Characterizing Individual Behavior from Interaction History

Patrick Perry
NYU Stern

Case Study: UCI Online Network

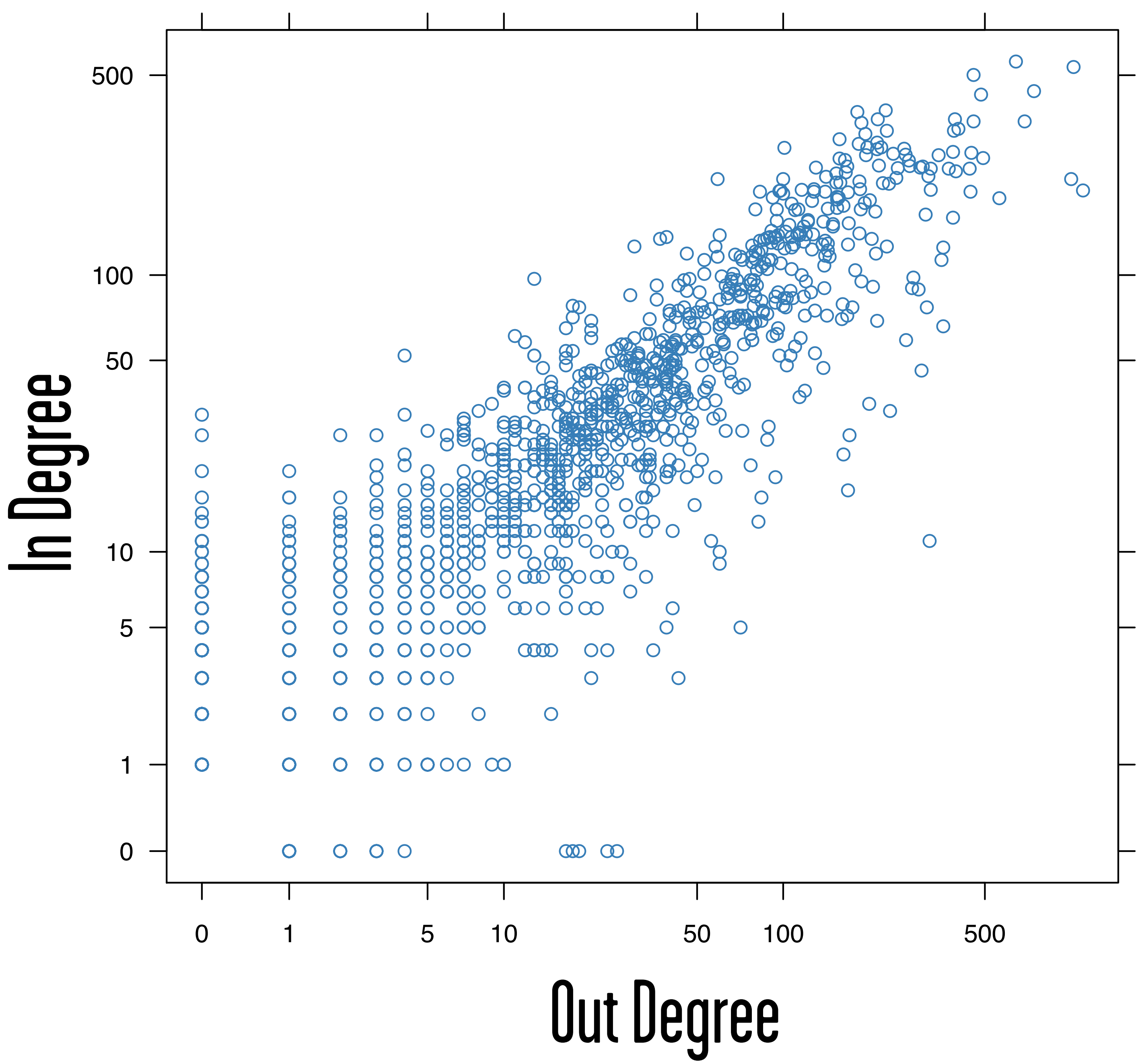
Online community for University of California, Irvine (Opsahl & Panzarasa, 2009)

Dataset covers seven-month period: April - October 2004

2000 users, 60K messages

Goal: Characterize user messaging behavior

Degrees Are Not Enough



Can we do better?

Agenda

1. Framework for studying interaction histories
2. Macroscopic behavior
3. Microscopic behavior

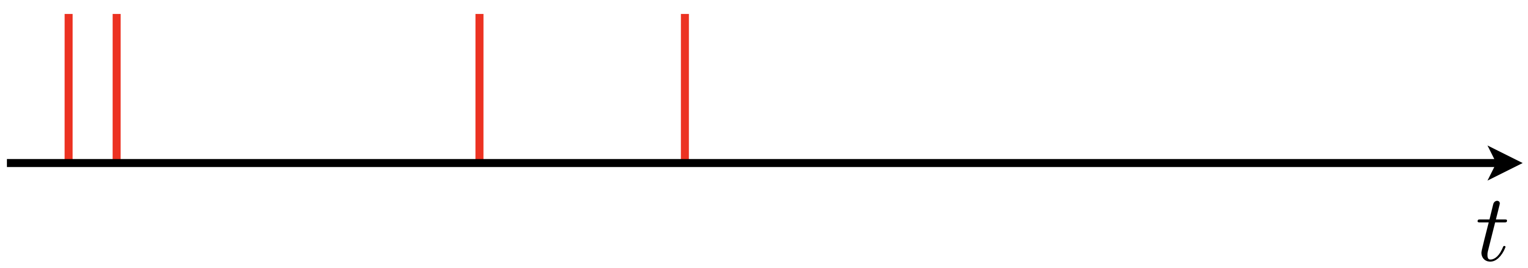
Events, Not Links

Messages

Time	Sender	Receiver
t_1	i_1	j_1
t_2	i_2	j_2
$\vdots$	$\vdots$	$\vdots$
t_n	i_n	j_n

Point Process Model

Messages from i to j :



Model via intensity, $\lambda_t(i,j)$:

$$\lambda_t(i,j)\,dt = \text{Prob}\{i \text{ sends to } j \text{ in } [t, t+dt)\}$$

