Modeling Dynamic Social Networks with Vertex Evolution via Latent Graphical Models

Ragupathyraj Valluvan
UC Irvine
rvalluva@uci.edu

Zack W. Almquist
UC Irvine
almquist@uci.edu

Animashree Anandkumar
UC Irvine
a.anandkumar@uci.edu

Carter T. Butts
UC Irvine
buttsc@uci.edu

Abstract

We here consider the problem of modeling network evolution with joint edge and vertex dynamics. It is natural to expect that the accuracy of vertex prediction strongly affect the ability to predict dynamic network evolution accurately. A latent graphical model is here employed to model vertex evolution. This model family can incorporate dependence in vertex co-presence, of the form found in many social settings (e.g., subgroup structure, selective pairing). Recent algorithms for learning latent tree graphical models and their extensions can be efficiently scaled for large graphs. Here, we introduce a novel latent graphical model based approach to the problem of vertex set prediction in dynamic social networks, combining it with a parametric model for covariate effects and a logistic model for edge prediction given the vertex predictions. We apply this approach to both synthetic data and a classic social network data set involving interactions among windsurfers on a Southern California beach. Experiments conducted show a significant improvement in prediction accuracy of the vertex and edge set evolution (about 45% for conditional vertex participation accuracy and 164% for overall edge prediction accuracy) over the existing dynamic network regression approach for modeling vertex co-presence.

Keywords: Social networks, dynamic networks, graphical models, latent variables, conditional random field.

1 Introduction

Over the last several decades there has been an increased interest in modeling and prediction of temporally evolving networks. This is especially relevant in the context of social networks. Examples include forming new organizational alliances, mass convergence of organizations in disasters, the formation of inter-firm networks within new industries, and interpersonal networks under strong external perturbations. Traditionally, network evolution has been studied with a fixed vertex set [e.g., 1, 2, 3, 4, 5]. This is unrealistic in many social and engineering processes e.g., addition and deletion of servers; entry and exit of students in a classroom; online social networks such as twitter friend/follower networks. In this paper, we consider co-evolution of vertex and edge sets in dynamic networks. An important consideration for any model is scalability, especially in the *high-dimensional* or *data-poor* regime. In other words, the number of samples may be far less than the parameters of the model, if we attempt to model all possible interactions between the vertices and edges. Recently, Almquist and Butts [6] introduced a family of simple network models that incorporates logistic models for vertex/edge dynamics (here on referred to as *Dynamic Network Regression*; DNR). However, the pure logistic model, while scalable, has limited ability to capture dependencies in vertex co-presence. Further, [6, 7] demonstrate the importance of obtaining the accurate vertex set predictions, noting that in many cases vertices are associated with covariates that strongly impact the edge structure. For instance, if one is modeling dynam-

ics of organizational collaboration during the hurricane Katrina disaster, predicting whether a prominent organization such as FEMA appears on a given day will greatly impact the network structure. Graphical models use undirected graphs to represent the conditional dependency structure [8] in a family of multivariate distributions. Recent work on latent tree learning [9, 10, 11] and their extension to learning loopy graphical models with latent variables [12] – have been proposed. These algorithms use conditional independence tests based on so-called "information distances" between pairs of variables to discover the latent variables in the model [13]. These approaches are scalable and possess consistency guarantees for learning the graph structure. Such developments in latent graphical models suggest their use as a mechanism for modeling dependence among the vertex set; these models can easily represent types of dependence expected in typical settings (e.g., subgroup structure, selective pairing), and can be efficiently scaled for very large graphs.

Summary of contributions Here, we propose a novel framework which combines a parametric model for the covariate effects with a latent graphical model (on to a conditional random field framework (CRF) [14]) as an approach to predict vertex evolution. Our approach first discovers the dependency structure with latent variables for vertex co-presence using the algorithms proposed in [9, 12]. Then we parametrize the edge and node potentials of the pairwise model represented by the discovered structure (with observed covariates). The model parameters are estimated using Expectation Maximization (EM) based on loopy propagation algorithm (LoopyBP) [15]. We consider various homogeneity assumptions to prevent over-fitting. We further explore the benefits of using Gaussian prior on parameters [16] to address over-fitting. Experiments conducted on synthetic datasets and real world data (a classic social network data set involving interactions among windsurfers on a California beach [6, 17]) show impressive performance in terms of statistical fit measured by AICc score, and in prediction accuracy of vertex and edge sets. Specifically, we observe about 45% improvement for vertex participation accuracy and 164% for edge prediction accuracy over the pure logistic approach for modeling vertex co-presence. At the same time, the approaches considered in this paper are scalable and fast to run on large datasets.

Related Work Temporal models for social network data tend to focus on either panel data (e.g., hourly, daily snapshots) or event data (e.g., sequences of tie formation/dissolution events). Here we focus on models for network panel data [for an introduction to event data models, see 18]. Currently, in the social network and statistical literatures there are three main families of models for panel data: the actor oriented models, which assume an underlying continuous-time model of network dynamics, where each latent event represents a single actor altering his or her outgoing links to optimize a function based on sufficient statistics [for details see, 1, 2, 19, 20]; latent variable approaches, such as dynamic latent feature models [e.g., 21] and dynamic latent space models [e.g., 22]; and the *temporal exponential-family random graph models* [TERGM; see, 3, 4, 5, 6, 7, 23]. Here, we focus on DNR, a subfamily of the TERGM approach and the only model family which currently incorporates endogenous vertex dynamics. The TERG model was extended to handle vertex dynamics by making a separable parameterization between the vertex set and edge set [6]. This was achieved through assumptions; i) given the vertex covariates, vertices are conditionally independent, and ii) given the vertex set and edge covariates, edges are conditionally independent. In our work, we relax conditional independence assumptions made in DNR which restricts the vertex process, by allowing dependence between vertex appearances via a latent graphical model.

CRF models have been studied extensively in areas such as natural language processing [24], image modeling [25], classification [16], and in dynamic CRF frameworks [26]. These models use the observed (a.k.a input) variables (in our setting covariates) to predict the unobserved states of output variables (in our setting a vertex presence/absence). The structure of these models varies from linear chains [24], trees [16, 27], to grids [25]. Few works address the issue of structure learning via greedy methods [28], convex programming [29] and local learning approaches based on conditional mutual information [27]. CRF models considering latent variables assume the existence and location are known along with corresponding covariates [16]. Our work does not make such strong assumptions; we test for evidence of their existence and incorporate them into the structure using CLGrouping and LocalCLGrouping algorithms. This enables us to discover the natural latent graph structure. Further, we do not restrict the choices of covariates for latent variables, which gives an additional flexibility towards dataset specific choices.

2 Statistical Models for Network Dynamics

Social networks are often represented as a *graph* (i.e., $G = (V, E)$, with V being a *vertex set* and E being a set of pairs or ordered pairs of vertices). This representation can be extended to model dynamic data with a time index, i.e. $G_t = (V_t, E_t)$, with G_t reflecting the state of the graph at time t.

