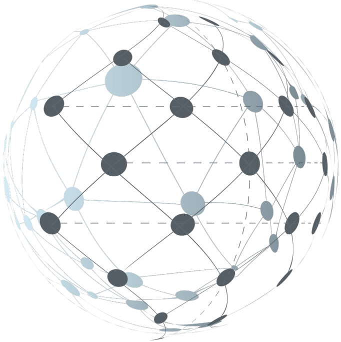
Extending the Relational Event Model

Full Day Workshop

Ernst C. Wit[1], Alessandro Lomi[1], Jürgen Lerner[2], Martina Boschi[1] Melania Lembo[1]

[1]Università della Svizzera italiana, [2]Universität Konstanz

Sunbelt 2025 - Paris, June 24th



Section	Presenter	Time	Abstract
Introduction to RHEMs[1]	Ernst	09:00–10:00	Reflects on relational and dynamic aspects of data, exploring how reframing them using REMs. It identifies opportunities and challenges in applying REMs to diverse research research contexts. Including Discussion Section.
(Hyper) Event Statistics	Jürgen	10:00–11:00	Explains how to compute relational hyper-event statistics by summarizing previous data and network attributes into explanatory variables. Including Computer Practical 1.
RHEMs: Inference	Martina	11:15–12:30	Examines several likelihood-based inference methods commonly used in relational event modeling, highlighting their limitations and interrelationships. Including Computer Practical 2.
Mixed Effect Additive RHEMs	Melania	13:30–14:30	Extends beyond linearity assumptions by introducing time-varying, non-linear, and random effects. Including Computer Practical 3.
Goodness of Fit of RHEMs	Ernst	14:30–15:15	Explores differences between model selection and goodness-of-fit (GOF), highlighting key techniques from current literature. Including Computer Practical 4.
Advanced RHEMs	Jürgen	15:30–16:30	Discusses some variations of RHEMs (Directed, Outcome, Labeled, Mixed multimode).

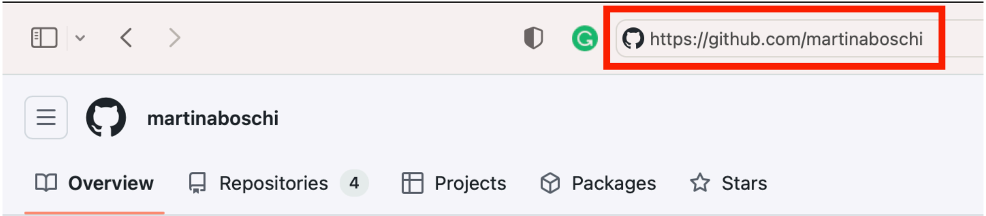
[1] Relational Hyper Event Models

Let's get started!

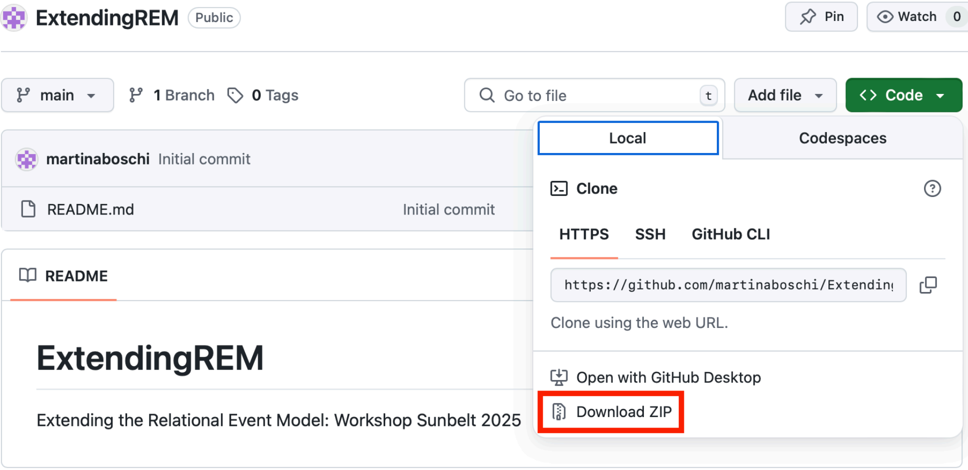
How to get materials?

Step 1: Visit GitHub

Visit `github.com/martinaboschi`:

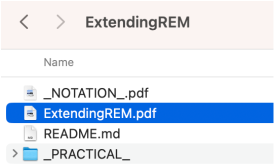


Step 2: Download the Folder



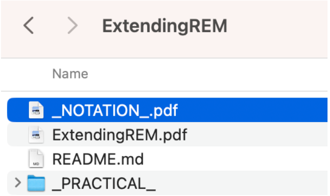
Step 3: Open the Folder

If you want to take notes on the slides...

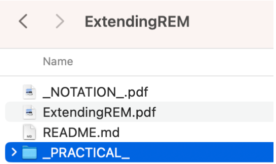


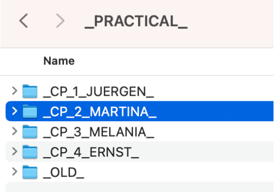
... Lost in ~~translation~~ notation?

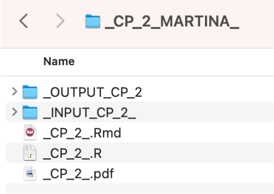
Source:Wikipedia



Step 4: Check the Practicals!







Introduction to R(H)EMs

Introduction to REMs

Abstract

Social networks are often analyzed as a collection of **static and dyadic** relations. This ignored that networks often consists of **dynamic** relations involving **more than two parties**.

In this section, we set up a general dynamic network model, called a **relational event model** to address these limitations.

Social networks: Static and Dyadic?

Social networks are **often** represented as a (un)directed graph

An edge (s, r) expresses a symmetric or asymmetric relation:

- s is friend of r
- s provides assistance to r
- s loans money to r

Social Networks: dynamic between more than 2 parties!

Social relationship often have **temporal component**:

- friendships are made and lost over time.
- emails are sent at a certain moment.
- alien species invade regions at specific times.

Social relations are **not only dyadic**:

- emails can be between 1 sender and multiple receivers.
- gossip is between multiple senders and 1 or multiple receivers.

What is a "longitudinal social network"?

Longitudinal social network

A social network where edges have a temporal component:

- **relational state:** edges appear and persist in time (e.g. friendship)
- **relational event:** edges are "instantaneous" (e.g. communication)

In this workshop we will focus[2] on

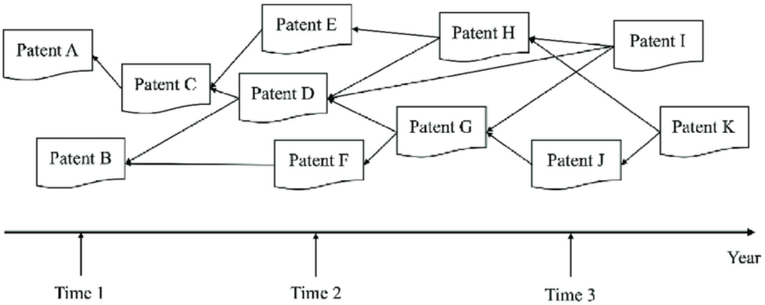
relational events models,

[2] However relational states can be thought of as two relational events:

- **friendship start:** a relational event marking beginning of state
- **friendship end:** a relational event marking end of state

Example 1: patent citations

When patent are deposited they need to cite state-of-art (i.e. previous patents)



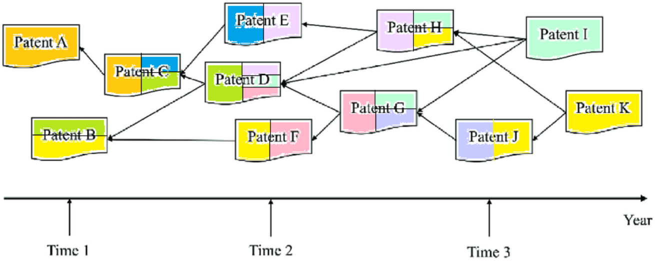
The citation network is collections of relational events, consisting of:

- **1 sender:** deposited patent.
- **multiple receivers:** cited patents in patent application.
- **time-stamp:** moment of deposition of patent.

Relational hyperevent data

A **relational hyperevent process** involves following components:

- Set of **actors** $V = \{1, 2, ..., p\}$.
- A relational hyperevent can be defined by a triple: $e = (s, r, t)$
  - $s \subset V$: Senders of interaction event e
  - $r \subset V$: Receivers of interaction event e
  - $t \in \mathbb{R}^+$: Time of interaction event



- Collection of time-stamped hyperevents $E = \{e_1, e_2, \dots e_n\}$.

Example 2: Meetings

Margeret Thatcher was British PM from 1980 until 1990.
She kept a diary from which her **meetings** were extracted.



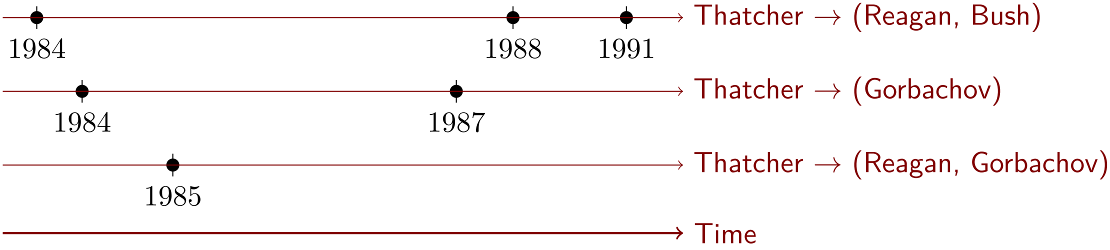
In this case, "sender set" is $s_1 = s_2 = s_3 =$ Thatcher, whereas "receiver set" changes:

- $t_1 =$ 22 Sept 1982, $r_1 = \{$Deng Xiaoping$\}$.
- $t_2 =$ 7 June 1984, $r_2 = \{$ Ronald Reagan, George Bush$\}$.
- $t_3 =$ 16 Dec 1984, $r_3 = \{$ Michael Gorbachov$\}$.

“Repeating edges”

Moreover, Thatcher met Gorbachov, Reagan and Bush several times:

- She met {Gorbachov,Reagan} in October 1985 in Reykjavik.
- She met {Gorbachov} again on 29 March 1987 in Moscow.
- She met {Reagan, Bush} on 2 June 1988 in the UK.
- She met {Reagan, Bush} again on 15 July 1991 at the G7.



Example 3: Gossiping as relational hyperevent

Gossiping

Consists of talking among a small group of people about a third person.

- Let V consist of a group of people, e.g., a class.
  - **senders:** $s \subset V$ be set of gossipers;
  - **receiver:** $r \in V$ be person about whom is gossiped;
  - **Time:** $t \in [0, T]$ moment that gossip took place.

Modelling longitudinal networks

Let $\{N_{sr}(t)\}_{SRt}$ be **longitudinal network**:

- $N_{sr}(t) =$ number of interactions from s to r until time t.
- $N_{sr}(0) = 0$

Mathematically, it is always possible to decompose:

$$N_{sr}(t) = \underbrace{\Lambda_{sr}(t)}_{\text{structural part}} + \underbrace{M_{sr}(t)}_{\text{noise part}}$$

If it exists, then the **hazard**

$$\lambda_{sr}(t) = \frac{d\Lambda_{sr}}{dt}(t)$$

is *instantaneous rate* that (s, r) occurs at time t.

What is a hazard?

Hazard

Hazard (also: rate, intensity) is instantaneous probability that event occurs now.

Mathematically:

$$P\left((s, r) \text{ occurs } (t, t+\delta t] \mid \text{did not occur till } t\right) \approx \lambda_{sr}(t)\delta t$$

So, if we know **hazard**:

- we know **entire probabilistic structure** of event network
- we can **simulate** relational events:
  - **Gillespie algorithm**: using exponential event times $\text{Exp}(\lambda_{sr})$.
  - **tau-leap algorithm**: using small time increments δt

Modelling hazard = modelling network

Let

$$\lambda_{sr}(t) = \underbrace{W_{sr}(t)}_{\text{Is } (s,r) \text{ at risk?}} \times \underbrace{\lambda_0(t)}_{\text{baseline hazard}} \times \underbrace{e^{f(x_{sr}(t)}}_{\text{edge specific risk factors}}$$

where

- $W_{sr}(t)$: edge specific risk indicator
  e.g. a past patent *cannot* cite a future patent.
- $\lambda_0(t)$: global risk determinants
  e.g. very few bike rides when it rains at time t.
- $f(x_{sr}(t)$: edge-specific risk determinants
  e.g. if r is a popular person, then (s, r) more likely.

How does a relational hyperevent network unfold?

Let's consider 3 individuals that enjoy **gossiping**:

- Two of them are **female, ages 10, 15** (1, 2), one of them is **male age 12** (3).
- Persons $s = (s_1, s_2)$ gossip about person r with rate,

$$\lambda_{sr}(t) = W_{sr}(t) e^{2 1_{\{\text{gender}(s_1)=\text{gender}(s_2)\}}}.$$

- Value 2 identifies
  - **relevance of sender gender:** it is not zero
  - **direction of sender gender:** same sender gender $\Rightarrow$ more gossips
- W_{sr} has value
  - 1 if set s consists of 2 people and r of 1,
  - 0 otherwise.

Distribution of event times

(1,2) gossip about 3

(1,3) gossip about 2

(2,3) gossip about 1

At each moment, there are 3 possible events. Let

$$\Delta T_{sr} = \text{waiting time until } (s_1, s_2) \text{ gossip about } r$$

The distribution of ΔT_{sr} is given as

$$\Delta T_{sr} \sim \text{Exp}(e^{2 \times 1_{\{\text{gender}(s_1)=\text{gender}(s_2)\}}})$$

When will first gossip happen?

The first gossip will happen at time T_1, where

$$T_1 = \min_{sr} \Delta T_{sr} = \min\{\Delta T_{12,3}, \Delta T_{23,1}, \Delta T_{13,2}\}$$

Using property that minimum of exponentials is exponential with sum of rates,

$$T_1 \sim \text{Exp}\left(\sum_{s,r} e^{2 \times 1_{\{\text{gender}(s_1)=\text{gender}(s_2)\}}} = e^2 + e^0 + e^0 \approx 9.4\right)$$

When will first gossip happen?

The first gossip will happen at time T_1, where

$$T_1 = \min_{sr} \Delta T_{sr} = \min\{\Delta T_{12,3}, \Delta T_{23,1}, \Delta T_{13,2}\}$$

Using property that minimum of exponentials is exponential with sum of rates,

$$T_1 \sim \mathsf{Exp}\left(\sum_{s,r} e^{2\times 1_{\{\mathsf{gender}(s_1)=\mathsf{gender}(s_2)\}}} = e^2 + e^0 + e^0 \approx 9.4\right)$$

Who is actually gossiping?

We now know that first event happened at t_1. But which one?

$$E_1 = \text{event that happened at time } t_1.$$

Using property that which exponential is minimum is multinomial,

$$E_1 \sim \text{Multinomial}\left(\frac{e^2}{e^2+e^0+e^0} = 0.72,\ \frac{e^0}{e^2+e^0+e^0} = 0.14,\ \frac{e^0}{e^2+e^0+e^0} = 0.14\right)$$

Who is actually gossiping?

We now know that first event happened at t_1. But which one?

$$E_1 = \text{event that happened at time } t_1.$$

Using property that which exponential is minimum is multinomial,

$$E_1 \sim \text{Multinomial}\left(\frac{e^2}{e^2+e^0+e^0} = 0.72,\ \frac{e^0}{e^2+e^0+e^0} = 0.14,\ \frac{e^0}{e^2+e^0+e^0} = 0.14\right)$$

When will second gossip event happen?

The gossip event will happen at time T_2, where

$$T_2 = \min\{t_1 + \Delta T_{12,3}, \Delta T_{23,1}, \Delta T_{13,2}\}$$

Using memoryless property of an exponential,

$$T_2 \sim t_1 + \text{Exp}\left(\sum_{s,r} e^{2\times 1_{\{\text{gender}(s_1)=\text{gender}(s_2)\}}} = e^2 + e^0 + e^0 \approx 9.4\right)$$

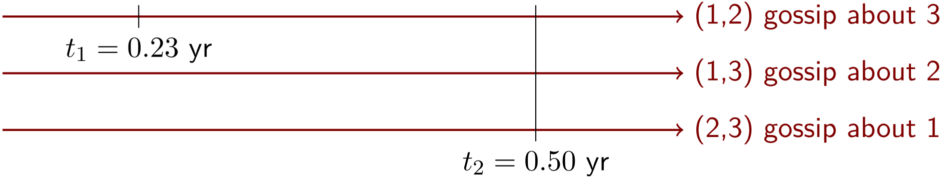
When will second gossip event happen?

The gossip event will happen at time T_2, where

$$T_2 = \min\{t_1 + \Delta T_{12,3}, \Delta T_{23,1}, \Delta T_{13,2}\}$$

Using memoryless property of an exponential,

$$T_2 \sim t_1 + \mathsf{Exp}\left(\sum_{s,r} e^{2 \times 1_{\{\mathsf{gender}(s_1) = \mathsf{gender}(s_2)\}}} = e^2 + e^0 + e^0 \approx 9.4 \right)$$



Which pair of individuals make up second request?

We know second event happened at $t_2 = 0.50$ yr. Which one?

$$E_2 = \text{event that happened at time } t_2.$$

Using property that which exponential is minimum is multinomial,

$$E_2 \sim \text{Multinomial}\,(0.72,\ 0.14,\ 0.14)$$

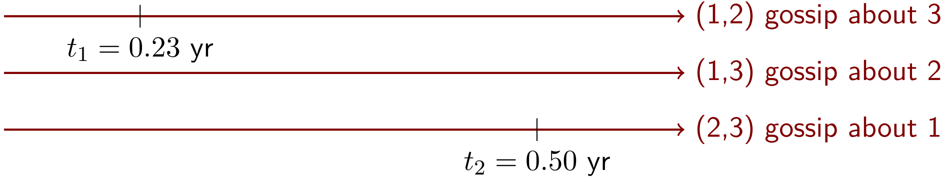
Which pair of individuals make up second request?

We know second event happened at $t_2 = 0.50$ yr. Which one?

$$E_2 = \text{event that happened at time } t_2.$$

Using property that which exponential is minimum is multinomial,

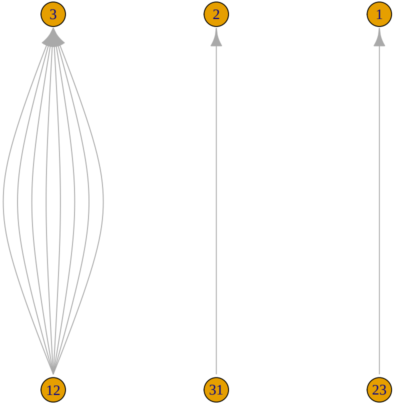
$$E_2 \sim \text{Multinomial}\,(0.72,\ 0.14,\ 0.14)$$



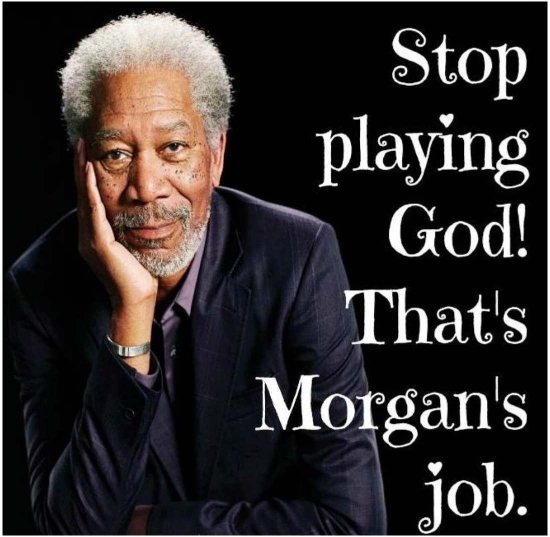
Repeat... to obtain a dynamic network

Repeating procedure another 8 times results in dynamic network with 10 hyperedges:

	time	gossiper1	gossiper2	gossipee
1	0.23	1	2	3
2	0.50	2	3	1
3	0.60	1	2	3
4	0.67	1	2	3
5	0.81	1	2	3
6	1.01	3	1	2
7	1.06	1	2	3
8	1.14	1	2	3
9	1.18	1	2	3
10	1.32	1	2	3



Let’s stop playing God



Can we use data to answer a sociological question?

Sociologists have two hypotheses about gossiping:

- People of same gender tend to gossip together.
- Older people get gossiped about more than younger people.

Translate these two hypotheses into a testable model:

Translation of 2 gossip hypothesis

Persons $s = (s_1, s_2)$ gossip about person r with rate,

$$\lambda_{sr}(t) = W_{sr}(t) e^{\beta_1 1_{\{\text{gender}(s_1)=\text{gender}(s_2)\}} + \beta_2 \text{age}_r}.$$

And our aim is to test 2 hypothesis:

$$\begin{aligned} H_0 &: \quad \beta_1 = 0 \\ H_1 &: \quad \beta_1 \neq 0 \end{aligned} \qquad\qquad \begin{aligned} H_0 &: \quad \beta_1 = 0 \\ H_1 &: \quad \beta_1 \neq 0 \end{aligned}$$

Can we use data to answer a sociological question?

Sociologists have two hypotheses about gossiping:

- People of same gender tend to gossip together.
- Older people get gossiped about more than younger people.

