Relational event models for the analysis of social networks

relational event modeling with `eventnet`

Alessandro Lomi   Jürgen Lerner

University of Exeter   University of Konstanz

University of Exeter, Business School   3–4 June 2019

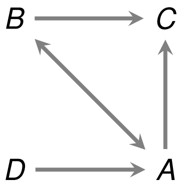
https://github.com/juergenlerner/eventnet

Discuss models for networks of relational events.

Time-stamped interaction among social actors

A	calls	B	at	t_1
B	calls	C	at	t_2
A	calls	C	at	t_3
B	calls	A	at	t_4
A	calls	C	at	t_5
D	calls	A	at	t_6

...



Examples: email, texting, telephone calls, online collaboration, social bookmarking, online social networks, product purchase or rating, automatically monitoring interaction, discourse networks, coding of documents/speech/video, ...

Dependencies in event network data.

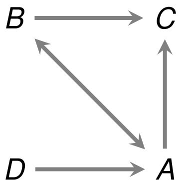
time dependence

A	calls	B	at	t_1
B	calls	C	at	t_2
A	calls	C	at	t_3
B	calls	A	at	t_4
A	calls	C	at	t_5
D	calls	A	at	t_6

$\ldots$

Events depend on **previous** events.

network dependence



Events depend on events that happened on **other dyads**.

Background literature.

small selection of papers on relational event models

General framework for relational event models.

- Butts (2008). **A Relational Event Framework for Social Action.** *Sociological Methodology*, 38(1), 155-200.

Conditional event-type models.

- Lerner, Bussmann, Snijders, and Brandes. (2013). **Modeling frequency and type of interaction in event networks.** *Corvinus Journal of Sociology and Social Policy*, 4(1), 3-32.

Actor-oriented models.

- Stadtfeld and Block (2017). **Interactions, actors, and time: Dynamic network actor models for relational events.** *Sociological Science*, 4, 318-352.

Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

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

Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

We observe **interaction** among social actors.

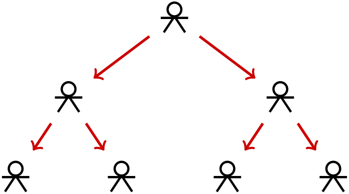
Can we understand or explain this?

Does this interaction have any **structure**?

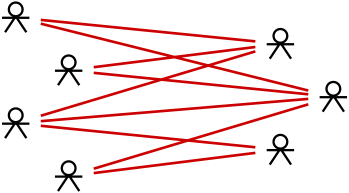
Does this interaction have any structure?

exemplarily, consider two competing hypotheses

pecking order



“us vs. them”



Social interaction and social structure.

Social interaction **shapes** social structure.

social interaction		**social structure**
(relational events)		(relational states)
friendly/hostile		sentiments
communication	*mutual* ⟷ *influence*	friendship
collaboration		opposition
fighting		dominance
...		like/dislike
		...

Social interaction **results from** social structure.

Testing hypotheses with relational event models.

pecking-order hypothesis (events induce/reveal dominance hierarchy)

past events

A hits *B* at time t_1
B hits *C* at time t_2
...

all events
before t

network of past events

A hit *B*
B hit *C*

hypotheses,
network effects

social structure

A dominates *B*
B dominates *C*
A dominates *C*

specifies

probability of next event(s)

A hits *C* at time t

has ***increased probability***

Testing hypotheses with relational event models.

us-vs.-them hypothesis (events induce/reveal polarization)

past events

A hits B at time t_1

B hits C at time t_2

$\ldots$

all events before t

network of past events

$A \xrightarrow{\text{hit}} B \xrightarrow{\text{hit}} C$

hypotheses, network effects

social structure

A — opponents — B

B — opponents — C

A - - allies - - C

specifies

probability of next event(s)

A hits C at time t

has ***decreased probability***

Social interaction and social structure.

Social interaction is ***assumed* to shape** social structure.

social interaction		**social structure**
(relational events)		(relational states)
friendly/hostile		sentiments
communication	*mutual influence*	friendship
collaboration		opposition
fighting		dominance
...		like/dislike
		...

Social interaction is ***assumed* to result from** social structure.

Outline.

- Basics of relational event modeling.
  - Testing hypotheses with event network data.
  - Underlying concepts.
  - Illustration: relational event modeling with eventnet.
- Variations in different directions.
  - Event types and weights; conditional event-type models.
  - Covariates of nodes, dyads, and networks.
  - One-mode, two-mode, multi-mode networks.
  - Analyzing large event networks (sampling).
  - Actor-oriented models.
  - Time-varying risk set.
- Case study: interaction events in international relations.

Relational event models (overview).

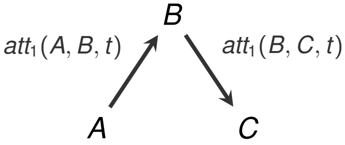
past events

$$e_1 = (A, B, t_1, x_1, w_1)$$
$$e_2 = (B, C, t_2, x_2, w_2)$$
$$\ldots$$

all events before t

attributes

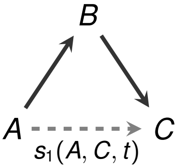
(recording past interaction)



network effects

statistics

(embedding of dyads)



specify

probability of next event(s)

how likely is it to observe

$$e = (A, C, t, x, w)$$

given the statistics $s_1(A, C, t), \ldots$

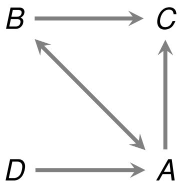
Relational events / input data.

Event network data.

Time-stamped interaction among social actors

A	calls	B	at	t_1
B	calls	C	at	t_2
A	calls	C	at	t_3
B	calls	A	at	t_4
A	calls	C	at	t_5
D	calls	A	at	t_6

...



Examples: email, texting, telephone calls, online collaboration, social bookmarking, online social networks, product purchase or rating, automatically monitoring interaction, discourse networks, coding of documents/speech/video, ...

Format of relational events.

Input data: time-ordered sequence of dyadic events

$$e_1 = (u_1, v_1, t_1, x_1, w_1)$$
$$e_2 = (u_2, v_2, t_2, x_2, w_2)$$
$$\ldots$$

- u **source** (initiator, sender) of the event;
- v **target** (addressee, receiver) of the event;
- t **time**, when the event happened;
- x **type** of the event (categorical information);
- w **weight** of the event (numerical information).

Who does when what to whom?

Not all of these components have to be given explicitly

(there may be no variation in, say, types or weights).

Relational event models (overview).

past events

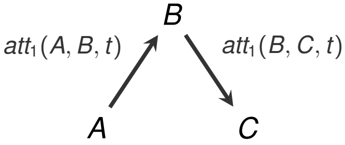
$$e_1 = (A, B, t_1, x_1, w_1)$$
$$e_2 = (B, C, t_2, x_2, w_2)$$
$$\dots$$

all events

before t

attributes

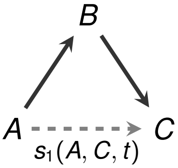
(recording past interaction)



network effects

statistics

(embedding of dyads)



specify

probability of next event(s)

how likely is it to observe

$$e = (A, C, t, x, w)$$

given the statistics $s_1(A, C, t), \dots$

Event network attributes.

Network attributes: recording past events.

We assume that actors ***remember past interaction***
(which may influence future interaction).

Network **attributes** record essential aspects of past events.

- Attributes are defined on dyads, nodes, or networks.
- Values may decay over time.

Network attributes: recording past events.

Attribute values at time t

- depend on past events, happening **before** t
  (depend potentially on the type and/or weight of past events)
- may decay over time with a given halflife.

Dyad-level attributes assign values to dyads (pairs of nodes).

- E. g., number of hostile events from A to B before t.

Node-level attributes assign values to nodes.

- E. g., number of hostile events initiated by A before t.

Network-level attributes assign values to the entire network.

- E. g., number of hostile events in the network before t.

Relational event models (overview).

past events

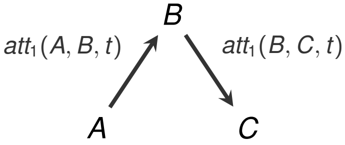
$$e_1 = (A, B, t_1, x_1, w_1)$$
$$e_2 = (B, C, t_2, x_2, w_2)$$
$$\ldots$$

all events

before t

attributes

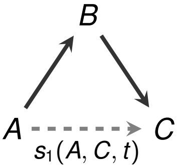
(recording past interaction)



network effects

statistics

(embedding of dyads)



specify

probability of next event(s)

how likely is it to observe

$$e = (A, C, t, x, w)$$

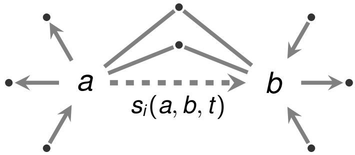
given the statistics $s_1(A, C, t), \ldots$

Event network statistics.

Event network statistics $s_i(a, b, t)$.

