Continuous-Time Regression Models for Longitudinal Networks

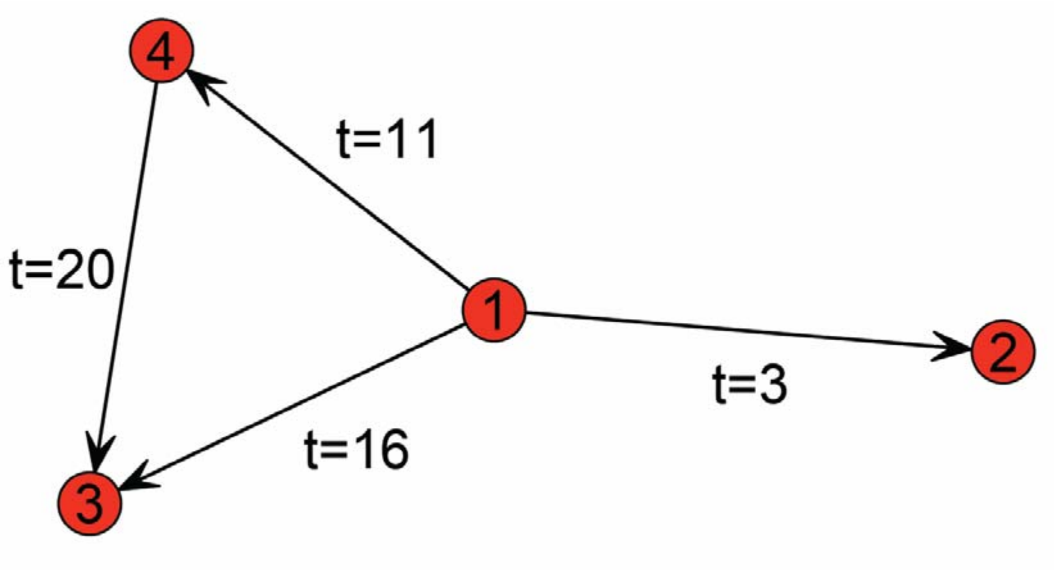
Duy Vu[1]
Arthur Asuncion[2,3]
David Hunter[1]
Padhraic Smyth[2]

[1]Dept. of Statistics, Penn State
[2]Dept. of Computer Science, UC Irvine
[3]Current Affiliation: Google, Inc.
Supported by ONR MURI Award N00014-08-1-1015

Neural Information Processing Systems, December 15, 2011

Motivation

- Statistical analysis of evolving network effects
- Prediction of future edge events
- Relatively little work on continuous-time network modeling



