Latent Stochastic Actor Oriented Models for Relational Event Data

J.A. Lospinoso[1,2] J.H. Koskinen[2] T.A.B. Snijders[2]

[1]Network Science Center
United States Military Academy
[2]Department of Statistics
University of Oxford

March 15, 2012

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **15 MAR 2012**	2. REPORT TYPE **N/A**	3. DATES COVERED
4. TITLE AND SUBTITLE **Latent Stochastic Actor Oriented Models for Relational Event Data**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **US Military Academy West Point, NY 10996**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**60**	

A convenient and powerful combination...

- Using relational event datasets to study social network dynamics can have many practical up-sides.
- SAOM effects (which are based on digraphs) are convenient (evidenced by the large amount of effects that have already been studied).

Practical considerations for relational event data

- Collecting high quality relational event datasets can be cheap and easy.

Practical considerations for relational event data

- Collecting high quality relational event datasets can be cheap and easy.
- Datasets can be quite large (and correspondingly contain lots of information).

Practical considerations for relational event data

- Collecting high quality relational event datasets can be cheap and easy.
- Datasets can be quite large (and correspondingly contain lots of information).
- We can "follow the sound of marching feet."

Practical considerations for “relational event history models”

Butts [2008] proposes to study relational event histories with a Cox model.

Practical considerations for "relational event history models"

Butts [2008] proposes to study relational event histories with a Cox model.
Devising good effects with relational event histories can be less convenient than with SAOMs:

- How do we handle the passing of time? How much more important is something that happened last week than last year?

Practical considerations for "relational event history models"

Butts [2008] proposes to study relational event histories with a Cox model.
Devising good effects with relational event histories can be less convenient than with SAOMs:

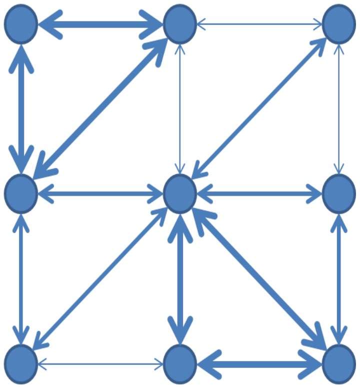
- How do we handle the passing of time? How much more important is something that happened last week than last year?
- How do we model "second order terms" like transitivity, balance, and betweenness?

Idea: model the evolution of the affective network with SAOMs

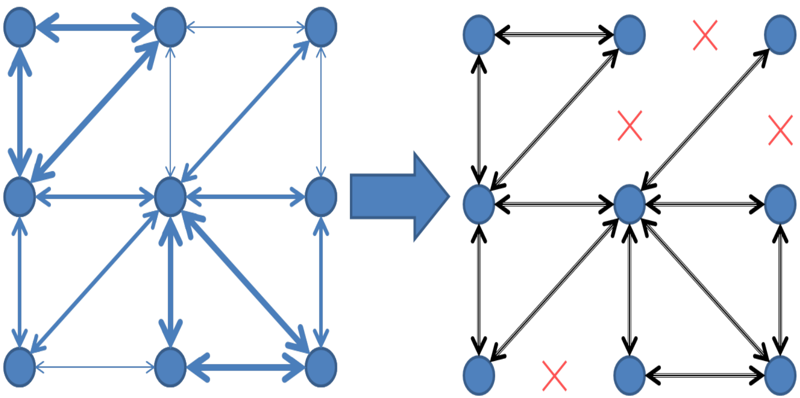
...but where are the digraphs?

- A fairly common practice is to do some "binning" to make data that looks like digraph panel data
- Intuitively, we suppose that more interaction in a time interval signals an affective tie.

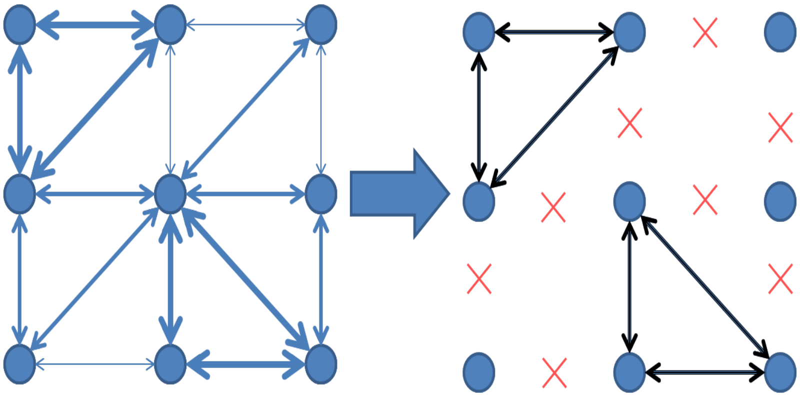
We can easily construct weighted networks over some time intervals...