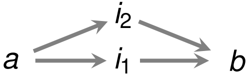
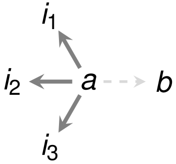
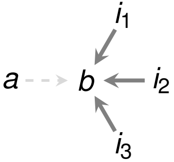
Statistics $s_i(a, b, t)$ assign time-varying values to dyads;

- are the variables which explain events on (a, b) at t;
- are functions of attributes $att_j(a', b', t)$, i. e., of past events happening before t on the same or other dyads;
- introduce dependence among dyadic observations.

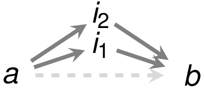
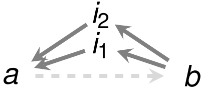
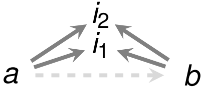


Examples: repetition, reciprocation, (in-/out-/mixed-)degrees, triadic effects, four-cycle effects, covariate effects, . . .

Examples: common statistics $s_i(a, b, t)$.

statistic	$s_i(a, b, t) =$	$a \dashrightarrow b$ depends on
`repetition`	$att(a, b, t)$	$a \longrightarrow b$
`reciproc.`	$att(b, a, t)$	$a \longleftarrow b$
`transitivity`	$\sum_i att(a, i, t) \cdot att(i, b, t)$	
`outDegSource`	$\sum_i att(a, i, t)$	
`inDegTarget`	$\sum_i att(i, b, t)$	

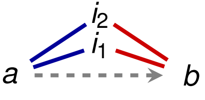
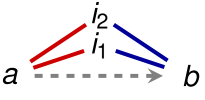
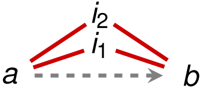
Examples: varying dyad direction.

statistic	$s_i(a, b, t) =$	$a \dashrightarrow b$ depends on
`transitivity`	$\sum_i att(a, i, t) \cdot att(i, b, t)$	
`cyclicality`	$\sum_i att(i, a, t) \cdot att(b, i, t)$	
`shared receiv.`	$\sum_i att(a, i, t) \cdot att(b, i, t)$	

Possible: **symmetrized** tie weights: $att(u, v, t) + att(v, u, t)$.

Examples: combining different types of attributes.

Given: attributes indicating friends and enemies.

statistic	$s_i(a,b,t) =$	$a \dashrightarrow b$ depends on
`enemy.of.friend`	$\sum_i F(a,i,t) \cdot E(b,i,t)$	
`friend.of.enemy`	$\sum_i E(a,i,t) \cdot F(b,i,t)$	
`enemy.of.enemy`	$\sum_i E(a,i,t) \cdot E(b,i,t)$	

Might be combined with variation in the dyad direction.

Relational event models (overview).

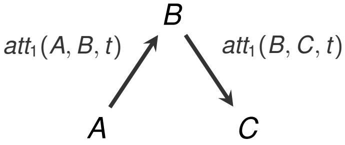
past events

$$e_1 = (A, B, t_1, x_1, w_1)$$
$$e_2 = (B, C, t_2, x_2, w_2)$$
$$\dots$$

all events before t →

attributes

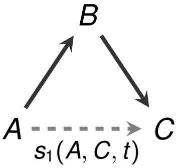
(recording past interaction)



↓ network effects

statistics

(embedding of dyads)



← specify

probability of next event(s)

how likely is it to observe

$$e = (A, C, t, x, w)$$

given the statistics $s_1(A, C, t), \dots$

Observations (events and non-events).

What do we observe (when observing events)?

A	**calls**	**B**	**at**	$\mathbf{t_1}$
B	calls	C	at	t_2
A	calls	C	at	t_3
B	calls	A	at	t_4
A	calls	C	at	t_5
D	calls	A	at	t_6

...

We observe **events**, e. g.,
A calls B at time t_1.

We also observe **non-events**:
A *does not call* C at t_1.
A *does not call* D at t_1.
B *does not call* A at t_1.
B *does not call* C at t_1.
...

Model estimation typically (also) uses information about non-events.

Dyads that ***could experience*** an event make up the **risk set** (can vary over time).

The definition of the risk set depends on the research question.

Development of the risk set over time.

The risk set might be **exogenously** defined:

- **Nodes** may enter, leave, re-enter, ... the risk set
  (risk set contains all pairs of nodes at risk).
- **Dyads** may enter, leave, re-enter, ... the risk set.

The risk set might **endogenously** result from observed events.

For instance, modeling start and termination of conversations.

- A conversation can only terminate when it is ongoing.
- A start-event puts the dyad at risk to experience a termination-event.
- A termination-event removes the dyad from the risk set.

The risk set (also) depends on the research question.

Explaining dyadic event rates.

- Risk set: all dyads that could experience an event.
- *Is an event on* (A, B) *more likely than on* (C, D)*, etc?*

Explaining actors' choices to select recipients of events.

- Given an observed event on (A, B),
- risk set: dyads (A, B'), for all nodes B'.
- *Given that A does initiate an event, is A more likely to direct the event at B rather than C, etc?*

Explaining preferences for specific *types* of events (e. g., cooperative vs. hostile).

- Given an observed event on (A, B),
- risk set: only the dyad (A, B); model the event type.
- *Given that there is an event on* (A, B)*, is it more likely to be a cooperative or a hostile event?*

Sampling from the risk set.

Typically the number of non-events is much larger than the number of events.

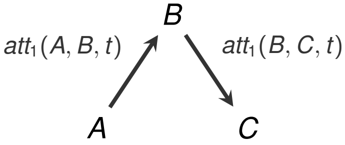
For instance, if a million nodes are at risk, there are one trillion non-events per event.

It is typically sufficient to sample few of the non-events for each event (**case-control sampling**).

Relational event models (overview).

attributes

(recording past interaction)



past events

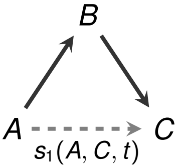
$$e_1 = (A, B, t_1, x_1, w_1)$$
$$e_2 = (B, C, t_2, x_2, w_2)$$
$$\ldots$$

all events before t

network effects

statistics

(embedding of dyads)



specify

probability of next event(s)

how likely is it to observe

$$e = (A, C, t, x, w)$$

given the statistics $s_1(A, C, t), \ldots$

Probability distribution of future events
(specifically: Cox proportional-hazard model).

Cox proportional-hazard model (CoxPH).

A CoxPH model does not specify the time to the next event.

A CoxPH model specifies relative event rates.

- *How likely is it that the next event happens on dyad* (A, B)*, rather than on any other dyad in the risk set?*
- *By which factor does the event rate increase/decrease if an explanatory variable changes its value by a given amount?*

The relative event rate λ is a parametric function of statistics

$$\lambda(a, b, t) = \exp\left(\theta_1 \cdot s_1(a, b, t) + \theta_2 \cdot s_2(a, b, t) + \dots\right)$$

Estimated parameters θ_i yield hypothesis tests,
e. g., *is there transitive closure in dyadic events?* ...

Estimation of Cox proportional-hazard model (CoxPH).

The relative event rate λ is a parametric function of statistics

$$\lambda(a,b,t) = \exp\left(\theta_1 \cdot s_1(a,b,t) + \theta_2 \cdot s_2(a,b,t) + \dots\right)$$

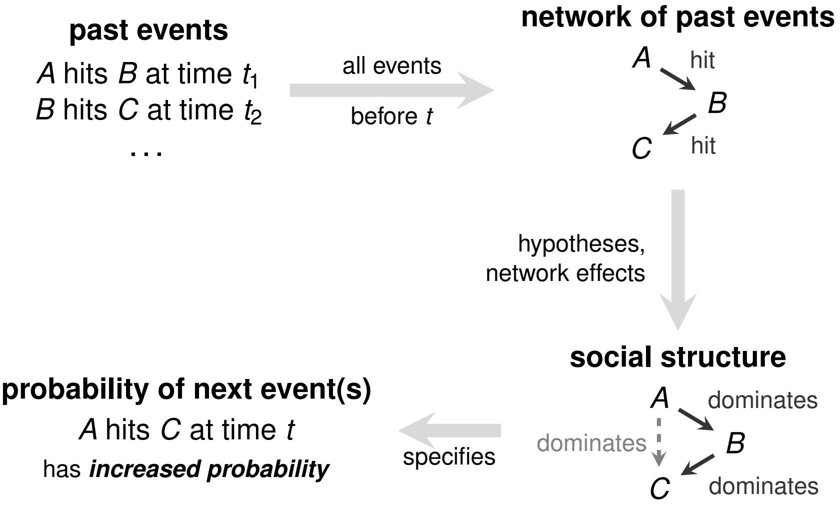
Parameters θ_i can be estimated with the function `coxph` in the R-package **survival**.

Needs table of statistics of all events and non-events.

Estimated parameters θ_i yield hypothesis tests, e. g., *is there transitive closure in dyadic events?* . . .

