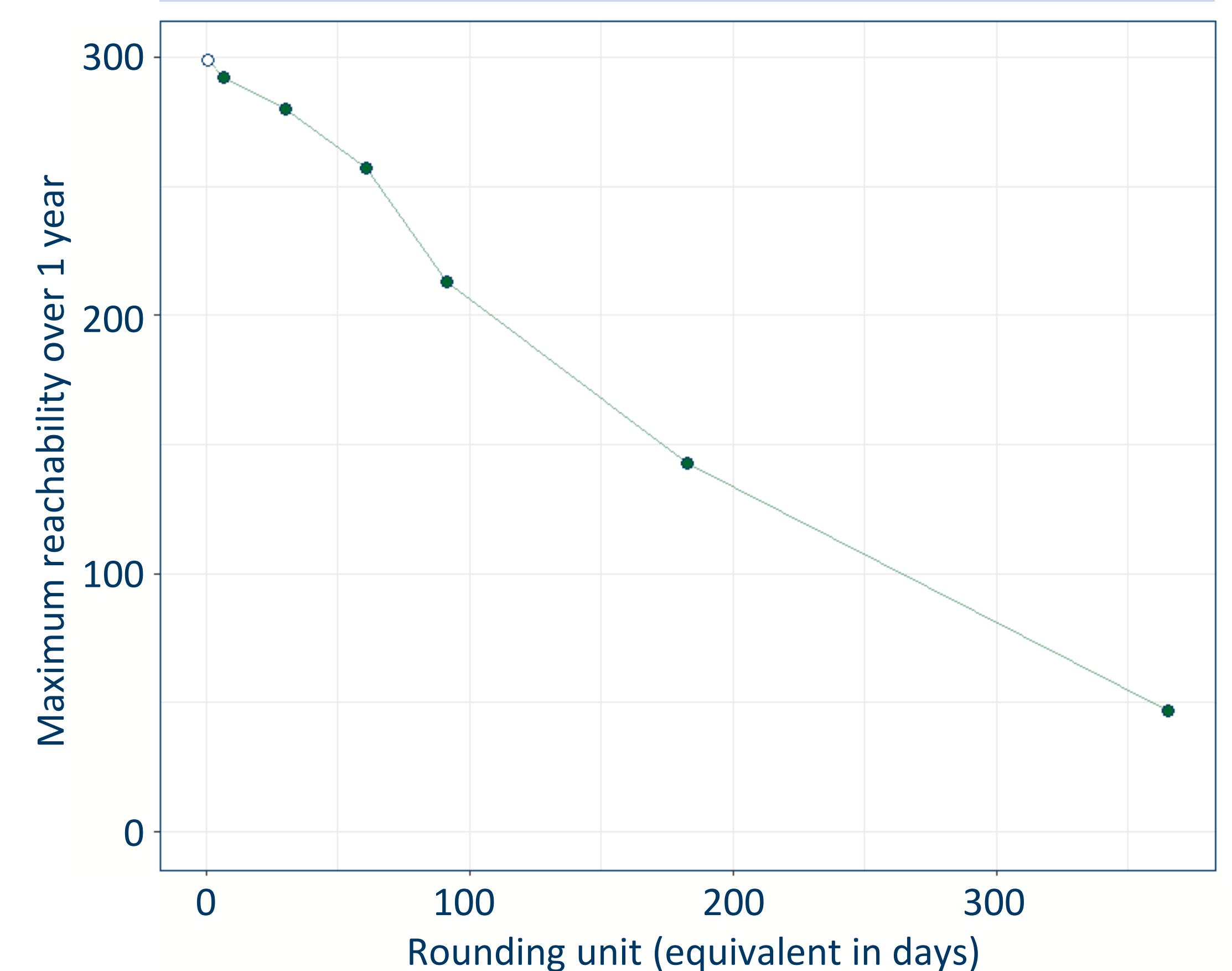
Greater jitter range → slightly overestimates overall (1-year) max. reachability