And use some “threshold” to yield a digraph:



Here's a higher threshold:



But really we've traded one set of difficulties for another:

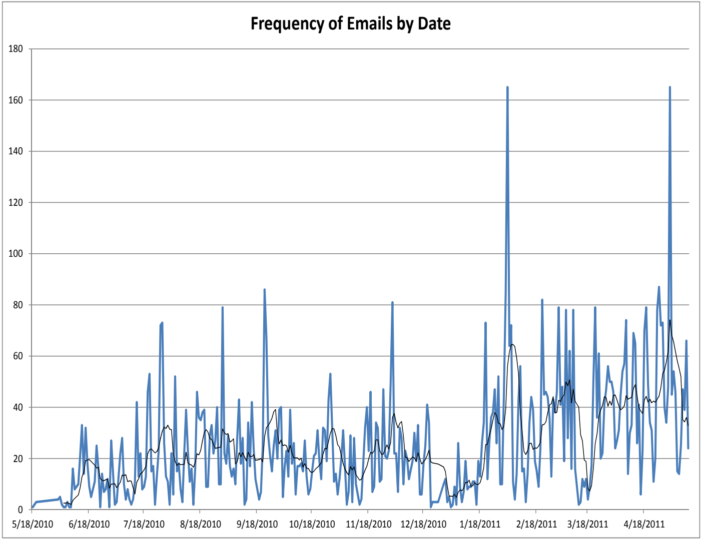
- How much interaction constitutes a tie?
- How do you choose intervals?
- How do we minimize information loss?
- How can we be sure that artifacts don't creep into our results?

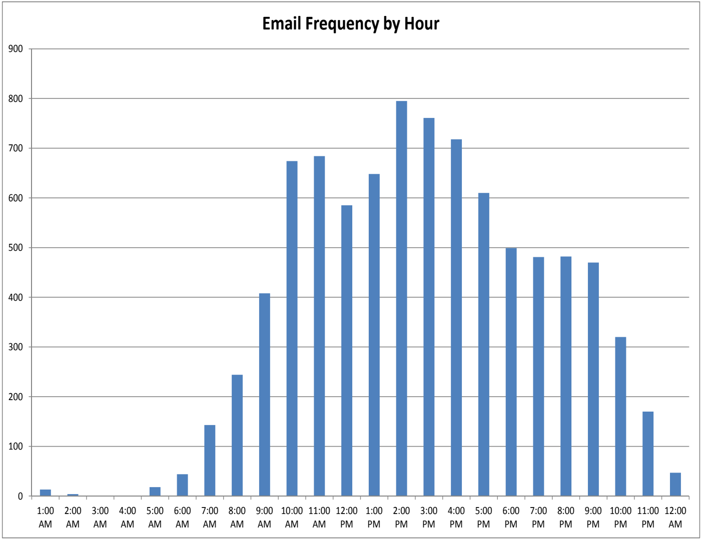
Data: The Ikenet Study

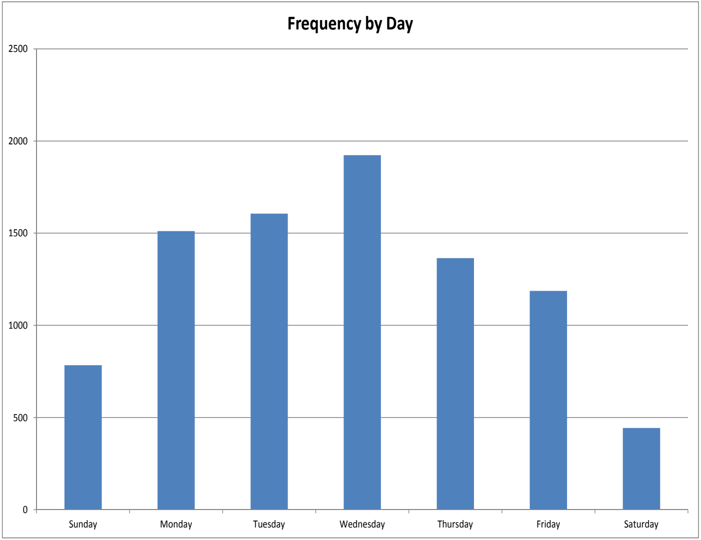
- Kate Coronges, in collaboration with the USMA Network Science Center, collected the email traffic among 22 mid-career Army officers over a 13-month period
- We will use one covariate which indicates whether the officer is a captain (lesser rank) or a major (greater rank).
- For emails that are not (a) broadcast emails (i.e. those emails sent to the whole group) or (b) self-sent emails, we add one "relational event" per recipient in a random order. This process yields 8819 events.

