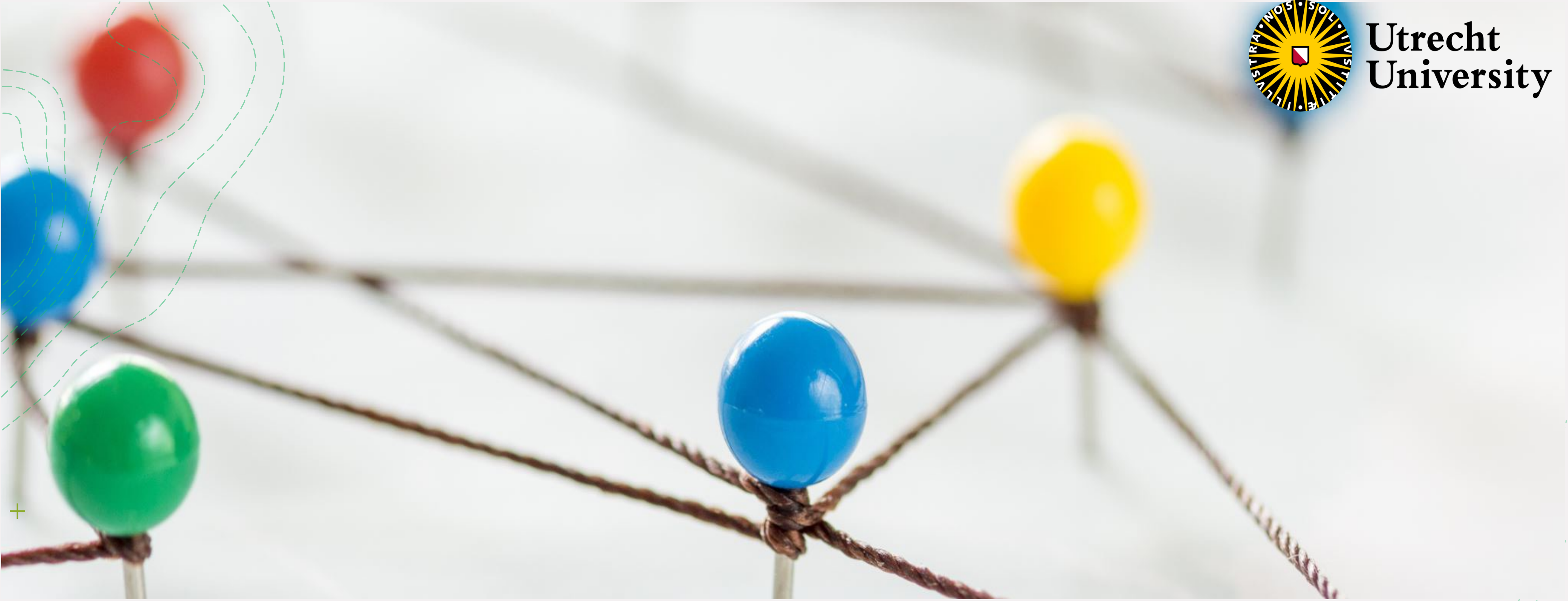


Utrecht University

Social network analysis

Mahdi Shafiee Kamalabad

Social network analysis

+ **14:00- 15:00**

Statistical approaches to network analysis

+ **15:00-15:15**

Break

+ **15:15-16:30**

Practical

+ **16:30- 17:00**

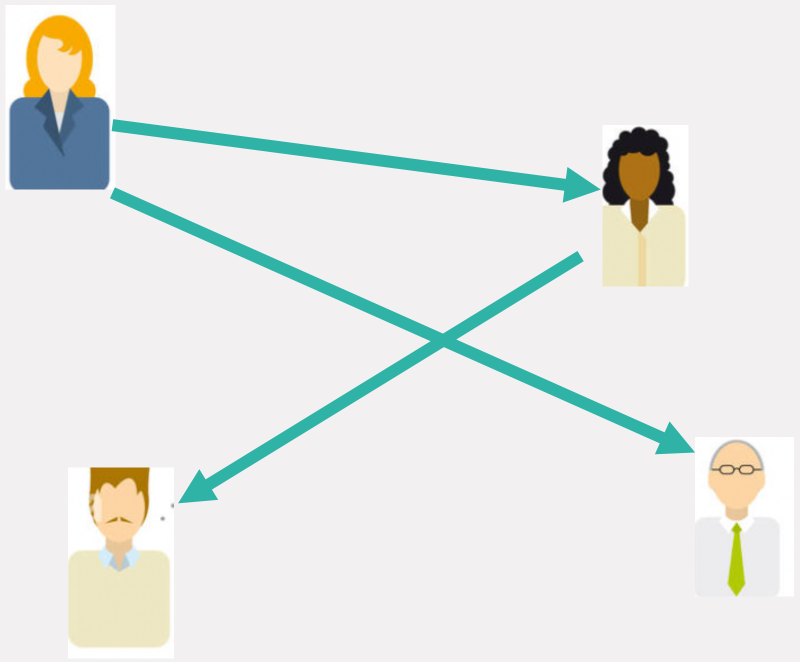
Discussion

Contents

- **Social networks**
- Relational Event History Data (**REH**)
- Relational Event Model (**REM**)
- **E**xponentioal **R**andom **G**raph **M**odel (**ERGM**)
- **S**tochastic **A**ctor **O**riented **M**odel (**SAOM**)
- Longitudinal network analysis: ERGMs – SOAM, TERGMs - REMs

Network

- Representations of **relational data**.
- **Nodes** (actors/vertices) represent **entities** while the **links** (edges/ties) connecting them represent any form of **interaction** or **connection** between the entities.



Social networks

+ A **Relation** defined on a collection of **individuals (actors).**

For example

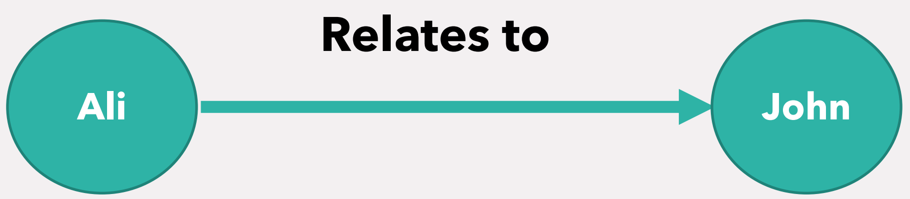
Ali goes to John for **advice**…

Ali considers John as a **friend**…

Ali sends an **email** to John…

Ali **calls** Join…

…



Social networks

+ A **Relation** defined on a collection of **individuals.**

Ali $\xrightarrow{\textbf{Relates to}}$ **John**

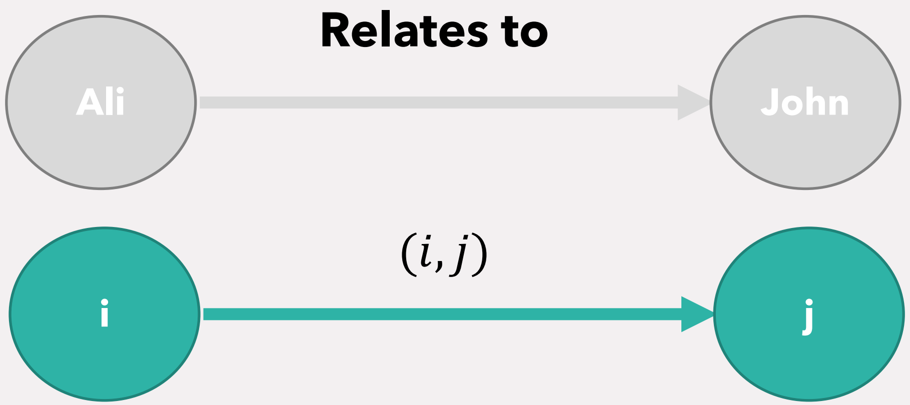
Tie present: On

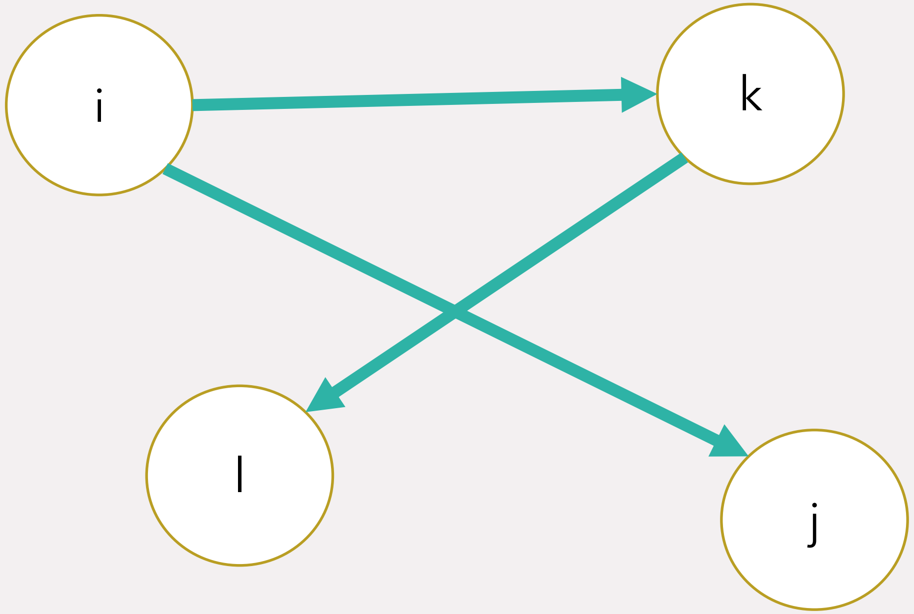
Tie absent: Off

Social networks

A network is a **Graph**: $G(V, E)$, on

Individuals/actors: $V=\{1,2,\ldots,n\}$
Relation/edges: $E \subseteq \{(i,j) : i,j \in V\}$





Tie present: On

Tie absent: Off

Apollo 13

Note that actors can be:

Such as:

- Countries
- Humans
- Animals
- Organiazations
- …

Relational Event History Data (REH data)

Relational Event History Data (REH data)

+ REH is a type of network data.
+ REH data contain detailed information ***what*** happened (message, email, talk, etc.), ***when*** it happened (time), and ***who*** were involved (sender, receiver).

+ They are time-stamped interactions.

+ REH data contains at least **receiver** (target), **sender** (sourse), and **time/order**.

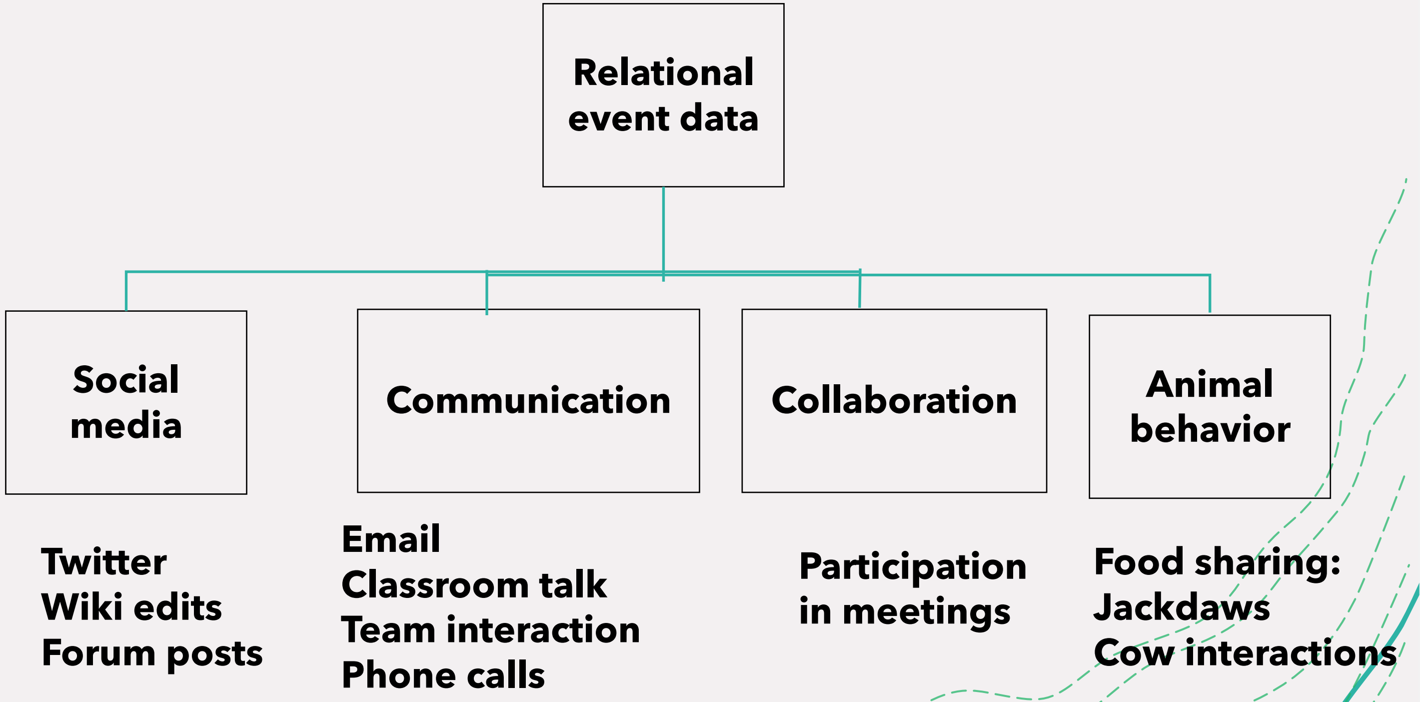
+ A **relational event**: "a **discrete event** generated by a social actor (the 'sender') and directed toward one or more targets (the 'receivers'), … **at a certain point in time**" (Butts, 2008, p. 159).
+ Event = (sender, receiver, time, …)

Why study relational events?