Key Insight: Use Past History

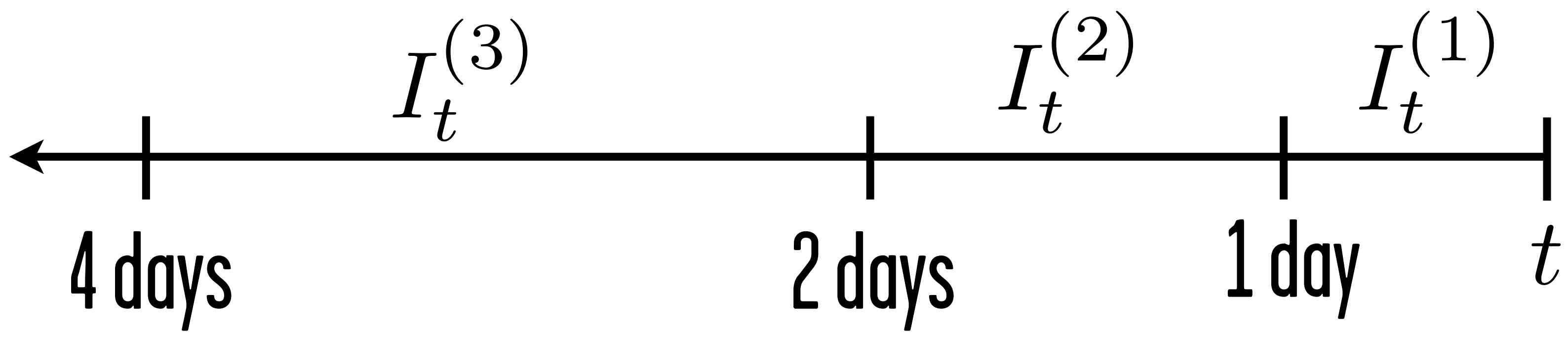
Hypotheses:

If you send me a message, I am likely to respond

If I have sent you a message in the past, I am likely to repeat this action in the future

These effects all decay with time.

History-Dependent Covariates



$$\mathbf{send}_t^{(k)}(i,j) = \#\{i \to j \text{ in } I_t^{(k)}\},$$
$$\mathbf{receive}_t^{(k)}(i,j) = \#\{j \to i \text{ in } I_t^{(k)}\};$$

Cox Proportional Intensity Model

$$\lambda_t(i,j) = \bar{\lambda}_t(i) \exp\{\beta^{\mathrm{T}} x_t(i,j)\}$$

$\lambda_t(i,j)\,dt$	Prob{i sends j a message in time [t,t+dt)}
$\bar{\lambda}_t(i)$	Baseline intensity for sender i
β	Vector of coefficients
$x_t(i,j)$	Vector of time-varying covariates

(Butts 2008 , Vu et al. 2011, POP & Wolfe 2013)

Interpretation

$$\lambda_t(i,j) = \bar{\lambda}_t(i) \exp\{\beta^{\mathrm{T}} x_t(i,j)\}$$

β_k Increasing $[x_t(i,j)]_k$ by one unit while holding all other covariates constant is associated with multiplying the message rate by e^{β_k} units.

$\bar{\lambda}_t(i)$ Treated as a nuisance parameter, estimated non-parametrically

Example: Self-Reinforcing Send

$$\lambda_t(i,j) = \bar{\lambda}_t(i) \exp\{1.8[x_t(i,j)]_1 + 0.7[x_t(i,j)]_2\}$$

$$[x_t(i,j)]_1 = \#\{i \to j \text{ in } [t - 1 \text{ day}, t)\}$$

$$[x_t(i,j)]_2 = \#\{i \to j \text{ in } [t - 1 \text{ week}, t - 1 \text{ day})\}$$

Every sent message is associated with an $e^{1.8}$-fold **increase** for 1 day, followed by an $e^{0.7}$-fold **increase** for 6 days (relative to the baseline).

After one week, the message is not associated with a change in rate

Example: Response Model

$$\lambda_t(i,j) = \bar{\lambda}_t(i) \exp\{1.8[x_t(i,j)]_1 - 0.3[x_t(i,j)]_2\}$$