Translate these two hypotheses into a testable model:

Translation of 2 gossip hypothesis

Persons $s = (s_1, s_2)$ gossip about person r with rate,

$$\lambda_{sr}(t) = W_{sr}(t) e^{\beta_1 1_{\{\text{gender}(s_1)=\text{gender}(s_2)\}} + \beta_2 \text{age}_r}.$$

And our aim is to test 2 hypothesis:

$$\begin{aligned} H_0 &: \quad \beta_1 = 0 \\ H_1 &: \quad \beta_1 \neq 0 \end{aligned} \qquad\qquad \begin{aligned} H_0 &: \quad \beta_1 = 0 \\ H_1 &: \quad \beta_1 \neq 0 \end{aligned}$$

Data

We want to use data to answer this question:

	time	gossiper1	gossiper2	gossipee	same gender	age receiver
1	0.23	1	2	3	1	12
2	0.50	2	3	1	0	10
3	0.60	1	2	3	1	12
4	0.67	1	2	3	1	12
5	0.81	1	2	3	1	12
6	1.01	3	1	2	0	15
7	1.06	1	2	3	1	12
8	1.14	1	2	3	1	12
9	1.18	1	2	3	1	12
10	1.32	1	2	3	1	12

But this dataset *hides* some information.

Data

We want to use data to answer this question:

	time	gossiper1	gossiper2	gossipee	same gender	age receiver
1	0.23	1	2	3	1	12
2	0.50	2	3	1	0	10
3	0.60	1	2	3	1	12
4	0.67	1	2	3	1	12
5	0.81	1	2	3	1	12
6	1.01	3	1	2	0	15
7	1.06	1	2	3	1	12
8	1.14	1	2	3	1	12
9	1.18	1	2	3	1	12
10	1.32	1	2	3	1	12

But this dataset *hides* some information.

Data (with riskset information)

Data should also explicitly state **which events did not happen!**

	time	gossiper1	gossiper2	gossipee	same gender	age receiver	event?
1	0.23	1	2	3	1	12	1
		2	3	1	0	10	0
		3	1	2	0	15	0
2	0.50	2	3	1	0	10	1
		1	2	3	1	12	0
		3	1	2	0	15	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
9	1.18	1	2	3	1	12	1
		2	3	1	0	10	0
		3	1	2	0	15	0
10	1.32	1	2	3	1	12	1
		2	3	1	0	10	0
		3	1	2	0	15	0

Inference: Cox Model inference

When data is presented as:

- a series of **events** that happen, competing with
- **non-events** that did not happen
- each with its own **exponential hazard** of happening,

then we can use **Cox Model Inference**.

Results: what about our 2 hypotheses?

```
> gossip.rem<-coxph(rel.event~same.gender+age.receiver)
> summary(gossip.rem)
Call:
coxph(formula = rel.event ~ same.gender + age.receiver)

  n= 30, number of events= 10

                 coef exp(coef)  se(coef)     z Pr(>|z|)
same.gender  2.079e+00 8.000e+00 8.031e-01 2.589  0.00962 **
age.receiver 2.577e-17 1.000e+00 2.828e-01 0.000  1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- $\beta_1 = 2.1$ (p-value $= 0.01$): people of same gender tend to gossip more.
- $\beta_2 = 0.0$ (p-value $= 1.00$): age of gossipee does not affect rate.

Results: what about our 2 hypotheses?

```
> gossip.rem<-coxph(rel.event~same.gender+age.receiver)
> summary(gossip.rem)
Call:
coxph(formula = rel.event ~ same.gender + age.receiver)

  n= 30, number of events= 10

                  coef exp(coef)  se(coef)     z Pr(>|z|)
same.gender  2.079e+00 8.000e+00 8.031e-01 2.589  0.00962 **
age.receiver 2.577e-17 1.000e+00 2.828e-01 0.000  1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- $\beta_1 = 2.1$ (p-value $=$ 0.01): people of same gender tend to gossip more.
- $\beta_2 = 0.0$ (p-value $=$ 1.00): age of gossipee does not affect rate.

Take-home messages

This tutorial considered **dynamic hyperevent networks**

1. ... as series of *relational hyperevents* in time,
2. ...between sets of Senders and sets of Receivers,
3. .... via a **general counting process**:
   - event times are (generalized) exponentially distributed
   - event types are multinomially distributed
4. Inference as **Cox Model**:
   - Each event (s_i, r_i, t_i) is contrasted with set of non-events $\{(s_i^*, r_i^*, t_i)\}$
   - Results can be used to evaluate sociological hypotheses.

DS1: Reframing Data via Relational Event Modeling

Abstract

In this discussion section, participants are invited to reflect on the relational and dynamic aspects of the data they work with and explore how these can be reframed using relational event modeling. Together, we'll identify opportunities and challenges in applying REMs to diverse research contexts.

(Hyper)event Statistics

Hyperevent Statistics

We have seen that crucial building block of REM is hazard:

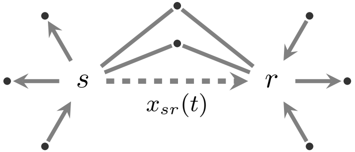
$$\lambda_{sr}(t) = \underbrace{W_{sr}(t)}_{\text{Is } (s,r) \text{ at risk?}} \times \underbrace{\lambda_0(t)}_{\text{baseline hazard}} \times \underbrace{e^{f(x_{sr}(t))}}_{\text{edge specific risk factors}}$$

In this section, we will focus on $x_{sr}(t)$, also called:

- Hyper-event statistics
- Covariates
- Explanatory variables
- “REM effects”

REM statistics

REM statistics $x_{sr}(t)$ are functions of **properties** of network **configurations** around the dyad (s, r).



Examples: repetition, reciprocation, (in-/out-/mixed-)degrees, triadic effects, four-cycle effects, exogenous attributes, ...

REM statistics

REM statistics $x_{sr}(t)$ are functions of **properties** of network **configurations** around the dyad (s, r).

Decisions to take:

1. Which property (network **attribute**)?
   - exogenous covariates: actors' age, gender; kinship relations; geographic distance; ...
   - (function of) past events: binary, count, aggregated weight, type, recency (time decay), ...
   - time since last event
2. Which configuration / structural effect (REM **statistic**)?
   - degree effects: out-star / in-star on source / target
   - dyadic effects: repetition, reciprocation
   - triadic effects: transitive / cyclic closure, shared neighbors
   - four-cycle effects, ...

More decisions to take: functional form, standardization, ...

Using statistics to test hypotheses with REM.

example: transitive closure in phone calls

past events

A calls B at time t_1
B calls C at time t_2
$\dots$

all events before t

network attributes z

$A \rightarrow B \rightarrow C$

$z_{AB}(t) = 1$

$z_{BC}(t) = 1$

network effects

network statistics x

A, B, C

z_{AB}, z_{BC}, $x_{AC}(t)$

x specify

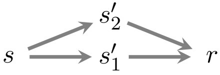
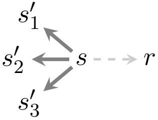
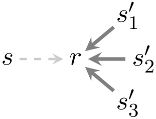
probability of next event(s)

A calls C at time t

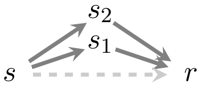
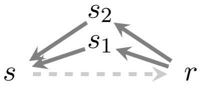
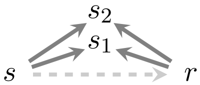
has ***increased probability***

Examples: statistics $x_{sr}(t)$.

N. of previous events from s to r: $\text{prior}(s,r,t) = |\{(s_i,r_i,t_i) \in E\colon s_i = s \wedge r_i = r \wedge t_i < t\}|$.

statistic	$x_{sr}(t) =$	$s \dashrightarrow r$ depends on
`repetition`	$\text{prior}(s,r,t)$	$s \longrightarrow r$
`reciproc.`	$\text{prior}(r,s,t)$	$s \longleftarrow r$
`transitive`	$\sum_{a \neq s,r} \text{prior}(s,a,t) \cdot \text{prior}(a,r,t)$	
`outDegSource`	$\sum_a \text{prior}(s,a,t)$	
`inDegTarget`	$\sum_a \text{prior}(a,r,t)$	

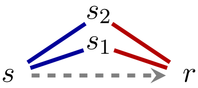
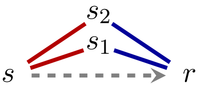
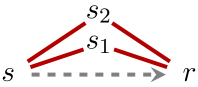
Examples: varying dyad direction.

statistic	$x(s,r,t) =$	$s \dashrightarrow r$ depends on
`trans.`	$\sum_{a \neq s,r} \text{prior}(s,a,t) \cdot \text{prior}(a,r,t)$	
`cycl.`	$\sum_{a \neq s,r} \text{prior}(a,s,t) \cdot \text{prior}(r,a,t)$	
`shared rec.`	$\sum_{a \neq s,r} \text{prior}(s,a,t) \cdot \text{prior}(r,a,t)$	

Possible: **symmetrized** tie weights: $\text{prior}(s', s'', t) + \text{prior}(s'', s', t)$.

Examples: combining different types of attributes.

Given: attributes indicating friends and enemies.

statistic	$x_{sr}(t) =$	$s \dashrightarrow r$ depends on
`enemy.of.friend`	$\sum_a \mathsf{F}(s,a,t) \cdot \mathsf{E}(r,a,t)$	
`friend.of.enemy`	$\sum_a \mathsf{E}(s,a,t) \cdot \mathsf{F}(r,a,t)$	
`enemy.of.enemy`	$\sum_a \mathsf{E}(s,a,t) \cdot \mathsf{E}(r,a,t)$	

Might be combined with variation in the dyad direction.

Time decay – recent events.

Often, network attributes count the number of prior events:

$$\begin{aligned}\mathsf{prior}(s,r,t) &= |\{(s_i,r_i,t_i)\in E\colon s_i=s\wedge r_i=r\wedge t_i<t\}| \\ &= \sum_{(s_i,r_i,t_i)\in E\colon s_i=s\wedge r_i=r\wedge t_i<t} 1\end{aligned}$$

To give recent events a higher weight we can use the attribute:

$$\mathsf{recent}(s,r,t) = \sum_{(s_i,r_i,t_i)\in E\colon s_i=s\wedge r_i=r\wedge t_i<t} \exp\left(-(t-t_i)\frac{\log 2}{T_{1/2}}\right)$$

for a given half life parameter $T_{1/2}>0$.

Temporal statistics – time since last event

Instead of aggregating past events on dyads, we can also record time since last event.

Examples of temporal statistics $x_{sr}(t)$:

$$x_{sr}(t) = t - \tau^*(t)$$

where $\tau^*(t)$ stands for one of the following:

- repetition: time of last event (s, r)
- reciprocation: time of last event (r, s)
- node activity: time of last event with source s
- transitive closure: time of last creation of transitive two-path: $s \to a \to r$
- ...

What is point of "temporal statistics"?

Describe temporal activation profile of repetition, reciprocity, ...

Practical aspects

What do we have to do to specify a REM?

Relational event models (REM).

For every **possible event dyad** $(s, r) \in \mathcal{R}_t$,

REM estimate the **rate of events** from s to r at t as a function of explanatory variables.

The good news: statistical software (free or commercial) has powerful packages for time-to-event / survival analysis.

However, it does not understand dependencies in event networks (cannot compute explanatory variables).

This gap can be filled by the **eventnet software**.

Relational event models (REM).

For every **possible event dyad** $(s, r) \in \mathcal{R}_t$,

REM estimate the **rate of events** from s to r at t as a function of explanatory variables.

The good news: statistical software (free or commercial) has powerful packages for time-to-event / survival analysis.

However, it does not understand dependencies in event networks (cannot compute explanatory variables).

This gap can be filled by the **eventnet software**.

Relational event models (REM).

For every **possible event dyad** $(s, r) \in \mathcal{R}_t$,

REM estimate the **rate of events** from s to r at t as a function of explanatory variables.

The good news: statistical software (free or commercial) has powerful packages for time-to-event / survival analysis.

However, it does not understand dependencies in event networks (cannot compute explanatory variables).

This gap can be filled by the **eventnet software**.

Workflow of eventnet.

`https://github.com/juergenlerner/eventnet`

Computes specified covariates for events and non-events.
Models can be fitted with standard software (`coxph`, ...).

observed events

A	B	t_1
B	C	t_2
A	C	t_3
...		

eventnet

statistics of events and non-events

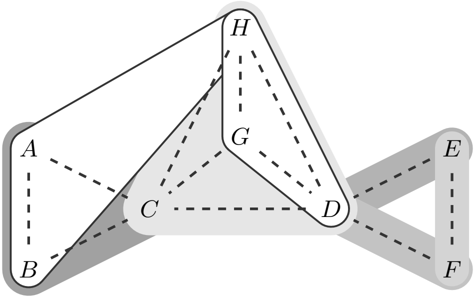
A	B	t_1	1	2.4	0.7
A	C	t_1	0	0.1	1.5
B	C	t_1	0	0.3	2.1
B	C	t_2	1	1.7	2.3
...					

configuration (file)
determines how statistics are computed from input events (reproducibility)

Configurations can be specified with the eventnet graphical user interface or via a configuration file (XML).

Template for RHEM effects (undirected).

$x_{\{A,B,H\}}(t)$ – should explain prob. of event $(\{A,B,H\},t)$.



- iterate over specific subsets $S' \subseteq \{A,B,H\}$
- compute numeric property $z_{S'}(t)$ (exogenous, endogenous)
- aggregate the $z_{S'}(t)$ over the specific subsets S'

Exact repetition of participant lists.

Tendency to repeat events with exactly the same participant list.

For a given hyperedge $S \subseteq V$, sum over all previous events $e_i = (S_i, t_i)$ with $S_i = S$.

$$\text{repetition}(S,t) \quad = \quad \sum_{e_i \,:\, t_i < t} \mathbf{1}(S_i = S)$$

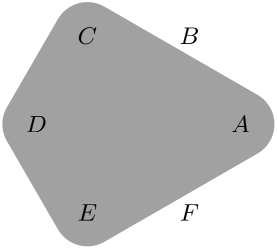
Might down-scale past events exponentially by elapsed time (same for many other effects).

$$\mathbf{1}(\text{condition}) := \begin{cases} 1 & \text{if condition is true} \\ 0 & \text{else} \end{cases}$$

Subset repetition of order k.

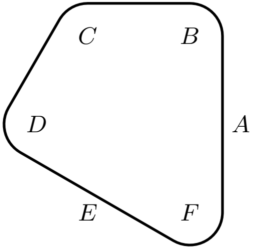
Average number of prior joint events over $S' \subseteq S$ of size k.

$$\text{sub.rep}^{(k)}(S,t) \quad = \quad \sum_{S' \in \binom{S}{k}} \frac{\text{hy.deg}(S',t)}{\binom{|S|}{k}} \quad .$$



$S = \{B, C, D, F\}$

$$\begin{aligned}\text{sub.rep}^{(1)}(S) &= 2/4\\ \text{sub.rep}^{(2)}(S) &= 1/6\\ \text{sub.rep}^{(3)}(S) &= 0/4\end{aligned}$$

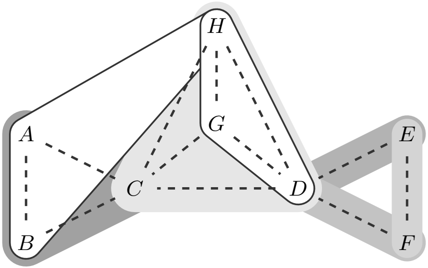


$$\text{hy.deg}(S',t) \quad = \quad \sum_{e_i\,:\,t_i<t} \mathbf{1}(S' \subseteq S_i)$$

Triadic closure.

Two-paths via third actors r for all $\{s, s'\} \subseteq S$.

$$\text{closure}(S,t) = \sum_{\{s,s'\}\in\binom{S}{2}\wedge r\neq s,s'} \frac{\min[\text{hy.deg}(\{s,r\},t), \text{hy.deg}(\{s',r\},t)]}{\binom{|S|}{2}} \,.$$

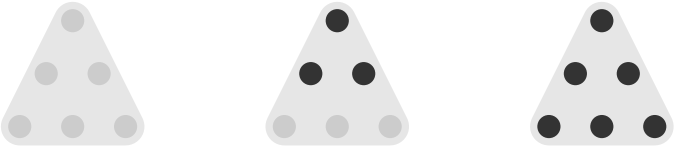


closure$(\{A,B,H\}) = 1$ closure$(\{D,G,H\}) = 2$

Covariate effects: main effect and homophily.

Given a node-attribute $z_S(t)$ (example below: binary)

$$\begin{aligned}\text{avg.z}(S,t) &= \sum_{s\in S} z_s(t)/|S| \\ \text{diff.z}(S,t) &= \sum_{\{s,s'\}\in\binom{S}{2}} \frac{|z_s(t)-z_{s'}(t)|}{\binom{|S|}{2}}\end{aligned}$$



$$\begin{aligned}\text{avg.z}(S_1) &= 0 \\ \text{diff.z}(S_1) &= 0\end{aligned} \qquad \begin{aligned}\text{avg.z}(S_2) &= 0.5 \\ \text{diff.z}(S_2) &= \tfrac{9}{15}\end{aligned} \qquad \begin{aligned}\text{avg.z}(S_3) &= 1 \\ \text{diff.z}(S_3) &= 0\end{aligned}$$

diff.z is measure of diversity; negative param $\Leftrightarrow$ homophily

Excursus: using RHEM as null models.

What can RHEM tell us about specific (groups of) individuals?

Illustrates subset repetition of different order.

Applied to the character co-appearance network
in Victor Hugo's *Les Misérables*.

Lerner, Hâncean, & Perc (2025). **Modeling temporal hypergraphs**. *arXiv preprint* `https://arxiv.org/abs/2506.01408`.

The three most prominent actors in *Les Misérables*.

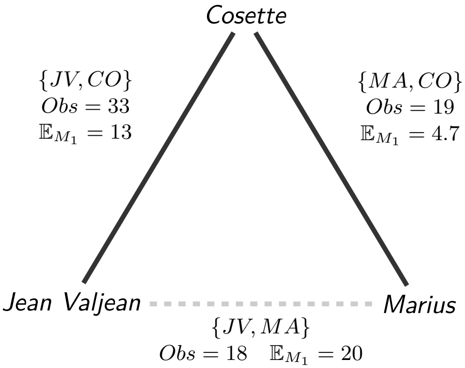
Jean Valjean, *Cosette*, and *Marius*

Do they form an open triangle, closed triangle, or filled triangle?

Data compiled by Donald Knuth: 80 actors in 288 chapters/hyperevents.

JV: 113 chapters; MA: 77 chapters; CO: 55 chapters

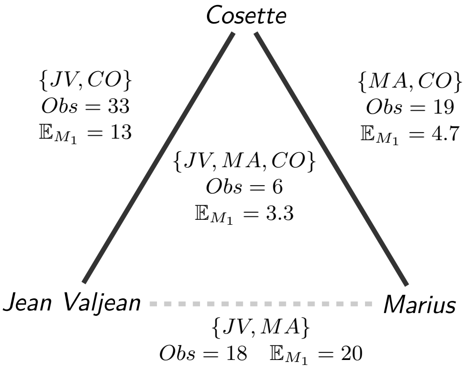
$$\text{Model 1: } \lambda_S(t) \quad = \quad \lambda_0(t) \exp[\beta_1 \cdot \text{node.activity}(S,t)]$$



Data compiled by Donald Knuth: 80 actors in 288 chapters/hyperevents.

JV: 113 chapters; MA: 77 chapters; CO: 55 chapters

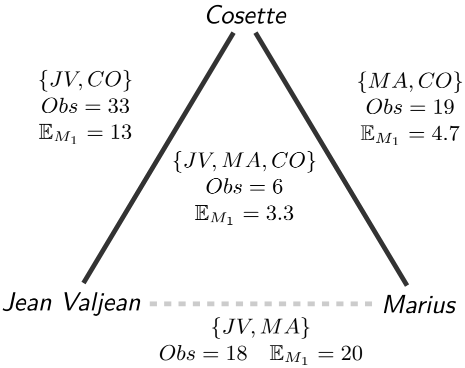
$$\text{Model 1: } \lambda_S(t) \quad = \quad \lambda_0(t) \exp[\beta_1 \cdot \text{node.activity}(S,t)]$$



Data compiled by Donald Knuth: 80 actors in 288 chapters/hyperevents.

JV: 113 chapters; MA: 77 chapters; CO: 55 chapters

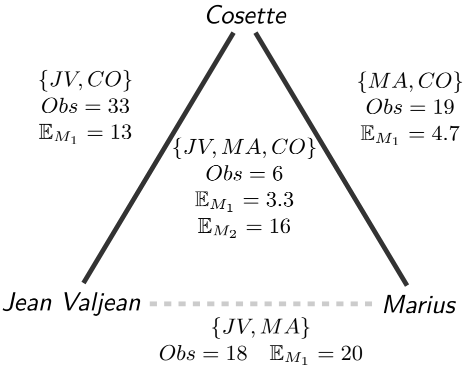
$$\text{Model 1: } \lambda_S(t) = \lambda_0(t) \exp[\beta_1 \cdot \text{node.activity}(S,t)]$$
$$\text{Model 2: } \lambda_S(t) = \lambda_0(t) \exp[\beta_1 \cdot \text{node.act}(S,t) + \beta_2 \cdot \text{dyad.act}(S,t)]$$



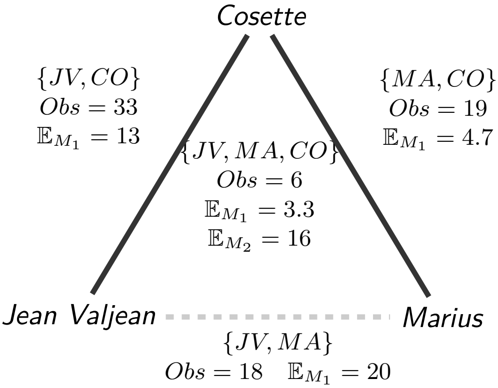
Data compiled by Donald Knuth: 80 actors in 288 chapters/hyperevents.