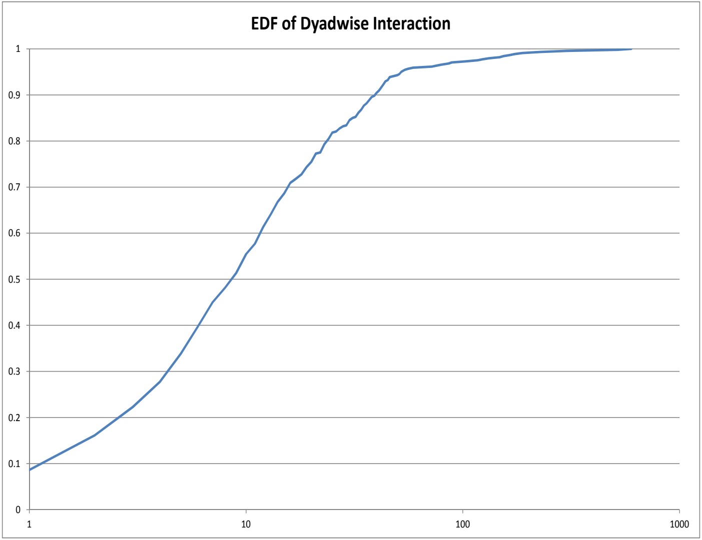
Data: The Ikenet Study

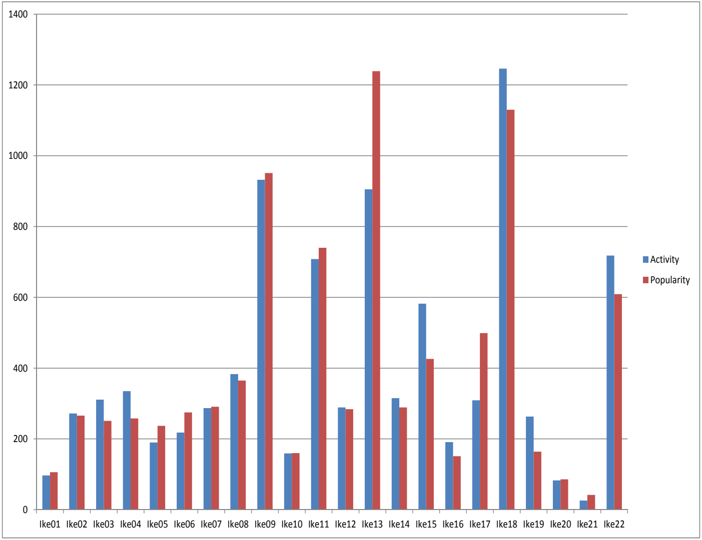
- Kate Coronges, in collaboration with the USMA Network Science Center, collected the email traffic among 22 mid-career Army officers over a 13-month period
- We will use one covariate which indicates whether the officer is a captain (lesser rank) or a major (greater rank).
- For emails that are not (a) broadcast emails (i.e. those emails sent to the whole group) or (b) self-sent emails, we add one "relational event" per recipient in a random order. This process yields 8819 events.
- Bonus: friendship surveys are collected at each month


Frequency of Emails by Date
180
160
140
120
100
80
60
40
20
0
5/18/2010
6/18/2010
7/18/2010
8/18/2010
9/18/2010
10/18/2010
11/18/2010
12/18/2010
1/18/2011
2/18/2011
3/18/2011
4/18/2011