$$[x_t(i,j)]_1 = \#\{j \to i \text{ in } [t - 1 \text{ day}, t)\}$$

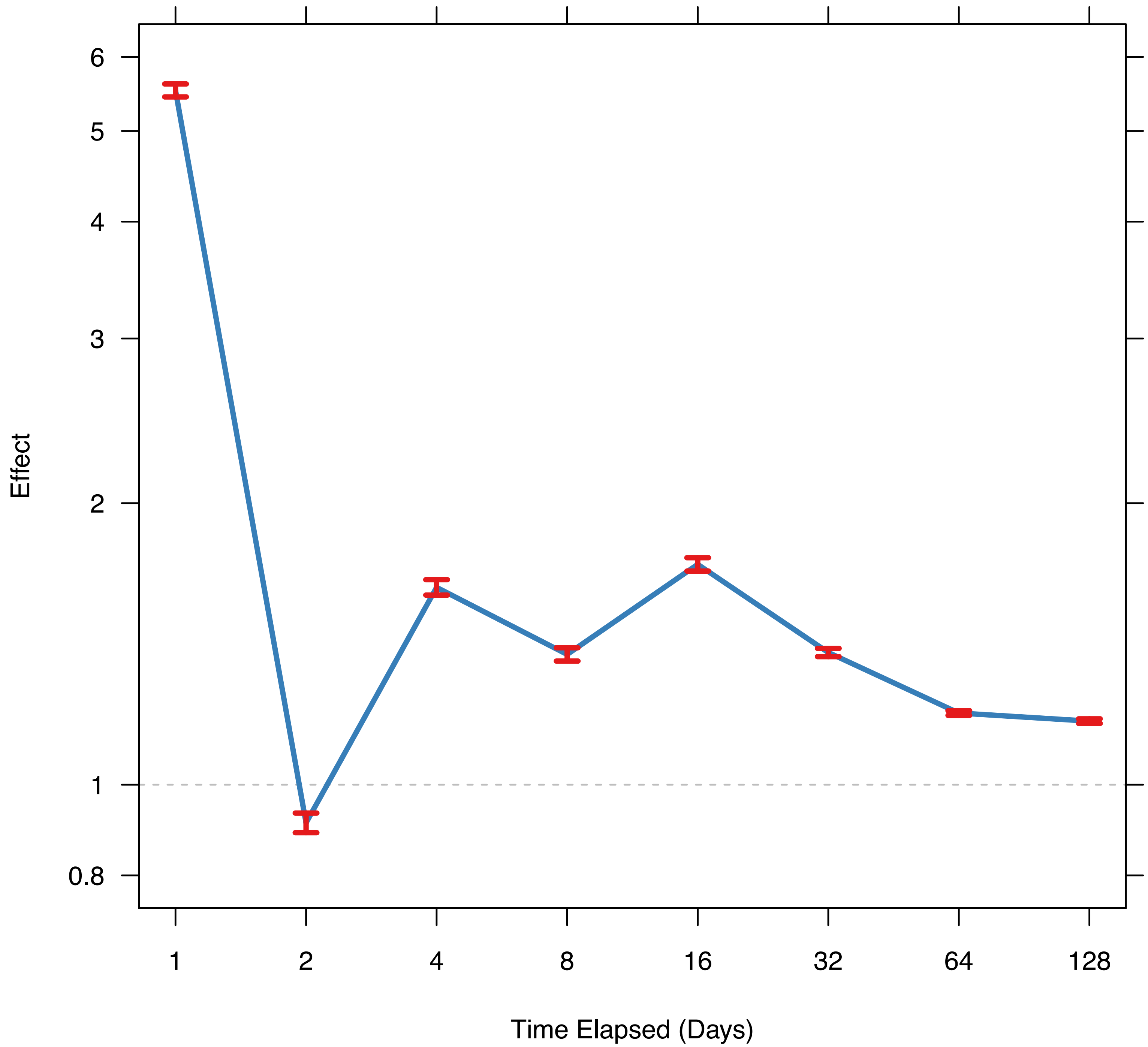
$$[x_t(i,j)]_2 = \#\{j \to i \text{ in } [t - 1 \text{ week}, t - 1 \text{ day})\}$$

Every received message is associated with an $e^{1.8}$-fold **increase** for 1 day, followed by an $e^{0.3}$-fold **decrease** for 6 days (relative to the baseline).