+ Relational events are everywhere, and ***increasingly available*** due to the development of technology.
+ Often in the form of Big Data

e.g.

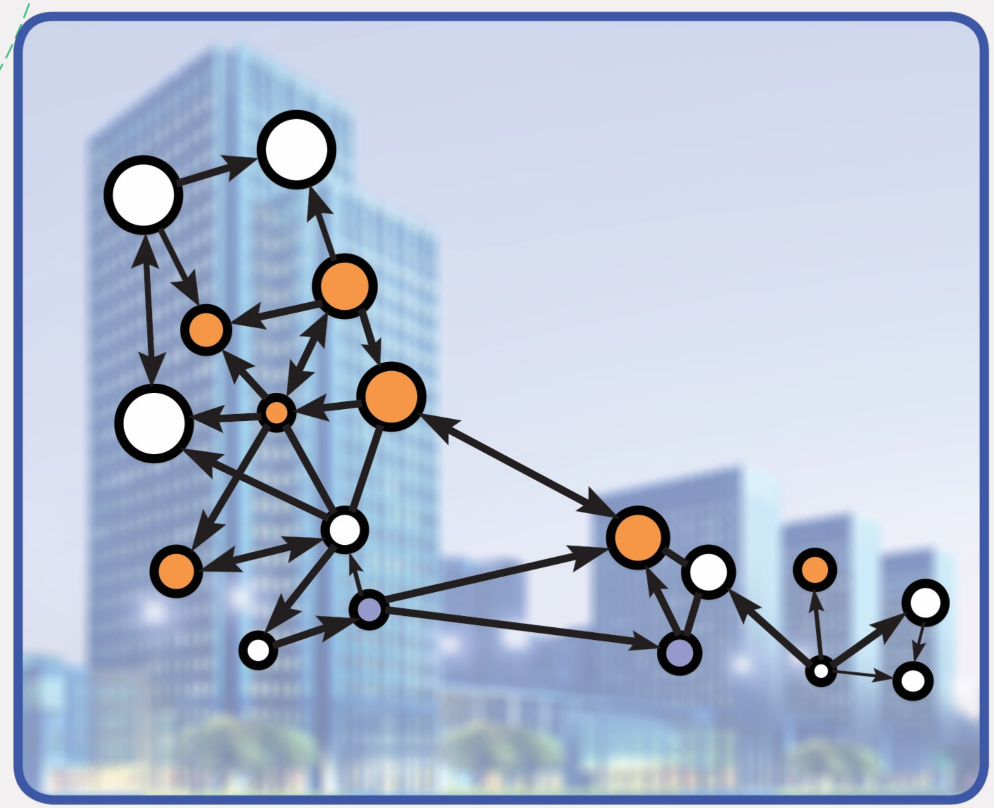
+ Sociometric badges, digital communication (email), video monitoring, etc.



What we are interested in knowing from these data

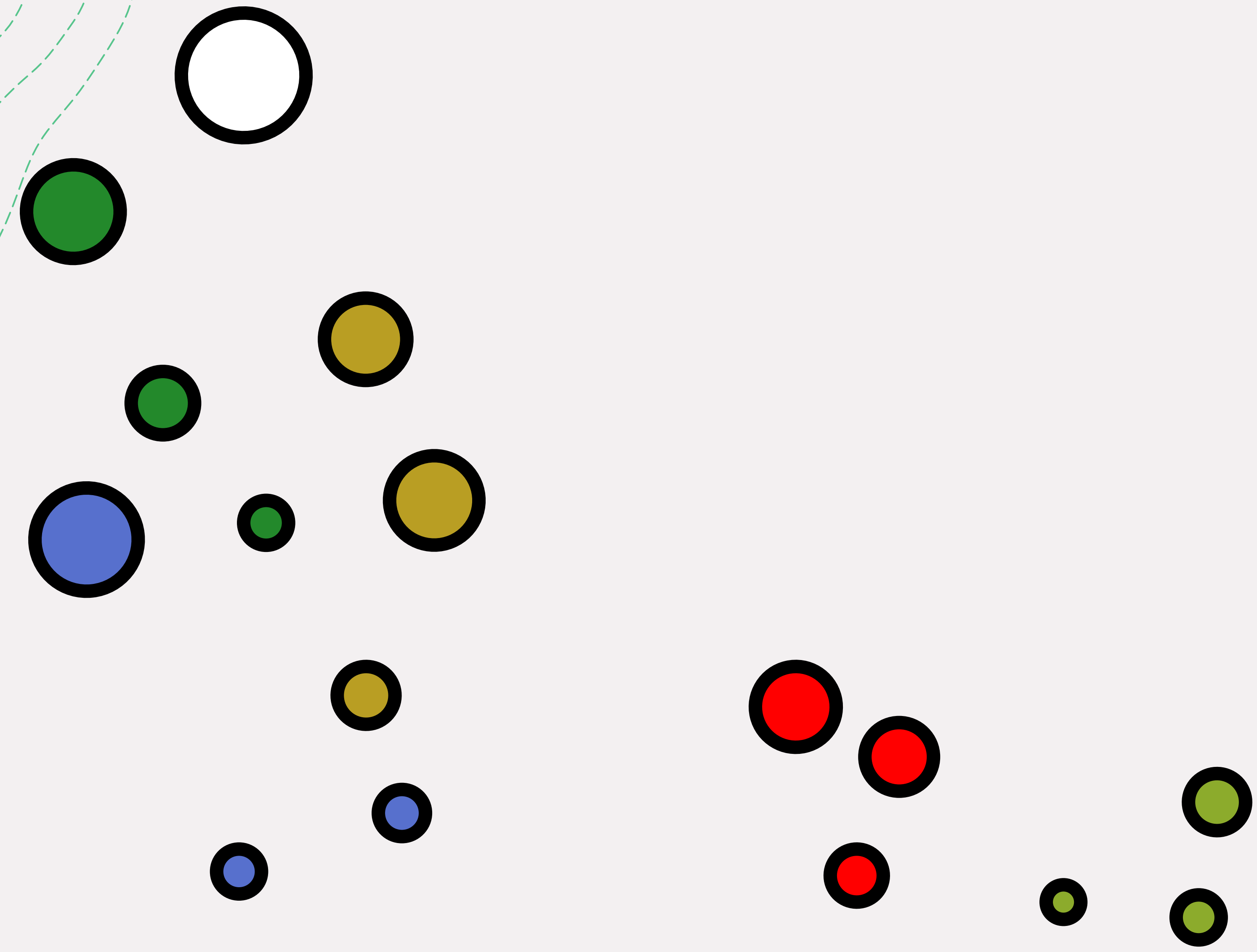
- **Complex interaction behavior** in dynamic social networks.
- **Who** does **what** and **when**?
- Predicting **what/when** is the next event? Or who is going to be the next sender/receiver and when?
- How **interaction dynamics change** over time.
- How the past **affects** the future.
- What are **true drivers of interaction** processes?
- ...

Example



Employees in organizations share information with each other via email.

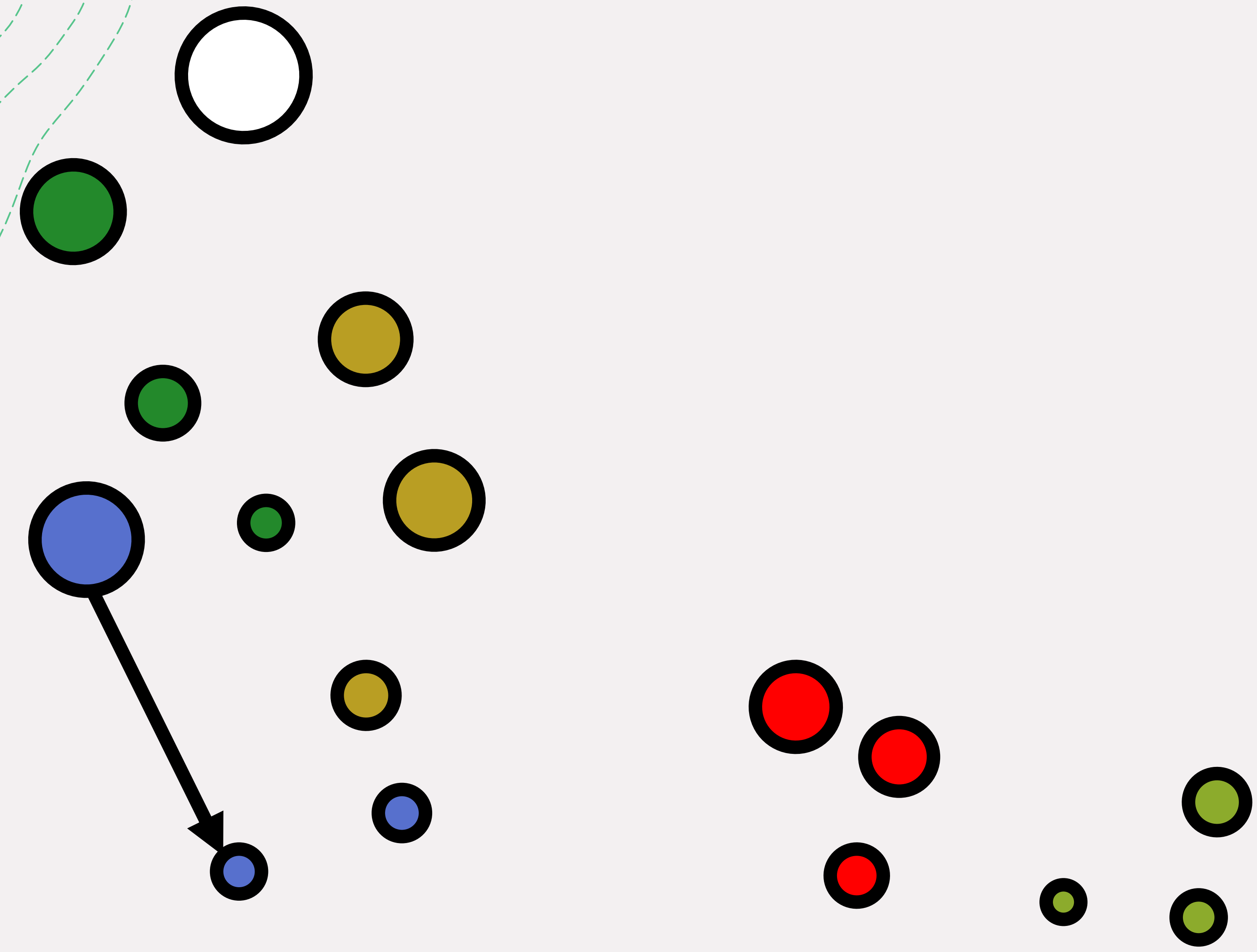
*How (**fast**) do employees share information with coworkers?*



