Logistic Network Regression for Scalable Analysis of Dynamic Relational Data

Zack W. Almquist and Carter T. Butts

Department of Sociology, University of California-Irvine

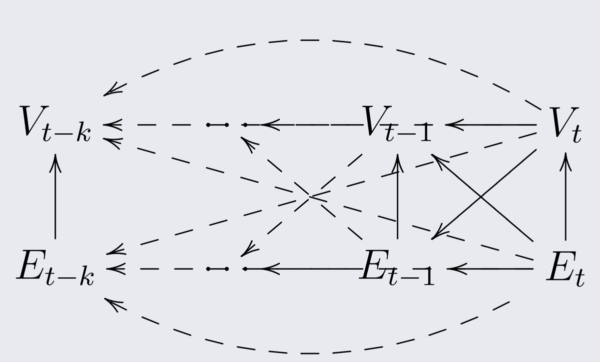
Abstract

Network dynamics may be viewed as a process of change in the edge structure of a network, in the vertex set on which edges are defined, or in both simultaneously. While early studies of such processes were primarily descriptive (e.g., Sampson, 1968), work on this topic in recent years has increasingly turned to formal statistical models (e.g., Snijders, 2001). While showing great promise, many of these modern dynamic models are computationally intensive and scale very poorly in the size of the network under study, making them difficult or impossible to apply to large networks in practical settings. Given this situation, there is a need for scalable approaches that – even if limited in various ways – can serve as a starting point for analysis of intertemporal network data at large scales. This paper explores the use of the well-known logistic network regression framework as a simple basis for the modeling of network dynamics with various orders of temporal dependence.

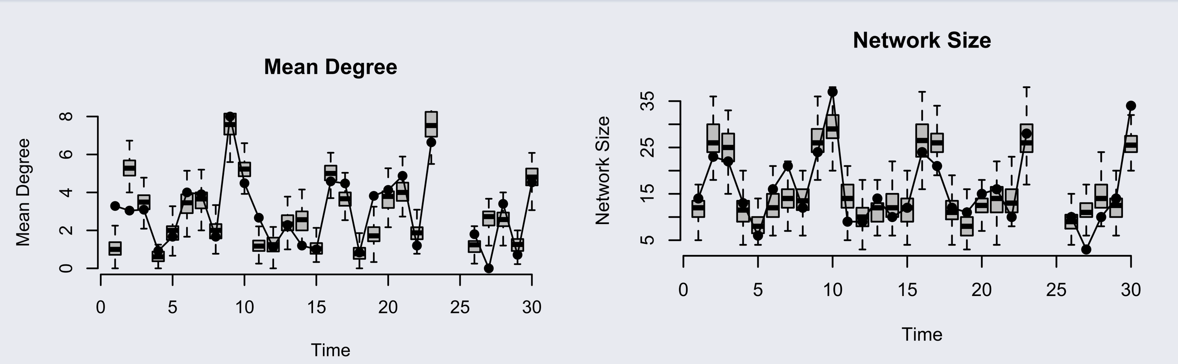
Dynamic Logistic Model

$$[\text{ERGM}]\ P(G=g \mid s,\theta) = \frac{\exp\left(\theta^T s(g)\right)}{\sum_{g'\in\mathcal{G}} \exp\left(\theta^T s(g')\right)} I_{\mathcal{G}}(g),$$

