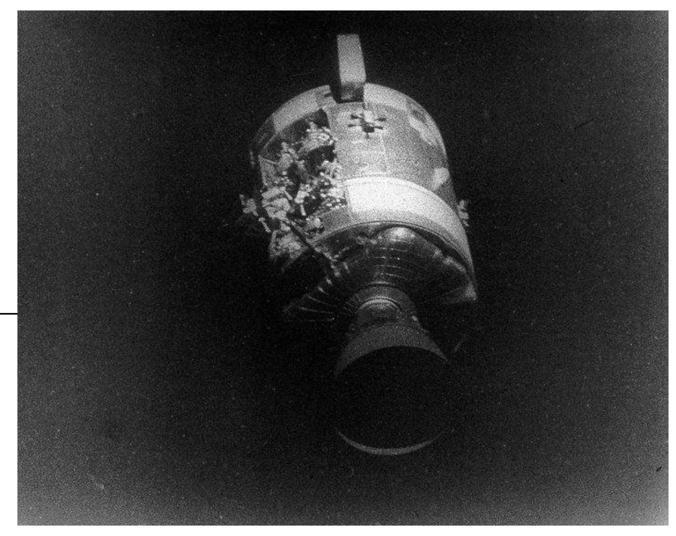
Utrecht University

Faculty of Science

(Multiple) Imputation in Relational Event History data: Missingness in Time, Sender, and/or Receiver



Student: Dominic Comerford
Supervisors: Dr. Mahdi Shafiee Kamalabad, Dr. Gerko Vink

Research aims:

1) To what extent does missingness in REH data (time, sender, and/or receiver of communications) introduce bias and
2) To what extent does imputation of missing values alleviate bias?

INTRODUCTION

- Widespread occurrence of missingness in social network analysis (SNA)
- Incomplete data, resulting from the exclusion of nodes (actors) or edges (associations) between nodes
- SNA typically from 'static' snapshots of networks
- A new type of dynamic SNA developed – Relational Event History (REH)
- REH contains
  - Timestamps of when interaction took place + Dyad of sender and receiver
- Why analyze missingness in REH?
  - High resolution and precise network data
  - Increasingly available but little to no research on the missingness problem
- Missingness simulation and imputation in time, sender and/or receiver
- Part of Apollo 13 communications among ground control and astronauts
  - 3882 relational events (communications) among 16 nodes (actors)
  - Missingness and imputation simulated 100 times
- Fully observed ('truth') & complete case analysis & imputed simulation
  - Check bias, coverage rate (CR) and absolute width (AW) compared to truth

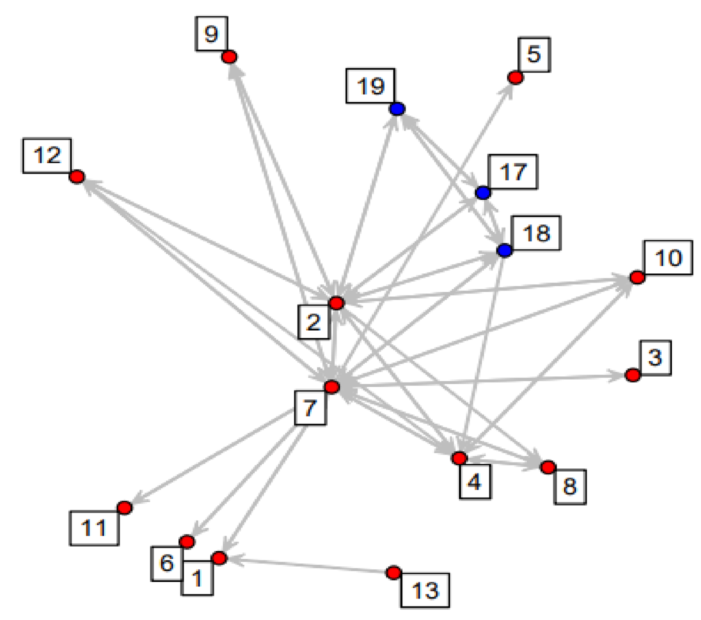
Table 1.