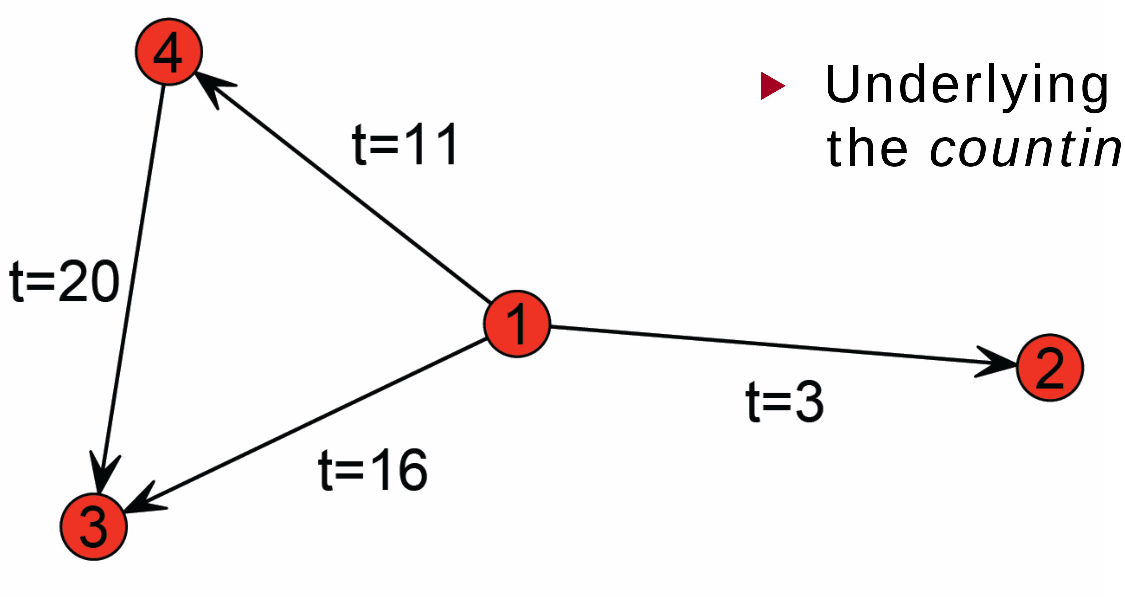
Outline

- Counting processes for evolving networks
  - Egocentric vs. Relational approach
  - Doob-Meyer decomposition
- Modeling the intensity process
  - Cox proportional hazards model
  - Aalen additive model with time-varying coefficients
- Inference methods and evaluation
  - Partial likelihood, least squares, and caching tricks
  - Predictive performance on network data

Outline

- Counting processes for evolving networks
  - Egocentric vs. Relational approach
  - Doob-Meyer decomposition
- Modeling the intensity process
  - Cox proportional hazards model
  - Aalen additive model with time-varying coefficients
- Inference methods and evaluation
  - Partial likelihood, least squares, and caching tricks
  - Predictive performance on network data

Counting Processes for Networks



- Underlying modeling mechanism: the *counting process*.

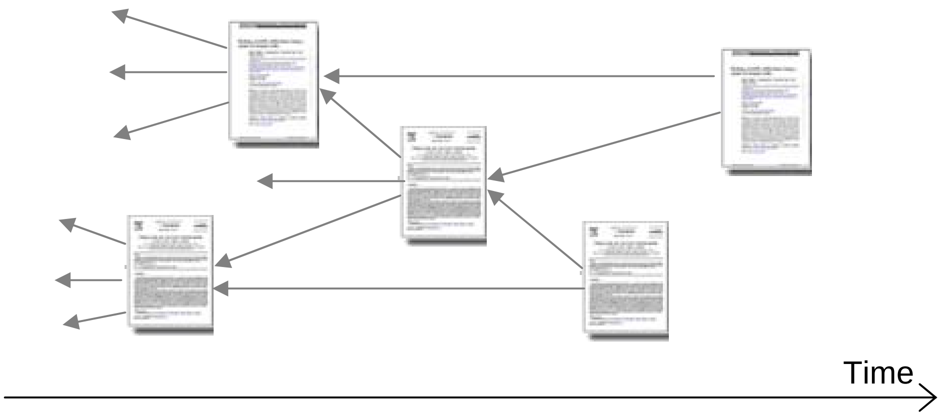
Two possible approaches (using terminology of Butts, 2008):

- **Egocentric:** The counting process $N_i(t)$ = cumulative number of "events" involving the ith node by time t.
- **Relational:** The counting process $N_{ij}(t)$ = cumulative number of "events" involving the (i,j)th node pair by time t.

Egocentric Example: Citation Networks

[Vu et al, ICML 2011]

- New papers join the network over time.
- At arrival, a paper cites others already in the network.

