Modeling Temporal Evolution and Multiscale Structure in Networks

Tue Herlau TUHE@DTU.DK
Morten Mørup MMOR@DTU.DK
Mikkel N. Schmidt MNSC@DTU.DK
Technical University of Denmark, DTU Compute, 2800, Richard Pedersens Plads, Lyngby, Denmark

Abstract

Many real-world networks exhibit both temporal evolution and multiscale structure. We propose a model for temporally correlated multifurcating hierarchies in complex networks which jointly capture both effects. We use the Gibbs fragmentation tree as prior over multifurcating trees and a change-point model to account for the temporal evolution of each vertex. We demonstrate that our model is able to infer time-varying multiscale structure in synthetic as well as three real world time-evolving complex networks. Our modeling of the temporal evolution of hierarchies brings new insights into the changing roles and position of entities and possibilities for better understanding these dynamic complex systems.

1. Introduction

Complex networks play a central role in the study of many complex systems. Most networks in practically all fields of research are dynamic. In biology temporal organizations arise in protein networks by mechanisms as diverse as their evolutionary history and cell cycle dynamics (Pastor-Satorras et al., 2003), and the functional connectivity of the brain changes during maturation and fluctuate spontaneously in awake rest (Dosenbach et al., 2010; Fox & Raichle, 2007). In social networks ties between individuals play different roles over the course of time: People move, change workplace as well as interests and these changes in turn affect how we interact with each other (Doreian, 1997), and global trends affect the fabric of society at the macro-scale. While networks traditionally have been modelled as static aggregated graphs, their temporal evolution carries important information about the structure of the system that is missed when considering only the time-aggregated network of interactions (Holme & Saramäki, 2012; Perra et al., 2012). Figure 1 illustrate some of the temporal effects presently considered.
In this work we explore two modelling assumptions:

Networks are organized at multiple scales Studies in network science suggest networks often exhibit multi-scale organization (represented as hierarchies over vertices), in which vertices are divided into groups which are further divided into subgroups. It has been shown that multiscale organization of networks account well for various network statistics such as scale-invariance, short paths lengths and a high degree of clustering (Clauset et al., 2008; Fortunato, 2010; Ravasz & Barabási, 2003), and by modeling networks in terms of hierarchies it is possible to account for the structures emerging at different scales (Herlau et al., 2012; Meunier et al., 2010; Ravasz et al., 2002; Roy & Teh, 2009; Roy et al., 2007; Sales-Pardo et al., 2007; Simon, 1962). Furthermore, cognitive science has long suggested that semantic knowledge is hierarchically organized, making it an attractive representation from an unsupervised learning perspective (Collins & Quillian, 1969)

Networks are temporally organized The study of dynamic networks has been at the forefront of statistical mechanics for more than a decade, focusing on network models which can take network growth into account. Common growth models have focussed on popularity (Barabási, 1999) or on how node similarity (McPherson et al., 2001; Watts et al., 2002) can explain the emergence of particular scaling properties ubiquitous in real networks (Dorogovtsev et al., 2000; Redner, 1998). Recent research has pointed out temporal evolution at two levels: The *connectivity-level* at which vertices enter or leave the system (i.e. capturing network growth), and at the *interaction-level*

Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28.

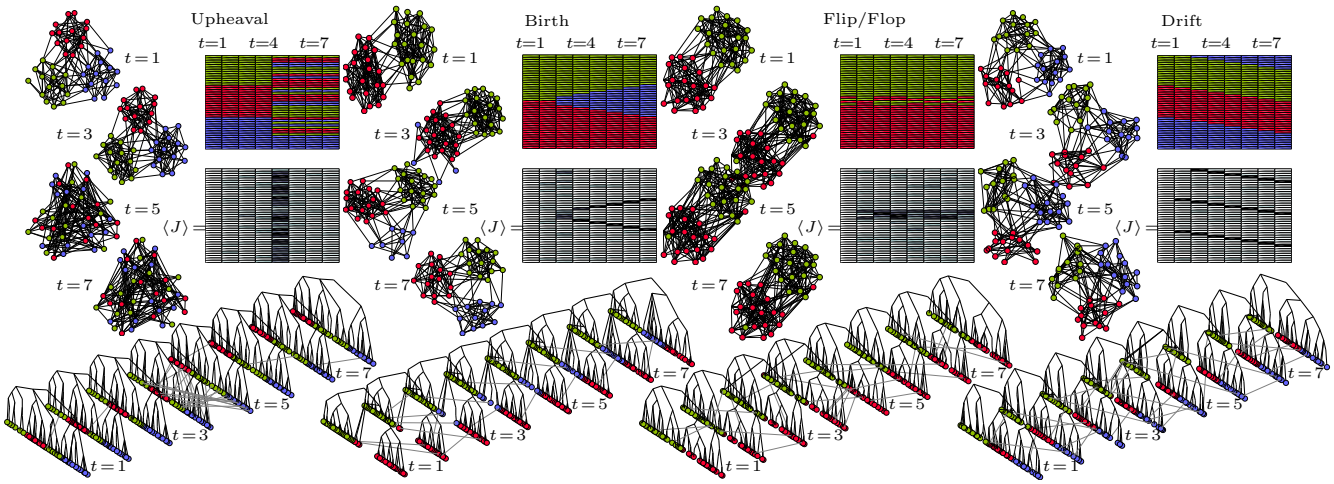

Figure 1. Four examples of temporal effect: Upheaval,network is reconfigured, Birth, where a new community emerges, Flip-Flop where a few members jump between communities and Drift where vertices consistently drift between communities. For each is shown snapshots of the actual network for times $t = 1, 3, 5, 7$ *(left)*, the true community structure *(top,right)*,average of inferred change point matrix $\langle J \rangle$ *(middle,right)* and the inferred temporal hierarchical structure *(bottom)* where movement of vertices between time slices is indicated by gray lines. As can be seen from $\langle J \rangle$ THRM recover each type of event (see section 4).