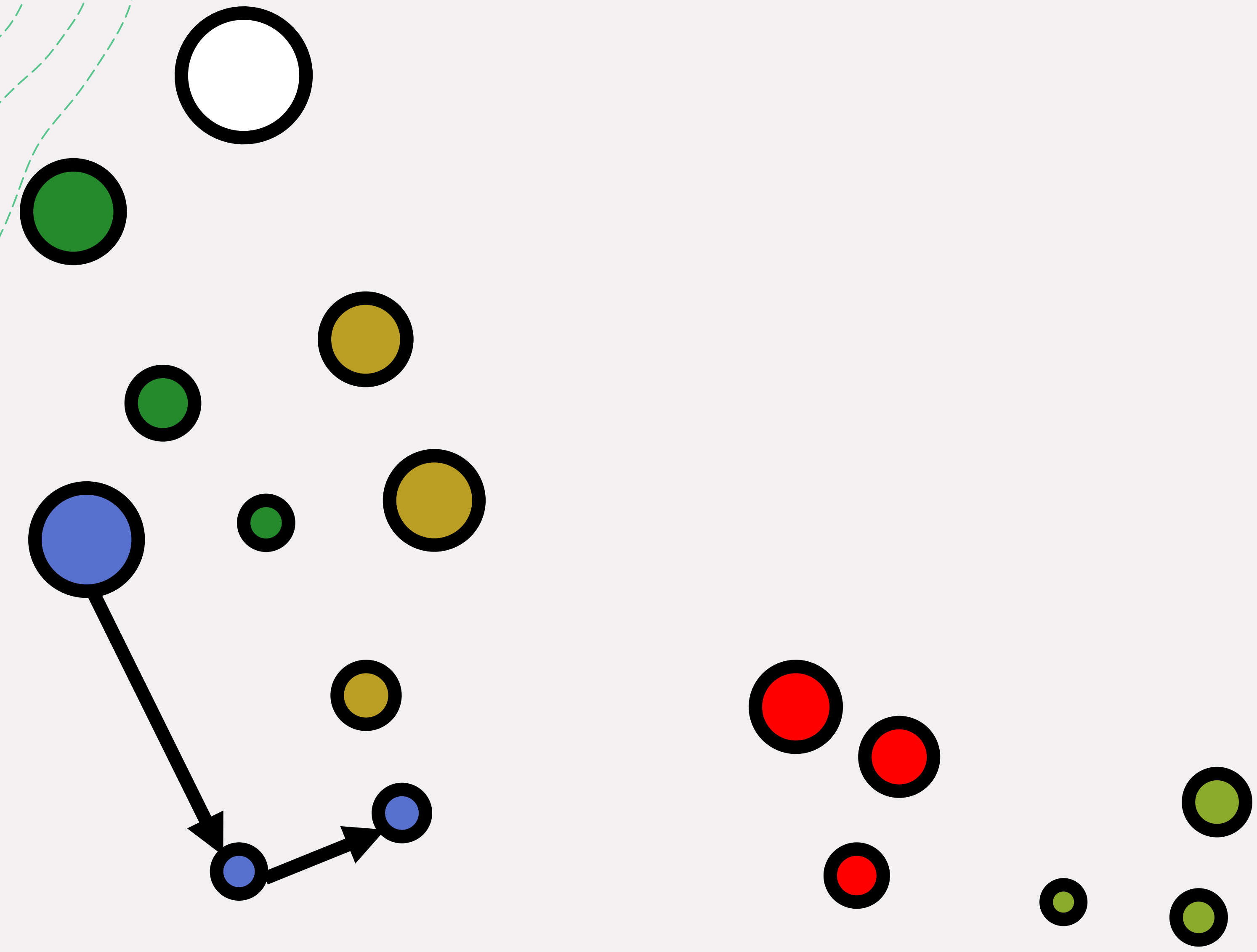
Relational Event History Data

time	sender	receiver

Relational Event History Data

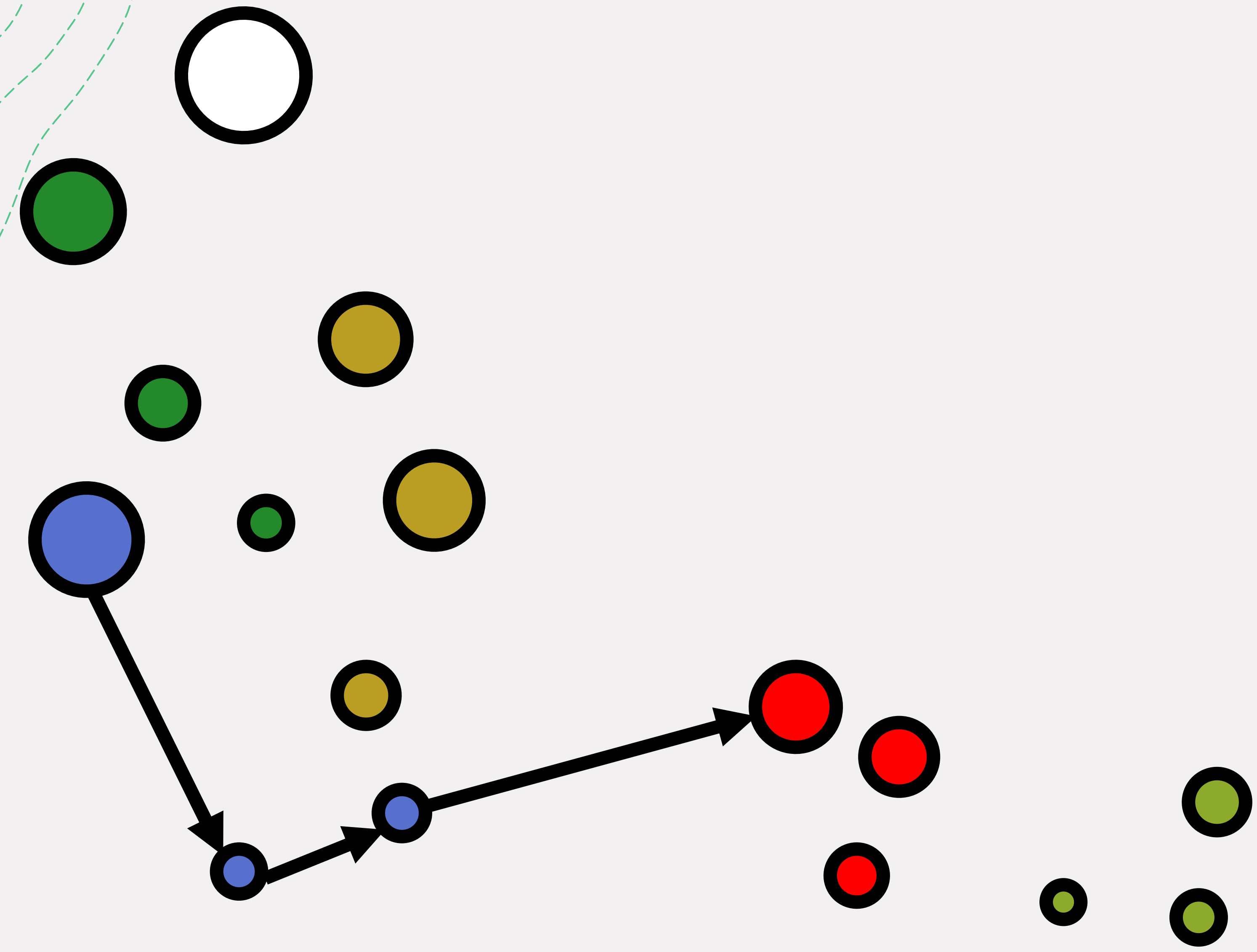


time	sender	receiver
00:03:34		

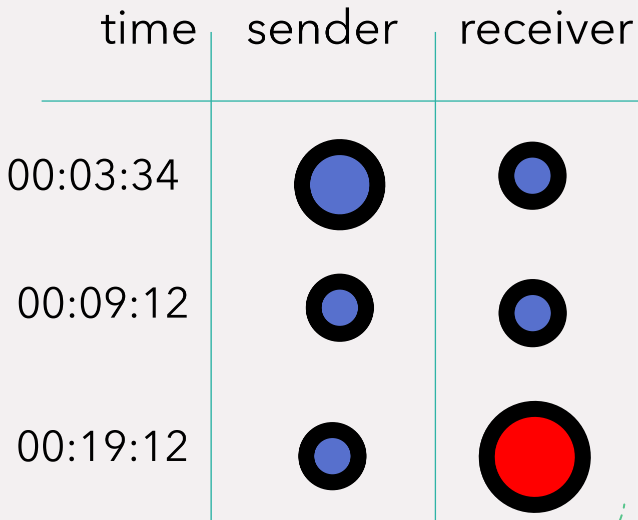


Relational Event History Data

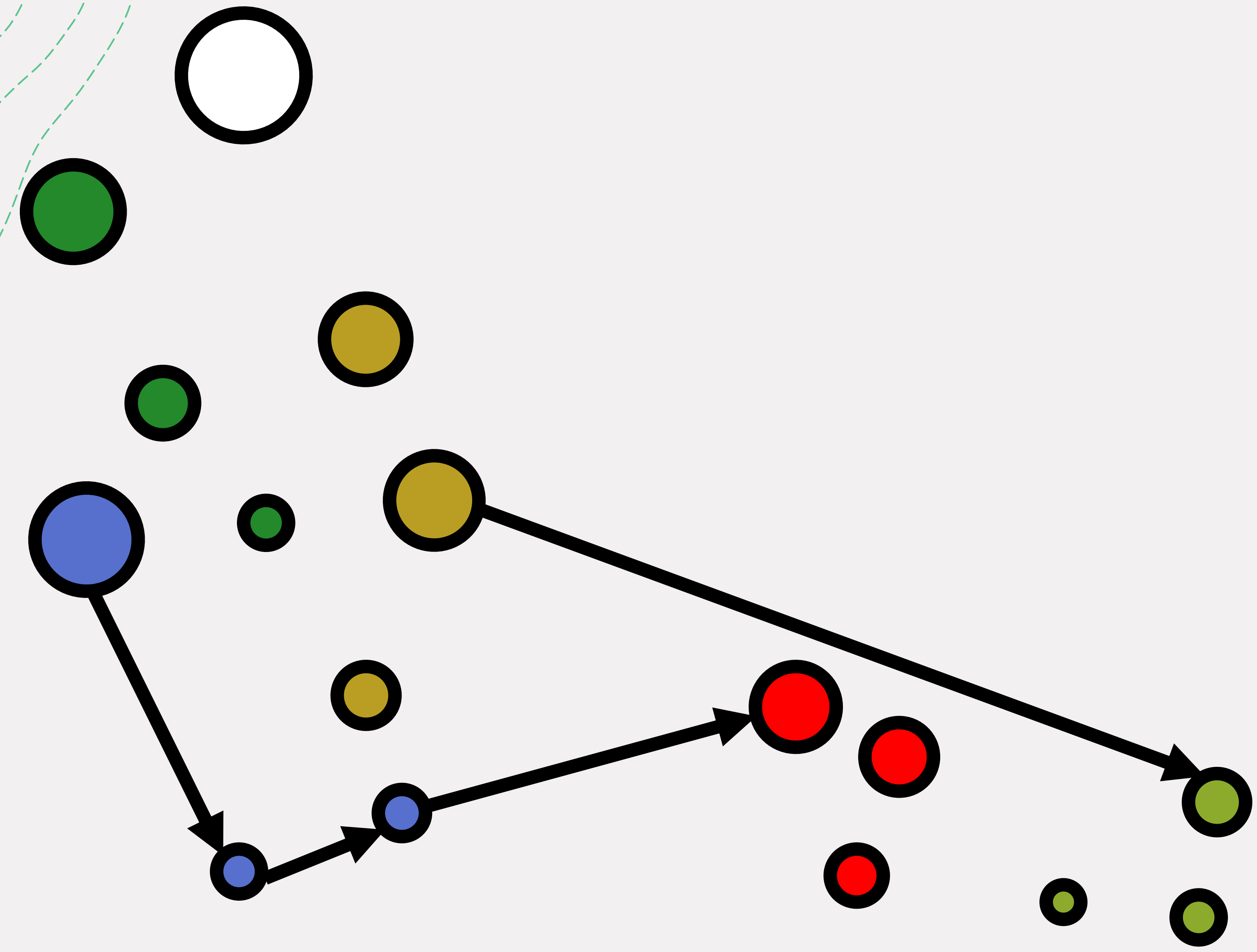
time	sender	receiver
00:03:34		
00:09:12		