Dependence Diagram



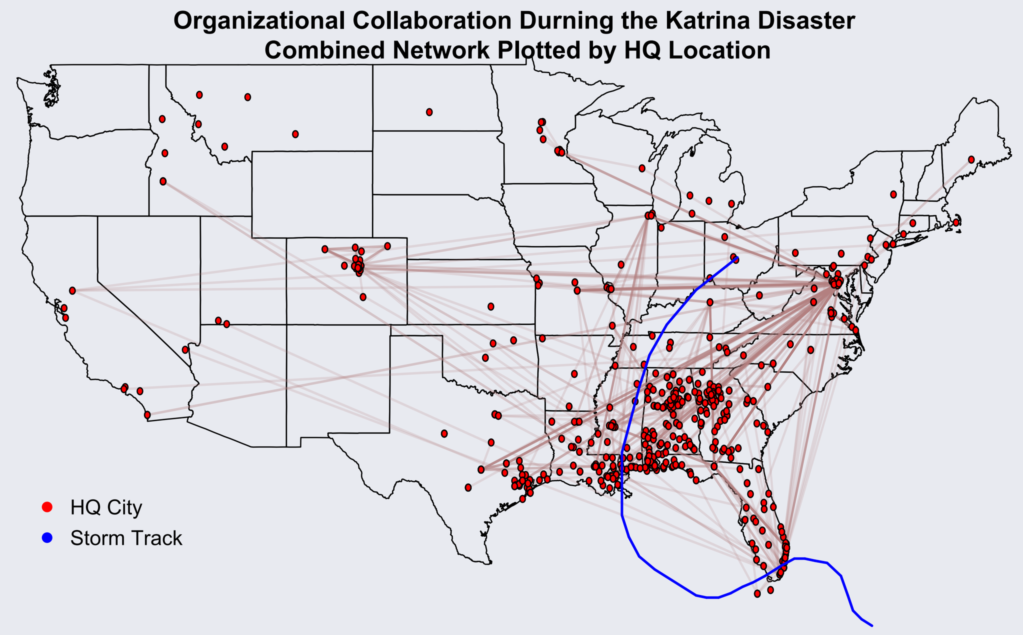
Edge and Vertex Prediction



Conversations between windsurfers from August to September 1986 (Freeman et al., 1988).

Contextual Effects: Org Collaboration Katrina

Data and Context: Butts et al. (2010)



Model

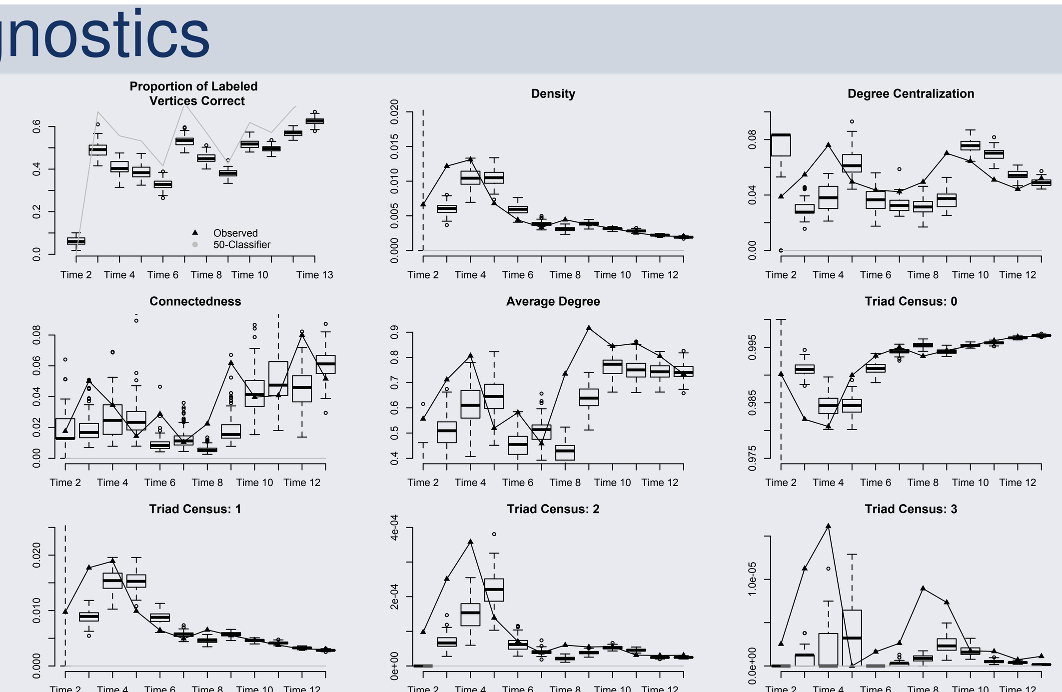
Vertex Parameter Estimates

	Model 5
BIC	31810.0173
Intercept	-4.5078*
	(0.0205)
Y_{t-1}	2.268*
	(0.0356)
$\log(n_{t-1})$	0.4273*
	(0.0035)
Degree	0.1989*
	(0.0309)
HQ State	-0.2274*
	(0.0206)
HQ City	0.3044*
	(0.0221)
FEMA Region	2.0954*
	(0.0206)
Type	0.4519*
	(0.0206)
Scale	-0.3264*
	(0.0206)
Sum of Lineage$_{t-1}$	-0.2943*
	(0.003)
Storm-track log Dist$_{t-1}$	0.0046*
	(0.001)