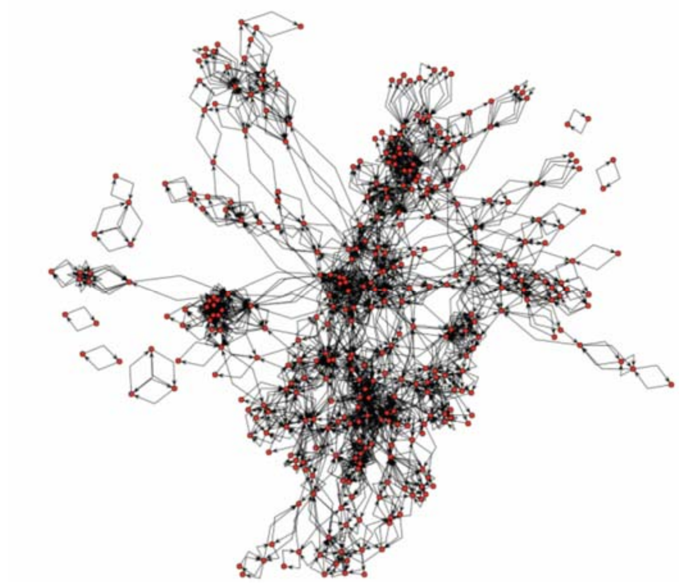


- Main dynamic development: Number of citations *received*.
- $N_i(t)$: Number of citations to paper i at time t.

Relational Example: Social Networks

- MetaFilter: Community weblog for sharing links and discussing content among users.
- Pattern of contacts: dynamically evolving network.
- Links are *non-recurrent*: $N_{ij}(t)$ is either 0 or 1.

contacter	contactee	date
1	14155	2004-06-15 12:00:00.000
1	2238	2004-06-15 12:00:00.000
1	14275	2004-06-15 12:00:00.000
...		
13099	7683	2004-06-17 16:31:51.040
15231	14752	2004-06-17 16:31:51.040
...		
45087	7610	2007-10-31 12:23:15.683
16719	61	2007-10-31 13:28:38.670
48758	1	2007-10-31 13:47:16.843

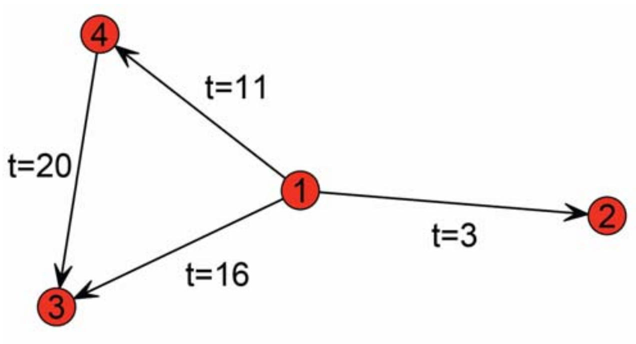


Continuous-Time Regression Models for Longitudinal Networks (Vu, Asuncion, Hunter, Smyth)

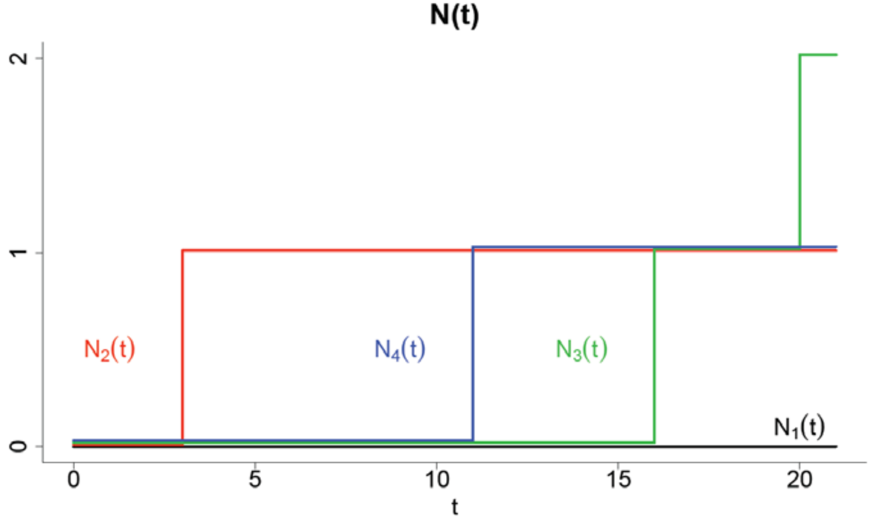
Multivariate Counting Process

- Combine the $N_i(t)$ to yield a multivariate counting process:

$$\mathbf{N}(t) = (N_1(t), \ldots, N_n(t))\,.$$



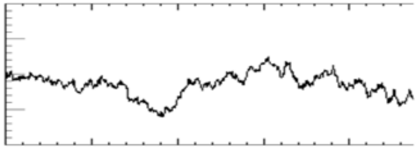
- Each $N_i(t)$ is nondecreasing in time, so **N**(t) may be considered a *submartingale.*