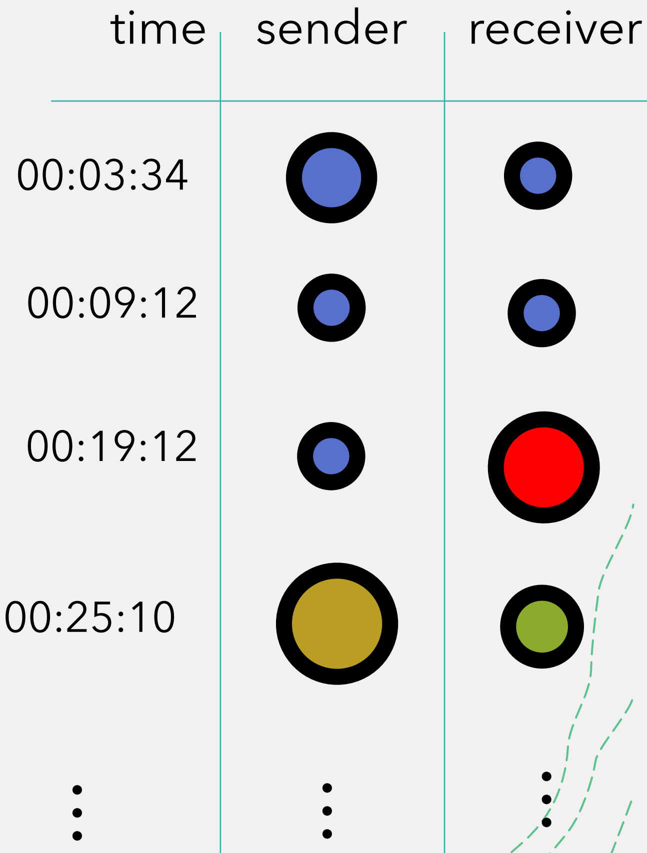


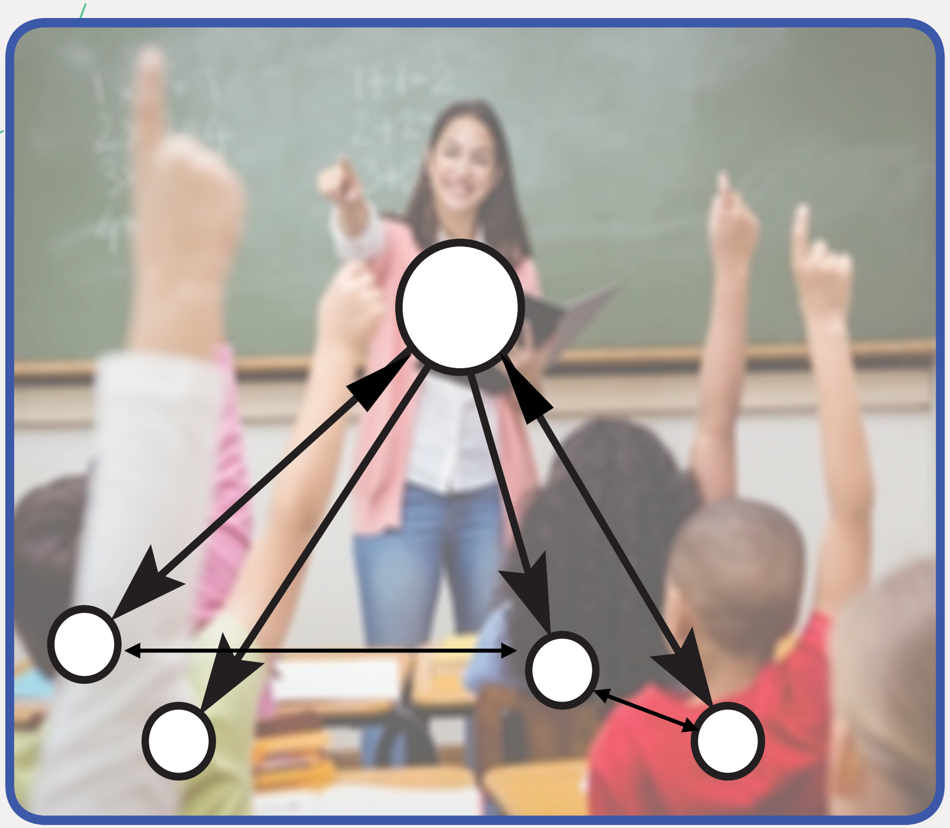
Relational Event History Data



Relational Event History Data





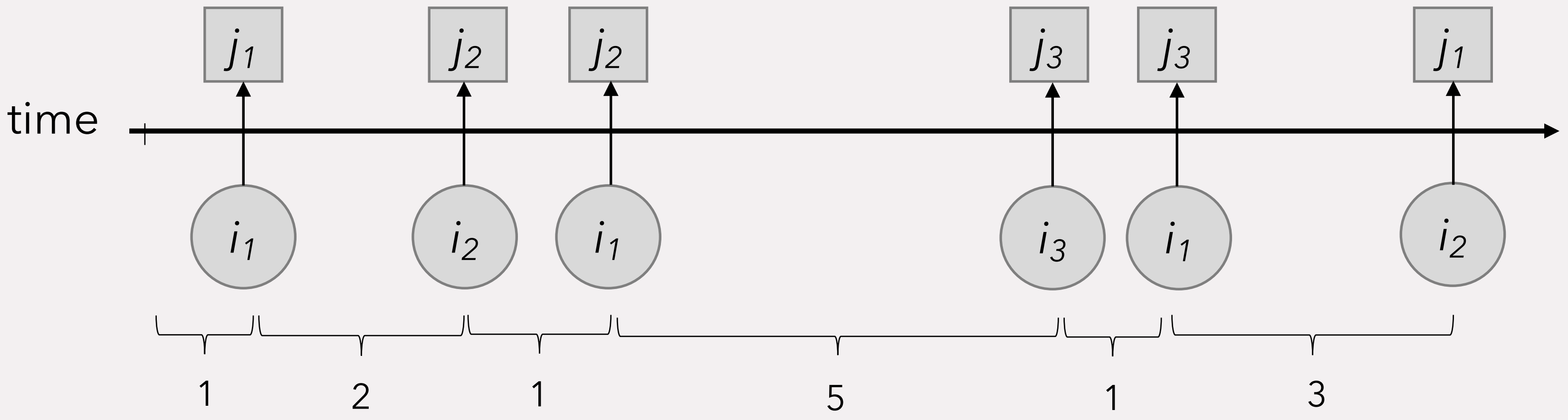


Teachers and students interact with each other in classrooms

How *do the teachers and students interact?*

Can we predict ***when*** *defiant behavior will occur?*

REH Data



Event sequence translated into a data frame:

time	sender/source	receiver/ target
0	i_1	j_1
3	i_2	j_2
4	i_1	j_2
9	i_3	j_3
10	i_1	j_3
13	i_2	j_1

Model to analyze REH data:

Relational Event Models (REMs)

R packages:

relevent

survival

remstats

Relational Event Models (REMs)

+ First introduced by Butts (2008)

+ Combination of **event history analysis** and **network** **analysis**

Relational Event Models (REMs)

- **The dependent variable**: ***what*** will happen next, ***when*** will it happen, and ***who*** will be involved.
- For each **dyad** (actor ***i***, actor ***j***) at time ***t*** there is a **rate parameter** λ which is a **loglinear function** of predictors.

+ The **propensity of an event** to occur is defined via its **hazard (rate)**.
+ Each event that is possible at a given moment has a **non-zero hazard.**
+ Larger hazards correspond to higher propensities (**the concept of risk set!**)

$$\log \lambda(i,j,t) = \beta_1\, x_1(i,j,t) + \beta_2\, x_2(i,j,t) + \beta_3\, x_3(i,j,t) + \ldots$$

Relational Event Models (REMs)

- **The predictor variables**:

  + **Actor characteristics** (tenure of employees x_1, hierarchy, gender,...)
  + **The past** (volume of past interactions x_2, ...)
  + **External factors** (epidemic situation x_3,...)

$$\log \lambda(i,j,t) = \beta_1\, x_1(i,j,t) + \beta_2\, x_2(i,j,t) + \beta_3\, x_3(i,j,t) + \ldots$$

Fitting a REM

+ requires the **estimation** of the **probability** that a particular sequence of events transpired as a function of exogenous and endogenous factors.

+ To capture this, **each event** is given **a rate, or a frequency of occurrence, $\lambda(i,j,t)$.**

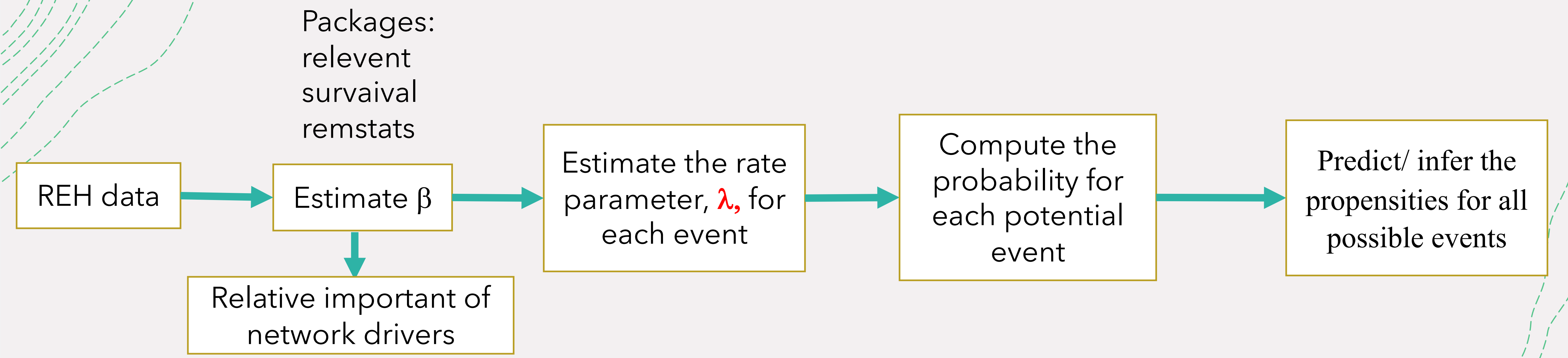
+ Events that are common have high rates, and events that are rare have low rates.

+ The **rate** for each dyad is a **function** of **statistics** such as **inertia**, **reciprocity**, **gender** of actors as well as **parameters** that represent the **sign** and **strength** of the statistics' effects.

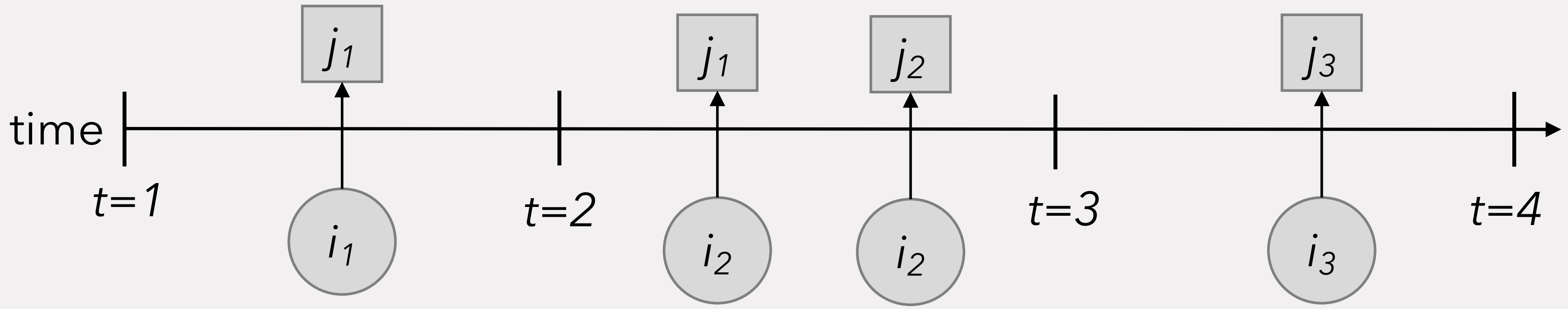
$$\log \lambda(i,j,t) = \beta_1\, x_1(i,j,t) + \beta_2\, x_2(i,j,t) + \beta_3\, x_3(i,j,t) + \ldots$$

β that represents the magnitude of a particular effect (Stadtfeld et al., 2018).

Relational Event Models (REMs)

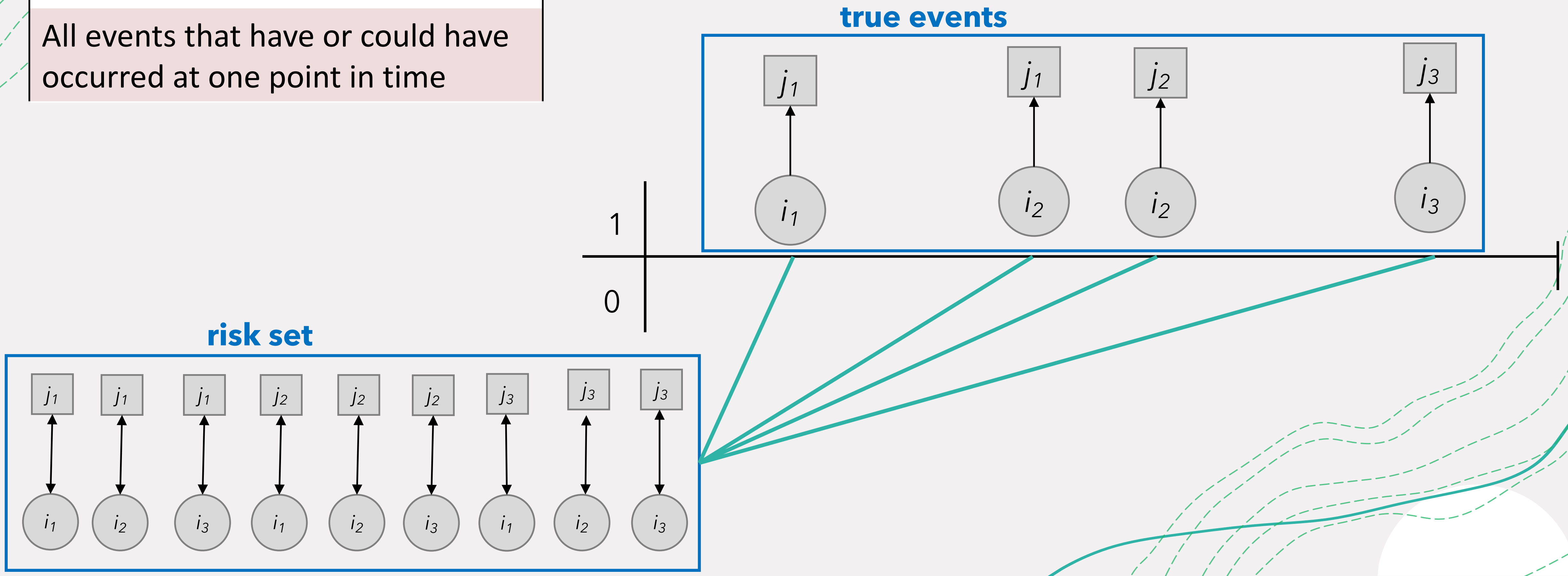


Risk set



risk set

All events that have or could have occurred at one point in time

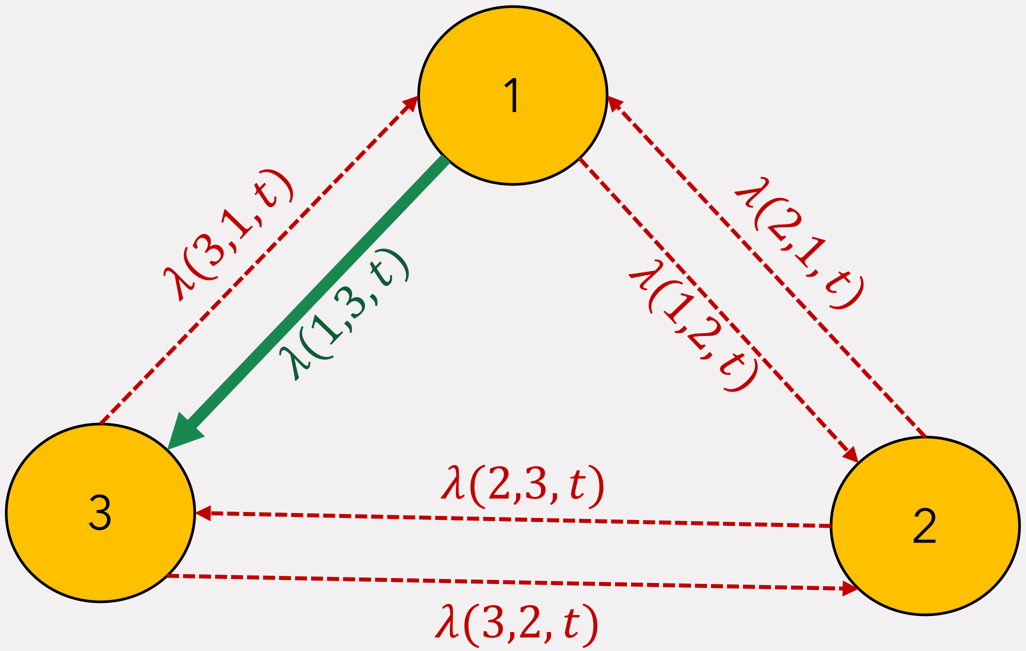


Example: Relational Event History data (REH)

Sender (i)	Receiver (j)	Time (t)
A	C	2
C	A	2.4
B	D	3

Sender	Receiver	Time (h)	$\hat{\beta}_1$ **Inertia**	$\hat{\beta}_2$ **Reciprocity**	$\hat{\beta}_3$ **Gender**	**Rate**	**Prob of event**
A	B	2	0	0	1	λ_{AB}	$\lambda_{AB}/\sum\lambda$
A	**C**	**2**	**0**	**0**	**1**	λ_{AC}	$\lambda_{AC}/\sum\lambda$
B	A	2	0	0	0	λ_{BA}	$\lambda_{BA}/\sum\lambda$
B	C	2	0	0	0	λ_{BC}	$\lambda_{BC}/\sum\lambda$
C	A	2	0	0	0	.	
C	B	2	0	0	0	.	
A	B	2.4	0	0	1	.	
A	C	2.4	1	0	1		
B	A	2.4	0	0	0		
B	C	2.4	0	0	0		
C	**A**	**2.4**	**0**	**1**	**0**		
C	B	2.4	0	0	0		
.	.	.	.	.			
.	.	.	.	.			