capturing processes whose timing and duration takes place at a shorter scale, typically reflecting interaction between the units forming the network (Albert et al., 1999; Holme & Saramäki, 2012; Perra et al., 2012). These effects can often be motivated on purely physical grounds (networks between physical entities do not form spontaneously), and compiling network datasets often involve the explicit removal of temporal information by edge aggregation or windowing (Holme & Saramäki, 2012).

Motivated by these two properties we propose the temporal hierarchical relational model (THRM) which describe the data by a temporally evolving multifurcating hieararchy. The THRM model assumes the temporal network data are organized in epochs indexed by t such that each epoch corresponds to the network observed at that time point.

At each epoch the multiscale relational structure is modelled using a hierarchical latent structure, and as a prior over hierarchies the Gibbs fragmentation tree (GFT) (McCullagh et al., 2008; Schmidt et al., 2012) is used.

Changes at the *connectivity-level* such as birth and death of vertices is accounted for by allowing vertices to be present or not at different epochs, while at the *interaction-level* nodes can change role and function in the multiscale organization to capture temporal effects at a shorter timing. To accomplish this each vertex has a change-point model that governs when vertices changes their placement in the hierarchical structure, see figure 1 for examples of inferred change-point structures. The model supports two limits: If no vertices have change-points the model reduces to a single, shared hierarchy for all vertices, and if all vertices have change-points at each epoch there is no temporal ordering in the hierarchical structure of each network.

In between these two extremes the model captures temporally evolving hierarchical network structures, sharing statistical strength between time epochs such that the closer two time instances are the more correlated their hierarchical structure will be. Inference in the model is performed using a distributed Gibbs sampler.

1.1. Related work

Time-evolving relational data has previously been considered with a variety of methods including multiway models (Peng & Li, 2010), information-theoretic approaches (Rosvall & Bergstrom, 2010), block-type models (Ishiguro et al., 2010) and change-point processes for edges (Vu et al., 2011). There is a growing literature on creating nonparametric priors for temporally evolving cluster structures using for instance connections between Dirichlet and Gamma processes (Lin et al., 2010) or fragmentation-coagulation processes (Teh et al., 2011). However, none of these works consider hierarchies as the organizational principle.

Our model most closely relate to existing Bayesian approaches to the modeling of hierarchical structure in networks, however these past models have not consid-

ered temporal evolution. In (Clauset et al., 2008) a network model with uniform prior over binary trees is proposed where the probability of generating an edge between two nodes is specified by a parameter at the level of their nearest common ancestor. In (Roy et al., 2007) each edge in the binary tree includes a weight defining the extend to which the network complies with the split, and in (Roy & Teh, 2009) the Mondrian process is proposed which for bipartite networks randomly bisects the nodes of the two modes until a stopping criterion is met, and parameters are then assigned to model the probability of links between each of the resulting pairs defined by the bisections. In (Herlau et al., 2012) a uniform prior over multifurcating trees is proposed where the leafs of the tree terminate at clusters generated from a Chinese Restaurant Process (CRP). In (Schmidt et al., 2012) the Gibbs fragmentation tree process is used as prior over trees when modeling relational data (in the following denoted HRM). The THRM currently proposed can be considered an extension of this framework to the modeling of time-evolving networks.

2. Methods

Before introducing our proposed model, we briefly review the infinite relational model (IRM). We then introduce relevant results for Gibbs fragmentation trees (GFT) and show how they can model networks in the hierarchical relational model (HRM). Using properties of Gibbs fragmentation trees we introduce temporal correlation to the HRM, leading to the proposed Temporal Hierarchical Relational Model (THRM).

2.1. Network models

As the simplest case, consider a simple graph of n vertices represented by the binary adjacency matrix $\boldsymbol{A}$ such that $A_{ij} = 1$ iff. there is an edge between vertex i and j. In the stochastic blockmodel (Holland et al., 1983) it is assumed that each vertex is assigned to one of k labelled communities, and the probability of an edge between vertex in community μ and community ν is given by the number $\eta_{\mu\nu}$. In a non-parametric Bayesian formulation (the IRM (Kemp et al., 2006)) the number of clusters is a priori unbounded and inferred from data.
Letting $z_i = \mu \in \{1, \ldots, k\}$ indicate the assignment to clusters, the IRM become

$$\begin{aligned} \mathbf{z} &\sim \mathrm{CRP}(\vartheta) && \textit{Assigment to clusters,} \\ \eta_{\mu\nu} &\sim \mathrm{Beta}(\eta_0^+, \eta_0^-) && \textit{interactions,} \\ A_{ij} &\sim \mathrm{Bernoulli}\left(\eta_{z_i z_j}\right) && \textit{links.} \end{aligned}$$

where $\mathrm{CRP}(\vartheta)$ is the Chinese Restaurant Process (Kemp et al., 2006) with parameter ϑ and Beta is the beta distribution. Intuitively the infinite relational model divides the adjacency matrix $\boldsymbol{A}$ into *blocks* (see figure 2 *top,left*), and each block is assigned a number $\eta_{\mu\nu}$ that determines the probability of forming edges on that region of the adjacency matrix (see figure 2). By conjugacy the η-parameters may be integrated out