JV: 113 chapters; MA: 77 chapters; CO: 55 chapters

$$\text{Model 1: } \lambda_S(t) = \lambda_0(t) \exp[\beta_1 \cdot \text{node.activity}(S,t)]$$
$$\text{Model 2: } \lambda_S(t) = \lambda_0(t) \exp[\beta_1 \cdot \text{node.act}(S,t) + \beta_2 \cdot \text{dyad.act}(S,t)]$$



$$\begin{aligned}\text{Model 1: } \lambda_S(t) &= \lambda_0(t)\exp[\beta_1 \cdot \mathsf{node.activity}(S,t)] \\ \text{Model 2: } \lambda_S(t) &= \lambda_0(t)\exp[\beta_1 \cdot \mathsf{node.act}(S,t) + \beta_2 \cdot \mathsf{dyad.act}(S,t)]\end{aligned}$$



Model 1: $\{JV, MA, CO\}$ is an over-represented triad.

Model 2: $\{JV, MA, CO\}$ is an under-represented triad.

end excursus.

Practical aspects

What do we have to do to specify a RHEM?

Workflow of eventnet.

https://github.com/juergenlerner/eventnet

Computes specified statistics for events and non-events; models can be fitted with standard software (More about this in Section 3...).

observed events

e_1	A	t_1
e_1	B	t_1
e_1	C	t_1
e_2	D	t_2
e_2	E	t_2
...		

eventnet

configuration (file)
determines how statistics are computed from input events (reproducibility)

statistics of events and non-events

e_1	$A\|B\|C$	t_1	1	2.4	0.7
e_1	$A\|B\|D$	t_1	0	0.1	1.5
e_1	$A\|B\|E$	t_1	0	0.5	1.9
e_1	$C\|D\|E$	t_1	0	0.3	0.9
e_2	$D\|E$	t_2	1	1.9	0.1
e_2	$A\|B$	t_2	0	2.1	1.9
e_2	$C\|E$	t_2	0	0.7.1	0.2
...					

Configurations can be specified with graphical user interface or via configuration file (XML).

Eventnet workflow – more details.

past events

A, B, C meet at t_1

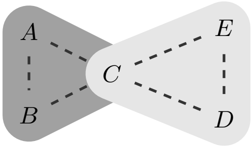
C, D, E meet at t_2

...

all events

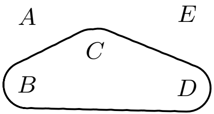
before t

hyperedge attributes



network effects

hyperedge statistics



partial rep, closure, ...

specify

prob of next event(s)

do B, C, D meet at t?

param. function of statistics

Computer Practical 1

Computer Practical 1 demonstrates the computation of the covariates using `eventnet`.

- Format of the input data.
- CSV and time settings.
- Specifying network attributes and statistics.
- Specifying the observed events, risk set, and sampling.

RHEM Inference

Descending into Dante's Inferno...

Abandon hope all ye who enter here...

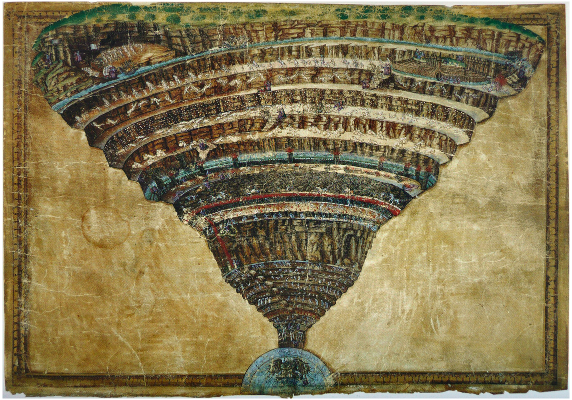
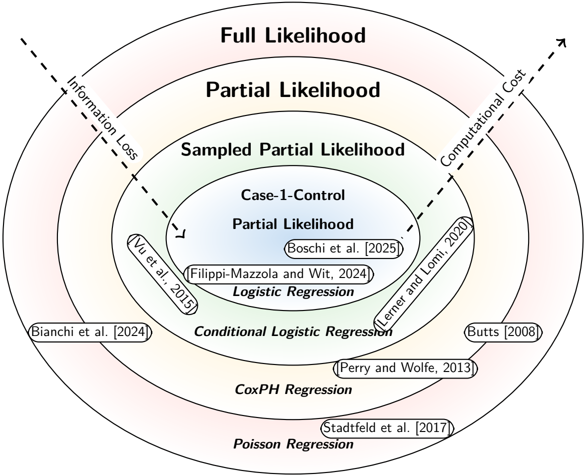


Figure: Watts, Barbara J.: "Sandro Botticelli's Drawings for Dante's Inferno"

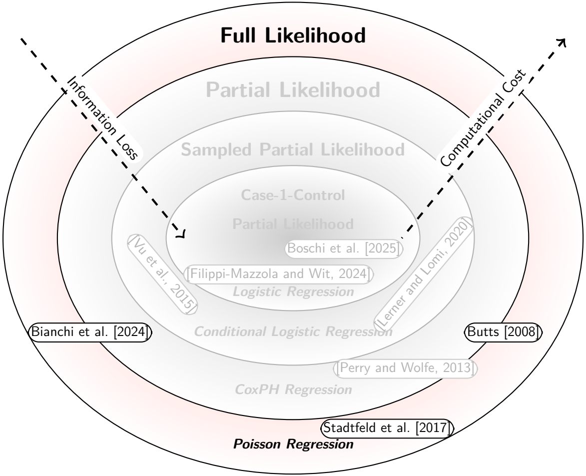
...or through rings of Likelihood?

Maintain ~~Abandon~~ hope all ye who enter here...



Full Likelihood Inference

Let's start from the beginning...



A quick recap!

Relational Data: stream of observed (hyper) events.

$$E = \{(S_i, R_i, t_i), i = 1, ..., n\}$$

REMARK: Dyadic events are a special case with one sender and one receiver:
$S_i = \{s_i\}, R_i = \{r_i\}$ for all i.

Parametric Model: linear (in log-hazard)... just for now!

$$\lambda_{SR}(t) = Y_{SR}(t) \times \lambda_0(t) \times \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{SR}(t)\}$$

Likelihood: constructed as joint probability of observing **data** under **model**.

Full Likelihood: Let's follow the definition...

Likelihood: constructed as joint probability of observing **data** under **model**.

$$\mathcal{L}(\boldsymbol{\beta}) = p\left(E; \boldsymbol{\beta}\right) \overbrace{\underbrace{=}_{\textbf{Multiplication Theorem}}}^{\text{Perry and Wolfe [2013]}} \prod_{i=1}^{n} \left[p\Big(t_i | \underbrace{t_1, (s_1, r_1), ..., t_{i-1}, (s_{i-1}, r_{i-1})}_{\text{previous events}}; \boldsymbol{\beta}\Big) \times \right.$$

$$\left. p\Big((s_i, r_i) | t_i, \underbrace{t_1, (s_1, r_1), ..., t_{i-1}, (s_{i-1}, r_{i-1})}_{\text{previous events}}; \boldsymbol{\beta}\Big) \right]$$

events are independent **conditional on history**

along with covariates, incorporated by $\mathbb{W} = \{\mathcal{W}_t\}_{t \in \mathbb{R}^+}$

$$= \prod_{i=1}^{n} p(t_i | \mathcal{W}_{t_i^-}; \boldsymbol{\beta}) p((s_i, r_i) | t_i, \mathcal{W}_{t_i^-}; \boldsymbol{\beta})$$

Full Likelihood [Bianchi et al., 2024]

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} p(t_i|\mathcal{W}_{t_i^-}) \times p((s_i, r_i)|t_i, \mathcal{W}_{t_i^-}; \boldsymbol{\beta})$$

How to account for fact that there are not other dyads observed in (t_{i-1}, t_i)?

$$\prod_{i=1}^{n} \underbrace{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i; \boldsymbol{\beta}) \exp\left[-\sum_{(s,r)\in\mathcal{R}_{t_i}} \int_{t_{i-1}}^{t_i} \lambda_{sr}(u; \boldsymbol{\beta}) \mathrm{d}u\right]}_{\textbf{Minimum Generalized Exponential Inter-arrival Time}} \times \underbrace{p((s_i, r_i)|t_i, \mathcal{W}_{t_i^-}; \boldsymbol{\beta})}_{\substack{\text{Multinomial Probability} \\ \frac{\lambda_{s_i r_i}(t_i; \boldsymbol{\beta})}{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i; \boldsymbol{\beta})}}}$$

Full Likelihood for... Piecewise Constant Rates (PCR) I

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \underbrace{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta}) \exp\left[-\sum_{(s,r)\in\mathcal{R}_{t_i}} \int_{t_{i-1}}^{t_i} \lambda_{sr}(u;\boldsymbol{\beta})\mathrm{d}u\right]}_{\text{Minimum Generalized Exponential Inter-arrival Time}} \times \underbrace{p((s_i, r_i)|t_i, \mathcal{W}_{t_i^-}; \boldsymbol{\beta})}_{\text{Multinomial Probability}}$$

$$\frac{\lambda_{s_i r_i}(t_i;\boldsymbol{\beta})}{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})}$$

What if hazard is assumed to be constant within inter-arrival times?

$$\prod_{i=1}^{n} \underbrace{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta}) \times \exp\left[-\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})(t_i - t_{i-1})\right]}_{\textbf{Exponential waiting time } \Delta T_i = T_i - T_{i-1} \textbf{ - Poisson N. of Events in } \Delta T_i} \times \frac{\lambda_{s_i r_i}(t_i;\boldsymbol{\beta})}{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})}$$

Full Likelihood for... Piecewise Constant Rates (PCR) II

$$\mathcal{L}(\boldsymbol{\beta}) \overset{\text{PCR}}{\overbrace{=}} \prod_{i=1}^{n} \cancel{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})} \times \exp\left[-\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})(t_i - t_{i-1})\right] \frac{\lambda_{s_i r_i}(t_i;\boldsymbol{\beta})}{\cancel{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})}}$$

$$\prod_{i=1}^{n} \lambda_{s_i r_i}(t_i;\boldsymbol{\beta}) \prod_{(s,r)\in\mathcal{R}_{t_i}} \exp\left[-\lambda_{sr}(t_i;\boldsymbol{\beta})\,(t_i - t_{i-1})\right]$$

Full Likelihood for... Piecewise Constant Rates (PCR) II

$$\mathcal{L}(\boldsymbol{\beta}) \overbrace{=}^{\text{PCR}} \prod_{i=1}^{n} \cancel{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})} \times \exp\left[-\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})(t_i - t_{i-1})\right] \frac{\lambda_{s_i r_i}(t_i;\boldsymbol{\beta})}{\cancel{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})}}$$

$$\prod_{i=1}^{n} \lambda_{s_i r_i}(t_i;\boldsymbol{\beta}) \prod_{(s,r)\in\mathcal{R}_{t_i}} \exp\left[-\lambda_{sr}(t_i;\boldsymbol{\beta})\,(t_i - t_{i-1})\right]$$

Call $\Delta N_{sr}(t_i) = N_{sr(t_i)} - N_{sr(t_{i-1})}$

$$\underbrace{\prod_{i=1}^{n} \prod_{(s,r)\in\mathcal{R}_{t_i}} \left(\lambda_{sr}(t_i;\boldsymbol{\beta})\right)^{\Delta N_{sr}(t_i)} \exp\left[-\lambda_{sr}(t_i;\boldsymbol{\beta})\,(t_i - t_{i-1})\right]}_{\text{Proportional to likelihood of a Poisson Regression}}$$

So, which statistical model should we focus on? (Pt. 1)

Full Log-Likelihood (Piecewise Constant Rates):

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{(s,r)\in\mathcal{R}_{t_i}} \Delta N_{sr}(t_i) \log\left(\lambda_{sr}(t_i;\boldsymbol{\beta})\right) - \lambda_{sr}(t_i;\boldsymbol{\beta})(t_i - t_{i-1})$$

Modeling:	**Inference**:
Relational Event Model	**Poisson Regression**
$\lambda_{sr}(t) = W_{sr}(t) \times \lambda_0(t) \times \exp\left\{\boldsymbol{\beta}'\boldsymbol{x}_{sr}(t)\right\}$	$\Delta N_{sr}(t_i) \mid \boldsymbol{x}_{sr}(t_i) \overset{\text{iid}}{\sim} \text{Poisson}\left(\mu_{sr}(t_i)\right)$ $\log(\mu_{sr}(t_i)) = \boldsymbol{\beta}'\boldsymbol{x}_{sr}(t_i) + \log(t_i - t_{i-1})$

Nice... But at what price?

1. Strong Assumption;
2. High Computational Cost;

So, which statistical model should we focus on? (Pt. 1)

Full Log-Likelihood (Piecewise Constant Rates):

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{(s,r)\in\mathcal{R}_{t_i}} \Delta N_{sr}(t_i) \log\left(\lambda_{sr}(t_i;\boldsymbol{\beta})\right) - \lambda_{sr}(t_i;\boldsymbol{\beta})(t_i - t_{i-1})$$

Modeling:	**Inference**:
Relational Event Model	**Poisson Regression**
$\lambda_{sr}(t) = W_{sr}(t) \times \lambda_0(t) \times \exp\left\{\boldsymbol{\beta}'\boldsymbol{x}_{sr}(t)\right\}$	$\Delta N_{sr}(t_i) \vert \boldsymbol{x}_{sr}(t_i) \overset{\text{iid}}{\sim} \text{Poisson}\left(\mu_{sr}(t_i)\right)$ $\log(\mu_{sr}(t_i)) = \boldsymbol{\beta}'\boldsymbol{x}_{sr}(t_i) + \log(t_i - t_{i-1})$

Nice... But at what price?

1. Strong Assumption;
2. High Computational Cost;

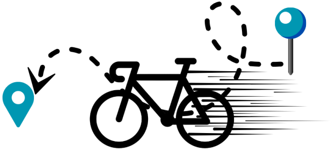
Full Likelihood in Practice I

Full Log-Likelihood (Piecewise Constant Rates):

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{(s,r)\in\mathcal{R}_{t_i}} \Delta N_{sr}(t_i) \log\left(\lambda_{sr}(t_i;\boldsymbol{\beta})\right) - \lambda_{sr}(t_i;\boldsymbol{\beta})(t_i - t_{i-1})$$

Peek-a-Data!

```
> head(poisson_dat, 10)
   sender receiver dist_min  dist   PT_dist random_intercept_s info event  start        stop log_dist     logtm
1   33202    31261     13.3  2766 3.2714157               3.46   80     1 0.0074 0.007485937 1.326013 -9.361895
2   31937    31937      0.0     0 1.7116404              -0.14    1     0 0.0074 0.007485937 0.000000 -9.361895
3   33200    31937     19.0  5258 4.1927036               0.85    2     0 0.0074 0.007485937 1.833861 -9.361895
4   32604    31937     38.9  9260 6.2539595              -7.03    3     0 0.0074 0.007485937 2.328253 -9.361895
5   31261    31937     17.6  4158 1.4511050               1.06    4     0 0.0074 0.007485937 1.640549 -9.361895
6   31246    31937     11.3  3333 1.4195458              -2.21    5     0 0.0074 0.007485937 1.466260 -9.361895
7   31285    31937     15.9  4465 2.3303198               0.46    6     0 0.0074 0.007485937 1.698364 -9.361895
8   31297    31937     13.7  3982 0.7931991               1.40    7     0 0.0074 0.007485937 1.605831 -9.361895
9   31218    31937     27.2  6001 1.9189675              -4.05    8     0 0.0074 0.007485937 1.946053 -9.361895
10  31966    31937     45.9 11088 1.8293264              -1.11    9     0 0.0074 0.007485937 2.492213 -9.361895
```

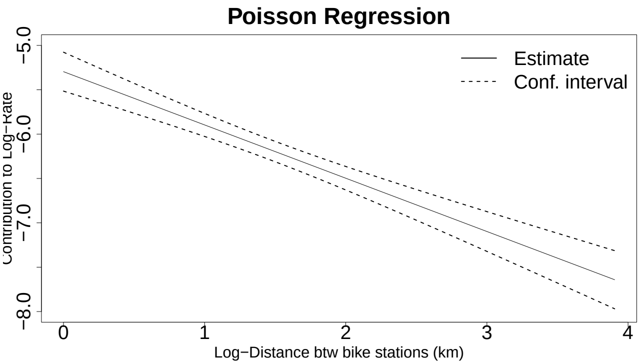


R Tool!

```
glm(event ~ log_dist + offset(logtm),
family = poisson(link = "log"), data = poisson_dat)
```

Full Likelihood in Practice II

Spoiler Alert!



Full Likelihood [Bianchi et al., 2024]

Full Likelihood via Poisson Regression... At what price?

1. Strong Assumption;
2. High Computational Cost;

Can we do better? Let's do one step back:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta}) \exp\left[-\sum_{(s,r)\in\mathcal{R}_{t_i}} \int_{t_{i-1}}^{t_i} \lambda_{sr}(u;\boldsymbol{\beta})\mathrm{d}u\right] \times \frac{\lambda_{s_i r_i}(t_i;\boldsymbol{\beta})}{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})}$$

Full Likelihood [Bianchi et al., 2024]

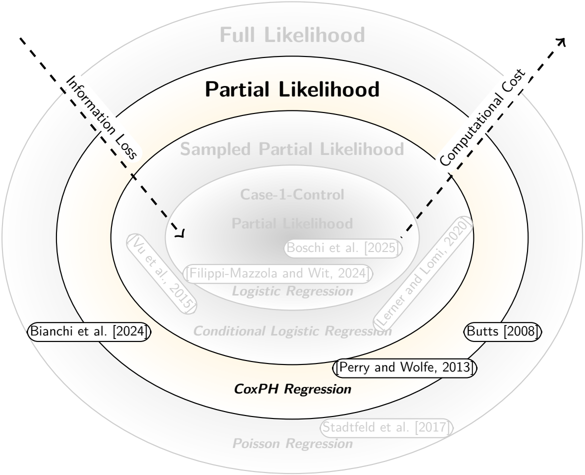
Full Likelihood via Poisson Regression... At what price?

1. Strong Assumption;
2. High Computational Cost;

Can we do better? Let's do one step back:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta}) \exp\left[-\sum_{(s,r)\in\mathcal{R}_{t_i}} \int_{t_{i-1}}^{t_i} \lambda_{sr}(u;\boldsymbol{\beta})\mathrm{d}u\right] \times \frac{\lambda_{s_i r_i}(t_i;\boldsymbol{\beta})}{\sum_{(s,r)\in\mathcal{R}_{t_i}} \lambda_{sr}(t_i;\boldsymbol{\beta})}$$

Partial Likelihood Inference



Inspired by Cox [1975]: REM's Partial Likelihood I

Partial likelihood focuses on **multinomial probability of observing a specific dyad among those at risk, given that an event has occurred**.

$$\mathcal{L}^P(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\lambda_{s_i r_i}(t_i; \mathcal{W}_{t_i^-}; \boldsymbol{\beta})}{\sum_{(s,r)\in \mathcal{R}_{t_i}} \lambda_{sr}(t_i; \mathcal{W}_{t_i^-}; \boldsymbol{\beta})} = \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}' \boldsymbol{x}_{s_i r_i}(t_i)\}}{\sum_{(s,r)\in \mathcal{R}_{t_i}} \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t_i)\}}$$

Inspired by Cox [1975]: REM's Partial Likelihood II

Partial likelihood focuses on **multinomial probability of observing a specific dyad among those at risk, given that an event has occurred**.

$$\mathcal{L}^P(\boldsymbol{\beta}) = \frac{\lambda_{s_i r_i}(t_i; \mathcal{W}_{t_i^-}; \boldsymbol{\beta})}{\sum_{(s,r)\in \mathcal{R}_{t_i}} \lambda_{sr}(t_i; \mathcal{W}_{t_i^-}; \boldsymbol{\beta})} = \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}' \boldsymbol{x}_{s_i r_i}(t_i)\}}{\sum_{(s,r)\in \mathcal{R}_{t_i}} \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t_i)\}}$$

Information Loss:

Leaving $\lambda_0(t)$ arbitrary results in **slight loss of information** about $\boldsymbol{\beta}$.

Dimensionality Reduction:

effective when providing a **simplification** of full likelihood by **eliminating nuisance parameters**.

Event Ordering:

Applicable when **only order of events is known**, not their exact timing.