Email Frequency by Hour
900
800
700
600
500
400
300
200
100
0
1:00 AM
2:00 AM
3:00 AM
4:00 AM
5:00 AM
6:00 AM
7:00 AM
8:00 AM
9:00 AM
10:00 AM
11:00 AM
12:00 PM
1:00 PM
2:00 PM
3:00 PM
4:00 PM
5:00 PM
6:00 PM
7:00 PM
8:00 PM
9:00 PM
10:00 PM
11:00 PM
12:00 AM


Frequency by Day
2500
2000
1500
1000
500
0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday


EDF of Dyadwise Interaction
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
10
100
1000


1400
1200
1000
800
600
400
200
0
Ike01
Ike02
Ike03
Ike04
Ike05
Ike06
Ike07
Ike08
Ike09
Ike10
Ike11
Ike12
Ike13
Ike14
Ike15
Ike16
Ike17
Ike18
Ike19
Ike20
Ike21
Ike22
Activity
Popularity

Binning

It's not obvious how we determine

- time intervals
- dichotimization thresholds

to create waves out of this data.

Binning

It's not obvious how we determine

- time intervals
- dichotimization thresholds

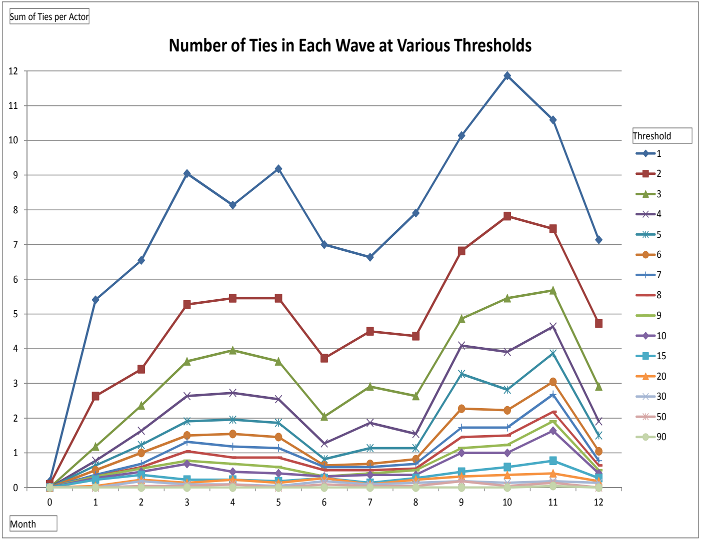
to create waves out of this data.
To simplify the example, we (quite arbitrarily) decide to create intervals by month of the year, yielding 13 digraphs in the panel.

Binning

It's not obvious how we determine

- time intervals
- dichotimization thresholds

to create waves out of this data.
To simplify the example, we (quite arbitrarily) decide to create intervals by month of the year, yielding 13 digraphs in the panel.
How do various threshold values affect these digraphs?


Sum of Ties per Actor
Number of Ties in Each Wave at Various Thresholds
12
11
10
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9
10
11
12
Month
Threshold
1
2
3
4
5
6
7
8
9
10
15
20
30
50
90

Model

Since there is no clear choice of a threshold at which the network density is uniquely plausible, we'll run estimations of the following model at various thresholds:

Model

Since there is no clear choice of a threshold at which the network density is uniquely plausible, we'll run estimations of the following model at various thresholds:

- Outdegree: "do I like to have ties?"
- Reciprocity: "do I like to have ties to match incoming ties?"

Model

Since there is no clear choice of a threshold at which the network density is uniquely plausible, we'll run estimations of the following model at various thresholds:

- Outdegree: "do I like to have ties?"
- Reciprocity: "do I like to have ties to match incoming ties?"
- Trans. Triplets: "do I like to have ties to ties-of-ties ?"
- Balance: "do I like my ties to match my current alters' ties?"

Model

Since there is no clear choice of a threshold at which the network density is uniquely plausible, we'll run estimations of the following model at various thresholds:

- Outdegree: "do I like to have ties?"
- Reciprocity: "do I like to have ties to match incoming ties?"
- Trans. Triplets: "do I like to have ties to ties-of-ties ?"
- Balance: "do I like my ties to match my current alters' ties?"
- In. Popularity: "do I like to tie to alters with lots of incoming ties?"
- In. Activity: "do I like to create more ties when I recieve lots of ties?"