$$p(\boldsymbol{A}|\boldsymbol{z}) = \prod_{\mu \leq \nu} \frac{\mathrm{B}\left(N_{\mu\nu}^+ + \eta_0^+, N_{\mu\nu}^- + \eta_0^-\right)}{\mathrm{B}\left(\eta_0^+, \eta_0^-\right)}, \qquad (1)$$

where $N_{\mu\nu}^+ = \sum_{ij} A_{ij}\delta_{\mu,z_i}\delta_{\nu,z_j}$ (and similar for $N_{\mu\nu}^-$ with A_{ij} replaced by $(1 - A_{ij})$) are called the *pseudo counts*, indicating the number of links (or nonlinks) in each block. $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$ is the beta function.

2.2. Multiscale Network Modelling

Multiscale network structure is naturally modelled using hierarchies over vertices: The root of the hierarchy corresponds to the coarsest resolution level indicating the large-scale structure, and moving further down the hierarchy corresponds to finer-scale structure between the fewer vertices partaking in the sub-hierarchies An intuitive view of this process is illustrated in figure 2, where the construction corresponds to modelling the entire network (at the top-level) by an IRM model, and then introducing new IRM models on the clusters along the diagonal in a recursive fashion. The construction is easiest formalized using notation from the GFT prior (McCullagh et al., 2008) reviewed below.

Gibbs Fragmentation Tree Prior Consider a rooted multifurcating tree T_B with leaf-set B. The leafs will later corresponds to vertices in the network so we write $n = |B|$. Each vertex $b \in T_B$ can be identified with the set of leafs off the subtree rooted at b, T_b, so without loss of generality we write $T_B = \{\text{leafs } T_b | b \in T_B\}$. Notice the collection contains B (corresponding to the root) and $|B|$ singleton sets corresponding to the leafs and is also called a *fragmentation* (McCullagh et al., 2008) of B, see figure 2 *(top,right)* for an illustration of a fragmentation of the set $\{A, B_1, B_2, C_1, C_2, D, E\}$.

For any subset A of B, consider the new tree T_A obtained by removing all leafs not in A, and then removing all vertices of degree 1. We call this operation the *projection* onto A written $\mathrm{proj}_A\ T_B$. In set notation it is given as $T_B \mapsto T_A = \{b \cap A \mid b \in T_B, b \cap A \neq \emptyset\}$. In figure 2 *(bottom,right)* is seen two projections onto the sets A, B_1, C_1, D, E and A, B_2, C_2, D, E. Also, for a vertex $b \in T_B$ such that $|b| \geq 2$, let $b_1, b_2, \ldots, b_k$ be

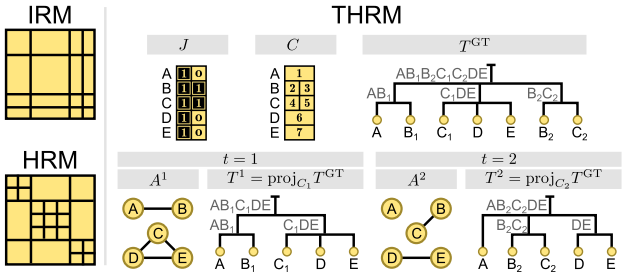

Figure 2. Left panel: Illustration of block-structure induced by IRM and HRM models. *Right panel:* The THRM for a system of 5 vertices at 2 time points where vertices B and C change hierarchical organization between time $t = 1$ and $t = 2$. In the middle is shown the change-point matrix J and the corresponding unique temporal states C, and further to the right the giant tree T^{GT} having the unique temporal states as leafs. At the bottom is shown the induced hierarchies at time $t = 1$ and $t = 2$ obtained by projecting onto the temporal states present at those epochs.

b's k children in T_B and $n_i = |b_i|$ their size. Defining the partition $\pi_b = \{b_1, \ldots, b_k\}$, we say T_B *fragments* b into π_b.

A random fragmentation model is a family of distributions over GFTs defined for all finite $B \subset \mathbb{N}$ having the following properties: *exchangeable*, meaning the distribution of T_B is invariant to labelling of B, *Markovian*, meaning for all partitions $B = B_1 \cup B_2 \cup \cdots \cup B_k$, the k tree $\{\text{proj}_{B_i} T_B\}$ are independently distributed as $T_1, \ldots, T_k$, and finally *consistent*, meaning for all $A \subset B$, $\text{proj}_A T_B$ is distributed as T_A.

McCullagh et al. (McCullagh et al., 2008) show all random fragmentation models have a representation as $p(T_B) = \prod_{b \in T_B} q(b_1, b_2, \ldots, b_k)$ and while the exact normalization is not given in the reference, for all but a few degenerate cases q takes the form

$$q(n_1, n_2, \ldots, n_k) = \frac{g_{k,\alpha}(\beta)}{g_{n,1}(\beta) - g_{n,1}(-\alpha)} \prod_{i=1}^{k} g_{n_i,1}(-\alpha) \tag{2}$$

for prior parameters $0 \leq \alpha < 0, \beta < -2\alpha$ (multifurcating trees with arbitrary block number) or $\alpha < 1, \beta = -2\alpha$ (binary trees). We have introduced the notation $g_{k,\alpha}(\beta) = \prod_{i=2}^{k-1}(\beta + i\alpha)$ and for leafs we simply have $q(1) = 1$. The two-parameter fragmentation process given by eq. (2) is denoted by $\text{GFP}(\alpha, \beta)$.