TERGM TERGM models typically begin with a conditional independence assumption. Given adjacency matrix Y_t, then we can make the assumption that Y_t is independent of $Y_1, \ldots, Y_{t-2}$ given Y_{t-1} [5]; this can be generalized further for arbitrary lags (e.g., $Y_{t-1}, \ldots, Y_{t-k} = Q_{t-k}$). Under this assumption, we can write down the standard TERG model:

$$\Pr(Y_t = y_t \mid Q_{t-k}, X) = \frac{\exp\left(\theta^T s(y_t, Q_{t-k}, X)\right)}{\sum_{y' \in \mathcal{Y}} \exp\left(\theta^T s(y_t', Q_{t-k}, X)\right)}, \tag{1}$$

for y_t belonging to the support, $\mathcal{Y}$. s here is a vector of real-valued sufficient statistics, with parameter vector θ. Notice that the denominator of (1) is intractable in the general case. See Appendix A for known issues with general TERGMs.

TERGM with Vertex Dynamics The TERG model in (1) is further extended to handle vertex dynamics by making a separable parameterization between the vertex set and edge [6]. Hence, TERG model with vertex dynamics could be given as:

$$\begin{aligned}\Pr(G_t = g_t \mid G_{t-1}, \cdots, G_{t-k}, X) &= \Pr(V_t = v_t \mid Z_{t-k}, X) \times \Pr(Y_t = y_t \mid V_t, Z_{t-k}, X) \\ &= \frac{\exp\left(\psi^T w(v_t, Z_{t-k}, X)\right)}{\sum_{v' \in \mathcal{V}} \exp\left(\psi^T w(v_t', Z_{t-k}, X)\right)} \times \frac{\exp\left(\theta^T u(y_t, V_t, Z_{t-k}, X)\right)}{\sum_{y' \in \mathcal{Y}} \exp\left(\theta^T u(y_t', V_t, Z_{t-k}, X)\right)},\end{aligned} \tag{2}$$

for observations in the support, where $Z_{t-k} = (V_{t-1}, Y_{t-1}), \cdots, (V_{t-k}, Y_{t-k})$.

Dynamic Network Regression [6] modeled $\Pr(V_t = v_t \mid Z_{t-k}, X)$ and $\Pr(Y_t = y_t \mid V_t, Z_{t-k}, X)$ as separable logistic processes. Under the necessary conditional independence, homogeneity, and temporal Markov assumptions one can derive the likelihood function for DNR, where the vertex likelihood is given by

$$\Pr(V_t \mid Z_{t-k}, X) = \prod_{i=1}^{n} B\left(\mathbb{I}(v_i \in V_t) \left| \text{logit}^{-1}\left(\psi^T w(i, Z_{i-k}, X)\right)\right.\right) \tag{3}$$

and the edge likelihood by

$$\Pr(Y_t \mid V_t, Z_{t-k}, X) = \prod_{(i,j) \in V_t \times V_t}^{n} B\left(Y_{tij} \left| \text{logit}^{-1}\left(\theta^T u(i, j, V_t, Z_{i-k}, X)\right)\right.\right), \tag{4}$$

where B is understood to be the Bernoulli pmf, $\mathbb{I}$ is the indicator function, X is a covariate set, u and w are sufficient statistics for the edge and vertex models (respectively), θ and ψ are the respective edge and vertex parameter vectors, and Z_i is defined as before. Here, we are proposing to relax these restrictions in the vertex process, allowing dependence between vertex appearances via a latent graphical model.

3 Latent Graphical Models

Graphical models A *binary graphical model* on a $\mathcal{G}_{\text{dep}} = (\mathcal{W}_{\text{dep}}, \mathcal{E}_{\text{dep}})$ is a family of multivariate distributions whose conditional dependence relations are expressed by a fixed undirected graph $\mathcal{G}_{\text{dep}}$ [8]. Each vertex in the graph $i \in \mathcal{W}_{\text{dep}}$ is associated with a random variable V_i taking value in $\{0, 1\}$. The edge set, $\mathcal{E}_{\text{dep}}$, captures the set of conditional independence relations among the random variables in $\mathcal{W}_{\text{dep}}$. We say that a set of random variables $\mathbf{V}$ with probability mass function (pmf) Pr is *Markov* on the graph $\mathcal{G}_{\text{dep}}$ if

$$\Pr(v_i | v_{\mathcal{N}(i)}) = \Pr(v_i | v_{\mathcal{W}_{\text{dep}} \setminus i}) \tag{5}$$

holds for all nodes $i \in \mathcal{W}_{\text{dep}}$, where $\mathcal{N}(i)$ are the neighbors of node i in graph $\mathcal{G}_{\text{dep}}$. The Hammersley-Clifford theorem [8] states that, under the positivity condition given by $\Pr(\mathbf{V}) > 0$ for all $\mathbf{V}_U \in \{0, 1\}^{|\mathcal{W}_{\text{dep}}|}$, a distribution Pr satisfies the Markov property according to a graph $\mathcal{G}_{\text{dep}}$ iff it factorizes according to the cliques of $\mathcal{G}_{\text{dep}}$. For instance, the distribution of a class of graphical models where the maximal clique size is 2 factorizes as

$$\Pr(\mathbf{V}) = \exp\left(\sum_{e \in \mathcal{E}_{\text{dep}}} \phi_{i,j} v_i v_j + \sum_{i \in \mathcal{W}_{\text{dep}}} \phi_i v_i - A(\boldsymbol{\theta})\right), \tag{6}$$

where $\boldsymbol{\phi}_e := \{\phi_{i,j}\}$ and $\boldsymbol{\phi}_n := \{\phi_i\}$ are respectively known as edge and the node potentials, $\boldsymbol{\theta} := \boldsymbol{\phi}_e \cup \boldsymbol{\phi}_n$ and $A(\boldsymbol{\theta})$ is known as the *log-partition function*, which normalizes the probability distribution.

Latent Graphical Models: A latent graphical model is a class of *graphical models* in which a subset of nodes is latent or hidden. We denote the hidden nodes by $\mathcal{H}_{\text{dep}} \subset \mathcal{W}_{\text{dep}}$ and the observed nodes by $\mathcal{V}_{\text{dep}} \subset \mathcal{W}_{\text{dep}}$.

Learning Latent Graphical Models: In general, learning latent graphical models consists of two tasks. The first task is to discover the presence of hidden variables corresponding to nodes $\mathcal{H}_{\text{dep}} \subset \mathcal{W}_{\text{dep}}$ and learn the unknown graph structure $\mathcal{G}_{\text{dep}}$, given N i.i.d. samples from observed variables corresponding to nodes $\mathcal{V}_{\text{dep}}$. Second task is is to estimate the model parameters $\widehat{\boldsymbol{\theta}}$ given the discovered structure $\widehat{\mathcal{G}}_{\text{dep}}$ and N i.i.d samples from observed variables at nodes $\mathcal{V}_{\text{dep}}$. Among the available approaches for latent tree learning, we build on algorithms based on a measure of statistical distance (a.k.a information distance),

$$d_{ij} =: -\log|\det(\widehat{\Pr}^{N}_{V_i,V_j})|, \tag{7}$$

where $\widehat{\Pr}^{N}_{V_i,V_j}$, is the empirical joint statistics of nodes i, j using N i.i.d samples, d is an additive tree-metric [13] satisfying the Markov property, and it forms the basis of scalable structure learning algorithms such as RG and CLGrouping in [9], and LocalCLGrouping (loopy graphs) [12] with provable guarantees. In our work, we will be using RG, CLGrouping and LocalCLGrouping for learning latent graph structure.

