Time-Varying Networks

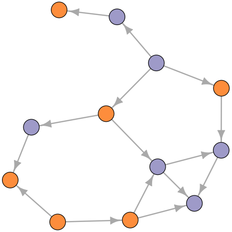
PJ Wolfe

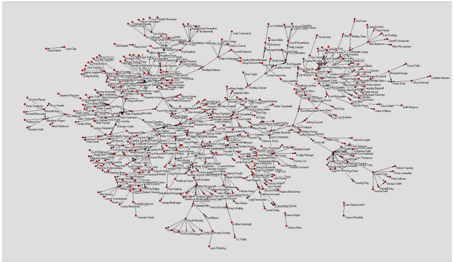
Chair of Statistics
Royal Society Research Fellow
University College London

IISc Workshop on High Dimensional Network Analytics
Bangalore, India, 18 December 2013

Prelude

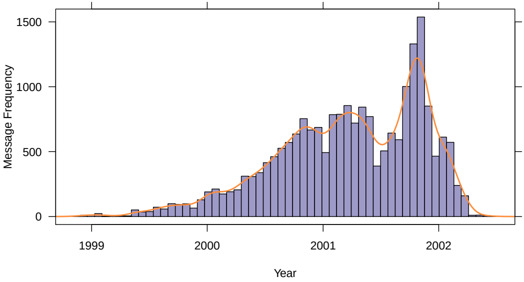
Motivation





- Networks represent high-dimensional yet sparse structure
- Information flows often come via repeated interactions
- Point processes are simple, flexible, and useful in this setting
- Example: analysis of a corporate e-mail data set

Point Process Approach



- Interaction data are often summarized as counts X_{ij}
- 'New social media', online messaging, transactions, etc. . .
- Interactions may have single or multiple receivers

Fitting a Simple Model

- Suppose we assume constant-rate Poisson ‘send’ processes, & constant-rate selection of a single receiver for each message
- Directed graph on N nodes $\Rightarrow 2N$ node-specific parameters
- Tabulate observed counts (X_{i+}, X_{+j}) per sender/receiver to fit

Theorem (Perry & W, 2011, Case of Binary, Undirected Data)

For $X_{ij} \in \{0,1\}$, let $\mathbb{P}\{X_{ij} = 1\} = p_{ij}$, and $\log p_{ij} = \alpha_i + \alpha_j$.

If all degrees X_{i+} satisfy $1 \leq X_{i+}^2 \leq \varepsilon_0 X_{++}$, with $\varepsilon_0 \leq 15^{-2}$, there exists a monotone transformation $\hat{\alpha}$ of the network degrees solving the likelihood equation such that $\|\hat{\alpha} - \hat{\alpha}_{\mathrm{ML}}\|_\infty \leq 10\,\varepsilon_0$.

Introduction

A Corporate Email Network

The Enron corpus: a large collection of email messages sent within the company between November 1998 and June 2002

21,635 messages
156 employees

A Typical Email Message

```
Message-ID:
<7303996.1075860726914.JavaMail.evans@thyme>
Date: Wed, 10 Oct 2001 08:51:16 -0700 (PDT)
From: kenneth.lay@enron.com
To: benjamin.rogers@enron.com
Subject: RE: Power Trading Group
Mime-Version: 1.0
Content-Type: text/plain; charset=us-ascii
Content-Transfer-Encoding: 7bit

Ben -

I likewise was glad to see you.  Sorry we didn't
have a chance to talk.

Good to hear you're doing well.  You're with a great
group and, yes, the company will soon be doing a lot
better.

Thanks,

Ken
```

Homophily in the Network?

- **Question:** Is group membership predictive of interaction?
  - Gender, Department, Seniority
- **Answer(?):** Contingency table analysis, homogeneity assumptions are violated:
  - Dependence, Time variation, Multi-way interactions

Other questions: Are past interactions predictive of future ones? Does this effect vary over time? How should multiple-receiver interactions be handled? Can these be treated as multiple pairwise interactions? . . .