Probability that a particular sequence of events transpired



- Time till the next event: $t_m - t_{m-1} \sim \mathrm{Exp}(\Sigma\lambda(s',r',t))$, $(s',r') \in R_t$
- Which relational event: $P(s, r) = \dfrac{\lambda(s\;\;,r\;\;,t)}{\Sigma\,\lambda(s',r',t)}$
- REM captures the probability of the full sequence by **tuning the rate parameters** and **maximizing the likelihood of each observed event**.

R packages for implementing REM

For estomating the parameters:

+ relevent -------(**rem.dyad(), rem()**)
+ survival --------(**coxph()**)

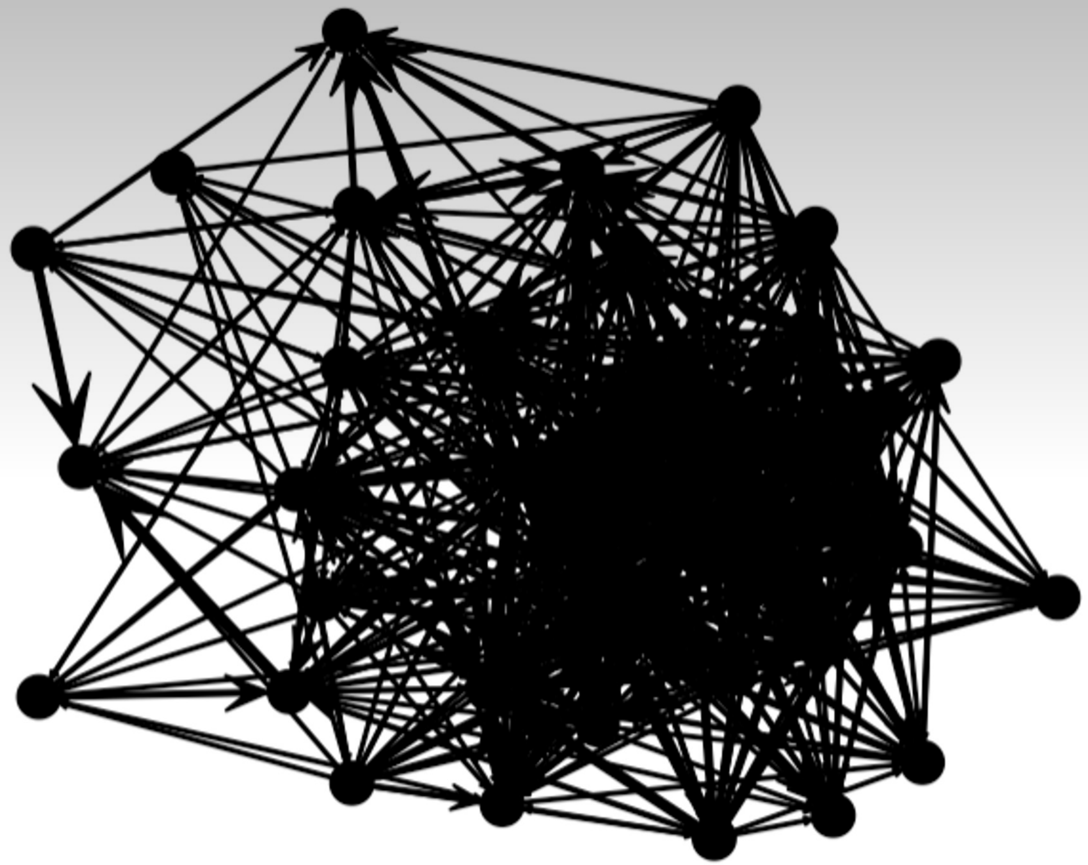
Estimation of the parameters

+ For computing statistics:
+ remstats -------- (**remstats()**)

Computing statistics

+ **Note**: For **<u>relevent::rem()</u>**, and **<u>survival::coxph()</u>** you need to compute the statistics first.

Twitter



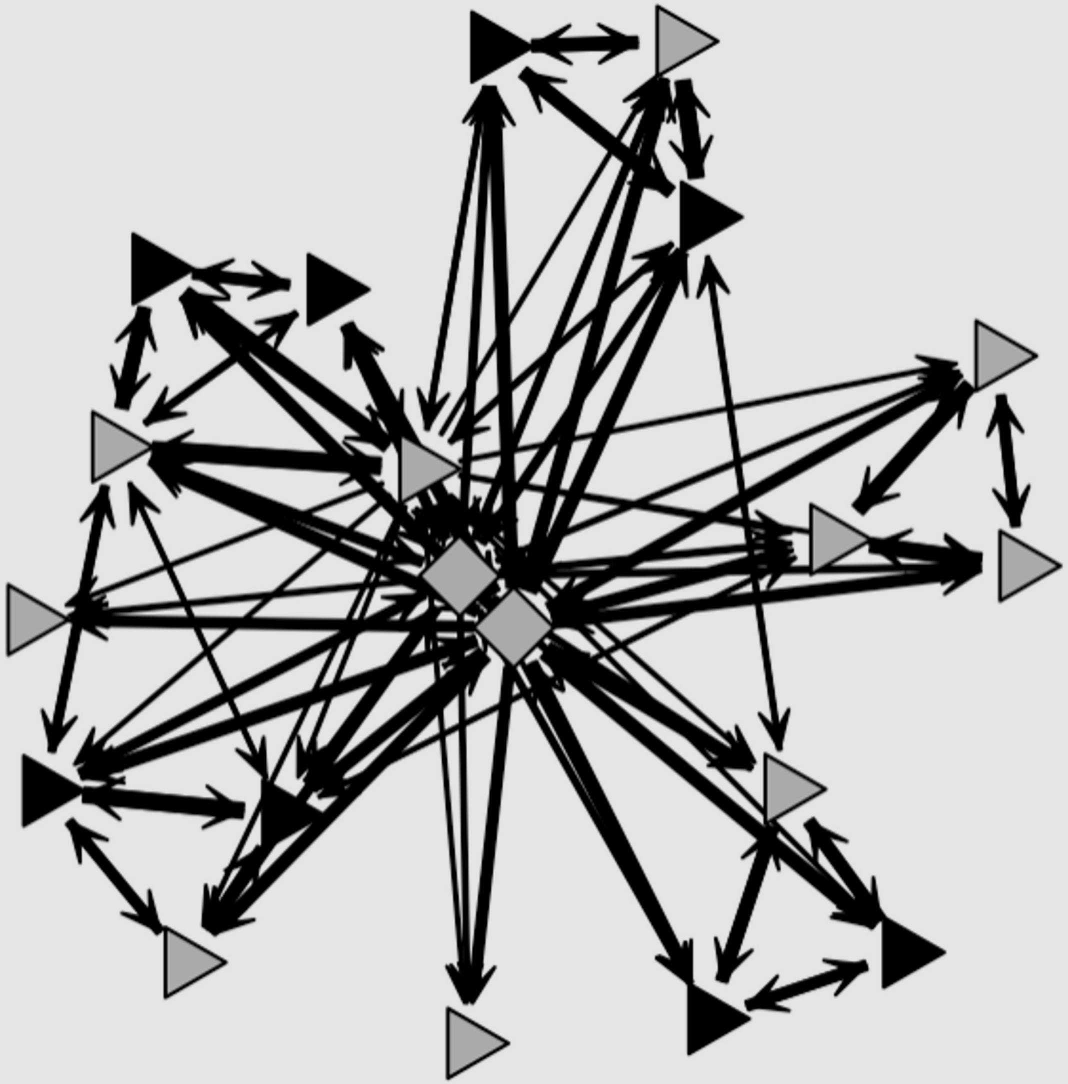
```
Twitter1 <-  rem.dyad(Twitter_data_rem3,n=39, effects = c("PSAB-BA", "PSAB-BY"), ordinal = FALSE, hessian = TRUE)

summary(twitter1)
```

Relational Event Model (Temporal Likelihood)

	Estimate	Std.Err	Z value	Pr(>|z|)
PSAB-BA	-6.757	0.5777.	-11.697	< 2.2e-16 ***
PSAB-BY	-8.285.	0.2040.	-40.611	< 2.2e-16 ***

Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

Class

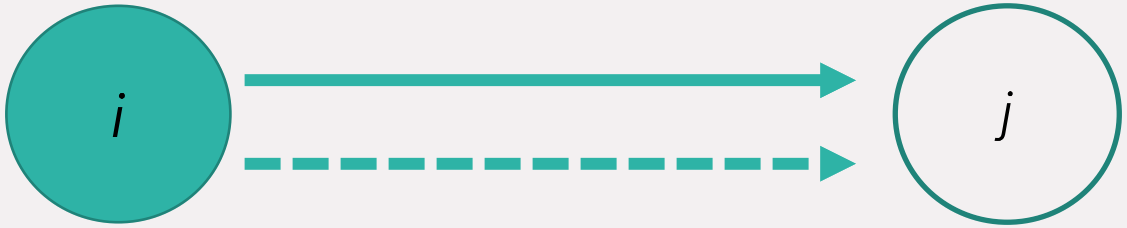


```
classfit5<-rem.dyad(Class,n=20,effects=c("CovSnd","CovRec","RRecSnd","RSndSnd",
"PSAB-BA","PSAB-AY","PSAB-BY"), covar=
list(CovSnd=cbind(ClassIntercept,ClassIsTeacher),
CovRec= cbind(ClassIsTeacher,ClassIsFemale))

summary(classfit5)
```

	Estimate	Std.Err	Z value	Pr(>|z|)
RRecSnd	2.429210	0.155367	15.6353	< 2.2e-16 ***
RSndSnd	-0.986720	0.144668	-6.8206	9.068e-12 ***
CovSnd.1	-5.003468	0.090610	-55.2197	< 2.2e-16 ***
CovRec.1	-0.722667	0.141950	-5.0910	3.562e-07 ***
PSAB-BA	4.622159	0.137602	33.5908	< 2.2e-16 ***
PSAB-BY	1.677639	0.164930	10.1718	< 2.2e-16 ***
PSAB-AY	2.869985	0.103114	27.8330	< 2.2e-16 ***

Inertia

The **tendency** of person *i* to continue to initiate events towards person *j*, as a function of the **volume of past events** from *i* to *j*.

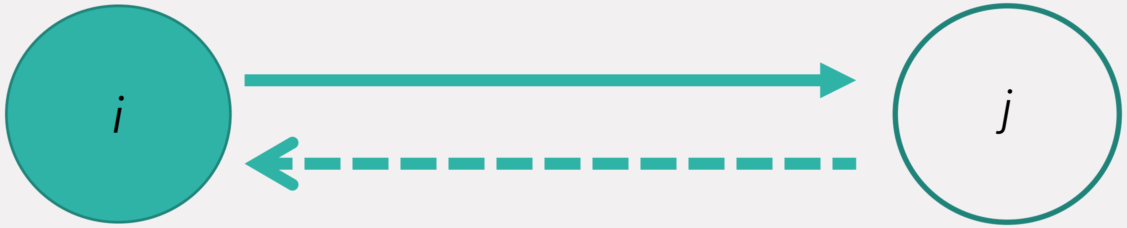


Example-- Inertia

- A teacher exhibits a tendency to ask students they have been frequently asked questions in the past. The true effect value $\beta_{INERTIA}$ of inertia statistics is positive, and the REM should find a positive and significant estimate.

Reciprocity

The tendency of person *i* to initiate events towards person *j*, as a function of the volume of past events *i* received from *j*.



Example-- Reciprocity

+ Teachers and students do reciprocate over time in ways that are not influenced by reciprocating in the past.

+ This can be understood as a teacher who does not show the tendency to ask the same student who has already asked a question more often.

+ When this is the case, the effect of the reciprocity statistic should be small.

+ The results would not show a tendency for reciprocating (i.e., connecting through a chain of questions and answers).

Example of Kitts et al. (2017)

+ **Research question:** Do hospitals engage in the social norm of reciprocity when exchanging patients, instead of sending them to the hospital that can offer the best service for the patient?