4 Conditional Random Field on Latent Graphical Models

A conditional random field (CRF) based approach [14] provides a modeling framework for label variables V_t according to an undirected graphical model $\mathcal{G}_{\text{dep}} = (\mathcal{W}_{\text{dep}}, \mathcal{E}_{\text{dep}})$ conditioned on observed covariate data $\mathbf{X}$. In our context, observed data will correspond to the covariates, and label variables will denote people in the dynamic network.

Definition 1 *Let $\mathcal{G}_{\text{dep}} = (\mathcal{W}_{\text{dep}}, \mathcal{E}_{\text{dep}})$, be a dependence graph on a binary graphical model. Then, ($\mathbf{X}, \mathbf{V}$) is a conditional random field, (CRF) if, when conditioned on $\mathbf{X}$, the random variables V_i obey the Markov property with respect to $\mathcal{G}_{\text{dep}}$; i.e. $\Pr(V_i|\mathbf{X}, V_{\mathcal{W}_{\text{dep}}\setminus i}) = \Pr(V_i|\mathbf{X}, \mathcal{N}(i))$, where $\{\mathcal{W}_{\text{dep}} \setminus i\}$ is the set of all nodes in the graph except i, and $\mathcal{N}(i)$ is the neighbors of i.*

Using the pairwise exponential form of the Hammersly-Clifford theorem, the distribution defined above could be given as;

$$\Pr(V|\mathbf{X}) = \exp\left(\sum_{ij\in\mathcal{E}_{\text{dep}}} \phi_{ij}(\mathbf{X}, \boldsymbol{\theta}) v_i v_j + \sum_{i\in\mathcal{W}_{\text{dep}}} \phi_i(\mathbf{X}, \boldsymbol{\theta}) v_i - A(\boldsymbol{\theta})\right), \tag{8}$$

where ϕ_{ij} and ϕ_i denote the potential functions of edge $(i,j) \in \mathcal{E}_{\text{dep}}$ and node $i \in \mathcal{W}_{\text{dep}}$ respectively, $\boldsymbol{\theta}$ is the set of model parameters. Given, covariates $\mathbf{X}$ and $\boldsymbol{\theta}$, with functions ϕ_{ij}, ϕ_i we can compute the corresponding edge and node potentials of the model. These functions could be chosen depending on the modeling requirements. For simplicity, lets assume that ϕ_{ij} and ϕ_i are linear functions of covariates, $\mathbf{X}$, i.e.

$$\phi_i = c_0 + c_1 x_{1,i} + c_2 x_{2,i} + ... + c_{K_n} x_{K_n,i}, \tag{9}$$

where c_k is the coefficient of kth covariate, K_n is the total number of covariates used for the parametrization, and $x_{k,i}$ is the k th covariate of i th vertex, and likewise

$$\phi_{ij} = e_0 + e_1 x_{1,ij} + e_2 x_{2,ij} + ... + e_{K_e} x_{K_e,ij}, \tag{10}$$

where e_k is the coefficient of kth covariate, K_e is the total number of covariates used for the parametrization, and $x_{k,ij}$ is the kth covariate of the edge (i,j).

The absence of prior information and interpretation makes the choice of covariates for latent variables challenging. A natural choice in a dynamic setting would be to use a set of shared covariates such as seasonality.

We consider a set of homogeneity assumptions to limit the number of free parameters, in order to avoid over-fitting in *high-dimensional* or *data-poor* settings. Benefits of using a Gaussian prior [16] on parameters is explored during the experiments.

Homogeneity assumptions: We can consider one of the following assumptions;

- A1 Edge and node coefficients are homogeneous throughout the entire graph, Or
- A2 Edge and node coefficient (e, c) within a connected component of the graph are homogeneous while the independent nodes in the forest have a separate set of homogeneous node coefficients c, Or
- A3 Edge and node coefficients (e, c) are homogeneous except the constant terms therein.

Among the set of assumptions above, (A1) represents the strongest level of homogeneity, while (A3) provides flexibility to allow for individual variation on the average of edge and node potentials. (A2) allows for heterogeneity across different disconnected components of a forest while maintaining homogeneity within each component. The model under assumption (A1) or (A2) could be given as:

$$\Pr(V|\mathbf{X};\boldsymbol{\theta}) = \exp(\sum_{i\in\mathcal{W}_{\text{dep}}} c^T x_i v_j + \sum_{ij\in\mathcal{E}_{\text{dep}}} e^T x_{ij} v_i v_j - A(\boldsymbol{\theta})). \tag{11}$$

Structure Learning Given N observations on $\mathbf{X}$ and V, we need to discover the structure of the graphical model. Structure learning algorithms in [9] and [12] propose a distance measure of the form defined in (7) to learn the structure using N i.i.d observations. In conditional random field models defined per definition 1, observation on V is independent given the observation on covariates $\mathbf{X}$. Hence, it necessary to use a distance measure which takes this conditional independence into account to satisfy the Markov property on the graph learned for CRF. We propose a distance measure conditioned on covariates;

$$[d_{ij}|\mathbf{x}] := \sum_{k=1}^{K_{ij}} w_{k,ij} d_{k,ij}, \tag{12}$$

where $w_{k,ij}$ are empirical probabilities of covariate pairs $(X_{k,i}, X_{k,j})$, such that $\sum_{k=1}^{K_{ij}} w_{k,ij} = 1$, K_{ij} is the total number of observed covariate pairs, and $d_{k,ij} := -\log|\widehat{\Pr}(V_i, V_j|(X_{k,i}, X_{k,j}))|$. This conditional distance measure could then be used in CLGrouping and LocalCLGrouping algorithms [9, 12] to learn latent graph structure from data. Learning loopy graph using LocalCLGrouping requires a distance threshold; $\mathsf{r_{th}} \in (r_{\min}, r_{\max})$ where,

$$r_{\min} := \min_{(i,j)\in V\times V} d_{ij},\ r_{\max} := \max_{(i,j)\in V\times V} d_{ij}, \tag{13}$$

as input to learn graphs with varying levels of density.

Forests Structures learnt using CLGrouping and LocalCLGrouping algorithms are connected graphs; even though there are marginally independent nodes, noisy data results in non-zero distance measures. The presence in the structure of such nodes, and nodes with very weak conditional dependencies, poses a problem: a small advantage in statistical fit comes at a cost in terms of more parameters (and hence over-fitting risk) and complexity. In order to address this, we remove weak edges from the structure learnt using CLGrouping and LocalCLGrouping, thus forming a forest. The threshold $\mathsf{r_{th}}$ defined in (13), is used for trimming weak edges, it gives an additional degree of freedom to systematically find an adequate structure which reduces over-fitting while increasing prediction performance.