Contingency Table Analysis

	Legal Jr	Legal Sr	Trading Jr	Trading Sr	Other Jr	Other Sr
Legal Jr	-0.07	2.8	-1.91	2.88	-0.3	-0.4
Legal Sr	1.39	0.3	2.58	-0.15	-1.0	0.9
Trading Jr	-0.15	-Inf	-0.76	1.05	1.3	2.3
Trading Sr	4.41	0.6	0.28	-0.07	0.7	0.1
Other Jr	0.36	0.9	-0.01	1.44	1.0	-1.3
Other Sr	0.82	1.6	1.23	-0.30	-0.4	0.6

	Legal Jr	Legal Sr	Trading Jr	Trading Sr	Other Jr	Other Sr
Legal Jr	""	"++++"	"----"	"++++"	""	""
Legal Sr	"++++"	"+++"	"++"	""	""	"++++"
Trading Jr	""	"---"	"---"	"+++"	"++"	""
Trading Sr	"++++"	""	""	""	"+++"	""
Other Jr	""	""	""	"++++"	"++++"	"----"
Other Sr	"+"	"++++"	"++"	""	"-"	"++++"

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

- Positive log-odds indicates homophily ('birds of a feather')
- Fisher's exact test yields significance levels
- Validity?

Dependence

```
Date: Wed, November 7, 2001 8:34 AM
From: Webb, Jay
To: Kitchen, Louise
Subject: Fw: 8:30 am trade count
Hi Louise,

We are having a typical trading pace so far today.  It is too early to tell
if any counterparty is really cutting back.  Like yesterday, however,
Aquila is buying longer dated physical gas and selling spot gas...
```

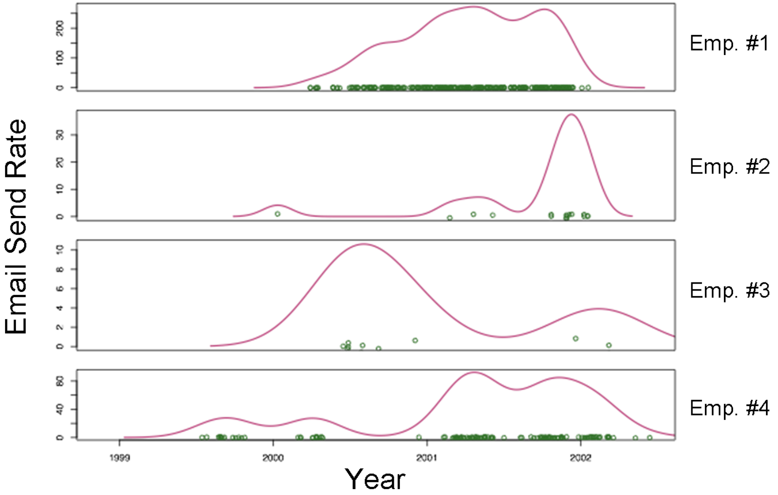
```
Date: Wed, November 7, 2001 10:14 AM
From: Kitchen, Louise
To: Arnold, John; Shilvey, Hunder; Neil, Scott; Martin, Tom; Grigsby, Mike
Subject: Fw: 8:30 am trade count


Note aquila.
```

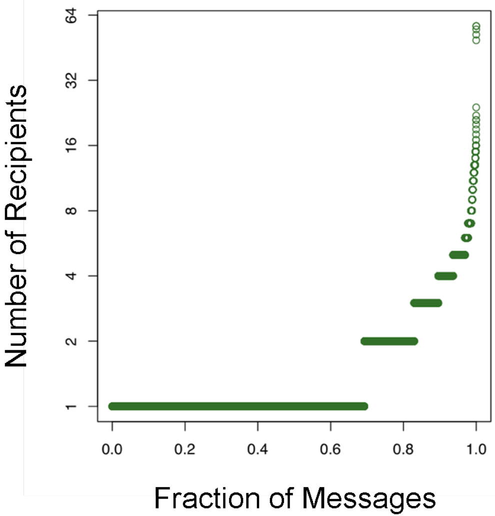
```
Date: Wed, November 7 2001 8:19 AM
From: Arnold, John
To: Kitchen, Loise; Webb, Jay
Subject: RE: 8:30 am trade count

fyi : Having more and more counterparties that will only deal on one side
of my market.
```