The Hierarchical Relational Model Returning to relational modelling, we can formalize the construction as follows. Consider again a network $\boldsymbol{A}$ and assume the n vertices $B = [n]$ are arranged in a GFT T_B.

At the root, T_B fragments B into a partition $\pi_B = \{B_1, B_2, \cdots, B_k\}$. Defining $\boldsymbol{z}$ by $z_i = \mu$ iff. $i \in B_\mu$ induces a block-structure on $\boldsymbol{A}$. Since each of the sets B_μ corresponding to diagonal elements in the block-structure (communities) these are further fragmented by T_{B_μ}. The construction can be applied recursively to each of these blocks creating a recursive refinement of the block-structure of $\boldsymbol{A}$. The construction continues over all vertices of T_B, see figure 2 *(bottom,left)*. Under this model the likelihood becomes

$$p(\boldsymbol{A}|\boldsymbol{\eta}, T) = \prod_{b \in T_B,\ |b| \geq 2} f_b(\boldsymbol{A}, \boldsymbol{\eta}^b, b). \tag{3}$$

Assume T_B fragments b into $\pi_b = \{b_1, \ldots, b_k\}$ and let μ, ν denote indices of each fragment. Using the Bernoulli likelihood we obtain the following form,

$$f_b(\boldsymbol{A}, \boldsymbol{\eta}_b, \pi_b) = \prod_{\substack{b_\mu, b_\nu \in \pi_b \\ \mu < \nu}} \prod_{i \in b_\mu, j \in b_\nu} \text{Bernoulli}(A_{ij}|\eta^b_{\mu\nu}),$$

where $\eta^b_{\mu\nu} \sim \text{Beta}(\eta_0^+, \eta_0^-)$.

2.3. Temporal Hierarchical Relational Model

Returning to temporal modelling, assume we are given a set of coordinates (i, t) (*vertex observations*) indicating vertex i was participating in the network at epoch t. All vertex observations are collected in a binary lattice $\mathcal{S}$, ie. if $\mathcal{S}_{it} = 1$ vertex i was present at epoch t and otherwise $\mathcal{S}_{it} = 0$. $\mathcal{S}_{it}$ is assumed to be known.

Next we encode if vertices between epochs t and $t+1$ may change their hierarchical organization. This is done using the binary change-point matrix J_{it} (*jumps*), defined for vertex observations (it) where $\mathcal{S}_{it} = 1$, with the interpretation that if $J_{it} = 0$ vertex i do not change hierarchical organization between time t and $t + 1$. In this case we say vertex i do not change *temporal state*, and if $J_{it} = 1$ vertex i is said to change temporal state. It is assumed different vertices are always in different temporal states and an exited temporal state cannot be reentered. We use the convention $J_{it} = 1$ if vertex i is not present at $t - 1$.

The collection of all temporal states is denoted by C. See figure 2 for illustration of J and C for a system of 5 vertices A, B, C, D, E, where vertex B and C change temporal states between epoch 1 and 2 giving the unique temporal states $A = 1, B_1 = 2, B_2 = 3, C_1 = 4, C_2 = 5, D = 6$ and $E = 7$.

To model the correlated hierarchical organization, all unique temporal states of the vertices according to their change-points are organized in a single hierarchy, the *giant tree*, denoted by T^{GT} (see figure 2 *(top,right)*). By projecting this tree to the set of tem-

poral states at each epoch, we arrive at the hierarchical organization for each observed network which can be modelled by an ordinary HRM. In figure 2 *(bottom,right)* we illustrate projections onto epoch $t = 1$ (states A, B_1, C_1, D, E) and $t = 2$ (A, B_2, C_2, D, E) corresponding to vertices B and C changing hierarchical organization.
For simplicity we consider simple life spans where each vertex has a single birth and death.

To formalize the construction, again let n denote the total number of vertices and T the total number of epochs. The network at epoch t is indicated by $\boldsymbol{A}^t$ (forming a simple graph) and all T networks simply by $\boldsymbol{A}$. Notice it is possible that no single vertex is present at all epochs, and that all networks contain less than n vertices.
Define J_{it} for vertex observations (it) st. $\mathcal{S}_{it} = 1$ as

$$J_{it} \sim \begin{cases} 1 & \text{if } \mathcal{S}_{i,t-1} = 0 \\ \text{Bernoulli}(\gamma) & \text{otherwise} \end{cases} \tag{4}$$

where γ is a parameter affecting the rate of change. The change-point deterministically give rise to the unique temporal states of all vertices: Different vertices $i \neq j$ are always in different temporal states: $c_{it_1} \neq c_{jt_2}$, temporal states change ($c_{it} \neq c_{i,t+1}$) if and and only if $J_{ij} = 1$, and finally once a state is exited it can never be reentered later (if $c_{it_1} = c_{it_2}$ and $t_1 < t < t_2$ then $c_{it_1} = c_{it}$).