Recall: Testing hypotheses with REM.

pecking-order hypothesis (events induce/reveal dominance hierarchy)



Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

Workflow of eventnet.

https://github.com/juergenlerner/eventnet

Computes specified statistics 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

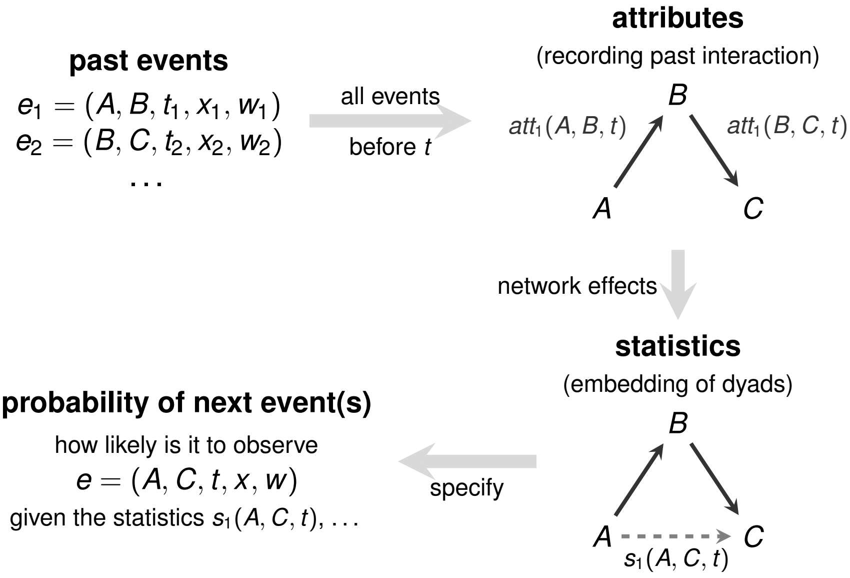
statistics of events and non-events

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).

Eventnet configurations specify the details.



Eventnet configurations (six main parts).

correspond to six tabs in the eventnet GUI

(files) specify input event files (or directories) in CSV format.

(events) specify which column holds which event component (source, target, . . .) and whether networks are one-mode or multi-mode.

(time) specify how time is encoded and interpreted.

(attributes) specify network attributes recording past events.

(statistics) specify statistics explaining future events.

(observations) specify risk sets and sampling (if applies).

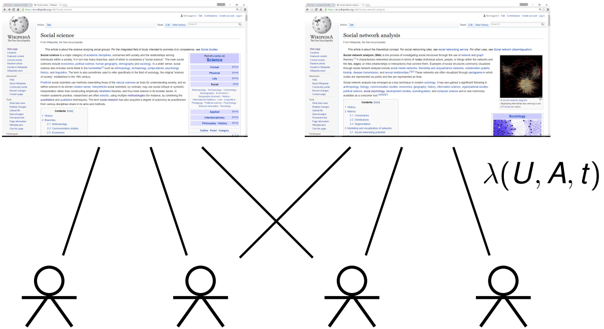
Example study: collaborative editing in Wikipedia.

Activity and attention in Wikipedia.

two-mode network of relational events

Contributing users edit Wikpedia articles at given points in time.

$$(u_1, a_1, t_1), (u_2, a_2, t_2), (u_3, a_3, t_3), \ldots$$



We model time-varying dyadic event rates $\lambda(U, A, t)$

(relative rate of edit events of user U on article A at time t).

Background: relevance of the Wikipedia network.

Wikipedia is one of the most popular sources of information;

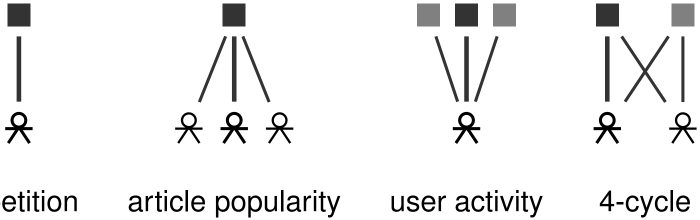
production community of volunteers writing an encyclopedia;

has no pre-defined organizational structure and no direct monetary rewards.

Claim: organizational structure and allocation of work results from emergent networks of task-oriented interaction.

Hypothetical network effects in this data.

Event rate on user-article pair depends on



Additionally: assortativity (interact popularity and activity).

Sub-network used in this course.

https://github.com/juergenlerner/eventnet/wiki/Basic-tutorial

All articles in `Category:Human migration` or in a sub-(sub-)category of it.

All registered users performing an edit to any of these articles.

$\geq$ 4,000 articles; 87,000 users; 816,000 relational events

Here: analyze the first 1,000 events (200 users, 155 articles).

Lerner and Lomi (2019): **Let's talk about refugees: Network effects drive contributor attention to Wikipedia articles about migration-related topics.** In: *Proc. Complex Networks 2018*.

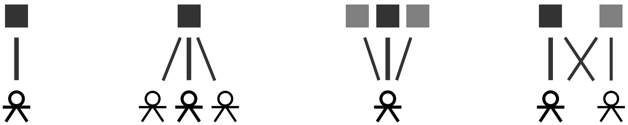
demo

Results (estimated parameters).

Cox proportional hazard model

$$\lambda(u,a,t) = \exp\left(\theta_1 \cdot s_1(u,a,t) + \theta_2 \cdot s_2(u,a,t) + \dots\right)$$

$s(u,a,t)$	parameter θ	$\exp(\theta)$
repetition	0.782 (0.034)*	2.185
article_popularity	0.010 (0.008)	1.010
user_activity	0.039 (0.004)*	1.040
four_cycle	0.049 (0.012)*	1.050
activity:popularity	−0.002 (0.001)*	0.998



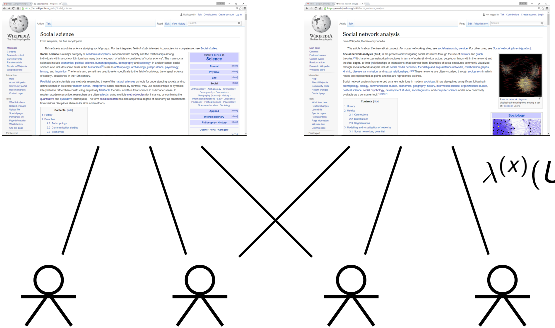
Exercise: editing and discussion in Wikipedia.

Exercise: editing and discussion in Wikipedia.

two-mode network of two types of relational events

Contributing users **edit** or **discuss** Wikpedia articles.

$$(u_1, a_1, t_1, x_1), (u_2, a_2, t_2, x_2), (u_3, a_3, t_3, x_3), \ldots$$



$\lambda^{(x)}(U, A, t)$

Question: does editing induce discussion; or does discussion induce editing?

Exercise: editing and discussion in Wikipedia.

https://github.com/juergenlerner/eventnet/wiki/Basic-tutorial

Data: all articles in `Category:Human migration` or in a sub-(sub-)category of it.

All registered users editing or discussing any of these articles.

$\geq$ 4,000 articles; 87,000 users; 950,000 relational events.

Question: does editing induce discussion; or does discussion induce editing?

Results (estimated parameters).

Cox proportional hazard models for edit events and talk events

	edit model	talk model
edit.repetition	5.018 (0.018)*	6.075 (0.075)*
talk.repetition	1.096 (0.020)*	5.803 (0.120)*
edit.popularity	0.924 (0.005)*	0.590 (0.022)*
edit.activity	0.967 (0.004)*	-0.205 (0.017)*
talk.popularity	0.027 (0.004)*	0.465 (0.021)*
talk.activity	-0.238 (0.004)*	1.328 (0.019)*
edit.4.cycle	0.034 (0.004)*	-0.370 (0.018)*
talk.4.cycle	0.343 (0.005)*	1.120 (0.024)*
edit.assortativity	-0.258 (0.003)*	-0.026 (0.012)
talk.assortativity	-0.088 (0.003)*	-0.655 (0.018)*
Num. obs.	4,892,946	828,101
Num. events	815,722	138,036

* $p < 0.001$

Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

Modeling typed or weighted relational events.

See details on models and data in Section
Case study: interaction events in international relations.

Data: events $e = (u, v, t, w)$ encoding

- u **source** actor (initiator);
- v **target** actor (recipient);
- t **timestamp**, the day on which e happened;
- w **weight** from -10 (most hostile) to $+10$ (most cooperative).

Decomposing the probability of typed events.

$$P(\text{negative interaction}) = P(\text{interaction}) \times P(\text{negative} \mid \text{interaction})$$

Models can be specified for

- the rate of positive events;
- the rate of negative events;
- the rate of all events (positive or negative);
- the conditional probability that a given event is positive;
- the conditional distribution of event weights;

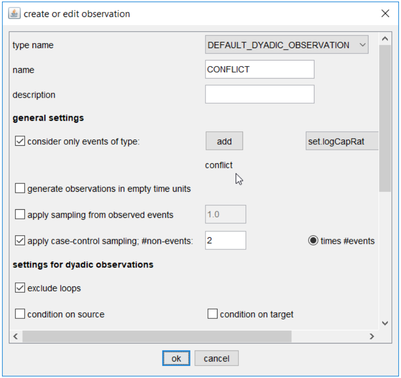
leading to different tests for, e. g., balance theory.