So, which statistical model should we focus on? (Pt. 2)

Partial Log-likelihood:

$$\ell^P(\boldsymbol{\beta}) = \log \mathcal{L}^P(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\boldsymbol{\beta}' \boldsymbol{x}_{s_i r_i}(t_i) - \log \left(\sum_{(s,r)\in\mathcal{R}_{t_i}} \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t_i)\} \right) \right],$$

Modeling:	**Inference**:
Relational Event Model	**CoxPH Regression**
$\lambda_{sr}(t) = W_{sr}(t) \times \lambda_0(t) \times \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t)\}$	$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ell^P(\boldsymbol{\beta})$

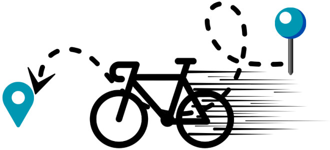
Partial Likelihood in Practice I

Partial Log-Likelihood:

$$\ell^P(\boldsymbol{\beta}) = \log \mathcal{L}^P(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\boldsymbol{\beta}' \boldsymbol{x}_{s_i r_i}(t_i) - \log \left(\sum_{(s,r) \in \mathcal{R}_{t_i}} \exp\left\{ \boldsymbol{\beta}' \boldsymbol{x}_{sr}(t_i) \right\} \right) \right],$$

Peek-a-Data!

```
> head(cox_dat, 10)
   sender receiver dist_min  dist   PT_dist random_intercept_s info event  start        stop log_dist
1   33202    31261     13.3  2766 3.2714157               3.46   80     1 0.0074 0.007485937 1.326013
2   31937    31937      0.0     0 1.7116404              -0.14    1     0 0.0074 0.007485937 0.000000
3   33200    31937     19.0  5258 4.1927036               0.85    2     0 0.0074 0.007485937 1.833861
4   32604    31937     38.9  9260 6.2539595              -7.03    3     0 0.0074 0.007485937 2.328253
5   31261    31937     17.6  4158 1.4511050               1.06    4     0 0.0074 0.007485937 1.640549
6   31246    31937     11.3  3333 1.4195458              -2.21    5     0 0.0074 0.007485937 1.466260
7   31285    31937     15.9  4465 2.3303198               0.46    6     0 0.0074 0.007485937 1.698364
8   31297    31937     13.7  3982 0.7931991               1.40    7     0 0.0074 0.007485937 1.605831
9   31218    31937     27.2  6001 1.9189675              -4.05    8     0 0.0074 0.007485937 1.946053
10  31966    31937     45.9 11088 1.8293264              -1.11    9     0 0.0074 0.007485937 2.492213
```

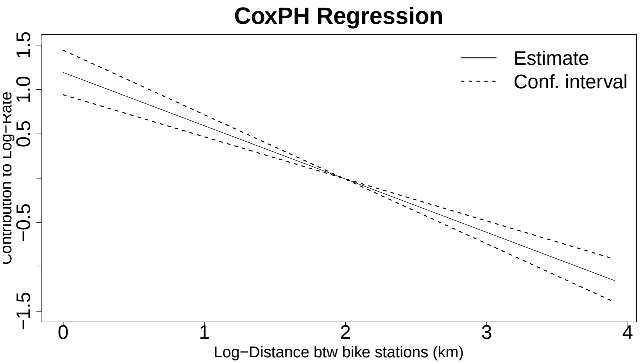


R Tool!

```
coxph(Surv(start,stop,event) ~ log_dist, data = cox_dat
```

Partial Likelihood in Practice II

Spoiler Alert!



Partial Likelihood in [Bianchi et al., 2024]

Full Likelihood via Poisson Regression... At what price?

1. ~~Strong Assumption~~;
2. High Computational Cost;

Can we do better? Let's do one step back:

$$\mathcal{L}^P(\boldsymbol{\beta}) = \frac{\lambda_{s_i r_i}(t_i; \mathcal{W}_{t_i^-}; \boldsymbol{\beta})}{\sum_{(s,r)\in \mathcal{R}_{t_i}} \lambda_{sr}(t_i; \mathcal{W}_{t_i^-}; \boldsymbol{\beta})} = \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}' \boldsymbol{x}_{s_i r_i}(t_i)\}}{\sum_{(s,r)\in \mathcal{R}_{t_i}} \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t_i)\}}$$

Partial Likelihood in [Bianchi et al., 2024]

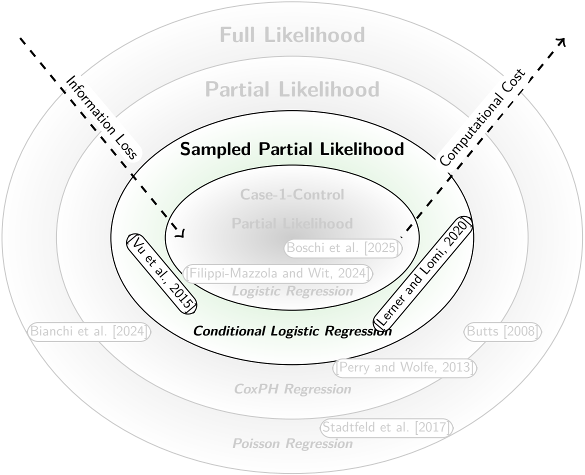
Full Likelihood via Poisson Regression... At what price?

1. ~~Strong Assumption~~;
2. High Computational Cost;

Can we do better? Let's do one step back:

$$\mathcal{L}^P(\boldsymbol{\beta}) = \frac{\lambda_{s_i r_i}(t_i; \mathcal{W}_{t_i^-}; \boldsymbol{\beta})}{\sum_{(s,r)\in \mathcal{R}_{t_i}} \lambda_{sr}(t_i; \mathcal{W}_{t_i^-}; \boldsymbol{\beta})} = \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}' \boldsymbol{x}_{s_i r_i}(t_i)\}}{\sum_{(s,r)\in \mathcal{R}_{t_i}} \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t_i)\}}$$

Sampled Partial Likelihood Inference



What changes when we sample? I

Relational Data: stream of observed events and sampled risk sets.

$$E = \{(s_i, r_i, t_i, \mathcal{SR}_{t_i}), i = 1, ..., n\},$$

where $\mathcal{SR}_{t_i} \subset \mathcal{R}_{t_i}$ is **sampled risk set** at time t_i.

Updated Filtration: incorporating history of events and sampling information.

$$\{\mathcal{F}_t\}_{t \geq 0}, \qquad \mathcal{F}_t = \mathcal{W}_t \cup \sigma(\mathcal{SR}_{t_i}; t_i \leq t)$$

Parametric Model: let's stay linear for the moment...

$$\lambda_{sr, \mathcal{SR}}(t) = \lambda_{sr}(t) \cdot \pi_t(\mathcal{SR}|(s,r)), \qquad \pi_t(\mathcal{SR}|(s,r)) = \frac{m}{|\mathcal{R}_t| - 1} \cdot 1_{\{(s,r) \in \mathcal{SR}, \mathcal{SR} \in \mathcal{R}_t\}}$$

m is the **number of sampled non-events** according to **nested case-control sampling**.

What changes when we sample? I

Relational Data: stream of observed events and sampled risk sets.

$$E = \{(s_i, r_i, t_i, \mathcal{SR}_{t_i}), i = 1, ..., n\},$$

where $\mathcal{SR}_{t_i} \subset \mathcal{R}_{t_i}$ is **sampled risk set** at time t_i.

Updated Filtration: incorporating history of events and sampling information.

$$\{\mathcal{F}_t\}_{t\geq 0}, \qquad \mathcal{F}_t = \mathcal{W}_t \cup \sigma(\mathcal{SR}_{t_i}; t_i \leq t)$$

Parametric Model: let's stay linear for the moment...

$$\lambda_{sr,\mathcal{SR}}(t) = \lambda_{sr}(t) \cdot \pi_t(\mathcal{SR}|(s,r)), \qquad \pi_t(\mathcal{SR}|(s,r)) = \frac{m}{|\mathcal{R}_t| - 1} \cdot 1_{\{(s,r)\in\mathcal{SR}, \mathcal{SR}\in\mathcal{R}_t\}}$$

m is the **number of sampled non-events** according to **nested case-control sampling**.

What changes when we sample? I

Relational Data: stream of observed events and sampled risk sets.

$$E = \{(s_i, r_i, t_i, \mathcal{SR}_{t_i}), i = 1, ..., n\},$$

where $\mathcal{SR}_{t_i} \subset \mathcal{R}_{t_i}$ is **sampled risk set** at time t_i.

Updated Filtration: incorporating history of events and sampling information.

$$\{\mathcal{F}_t\}_{t \geq 0}, \qquad \mathcal{F}_t = \mathcal{W}_t \cup \sigma(\mathcal{SR}_{t_i}; t_i \leq t)$$

Parametric Model: let's stay linear for the moment...

$$\lambda_{sr, \mathcal{SR}}(t) = \lambda_{sr}(t) \cdot \pi_t(\mathcal{SR} | (s, r)), \qquad \pi_t(\mathcal{SR} | (s, r)) = \frac{m}{|\mathcal{R}_t| - 1} \cdot 1_{\{(s,r) \in \mathcal{SR}, \mathcal{SR} \in \mathcal{R}_t\}}$$

m is the **number of sampled non-events** according to **nested case-control sampling**.

What changes when we sample? II

Sampled Partial likelihood:

$$\mathcal{L}^S(\boldsymbol{\beta}) = \frac{\lambda_{s_i r_i}(t_i; \mathcal{F}_{t_i^-}; \boldsymbol{\beta}) \cdot \cancel{\pi_{t_i}(\mathcal{SR}_{t_i} | (s_i, r_i))}}{\sum_{(s,r) \in \mathcal{SR}_{t_i}} \lambda_{sr}(t_i; \mathcal{F}_{t_i^-}; \boldsymbol{\beta}) \cdot \cancel{\pi_{t_i}(\mathcal{SR}_{t_i} | (s, r))}} = \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}' \boldsymbol{x}_{s_i r_i}(t_i)\}}{\sum_{(s,r) \in \mathcal{SR}_{t_i}} \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t_i)\}}$$

Look Familiar?

Conditional logistic regression

1 language

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

Conditional logistic regression is an extension of logistic regression that allows one to account for stratification and matching. Its main field of application is observational studies and in particular epidemiology. It was devised in 1978 by Norman Breslow, Nicholas Day, Katherine Halvorsen, Ross L. Prentice and C. Sabai.[1] It is the most flexible and general procedure for matched data.

Background [edit]

Observational studies use stratification or matching as a way to control for confounding.

Logistic regression can account for stratification by having a different constant term for each stratum. Let us denote $Y_{i\ell} \in \{0, 1\}$ the label (e.g. case status) of the ℓth observation of the ith stratum and $X_{i\ell} \in \mathbb{R}^p$ the values of the corresponding predictors. We then take the likelihood of one observation to be

$$\mathbb{P}(Y_{i\ell} = 1 | X_{i\ell}) = \frac{\exp(\alpha_i + \boldsymbol{\beta}^\top X_{i\ell})}{1 + \exp(\alpha_i + \boldsymbol{\beta}^\top X_{i\ell})}$$

Figure: Source: Wikipedia

Stratification...

with respect to what?

By case!

So, which statistical model should we focus on? (Pt. 3)

Sampled Partial likelihood:

$$\mathcal{L}^S(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i r_i}(t_i)\}}{\sum_{(s,r)\in \boxed{\mathcal{SR}_{t_i}}} \exp\{\boldsymbol{\beta}'\boldsymbol{x}_{sr}(t_i)\}}$$

Modeling:

Relational Event Model

$$\lambda_{sr}(t) = W_{sr}(t) \times \lambda_0(t) \times \exp\{\boldsymbol{\beta}'\boldsymbol{x}_{sr}(t)\}$$

Inference:

Conditional Logistic Regression

$$\Delta N_{sr}(t_i) | \sum_{s'r' \in \mathcal{SR}_i} \Delta N_{s'r'}(t_i) = 1 \sim \text{Categorical}\,(\pi_{sr}(t_i))$$

$$\pi_{sr}(t_i) = \frac{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{sr}(t_i)\}}{\sum_{(s',r')\in \boxed{\mathcal{SR}_{t_i}}} \exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s'r'}(t_i)\}}$$

Sampled Partial Likelihood in Practice I

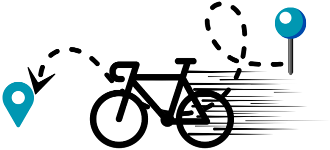
Sampled Partial Log-Likelihood:

$$\ell^S(\boldsymbol{\beta}) = \log \mathcal{L}^S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[\boldsymbol{\beta}' \boldsymbol{x}_{s_i r_i}(t_i) - \log \left(\sum_{(s,r) \in \mathcal{SR}_{t_i}} \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t_i)\} \right) \right],$$

Peek-a-Data!

```
> head(sampled_dat, 10)
```

	sender	receiver	dist_min	dist	PT_dist	random_intercept_s	info	event	start	stop	log_dist
1	33202	31261	13.3	2766	3.2714157	3.46	80	1	0.0074	0.007485937	1.326013
2	31297	31218	15.9	3630	0.7931991	1.40	147	0	0.0074	0.007485937	1.532557
3	31255	31419	19.5	4453	6.7501025	-0.15	318	0	0.0074	0.007485937	1.696166
4	31071	32604	50.3	12994	1.6706971	3.96	54	0	0.0074	0.007485937	2.638629
5	31272	31071	34.6	8183	2.2255955	0.14	277	0	0.0074	0.007485937	2.217354
6	32217	31071	124.3	33190	3.8203727	-3.86	279	0	0.0074	0.007485937	3.531933
7	31324	31262	11.2	2435	1.3001624	6.14	250	1	0.0132	0.013290035	1.234017
8	32217	31297	117.9	44455	3.8203727	-3.86	139	0	0.0132	0.013290035	3.816723
9	31255	32604	49.4	12196	6.7501025	-0.15	58	0	0.0132	0.013290035	2.579914
10	32604	31218	63.9	15186	6.2539595	-7.03	143	0	0.0132	0.013290035	2.784147

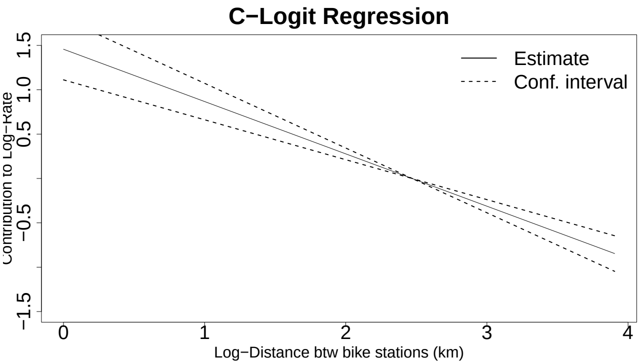


R Tool!

```
clogit(event ∼ log_dist + strata(stop), data = sampled_dat)
```

Sampled Partial Likelihood in Practice II

Spoiler Alert!



Who first used this approach in REMs?

Social Networks
Volume 43, October 2015, Pages 121-135

Relational event models for social learning in MOOCs

Duy Vu [a], Philippa Pattison [b], Garry Robins [a]

Network Science

Reliability of relational event model estimates under sampling: How to fit a relational event model to 360 million dyadic events

Published online by Cambridge University Press: **22 November 2019**

Jürgen Lerner and Alessandro Lomi

Show author details

Article Metrics

Towards Case-control sampling via GLM

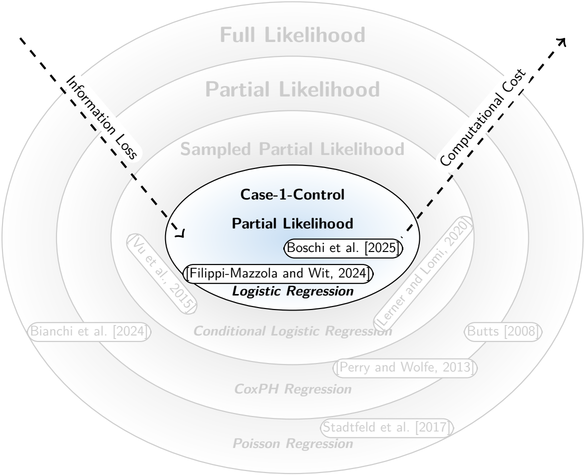
Full Likelihood via Poisson Regression... At what price?

1. ~~Strong Assumption~~;
2. ~~High Computational Cost~~;

Well, we can we even better!

Case-1-control Sampling Inference

Can we do more?



How to get a Logistic Regression... I

Sampled Partial likelihood with $m = 1$:

$$\mathcal{L}^S(\boldsymbol{\beta}) \overset{m=1}{=} \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i r_i}(t_i)\}}{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i r_i}(t_i)\} + \exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i^* r_i^*}(t_i)\}}$$

How to get a Logistic Regression... II

Sampled Partial likelihood with $m = 1$:

$$\mathcal{L}^S(\boldsymbol{\beta}) \overset{m=1}{=} \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i r_i}(t_i)\}}{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i r_i}(t_i)\} + \exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i^* r_i^*}(t_i)\}}$$

$$= \prod_{i=1}^{n} \frac{\dfrac{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i r_i}(t_i)\}}{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i^* r_i^*}(t_i)\}}}{\dfrac{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i r_i}(t_i)\}}{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i^* r_i^*}(t_i)\}} + \dfrac{\cancel{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i^* r_i^*}(t_i)\}}}{\cancel{\exp\{\boldsymbol{\beta}'\boldsymbol{x}_{s_i^* r_i^*}(t_i)\}}}} = \prod_{i=1}^{n} \frac{\exp\{\boldsymbol{\beta}'\left(\boldsymbol{x}_{s_i r_i}(t_i) - \boldsymbol{x}_{s_i^* r_i^*}(t_i)\right)\}}{1 + \exp\{\boldsymbol{\beta}'\left(\boldsymbol{x}_{s_i r_i}(t_i) - \boldsymbol{x}_{s_i^* r_i^*}(t_i)\right)\}}$$

$$= \prod_{i=1}^{n} \text{logistic}(\boldsymbol{\beta}'\Delta\boldsymbol{x}_i) \quad \Delta\boldsymbol{x}_i = \boldsymbol{x}_{s_i r_i}(t_i) - \boldsymbol{x}_{s_i^* r_i^*}(t_i)$$

So, which statistical model should we focus on? (Pt. 4)

Sampled Partial likelihood with $m = 1$:

$$\mathcal{L}^S(\boldsymbol{\beta}) \overset{m=1}{=} \prod_{i=1}^{n} \text{logistic}(\boldsymbol{\beta}' \Delta \boldsymbol{x}_i) \quad \Delta \boldsymbol{x}_i = \boldsymbol{x}_{s_i r_i}(t_i) - \boldsymbol{x}_{s_i^* r_i^*}(t_i)$$

Modeling:

Relational Event Model

$$\lambda_{sr}(t) = W_{sr}(t) \times \lambda_0(t) \times \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t)\}$$

Inference:

Logistic Regression

$$Y_i | \Delta \boldsymbol{x}_i \overset{\text{iid}}{\sim} \text{Bernoulli}(\pi_i)$$

$$\text{logit}(\pi_i) = \boldsymbol{\beta}' \Delta \boldsymbol{x}_i$$

$y_i = 1 \quad i = 1, \ldots, n;$ no intercept

Wondering what advantage is? Stay tuned for Melania's presentation!

SPOILER ALERT: It's all about flexibility!

So, which statistical model should we focus on? (Pt. 4)

Sampled Partial likelihood with $m = 1$:

$$\mathcal{L}^S(\boldsymbol{\beta}) \overset{m=1}{=} \prod_{i=1}^{n} \text{logistic}(\boldsymbol{\beta}' \Delta \boldsymbol{x}_i) \quad \Delta \boldsymbol{x}_i = \boldsymbol{x}_{s_i r_i}(t_i) - \boldsymbol{x}_{s_i^* r_i^*}(t_i)$$

Modeling:

Relational Event Model

$$\lambda_{sr}(t) = W_{sr}(t) \times \lambda_0(t) \times \exp\{\boldsymbol{\beta}' \boldsymbol{x}_{sr}(t)\}$$

Inference:

Logistic Regression

$$Y_i | \Delta \boldsymbol{x}_i \overset{\text{iid}}{\sim} \text{Bernoulli}(\pi_i)$$
$$\text{logit}(\pi_i) = \boldsymbol{\beta}' \Delta \boldsymbol{x}_i$$

$y_i = 1 \quad i = 1, \ldots, n;$ no intercept

Wondering what advantage is? Stay tuned for Melania's presentation!
SPOILER ALERT: It's all about flexibility!