+ Using REM, they examined reciprocity in over 4,000 patient exchanges between 21 hospitals in a region of Italy, spanning 5 years.

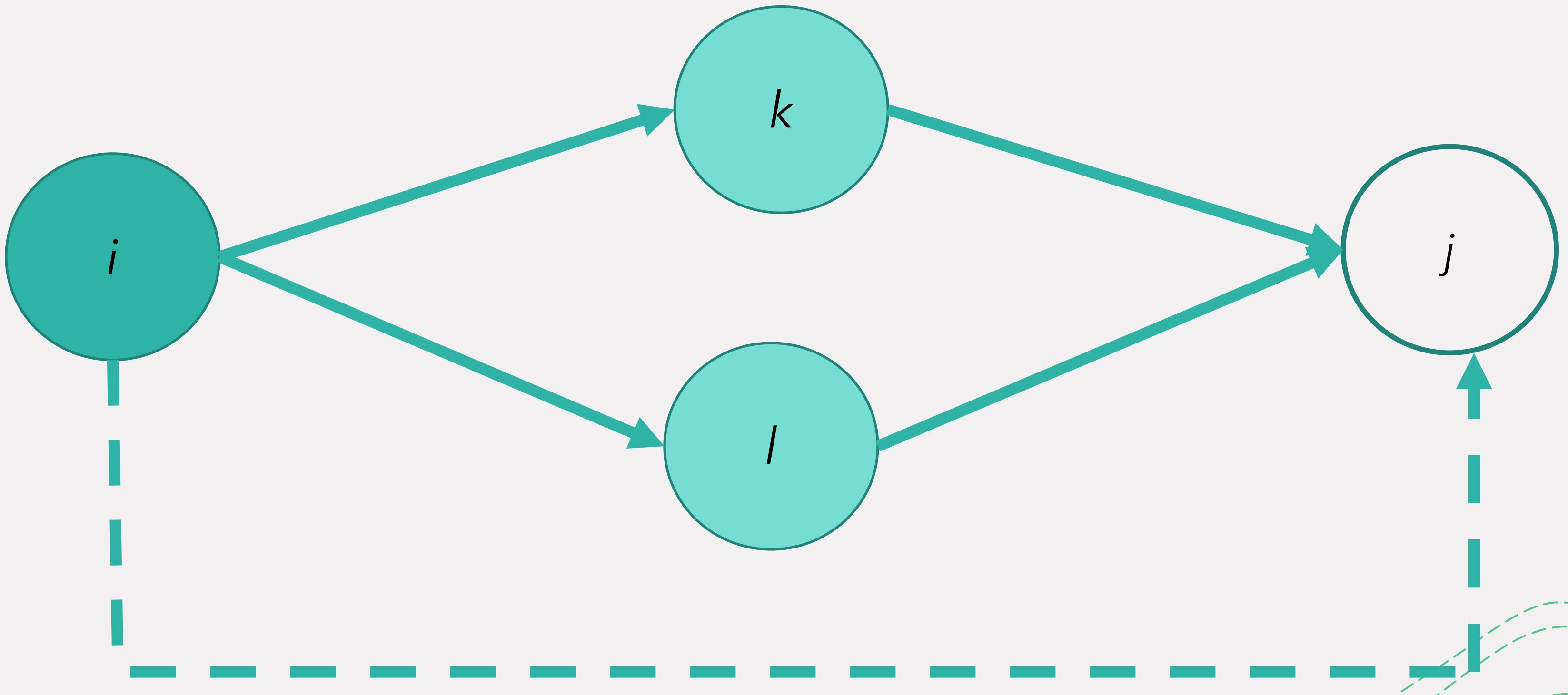
+ **Result**: hospitals do reciprocate patient exchanges over time in ways that are not explained by the availability of beds, the quality of service, or the specialization of hospitals.

The predictor variables: $(X_1, X_2, X_3, \ldots)$

+ Inertia…
+ Reciprocity
+ Transitivity
+ In(out) degree sender/receiver
+ …
+ Age
+ Hirarchy
+ Same location
+ …

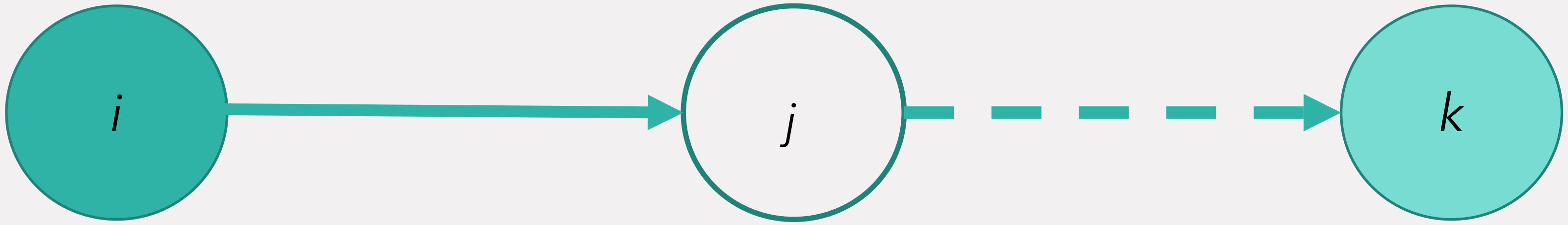
Transitivity

The tendency of person *i* to initiate events towards person *j*, as a function of the volume of past events *j* received from others to whom *i* had sent events.



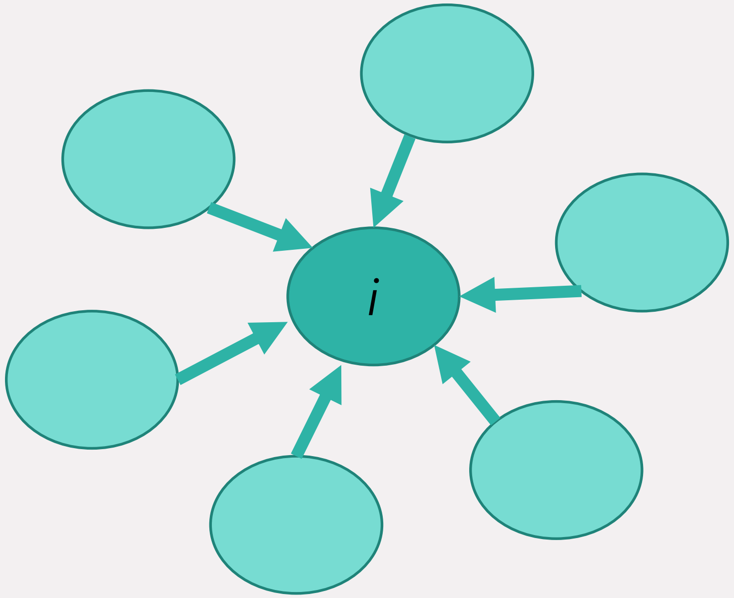
Participation shift AB-BY ("turn receiving")

The tendency of an initial receiver *j* of an event to, **in turn**, direct the next event to another person *k*.



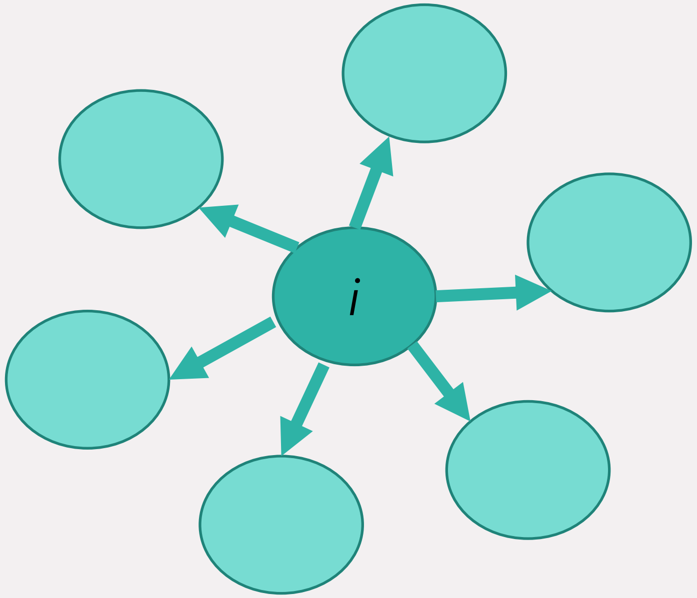
In-degree

In-degree is the number of connections that point inward at a vertex. Actors with high in-degree are impacted by multiple other actors.



Outdegree

Out-degree is the number of connections that originate at a vertex and point outward to other vertices.



In a nutshell, REM is suitable for

- **Estimating** the relative important of *network driver effects* β.
- **Testing** *temporal social theories* via competing statistical models.

- **Predicting** future events, *what* will happen next, *when* it will happen, and *who* will be involved. REMs predict the occurrence of the next event in a temporally distributed sequence of events (Marcum & Butts, 2015)
- This means that, in REM, the dependent variable can be the occurrence of the next event in a sequence, which is modelled as a function of the sequence of past events.

- **Understanding** how interaction behavior *changes in continuous time*.
- Why did some node tie to another at this point in time and not previously?

Longitudinal network analysis: ERGMs - TERGMs – SAOMS and REMs

+ The **choice** of network inference model depends on **how time is recorded**.

Four main network inference models:

+ Exponential Random Graph Models
+ Stochastic Actor Oriented Model (SAOM)
+ Temporal Exponential Random Graph Models and
+ **Relational Event Models**

Exponential Random Graph Model (ERGM)

R package: ergm

ERGM: **E**xponential **R**andom **G**raph **M**odel

+ Goal "**describe parsimoniously the <u>local selection forces</u> that shape the global structure of a network**" (Hunter et al. 2008). (the processes that influence link creation).

+ ERGMs are tie-based statistical models for understanding **how and why social network ties arise**.

ERGM

- Let say **G** is a graph.
- **Summary measures *z*(*G*):** or "network statistics,"

- Network statistics such as the **number of edges** in *G*, **triad census**, and so on.

- The ERGM assigns probability to graphs according to these statistics:

$$\mathbf{P_{\theta}\ (\mathit{G}) = c\mathit{e}^{\theta_1 z_1(\mathit{G}) + \theta_2 z_2(\mathit{G}) + \cdots + \theta_p z_p(\mathit{G})}}$$

c is a normalizing constant.

This is the probability of a given network.

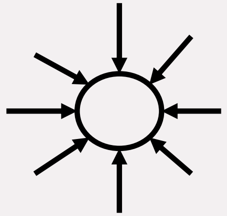
Remark: Network Statistics

- The network statistics are **counts of the number of network configurations** in the given network G, or some function of those counts.
- These configurations are **small, local subgraphs** in the network.
- The **probability** of the network depends on **how many of those configurations** are present.
- The parameters inform us of the importance of each configuration.

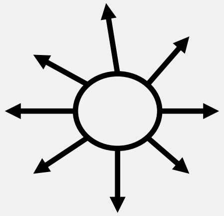
Different considered network statistics. More detailed explanations can be found in Lusher et al. (2013)

Reciprocity

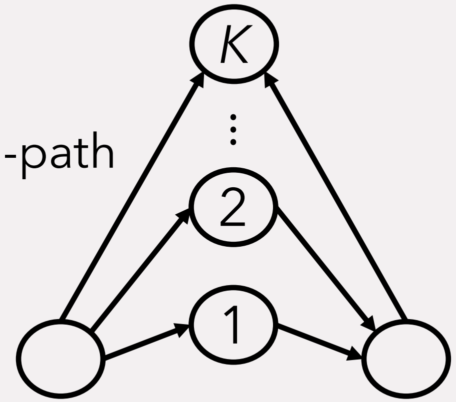
Popularity (K in-stars)



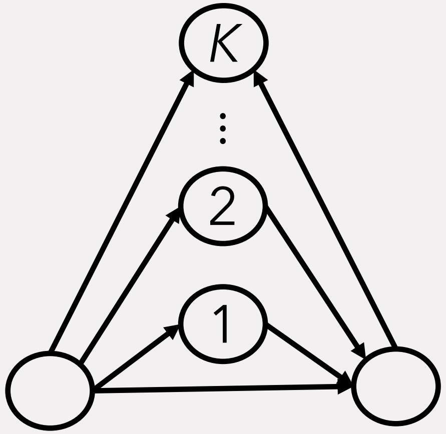
Activity (K out-stars)



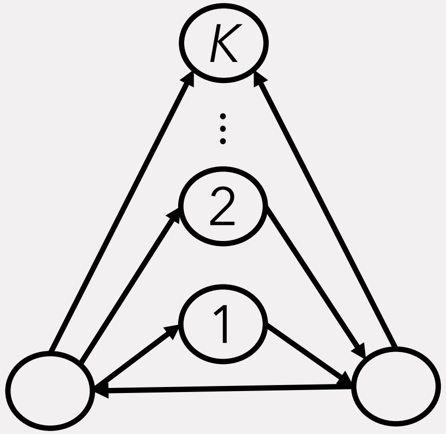
K 2-path



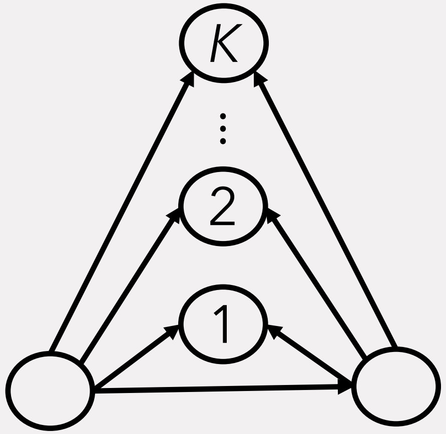
Path closure AT-T



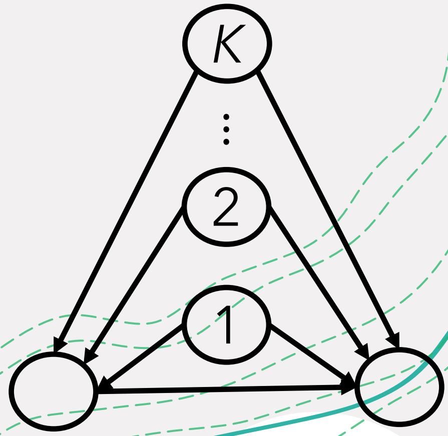
Cyclic closure AT-C



Activity closure AT-U



Popularity closure AT-D



In nutshell:

- Include variables in the model that are hypothesised to explain the observed network.