Doob-Meyer Decomposition

- Any submartingale may be uniquely decomposed in the following manner [Doob, 1953; Meyer, 1963]:

$$\underbrace{\mathbf{N}(t)}_{\text{Observed counts}} = \underbrace{\int_0^t \lambda(s)ds}_{\text{"Signal" (intensity process)}} + \underbrace{\mathbf{M}(t)}_{\text{"Noise" (Martingale)}}.$$



- Intensity function $\lambda_{ij}(t)$: instantaneous rate of event occurring between *(i,j)* pair at time *t*.

Outline

- Counting processes for evolving networks
  - Egocentric vs. Relational approach
  - Doob-Meyer decomposition
- Modeling the intensity process
  - Cox proportional hazards model
  - Aalen additive model with time-varying coefficients
- Inference methods and evaluation
  - Partial likelihood, least squares, and caching tricks
  - Predictive performance on network data

Modeling the Intensity Process

The intensity process for node pair *(i,j)* is given by

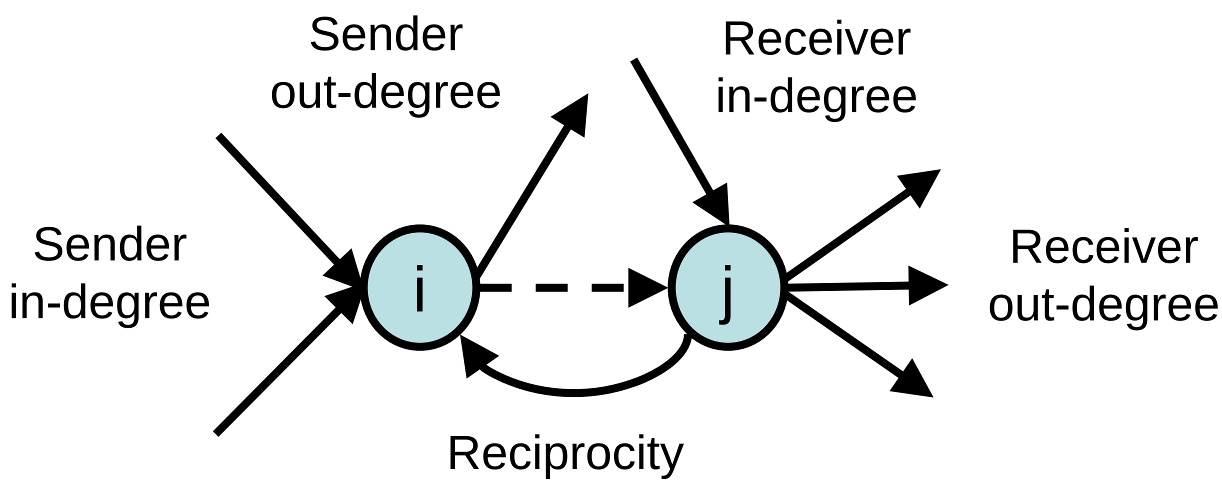
- Cox proportional hazards model [Cox, 1972]:

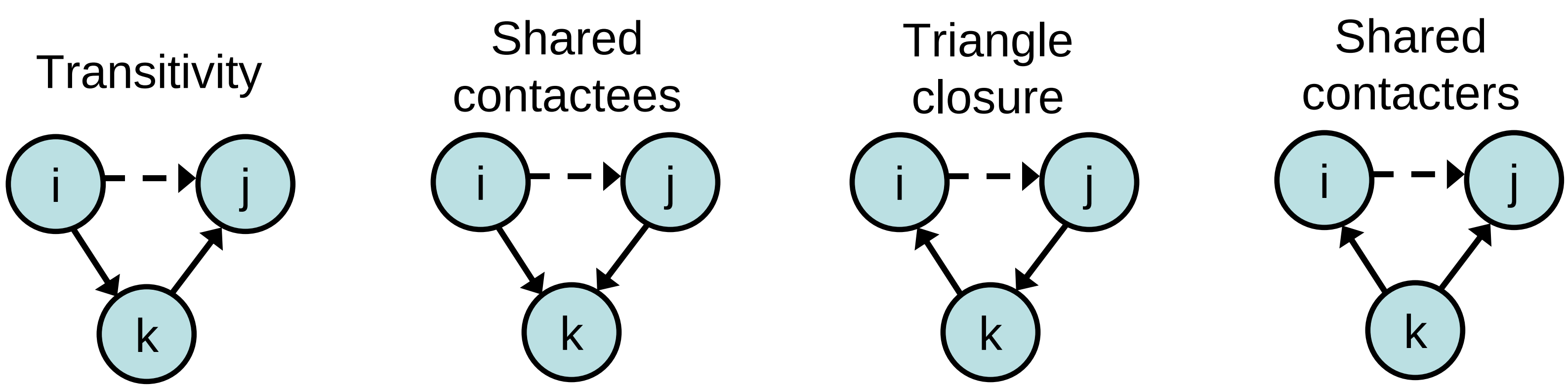
$$\lambda_{ij}(t \mid \mathbf{H}_{t-}) = R_{ij}(t)\alpha_0(t)\,\mathsf{exp}\left(\beta^{\mathsf{T}}\mathbf{s}_{ij}(t)\right).$$

- Aalen additive model [Aalen et al, 2008]:

$$\lambda_{ij}(t \mid \underbrace{\mathbf{H}_{t-}}_{\text{Past history of network up to time } t}) = \underbrace{R_{ij}(t)}_{\text{“At-risk” indicator}}\left(\underbrace{\beta_0(t)}_{\text{Baseline “hazard” function}} + \underbrace{\beta(t)^{\mathsf{T}}}_{\text{Coefficients to estimate}}\,\underbrace{\mathbf{s}_{ij}(t)}_{\text{Statistics for } (i,j) \text{ pair}}\right).$$

Network Statistics





Outline

- Counting processes for evolving networks
  - Egocentric vs. Relational approach
  - Doob-Meyer decomposition
- Modeling the intensity process
  - Cox proportional hazards model
  - Aalen additive model with time-varying coefficients
- Inference methods and evaluation
  - Partial likelihood, least squares, and caching tricks
  - Predictive performance on network data

Partial Likelihood (fitting the Cox model)

- We treat $\alpha_0(t)$ as a nuisance parameter and take a partial likelihood approach as in Cox, 1972. Maximize:

$$\mathcal{L}(\beta) = \prod_{e=1}^{m} \frac{\exp\left(\beta^\mathsf{T} \mathbf{s}_{i_e}(t_e)\right)}{\sum_{j=1}^{n} R_j(t_e) \exp\left(\beta^\mathsf{T} \mathbf{s}_j(t_e)\right)}$$

Partial Likelihood (fitting the Cox model)

- We treat $\alpha_0(t)$ as a nuisance parameter and take a partial likelihood approach as in Cox, 1972. Maximize:

$$\mathcal{L}(\beta) = \prod_{e=1}^{m} \frac{\exp\left(\beta^{\mathsf{T}} \mathbf{s}_{i_e}(t_e)\right)}{\sum_{j=1}^{n} R_j(t_e) \exp\left(\beta^{\mathsf{T}} \mathbf{s}_j(t_e)\right)} = \prod_{e=1}^{m} \frac{\exp\left(\beta^{\mathsf{T}} \mathbf{s}_{i_e}(t_e)\right)}{\kappa(t_e)}.$$

- Computational Trick: Write $\kappa(t_e) = \kappa(t_{e-1}) + \Delta\kappa(t_e)$ and optimize $\Delta\kappa(t_e)$ calculation (by using caching and exploiting sparsity).