Model

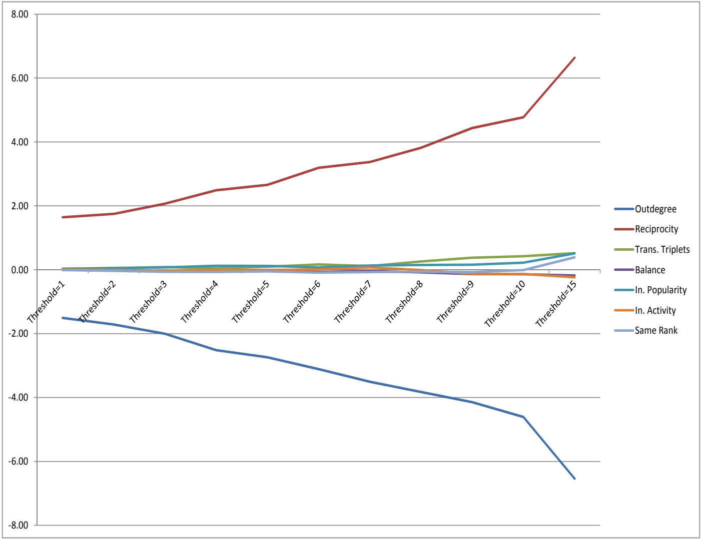
Since there is no clear choice of a threshold at which the network density is uniquely plausible, we'll run estimations of the following model at various thresholds:

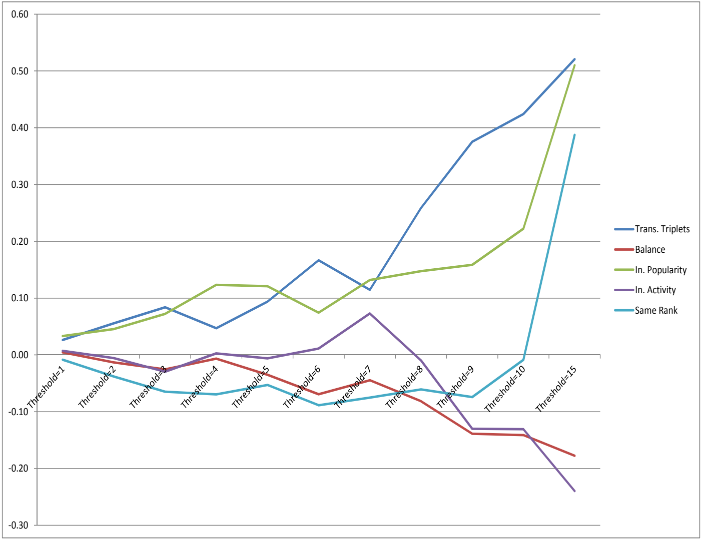
- Outdegree: "do I like to have ties?"
- Reciprocity: "do I like to have ties to match incoming ties?"
- Trans. Triplets: "do I like to have ties to ties-of-ties ?"
- Balance: "do I like my ties to match my current alters' ties?"
- In. Popularity: "do I like to tie to alters with lots of incoming ties?"
- In. Activity: "do I like to create more ties when I recieve lots of ties?"
- Same Rank: "do I like to create ties with other officers of the same rank?"