Varying Rates



Multiple Recipients



Modeling

Proportional Intensity Model

Model pairwise interactions $i \to j$ via stochastic intensity $\lambda_t(i,j)$:

$$\lambda_t(i,j)\,dt = \Pr\{\text{interaction } i \to j \text{ occurs in time } [t, t+dt)\}.$$

Sender i interacts with receiver j at a baseline rate $\bar{\lambda}_t(i)$ modulated up or down according to the pair's covariate vector, $x_t(i,j)$:

$$\lambda_t(i,j) = \bar{\lambda}_t(i) \cdot \exp\{\beta_0^T x_t(i,j)\} \cdot 1\{j \in \mathcal{J}_t(i)\}.$$

- $\mathcal{J}_t(i)$ is the receiver set of sender i at time t
- $\bar{\lambda}_t(i)$ denotes the baseline intensity of sender i
- $x_t(i,j) \in \mathbb{R}^p$ comprises covariates; coefficient vector β_0

Covariate Possibilities

- Group-Level Covariates: same gender, dept, seniority...

$$1\{i \text{ and } j \text{ belong to the same group}\}$$

- Network Covariates: received from j last hour, day, week...

$$1\{\text{interaction } j \to i \text{ occurred in } [t - \delta_l, t)\}$$

Any process depending only on the past is a valid covariate; e.g.,

$$1\{\text{for some } k, \text{ interactions } i \to k \text{ and } k \to j \text{ occurred in } [t - \delta_l, t)\}$$

Treat $\bar{\lambda}_t(i)$ as a nuisance parameter (Cox's partial likelihood):

a) Log partial likelihood at time t, evaluated at β:

$$\log PL_t(\beta) = \sum_{t_m \le t} \left\{ \beta^T x_{t_m}(i_m, j_m) - \log \left[\sum_{j \in \mathcal{J}_{t_m}(i_m)} \exp\{\beta^T x_{t_m}(i_m, j)\} \right] \right\}$$

b) Approximate "multicast" likelihood:

$$\log \widetilde{PL}_t(\beta) = \sum_{t_m \le t} \left\{ \sum_{j \in J_m} \beta^T x_{t_m}(i_m, j) - |J_m| \log \left[\sum_{j \in \mathcal{J}_{t_m}(i_m)} \exp\{\beta^T x_{t_m}(i_m, j)\} \right] \right\}$$

NB: Maximizing $\log \widetilde{PL}_t(\cdot)$ instead of $\log PL_t(\cdot)$ introduces bias

Asymptotics

Different asymptotic regime than traditional proportional hazards

For pairwise interactions, under suitable regularity conditions:

Theorem (Perry & W, 2010)

As the number n of interactions grows,

i) The maximum likelihood estimator $\hat{\beta}_n$ of β_0 is consistent; i.e., it converges in probability to β_0;

ii) The quantity $\sqrt{n}\,(\hat{\beta}_n - \beta_0)$ converges in distribution to a zero-mean Normal random variable whose covariance can also be consistently estimated.

Results also extend to the case of multiple recipients (more work)

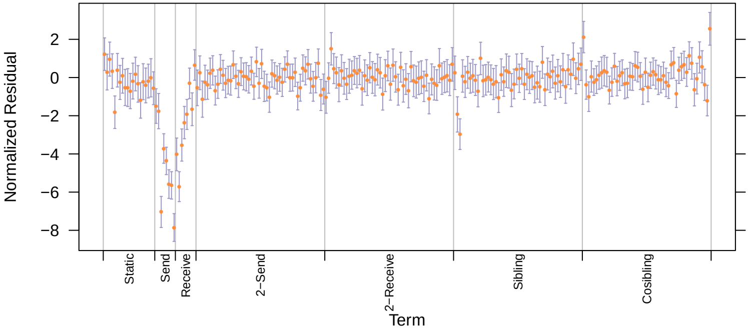
Results

Goodness of Fit

Term	Df	Deviance	Resid. Df	Resid. Dev
Null			32261	325412
Static	20	50365	32241	275047
Send	8	107942	32233	167105
Receive	8	5919	32225	161186
Sibling	50	3601	32175	157585
2-Send	50	516	32125	157069
Cosibling	50	1641	32075	155428
2-Receive	50	158	32025	155270