Least Squares (fitting the Aalen model)

- We do inference not for β_k , but rather for their time-integrals:

$$B_k(t) = \int_0^t \beta_k(s)ds.$$

- Estimation is similar to least squares [Aalen, 2008]:

$$\hat{\mathbf{B}}(t) = \sum_{t_e \le t} J(t_e) \left[\mathbf{W}(t_e)^\top \mathbf{W}(t_e)\right]^{-1} \mathbf{W}(t_e)^\top \Delta \mathbf{N}(t_e).$$

Indicator that **W**(t) has full column rank

N(N-1) x p matrix with (i,j)th row $R_{ij}(t)\ s_{ij}(t)^\top$

N(N-1) vector denoting change in count from t_{e-1} to t_e

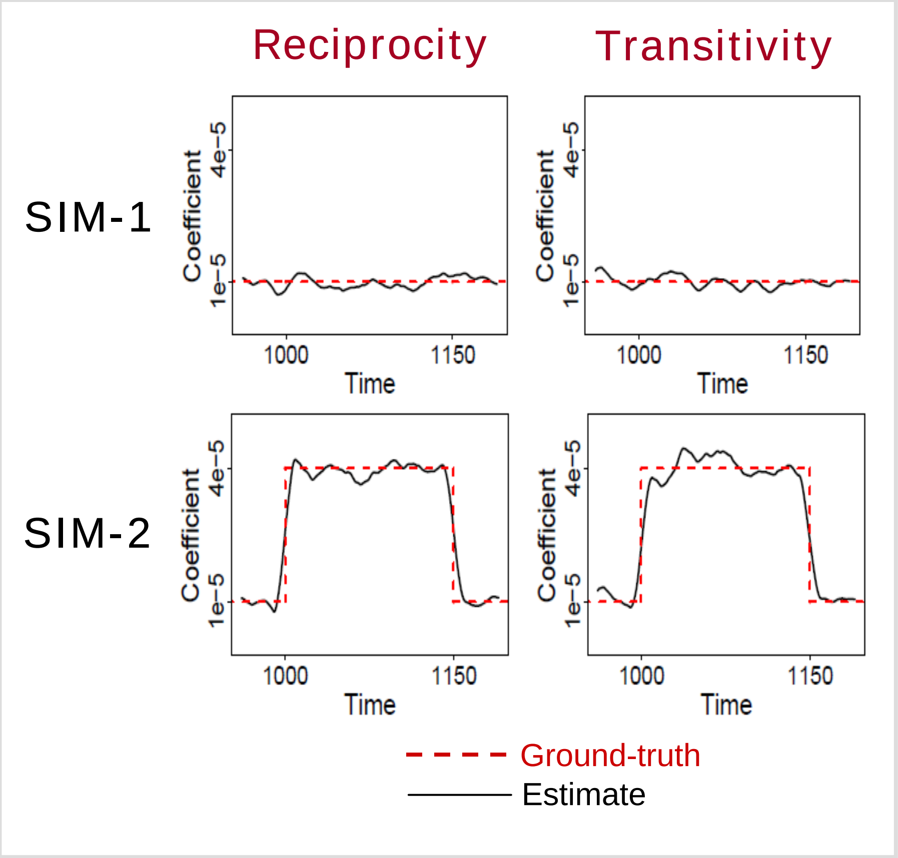
Network Data Sets

- Simulated data (SIM-1, SIM-2)
- Real networks:
  - **Irvine**: an online social network at UC Irvine (4/2004 to 10/2004).
  - **MetaFilter**: a community weblog contact network (8/2007 to 2/2011).