Lerner (2016). **Structural balance in signed networks: Separating the probability to interact from the tendency to fight**. *Social Networks*.

Specifying models for typed events in `eventnet`.

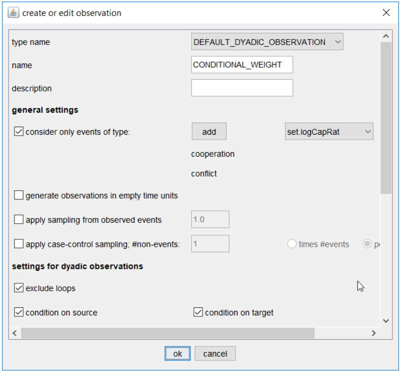
varying the settings for observation generators

rate of conflictive events



events: only conflictive
controls: entire risk set

conditional event type



events: coop. or confl.
controls: none

Outline.

Covariates of nodes, dyads, and networks.

Covariates: exogenous information that may influence events.

- **(node covariates)** gender, age, behavior, ...
- **(dyad covariates)** geographic distance, kinship ties, ...
- **(network cov)** day of week, environmental conditions, ...

First way to use covariates in eventnet.

In many cases, covariates can be easily merged into the statistics table produced by eventnet before estimating models.

statistics of events and non-events

src	trg	time	obs	s_1	s_2	...
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	
...						

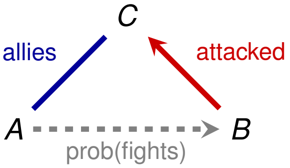
Applies whenever the triple (source, target, time) determines the value of the covariate.

Second way to use covariates in eventnet.

Sometimes covariates are assumed to impact events in a more complex way.

Example: model conflictive events; have formal alliances among actors given as a dyadic covariate.

Actors may be likely to fight those who attacked their allies.



Then, covariates have to be entered into eventnet (via dummy events) before computing statistics.

Entering covariates via dummy events.

A dummy event of type `set.allies` can set a dyad-level attribute recording which actors are allies.

Time	Source	Target	Weight	Type
790101	FRN	GER	1.0	set.allies
…				

Such events are (typically) not modeled but may explain the distribution of cooperative or conflictive events.

Outline.

Multi-mode networks: subsets of nodes.

The nodes may be partitioned into any number of subsets (modes); events of different types connect different sets.

(one-mode) such as person-person communication

- persons communicate with persons.

(two-mode) such as Wikipedia user-article network

- users edit articles;
- users discuss with users;
- articles link to articles.

(three-mode) such as actors-organizations-projects

- ...

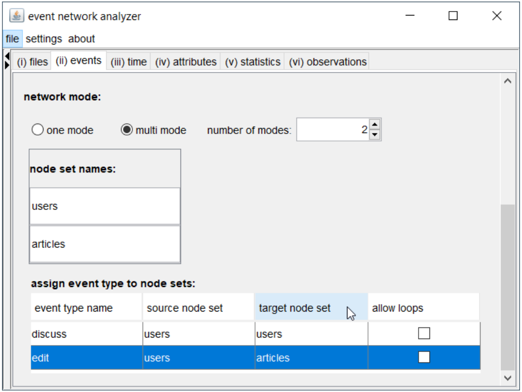
Multi-mode networks in eventnet.

Modes constrain the risk set for types of events.

Set number of modes.

Assign node set names.

Map event types to pairs of node sets.



Events may depend on events of different types, connecting different modes.

Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

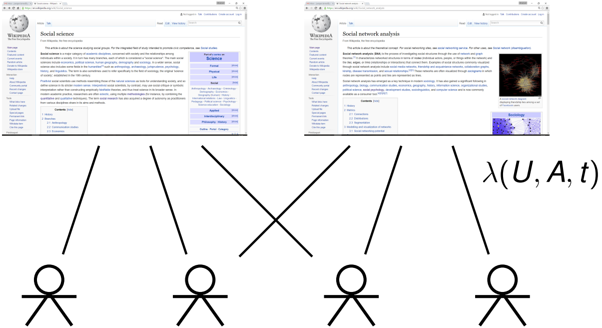
Example data: collaborative editing in Wikipedia.

Activity and attention in Wikipedia.

two-mode network of relational events

Contributing users edit Wikpedia articles at given points in time.

$$(u_1, a_1, t_1), (u_2, a_2, t_2), (u_3, a_3, t_3), \ldots$$

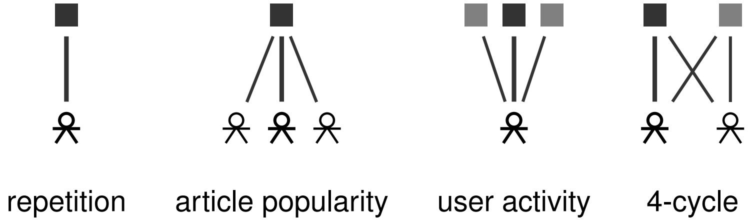


We model time-varying dyadic event rates $\lambda(U, A, t)$

(relative rate of edit events of user U on article A at time t).

Hypothetical network effects in this data.

Event rate on user-article pair depends on



Additionally: assortativity (interact popularity and activity).

Data: complete Wikipedia collaboration network.

https://github.com/juergenlerner/eventnet/wiki/Large-event-networks-(tutorial)

All edits from any registered user to any Wikipedia article from January 2001 to January 2018. More than

- 6 million users;
- 5 million articles;
- 360 million events.

Risk set at the end of the observation period contains more than 30 trillion dyads.

Perform two types of sampling

- case-control sampling of non-events;
- uniform sampling of observed events.

Case-control sampling from non-events.

Often applied in epidemiological studies of rare diseases.

> *"Loosely, if the disease of interest is rare, the contribution of the nonfailures, in terms of the statistical power of the study, will be negligible compared to that of the failures. Thus cohort sampling methods which include all the failures and a portion of the nonfailures are highly desirable."* [Borgan et al. (1995)]

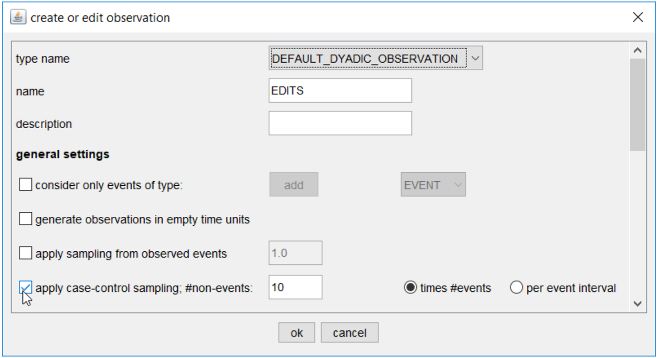
Sample only a fixed number of non-events per event.

Estimating CoxPH on this sampled data is consistent estimator.

Borgan, Goldstein, and Langholz (1995). **Methods for the analysis of sampled cohort data in the Cox proportional hazards model.** *The Annals of Statistics*, 23(5).

Case-control sampling in `eventnet`.

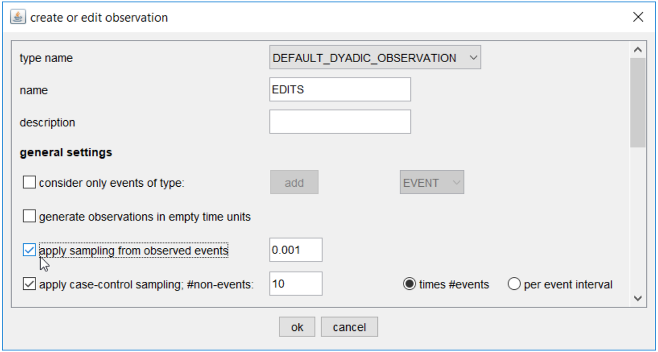
Observations specify sampling from non-events / number of controls.



Can use several observations in the same study with the same or different numbers of non-events to **assess variation** introduced by sampling.

Uniform sampling from events in `eventnet`.

Specify sampling from events (prob. to include in output).



Important: statistics are computed as a function of ***all*** previous events – not just the sampled previous events.

Reliability of parameter estimates under sampling

(variability introduced by sampling).

Reliability of parameter estimates under sampling.

Sample repeatedly from events and non-events.