Edge Parameter Estimates

	Model 5
BIC	31810.0173
Density	-4.3685*
	(0.0293)
Y_{t-1}	5.8815*
	(0.0639)
$\log(n_{t-1})$	-0.5323*
	(0.0049)
Two-path	-0.1214*
	(0.0297)
Mean Degree	0.1877*
	(0.0061)
HQ State	1.2508*
	(0.04)
HQ City	-0.3382*
	(0.055)
FEMA Region	-0.3715*
	(0.037)
Type	0.6179*
	(0.0335)
Scale	0.0735
	(0.0436)
Lineage	1.9084*
	(0.1021)
Log Dist HQ city	-0.1539*
	(0.0056)

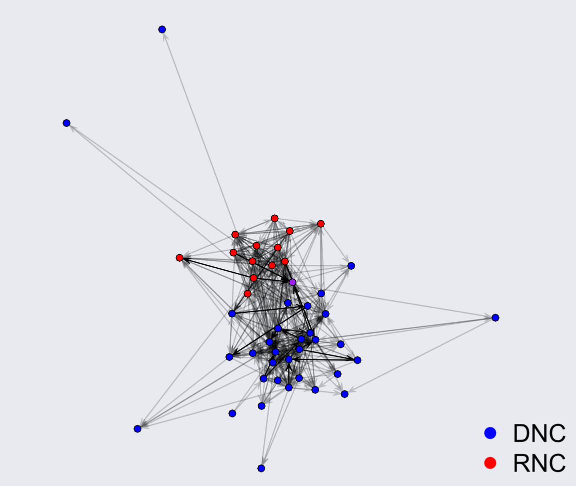
Diagnostics

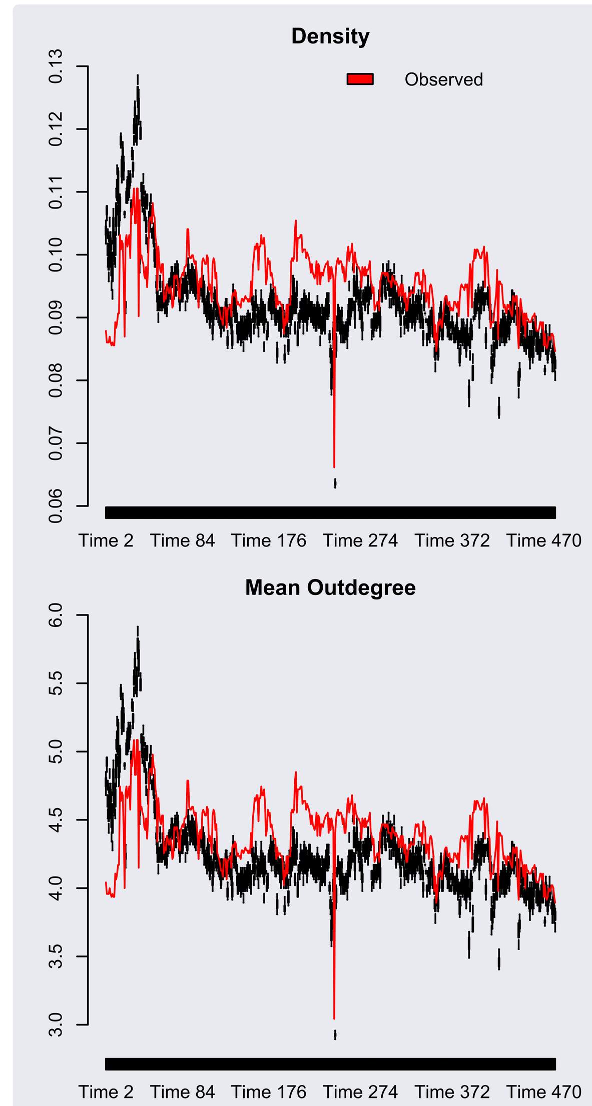


Logistic Choice

Data: Butts and Cross (2009)

Aggregate Network





Model

Assumptions

At its crudest level, a blog is a web page with dynamically updated links to other online resources.

1. The state of outgoing edges at each observation of the blog network is assumed to result from the choices of the sending blog;
2. Each blog in the network may send an edge to any number of other blogs in the network at any time;
3. The decision of a given blog regarding the state of a given edge is made myopically, and in isolation;
4. The decision of a given blog regarding the state of a given edge may depend upon the past history of the blog network, or of the current external context.

Parameters

- Includes an effect for clique counts from Eppstein et al. (2010).