	Nodes	Edges	Stats-Building Phase	Training Phase	Test Phase
Irvine	1,899	20,296	7,073	7,646	5,507
MetaFilter	51,362	76,791	60,376	8,763	7,620

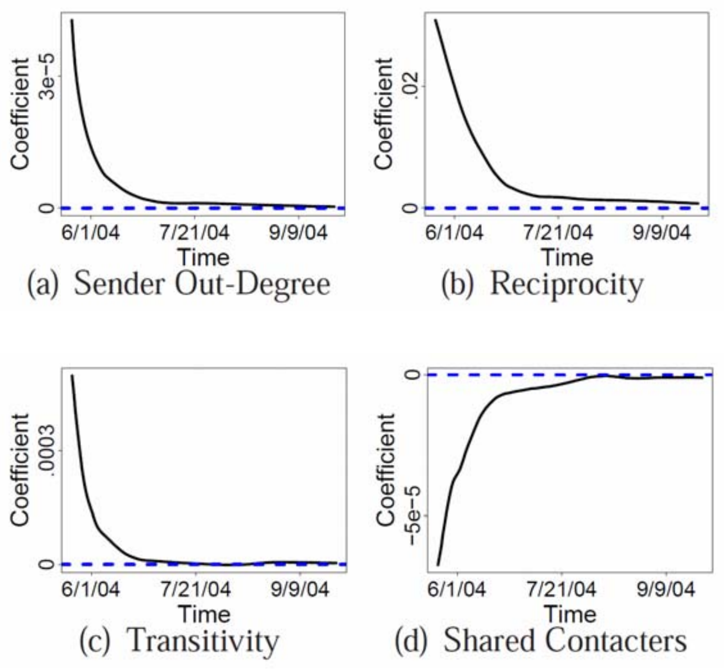
Recovering time-varying coefficients

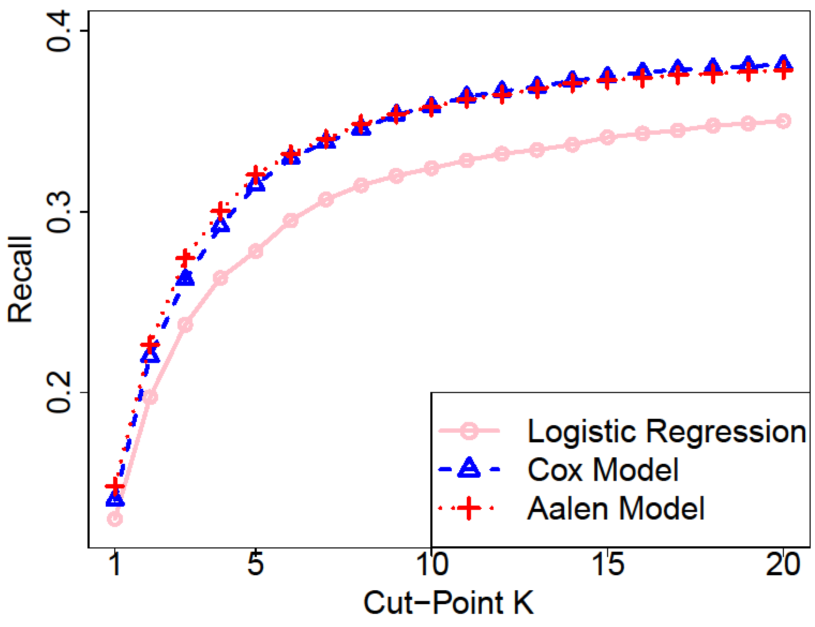
- We simulated data from ground-truth coefficients:
  - **SIM-1**: Constant coefficients for reciprocity, transitivity.
  - **SIM-2**: Varying coefficients for reciprocity, transitivity.
- We learned time-varying coefficients of Aalen model on simulated data.