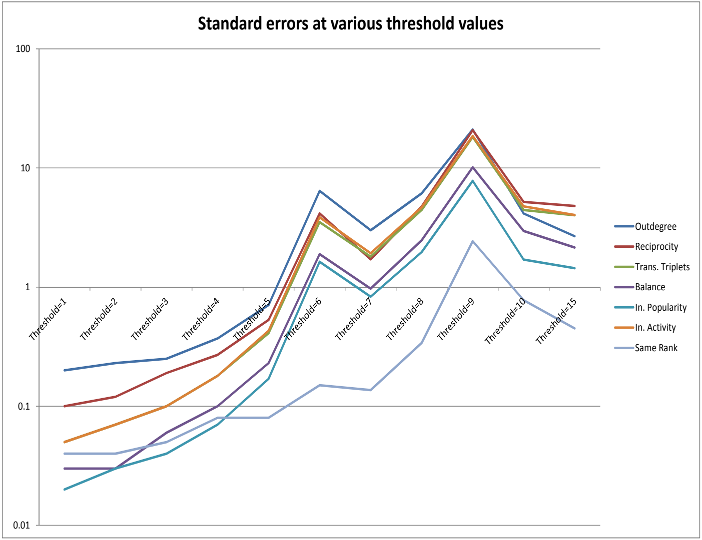
MoM Estimation [Snijders, 2001] Results

Threshold	1	2	3	4	5	6
Outdegree	-1.51	-1.71	-2.00	-2.52	-2.74	-3.11
Reciprocity	1.64	1.75	2.07	2.49	2.66	3.19
Trans. Triplets	0.03	0.06	0.08	0.05	0.09	0.17
Balance	0.00	-0.01	-0.03	-0.01	-0.04	-0.07
In. Popularity	0.03	0.05	0.07	0.12	0.12	0.07
In. Activity	0.01	-0.01	-0.03	0.00	-0.01	0.01
Same Rank	-0.01	-0.04	-0.06	-0.07	-0.05	-0.09

Threshold	7	8	9	10	15
Outdegree	-3.51	-3.83	-4.14	-4.61	-6.54
Reciprocity	3.37	3.82	4.44	4.77	6.64
Trans. Triplets	0.11	0.26	0.38	0.42	0.52
Balance	-0.04	-0.08	-0.14	-0.14	-0.18
In. Popularity	0.13	0.15	0.16	0.22	0.51
In. Activity	0.07	-0.01	-0.13	-0.13	-0.24
Same Rank	-0.08	-0.06	-0.07	-0.01	0.39


8.00
6.00
4.00
2.00
0.00
-2.00
-4.00
-6.00
-8.00
Threshold=1
Threshold=2
Threshold=3
Threshold=4
Threshold=5
Threshold=6
Threshold=7
Threshold=8
Threshold=9
Threshold=10
Threshold=15
Outdegree
Reciprocity
Trans. Triplets
Balance
In. Popularity
In. Activity
Same Rank


0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
Threshold=1
Threshold=2
Threshold=3
Threshold=4
Threshold=5
Threshold=6
Threshold=7
Threshold=8
Threshold=9
Threshold=10
Threshold=15
Trans. Triplets
Balance
In. Popularity
In. Activity
Same Rank


Standard errors at various threshold values
100
10
1
0.1
0.01
Threshold=1
Threshold=2
Threshold=3
Threshold=4
Threshold=5
Threshold=6
Threshold=7
Threshold=8
Threshold=9
Threshold=10
Threshold=15
Outdegree
Reciprocity
Trans. Triplets
Balance
In. Popularity
In. Activity
Same Rank

Summary on binning

We get convenient SAOM-based inference (easy to interpret effects) out of relational event data, but...

- Choice of binning parameters is not obvious and may have important consequences on results
- How much information was lost by binning?
- Did we introduce artifacts?

Bringing relational events into the SAOM

Let's consider how we might bring the relational events into the SAOM:

- We consider the relational event *mode* as just another aspect over which *ego* has complete control.

Bringing relational events into the SAOM

Let's consider how we might bring the relational events into the SAOM:

- We consider the relational event *mode* as just another aspect over which *ego* has complete control.
- An ego may then choose to modify her outgoing ties, or to send a relational event to an alter.

Bringing relational events into the SAOM

Let's consider how we might bring the relational events into the SAOM:

- We consider the relational event *mode* as just another aspect over which *ego* has complete control.
- An ego may then choose to modify her outgoing ties, or to send a relational event to an alter.
- As the network is unobserved, this is a *hidden Markov model*

We can think of this process in terms of ordered *ministep* tuples $v = (i, j, k, t) \in \mathcal{V}$ where:

- $i \in \mathcal{N}$ indexes the ego
- $j \in \mathcal{N}$ indexes the alter
- $k \in \mathcal{K}$ indexes the aspect
- $t \in \mathfrak{R}$ indexes time

where $\mathcal{N}$ is the actor set, $\mathcal{K}$ is the set of networks and modes.

We can think of this process in terms of ordered *ministep* tuples $v = (i, j, k, t) \in \mathcal{V}$ where:

- $i \in \mathcal{N}$ indexes the ego
- $j \in \mathcal{N}$ indexes the alter
- $k \in \mathcal{K}$ indexes the aspect
- $t \in \mathfrak{R}$ indexes time

where $\mathcal{N}$ is the actor set, $\mathcal{K}$ is the set of networks and modes.