After one week, the message is not associated with a change in rate

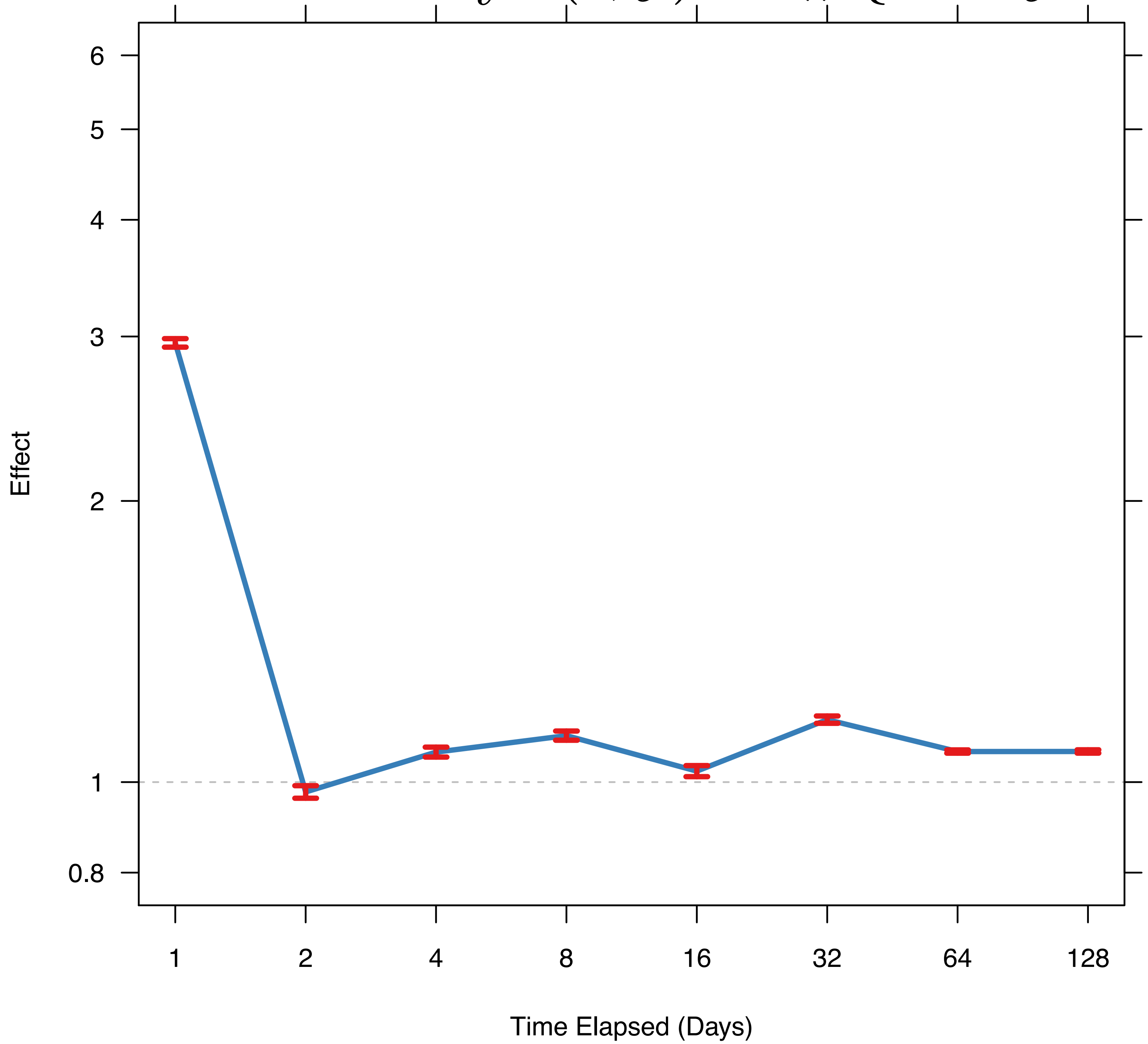
Users Respond to Messages

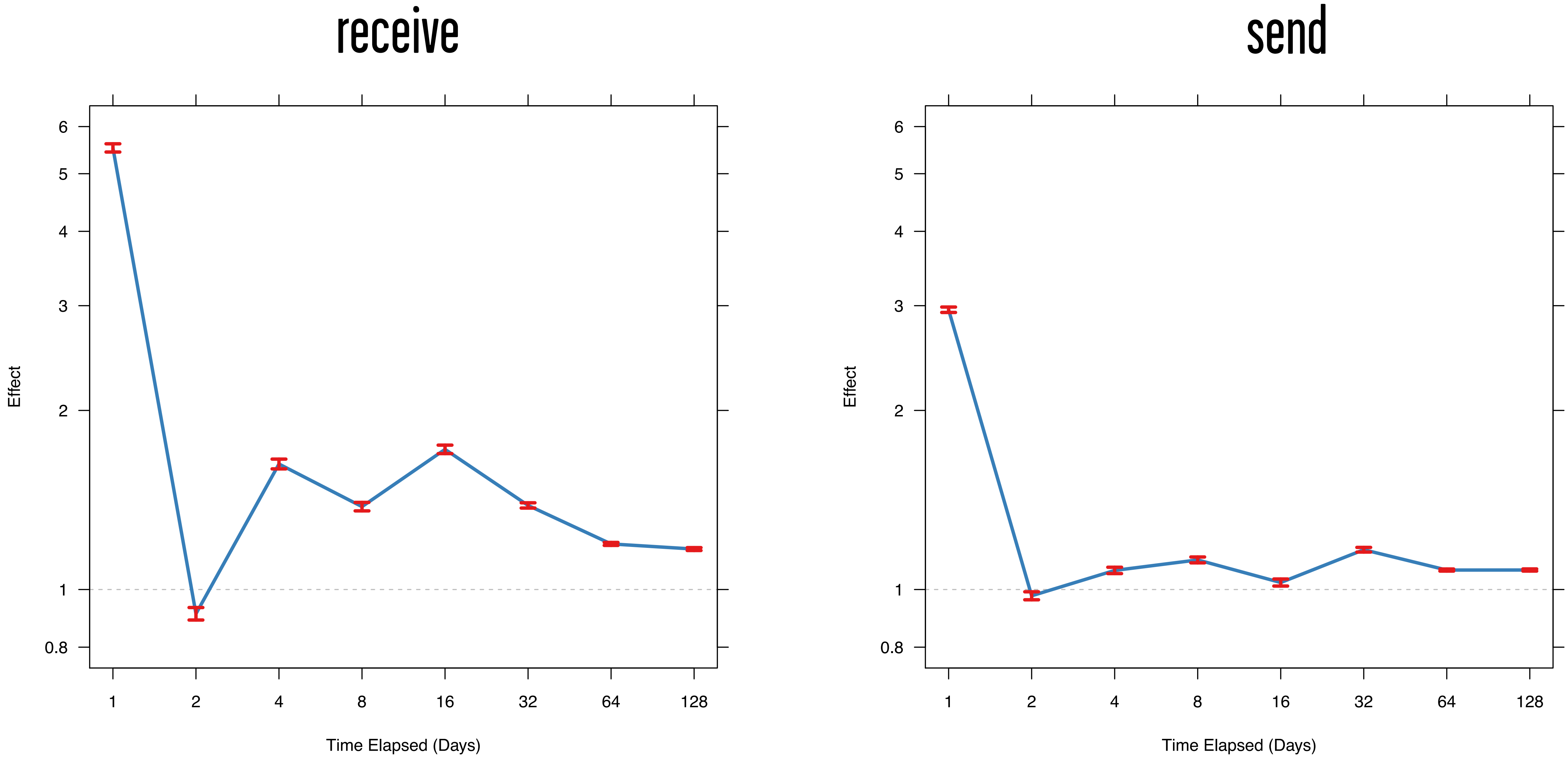
$$\text{Coefficient of } \mathbf{receive}_t^{(k)}(i,j) = \#\{j \to i \text{ in } I_t^{(k)}\}$$



Users Repeat Past Behavior

$$\text{Coefficient of } \mathbf{send}_t^{(k)}(i,j) = \#\{i \to j \text{ in } I_t^{(k)}\}$$





(1) receiving is associated with responding
(2) users repeat their past behaviors
(3) effect (2) decays faster than effect (1)

Same behavior for each user?

Micro-level Model

Old Model:

$$\lambda_t(i,j) = \bar{\lambda}_t(i) \exp\{\beta^{\mathrm{T}} x_t(i,j)\}$$

New Model:

$$\lambda_t(i,j) = \bar{\lambda}_t(i) \exp\{\beta_i^{\mathrm{T}} x_t(i,j)\}$$

$$\beta_i \sim \mathrm{Normal}(\mu, \Sigma)$$

(Related model: DuBois et al. 2013)

Estimating User-Specific Coefficients

Fitting time: 3 CPU hours

2000 sets of coefficients (one set for each user)