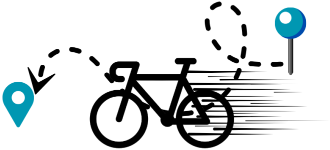
Case-1-control Sampling Inference via GLM in Practice I

Sampled Partial likelihood with $m = 1$:

$$\mathcal{L}^S(\boldsymbol{\beta}) \overset{m=1}{=} \prod_{i=1}^{n} \text{logistic}(\boldsymbol{\beta}' \Delta \boldsymbol{x}_i) \quad \Delta \boldsymbol{x}_i = \boldsymbol{x}_{s_i r_i}(t_i) - \boldsymbol{x}_{s_i^* r_i^*}(t_i)$$

Peek-a-Data!

```
> head(dat_gam, 10)
   y        time    s1    r1 log_dist1    s2    r2 log_dist2 delta_log_dist
1  1 0.007485937 33202 31261 1.3260134 33200 33202  1.229348    0.096665320
2  1 0.013290035 31324 31262 1.2340169 31966 31255  2.766130   -1.532113487
3  1 0.016123561 31261 33200 0.6333972 31272 31071  2.217354   -1.583956772
4  1 0.016833676 31218 31218 0.0000000 31218 32604  2.821557   -2.821557442
5  1 0.027780092 31285 31246 0.7788661 31297 31071  1.876101   -1.097234562
6  1 0.036779525 31272 31023 2.3787125 31966 32050  3.278728   -0.900015931
7  1 0.045849812 31218 31218 0.0000000 31272 31262  1.149305   -1.149305403
8  1 0.047972385 33202 31966 2.8733951 31255 31937  1.585350    1.288045017
9  1 0.052986957 31272 31419 1.9516082 33200 31313  1.925853    0.025754966
10 1 0.061285646 31071 31023 1.9742199 31071 31246  1.968091    0.006129007
```

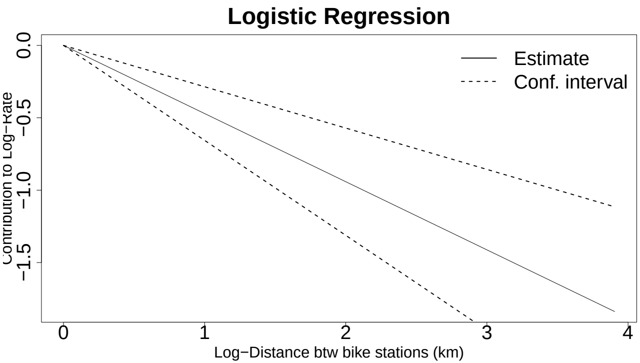


R Tool!

```
gam(y ∼ -1 + delta_log_dist,
family="binomial"(link = 'logit'), data = dat_gam)
```

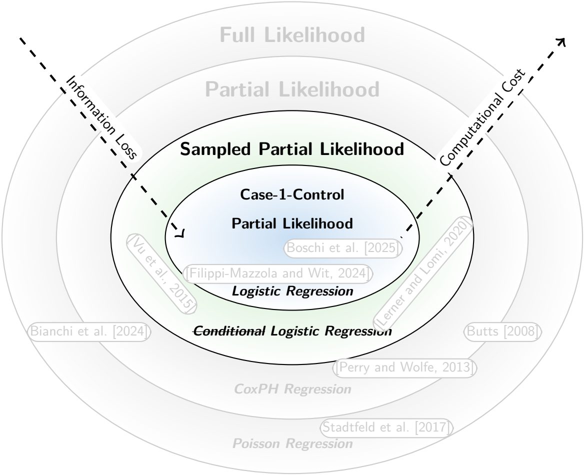
Case-1-control Sampling Inference via GLM in Practice I

Spoiler Alert!



Challenge I

Can we generalize flexible approach for any m?

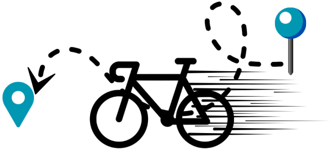


Challenge II

... Yes, we can!

This time, let's first look at the data. Peek-a-Data!

```
> head(dat_gam_m, 10)
   y event_time    s1    r1 log_dist1    s2    r2 log_dist2 delta_log_dist
1  1 0.007485937 33202 31261  1.326013 31297 31218  1.532557   -0.206543438
2  1 0.007485937 33202 31261  1.326013 31255 31419  1.696166   -0.370152486
3  1 0.007485937 33202 31261  1.326013 31071 32604  2.638629   -1.312615236
4  1 0.007485937 33202 31261  1.326013 31272 31071  2.217354   -0.891340518
5  1 0.007485937 33202 31261  1.326013 32217 31071  3.531933   -2.205919773
6  1 0.013290035 31324 31262  1.234017 32217 31297  3.816723   -2.582705900
7  1 0.013290035 31324 31262  1.234017 31255 32604  2.579914   -1.345896828
8  1 0.013290035 31324 31262  1.234017 32604 31218  2.784147   -1.550129745
9  1 0.013290035 31324 31262  1.234017 32604 31071  2.650704   -1.416686609
10 1 0.013290035 31324 31262  1.234017 31262 31297  1.243578   -0.009561133
```



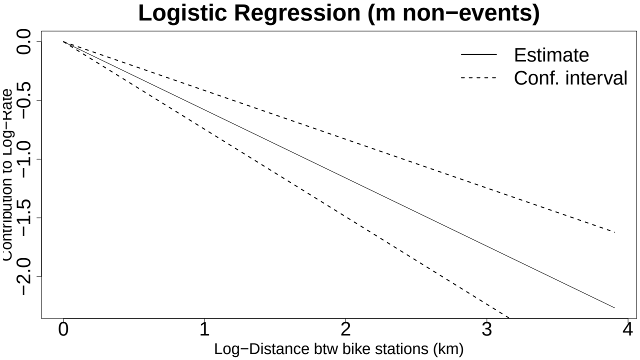
How was this dataframe created?

1. Sample m non-events for each event;
2. Compute the difference between the event and each of the related non-events.

... this means that m **rows refer to the same event**!

Case-m-control Sampling Inference via GLM in Practice

Look at the effect...



Variance Correction for Case-m-control Sampling

... but pay attention to the variance!

3 important points:

1. Increasing m reduces the variance;
2. The variance of the estimates is larger than the one obtained from `gam`;
3. ... indeed, observations within the same event are not independent!

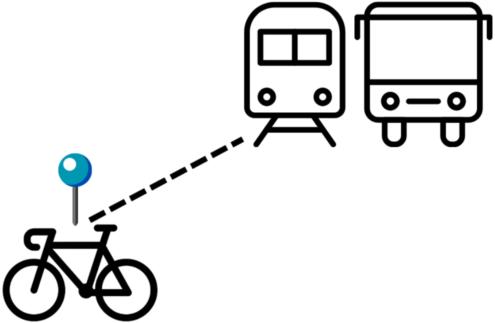
...TO BE CONTINUED...

Computer Practical 2

File _CP_2_.Rmd is currently in Folder _PRACTICAL_/_CP_2_MARTINA_.
Computer Practical 2 aims to apply the concepts reviewed in the previous section to a hyperevent dataset. The dataset has been created in Computer Practical 1.

Mixed effect additive RHEMs

All things linear is nice, right?



LET'S TALK BIKE SHARING DYNAMICS

How does distance between a bike station and a public transportation hub affect the number of bike rides? Is this effect going to be the same at 8AM or 6PM compared to 2 PM?

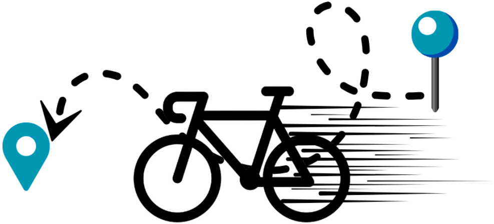
We need **time-varying effects**: the risk factor x_{sr} has **different effect across time**

$$\beta'(t)\, x_{sr}(t)$$

All things linear is nice, right?

LET'S TALK BIKE SHARING DYNAMICS

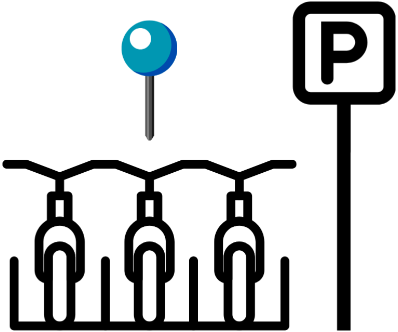
Do people bike far? Is it more likely to observe a bike ride between 2 stations that are 2km from each other or between two that are 20km?



We need non-linear effects: **different effect for different values** assumed by the risk factor x_{sr}

$$f(x_{sr}(t))$$

All things linear is nice, right?



LET'S TALK BIKE SHARING DYNAMICS

Not all bike stations are the same! Some have higher/lower usage due to aspects that have nothing to do with the predictors we account for.

We need **random effects**: accounts for the **heterogeneity** arising from the system features (e.g. bike availability, proximity to tourist attractions...)

$$\gamma' z_{sr}(t)$$

Back to the hazard

$$\lambda_{sr}(t) = \underbrace{Y_{sr}(t)}_{\text{Is } (s,r) \text{ at risk?}} \times \underbrace{\lambda_0(t)}_{\text{baseline hazard}} \times \underbrace{e^{f_{sr}(x,z)}}_{\text{edge specific risk factors}}$$

$$f_{sr}(x(t), z(t)) = \underbrace{\beta' x_{sr}^{(1)}(t)}_{\text{linear}} + \underbrace{\beta'(t) x_{sr}^{(2)}(t)}_{\text{time-varying}} + \underbrace{f(x_{sr}^{(3)}(t))}_{\text{non-linear}} + \underbrace{\gamma' z_{sr}(t)}_{\text{random}}$$

1. linear: β is independent on time and $x_{sr}^{(1)}(t)$
2. time-varying: $\beta'(t)$ is (non-linear) function of time
3. non-linear: $f(x_{sr}^{(3)}(t))$ is a (non-linear) function of $x_{sr}^{(3)}(t)$
4. random: γ is a (Gaussian) random variable

From GLMs to GAMs

Reminder!

Linear effects are fitted using a **logistic regression** with:

- only successes $(y_i = 1, i = 1, ..., n)$;
- no intercept;
- covariates: $\Delta x_i := x_{s_i r_i} - x_{s_i^* r_i^*}$;

We extend this approach by **relaxing the assumption of linearity**, we exploit **additive models** instead of linear ones

mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation

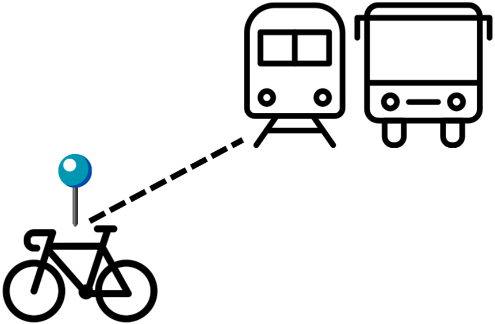
Generalized additive (mixed) models, some of their extensions and other generalized ridge regression with multiple smoothing parameter estimation by (Restricted) Marginal Likelihood, Generalized Cross Validation and similar, or using iterated nested Laplace approximation for fully Bayesian inference. See Wood (2017) <doi:10.1201/9781315370279> for an overview. Includes a gam() function, a wide variety of smoothers, 'JAGS' support and distributions beyond the exponential family.

Modeling Time-Varying Effects

Time-Varying Effect $\beta(t)$ of **distance**

of a bike station from a public transportation hub $x_{sr}^{(2)}(t)$

on bike sharing dynamics



Modeling Time-Varying Effects

- Effect of $x_{sr}^{(2)}(t)$ as $\rightarrow$ Spline function of Time

$$\beta(t) \approx \sum_{j=1}^{q} \theta_j b_j(t)$$

Choices to be made:

- **basis dimension** q
- a **smooth basis functions** $\{b_j\}_j$ e.g. *B-splines, thin plate regression splines,...*
- given range of variable, **basis evaluation points** (e.g. *uniform, quantile of a distribution,...*) to evaluate q **smooth non-linear functions** of time $b_j(t)$.

Estimating $\beta(t) \Longleftrightarrow$ Estimating θ_j-s

TVE: contribution of distance from public transport

- ... to **log-hazard of an interaction**

$$f_{sr} = \dots + \underbrace{\sum_{j=1}^{q} \theta_j b_j(t)}_{\text{time-varying effect}} \times \underbrace{x_{sr}^{(2)}(t)}_{\text{covariate}} + \dots$$

- ... to **sampled partial likelihood (m=2)**

$$\mathcal{L}^S(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{e^{\dots + \beta(t_i)[x^{(2)}_{s_i r_i}(t_i) - x^{(2)}_{s_i^* r_i^*}(t_i)] + \dots}}{1 + e^{\dots + \beta(t_i)[x^{(2)}_{s_i r_i}(t_i) - x^{(2)}_{s_i^* r_i^*}(t_i)] + \dots}}$$

$$\downarrow$$

$$\mathcal{L}^S(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{e^{\dots + \left[\sum_{j=1}^{q} \theta_j b_j(t_i)\right] \Delta x_i^{(2)} + \dots}}{1 + e^{\dots + \left[\sum_{j=1}^{q} \theta_j b_j(t_i)\right] \Delta x_i^{(2)} + \dots}} \quad (+\text{ smoothing penalty})$$

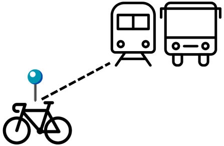
TVE: contribution of distance from public transport

- … to **sampled partial likelihood (m=2)**

$$\mathcal{L}^S(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{e^{\cdots} + \left[\sum_{j=1}^{q} \theta_j b_j(t_i)\right]\Delta x_i^{(2)} + \ldots}{1 + e^{\cdots} + \left[\sum_{j=1}^{q} \theta_j b_j(t_i)\right]\Delta x_i^{(2)} + \ldots} \text{ (+ smoothing penalty)}$$

Peek-a-Data!

```
> head(data)
  y        time    s1    r1 PT_dist1    s2    r2 PT_dist2    PT_dist
1 1 0.001451090 32217 32604 3.820373 32604 31937 6.2539595 -2.4335868
2 1 0.001584378 31937 31285 1.711640 32217 31246 3.8203727 -2.1087323
3 1 0.008190801 31324 31218 1.300162 31272 31261 2.2255955 -0.9254332
4 1 0.011731796 31313 31272 2.346278 31297 32604 0.7931991  1.5530784
5 1 0.014548734 31285 32217 2.330320 31261 31285 1.4511050  0.8792148
6 1 0.018378723 33200 31313 4.192704 31262 31419 1.1147747  3.0779289
```



R Tool!

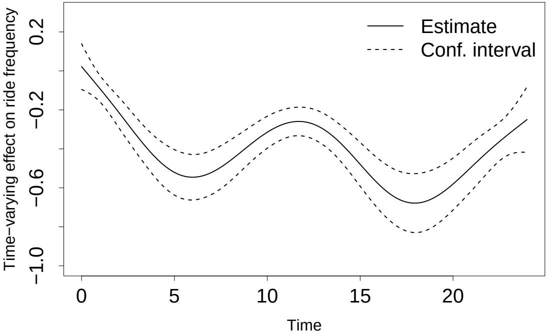
```
gam(y ~ -1+ s(time, by = PT_dist), family= "binomial", data = data)
```

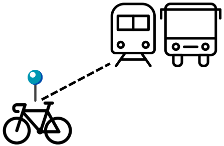
TVE: contribution of distance from public transport

- ... to **sampled partial likelihood (m=2)**

$$\mathcal{L}^S(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{e^{\cdots + \left[\sum_{j=1}^{q} \theta_j b_j(t_i)\right] \Delta x_i^{(2)} + \ldots}}{1 + e^{\cdots + \left[\sum_{j=1}^{q} \theta_j b_j(t_i)\right] \Delta x_i^{(2)} + \ldots}} \text{ (+ smoothing penalty)}$$

Result...





Modeling Non-Linear Effects

Non-linear Effect f of **distance between two bike stations** $x_{sr}^{(3)}(t)$

on bike sharing dynamics

Modeling Non-Linear Effects

- Effect of $x_{sr}^{(3)}(t)$ as $\rightarrow$ Spline function of Covariate

$$f(x_{sr}^{(3)}(t)) \approx \sum_{j=1}^{q} \theta_j b_j(x_{sr}^{(3)}(t))$$

Choices to be made:

- **basis dimension** q
- a **set of smooth basis functions** $\{b_j\}_j$ e.g. *B-splines, thin plate regression splines,...*
- given range of variable, **basis evaluation points** (e.g. *uniform, quantile of a distribution,...*) to evaluate q **smooth non-linear functions** of covariate $b_j(x_{sr}^{(3)}(t))$.

Estimating $f \iff$ Estimating θ_j-s

NLE: contribution of distance between two bike stations

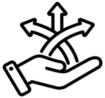
- ... to **log-hazard of an interaction**

$$f_{sr} = \dots \; + \underbrace{\sum_{j=1}^{q} \theta_j b_j(x_{sr}^{(3)}(t))}_{\text{non-linear effect}} + \; \dots$$

- ... to **sampled likelihood function (m=2)**

$$\mathcal{L}^S(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{e^{\cdots \, + \left[\sum_{j=1}^{q} \theta_j [b_j(x_{s_i r_i}^{(3)}(t_i)) - b_j(x_{s_i^* r_i^*}^{(3)}(t_i))]\right] + \dots}}{1 + e^{\cdots \, + \left[\sum_{j=1}^{q} \theta_j [b_j(x_{s_i r_i}^{(3)}(t_i)) - b_j(x_{s_i^* r_i^*}^{(3)}(t_i))]\right] + \dots}} \; (+ \text{ smoothing penalty})$$

NLE: contribution of distance between two bike stations



- … to **sampled likelihood function (m=2)**

$$\mathcal{L}^S(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{e^{\cdots + \left[\sum_{j=1}^{q} \theta_j [b_j(x^{(3)}_{s_i r_i}(t_i)) - b_j(x^{(3)}_{s^*_i r^*_i}(t_i))]\right] + \cdots}}{1 + e^{\cdots + \left[\sum_{j=1}^{q} \theta_j [b_j(x^{(3)}_{s_i r_i}(t_i)) - b_j(x^{(3)}_{s^*_i r^*_i}(t_i))]\right] + \cdots}} \quad (+ \text{ smoothing penalty})$$

Peek-a-Data!

```
> head(data)
  y        time    s1    r1  dist1    s2    r2  dist2 ev.i nonev.i
1 1 0.001377150 33202 32604 14.073 32604 31937  9.260    1      -1
2 1 0.001453772 31937 31285  4.404 32217 31246 43.790    1      -1
3 1 0.001979046 33200 31937  5.258 31255 31261  1.881    1      -1
4 1 0.004495738 31246 31937  3.333 31218 31937  6.001    1      -1
5 1 0.007457826 31324 31218  3.970 31272 31261  3.300    1      -1
6 1 0.010580056 31272 31324  4.492 31272 31937  6.689    1      -1
```

R Tool!

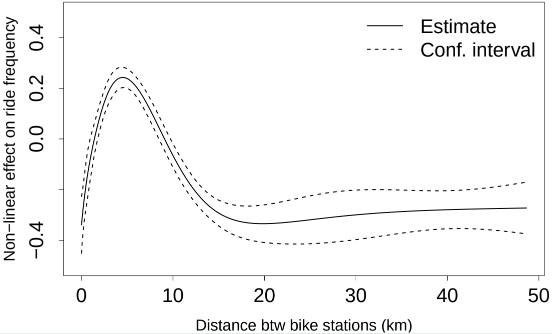
```
gam(y ~ -1+ s(cbind(dist1, dist2),by = cbind(ev.i, nonev.i),
             family= "binomial", data = data)
```

NLE: contribution of distance between two bike stations

- … to **sampled likelihood function (m=2)**

$$\mathcal{L}^S(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{e^{\cdots \,+\, \left[\sum_{j=1}^{q} \theta_j \left[b_j(x^{(3)}_{s_i r_i}(t_i)) - b_j(x^{(3)}_{s_i^* r_i^*}(t_i))\right]\right] \,+\, \cdots}}{1 + e^{\cdots \,+\, \left[\sum_{j=1}^{q} \theta_j \left[b_j(x^{(3)}_{s_i r_i}(t_i)) - b_j(x^{(3)}_{s_i^* r_i^*}(t_i))\right]\right] \,+\, \cdots}} \quad \text{(+ smoothing penalty)}$$

Result…

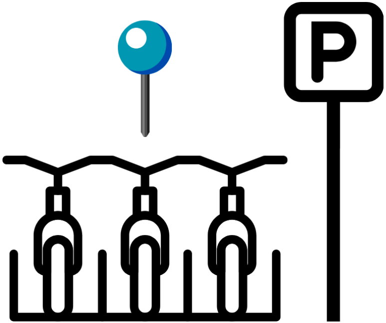


Modeling Random Effects

Random Effect γ of

a bike station $z_{sr}(t) = \mathbb{1}\{\text{start station} = s\}$

on bike sharing dynamics



Modeling Random Effects

- Effect of $z_{sr}(t) = \mathbb{1}\{\text{start station} = s\}$ as $\rightarrow$ **0-dim spline**

$$\gamma' z_{sr}(t) \approx \sum_{s' \in \mathcal{S}} \gamma_{s'} \mathbb{1}\{s' = s\} \qquad \boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{|\mathcal{S}|})$$

- Consider a model with Gaussian random effects, $\gamma \sim \mathcal{N}(0, \sigma^2 I_{|\mathcal{S}|})$;
- Same as fitting a ridge regression model (quadratic penalty term on $\boldsymbol{\gamma}$)

Estimating $\gamma \Longleftrightarrow$ Estimating $\gamma_{s'}$-s (and σ^2)

RE: contribution of a bike station

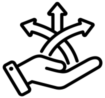
- ... to **log-hazard of an interaction**

$$f_{sr} = \ \dots \ + \underbrace{\sum_{s' \in \mathcal{S}} \gamma_{s'} \mathbb{1}\{s = s'\}}_{\text{random effect}} + \ \dots$$

- ... to **sampled likelihood function (m=2)**

$$\mathcal{L}^S(\gamma) = \prod_{i=1}^{n} \frac{e^{\dots + \gamma_{s_i} - \gamma_{s_i^*} + \dots}}{1 + e^{\dots + \gamma_{s_i} - \gamma_{s_i^*} + \dots}} - \frac{1}{2\sigma^2_{\text{station}}} \gamma' \gamma$$

RE: contribution of a bike station

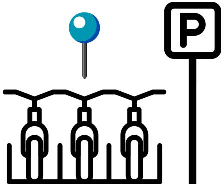


- … to **sampled likelihood function (m=2)**

$$\mathcal{L}^S(\gamma) = \prod_{i=1}^{n} \frac{e^{\cdots + \gamma_{s_i} - \gamma_{s_i^*} + \cdots}}{1 + e^{\cdots + \gamma_{s_i} - \gamma_{s_i^*} + \cdots}} - \frac{1}{2\sigma^2_{\text{station}}} \gamma' \gamma$$

Peek-a-Data!

```
> head(data)
  y        time    s1    s2 ev.i nonev.i
1 1 0.003427363 31324 31023    1       1
2 1 0.041364621 31324 31262    1       1
3 1 0.042194513 31324 31272    1       1
4 1 0.049880008 31324 31262    1       1
5 1 0.051420690 31324 31071    1       1
6 1 0.051820703 31324 31419    1       1
```



R Tool!