Prediction The dynamic network model attempts to predict the states of the vertices V given the covariates and model parameters in a maximum likelihood fashion. It is possible to estimate the most likely states of the vertices resulting in a MAP estimate under uniform priors. However, in our work we simulate instances by drawing samples from $\Pr(V|X_t;\boldsymbol{\theta})$, giving us a more complete sense of predictive uncertainty. We use Gibbs sampling to draw samples from this model to predict several likely instances. The model parameters $\boldsymbol{\theta}$ and observed covariates X_t corresponding to prediction instance t are used to compute the node and edge potentials using (9) and (10).

5 Experiments

We conduct extensive experiments on synthetic and real world data. We use a penalized likelihood score, the Akaike Information Criterion with corrections (AICc) for finite sample size, to measure goodness of statistical fit. Predictive accuracy is evaluated using the estimate of prediction accuracy. These model selection criteria are used to compare the performances of DNR vertex prediction and for prediction of social network structure more generally.

Synthetic Data We use a randomly generated tree structure to build a latent graphical model with 55 nodes (i.e. $|\mathcal{W}_{\rm dep}| = 55$), where we parametrize the node and edge potentials with covariates of lengths $|\phi_i| = 10$, $|\phi_{ij}| = 9$ for nodes and edges respectively. Among these covariates, we included indicator variables to introduce seasonality. Model parameters $\boldsymbol{\theta}$ are arbitrarily generated from uniform distributions and samples are drawn from the model using a Gibbs sampler. This data set has observed variables $|V| = 50$ and a sample size $N = 200$ with a sparsity of 0.19.

Beach Data As a comparison case we use the same data and model as in [6]. This data involves a dynamically evolving network of interpersonal communication among individuals congregating on a beach in Southern California over a one-month observation period [17]. This network was collected daily (aggregated over a morning and an afternoon observation period) for 31 days (August 28, 1986 to September 27, 1986).[1] Individuals were tracked with a unique ID, and were divided by Freeman et al. into *"regulars"* ($n = 54$) – frequent attendees who were well-integrated into the social life of the beach community – and *"Irregulars"* ($n = 41$) on ethnographic grounds. The *"regulars"* were further broken into two groups, *group 1* ($n = 22$) and *group 2* ($n = 21$), with 11 individuals not classified as belonging to either group 1 or group 2. The union of these sets (V) consists of 95 individuals. On any given day during the observation period, the number of windsurfers appearing on the beach ranged from 3 to 37, with the number of communication ties per day ranging from 0 to 95.

Experimental setup For each simulated data set, a set of different structures are learnt using the information distances defined in (7) and (12). Given the estimated information distance, a threshold $\mathsf{r_{th}}$ defined in (13) is chosen to form forests, where $\mathsf{r_{th}} \in (r_{\min}, r_{\max})$; and resulting forests are expected to have independent nodes, trees, loopy sub-graphs as components . Then EM algorithm based on LoopyBP and Gradient Ascent method is used to find the maximum likelihood estimate of the coefficients for node and edge parametrization. For models incorporating a prior on parameters, a Gaussian prior with $\sigma_p > 1$ is used. The set of thresholds $\mathsf{r_{th}}$ (defined in 13), considered for beach data is in the interval, $\mathsf{r_{th}} \in [0, 23]$, for models using structure leant from unconditioned information distance defined in (7). For structures using conditioned distance defined in (12), $\mathsf{r_{th}} \in [0, 9]$. Experiments on synthetic dataset learn a single tree instead of forest, since the underlying structure is a latent tree.

Model Selection The corrected Akaike Information criterion (AICc) is used for model selection.

$$\mathsf{AICc} := 2p - 2\mathcal{L} + \frac{2p(p+1)}{N-p-1}, \quad \mathcal{L}(V|\mathbf{X};\boldsymbol{\theta}) = \Sigma_t \log(\Pr(V_t|X_t;\boldsymbol{\theta})), \tag{14}$$

where p is the number of parameters in the model, N is the number of samples in the dataset, $\mathcal{L}$ is the log-likelihood of the data set, $\boldsymbol{\theta}$ is the set of model parameters, and t is the time index.

Prediction Scores We assess prediction accuracy by comparing the observed values for vertex presence at each time point with the corresponding predicted state. We estimate the correct number of predictions for those who appear and who do not appear. In addition, we also evaluate the correct number of edges predicted using DNR edge model which uses the predicted vertex set from our models;

$$\mathsf{PredPresent} := \sum_{t=1}^{N}\sum_{m=1}^{M} \frac{\mathbb{I}(\widehat{V}_{i,t} = 1|V_{i,t} = 1)_m}{MN}, \quad \mathsf{PredAbsent} := \sum_{t=1}^{N}\sum_{m=1}^{M} \frac{\mathbb{I}(\widehat{V}_{i,t} = 0|V_{i,t} = 0)_m}{MN} \tag{15}$$

$$\mathsf{VertPredAccuracy} := \sum_{t=1}^{N}\sum_{m=1}^{M} \frac{\mathbb{I}(\widehat{V}_{i,t} = V_{i,t})_m}{MN}, \quad \mathsf{EdgePredAccuracy} := \sum_{t=1}^{N}\sum_{m=1}^{M} \frac{\mathbb{I}(\widehat{Y}_{ij,t} = Y_{ij,t})_m}{MN}, \tag{16}$$

where N is the number of predicting instances or samples, M is the number of samples drawn for likely vertex set presence on a given day, PredPresent is the conditional prediction accuracy given an individual is present, PredAbsent is the conditional accuracy given an individual is not present, VertPredAccuracy is the overall accuracy on how well we predict given the individual is absent or present, EdgePredAccuracy is the overall accuracy of predicting edges (conversation among the individuals present at beach), $V_{i,t} \in V_t$ is the corresponding vertex, $Y_{ij,t} \in E_t$ is the corresponding edge in the dynamic network, and t is the corresponding time index (or sample point in test set).

Outcomes A summary of experimental outcomes for synthetic data is given in Table 1. The outcomes for synthetic data given in Table 1 corresponding to $\mathsf{r_{th}} = r_{\max}$, forms a tree. A selected set of outcomes for beach data is given in Table 2, these outcomes correspond to a set of selected thresholds $\mathsf{r_{th}}$.

[1] Unfortunately, one day (September 21st) is missing due to a race on a different beach, which precluded data collection. Thus, complete data is available for 30 days during the observation period.

	$N = 20$				$N = 80$				$N = 160$			
Model	CP %	CA %	VP %	AICc	CP %	CA %	VP %	AICc	CP %	CA %	VP %	AICc
UC-H	36.8	85.6	76.3	8346	36.7	85.56	76.2	6717	39.6	85.4	76.5	4962
UC-NH	30.6	89.6	78.4	19492	38.5	86.85	77.6	9424	42.0	87.1	78.3	5282
C-H	35.6	85.6	76.1	7311	36.22	85.1	75.8	6448	35.9	85.2	75.7	4878
C-NH	38.0	87.0	77.7	11605	41.78	86.5	78.0	7559	41.8	87.0	78.3	5042
DNR-V	36.1	82.4	73.6	7941	35.84	81.6	72.9	7063	35.7	81.7	72.8	5340

Table 1: Prediction performance of different models compared with baseline, DNR-Vertex model in terms PredPresent (CP) in (15), PredAbsent (CA) in (15), VertPredAccuracy (VP) in (16),and AICc score (14) using $(200 - N)$ test samples of the synthetic data with varying number of training samples N, for models with homogeneity(H) and non-homogeniety(NH) assumption, graph structures(tree) using conditioned(C) in (12) and unconditioned(UC) in (7) information distances in Synthetic data with sparsity $= 0.19$. Tree models beat the DNR-V baseline as training set grows.