Greater jitter range or rounding unit → misrepresents the relative monthly reachability of holdings

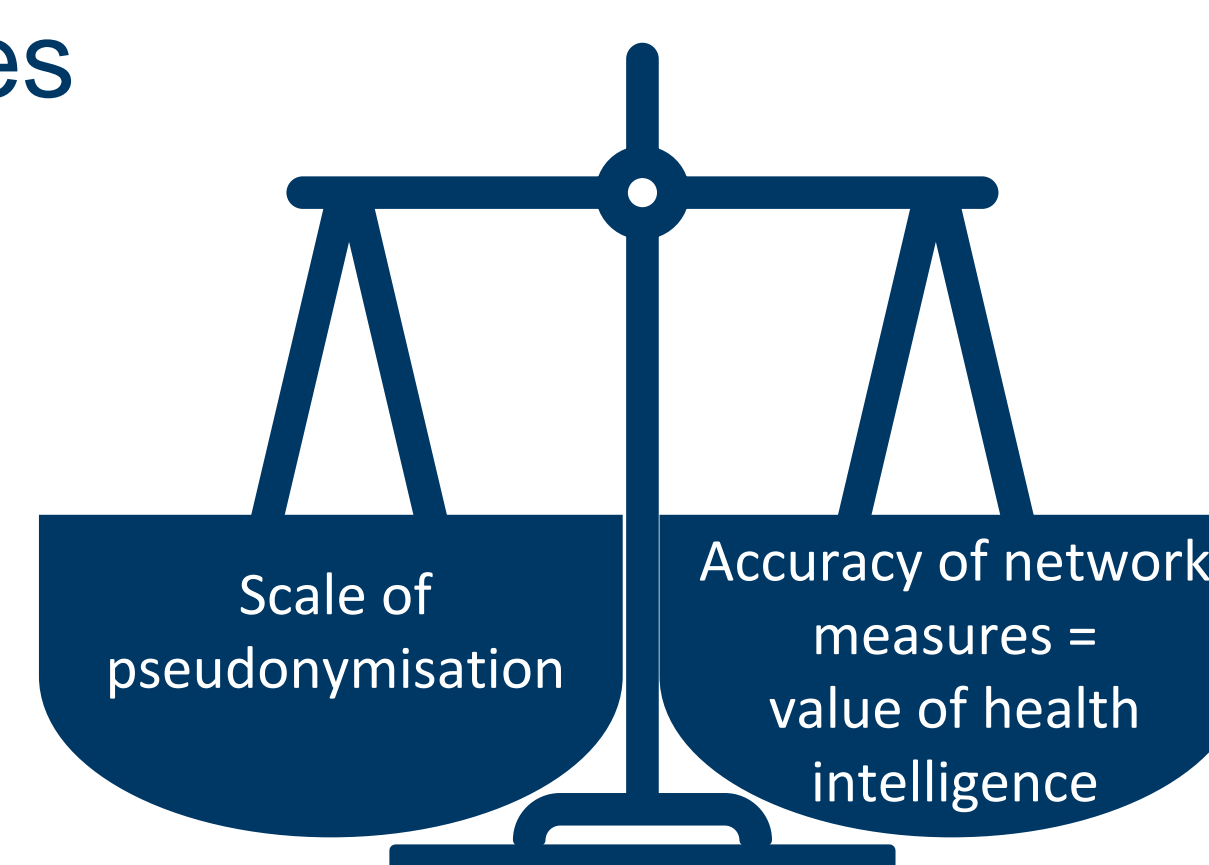


Greater rounding unit → severely underestimates overall (1-year) max. reachability



Conclusions

- For a more realistic network, choose jittering over rounding dates
- The greater the scale of pseudonymisation, the less accurate:
 - the network properties → risk of misassessing outbreak growth potential
 - the properties of specific holdings in the network → risk of misassessing effects of targeted control measures



In the DigiVet project, we are developing *Movenet*, an R package and web app for exploring and pseudonymising livestock movement networks.

[Movenet Github](#) →