- Group-level (static) effects account for 15% of the residual deviance; network effects account for 37%
- Residual deviance is about 5$\times$ residual Df (overdispersion)

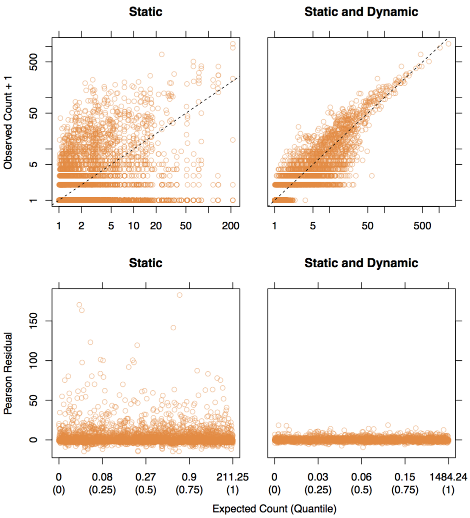
Multicast Bias Correction



Bootstrap residuals normalized by standard errors

- Note (correctable) negative bias in the coefficient estimates

Group-Level vs. Network Effects



Goodness of fit (Observed-vs.-expected, Pearson residuals)

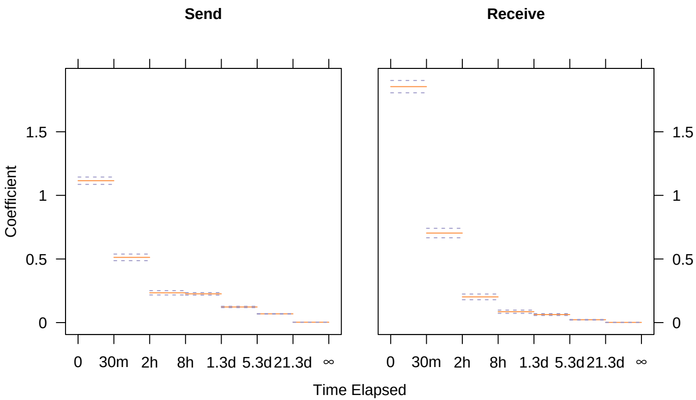
Group-Level Effects

	Receiver			
Sender	L	T	J	F
1	-0.91 (0.04)	-0.36 (0.04)	-0.34 (0.04)	0.04 (0.03)
L	0.63 (0.05)	0.28 (0.05)	0.22 (0.04)	0.15 (0.04)
T	0.32 (0.07)	0.43 (0.05)	0.27 (0.05)	-0.07 (0.05)
J	0.06 (0.05)	0.28 (0.04)	0.37 (0.03)	-0.13 (0.03)
F	0.59 (0.05)	-0.21 (0.05)	-0.09 (0.04)	0.15 (0.03)

Estimated group-level effects: Row(sender) $\cdot$ Col(receiver)

Network Effects

Variate	1{send}	1{rec}	1{2-send}	1{2-rec}	1{sibling}	1{cosib}
Coefficient	3.26	0.97	0.67	0.01	1.06	0.09
(SE)	(0.03)	(0.02)	(0.05)	(0.04)	(0.05)	(0.04)



Conclusion

Big Picture

- Flows over networks represented by repeated interactions
- Point process representation is simple, flexible, and useful
- Modeling message exchanges in a corporate e-mail network
  - Characteristics & behaviors predictive of interaction
  - Enables quantitative description of network effects

Further details: See "Point process modeling for directed interaction networks," J. Roy. Stat. Soc., B (arXiv:1011.1703). NSF-DMS/MSBS/CISE, DARPA, ONR, ARO MURI and PECASE support gratefully acknowledged