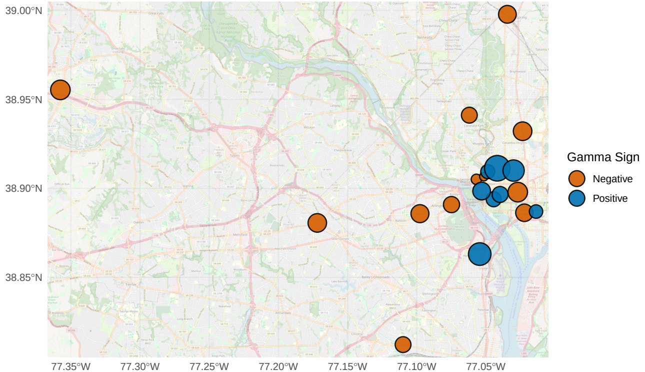
```
stat_pop <- matrix(factor(c(data$s1, data$s2)), ncol = 2)

gam(y ~ -1+ s(stat_pop, by = cbind(ev.i, nonev.i), bs = "re"),
        method = "REML", family="binomial", data = data)
```

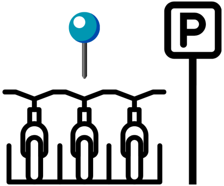
RE: contribution of a bike station

- … to **sampled likelihood function (m=2)**

$$\mathcal{L}^S(\gamma) = \prod_{i=1}^{n} \frac{e^{\cdots + \gamma_{s_i} - \gamma_{s_i^*} + \cdots}}{1 + e^{\cdots + \gamma_{s_i} - \gamma_{s_i^*} + \cdots}} - \frac{1}{2\sigma^2_{\text{station}}} \gamma' \gamma$$

Result…

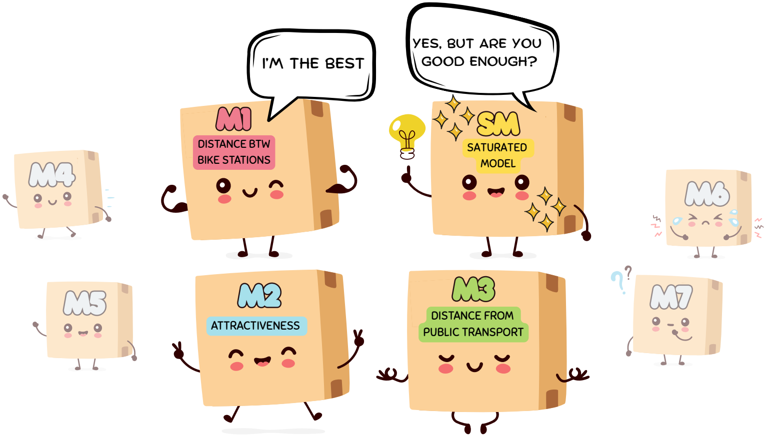




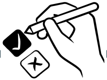
Computer Practical 3

File _CP_3_.Rmd is currently in Folder _PRACTICAL_/_CP_3_MELANIA___.
Computer Practical 3 aims to apply the concepts reviewed in the previous section to a hyperevent dataset The dataset has been created in Computer Practical 1 and has been reworked in Computer Practical 2.

Goodness-of-fit of RHEMs



A quick recap!



Data Collection:

- Stream of **time-stamped relational events** ...
- ... Arises from an underlying **data generating process** (DGP);

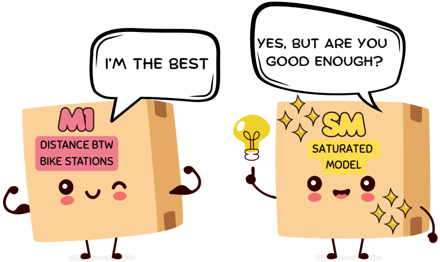
Model Fitting:

- **Multiple candidate models** based on available information ...
- ... may differ in **included covariates**, **types of effects**, sensitivity to **outliers**;

Model Selection vs...

Model Selection:

- Select **best model** for **predictive power** or **posterior model probability**;
- Inclusion of **significant drivers**;



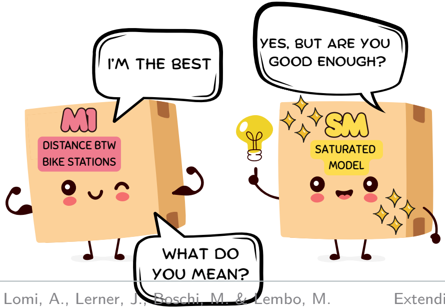
Model Selection vs Goodness of Fit

Model Selection:

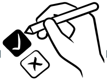
- Select **best model** for **predictive power** or **posterior model probability**
- Inclusion of **significant drivers**;

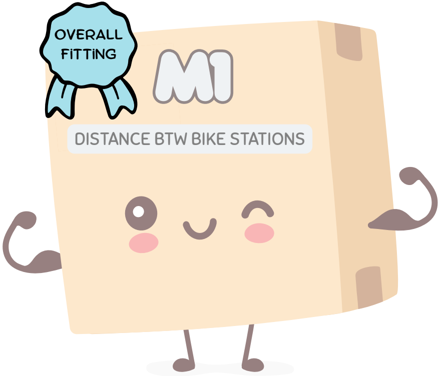
Goodness of Fit:

- Test **adequacy of fitted model**;



GOF: Overall fitting





Overall model fit

A model fits adequately if misfit is completely explained by stochastic nature of response.

Usual tests employed:

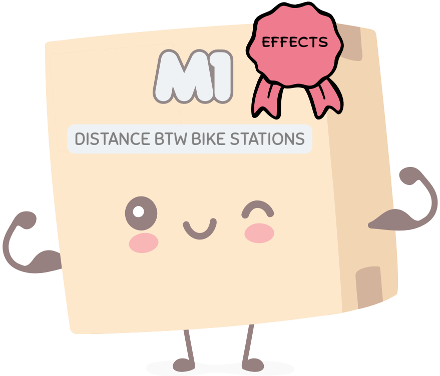
- Deviance test
- R^2

GOF: Covariates

Covariate GOF

A covariate provides adequate fit if expected coviate value matches observed covariate value.

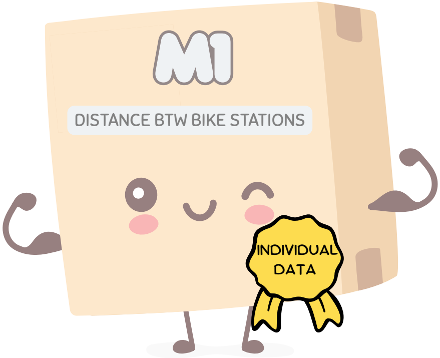
GOF: Effects



Effect GOF

Functional form of individual effects are well-specified.

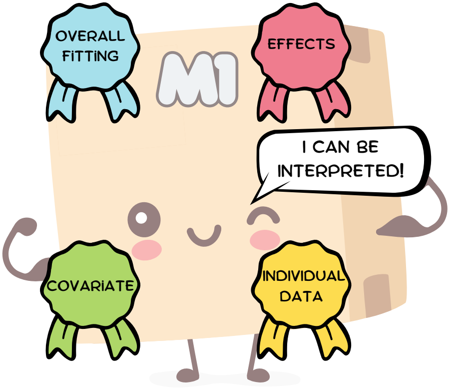
GOF: Individual Data



Individual observation GOF

Are data contaminated by observations that do not follow global model?

Interpretability



Interpretability

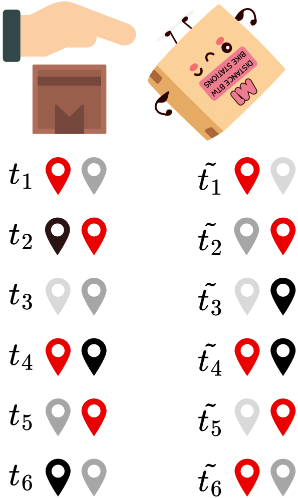
Only adequate models should be interpreted (...rest should be treated with caution).

GOF: State of the art (Part 1)

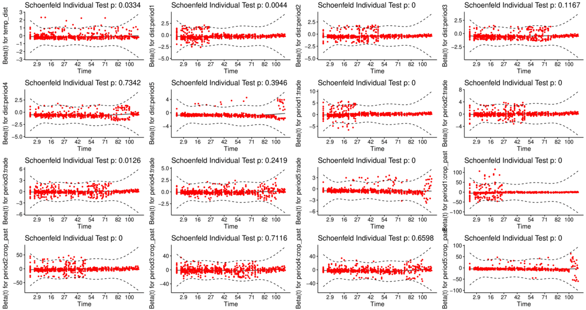
Simulation-based techniques:

- Comparing **simulated and observed** events;
- ☹: $\uparrow$ computational cost;

Q: What type of GOF are we performing?



GOF: State of the art (Part 2)

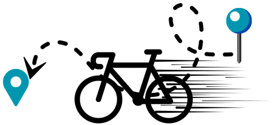


Cox regression techniques:

- **Schoenfeld and deviance residuals**;
- ☹: Missing formulation for extended models;
- ☹: Exploratory, not formal;

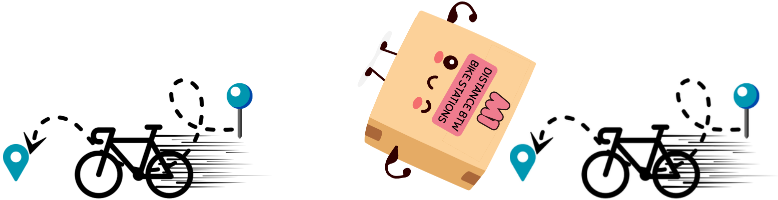
Q: What type of GOF are we performing?

GOF via Weighted Sum of Martingale Residuals



$$x_{sr}(t)$$

GOF via Weighted Sum of Martingale Residuals



$$x_{sr}(t) \quad - \quad \mathbb{E}[x_{sr}(t) \,|\, \text{fitted model}]$$

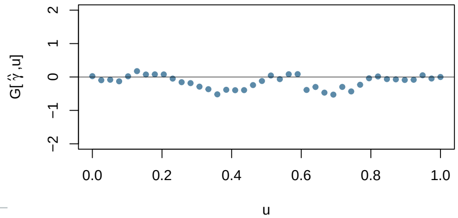
GOF via Weighted Sum of Martingale Residuals

$$G[\hat{\beta}, u|\mathcal{E}] = \sum_{k \le \lfloor nu \rfloor} \left[\underbrace{x_{s_k r_k}(t_k)}_{\text{observed covariate}} - \underbrace{\frac{\sum_{(s,r)\in SR_k} x_{sr}(t_k) \cdot \exp\left[\hat{\beta}^T x_{sr}(t_k)\right]}{\sum_{(s,r)\in SR_k} \exp\left[\hat{\beta}^T x_{sr}(t_k)\right]}}_{\text{expected value of covariate under fitted model}} \right], \quad u \in [0,1]$$

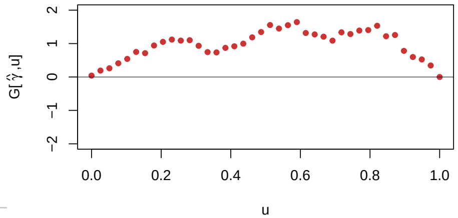
GOF via Weighted Sum of Martingale Residuals

$$G[\hat{\beta}, u|\mathcal{E}] = \sum_{k \le \lfloor nu \rfloor} \left[\underbrace{x_{s_k r_k}(t_k)}_{\text{observed covariate}} - \underbrace{\frac{\sum_{(s,r)\in SR_k} x_{sr}(t_k) \cdot \exp\left[\hat{\beta}^T x_{sr}(t_k)\right]}{\sum_{(s,r)\in SR_k} \exp\left[\hat{\beta}^T x_{sr}(t_k)\right]}}_{\text{expected value of covariate under fitted model}} \right], \quad u \in [0,1]$$

G Process over time − Adequate Model



G Process over time: Misspecified Model



GOF via Weighted Sum of Martingale Residuals

$$G[\hat{\beta}, u|\mathcal{E}] = \sum_{k \leq \lfloor nu \rfloor} \left[\underbrace{x_{s_k r_k}(t_k)}_{\text{observed covariate}} - \underbrace{\frac{\sum_{(s,r)\in SR_k} x_{sr}(t_k) \cdot \exp\left[\hat{\beta}^T x_{sr}(t_k)\right]}{\sum_{(s,r)\in SR_k} \exp\left[\hat{\beta}^T x_{sr}(t_k)\right]}}_{\text{expected value of covariate under fitted model}} \right], \quad u \in [0,1]$$

- $\downarrow$ **Computational cost**: ☺: **no simulation** of relational process;
- **Formal test**: ☺: based on **theoretical properties of process**;
- Used in **extended models**: ☺: with **non-linear & random effects**;

Towards a... (Flexible) Formal test

In Linear case, **variance** of the process is:
$\widehat{J} = \frac{\mathcal{I}[\hat{\beta}]_{j,j}}{n}$, where $\mathcal{I}[\hat{\beta}]$: **observed information matrix**;

If model is adequate,

$$\widehat{W}[\hat{\beta}, \cdot] = \sqrt{\mathcal{I}[\hat{\beta}]_{j,j}} \times G[\hat{\beta}, \cdot] \xrightarrow{d} Z^0(\cdot)$$

where $Z^0(\cdot)$: **standard univariate Brownian bridge**

We consider **statistical test of Kolmogorov–Smirnov type**:

$$T_x = \sup_{u \in [0,1]} |\widehat{W}[\hat{\beta}, u]|$$

which follows a **Kolmogorov distribution**. Then, **Exact p-value** can be computed.

Towards a... (Flexible) Formal test

In Linear case, **variance** of the process is:
$\widehat{J} = \frac{\mathcal{I}[\hat{\beta}]_{j,j}}{n}$, where $\mathcal{I}[\hat{\beta}]$: **observed information matrix**;

If model is adequate,

$$\widehat{W}[\hat{\beta}, \cdot] = \sqrt{\mathcal{I}[\hat{\beta}]_{j,j}} \times G[\hat{\beta}, \cdot] \xrightarrow{d} Z^0(\cdot)$$

where $Z^0(\cdot)$: **standard univariate Brownian bridge**

We consider **statistical test of Kolmogorov–Smirnov type**:

$$T_x = \sup_{u \in [0,1]} |\widehat{W}[\hat{\beta}, u]|$$

which follows a **Kolmogorov distribution**. Then, **Exact p-value** can be computed.

Towards a... (Flexible) Formal test

In Linear case, **variance** of the process is:
$\widehat{J} = \frac{\mathcal{I}[\hat{\beta}]_{j,j}}{n}$, where $\mathcal{I}[\hat{\beta}]$: **observed information matrix**;

If model is adequate,

$$\widehat{W}[\hat{\beta}, \cdot] = \sqrt{\mathcal{I}[\hat{\beta}]_{j,j}} \times G[\hat{\beta}, \cdot] \xrightarrow{d} Z^0(\cdot)$$

where $Z^0(\cdot)$: **standard univariate Brownian bridge**

We consider **statistical test of Kolmogorov–Smirnov type**:

$$T_x = \sup_{u \in [0,1]} |\widehat{W}[\hat{\beta}, u]|$$

which follows a **Kolmogorov distribution**. Then, **Exact p-value** can be computed.

Towards a... (Flexible) Formal test

In general (Non-Linear + Random) case,
We can adjust...

- $\boldsymbol{G}[\hat{\gamma}, \cdot]$ is multivariate $-$ $\boldsymbol{b}_{sr}$ and $\boldsymbol{\gamma}_{sr}$ are set of basis functions.
- We do have a **penalty term**... We need to center each term!
- We cannot rely anymore on $\widehat{\boldsymbol{J}}$: empirical variance-covariance matrix;

If the model is adequate $\widehat{\boldsymbol{W}}[\hat{\gamma}, \cdot] = \hat{\boldsymbol{J}}^{-\frac{1}{2}} \times n^{-\frac{1}{2}} \times \boldsymbol{G}[\hat{\gamma}, \cdot] \xrightarrow{d} \boldsymbol{Z}^0(\cdot)$
where $\boldsymbol{Z}^0(u)$: *q-multivariate Brownian bridge.*

... still a **statistical test of Kolmogorov–Smirnov type**!

Empirical p-value can be computed.

Towards a... (Flexible) Formal test

In general (Non-Linear + Random) case,
We can adjust...

- $\boldsymbol{G}[\hat{\gamma}, \cdot]$ is multivariate – $\boldsymbol{b}_{sr}$ and $\boldsymbol{\gamma}_{sr}$ are set of basis functions.
- We do have a **penalty term**... We need to center each term!
- We cannot rely anymore on $\widehat{\boldsymbol{J}}$: empirical variance-covariance matrix;

If the model is adequate $\widehat{\boldsymbol{W}}[\hat{\gamma}, \cdot] = \widehat{\boldsymbol{J}}^{-\frac{1}{2}} \times n^{-\frac{1}{2}} \times \boldsymbol{G}[\hat{\gamma}, \cdot] \xrightarrow{d} \boldsymbol{Z}^0(\cdot)$
where $\boldsymbol{Z}^0(u)$: *q-multivariate Brownian bridge*.

... still a **statistical test of Kolmogorov–Smirnov type**!

Empirical p-value can be computed.

Towards a... (Flexible) Formal test

In general (Non-Linear + Random) case,
We can adjust...

- $\boldsymbol{G}[\hat{\gamma}, \cdot]$ is multivariate – $\boldsymbol{b}_{sr}$ and $\boldsymbol{\gamma}_{sr}$ are set of basis functions.
- We do have a **penalty term**... We need to center each term!
- We cannot rely anymore on $\widehat{\boldsymbol{J}}$: empirical variance-covariance matrix;

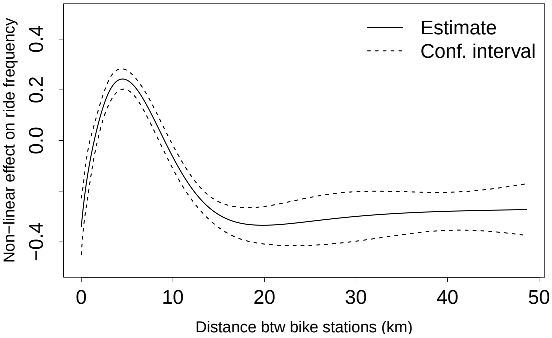
If the model is adequate $\widehat{\boldsymbol{W}}[\hat{\gamma}, \cdot] = \hat{\boldsymbol{J}}^{-\frac{1}{2}} \times n^{-\frac{1}{2}} \times \boldsymbol{G}[\hat{\gamma}, \cdot] \xrightarrow{d} \boldsymbol{Z}^0(\cdot)$
where $\boldsymbol{Z}^0(u)$: *q-multivariate Brownian bridge*.

... still a **statistical test of Kolmogorov–Smirnov type**!

Empirical p-value can be computed.

NLE: contribution of distance between two bike stations

Result...



Let's test it!

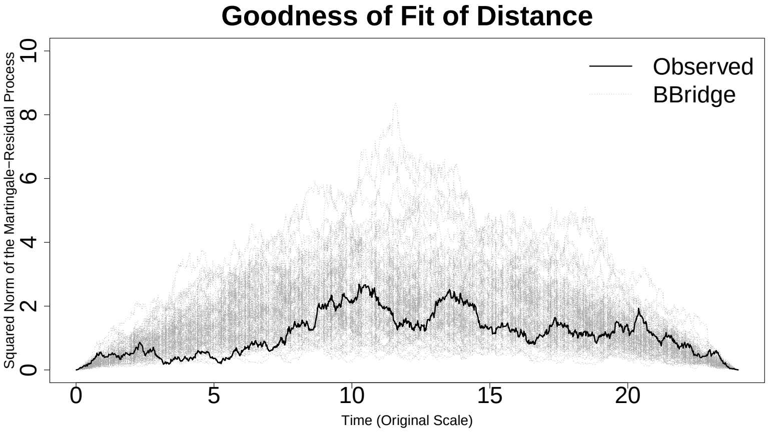
R Tool!

```
gam.fit = gam(y ~ -1+ s(cbind(dist1, dist2),
            by = cbind(ev.i, nonev.i),
      family= "binomial", data = data)
```

```
                 BB9 <- BB.single(dim.k = 9)
GOF = GOF_multivariate(gam.fit = gam_fit, index = 1:9,
                     BB.stat = BB9[[2]])
```

How does it look like?

Spoiler Alert!