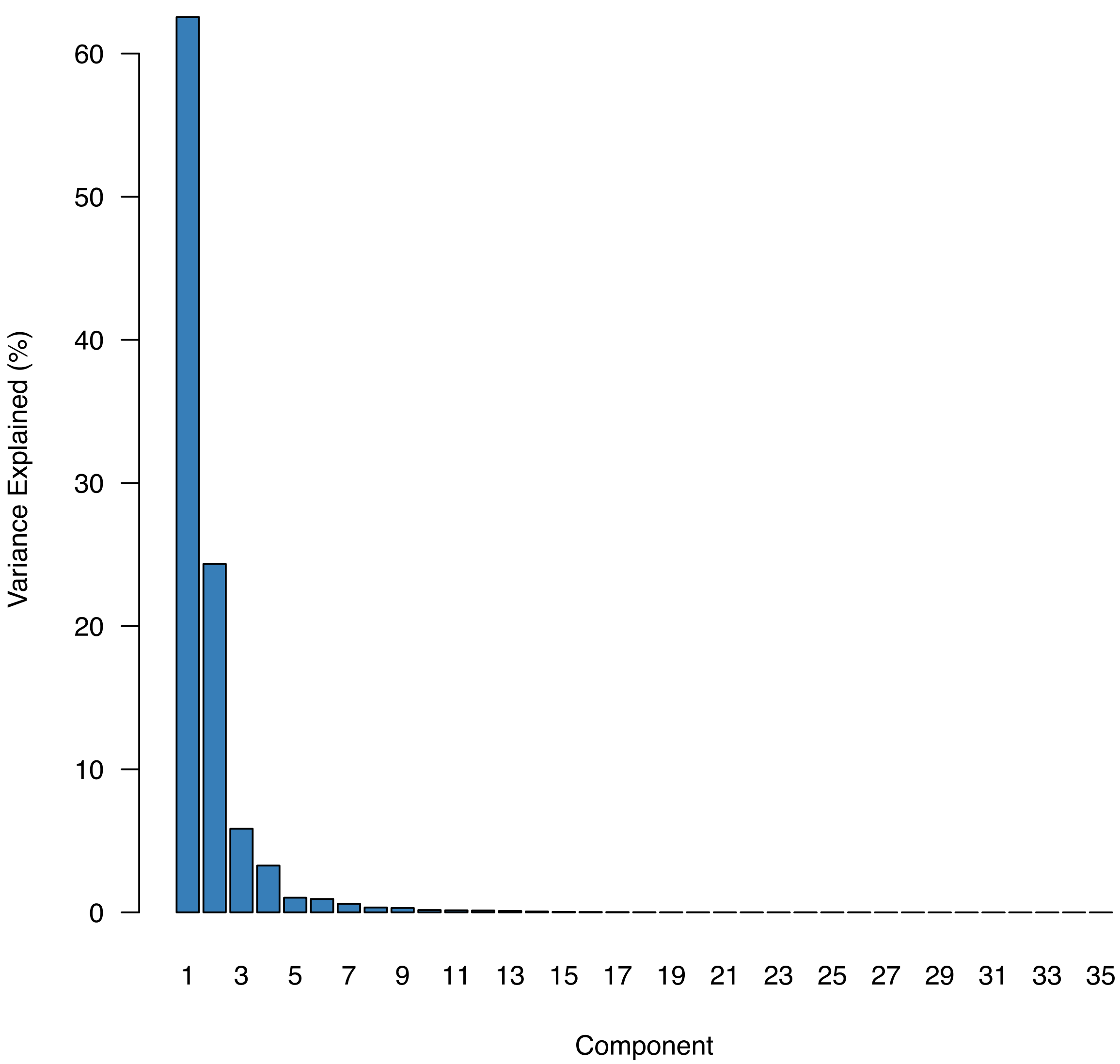
Need summarization method to visualize

Visualize by Factor Analysis

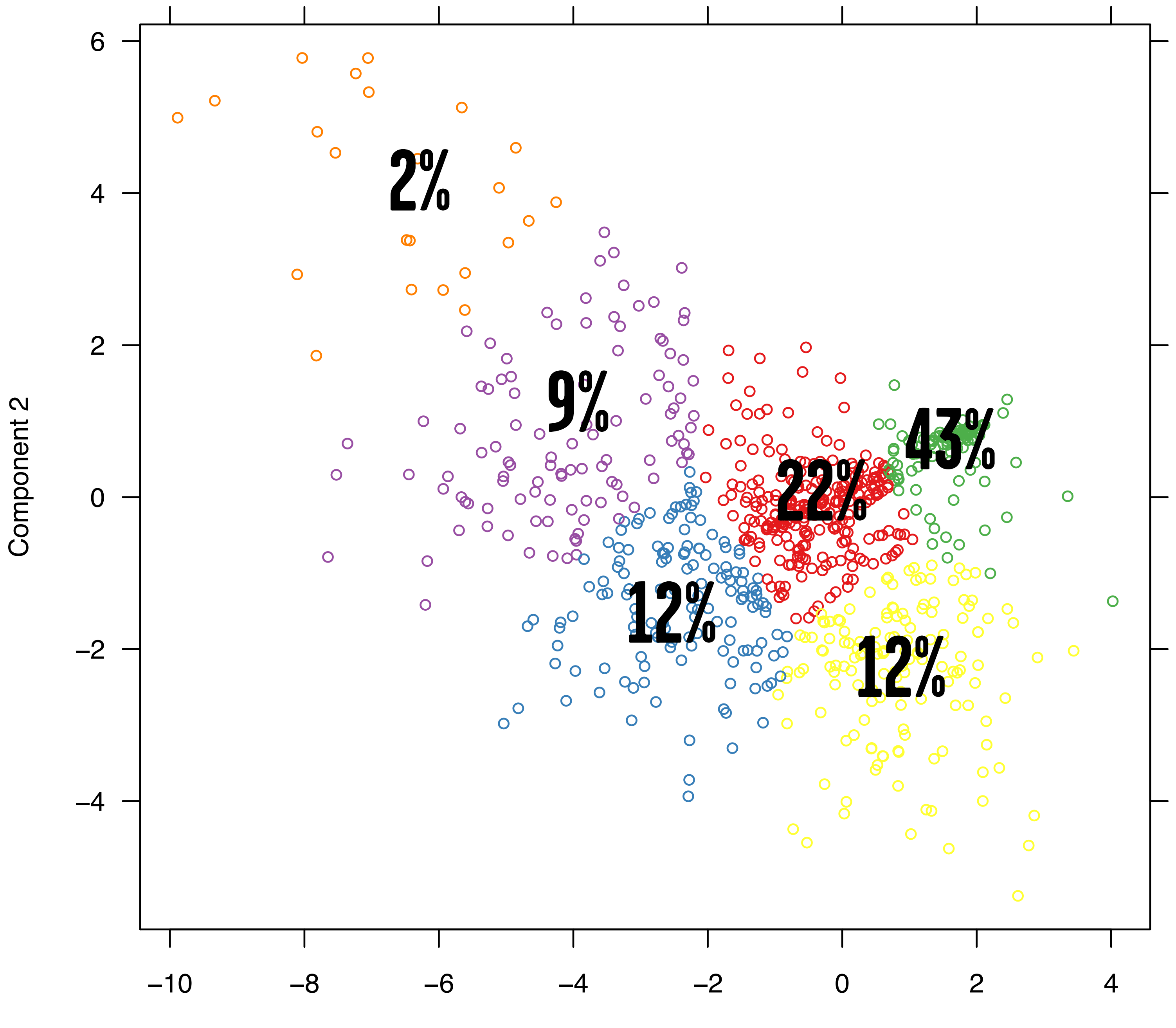
2000 sets of coefficients (one set for each user)