No.	Model	r_{th}	σ_p	AICc	CP (%)	CA(%)	VP (%)	EP (%)
1	UC-NH	0.9	na	2000.08	43.91	88.40	80.86	11.99
3	UC-H	0.9	na	1928.10	41.59	88.02	80.15	9.79
4	UC-H	1	na	1935.62	41.97	88.34	80.48	10.18
5	C-NH	9	na	2207.52	45.19	85.52	78.68	11.17
7	UC-NH	1.1	na	1995.3	39.03	91.10	82.27	11.30
8	UC-NH	1.1	1.6	2013.4	45.21	89.13	81.68	13.26
10	UC-H	1.1	1.95	1950.8	41.64	88.13	80.25	10.27
DNR-V	na	na	na	2037.80	31.16	86	76.75	5.03

Table 2: Comparison of performance in terms of AICc (14) score and prediction score for correct vertex prediction (VP) in (16), correct conditional prediction of people at beach (CP) in (15), people not at beach (CA) in (15), and conversation between people (EP) in (16) for models with homogeneity (H) and non-homogeniety (NH) assumption, graph structures (tree) using conditioned (C) (12) and unconditioned (UC) (7) information distances. Models correspond to threshold r_{th} values in Beach data (resulting in latent graphs) and standard deviation σ_p for Gaussian prior on parameters. Dynamic Network Regression-vertex (DNR-V) baseline is substantially outpredicted by all tree models in getting the labels correct.

Graph Structure For Beach data, a number of different structures resulted from the threshold choices made above. Structures learnt using distance in (7) produced a diverse set of forests with multiple tree components while using distance in (12) produced forests with a single tree and independent nodes. This difference indicates that covariates in Beach data capture significant dependencies, resulting in a weak structure when conditioned on (figures 4 and 5.). Further, these structures show that vertices classified as *"regulars"* have more clusters among themselves compared to *"Irregulars"*. *"Irregulars"*, while having small clusters among themselves, show conditional dependency on *"regulars"*. This could be attributed to the sparsity of attendance seen in *"Irregulars"*, while *"regulars"* show up more regularly.

Synthetic Data The outcomes for the synthetic data experiments are shown in Table 1. The baseline DNR-V model performs better for very small sizes (e.g. $N = 20$). However, as the sample size increases, more complex models show better performance. It should be noted that when $N < |\mathcal{V}_{\mathrm{dep}}|$, our models perform equally or slightly worse than the DNR-V model. In addition, we note that latent structure models with homogeneous assumptions perform as well as DNR-V baseline (besting it, as do the inhomogeneous models, when fit to a larger training set). Hence, this shows our homogeneity assumptions do in fact reduce over-fitting, as predicted. We emphasize that the scores in Table 1 are obtained using separate set of data unseen by the model (out of sample) and out perform DNR-V model in vertex prediction accuracy VertPredAccuracy in (16) for all sample sizes. Thus, it is clear that our model provides a substantial improvement of predictive performance versus pure logistic regression for larger data sets, while providing essentially similar performance for homogeneous models in the small data regime.

Beach Data: Homogeneous Models Among the models shown in Table 2, homogeneous models using structure learnt from unconditioned distance (7) perform well in terms of AICc score; however the prediction performance is

low compared to respective non-homogeneous counterparts. It should be noted that these prediction scores given for Beach data are obtained using the training set, due to small number of samples. However, the outcome given for synthetic data in Table 1 shows that homogeneous models (including the DNR-V model) perform better than non-homogeneous models with small number of samples. In Table 2, homogeneous models with $\mathsf{r}_{\mathsf{th}} = 0.9$, and 1 have better AICc scores than non-homogeneous model with $\mathsf{r}_{\mathsf{th}} = 0.9$ while performing equally well in terms of prediction (VertPredAccuracy). In general, number of tree/loopy components in $\mathsf{r}_{\mathsf{th}} < r_{\max}$ increases as we reduce the threshold, r_{th}. Hence, homogeneity assumptions in some cases enable us to find a homogeneous model which performs equally well as a non-homogeneous model with fewer free parameters. Figures 2a and 2b show variation of VertPredAccuracy of individuals for DNR-V model and a homogeneous model with $\mathsf{r}_{\mathsf{th}} = 0.9$ respectively, our model shows a significant improvement over DNR-V model. In addition, figures 1a and 1b show variation in aggregates of predicted vertices for the same models; our model performs equally well and in some instance better than DNR-V model.

Beach Data: Non-Homogeneous Models From Table 2, non-homogeneous models exhibit better performance in terms of prediction accuracy; however they perform poorly in terms of AICc score. The larger number of free parameters compared to homogeneous models lead to the poor AICc score due to over-fitting. A similar trend could be seen in Table 1, $N < |\mathcal{V}_{\mathrm{dep}}|$ leads to over-fitting resulting in poor prediction performance of correct labels on unseen data. However, as the sample size N gets closer to the number of nodes $|\mathcal{V}_{\mathrm{dep}}|$, the performance increases, eventually out-performing the other models. However, as results from Beach data in Table 2 show, models formed according to homogeneity assumptions enable us to find non-homogeneous model which performs closer to a homogeneous model with a small compromise in predictive performance. This approach would prove fruitful in the *high-dimensional* regime.

Beach Data: Edge Prediction Vertex set predicted is used in DNR-Edge model (4) to predict the edge set (i.e. E_t). These outcomes are given in Table 2, and figure 3 shows the variation in percent of correctly predicted edges. A significant performance improvement is observed in terms of EdgePredAccuracy (16) in table 2. A peak performance of 13.26% is shown by model (8) which is non-homogeneous using unconditional information distance (7) with a prior on parameters, compared to 5.03% performance of DNR-V model.

In summary, the above experimental results on the Synthetic and Beach data demonstrate models resulting from the proposed scalable dynamic vertex set prediction framework with conditional random fields, defined on latent graphical models and with a systematic approach on forming forests along with specified homogeneity assumptions. This work resulted in significantly improved performance on predicting vertex co-presences and edge dynamics in the experimental cases.

Acknowledgement

This work was supported in part by ONR award N00014-08-1-1015, NSF award BCS-0827027, NIH/NICHD award 1R01HD068395-01, NSF Award CCF-1219234, AFOSR Award FA9550-10-1-0310, ARO Award W911NF-12-1-0404, and setup funds at UCI.

References

[1] Tom A.B. Snijders. Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology*, 21(1–2):149–172, 1996.