Let $C = \cup_{it}\{c_{it}\}$ be the set of unique temporal states. Notice $|C| = |J|$. We let the elements in C form the leafs of a hierarchy T_C^{GT} (i.e., the *giant tree*) distributed as a $\text{GFT}(\alpha, \beta)$. It is this hierarchy that induces the hierarchies used for each epoch simply by projecting the giant tree onto the temporal observations at that epoch. Formally, defining $C_t = \{c_{it} |\ i \text{ st.} \mathcal{S}_{it} = 1\}$, the hierarchy at time epoch t is found as $\text{proj}_{C_t} T_C^{\text{GT}}$, see figure 2. Conditional on these hierarchies we simply model each network according to a HRM. Generatively the model becomes

$$\begin{aligned}
\gamma &\sim \text{Beta}(\gamma_0^+, \gamma_0^-) && \textit{Change rate,} \\
J_{it}|\mathcal{S} &\sim \text{According to eq. (4)} && \textit{change points,} \\
c_{it} &= \text{Deterministically given } J && \textit{giant tree leafs,} \\
T_C^{\text{GT}} &\sim \text{GFT}(\alpha, \beta) && \textit{giant tree,} \\
T^t &= \text{proj}_{C_t}\, T_C^{\text{GT}} && \textit{tree at each epoch,} \\
\boldsymbol{A}^t|T^t &\sim \text{According to eq.(3)} && \textit{edges.}
\end{aligned} \tag{5}$$

To put the model in words, we first generate the change points independently, thus determining when vertices may change roles in the network. Next all unique temporal states are arranged in a giant Gibbs fragmentation tree, and by projecting this tree onto each epoch the networks at each epoch is conditionally independently generated using the HRM.
The hyperparameters were selected as $\eta_0^+ = \eta_0^- = \gamma_0^+ = \gamma_0^- = 1$ and $\alpha = \beta = 0.5$ and $\vartheta = 1$ reflecting uniform beta distributions. Preliminary experiments suggest the results are robust to these choices.

Properties of the model Consider a family of models p_n of n parameters $x_1, \ldots, x_n$. Such a family is *simply exchangeable* if it is invariant under relabelling and *marginally consistent* if

$$\sum_{x_{n+1}} p_{n+1}(x_1, \ldots, x_n, x_{n+1}) = p_n(x_1, \ldots, x_n) \tag{6}$$

and a family of models having these properties is said to be *completely exchangeable* (McCullagh et al., 2008).

Qualitatively, marginal consistency means inference is not affected by knowing there exists additional data which is unobserved. This is particulary important for temporal models, since the number of observed vertices at each epoch may be very different, and we do not wish this fact alone to drive the marginal distributions. While there may be other ways to correlate hierarchical structure, we take marginal consistency to be a primary goal of a temporal hierarchical models in general and the primary justification for our choice of prior.

With respect to full exchangeability, one should not *hope* the model to be fully invariant under relabelling, since that would imply no causal structure across time. With these comments in mind we summarize the properties of our model in the following claims:

- **c1** invariance under relabeling of vertices.
- **c2** invariance under temporal translation of all vertex observations, $(it) \mapsto (i, t + \Delta t)$.
- **c3** marginal consistency when integrating over a single temporal observation (it), see eq.(6).
- **c4** correlation of induced hierarchies at epochs t and s decrease as $|t - s|$ increase.
- **c5** marginally distributed as the HRM at each epoch.

The last property implies if there are several epochs in a model, but we only observe one, it is distributed as a single HRM. Proofs of the claims can be found in the supplementary material.

2.4. Evaluation method

A principal quantitative way of evaluating network models is by link prediction (Clauset et al., 2008; Miller et al., 2009; Mørup & Schmidt, 2012). Consider a triplet (i, j, t) indicating a potential edge observation between vertex i and j at time t. Let