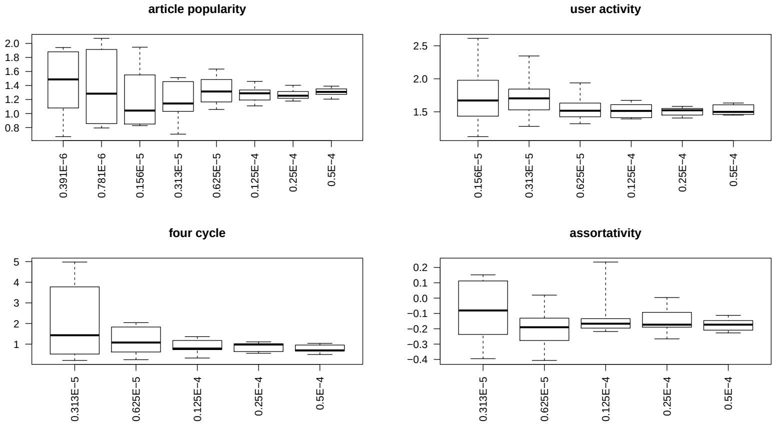
Let the probability to sample events p and the number of non-events per event m vary within useful ranges.

Assess variability of parameter estimates dependent on p and m.

- Which numbers of events and non-events are sufficient to get reliable estimates?
- Given a limited budget of observations, which is the optimal ratio of events to non-events?

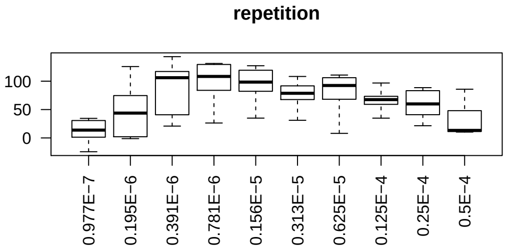
Reliability of parameter estimates (some results).

setting $m = 5$ and letting p increase; y-axis displays distribution of parameters



The repetition parameter behaves differently.

setting $m = 5$ and letting p increase; y-axis displays distribution of parameters



The repetition parameter behaves differently.

Compute density (proportion of dyads with non-zero values) for all observations, for events, and for non-events.

statistic	density.obs	density.events	density.non-ev
repetition	0.084385	0.506282	**0.000006**
popularity	0.780180	0.937911	0.748634
activity	0.378371	0.977290	0.258588
four.cycle	0.163034	0.766557	0.042329
assortativity	0.332671	0.917340	0.215737

Repetition is **extremely sparse among the non-events** (six in a million have non-zero value). This causes the high variability of parameter estimates.

Reliability of parameter estimates (lessons learned).

Given a limited budget of observations, it is preferable to **sample more events and fewer non-events**.

In our given data, **a few hundred events** (and five times as many non-events) gave very reliable estimates of the **degree effects** (activity and popularity);

some thousand events for four-cycle and assortativity.

Estimation of the repetition parameter was the most unreliable due to extreme sparsity **among the non-event dyads**.

In general, it is advisable to check distributions of statistics **separately for events and non-events**.

Lerner and Lomi (2019). **Reliability of relational event model estimates under sampling.** arXiv:1905.00630

Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

Actor-oriented models

So far, we modeled **dyadic** event rates $\lambda(A, B, t)$.

Actor-oriented models separate these into

- decisions of actors to become active (initiate events);
- conditional choice of event targets.

Stadtfeld and Block (2017). **Interactions, actors, and time: Dynamic network actor models for relational events.** *Sociological Science*, 4, 318-352.

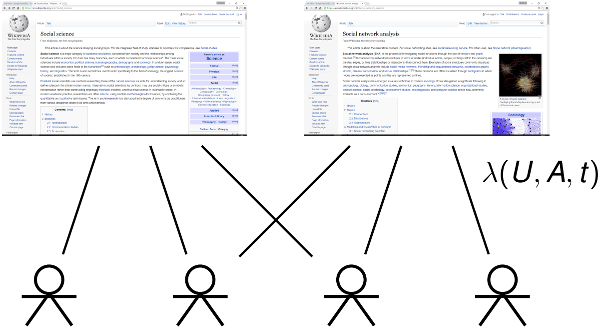
Example data: collaborative editing in Wikipedia.

Activity and attention in Wikipedia.

two-mode network of relational events

Contributing users edit Wikpedia articles at given points in time.

$$(u_1, a_1, t_1), (u_2, a_2, t_2), (u_3, a_3, t_3), \ldots$$



We model time-varying dyadic event rates $\lambda(U, A, t)$

(relative rate of edit events of user U on article A at time t).

Hypothetical network effects in this data.

Event rate on user-article pair depends on

repetition article popularity user activity 4-cycle

Additionally: assortativity (interact popularity and activity).

Sub-network used in this course.

https://github.com/juergenlerner/eventnet/wiki/Basic-tutorial

All articles in `Category:Human migration` or in a sub-(sub-)category of it.

All registered users performing an edit to any of these articles.

$\geq$ 4,000 articles; 87,000 users; 950,000 relational events

Here: analyze the first 1,000 events (200 users, 155 articles).

Lerner and Lomi (2019): **Let's talk about refugees: Network effects drive contributor attention to Wikipedia articles about migration-related topics.** In: *Proc. Complex Networks 2018*.

Separating user activity from choice of article to edit.

So far we modeled the relative **dyadic** event rates $\lambda(U, A, t)$.

Is the event rate on (U_1, A_1) *higher or lower than on* (U_2, A_2)*?*

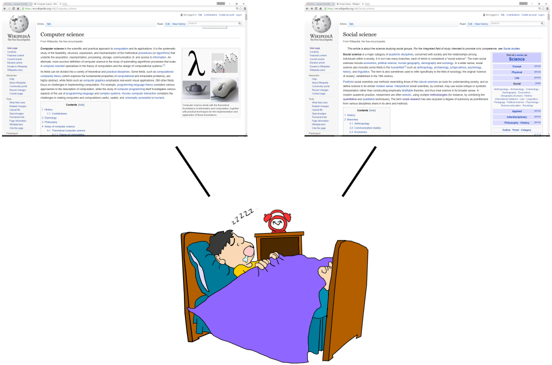
This potentially mixes up two separate processes:

- Who is the next user that becomes active (edits any article)?
- Given that U is the next user that becomes active, what is the probability that U chooses to contribute to article A?

Separating user activity from choice of article to edit.

So far we modeled the relative **dyadic** event rates $\lambda(U, A, t)$.

Is the event rate on (U_1, A_1) *higher or lower than on* (U_2, A_2)*?*



Realistically, actors do not constantly think about initiating events.

Separating user activity from choice of article to edit.

Modeling the dyadic event rate might mix up processes that

- drive user **activity**
- determine the choice of articles to edit (**popularity**).

Actor-oriented models keep these separate by design.

Stadtfeld and Block (2017). **Interactions, actors, and time: Dynamic network actor models for relational events.** *Sociological Science*, 4, 318-352.

Separating activity and conditional choice of target.

Rather than modeling the dyadic event rate $\lambda(U, A, t)$,

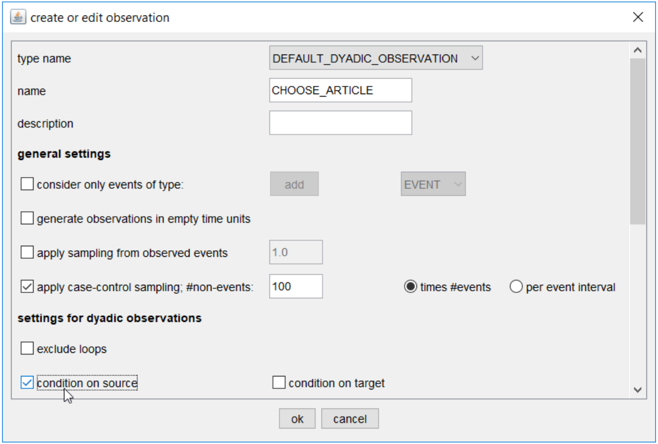
we model the node-level rate $\lambda(U, \cdot, t)$ to initiate events

and the conditional dyadic rate $\lambda(U, A, t|U)$, given the source,

independently of each other.

Conditional choice of target in `eventnet`.

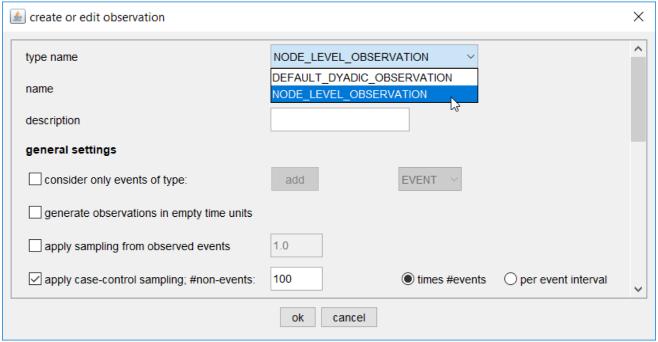
We can condition the risk set on the observed source (or target, or both) of events.



$\Rightarrow$ compare relative event rates only for dyads with the same observed source node.

Modeling node-level events in `eventnet`.

Node-level observations can model the activity (irrespective of the target) or popularity (irrespective of the source).



Estimation must be based only on statistics that do not depend on the target (or source, respectively)!

Results on choosing articles.