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

[3] Bruce A. Desmarais and Skyler J. Cranmer. Consistent confidence intervals for maximum pseudolikelihood estimators. *Neural Information Processing Systems 2010 Workshop on Computational Social Science and the Wisdom of Crowds*, 2010.

[4] Bruce A. Desmarais and Skyler J. Cranmer. Statistical mechanics of networks estimation and uncertainty. *Physica A*, 391(4):1865–1876, 2011.

[5] Steve Hanneke, Wenjie Fu, and Eric P. Xing. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605, 2010.

[6] Zack W. Almquist and Carter T. Butts. Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. *Sociological Methodology*, forthcoming, 2012.

[7] Zack W. Almquist and Carter T. Butts. Evolving context: Evidence from temporal change in organizational collaboration over the course of the 2005 katrina disaster. forthcoming as an IMB Technical Report, University of California, Irvine, 2012.

[8] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[9] Myung Jin Choi, Vincent Y.F. Tan, Animashree Anandkumar, and Alan S. Willsky. Learning latent tree graphical models. *Journal of Machine Learning Research*, 12, 2011.

[10] P. L. Erdös, L. A. Székely, M. A. Steel, and T. J. Warnow. A few logs suffice to build (almost) all trees: Part i. *Random Structures and Algorithms*, 14:153–184, 1999.

[11] E. Mossel. Distorted metrics on trees and phylogenetic forests. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 108–116, 2007.

[12] Animashree Anandkumar and Ragupathyraj Valluvan. Learning loopy graphical models with latent variables: Efficient methods and guarantees. arxiv:1203.3887, 2012.

[13] James Lake. Reconstructing evolutionary trees from dna and protein sequences: paralinear distances. *Proceedings of the National Academy of Sciences*, 91(4):1455–1459, 1994.

[14] John Lafferty. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. pages 282–289. Morgan Kaufmann, 2001.

[15] Kevin P. Murphy, Yair Weiss, and Michael I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *In Proceedings of Uncertainty in AI*, pages 467–475, 1999.

[16] A. Quattoni, S. Wang, L. p Morency, M. Collins, T. Darrell, and Mit Csail. Hidden-state conditional random fields. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007.

[17] Linton C. Freeman, Sue C. Freeman, and Alaina G. Michaelson. On human social intelligence. *Journal of Social Biological Structure*, 11:415–425, 1988.

[18] Carter T. Butts. A relational event framework for social action. *Sciological Methodology*, 38(1):155–200, 2008.

[19] Tom A.B. Snijders and M.A.J. Van Duijn. Simulation for statistical inference in dynamic network models. In R. Conte, R. Hegselmann, and P. Terna, editors, *Simulating Social Phenomena*, pages 493–512. Springer, Berlin, 1997.

[20] Tom A.B. Snijders. Models for longitudinal network data. In P. Carrington, J. Scott, and S. Wasserman, editors, *Models and methods in social network analysis*, chapter 11. Cambridge University Press, New York, 2005.

[21] James Foulds, Christoher DuBois, Arthur U. Asuncion, Carter T. Butts, and Padraic Smyth. A dynamic relational infinite feature model for longitudinal social networks. In *Proceedings of the 14th International Conference on Articial Intelligence and Statistics (AISTATS)*, volume 15, pages 287–295, Fort Lauderdale, FL, 2011.

[22] Purnamrita Sarkar and Andrew W. Moore. Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems (NIPS)*, volume 18, pages 1145–1152, Cambridge, MA,, 2005. MIT Press.

[23] Steve Hanneke and Eric P. Xing. *Statistical Network Analysis: Models, Issues, and New Directions: ICML 2006 Workshop on Statistical Network Analysis, Pittsburgh, PA, USA, June 29, 2006, Revised Selected Papers*, volume 4503 of *Lecture Notes in Computer Science*, chapter Discrete Temporal Models of Social Networks, pages 115–125. Springer-Verlag, 2007.

[24] Fei Sha and Fernando Pereira. Shallow parsing with conditional random fields. pages 213–220, 2003.

[25] Sanjiv Kumar and Martial Hebert. Discriminative fields for modeling spatial dependencies in natural images. In *In NIPS*. MIT Press, 2003.

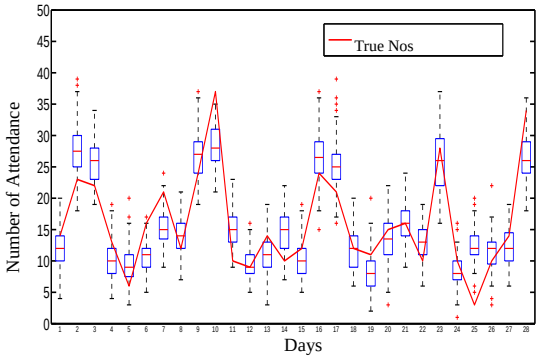
[26] Charles Sutton., Andrew McCallum., and Khashayar Rohanimanesh. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research*, 8, 2007.

[27] Joseph K. Bradley and Carlos Guestrin. Learning tree conditional random fields. In *International Conference on Machine Learning (ICML 2010)*, Haifa, Israel, June 2010.

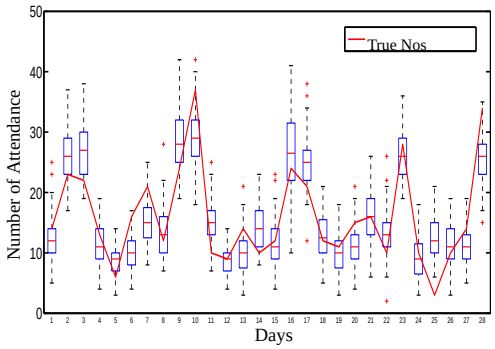
[28] Antonio Torralba, Kevin P. Murphy, and William T. Freeman. Contextual models for object detection using boosted random fields. In *In NIPS*, 2004.

[29] Mark Schmidt, Kevin Murphy, Glenn Fung, and Rmer Rosales. Structure learning in random fields for heart motion abnormality detection. In *In CVPR*, 2008.

[30] Michael Schweinberger. Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496):1361–1370, 2011.



(a) DNR Vertex Model.

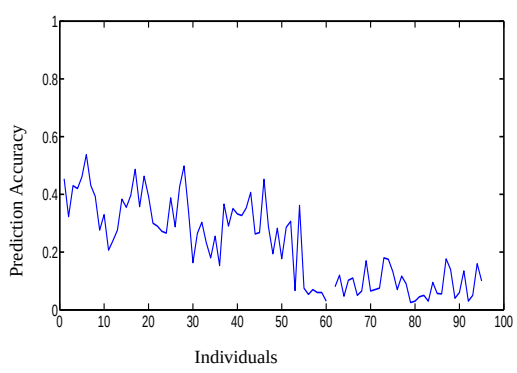


(b) Non-Homogeneous Model.