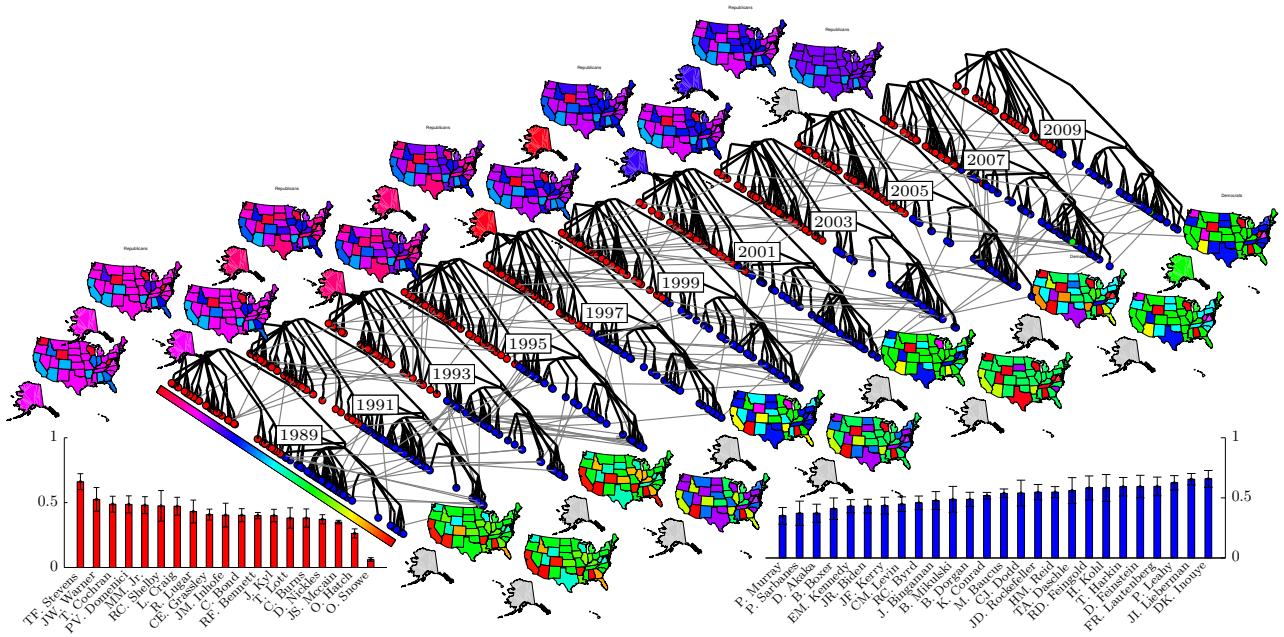

Figure 3. US Senate voting record data for 101st seating to 111th seating. In the middle is shown the temporal hierarchy for all senators, each senator corresponds to a leaf and party membership is indicated by blue/red as democrat/republican and "others" are green. To illustrate factions within parties, for each epoch t we restricted the giant tree to the democratic/republican senators, grouped senators within parties according to how they were fragmented by the giant tree and assigned each a color depending on the (horizontal) position of the group in the giant tree (see inserted colorbar). This was used to color their home state. Thus the colors on the US map do not indicate party membership, but illustrate factions within the two parties for republicans *(top,left)* and democrats *(bottom,right)*. The bottom histograms show the probability of changing position in the giant tree for each of the senators seated at least 8 epochs.

E^{mis} be a set of such triplets and $\overline{\boldsymbol{A}}$ the connectivity matrix with E^{mis} treated as missing. We consider both the *accuracy* and *log-loss* which for every $e=(ijt)\in E^{\text{missing}}$ is $\langle A^t_{ij}\eta^t_{ij}+(1-A^t_{ij})(1-\eta^t_{ij})\rangle_{p(\cdot|\overline{\boldsymbol{A}})}$ and $\langle A^t_{ij}\log\eta^t_{ij}+(1-A^t_{ij})\log(1-\eta^t_{ij})\rangle_{p(\cdot|\overline{\boldsymbol{A}})}$ respectivtely with $\eta^t_{ij}=p(A^t_{ij}=1|\cdot)$.

To overcome class imbalance E^{mis} consists of 5% of the edges in each epoch and a similar number of non-edges. We also include AUC score of the ROC characteristics for the test data (Clauset et al., 2008; Miller et al., 2009; Mørup & Schmidt, 2012).

3. Implementation

Inference in the model was performed by Gibbs sampling. Both the change rate γ and the $\boldsymbol{\eta}$-parameters for the networks at each epoch can be integrated out (Schmidt et al., 2012). This leaves only the giant tree T^{GT} and the change points J to be sampled since the other operations in eqn. (5) are deterministic. Sampling is done using two Gibbs-type operations.

Subtree regrafting A vertex in the giant tree $b\in T^{\text{GT}}$ is selected at random conditional on b not being a leaf. Denote by $T^{\text{GT}}_{\setminus b}$ the giant tree with b removed and T_b the subtree rooted at b. T_b is now inserted at all possible sites in T^{GT}, that is, (i) as a child to all vertices with at least two children and (ii) by selecting a vertex $a\in T^{\text{GT}}_{\setminus v}$ and replacing the subtree rooted at a with a new tree having T_a and T_b as it's two children, ie. $T_a\cup T_b\cup\{a\cup b\}$ in fragmentation-notation. There are as many potential moves of the first type as there are vertices which are not leafs in $T^{\text{GT}}_{\setminus b}$, and as many of the later type as there are vertices in $T^{\text{GT}}_{\setminus b}$. All moves of both types is considered as a single Gibbs update.

Change-point estimation The sampler first selects a temporal observation (it). The sampler now consider all transitions which can be obtained by setting $J_{it}=1$ and considering all moves of the previous type for the vertex $b=\{c_{it}\}\in T^{\text{GT}}$, and if $\mathcal{S}_{i,t-1}=1$ also the move obtained by letting $J_{it}=0$, ie. the state now inherits it's position in the giant tree from $c_{i,t-1}$. Again all possible moves are computed and sampled from.

Both of these moves were implemented using a two-stage procedure: First, the Gibbs move at the giant-tree level is translated into a set of moves on the projection of T^{GT} onto the temporal epochs. These moves may be computed independently, and the transition

probabilities are combined to form the set of transition probabilities relevant for operations on the giant tree.

When combining the transition probability in the giant tree (requiring computation of the likelihood, eq. (2)) and when computing the contributions to the likelihood at each subtree (eq. (3)), the changes to the likelihood can be reused according to the tree-structure: If b_1 is an ancestor of b_2 and b_2 is an ancestor of b_3, performing a change (for instance insertion of subtree) at b_2 requires changes at the level of b_1, but these changes can be reused when computing changes to the tree at level b_3. In our implementation the parallel computations at the level of epochs were performed using distributed computing resources.

Previous work in HRMs (Schmidt et al., 2012) only considered Metropolis-Hastings pruning and regrafting. To compare to the THRM, consider the total number of (new) states considered by the Gibbs sampler in one sweep. If there are $n = 200$ vertices observed over $T = 30$ time-steps and a change rate $\gamma = \frac{1}{3}$, there will be $\sim 200 \cdot 30/3$ leafs in the giant tree. Including internal vertices this give ~ 3000 vertices in total and considering only Gibbs sampling of leafs $(3000 + 1000)2000 = 8 \cdot 10^6$ regrafting steps at each sweep due to the 2 types of regraft moves (not considering regrafting of subtrees and changes in J). 100 sweeps is then $\sim 10^9$ regrafting operations, compared to about $\sim 10^6$ subtree regrafting operations considered by Schmidt et.al.(Schmidt et al., 2012). Despite these efforts, neither the IRM or THRM mix well on the larger datasets.