Computer Practical 4

File _CP_4_.Rmd is currently in Folder _PRACTICAL_/_05_: Computer Practical 4 aims to show the potential of GOF techniques via Martingale Residuals.

Advanced RHEMs

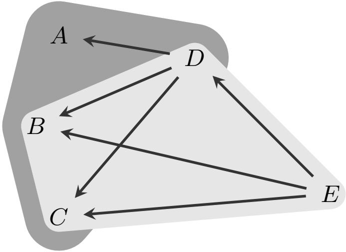
Discuss the following variations of RHEM.

- **Directed RHEM:** hyperevents have senders and receivers (e. g., email communication).
- **Relational hyperevent outcome models (RHOM):** hyperevents represent team work that has an outcome (e. g., coauthoring).
- **Labeled RHEM / two-mode RHEM:** hyperevents have a set of types (e. g., keywords of a published paper).
- **Mixed multimode RHEM:** e. g., coevolution of coauthoring an citation networks.

RHEM for directed multicast communication networks

Polyadic interaction: events involving several nodes.

$$e_1 = (D, \{A, B, C\}, t_1)$$
$$e_2 = (E, \{B, C, D\}, t_2)$$



Directed polyadic interaction:

- multicast (one-to-many) communication, email, texting
- citation networks: papers citing lists of references
- contact-nomination networks (contact diaries of several egos)

RHEM for directed hyperevents.

Here: only for events with a single source and arbitrary number of targets.

Hyperedge: can connect any number of nodes.

Hyperevent: hyperedge (event participants) with time stamp (event time).

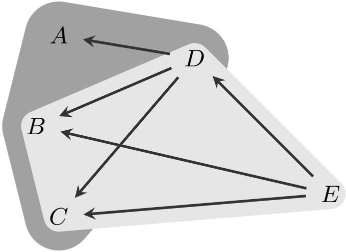
Lerner and Lomi (2023). **Relational hyperevent models for polyadic interaction networks.** *J RSSA*.

Observed data.

Directed hyperevents $(s_1, R_1, t_1), \ldots, (s_n, R_n, t_n)$,
where for $e_i = (s_i, R_i, t_i)$

- t_i is the **time** of event e_i;
- $s_i \in V_{1,t_i}$ is the **sender** of event e_i, taken from a set of possible senders V_{1,t_i};
- $R_i \subseteq V_{2,t_i}(s_i)$ is the **set of receivers** of event e_i, taken from a set of possible receivers $V_{2,t_e}(s_i)$.

$$e_1 = (D, \{A, B, C\}, t_1)$$
$$e_2 = (E, \{B, C, D\}, t_2)$$



RHEM for directed hyperevents (CoxPH model).

Intensity $\lambda_{sR}(t) \approx$ expected $\#$ events send by s to R at t

$$\lambda_{sR}(t) \quad = \quad \overline{\lambda}_{s,|R|}(t) \exp\left\{\beta^{\mathrm{T}} x_{sR}(t)\right\} \mathbf{1}\{R \subseteq V_{2,t}(s)\} \quad .$$

- **baseline intensity** $\overline{\lambda}_{s,|R|}(t) \approx$ "average" $\#$ events send by s to any set of receivers of size $|R|$
  Baseline intensity is not explained by the model (usually).
- **hyperedge covariates** $x_{sR}(t)$, explain relative intensity of R compared to average receiver set of the same size.

Estimate parameters $\boldsymbol{\beta}$ by maximizing partial likelihood:

$$\mathcal{L}^{P}(\boldsymbol{\beta}, t) \quad = \quad \prod_{e_i : t_i \leq t} \frac{\beta^{\mathrm{T}} x_{s_i R_i}(t_i)}{\sum_{R \in \binom{V_{2,t_i}(s_i)}{|R_i|}} \exp\left\{\beta^{\mathrm{T}} x_{s_i R}(t_i)\right\}}$$

Usually: sample from the risk set (case-control sampling).

Compare RHEM and dyadic REM.

Perry and Wolfe (2013)

$$\lambda_{sR}^{\text{(dyadic)}}(t) = \overline{\lambda}_{s,|R|}(t) \exp\left\{\beta^{\mathrm{T}} \sum_{r \in J} x_{sR}(t)\right\} \prod_{r \in R} \mathbf{1}\{r \in V_{2,t}(s)\}$$

$$\lambda_{sR}^{\text{(rhem)}}(t) = \overline{\lambda}_{s,|R|}(t) \exp\left\{\beta^{\mathrm{T}} x_{sR}(t)\right\} \mathbf{1}\{R \subseteq V_{2,t}(s)\} \quad ,$$

Dyadic REM for hyperevents require the decomposition:

$$x_{sR}(t) = \sum_{r \in R} x_{sr}(t) \quad .$$

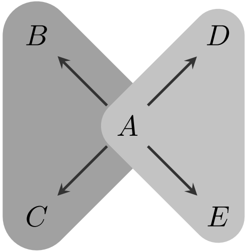
Suitability of r as a receiver is assumed to be independent of other receivers $r' \in R$.

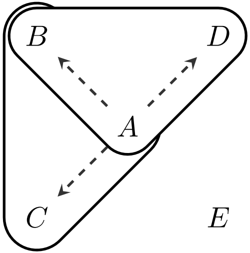
RHEM don't impose that condition & allow more general **hyperedge covariates** $x_{sR}(t)$.

Perry & Wolfe (2013). **Point process modelling for directed interaction networks**. *J RSSB.*

Insufficiency of dyadic effects (I).

Actor A sent two messages: $(A, \{B, C\})$ and $(A, \{D, E\})$.

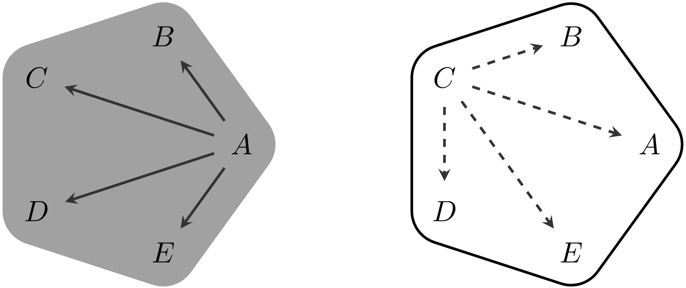




Purely dyadic effects would consider a future message $(A, \{B, C\})$ as likely as a message $(A, \{B, D\})$.

Insufficiency of dyadic effects (II).

"Reply-to-all" in email communication:



Such patterns cannot be captured with purely dyadic covariates.

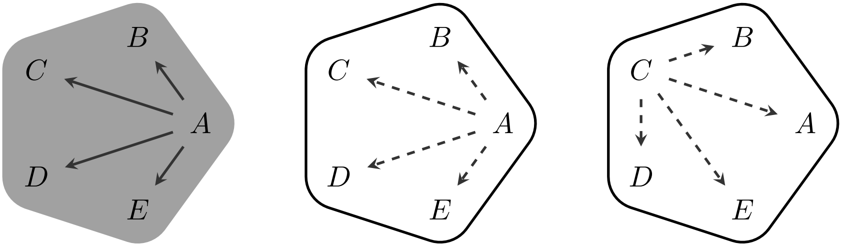
RHEM effects: directed hyperedge covariates.

Here: only for events with a single sender and arbitrary number of receivers.

Exact repetition and undirected exact repetition.

$$\text{repetition}(s,R,t) = \sum_{e_i : t_i < t} w(t-t_i) \cdot \mathbf{1}(s_i = s \wedge R_i = r) \ .$$

$$\text{undir.rep}(s,R,t) = \sum_{e_i : t_i < t} w(t-t_i) \cdot \mathbf{1}(\{s_i\} \cup R_i = \{s\} \cup R) \ .$$

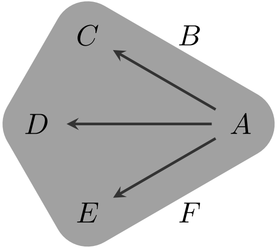


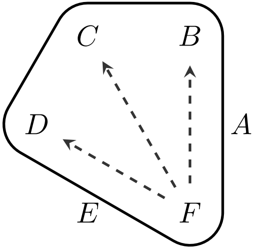
$$w(t-t_i) := \exp\left(-(t-t_i)\frac{\log 2}{T_{1/2}}\right).$$

Partial receiver set repetition.

clustering in space of possible receivers

$$\text{r.sub.rep}^{(k)}(s, R, t) \quad = \quad \sum_{R' \in \binom{R}{k}} \frac{\text{hy.deg}^{(in)}(R', t)}{\binom{|J|}{p}} \quad .$$



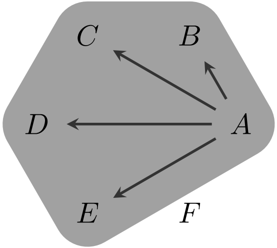


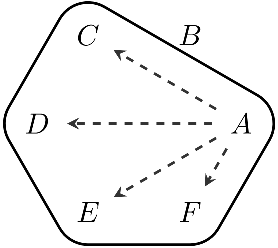
$$\text{hy.deg}^{(in)}(R', t) \quad = \quad \sum_{e_i : t_i < t} w(t - t_i) \cdot \mathbf{1}(R' \subseteq R_i) \quad .$$

Sender-specific partial receiver set repetition.

sender-specific clustering in space of possible receivers

$$\text{s.r.sub.rep}^{(k)}(s, R, t) \quad = \quad \sum_{R' \in \binom{R}{k}} \frac{\text{hy.deg}(s, R', t)}{\binom{|R|}{k}} \; .$$



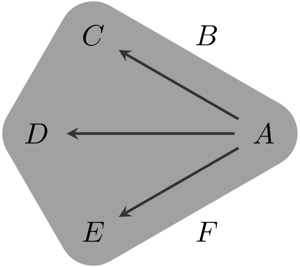


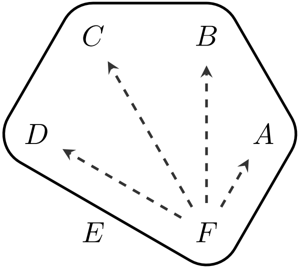
$$\text{hy.deg}(s, R', t) \quad = \quad \sum_{e_i : t_i < t} w(t - t_i) \cdot \mathbf{1}(s = s_i \wedge R' \subseteq R_i) \; .$$

Interaction among receivers.

for instance, citing a paper and some of its references

$$\text{interact.receivers}^{(k)}(s,R,t) \quad = \sum_{r\in R,\, R'\in\binom{R\setminus\{r\}}{k}} \frac{\text{hy.deg}(r,R',t)}{|R|\cdot\binom{|R|-1}{k}} \quad .$$

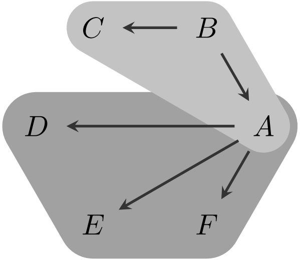


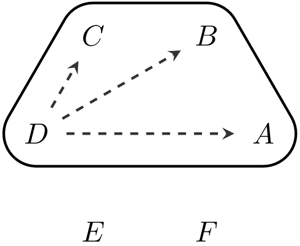


(Generalized) reciprocation.

$$\text{reciprocation}(s, R, t) = \sum_{r \in R} \text{hy.deg}(r, \{s\}, t)/|R|$$

$$\text{gen.recip}(s, R, t) = \sum_{r \in R} \text{deg}^{(out)}(r, t)/|R|$$





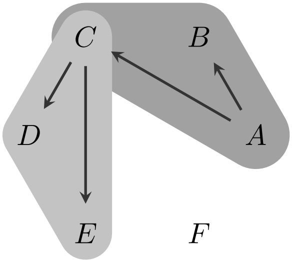
$$\text{deg}^{(out)}(s', t) = \sum_{e_i : t_i < t} w(t - t_i) \cdot \mathbf{1}(s' = s_i)$$

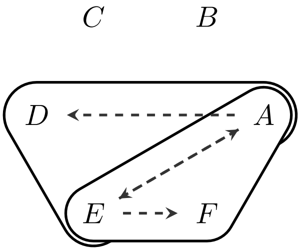
Closure.

$$
\begin{aligned}
\text{trans.closure}(s, R, t) &= \sum_{r \in R,\, a \neq s, r} \frac{\min\{\text{hy.deg}(s, \{a\}, t), \text{hy.deg}(a, \{r\}, t)\}}{|R|} \\
\text{cyclic.closure}(s, R, t) &= \sum_{r \in R,\, a \neq s, r} \frac{\min\{\text{hy.deg}(a, \{s\}, t), \text{hy.deg}(r, \{a\}, t)\}}{|R|} \\
\text{shared.sender}(s, R, t) &= \sum_{r \in R,\, a \neq s, r} \frac{\min\{\text{hy.deg}(a, \{s\}, t), \text{hy.deg}(a, \{r\}, t)\}}{|R|} \\
\text{shared.receiver}(s, R, t) &= \sum_{r \in R,\, a \neq s, r} \frac{\min\{\text{hy.deg}(s, \{a\}, t), \text{hy.deg}(r, \{a\}, t)\}}{|R|} \quad .
\end{aligned}
$$

Closure: visual illustration (I).

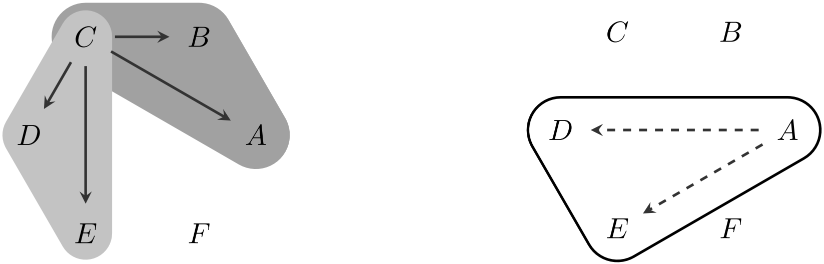
transitive closure and cyclic closure





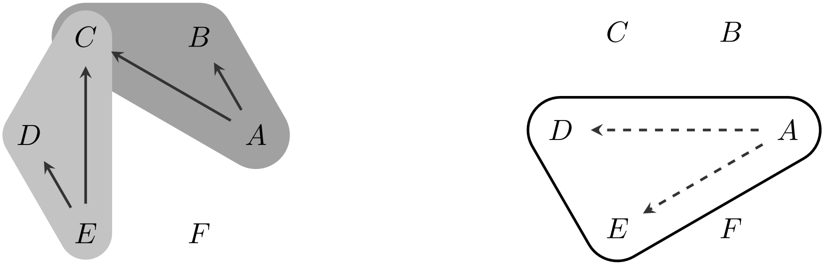
Closure: visual illustration (II).

shared sender (source)



Closure: visual illustration (III).

shared receiver (target)



Actor attribute effects.

Given actor attribute $z \in \mathbb{R}$

Hyperedge covariates $x_{sR}(t)$ dependent on z can measure

- attribute value of the sender z_s (not in sender-conditional models)
- summary measure of the distribution of attribute values of the receivers, e. g., $\text{mean}_{r\in R}[z_r]$, $\text{sd}_{r\in R}[z_r]$
- summary measure of the distribution of attribute values of the receivers in relation to the sender, e. g., $\text{mean}_{r\in R}[|z_r - z_s|]$,

Case study: email network.

Enron email corpus.

https://www.cs.cmu.edu/~enron/

Collection of 21,635 emails among 156 employees of Enron Corporation, cleaned and compiled by Zhou et al. (2007).

Emails (hyperevents) have one sender and between one and 57 receivers.

Actor-level attributes: gender, seniority, and department (legal, trading, other).

https://github.com/patperry/interaction-proc/tree/master/data/enron

https://github.com/juergenlerner/eventnet/tree/master/data/enron

Receiver set size distribution.

Number of receivers between 1 and 57.

About 30% have more than one receiver.

Mean number of receivers is 1.77.

num. receivers $\lvert J \rvert$	frequency
1	14,985
2	2,962
3	1,435
4	873
5	711
6	180
7	176
8	61
9	24
10	29
> 10	199
all	21,635

	RHEM	Dyadic REM
r.avg.female	0.220 (0.024)***	0.261 (0.024)***
s.r.abs.diff.gender	−0.184 (0.023)***	−0.232 (0.023)***
r.pair.abs.diff.gender	−0.253 (0.065)***	
r.avg.seniority	0.294 (0.024)***	0.417 (0.024)***
s.r.abs.diff.seniority	−0.424 (0.022)***	−0.496 (0.022)***
r.pair.abs.diff.seniority	−0.795 (0.068)***	
r.avg.in.legal	0.057 (0.032)	0.095 (0.032)**
r.avg.in.trading	−0.074 (0.028)**	−0.180 (0.029)***
s.r.frac.diff.department	−0.761 (0.023)***	−0.922 (0.023)***
r.pair.frac.diff.department	−1.152 (0.066)***	
repetition	−0.221 (0.011)***	
undirected.repetition	0.391 (0.013)***	
r.sub.rep.1	0.089 (0.018)***	0.053 (0.018)**
r.sub.rep.2	0.110 (0.009)***	
r.sub.rep.3	0.139 (0.020)***	
r.sub.rep.4	0.252 (0.054)***	
s.r.sub.rep.1	0.674 (0.012)***	0.888 (0.007)***
s.r.sub.rep.2	0.515 (0.024)***	
s.r.sub.rep.3	1.225 (0.166)***	
receiver.outdeg	0.049 (0.016)**	0.101 (0.015)***
reciprocation	0.062 (0.009)***	0.227 (0.006)***
interact.receivers.1	0.164 (0.007)***	
interact.receivers.2	0.290 (0.037)***	
interact.receivers.3	0.630 (0.145)***	
shared.sender	0.352 (0.016)***	0.384 (0.015)***
shared.receiver	−0.009 (0.016)	0.001 (0.014)
transitive.closure	−0.025 (0.019)	0.120 (0.017)***
cyclic.closure	−0.121 (0.014)***	−0.185 (0.013)***
AIC	74,670.703	85,999.084

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Qualitative findings

- Found relevant effects that do not admit a dyadic decomposition (higher-order effects).

- Higher-order effects are typically significant.
- Effect sizes of dyadic effects typically decrease when controling for higher order effects.
- In some cases: effect significant in dyadic model but not in RHEM.
- RHEM have better model fit.

Some structural effects

Negative repetition and positive undirected repetition.

- Turn-taking within emergent conversation groups.
- Alternatively: effect of reply-to-all functionality.

Partial repetition of receiver sets.

- Clustering in space of actors: subsets of actors likely to receive joint messages.

Sender-specific partial repetition of receiver sets.

- Subsets of actors likely to receive joint messages from a given sender (sender-specific clustering).

Practical aspects

How to specify directed RHEM with eventnet?

`https://github.com/juergenlerner/eventnet`

Conclusion

Higher-order effects can be found in empirical data.

Ignoring them can decrease model fit and yield potentially spurious findings.

Higher-order dependencies should not be considered merely an annoyance to be controlled away.

They allow to develop and test additional theories.

Lerner and Lomi (2023). **Relational hyperevent models for polyadic interaction networks.** *J RSSA*.

`https://github.com/juergenlerner/eventnet`

Relational events that have an outcome.

Sometimes, relational (hyper)events produce a tangible outcome with associated "success" or "performance".

- Teams of scientists publishing an article → impact.
- Filmmakers producing a film → revenue, popularity, evaluation, awards.
- Medical teams performing an operation → cost, duration, benefits for the patient.
- Two soccer teams meeting in a match.

Past success may explain future collaboration (RHEM).

Past success may explain future success (RHOM).

Lerner & Hâncean (2023). **Micro-level network dynamics of scientific collaboration and impact: relational hyperevent models for the analysis of coauthor networks.** *Network Science*.

Data: relational events that have an outcome.

A relational hyperevent with an outcome is given by

$$e_i = (S_i, t_i, x_i, y_i) \ \ ,$$

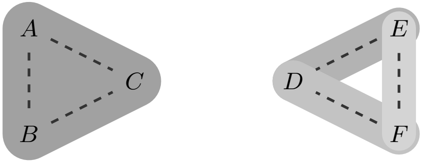
- t_i is the time of the event;
- $S_i \subseteq V_{1,t_i}$ event participants (e. g., coauthors);
- x_i covariates of the event (type, weight);
- y_i is the *relational outcome* (success, performance, impact).

Past success as explanatory variable.

Recall: subset repetition of order k.

Average number of prior joint events over $S' \subseteq S$ of size S.

$$\text{sub.rep}^{(k)}(S,t) \quad = \quad \sum_{S' \in \binom{S}{k}} \frac{\text{hy.deg}(S',t)}{\binom{|S|}{k}} \quad .$$



$\text{sub.rep}^{(1)}(\{A,B,C\}) = 1$ $\quad$ $\text{sub.rep}^{(1)}(\{D,E,F\}) = 2$

$\text{sub.rep}^{(2)}(\{A,B,C\}) = 1$ $\quad$ $\text{sub.rep}^{(2)}(\{D,E,F\}) = 1$

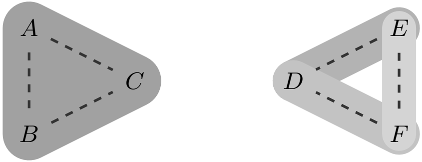
$\text{sub.rep}^{(3)}(\{A,B,C\}) = 1$ $\quad$ $\text{sub.rep}^{(3)}(\{D,E,F\}) = 0$

$$\text{hy.deg}(S',t) \quad = \quad \sum_{e_i : t_i < t} \mathbf{1}(S' \subseteq S_e)$$

Prior (shared) success of order k.