- The ERGM will provide information relative to the statistical significance of the included variable.

- Choose a set of configurations of theoretical interest.
- Estimate the parameters by applying ERGM
- Do inferences about the configurations – the network patterns – in the data.
- Do inferences about the type of social processes that are important in creating and sustaining the network

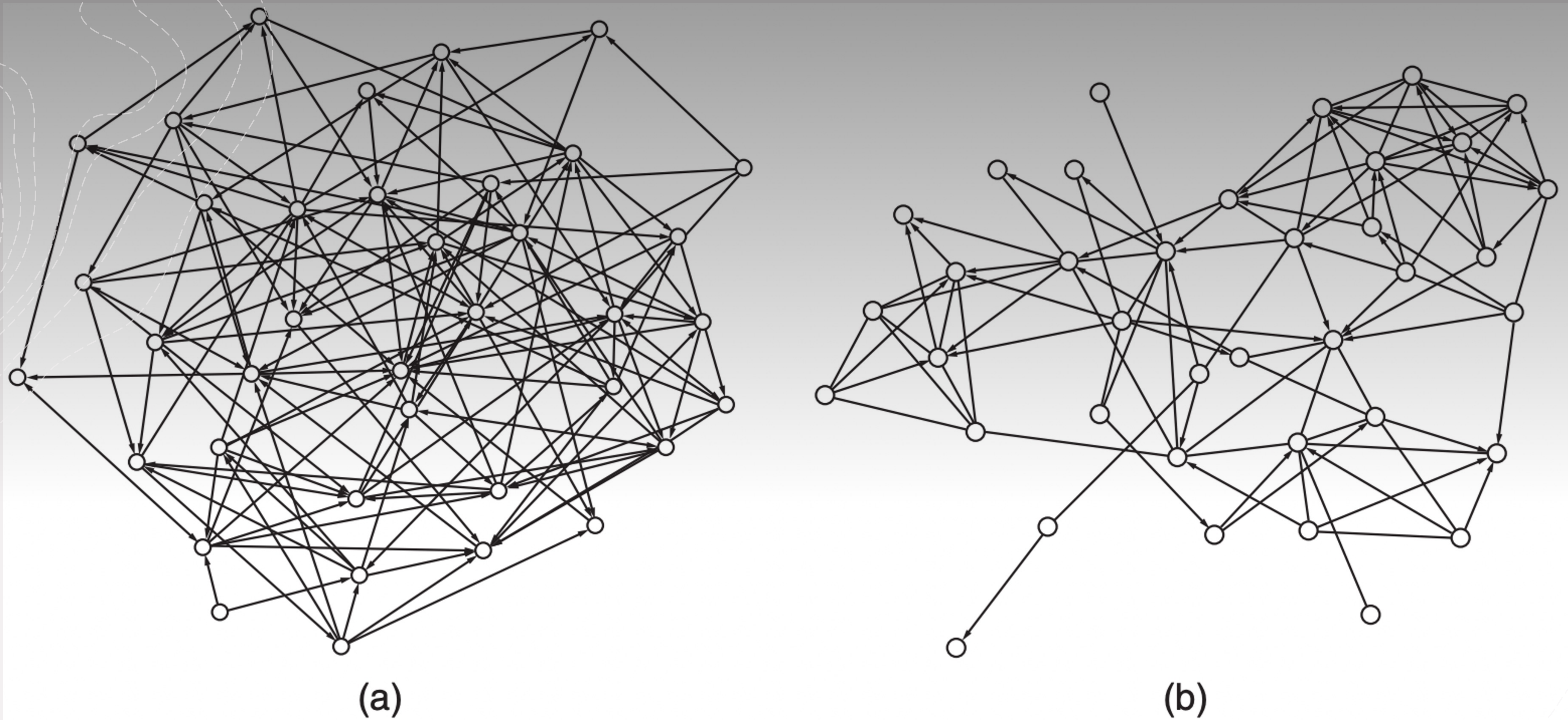


Figure 4.1. (a) Simple random network and (b) empirical communication network.

Lusher et al. (2013)

Table 4.1. *Selected network statistics for networks in Figure 4.1*

	Random network	Communication network
Actors	38	38
Arcs	146	146
Reciprocated arcs	6	44
Transitive triads	53	212
In-2-stars	292	313
Out-2-stars	254	283

Examples 1

- **The presence of triangles** : There is a process that generates a significant number of **triangles** that **is not** the result of **random link creation** e.g., a tendency to create a link between common friends.

Examples 2

- As function of individual covariates, e.g. **Are girls more popular than boys?**

- As function of network structures, e.g. If Adam is friends with Bill, and Bill is friends with Carl, what can we say about the chances of Adam and Carl being friends?

Stochastic Actor Oriented Models (SAOMs)

(Snijders, 1996; Snijders et al., 2010)
R package: RSiena

SAOMs

- Models for network **dynamics** and network **panel data**

Network dynamic through **simulations**.

Network panel data are common for representing relations like **friendship**, **advice**, **collaboration**, exchange which can be regarded as *states* rather than *events*.

Application of SAOMs

a wide variety of domains:

- Study of selection patterns in school classrooms ,
- The evolution of communication networks in high-risk social-ecological systems,
- The role of teen drinking behaviour in friendship selection,
- ...

SAOMs

- developed for the analysis of **longitudinal social network** data, collected by taking two or more "snapshots" ("**panels**" or "**waves**") of a network as it evolves over time.

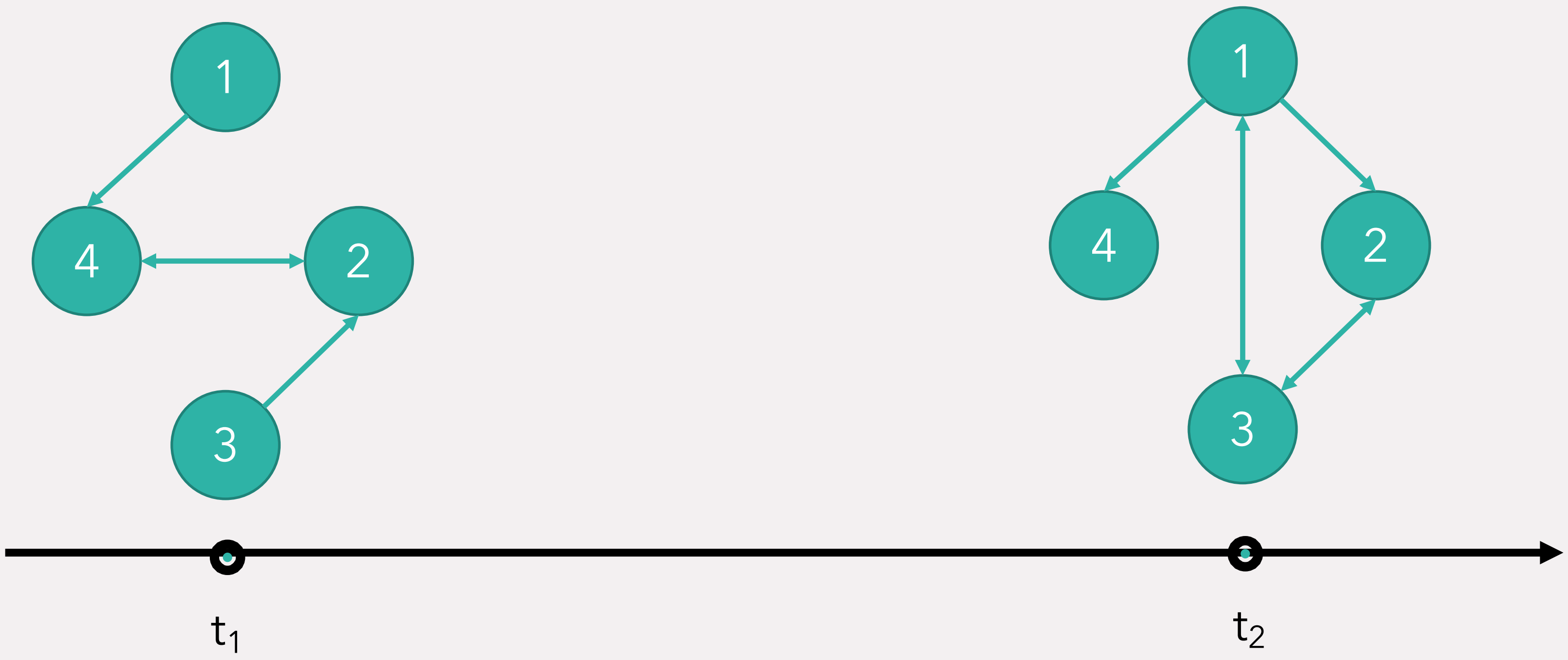
- **agent-based ('actor-oriented')** : They model changes from the perspective of the actors (**creating, maintaining or terminating ties** to other actors (a series of "**choices**")) within a (potentially) changing network.

SAOM as a model of the network evolution

- All network changes are decomposed into very small steps, so-called **ministeps**, in which one actor creates or terminates one outgoing tie.
- These **ministeps** are **probabilistic** and made sequentially.
- The **transition** from the observation at **one wave** to the **next** is done by means of normally a large number of ministeps. These **changes** are not individually observed, but they are **simulated.**
- This simulation model implements the statistical model for the **network dynamics**.