4. Results

p The synthetic datasets shown in figure 1 were created by planting each of the four cluster structures using a simple model with within-cluster edge density 0.9 and between-cluster edge density 0.1. All networks had $n = 42$ and $T = 8$. The sampler was evaluated for 80 iterations (random initialization, half of which were discarded as burn-in. Shown are the true cluster-membership matrices and inferred J-matrices averaged over 10 networks. For *upheaval* and *flip-flop* the average can be expected to be less clear due to different sets of vertices participating for different networks, but as can be seen the sampler correctly infers the timing and allows identification of the type of change.
In addition, we also analyzed three larger datasets.

Enron email network: The Enron email network(Cohen, 2009) consists of the email communication of 184 Enron executives. The email communication was binned into months giving 44 epochs corresponding to 1st November, 1998 to 1st July, 2002. For each epoch directed edges was added between executive i and j if i send j an email in that month. The resulting network was directed and all simulations were performed using directed versions of the models, ie. both using the upper and lower parts of the $\boldsymbol{\eta}$-matrix as individually defined entries. The edge-density was 2.3%.

US Senate voting records: The US senate voting records dataset (Lewis et al., 2010) covers voting records for 235 senators in the 101st congress (1989-1991, first two years of the presidency of G.W. Bush), to the 111th congress (2009-2011,first 2 years of the B.H. Obama presidency). Senators were included from their first registered votes to their last, and at each epoch the network was constructed by computing the Jaccard index between each senators binary yes/no voting record in the particular senate seating and including the top 20% values as edges.

NIPS author collaboration: The NIPS dataset (Roweis, 2002) covers co-authorship of the 1-12th NIPS conference. Edges indicate co-authorship and authors were included who participated in at least 3 conferences giving 360 authors. The edge-density was 0.9%.

As comparison we considered two ways for applying the IRM and HRM to temporal data. One was simply modelling each time epoch independently (Sliced), and the other was fixing cluster-assignments/hierarchy in the IRM/HRM across all epochs (Joint), forcing all temporal epochs to share structure while $\boldsymbol{\eta}^t$ is epoch specific. In the later case we treated all edges incident to vertices not present at the given epoch as unobserved to avoid bias. Finally as a base reference we considered a naive model (Past) which predicted edges based on them being present in the previous time epoch, $p(A^t_{ij} = A^{t-1}_{ij}) = 1$.

In table 1 is shown results of all three scores for the three considered datasets. For each entry the appropriate sampler was evaluated for 2000 Gibbs sweeps in case of the IRM and 100 sweeps for the hierarchical-based models. Half the samples were discarded as burn-in and 10 equidistant samples was used. Shown is the mean and standard deviation of the mean for four simulation, and for the naive model the standard deviation of the mean was computed based on splitting the (balanced) set of edges/non-edges into four.

Focusing on AUC, for Enron and NIPS all methods

Table 1. Results for the considered models on the Enron, US Senate and NIPS author collaboration networks.

		IRM (Sliced)	IRM (Joint)	HRM (Sliced)	HRM (Joint)	THRM	Past
Enron	Accur.	0.567(08)	0.515(10)	**0.630(09)**	**0.625(09)**	**0.631(10)**	–
	log-loss	−1.422(40)	−1.842(75)	−1.312(54)	−1.157(44)	−1.274(49)	–
	AUC	0.870(03)	0.823(12)	0.855(13)	**0.884(06)**	0.859(09)	0.742(03)
US Senate	Accur.	0.865(03)	0.577(02)	0.865(02)	0.840(03)	**0.872(03)**	–
	log-loss	−0.208(04)	−1.011(11)	−0.257(06)	−0.265(04)	−0.239(06)	–
	AUC	0.973(01)	0.955(01)	0.977(02)	0.972(02)	**0.981(02)**	0.822(02)
NIPS	Accur.	0.544(06)	0.502(01)	0.633(12)	0.658(08)	**0.685(10)**	–
	log-loss	−2.169(27)	−3.245(17)	−1.793(84)	−1.224(72)	−1.127(44)	–
	AUC	0.621(13)	0.633(23)	0.762(05)	0.891(18)	**0.908(12)**	0.708(10)

perform well, joint IRM perhaps slightly underperforming. The NIPS network shows more variation, with hierarchical methods outperforming IRM methods and THRM scoring higher than sliced HRM, possibly also better than joint HRM. Model-wise, the hierarchical models seem to perform more consistently than the IRM. The NIPS network is very sparse, and results are indicative that hierarchical methods are better suited for this type of data. This is of particular interest from the perspective of temporal network modelling, since in the very sparse limit shared information across epochs can be expected to have the largest effect. All models (except IRM on NIPS) outperform the naive link prediction method, indicating methods based on inferring structure carry good explanatory power.

In figure 3 is shown the temporal hierarchy for the MAP sample of the US Senate dataset. The giant tree is first laid out and then projected onto each epoch while keeping the leaf-locations fixed. Gray lines indicate movement across epochs and vertices are colored according to party. Republicans and democrats are consistently split, showing a strong tendency to vote similarly along these two factions.