Relational events of Apollo 13 data

Time	S-ID	R-ID
11849.2	18	2
11854.2	2	18
…	…	…
50012.8	7	4
50014.8	4	7

Figure 1.

Network graph Apollo-13 data



THEORY AND APPROACH

Relational Event Model (REM)

- Framework to model REH data
  - Events occurs in discrete moments
  - Ties exists for short moments
  - Understanding of order and duration of events
- Enables estimation of *what* dyad will communicate and *when* future communication occurs (event rate) in log-linear function:

$$\log \lambda(s,r,t) = \sum \beta_p X_p(s,r,t)$$

- β_p refers to the impact of the *p*-th statistic $X_p(s,r,t)$ on the event rate.
- Endogenous statistics (past communication)
  - Include past communication in determining event rate
    - Reciprocity
    - In-degree sender
    - Out-degree receiver
- Exogenous statistics
  - Include external factors in determining event rate
    - Same location

Not data-dependent (NDD) Missingness

- Missingness not related to observed/unobserved characteristics of the data → random missingness across data
- Most convenient missingness mechanism
- Benchmark for more complex context (SDD, UDD)

(Multiple) Imputation

- Missing 'time' interpolated with single imputation
- Missing in sender and receiver multiple imputed
  - replacement of a missing value by a plausible one multiple times and pooling the results

RESULTS

Fully Observed Data - TRUTH

- In-degree sender (β = 4.31^{-04}, p < .001) and same location (β = -.86, p < .001) statistically significant predictors of event rate
- Reciprocity (β = 2.33^{-02}, *p* = .209) and out-degree receiver (β = -9.02^{-05}, *p* = .225) not
- Those who receive more communications more likely to initiate contact, and communication more likely to take place between nodes in different location

Complete Case Analysis

- Overestimated reciprocity (β = 2.77^{-02}) and in-degree sender (β = 5.68^{-04}) and underestimated out-degree receiver (β = 1.21^{-04}) and same location (β = -1.35) but no change in statistical significance

Imputed Simulations

- Less bias BUT…
- False statistically significant
  - Reciprocity
  - Out-degree receiver

Statistic	*β*	*SE*	*p*	*CR*	*Bias*	*PB*
Reciprocity	2.52^{-02}	1.076^{-03}	< .001	.75	1.83^{-03}	7.87
ID sender	4.18^{-04}	9.946^{-06}	< .001	.88	-1.37^{-05}	3.31
OD receiver	-9.31^{-05}	4.053^{-0}	< .001	.89	-2.89^{-06}	4.52
Same Location	-.91	1.181^{-02}	< .001	.17	-4.76^{-02}	5.52

DISCUSSION

- Various sensitivity analyses (other interpolation methods for time and 'same location' data in imputation yielder similar conclusions
- Imputation led to better estimates of effect sizes
- However, small standard errors led to false statistically significant results and suboptimal coverage rates

Limitations and directions for research

- Time interpolated and imputed as single value
- Content of communication is overlooked
- Convenient NDD-mechanism for missingness
- Samples drawn from a finite population (Apollo 13 data)
  - No sampling variance resulting in smaller standard errors

- Butts, C. T. (2008). A relational event framework for social action. *Sociological methodology*, *38*(1), 155-200. https://doi.org/10.1111/j.1467-9531.2008.00203.x
- Kamalabad, M. S., Leenders, R., & Mulder, J. (2023). What is the Point of Change? Change Point Detection in Relational Event Models. *Social Networks*, *74*, 166-181. https://doi.org/10.1016/j.socnet.2023.03.004
- Meijerink-Bosman, M., Back, M., Geukes, K., Leenders, R., & Mulder, J. (2023). Discovering trends of social interaction behavior over time: An introduction to relational event modeling: Trends of social interaction. *Behavior Research Methods*, *55*(3), 997-1023. https://doi.org/10.3758/s13428-024-02423-2