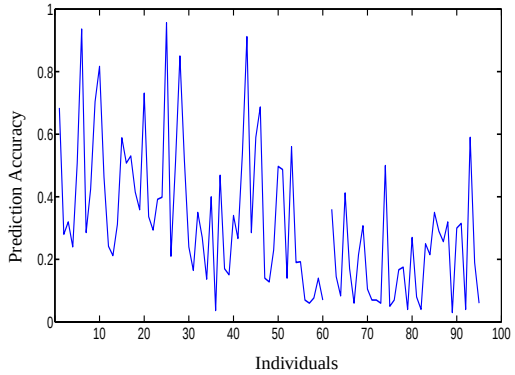
Figure 1: Variation on the number of predicted vertices (i.e. predicted number of individuals at beach) from Dynamic Network Regression-vertex model as a baseline model and Model 3 in table 3, a homogeneous model corresponding to threshold $\mathsf{r}_{\mathsf{th}} = 0.9$, using unconditioned information distance.

No.	Model	r_{th}	σ_p	AICc	CP (%)	CA(%)	VP (%)	EP (%)
1	UC-NH	0.9	na	2000.08	43.91	88.40	80.86	11.99
2	UC-H	0.8	na	1931.08	41.40	88.05	80.14	9.79
3	UC-H	0.9	na	1928.10	41.59	88.02	80.15	9.79
4	UC-H	1	na	1935.62	41.97	88.34	80.48	10.18
5	C-NH	9	na	2207.52	45.19	85.52	78.68	11.17
6	UC-NH	25	na	1687.4	38.04	89.09	80.44	9.52
7	UC-NH	1.1	na	1995.3	39.03	91.10	82.27	11.30
8	UC-NH	1.1	1.6	2013.4	45.21	89.13	81.68	13.26
9	UC-NH	0.8	1.7	2002.5	42.86	88.64	80.88	10.84
10	UC-H	1.1	1.95	1950.8	41.64	88.13	80.25	10.27
11	C-H	9	1.95	2013.4	31.79	86.14	76.92	5.48
12	C-H	6.5	1.95	2039.6	32.70	86.26	77.18	5.66
13	C-H	5.5	1.95	1987.7	33.57	86.14	77.22	6.31
14	C-NH	5.5	1.95	2182.5	41.33	88.31	80.31	10.35
DNR-V	na	na	na	2037.80	31.16	86	76.75	5.03

Table 3: Comparison performance in terms of AICc (14) score and prediction score for correct vertex prediction (VP) in (16), correct conditional prediction of people at beach(CP) in (15), people not at beach (CA) in (15), and conversation between people (EP) in (16) for models with homogeneity(H) and non-homogeniety(NH) assumption, graph structures(tree) using conditioned(C) in (12) and unconditioned(UC) in (7) information distances. Models correspond to threshold r_{th} values in Beach data (resulting in Late latent nodes) and standard deviation σ_p for Gaussian prior on parameters. Dynamic Network Regression-vertex (DNR-V) baseline is substantially outpredicted by all tree models in getting the labels correct.



(a) DNR Vertex Model.



(b) Non-Homogeneous Model.

Figure 2: Vertex prediction accuracy in terms of conditional prediction score of Dynamic Network Regression-vertex(DNR-V) model and Model 3 in table 3, a homogeneous model with graph structure corresponding to threshold $\mathsf{r}_{\mathsf{th}} = 0.9$ using unconditioned information distance

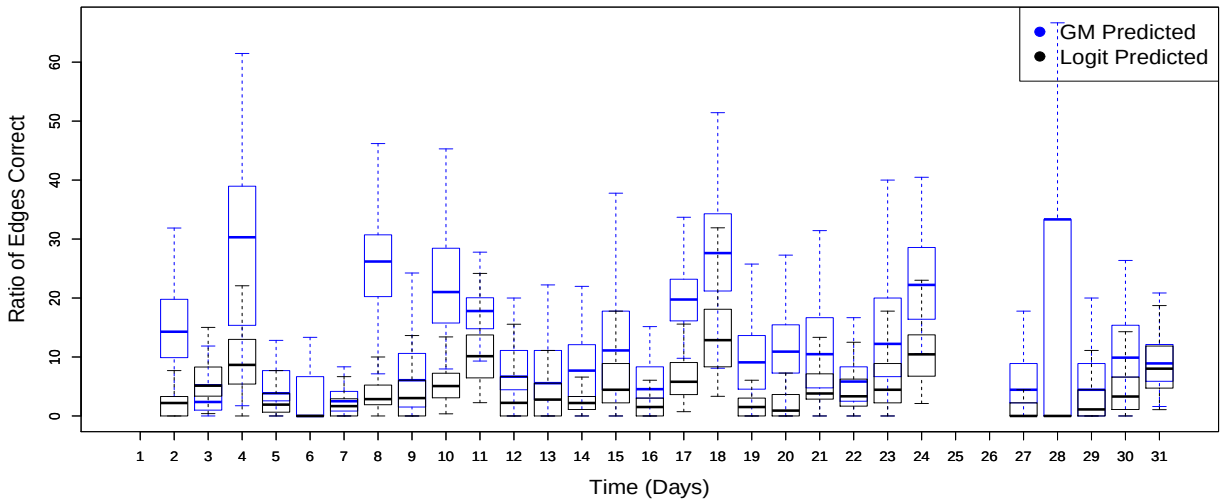


Figure 3: Correctly predicted edges for beach data (i.e. predicted conversation between individuals at beach on a given day)in (% using EdgePredAccuracy in (16)) using the vertex sets predicted by Model 8 (in table 3), a non-homogeneous model assuming a Gaussian prior (with $\sigma_p = 1.6$) on parameters $\boldsymbol{\theta}$ with structure corresponding to $\mathsf{r}_{\mathsf{th}} = 1.1$ learnt using unconditioned information distance in (7) with baseline, Dynamic Network Regression Vertex model.