Average weight of prior joint events over $S' \subseteq S$ of size k.

$$\text{prior.success}^{(k)}(S,t) = \frac{\sum_{S' \in \binom{S}{k}} \text{performance}(S',t)}{\sum_{S' \in \binom{S}{k}} \text{hy.deg}(S',t)} \ .$$

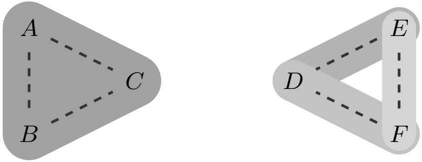


$$\text{performance}(S',t) = \sum_{e_i : t_i < t} y_i \cdot \mathbf{1}(S' \subseteq S_i) \ .$$

Related: prior success disparity.

Individual success vs. shared success.

Assume that two teams are composed of individuals that have been equally successful in the past.



It still makes a difference whether they have **jointly** achieved this success.

A team is more than the sum of its members.

Success of the event as the response variable.

Recall: relational events that have an outcome.

A relational hyperevent with an outcome is given by

$$e_i = (S_i, t_i, y_i) \ ,$$

- t_i is the time of the event;
- $S_i \subseteq V_{1,t_i}$ event participants (e. g., coauthors);
- y_i is the *relational outcome* (success, performance, impact).

RHEM explain the participants S_i (and/or the time t_i).

RHOM explain the outcome y_i, given that there is an event (S_i, t_i).

Relational hyperevent outcome models (RHOM).

Given of sequence of relational hyperevents with an outcome

$$E = [(S_1, t_1, y_1), \ldots, (S_n, t_n, y_n)]$$

RHOM specify regression models (OLS, logit, ...) for

$$y_i \sim f(t_i, S_i; \mathcal{W}_{t_i^-})$$

Allows to control for certain forms of dependence:

- repeated observations by the same team;
- observations with overlapping teams;
- ...

Example study.

Lerner & Hâncean (2023). **Micro-level network dynamics of scientific collaboration and impact: relational hyperevent models for the analysis of coauthor networks.** *Network Science*.

Data: publications of European researchers.

More than 300,000 papers by more than 500,000 researchers from Physics, Medicine, and the Social Sciences.

Outcome variable (y_i): normalized number of citations (by discipline and year of publication).

(Not ideal – but in widespread use.)

Results (RHEM: CoxPH; RHOM: OLS).

	RHEM (team formation)	RHOM (performance)
sub.rep$^{(1)}$	$0.317\ (0.000)^{***}$	$-0.217\ (0.167)$
sub.rep$^{(2)}$	$-0.033\ (0.000)^{***}$	$-0.614\ (0.165)^{***}$
sub.rep$^{(3)}$	$0.048\ (0.000)^{***}$	$-0.634\ (0.147)^{***}$
closure	$-0.042\ (0.001)^{***}$	$-0.071\ (0.105)$
prior.succ$^{(1)}$	$-0.224\ (0.002)^{***}$	$22.231\ (0.491)^{***}$
prior.succ$^{(2)}$	$0.025\ (0.001)^{***}$	$1.954\ (0.172)^{***}$
prior.succ$^{(3)}$	$0.034\ (0.001)^{***}$	$0.814\ (0.116)^{***}$
succ.disparity	$0.086\ (0.000)^{***}$	$-1.825\ (0.136)^{***}$
num.auth		$0.727\ (0.023)^{***}$
(Intercept)		$-0.673\ (0.411)$
Num. obs.	$3,675,744$	$355,977$

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Practical aspects

how to specify RHOM with eventnet?

`https://github.com/juergenlerner/eventnet`

Conclusion.

REM or RHEM typically applied to explain interaction.

A slight variation can be used to explain **outcome** (success, performance, . . .) of interaction events $\rightarrow$ relational (hyperevent) outcome models (ROM, RHOM).

R(H)EM and R(H)OM can be specified with the same explanatory variables, allowing to compare effects on **team selection** and **team performance**.

Effects can be consistent or inconsistent ("maladaptive").

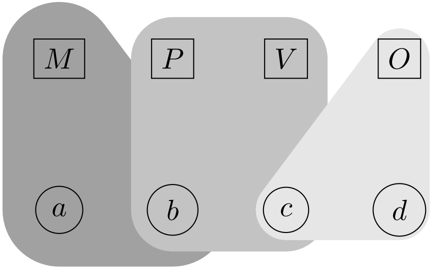
Lerner & Hâncean (2023). **Micro-level network dynamics of scientific collaboration and impact: relational hyperevent models for the analysis of coauthor networks.** *Network Science*.

`https://github.com/juergenlerner/eventnet`

Relational events that are associated with labels.

Actors co-participating in crime events of certain type:
market (M), property (P), violence (V), "other" (O).

$$e_1 = ([\{a, b\}, \{M\}], t_1)$$
$$e_2 = ([\{b, c\}, \{P, V\}], t_2)$$
$$e_3 = ([\{c, d\}, \{O\}], t_3)$$



In general, meeting, communication, or team work events might have associated "labels" (purposes, objectives, topics, ...).

Another study: the Andy Warhol diaries.

Andy Warhol mentions other persons ("event participants") in the context of topics, purposes, places, . . . ("labels").

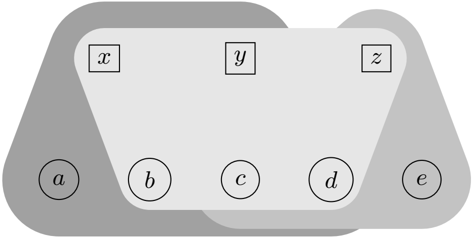
> *Catherine Guinness called New York, to Jodie Foster's place, to confirm the interview she and I were supposed to do that afternoon (. . .). Picked up Catherine Guinness and walked over to the Pierre Hotel to meet Jodie Foster. Said hello to lots of people who said hello to me. At the Pierre I saw a beautiful woman staring at me and it turned out to be Ingrid Bergman. . . .*
>
> *[AWD, November 28, 1976]*

joint work with Alessandro Lomi, Moses Boudorides, and Vitaliano Barberio

Two-mode relational hyperevent network.

Actors (circles: $a, b, \ldots$) participate in events having labels (squares: $x, y, \ldots$).

$$
\begin{aligned}
e_1 &= ([\{a, b, c, d\}, \{x, y\}], t_1) \\
e_2 &= ([\{c, d, e\}, \{z\}], t_2) \\
e_3 &= ([\{b, c, d\}, \{x, y, z\}], t_3)
\end{aligned}
$$



Many additional effects possible:
Partial repetition: of actors, labels, and mixed.
Closure: in the space of actors, labels, and mixed.

Andy Warhol diaries: selected results (prelim!).

actor activity	$0.151\ (0.033)^{***}$
actor-pair co-attendance	$0.850\ (0.057)^{***}$
actor-triple co-attendance	$0.650\ (0.051)^{***}$
actor closure	$-1.173\ (0.085)^{***}$
label popularity	$0.504\ (0.052)^{***}$
label-pair co-appearance	$0.179\ (0.029)^{***}$
label-triple co-appearance	$0.005\ (0.030)$
label closure	$0.166\ (0.098)^{\cdot}$
actor label repetition	$0.552\ (0.047)^{***}$
actor-pair label repetition	$-0.216\ (0.049)^{***}$
actor label-pair repetition	$-0.051\ (0.037)$
actor closure by label	$1.589\ (0.117)^{***}$
label closure by actor	$0.559\ (0.084)^{***}$
assortativity	$0.051\ (0.019)^{**}$
AIC	4937.124
Num. events	1,924
Num. obs.	194,224

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Practical aspects: how to specify labeled RHEM with eventnet?

`https://github.com/juergenlerner/eventnet`

Conclusion.

Labeled relational hyperevents: hyperevents occuring in a two-mode network of actors and “labels”.

The labels characterize the events, the actors are the event participants.

Gives a much richer set of possible effects.

Specified in eventnet as directed RHEM for two-mode networks.

Coevolution of collaboration and references to prior work

Research settings about

- **teams of actors** collaboratively doing **work**
- building on **prior work** (of the same or different actors).

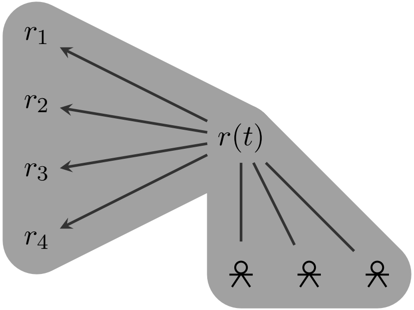
Examples include:

- Scientists publishing papers that cite previous papers.
- Inventors filling patents referencing other patents.
- Artists producing music / movies making implicit or explicit references to other music / movies.

The unit of analysis.

collaboration and references to previous work

Work $r(t)$ produced at time t by **team** of actors $\{s_1, s_2, s_3, \ldots\}$ referencing a **list** of previous works $\{r_1, r_2, r_3, r_4, \ldots\}$.



Previous works produced by same, overlapping or disjoint teams; referencing other previous works, produced by ...

RHEM for publication events.

RHEM: relational hyperevent models

Counting processes $N_{SR}(t)$ ("number of publications with authors S citing papers R").

Intensity $\lambda_{SR}(t)$ specified as **CoxPH model**

$$\lambda_{SR}(t) = \overline{\lambda}_{|S|,|R|}(t) \cdot \exp\{\beta \cdot x_{SR}(t)\}$$

Observed list of publications $(r(t_1), S_1, R_1, t_1), \ldots, (r(t_n), S_n, R_n, t_n)$ gives **partial likelihood**

$$\mathcal{L}^P(\boldsymbol{\beta}, t) = \prod_{t_i \leq t} \frac{\exp\left\{\beta^{\mathrm{T}} x_{t_i}(S_i, R_i)\right\}}{\sum_{SR \in \binom{V_{1,t_i}}{|S_i|} \times \binom{V_{2,t_i}}{|R_i|}} \exp\left\{\beta^{\mathrm{T}} x_{t_m}(I, J)\right\}}$$

Sample from the risk set (case-control sampling).

Explains relative rate of observed paper, compared to alternatives (different authors, different references).

Types of effects / research questions.

incomplete list

Effects in the collaboration network.

- Activity of authors; homophily effects.
- Repeated collaboration among pairs, triples, . . . of authors.
- Closure: collaborating with collaborators' collaborators.

Effects in the reference network.

- Popularity of papers.
- Repeated co-references to pairs, triples, . . . of papers.

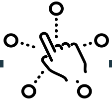
Effects in the mixed two-mode network.

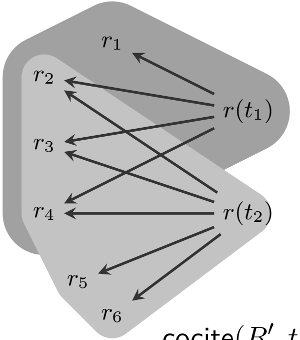
- Referencing the own paper (self-citation).
- Referencing the papers/references of collaborators.
- Collaborating with authors referencing own/same papers.
- Reference repetition / reciprocation on the author level.

Graphical illustration of exemplary effects.

Partial repetition of citation lists (“subset repetition”).

indegree popularity and repeated cocitation – many structural variants exist





Partial repetition (of order 3): the ***three*** papers $R_1 = \{r_2, r_3, r_4\}$ have been co-cited by $r(t_1)$.

$\Rightarrow R_2 = \{r_2, r_3, r_4\}$ are co-cited again by $r(t_2)$

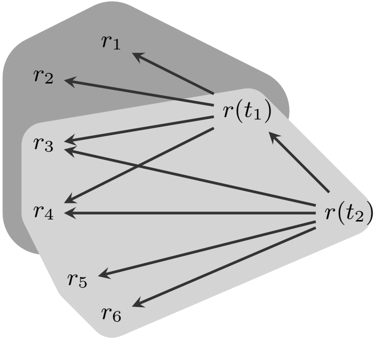
Generalizes to partial repetition of order 1, 2, 3, 4, ...

$$\text{cocite}(R', t) = \sum_{e_i : t_i < t} w(t - t_i) \cdot \mathbf{1}(R' \subseteq R_i) \ .$$

$$\text{cocite rep}^{(k)}(R, t) = \sum_{(R') \in \binom{R}{k}} \frac{\text{cocite}(R', t)}{\binom{|R|}{k}} \ .$$

Citing paper and part of its references.

variant of subset repetition

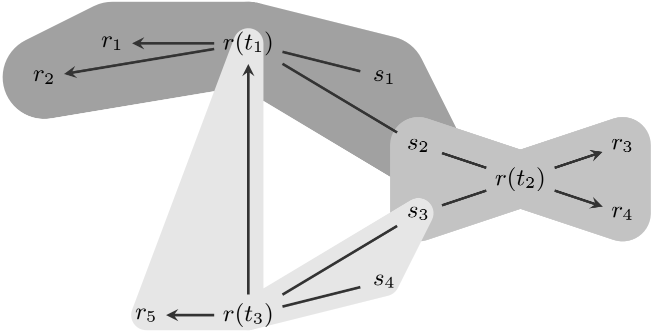


Citing a paper and part of its references: paper $r(t_2)$ cites $r(t_1)$ and some of its references $\{r_3, r_4\}$.

$$\text{cite paper and its refs}(R,t) = \sum_{\{r,r'\}\in\binom{R}{2}} \frac{\text{cite}(r,r',t) + \text{cite}(r',r,t)}{\binom{|R|}{2}} .$$

Other variants of subset repetition take authors into account: repeated co-authorship, citation repetition, . . .

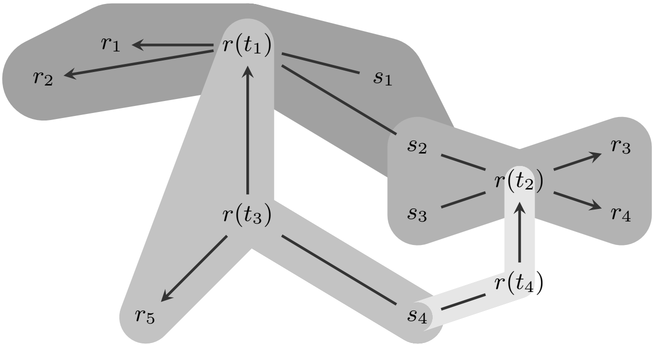
Cite coauthor's paper.



Cite coauthor's paper: author s_3 cites paper of her coauthor s_2.

$$\text{cite coauthor}(S, R, t) = \sum_{s \in S \wedge r \in R \wedge s' \neq s} \frac{\min[\text{coauth}(s, s', t), \text{author}(s', r, t)]}{|S| \cdot |R|} .$$

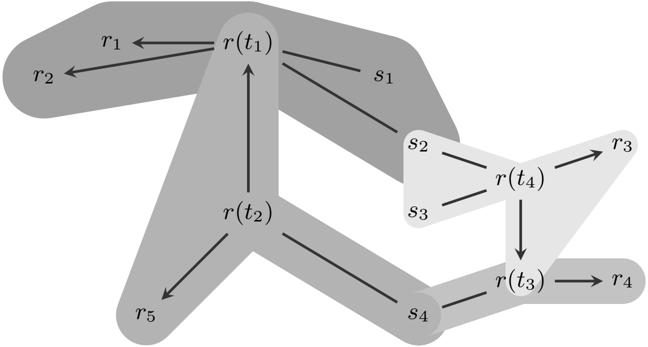
Citation repetition at the author level.



Citation repetition at the author level: author s_4 repeatedly cites $\neq$ papers of s_2.

$$\text{auth cite repet}(S,R,t) = \sum_{s \in S \wedge r \in R \wedge s' \neq s} \frac{\min[\text{cite}^{(\text{aa})}(s,s',t), \text{author}(s',r,t)]}{|S| \cdot |R|}.$$

Citation reciprocation at the author level.



Citation reciprocation at author level: s_4 cites paper of s_2; then s_2 cites paper of s_4.

$$\text{auth cite recipr}(S,R,t) = \sum_{s\in S\wedge r\in R\wedge s'\neq s} \frac{\min[\text{cite}^{(\text{aa})}(s',s,t), \text{author}(s',r,t)]}{|S|\cdot|R|} .$$

Illustrative case study and results.

Lerner, Hâncean, & Lomi: **Relational hyperevent models for the coevolution of coauthoring and citation networks**. *JRSS-A*, 2024.

`https://github.com/juergenlerner/eventnet`

Case study: data.

Data: `https://www.aminer.org/citation`.

DBLP-Citation-network V14, restricted to journal papers.

1,416,353 papers written by 1,286,941 unique authors

	parameters
num prior papers	$0.129\ (0.003)^{***}$
diff prior papers	$0.185\ (0.002)^{***}$
num prior joint papers	$0.823\ (0.004)^{***}$
author citation popularity	$-0.017\ (0.001)^{***}$
diff auth cite pop	$-0.007\ (0.001)^{***}$
collab w citing author	$0.318\ (0.003)^{***}$
closure: common coauthor	$-0.014\ (0.001)^{***}$
closure: citing common paper	$-0.112\ (0.002)^{***}$
closure: citing common author	$0.001\ (0.001)$
closure: cited by common author	$0.007\ (0.001)^{***}$
paper citation popularity	$0.170\ (0.000)^{***}$
paper-pair cocitation	$0.032\ (0.001)^{***}$
paper-triple cocitation	$-0.170\ (0.004)^{***}$
author citation repetition	$0.643\ (0.008)^{***}$
paper outdegree popularity	$0.195\ (0.001)^{***}$
cite paper and its refs	$7.360\ (0.010)^{***}$
adopt citation of coauthor	$-0.264\ (0.003)^{***}$
self citation	$1.799\ (0.012)^{***}$
cite coauthor's paper	$0.291\ (0.007)^{***}$
author-author citation repetition	$-0.381\ (0.007)^{***}$
author-author citation reciprocation	$-0.343\ (0.007)^{***}$

[-ll(null)=11,063,051; ll(full)-ll(null)=1,599,202]	contrib. to log likelihood		
	to null	in full	type
cite paper and its refs	1,150,680	237,136	R
paper citation popularity	629,911	115,535	R
author citation repetition	614,921	3,495	M
self citation	537,228	9,967	M
num prior joint papers	508,556	21,586	A
num prior papers	508,359	1,230	A
author-author citation repetition	441,856	1,545	M
collab w citing author	415,377	4,324	M
diff num prior papers	402,533	5,537	A
paper-pair cocitation	400,152	278	R
author-author citation reciprocation	359,472	1,331	M
cite coauthor's paper	355,529	849	M
adopt citation of coauthor	329,170	4,649	M
author citation popularity	251,639	198	M
diff auth cite pop	217,229	55	M
paper-triple cocitation	212,057	1,123	R
closure: citing common paper	202,071	1,682	M
closure: common coauthor	189,043	67	A
paper outdegree popularity	168,659	12,428	R
closure: citing common author	160,407	0	M
closure: cited by common author	112,673	66	M

Practical aspects: how to specify mixed two-mode RHEM with eventnet?

`https://github.com/juergenlerner/eventnet`

Summary.

Scientific publications give rise to mixed two-mode networks:

- author nodes and paper nodes;
- two relations: authors publish papers; papers cite papers;
- mixed effects (combining the two relations) make significant contributions in explaining publication events.

Conclusion.

Relational hyperevent models (RHEM) are a general model for time-ordered multi-actor events: meetings, team-work, communication, ...

`https://github.com/juergenlerner/eventnet`

Key Takeaways

1. **From Dyads to Hyper Events**:
   Extended the notion of events and related statistics for **hyper-events**, relaxing strict pairwise assumptions.
2. **REM Inference: an Overview**:
   Unified commonly used inference procedures into a **unified framework** for R(H)EMs with **practical implementation tools**.
3. **Relaxing Linearity Assumption**:
   Allowed for more flexibility by incorporating **time-varying, nonlinear, and random effects** into REMs.
4. **Advanced RHEMs**:
   Developed advanced models **with labels, outcomes, and directed hyper-events** to capture more complex dynamics.

References I

Federica Bianchi, Edoardo Filippi-Mazzola, Alessandro Lomi, and Ernst C. Wit. Relational event modeling. *Annual Review of Statistics and Its Application*, 11, 2024.

Martina Boschi, Rūta Juozaitienė, and Ernst C Wit. Mixed additive modelling of global alien species co-invasions of plants and insects. *Journal of the Royal Statistical Society Series C: Applied Statistics*, page qlaf034, 2025.

Carter T. Butts. A relational event framework for social action. *Sociological Methodology*, 38(1):155–200, 2008.

David R Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

Edoardo Filippi-Mazzola and Ernst C Wit. A stochastic gradient relational event additive model for modelling us patent citations from 1976 to 2022. *Journal of the royal statistical society series C: applied statistics*, 73(4):1008–1024, 2024.

References II

Jürgen Lerner and Alessandro Lomi. Reliability of relational event model estimates under sampling: How to fit a relational event model to 360 million dyadic events. *Network Science*, 8(1):97–135, 2020.

Patrick O. Perry and Patrick J. Wolfe. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75 (5):821–849, 2013.

Christoph Stadtfeld, James Hollway, and Per Block. Dynamic network actor models: Investigating coordination ties through time. *Sociological Methodology*, 47(1):1–40, 2017.

Duy Vu, Philippa Pattison, and Garry Robins. Relational event models for social learning in moocs. *Social Networks*, 43:121–135, 2015.