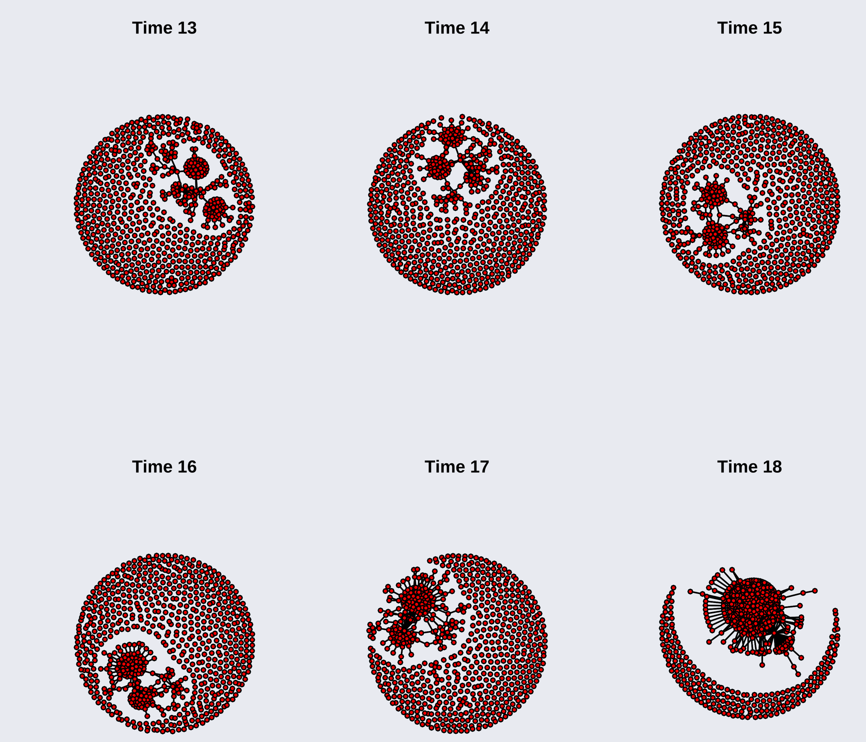
- *utility function*, u
- $A_{ij,t}=0$ or $A_{ij,t}=1$, the odds that i will choose $A_{ij,t}=1$ are strictly increasing in $u_i(A|A_{ij,t}=1)/u_i(A|A_{ij,t}=0)$
- *stochastic choice* process
- We employ a *logistic choice* model

$$\Pr(A_{ij,t}=1) = \frac{\exp\left[u_i\left(A|A_{ij,t}=1\right)\right]}{\exp\left[u_i\left(A|A_{ij,t}=1\right)\right]+\exp\left[u_i\left(A|A_{ij,t}=0\right)\right]},$$

or, equivalently, that

$$\operatorname{logit}(A_{ij,t}) = \ln\frac{\Pr(A_{ij,t}=1)}{\Pr(A_{ij,t}=0)} = u_i\left(A|A_{ij,t}=1\right) - u_i\left(A|A_{ij,t}=0\right)$$

N-Step Prediction of Katrina



Conclusions

1. Successfully applied Dynamic Logistic Regression to:
   - Interpersonal Collaboration (Conversations on a Beach)
   - Online Interaction (Blog Networks)
   - Organizational Collaboration (Katrina Disaster 2005)
2. Modeled both the Vertex and Edge Dynamics
3. Scalable Model through Simplified Assumptions
4. Behavioral and Utility Interpretations

References

Butts, Carter T., Ryan M. Acton, and Christopher Steven Marcum. 2010. "Interorganizational Collaboration In the Hurricane Katrina Response." *Journal of Social Structure* .

Butts, Carter T. and B. Remy Cross. 2009. "Change and External Events in Computer-Mediated Citation Networks: English Language Weblogs and the 2004 U.S. Electoral Cycle." *The Journal of Social Structure* 10.

Eppstein, David, Maarten Löffler, and Darren Strash. 2010. "Listing all maximal cliques in sparse graphs in near-optimal time." In *Proc. 21st Int. Symp. on Algorithms and Computation.* to appear.

Freeman, Linton C., Sue C. Freeman, and Alaina G. Michaelson. 1988. "On Human Social Intelligence." *Journal of Social Biological Structure* 11:415–425.

Sampson, S. F. 1968. *A Novitiate in a Period of Change: An Experimental and Case Study of Relationships.* Ph.D. thesis, Cornell University.

Snijders, Tom A.B. 2001. "The Statistical Evaluation of Social Network Dynamics." In *Sociological Methodology*, edited by M.E. Sobel and M.P. Becker Boston, pp. 361–395. London: Basil Blackwell.

UCIRVINE | DEPARTMENT OF SOCIOLOGY

NCASD
NETWORKS, COMPUTATION, and SOCIAL DYNAMICS

This material is based on research supported by the Office of Naval Research under award N00014-08-1-1015.

Scalable Methods for the Analysis of Network-Based Data