Reduce dimensionality via principle components

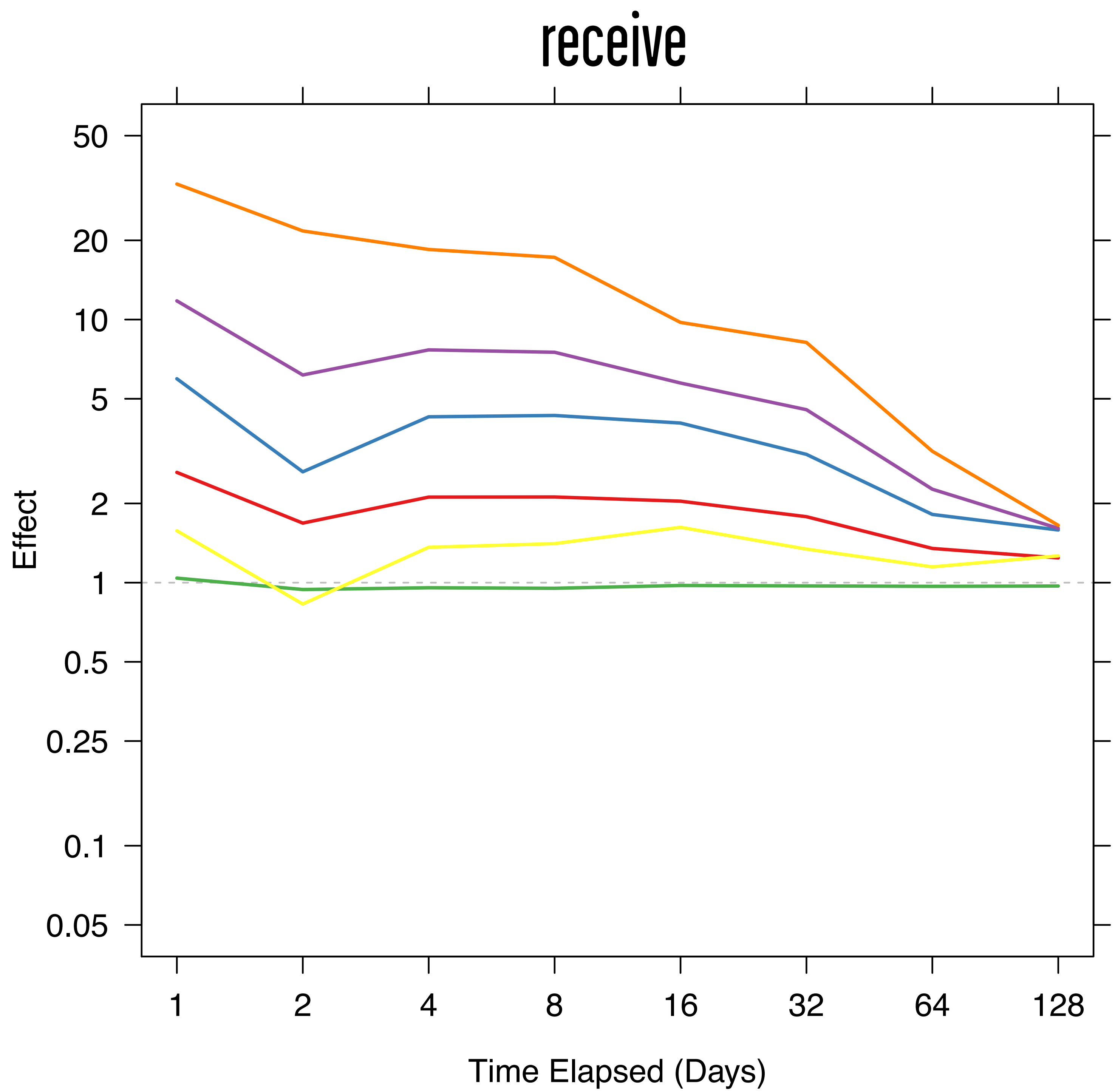
First 2 components explain 87% of variance