Cox proportional hazard model, conditional on the active user.

$$\lambda(u, a, t|u) = \exp\left(\theta_1 \cdot s_1(u, a, t) + \theta_2 \cdot s_2(u, a, t) + \dots\right)$$

$s(u, a, t)$	parameter θ	$\exp(\theta)$
repetition	0.736 (0.032)*	2.088
four_cycle	0.084 (0.016)*	1.088
article_popularity	0.019 (0.009)*	1.019
article_popularity:user_activity	$-$0.004 (0.001)*	0.996

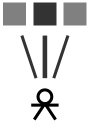
Results on becoming active.

Cox proportional hazard model for nodes initiating events.

$$\lambda(u, \cdot, t) = \exp\left(\theta_1 \cdot s_1(u, \cdot, t) + \theta_2 \cdot s_2(u, \cdot, t) + \ldots\right)$$

We have just one statistic that is independent of the target.

$s(u, \cdot, t)$	parameter θ	$\exp(\theta)$
user_activity	0.046 (0.002)***	1.047



Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

What do we observe (when observing events)?

A	**calls**	**B**	**at**	$\mathbf{t_1}$
B	calls	C	at	t_2
A	calls	C	at	t_3
B	calls	A	at	t_4
A	calls	C	at	t_5
D	calls	A	at	t_6

...

We observe **events**, e. g.,
A calls B at time t_1.

We also observe **non-events**:
A does not call C at t_1.
A does not call D at t_1.
B does not call A at t_1.
B does not call C at t_1.
...

Model estimation typically (also) uses information about non-events.

Dyads that ***could experience*** an event make up the **risk set** (can vary over time).

Global structure of risk set.

Explaining dyadic event rates.

- Risk set: all dyads that could experience an event.
- *Is an event on* (A, B) *more likely than on* (C, D)*, etc?*

Explaining actors' choices to select recipients of events.

- Given an observed event on (A, B),
- risk set: dyads (A, B'), for all nodes B'.
- *Given that A does initiate an event, is A more likely to direct the event at B rather than C, etc?*

Explaining preferences for specific *types* of events (e. g., cooperative vs. hostile).

- Given an observed event on (A, B),
- risk set: only the dyad (A, B); model the event type.
- *Given that there is an event on* (A, B)*, is it more likely to be a cooperative or a hostile event?*

Development of the risk set over time.

The risk set might be **exogenously** defined:

- **Nodes** may enter, leave, re-enter, ... the risk set (risk set contains all pairs of nodes at risk).
- **Dyads** may enter, leave, re-enter, ... the risk set.

The risk set might **endogenously** result from observed events.

For instance, modeling start and termination of conversations.

- A start-event puts the dyad at risk to experience a termination-event.
- A termination-event removes the dyad from the risk set.

Specifying time-varying risk sets in eventnet.

Time-varying risk sets are defined via node- or dyad-attributes.

Can constrain the risk set to dyads

- whose source/target have a non-zero value on a specified node-attribute;
- having a non-zero value on a specified dyad-attribute.

Attribute values can change as a function of input events (potentially dummy events created for the sole purpose of managing the risk set).

Outline.

Basics of relational event modeling.

- Testing hypotheses with event network data.
- Underlying concepts.
- Illustration: relational event modeling with eventnet.

Variations in different directions.

- Event types and weights; conditional event-type models.
- Covariates of nodes, dyads, and networks.
- One-mode, two-mode, multi-mode networks.
- Analyzing large event networks (sampling).
- Actor-oriented models.
- Time-varying risk set.

Case study: interaction events in international relations.

Computational Event Data System.

http://eventdata.parusanalytics.com/

Tool that generates sequences of events from news reports.

Event $e = (u, v, t, w)$ encodes

- u **source** actor (initiator);
- v **target** actor (recipient);
- t **timestamp**, the day on which e happened;
- w **weight** from -10 (most hostile) to $+10$ (most cooperative).

Examples of event-types and their associated weights:

OPTIMIST COMMENT	0.4	PESSIMIST COMMENT	-0.4
VISIT	1.9	ACCUSE	-2.0
PROMISE	4.0	REJECT	-4.0
AGREE	6.0	THREATEN	-6.0
EXTEND MIL AID	8.3	MILITARY DEMO	-7.6
MERGE, INTEGRATE	10.0	MILITARY ENGAGEMENT	-10.0

Gulf dataset.

http://eventdata.parusanalytics.com/ and
http://www.correlatesofwar.org/

We restrict the analysis to **state actors** (sovereign countries), excluding, e. g., organizations or ethnic groups.

data	time period	actors	events
FULL	1979/04/15 – 1999/03/31	202	304, 000
STATE ACTORS	1979/04/15 – 1999/03/31	168	218, 000

```
800923 IRQ IRN 223 MIL ENGAGEME
800923 IRN IRQ 223 MIL ENGAGEME
800924 USA IRN 121 CRITICIZE
800924 USA IRQ 121 CRITICIZE
800924 USA IRN 192 CUT ROUTINE
800924 USA IRQ 192 CUT ROUTINE
800924 UNO IRN 095 PLEAD
800924 UNO IRQ 095 PLEAD
800924 IRQ IRN 222 NONMIL DESTR
800924 IRN IRQ 223 MIL ENGAGEME
800924 IRQ IRN 122 DENIGRATE
800924 IRN IRQ 122 DENIGRATE
800924 IRQ IRN 031 MEET
...
```

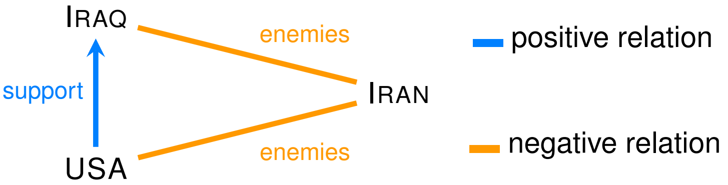
Additionally use node-level and dyad-level **covariates** (contiguity, alliances, capability, trade, democracy score, . . .).

Exemplary hypothesis to be tested.

Structural balance theory: do actors collaborate with the enemies of their enemies, fight the enemies of their friends, . . .

Anecdotal illustration of structural balance.

In the 1980s the USA provided support to the Iraq, although Iraq is not a typical ally of the USA.



Potential explanation: USA supported enemy of an enemy.

Here: statistical tests of structural balance theory in event data.

Hypotheses derived from structural balance theory.

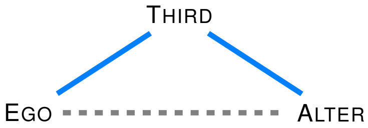
(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\text{The friend of my friend}}^{\textit{explanatory variable}}$ is my $\overbrace{\text{friend}}^{\textit{dep. var.}}$.
- The enemy of my friend is my enemy.
- The friend of my enemy is my enemy.
- The enemy of my enemy is my friend.

Hypotheses derived from structural balance theory.

(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\textbf{The friend of my friend}}^{\textit{explanatory variable}}$ **is my** $\overbrace{\textbf{friend}}^{\textit{dep. var.}}$.
- The enemy of my friend is my enemy.
- The friend of my enemy is my enemy.
- The enemy of my enemy is my friend.

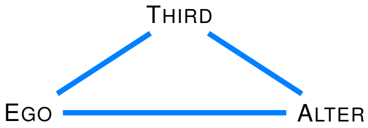


$\Rightarrow$ higher probability for cooperation, lower probability for conflict

Hypotheses derived from structural balance theory.

(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\textbf{The friend of my friend}}^{\textit{explanatory variable}}$ **is my** $\overbrace{\textbf{friend}}^{\textit{dep. var.}}$**.**
- The enemy of my friend is my enemy.
- The friend of my enemy is my enemy.
- The enemy of my enemy is my friend.

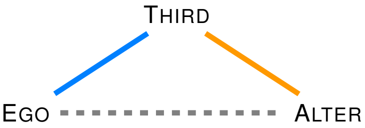


$\Rightarrow$ higher probability for cooperation, lower probability for conflict

Hypotheses derived from structural balance theory.

(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\text{The friend of my friend}}^{\textit{explanatory variable}}$ is my $\overbrace{\text{friend}}^{\textit{dep. var.}}$.
- **The enemy of my friend is my enemy.**
- The friend of my enemy is my enemy.
- The enemy of my enemy is my friend.

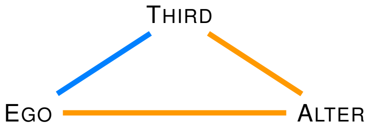


⇒ lower probability for cooperation, higher probability for conflict

Hypotheses derived from structural balance theory.

(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\text{The friend of my friend}}^{\textit{explanatory variable}}$ is my $\overbrace{\text{friend}}^{\textit{dep. var.}}$.
- **The enemy of my friend is my enemy.**
- The friend of my enemy is my enemy.
- The enemy of my enemy is my friend.

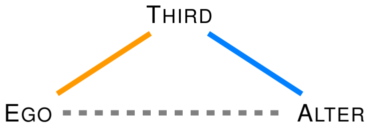


⇒ lower probability for cooperation, higher probability for conflict

Hypotheses derived from structural balance theory.