For convenience, these indices can be used as functions, e.g. $i(v_5)$ refers to the ego of the 5-th ministep.

We can think of this process in terms of ordered *ministep* tuples $v = (i, j, k, t) \in \mathcal{V}$ where:

- $i \in \mathcal{N}$ indexes the ego
- $j \in \mathcal{N}$ indexes the alter
- $k \in \mathcal{K}$ indexes the aspect
- $t \in \mathfrak{R}$ indexes time

where $\mathcal{N}$ is the actor set, $\mathcal{K}$ is the set of networks and modes.

For convenience, these indices can be used as functions, e.g. $i(v_5)$ refers to the ego of the 5-th ministep.

Also for convenience, let $\mathcal{V}_\tau = \{v_b \in \mathcal{V} : t(V_b) < \tau\}$

If we let

- $p_0(X^*) = p_0(X^*|\theta)$ represent the PMF of the initial network
- $f_v(v_b) = f_v(v_b|\mathcal{V}_{t(v_b)}, x^*, \theta)$ represent the conditional PDF given all ministeps occurring before v_b *and* the initial network.

If we let

- $p_0(X^*) = p_0(X^*|\theta)$ represent the PMF of the initial network
- $f_v(v_b) = f_v(v_b|\mathcal{V}_{t(v_b)}, x^*, \theta)$ represent the conditional PDF given all ministeps occurring before v_b *and* the initial network.

then the likelihood is

$$L(\theta|\mathcal{V}, x^*) = p_0(x^*) \prod_{v_b \in \mathcal{V}} f_v(v_b) \qquad (1)$$

For $p_0(x^*)$, we are free to choose anything. In the context of the Ikenet data, we use a simple, dyad-independent model:

$$p_0(X) = \prod_{i,j \neq i \in \mathcal{N}} p(X_{ij} = x_{ij}, X_{ji} = x_{ji} | \theta) \tag{2}$$

For $p_0(x^*)$, we are free to choose anything. In the context of the Ikenet data, we use a simple, dyad-independent model:

$$p_0(X) = \prod_{i,j \neq i \in \mathcal{N}} p(X_{ij} = x_{ij}, X_{ji} = x_{ji} | \theta) \tag{2}$$

We include terms in $p(X_{ij} = 1 | x_{ji}, \theta)$ for density and reciprocity.

For $f_v(v_b)$, we decompose the joing wait-time/ego/aspect density, and alter selection density:

$$f_v(v_b) = f_{ikt}\left\{i(v_b), k(v_b), t(v_b)\right\} f_{j|ikt}\left\{j(v_b)\right\}$$

For $f_v(v_b)$, we decompose the joing wait-time/ego/aspect density, and alter selection density:

$$f_v(v_b) = f_{ikt}\left\{i(v_b), k(v_b), t(v_b)\right\} f_{j|ikt}\left\{j(v_b)\right\}$$

Just as in the regular SAOM, we suppose f_{ikt} is a negative exponential process, and we have a different rate for each aspect $\mathcal{K}$.

For a network aspect k, our alter selection density is the SAOM alter selection density:

$$f_{j|ikt}\{j\} = \frac{g(i \rightsquigarrow j)}{\sum_{n \in \mathcal{N}} g(i \rightsquigarrow k)}$$

For a relational event mode aspect k, our alter selection density can be anything, but we choose a parsimonious "observation function"

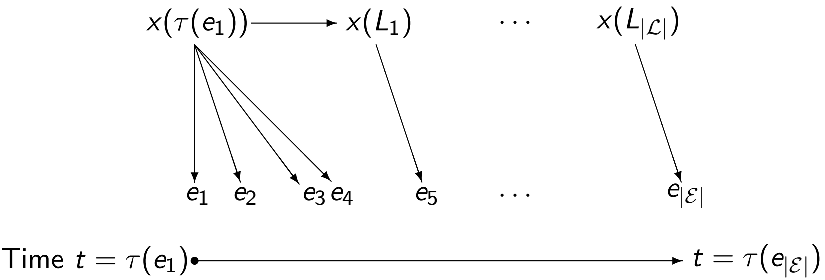
$$f_{j|ikt}\{j\} = \frac{\exp\{\gamma x_{ij}(t)\}}{\sum_{n \neq i \in \mathcal{N}} \exp\{\gamma x_{in}(t)\}} .$$

For a relational event mode aspect k, our alter selection density can be anything, but we choose a parsimonious "observation function"

$$f_{j|ikt}\{j\} = \frac{\exp\{\gamma x_{ij}(t)\}}{\sum_{n \neq i \in \mathcal{N}} \exp\{\gamma x_{in}(t)\}} \; .$$

The "network effect" γ represents the tendency for an ego to want to send relational events to alters with whom they have a network tie.