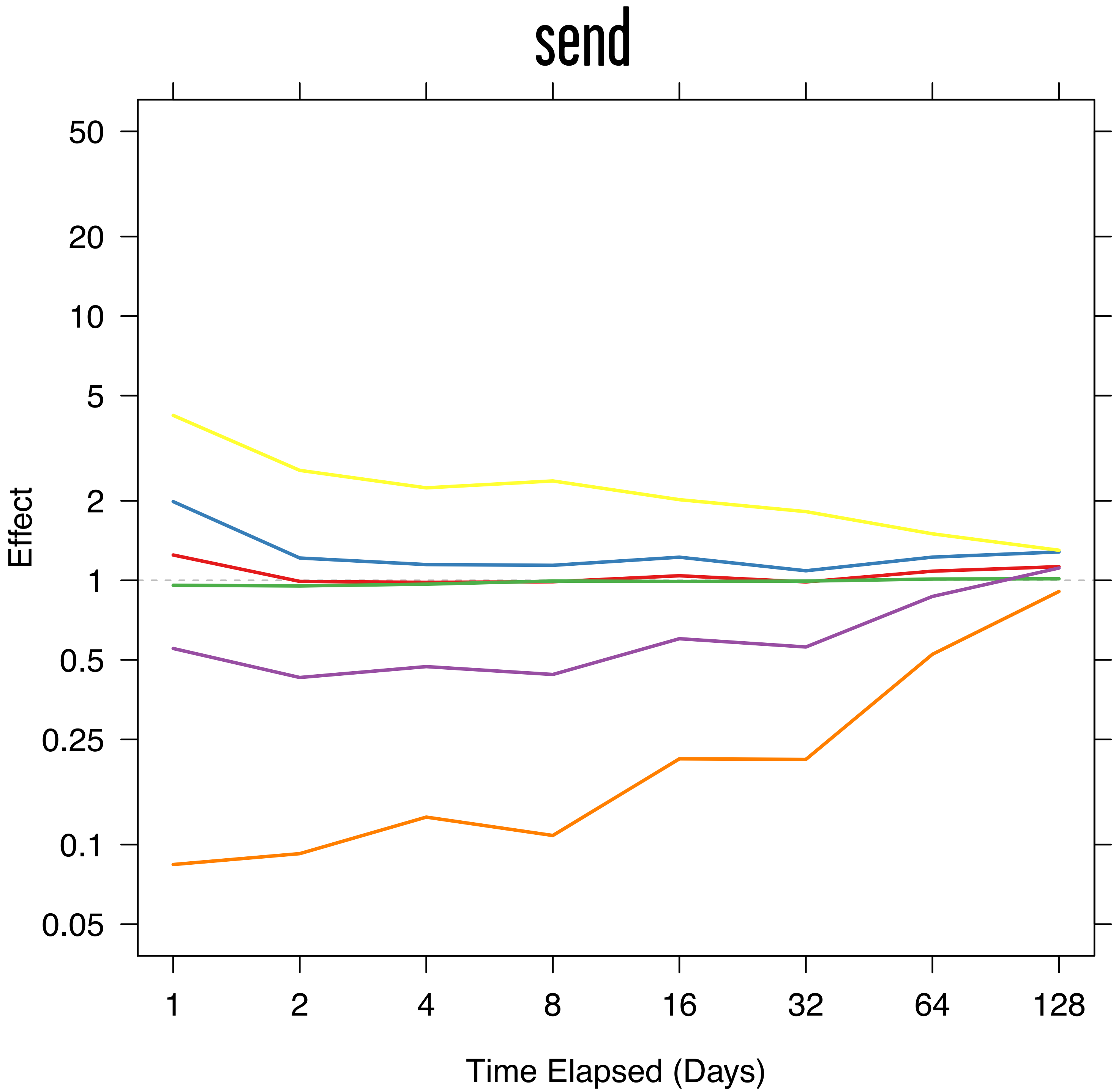
User-specific Principle Component Scores