(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\text{The friend of my friend}}^{\textit{explanatory variable}}$ is my $\overbrace{\text{friend}}^{\textit{dep. var.}}$.
- The enemy of my friend is my enemy.
- **The friend of my enemy is my enemy.**
- The enemy of my enemy is my friend.

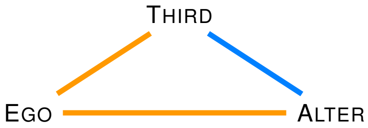


$\Rightarrow$ lower probability for cooperation, higher probability for conflict

Hypotheses derived from structural balance theory.

(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\text{The friend of my friend}}^{\textit{explanatory variable}}$ is my $\overbrace{\text{friend}}^{\textit{dep. var.}}$.
- The enemy of my friend is my enemy.
- **The friend of my enemy is my enemy.**
- The enemy of my enemy is my friend.

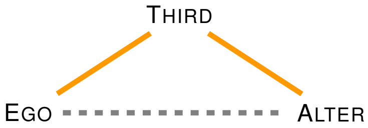


$\Rightarrow$ lower probability for cooperation, higher probability for conflict

Hypotheses derived from structural balance theory.

(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\text{The friend of my friend}}^{\textit{explanatory variable}}$ is my $\overbrace{\text{friend}}^{\textit{dep. var.}}$.
- The enemy of my friend is my enemy.
- The friend of my enemy is my enemy.
- **The enemy of my enemy is my friend.**

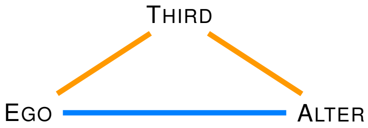


$\Rightarrow$ higher probability for cooperation, lower probability for conflict

Hypotheses derived from structural balance theory.

(Heider, 1946; Cartwright and Harary, 1956)

- $\overbrace{\text{The friend of my friend}}^{\textit{explanatory variable}}$ is my $\overbrace{\text{friend}}^{\textit{dep. var.}}$.
- The enemy of my friend is my enemy.
- The friend of my enemy is my enemy.
- **The enemy of my enemy is my friend.**



$\Rightarrow$ higher probability for cooperation, lower probability for conflict

What counts as cooperation or conflict?

Hypothesis: actors who are enemies of enemies have a higher probability to cooperate and a lower probability to engage in conflict.

Data: typed and weighted events.

- **type** of interaction (more than 100 different types).
- **weight** of interaction (numeric, ranging from -10 to 10).

How to assess the **probability of cooperation (conflict)**?

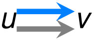
Modeling the rate of cooperative/conflictive events

(cutoff at weight equal to 0.0).

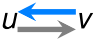
Statistics (I): inertia and reciprocity.

explanatory variables: dependent var.: $u \rightarrow v$;

dependence on dyad history

$\texttt{repetition}^{\pm}(u, v; G[E; t])$ 

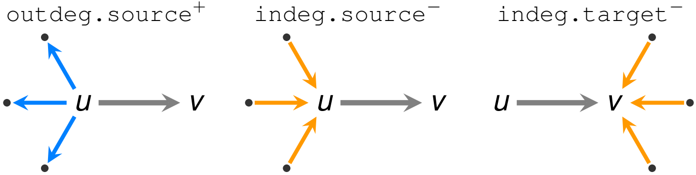
dependence on reverse dyad history

$\texttt{reciprocity}^{\pm}(u, v; G[E; t])$ 

Statistics (II): activity and popularity.

explanatory variables: dependent var.: $u \rightarrow v$;

accounting for differences in roles and positions (degree)

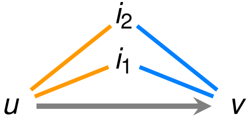


Varying source/target, positive/negative, and popularity/activity yields eight statistics.

Statistics (III): structural balance.

explanatory variables: dependent var.: $u \longrightarrow v$;

$$\texttt{friend.of.enemy}(u, v; G[E; t]) = \sum_{\text{actors: } i} \min(a^{-}(u, i; t), a^{+}(i, v; t))$$



Similar for friends of friends, enemies of friends, and enemies of enemies.

Statistics (IV): covariates.

Events on a dyad (u, v) depend on exogenous covariates:

- u and v are allies or not;
- geographical distance between u and v; joint border
- democracy level of u and v;
- capability scores (size, military and economic power);
- u or v (or both) are a “major power”;
- trade between u and v;
- number of joint IGOs of u and v

Results: rate of cooperative and conflictive events.

Cox proportional-hazard model

	cooperation	cooperation	conflict	conflict
pos.repetition	$-0.16\ (0.01)^{***}$	$-0.14\ (0.01)^{***}$	$-0.25\ (0.01)^{***}$	$-0.21\ (0.01)^{***}$
neg.repetition	$-0.17\ (0.01)^{***}$	$-0.14\ (0.01)^{***}$	$-0.10\ (0.01)^{***}$	$-0.10\ (0.01)^{***}$
pos.reciprocation	$-0.14\ (0.01)^{***}$	$-0.13\ (0.01)^{***}$	$-0.20\ (0.01)^{***}$	$-0.18\ (0.01)^{***}$
neg.reciprocation	$-0.10\ (0.01)^{***}$	$-0.08\ (0.01)^{***}$	$-0.14\ (0.01)^{***}$	$-0.13\ (0.01)^{***}$
pos.outdeg.source	$0.54\ (0.02)^{***}$	$0.67\ (0.02)^{***}$	$0.36\ (0.02)^{***}$	$0.48\ (0.02)^{***}$
neg.outdeg.source	$0.20\ (0.02)^{***}$	$0.20\ (0.02)^{***}$	$0.47\ (0.02)^{***}$	$0.40\ (0.02)^{***}$
pos.indeg.source	$0.49\ (0.02)^{***}$	$0.40\ (0.02)^{***}$	$0.46\ (0.02)^{***}$	$0.42\ (0.02)^{***}$
neg.indeg.source	$0.10\ (0.02)^{***}$	$0.06\ (0.02)^{**}$	$0.22\ (0.02)^{***}$	$0.15\ (0.02)^{***}$
pos.outdeg.target	$0.38\ (0.02)^{***}$	$0.43\ (0.02)^{***}$	$0.36\ (0.02)^{***}$	$0.40\ (0.02)^{***}$
neg.outdeg.target	$0.14\ (0.02)^{***}$	$0.10\ (0.02)^{***}$	$0.33\ (0.02)^{***}$	$0.24\ (0.02)^{***}$
pos.indeg.target	$0.63\ (0.02)^{***}$	$0.62\ (0.02)^{***}$	$0.45\ (0.02)^{***}$	$0.50\ (0.02)^{***}$
neg.indeg.target	$0.13\ (0.02)^{***}$	$0.12\ (0.02)^{***}$	$0.38\ (0.02)^{***}$	$0.32\ (0.02)^{***}$
friend.of.friend	$\mathbf{0.93\ (0.03)}^{***}$	$\mathbf{0.51\ (0.03)}^{***}$	$\mathbf{0.48\ (0.03)}^{***}$	$\mathbf{0.13\ (0.03)}^{***}$
friend.of.enemy	$\mathbf{-0.62\ (0.02)}^{***}$	$\mathbf{-0.49\ (0.02)}^{***}$	$\mathbf{-0.39\ (0.03)}^{***}$	$\mathbf{-0.31\ (0.03)}^{***}$
enemy.of.friend	$\mathbf{-0.64\ (0.02)}^{***}$	$\mathbf{-0.46\ (0.02)}^{***}$	$\mathbf{-0.41\ (0.03)}^{***}$	$\mathbf{-0.30\ (0.03)}^{***}$
enemy.of.enemy	$\mathbf{0.46\ (0.02)}^{***}$	$\mathbf{0.36\ (0.02)}^{***}$	$\mathbf{0.28\ (0.03)}^{***}$	$\mathbf{0.30\ (0.03)}^{***}$
logDistance		$-0.29\ (0.01)^{***}$		$-0.32\ (0.01)^{***}$
logCapRat		$-0.11\ (0.01)^{***}$		$-0.07\ (0.00)^{***}$
majorPower		$0.23\ (0.01)^{***}$		$0.35\ (0.01)^{***}$
contiguity		$0.08\ (0.01)^{***}$		$0.17\ (0.01)^{***}$
logTrade		$0.16\ (0.01)^{***}$		$0.15\ (0.00)^{***}$
polityWeakLink		$-0.03\ (0.00)^{***}$		$-0.01\ (0.00)^{***}$
logJointIGO		$-0.14\ (0.01)^{***}$		$-0.12\ (0.01)^{***}$
allies		$0.16\ (0.01)^{***}$		$0.11\ (0.01)^{***}$
Num. events	97557	97557	119922	119922
Num. obs.	236714	236714	259992	259992

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

Summary: rate of cooperative and conflictive events.