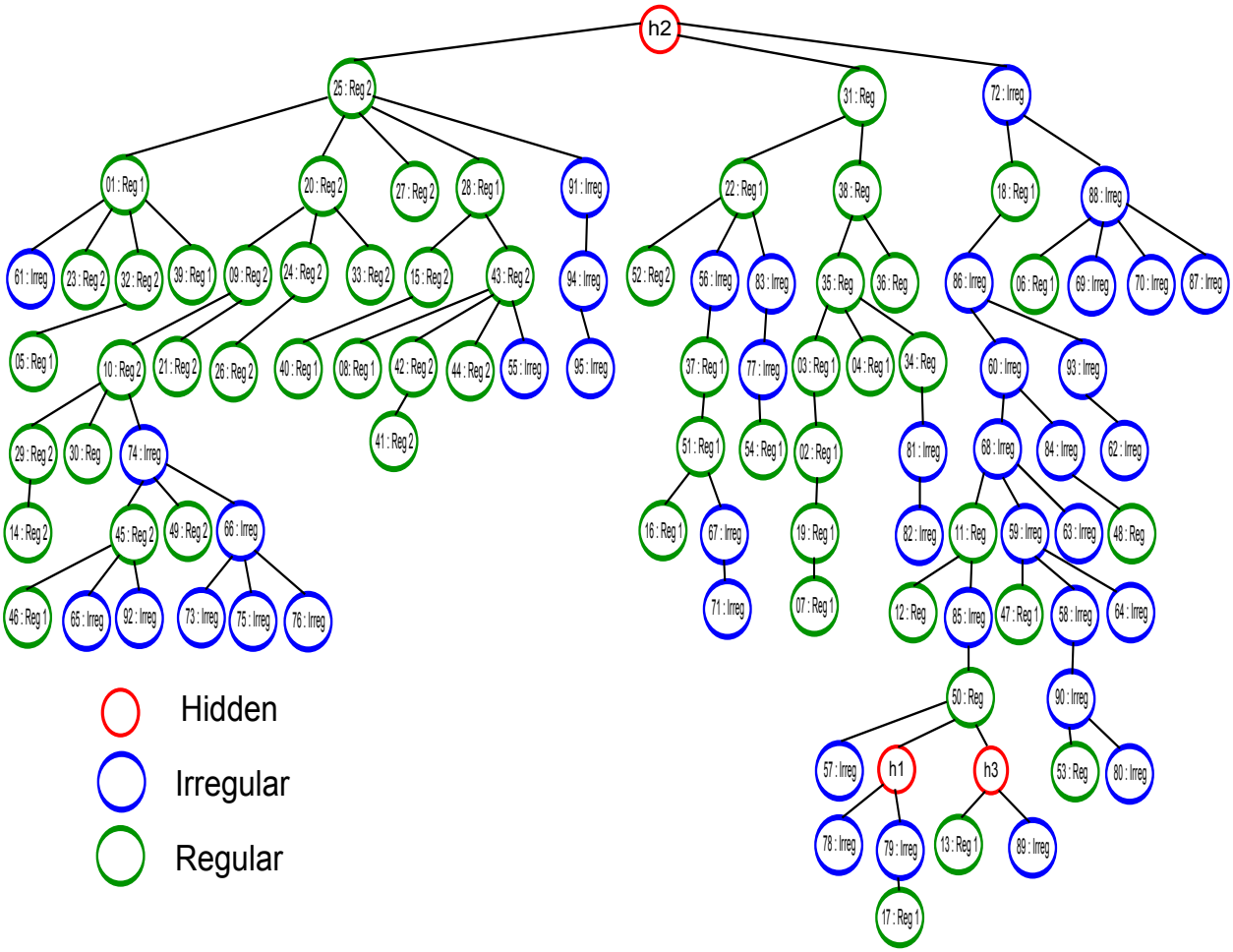


Figure 4: Latent tree graph structure corresponding to $r_{th} = 25$ learnt using unconditioned information distance in (7) for beach data.

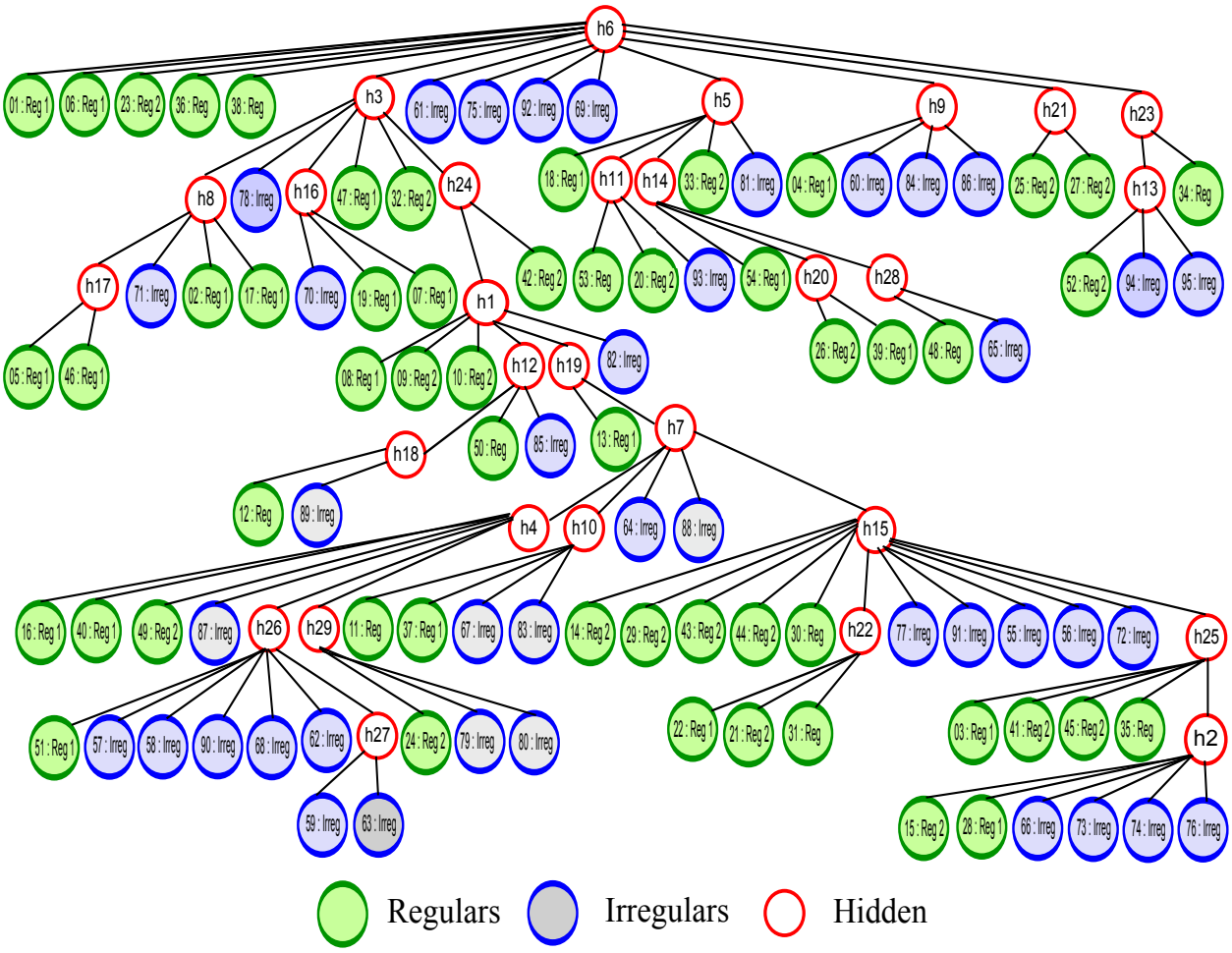


Figure 5: Latent tree graph structure corresponding to $r_{th} = 9$ learnt using conditioned information distance in (12) for beach data.

A Known Issues with General TERGMs

The general TERGM framework, like the corresponding cross-sectional framework (*exponential random graph models* or ERGMs) can parametrize an extremely broad class of models, not all of which are statistically or computationally tractable. Indeed, poorly chosen ERGMs are known to have issues of instability, sensitivity, degeneracy, and scalability, which are often intractable problems for many applications of interest [for a current review, see 30]. Further [5, 23] have shown that the general TERGM case includes model families with similar issues of instability, sensitivity, degeneracy, and scalability, but also that under certain conditions the temporal structure stabilizes these models; this work naturally leads to the assumptions underlying dynamic network regression.