DAG for the L-SAOM



In principle, estimation is a straightforward MCMC-MLE scheme [Snijders et al., 2010].

In principle, estimation is a straightforward MCMC-MLE scheme [Snijders et al., 2010].

We've added a few proposals to the MCMC scheme to account for the lack of so-called "parity" conditions. The details are pretty involved.

In principle, estimation is a straightforward MCMC-MLE scheme [Snijders et al., 2010].

We've added a few proposals to the MCMC scheme to account for the lack of so-called "parity" conditions. The details are pretty involved.

For the sake of time, we'll surf right over these details. They are implemented in a .NET port of RSiena:

`github.com/JLospinoso/sie`

Estimation results

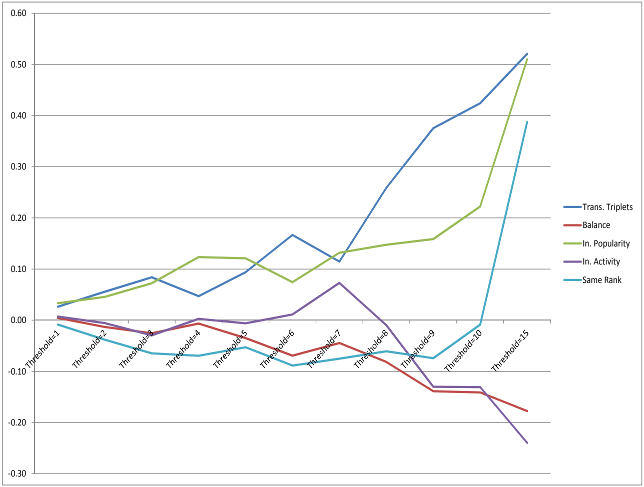
	$\hat{\theta}$	s.e. $\hat{\theta}$		$\hat{\theta}$	s.e. $\hat{\theta}$
Network Dynamics			*Observation*		
Outdegree	-3.19	0.70	Network	2.30	0.71
Reciprocity	0.84	0.15	*Initial Network*		
Trans. Triplets	-0.03	0.06	Outdegree	-3.49	0.53
Balance	0.04	0.07	Reciprocity	7.24	0.94
In. Popularity	0.07	0.01	*Pacing*		
In. Activity	0.07	0.08	Network	4.73	1.05
Same Rank	0.15	0.06	Email	5.99	0.27

How do these compare with the actual friendship survey data?

How do these compare with the actual friendship survey data?

	Email		Friendship	
	$\hat{\theta}$	s.e. $\hat{\theta}$	$\hat{\theta}$	s.e. $\hat{\theta}$
Outdegree	-3.19	0.70	-0.95	0.20
Reciprocity	0.84	0.15	1.12	0.19
Trans. Triplets	-0.03	0.06	-0.24	0.23
Balance	0.04	0.07	0.21	0.13
In. Popularity	0.07	0.01	0.09	0.03
In. Activity	0.07	0.08	0.03	0.10
Same Rank	0.15	0.06	0.24	0.09

Binning results again:



Summary

- Using SAOM effects to model relational event data is a convenient and powerful combination.
- Binning might not be such a great idea.
- We can extend SAOMs to include a relational event mode and use an observation function.
- In our study, inferential results are very similar for friendship and email data.

Carter T. Butts. A relational event framework for social action. *Sociological Methodology*, 381(1):155–200, 2008.

T.A.B. Snijders. The statistical evaluation of social network dynamics. In M.E. Sobel and M.P. Becker, editors, *Sociological Methodology*, pages 361–395. Basil Blackwell, Boston and London, 2001.

T.A.B. Snijders, J. Koskinen, and Michael Schweinberger. Maximum likelihood estimation for social network dynamics. *Annals of Applied Statistics*, page In press, 2010.