State actors **cooperate more** frequently with the friends of their friends and with the enemies of their enemies;
and they **cooperate less** frequently with the friends of their enemies and with the enemies of their friends;

$\Rightarrow$ **supporting balance theory.**

	cooperation	cooperation	conflict	conflict
other network effects				
friend.of.friend	**0.93 (0.03)*****	**0.51 (0.03)*****	**0.48 (0.03)*****	**0.13 (0.03)*****
friend.of.enemy	**−0.62 (0.02)*****	**−0.49 (0.02)*****	**−0.39 (0.03)*****	**−0.31 (0.03)*****
enemy.of.friend	**−0.64 (0.02)*****	**−0.46 (0.02)*****	**−0.41 (0.03)*****	**−0.30 (0.03)*****
enemy.of.enemy	**0.46 (0.02)*****	**0.36 (0.02)*****	**0.28 (0.03)*****	**0.30 (0.03)*****
		covariates		*covariates*

But they also **fight more** frequently the friends of their friends and the enemies of their enemies;
and they **fight less** frequently the friends of their enemies and the enemies of their friends;

$\Rightarrow$ **contradicting balance theory.**

Conditional type (or weight) models.

Decomposing the probability of typed events.

$$P(\text{negative interaction}) = P(\text{interaction}) \times P(\text{negative} \mid \text{interaction})$$

Lerner (2016). **Structural balance in signed networks: Separating the probability to interact from the tendency to fight**. *Social Networks*.

Decomposing the probability of typed events.

$$P(\text{negative interaction}) = P(\text{interaction}) \times P(\text{negative} \mid \text{interaction})$$

Probability to interact at all

- is very skewed over the set of dyads;
- might depend on the same variables (e. g., being enemy of enemy) as the type of events.

Modeling the marginal probability of typed events confounds processes shaping the frequency of interaction with processes shaping the type of interaction.

Lerner (2016). **Structural balance in signed networks: Separating the probability to interact from the tendency to fight**. *Social Networks*.

Related work.

modeling positive or negative reviews among literary authors and critics

de Nooy (2008):

> *In my case, the **presence or absence of a line** (literary evaluation) is **not the important phenomenon** to be explained because it depends on events and constraints outside the power of the actors in the network.* [. . .] *As we will see, it is possible and interesting to predict the **sign of an evaluation, conditional on the presence** of an evaluation, from the pattern of signs of previous evaluations.*

Here we argue: even when the occurrence of ties could be explained, the conditional sign may be more appropriate.

de Nooy (2008). **Signs over time**. *Journal of Social Structure*.

Testing balance theory with conditional type models.

Rather than modeling the rate of signed events

$$P(\text{signed interaction}) = P(\text{interaction}) \times P(\text{sign} \mid \text{interaction})$$

model the conditional probability that an event has positive/negative sign.

Alternatively: model the conditional event weight.

Lerner (2016). **Structural balance in signed networks: Separating the probability to interact from the tendency to fight**. *Social Networks*.

Testing balance theory with conditional type models.

Logit model for conditional probability of cooperative events

$$prob(coop|event) = \text{logit}^{-1}(\sum_i \alpha_i \cdot s_i) \ .$$

Parameters α are estimated on all events (not on non-events).

Linear regression for conditional event weight

$$f(weight|event) = \mathcal{N}(\sum_i \beta_i \cdot s_i, \sigma) \ .$$

Parameters β are estimated on all events (not on non-events).

Results: modeling conditional event type / weight.

	(logit regression) cond. prob. cooperation		(linear regression) conditional event weight	
(Intercept)	$-0.23\ (0.00)^{***}$	$-0.04\ (0.01)^{**}$	$-1.05\ (0.01)^{***}$	$-0.78\ (0.03)^{***}$
pos.repetition	$0.16\ (0.02)^{***}$	$0.16\ (0.02)^{***}$	$0.57\ (0.04)^{***}$	$0.54\ (0.04)^{***}$
neg.repetition	$-0.31\ (0.02)^{***}$	$-0.23\ (0.02)^{***}$	$-1.03\ (0.04)^{***}$	$-0.84\ (0.04)^{***}$
pos.reciprocation	$0.05\ (0.01)^{***}$	$0.02\ (0.01)$	$0.17\ (0.03)^{***}$	$0.07\ (0.03)^{*}$
neg.reciprocation	$-0.33\ (0.02)^{***}$	$-0.20\ (0.02)^{***}$	$-0.69\ (0.03)^{***}$	$-0.41\ (0.03)^{***}$
pos.outdeg.source	$0.26\ (0.02)^{***}$	$0.14\ (0.02)^{***}$	$0.66\ (0.04)^{***}$	$0.12\ (0.04)^{**}$
neg.outdeg.source	$-0.21\ (0.02)^{***}$	$-0.14\ (0.02)^{***}$	$-0.45\ (0.04)^{***}$	$-0.16\ (0.04)^{***}$
pos.indeg.source	$-0.01\ (0.02)$	$0.04\ (0.02)^{*}$	$-0.15\ (0.04)^{***}$	$0.17\ (0.04)^{***}$
neg.indeg.source	$0.05\ (0.02)^{**}$	$-0.00\ (0.02)$	$0.04\ (0.04)$	$-0.23\ (0.04)^{***}$
pos.outdeg.target	$0.13\ (0.01)^{***}$	$0.01\ (0.02)$	$0.33\ (0.03)^{***}$	$-0.11\ (0.03)^{**}$
neg.outdeg.target	$-0.09\ (0.02)^{***}$	$-0.01\ (0.02)$	$-0.06\ (0.04)$	$0.20\ (0.04)^{***}$
pos.indeg.target	$-0.06\ (0.02)^{**}$	$0.00\ (0.02)$	$-0.19\ (0.05)^{***}$	$0.10\ (0.05)^{*}$
neg.indeg.target	$-0.04\ (0.02)^{*}$	$-0.09\ (0.02)^{***}$	$-0.06\ (0.05)$	$-0.32\ (0.05)^{***}$
friend.of.friend	$\mathbf{0.49\ (0.03)}^{***}$	$\mathbf{0.43\ (0.03)}^{***}$	$\mathbf{1.44\ (0.05)}^{***}$	$\mathbf{1.25\ (0.05)}^{***}$
friend.of.enemy	$\mathbf{-0.40\ (0.02)}^{***}$	$\mathbf{-0.33\ (0.02)}^{***}$	$\mathbf{-1.06\ (0.05)}^{***}$	$\mathbf{-0.86\ (0.05)}^{***}$
enemy.of.friend	$\mathbf{-0.53\ (0.02)}^{***}$	$\mathbf{-0.45\ (0.02)}^{***}$	$\mathbf{-1.39\ (0.05)}^{***}$	$\mathbf{-1.17\ (0.05)}^{***}$
enemy.of.enemy	$\mathbf{0.49\ (0.02)}^{***}$	$\mathbf{0.36\ (0.02)}^{***}$	$\mathbf{1.19\ (0.04)}^{***}$	$\mathbf{0.86\ (0.04)}^{***}$
logDistance		$0.01\ (0.01)$		$0.34\ (0.02)^{***}$
logCapRat		$0.03\ (0.01)^{***}$		$0.00\ (0.01)$
majorPower		$-0.37\ (0.02)^{***}$		$-0.64\ (0.04)^{***}$
contiguity		$-0.33\ (0.02)^{***}$		$-0.52\ (0.03)^{***}$
logTrade		$0.26\ (0.01)^{***}$		$0.56\ (0.02)^{***}$
polityWeakLink		$-0.02\ (0.01)^{***}$		$-0.15\ (0.01)^{***}$
logJointIGO		$-0.09\ (0.01)^{***}$		$-0.13\ (0.01)^{***}$
allies		$0.42\ (0.01)^{***}$		$0.95\ (0.03)^{***}$
Num. obs.	217,479	217,479	217,479	217,479

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

Summary: modeling conditional event type / weight.

Given that state actors interact with the friends of their friends, or with the enemies of their enemies, their interaction is **more likely to be cooperative** and has a **higher expected weight.**

$\Rightarrow$ **supporting balance theory.**

	(logit regression) cond. prob. cooperation		(linear regression) conditional event weight	
other network effects				
friend.of.friend	**0.49 (0.03)*****	**0.43 (0.03)*****	**1.44 (0.05)*****	**1.25 (0.05)*****
friend.of.enemy	$-$**0.40 (0.02)*****	$-$**0.33 (0.02)*****	$-$**1.06 (0.05)*****	$-$**0.86 (0.05)*****
enemy.of.friend	$-$**0.53 (0.02)*****	$-$**0.45 (0.02)*****	$-$**1.39 (0.05)*****	$-$**1.17 (0.05)*****
enemy.of.enemy	**0.49 (0.02)*****	**0.36 (0.02)*****	**1.19 (0.04)*****	**0.86 (0.04)*****
		covariates		*covariates*
Num. obs.	217,479	217,479	217,479	217,479

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

Given that state actors interact with the friends of their enemies, or with the enemies of their friends, their interaction is **less likely to be cooperative** and has a **lower expected weight.**

$\Rightarrow$ **supporting balance theory.**