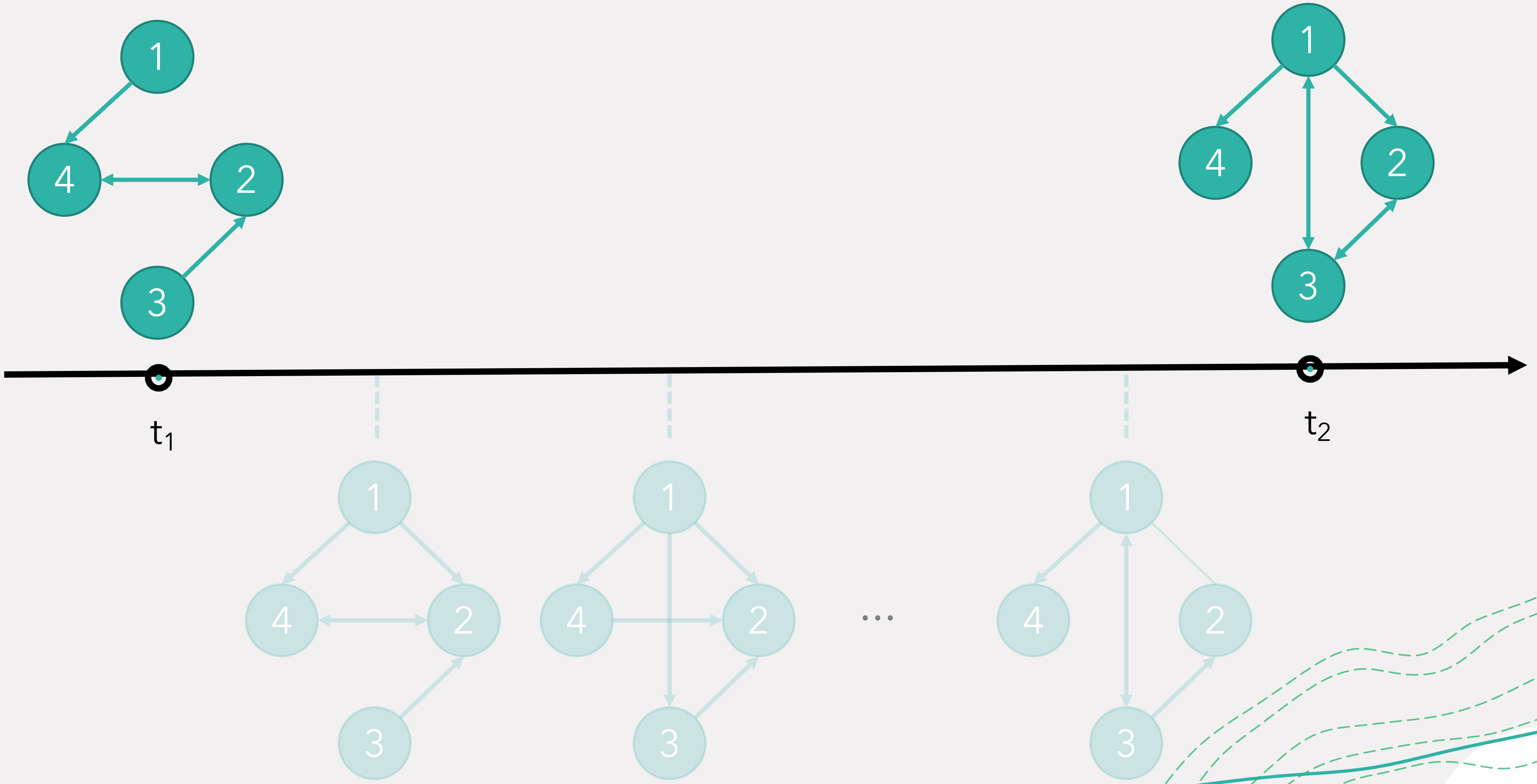
SAOMs

Model assumptions: consequences



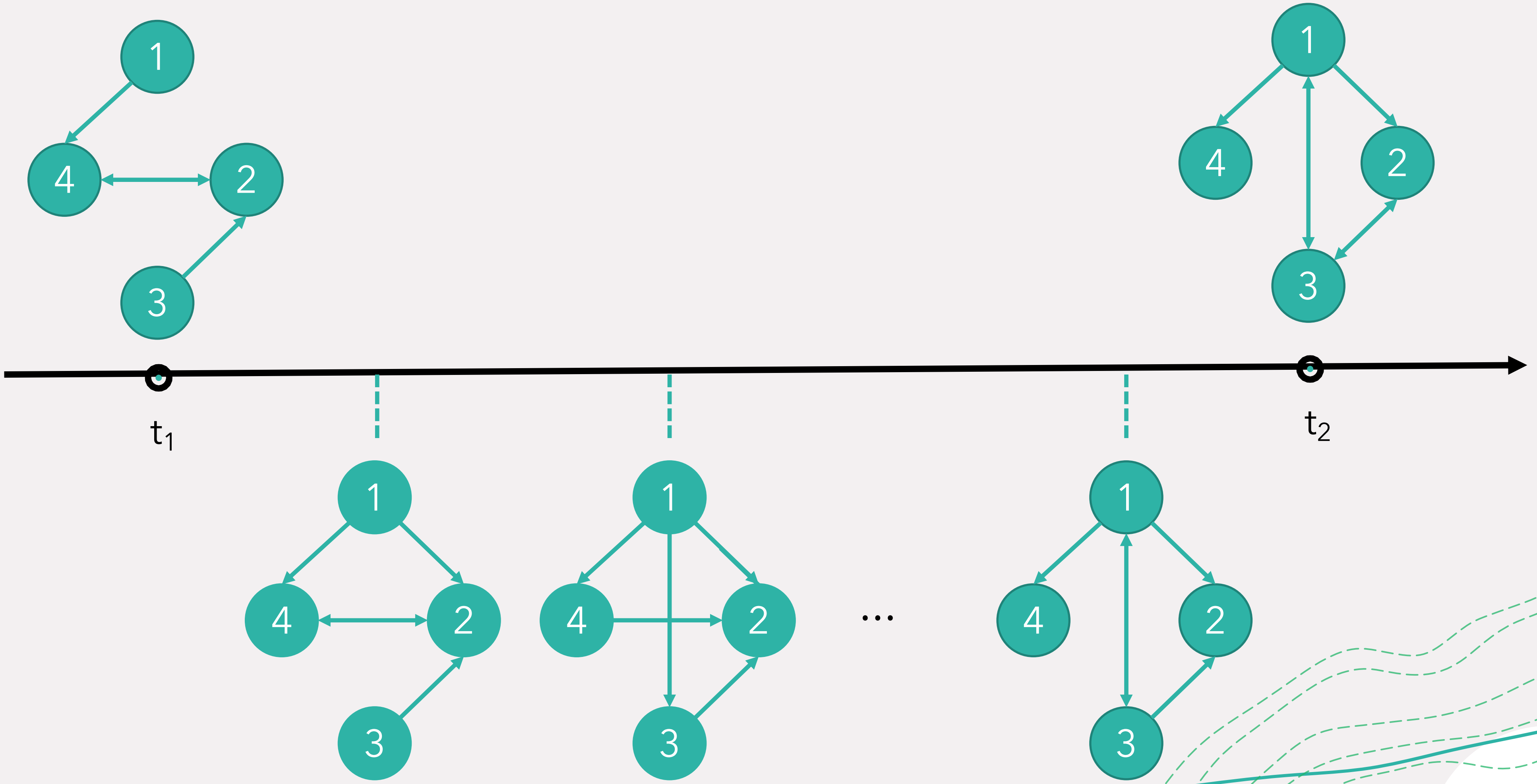
SAOMs

Model assumptions : illustration



SAOMs

Model assumptions: illustration



Modeling tie changes

Who gets the opportunity for a tie change and when?

A **person** from the network is chosen to make a change according to **the rate function**.
For actor i, the waiting time until the next opportunity for change is exponentially distributed with rate parameter

$$\boldsymbol{\lambda_i(x, v) = \exp(\textstyle\sum_k \alpha_k\, r_{i,k}(\mathrm{x}, \mathrm{v}))}$$

To whom?

Next, we model **which tie change** is made. This is modelled in the objective function:

$$\boldsymbol{f(\beta, x, v, w) = exp(\textstyle\sum_k \beta_k\, s_{i,k}(\mathrm{x}, \mathrm{v}, \mathrm{w})}$$

At each time step, the actors move in a direction that **maximizes their particular objective function**

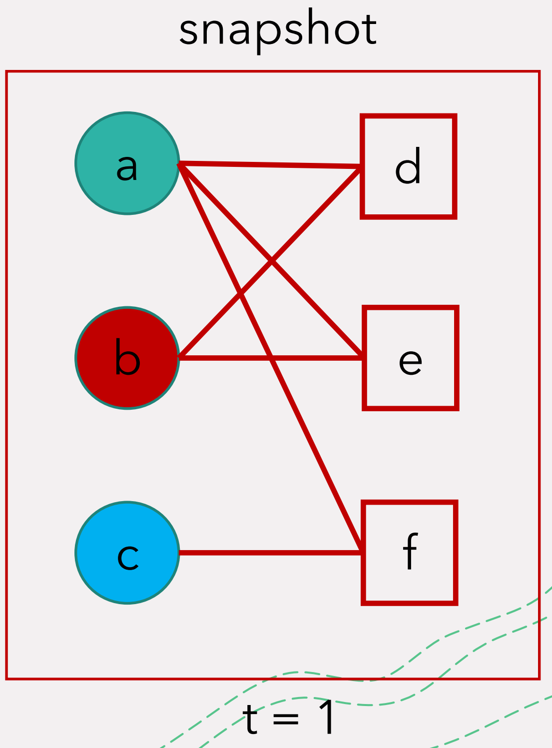
REM, ERGMs,TERGMs and SAOMs

- For **ERGMs, TERGMs or SAOMs** the issue of **dependence between observations** is critical (Kalish, 2020; Lusher et al., 2013).

- In REM, each event is considered to be **conditionally independent** of all other events in the sequence.
- REM assumes temporal dependence.

Longitudinal network analysis: ERGMs – SOAM, TERGMs - REMs

- If you have 1 snapshot of your network → run an ERGM
- ERGM = exponential random graph model



Research question

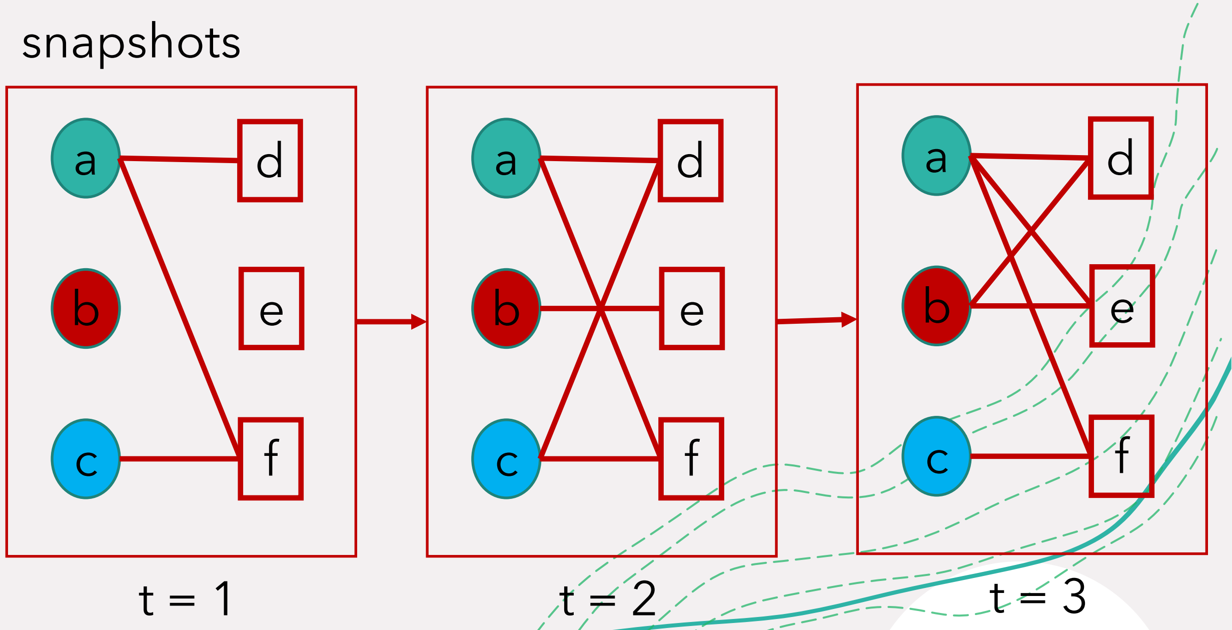
Which factors affect the structure of the network?

Longitudinal network analysis: ERGMs – SOAM, TERGMs - REMs

- If you have multiple snapshots of your network → run an TERGM or SAOM
- tERGM = temporal exponential random graph model

Research question

Which factors affect the structure of the networks and how do networks change over time?

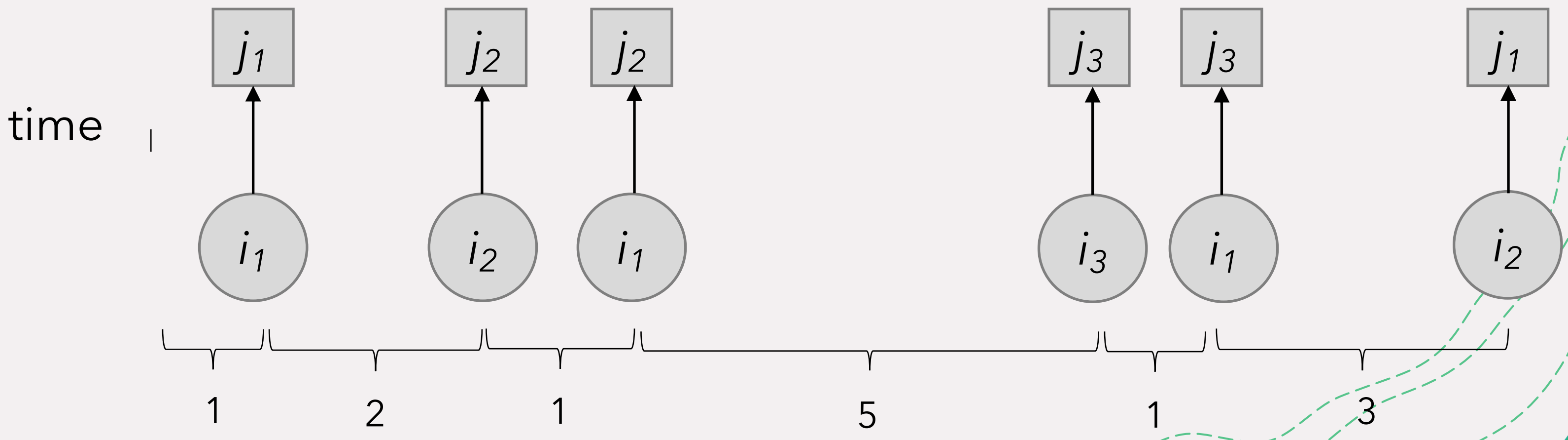


Longitudinal network analysis: ERGMs – SOAM, TERGMs - REMs

+ If you know the **time/order** each tie is created in a network → run a REM
+ … recorded in exact time or ordered

Research question

Which factors affect the probability of an edge forming at time point *t*?



References

+ Cranmer, S., Desmarais, B., & Morgan, J. (2020). *Inferential Network Analysis* (Analytical Methods for Social Research). Cambridge: Cambridge University Press. https:// 10.1017/9781316662915.
+ Butts, C. T. (2008). 4. A Relational Event Framework for Social Action. Sociological Methodology, 38(1), 155–200. https://doi.org/10.1111/j.1467-9531.2008.00203.x
+ Meijerink-Bosman, M., Back, M., Geukes, K., Leenders, R., & Mulder, J. (2022). Discovering trends of social interaction behavior over time: An introduction to relational event modeling. Behavior Research Methods. https://doi.org/10.3758/s13428-022-01821-8
+ Mulder, J., & Leenders, R. T. (2019). Modeling the evolution of interaction behavior in social networks: A dynamic relational event approach for real-time analysis. Chaos, Solitons & Fractals, 119, 73–85. https://doi.org/10.1016/j.chaos.2018.11.027
+ Lusher, D., Koskinen, J. and Robins, G. (2012), Exponential random graph models for social networks: Theory, methods, and applications, Cambridge University Press.
+ Snijders, T.A.B. (2011) Statistical models for social networks. Annual Review of Sociology, 37, 131–153.
+ Snijders, T.A.B., van de Bunt, G.G. & Steglich, C.E.G. (2010) Introduction to stochastic actor-based models for network dynamics. Social Networks, 32, 44–60
+ Snijders, T. A. B. 2017. "Stochastic actor-oriented models for network dynamics." Annual Review of Statistics and Its Application 4:343–363