Irvine data set

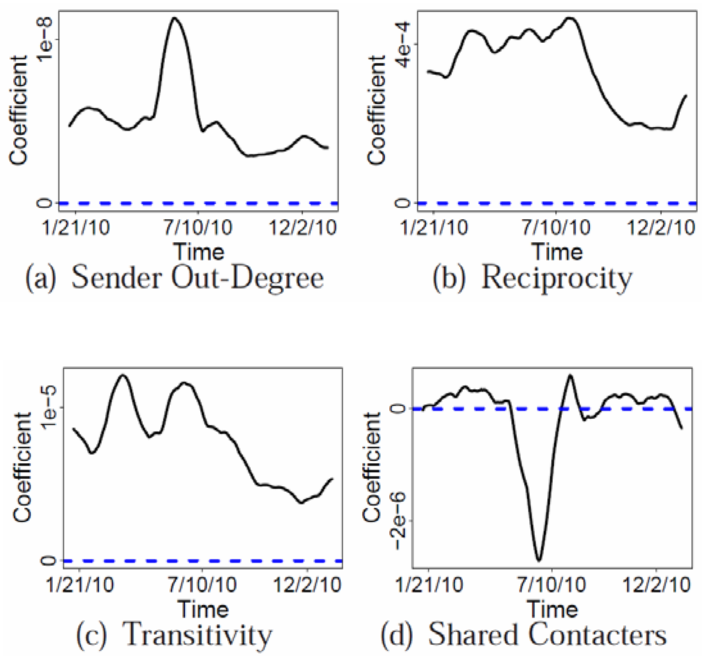
- Aalen coefficients suggest two distinct phases of network evolution, consistent with an independent analysis [Panzarasa et al, 2009].
- On prediction experiments, Aalen/Cox outperforms logistic regression.





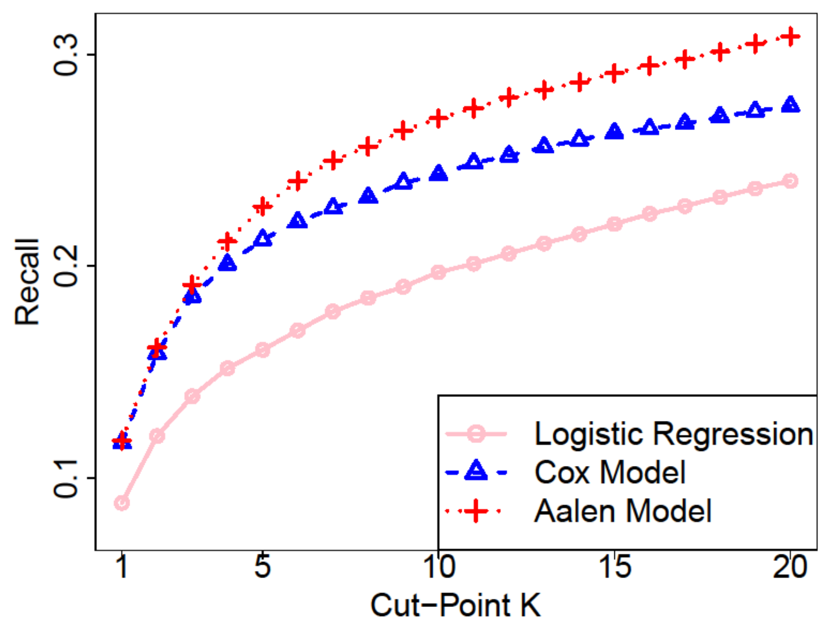
MetaFilter data set

- Network effects continuously change over time.
- Time-varying Aalen model outperforms Cox model.



(a) Sender Out-Degree (b) Reciprocity

(c) Transitivity (d) Shared Contacters



Summary

- Counting processes for evolving networks
  - Egocentric vs. Relational approach
  - Doob-Meyer decomposition
- Modeling the intensity process
  - Cox proportional hazards model
  - Aalen additive model with time-varying coefficients
- Inference methods and evaluation
  - Partial likelihood, least squares, and caching tricks
  - Predictive performance on network data

References

- Aalen, O.O., Borgan, O., and Gjessing, H.K.
  Survival and Event History Analysis: A Process Point of View
  Springer, 2008.

- Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N.
  Statistical Models Based on Counting Processes
  Springer, 1993.

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

- Cox, D.R.
  Regression models and life-tables
  Journal of the Royal Statistical Society, Series B, 34:187-220, 1972.

- Martinussen, T. and Sheike, T.H.
  Dynamic Regression Models for Survival Data
  Springer, 2006.

- Vu, D.Q., Asuncion, A.U., Hunter, D.R., and Smyth, P.
  Dynamic Egocentric Models for Citation Networks
  Proceedings of the 28th International Conference on Machine Learning, 857-864, 2011.