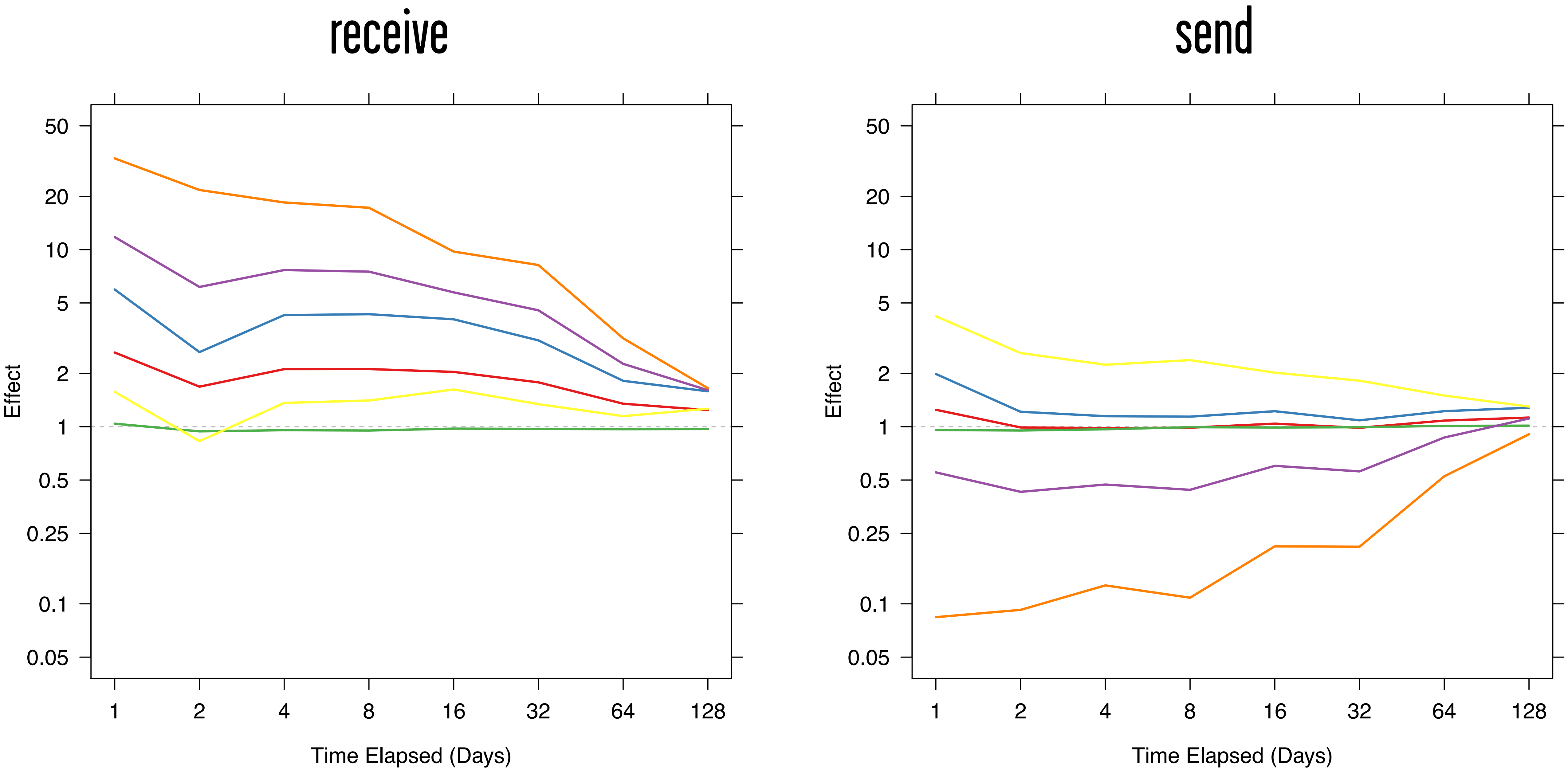


Variation in Response



Variation in Repetition





(1) two dimensions of behavior

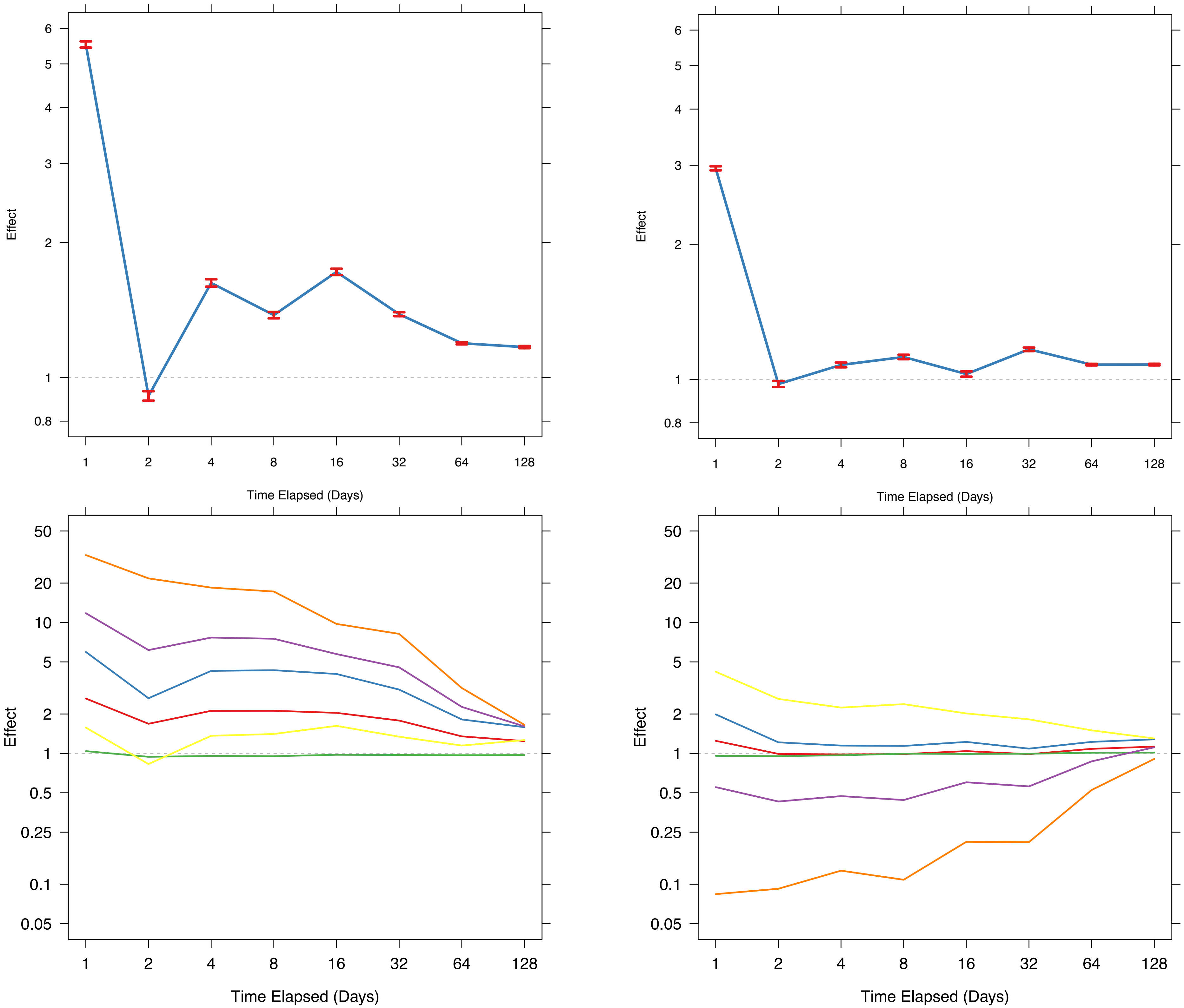
(2) large range of response rates, similar qualitative patterns

(3) some users repeat, others innovate; big effects in both directions

Comparing Macro and Micro

receive

send



Theory for Macro Case

Theorem (POP & Wolfe): Under regularity conditions, MPLE satisfies:

1. $$\hat{\beta}_n \xrightarrow{P} \beta$$
2. $$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \text{Normal}\big(0,\ \Sigma(\beta)\big)$$

Related results:

Cox (1975): heuristic argument ("under mild conditions implying some degree of independence… and that the information values are not too disparate")

Andersen & Gill (1982): survival analysis, fixed time interval

Implementation

$$PL_{t_n}(\beta) = \prod_{t_m \le t_n} \frac{e^{\beta^{\mathrm{T}} x_{t_m}(i_m, j_m)}}{\sum_j e^{\beta^{\mathrm{T}} x_{t_m}(i_m, j)}}$$

Loop over all messages

Loop over all receivers

Na ve: O(messages × receivers)

With bookkeeping: O(messages + receivers)

Implementation Trick: Sparsity

Inner sum: $$\sum_j e^{\beta^{\mathrm{T}} x_t(i,j)} = \sum_j e^{\beta^{\mathrm{T}} x_0(i,j)} + \left[\sum_j e^{\beta^{\mathrm{T}} x_t(i,j)} - e^{\beta^{\mathrm{T}} x_0(i,j)} \right]$$

Note! $$x_t(i,j) = x_0(i,j) + d_t(i,j)$$

Implementation Trick: Structure

Initial sum: $$\sum_j e^{\beta^{\mathrm{T}} x_0(i,j)}$$

Redundancy in $$\left\{ \left(x_0(i,1), x_0(i,2), \ldots, x_0(i,J) \right) \right\}_{i=1}^{I}$$

More Details

Computing $d_t(i, j)$

Self-loops

Similar tricks for gradient, Hessian

Numerical overflow

R package forthcoming

Summary

1. Events, not links
2. Point process model captures behavior
3. User-specific coefficients allow for heterogeneity