To better illustrate the multi-scaled structure of voting patterns, for each epoch the set of democrats and republicans was fragmented according to the tree structure into sets of size less than 30. We used the set of senators mean location on the inserted color-bar to color-code their home states. The results indicate some consistency, for the projection of republicans notice California, Arizona and Utah Arkansas and Alabama seem to form a consistent pattern across all epochs contrasted to a faction consisting of Texas along with several central states. For democrats the CA,AZ,UT,AL,AB pattern seem to be present during the first three epochs and then become less consistent during the B. Clinton presidency.

To illustrate the possibility for the model to explain faction changes, we extracted the senators present during at least 8 epochs and plotted the fraction of times they change position in the tree as well as standard deviations. Care need to be taken for this result due to the multiple-comparison problem, but notice the high value for J. Lieberman (D) who became an independent democrat during the 111th senate. Mind this is not indicating lower or higher degree of similarity of voting pattern, but that the voting pattern is not stable.

5. Conclusion

We have proposed the temporal hierarchical relational model which incorporates two important aspects of complex networks: Temporal evolution and multiscale structure. The model is thereby able to express a number of important dynamic network effects such as the emergence of new factions or vertices changing from one faction to another at multiple scales. The model is invariant under relabeling of vertices, invariant under temporal translation of all coordinates, marginal consistent, and marginally distributed as the HRM at each epoch.

We further described how inference in the model can make use of distributed Gibbs sampling where computations are reused by exploiting computational aspects of the inferred hierarchical structure during inference. Experiments on real datasets demonstrate that hierarchical models perform on par or better than the IRM model, while the proposed THRM provides a novel dynamic description of the data.

References

Albert, Reka, Jeong, Hawoong, and Barabasi, Albert-Laszlo. Internet: Diameter of the World-Wide Web. 401 (6749):130–131, September 1999. ISSN 0028-0836.

Barabási, A. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999. ISSN 00368075.

Clauset, A., Moore, C., and Newman, M.E.J. Hierarchical

structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.

Cohen, William W. Enron Dataset, 2009. URL http://www.cs.cmu.edu/~enron/.

Collins, Allan M. and Quillian, M. Ross. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2):240–247, April 1969. ISSN 00225371.

Doreian, P. *Evolution of social networks*, volume 1. Routledge, 1997.

Dorogovtsev, S., Mendes, J., and Samukhin, A. Structure of Growing Networks with Preferential Linking. *Physical Review Letters*, 85(21):4633–4636, November 2000. ISSN 0031-9007.

Dosenbach, N.U.F., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., et al. Prediction of individual brain maturity using fmri. *Science*, 329(5997): 1358–1361, 2010.

Fortunato, S. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

Fox, M.D. and Raichle, M.E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9):700–711, 2007.

Herlau, T., Mørup, M., Schmidt, M. N., and Hansen, L. K. Detecting hierarchical structure in networks. *Proceedings of Cognitive Information Processing*, 2012.

Holland, Paul W., Laskey, Kathryn Blackmond, and Leinhardt, Samuel. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109 – 137, 1983.

Holme, Petter and Saramäki, Jari. Temporal networks. *Physics Reports*, 519(3):97–125, October 2012. ISSN 03701573.

Ishiguro, K., Iwata, T., Ueda, N., and Tenenbaum, J. Dynamic infinite relational model for time-varying relational data analysis. *Advances in Neural Information Processing Systems*, 23, 2010.

Kemp, Charles, Tenenbaum, Joshua B, Griffiths, Thomas L, Yamada, Takeshi, and Ueda, Naonori. Learning Systems of Concepts with an Infinite Relational Model. In *AAAI*, pp. 381–388, 2006.

Lewis, Jeff, Poole, Keith, and McCarty, Nolan. US Senate Dataset, 2010. URL http://www.voteview.com/downloads.asp.

Lin, Dahua, Grimson, Eric, and Fisher, John. Construction of Dependent Dirichlet Processes based on Poisson Processes. 2010.

McCullagh, Peter, Pitman, Jim, and Winkel, Matthias. Gibbs fragmentation trees. *Bernoulli*, 14(4):988–1002, 2008.

McPherson, Miller, Smith-Lovin, Lynn, and Cook, James M. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444, August 2001. ISSN 0360-0572.

Meunier, D, Lambiotte, R, and Bullmore, E T. Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, 4, 2010.

Miller, K.T., Griffiths, T.L., and Jordan, M.I. Nonparametric latent feature models for link prediction. *Advances in Neural Information Processing Systems (NIPS)*, pp. 1276–1284, 2009.

Mørup, Morten and Schmidt, Mikkel N. Bayesian community detection. *Neural Computation*, 24(9):2434–2456, 2012.

Pastor-Satorras, R., Smith, E., Solé, R.V., et al. Evolving protein interaction networks through gene duplication. *Journal of Theoretical biology*, 222(2):199–210, 2003.

Peng, Wei and Li, Tao. Temporal relation co-clustering on directional social network and author-topic evolution. *Knowledge and Information Systems*, 26(3):467–486, March 2010. ISSN 0219-1377.

Perra, N, Gonçalves, B, Pastor-Satorras, R, and Vespignani, A. Activity driven modeling of time varying networks. *Scientific reports*, 2:469, January 2012. ISSN 2045-2322.

Ravasz, E and Barabási, A L. Hierarchical organization in complex networks. *Physical Review E*, 67(2):26112, 2003.

Ravasz, E, Somera, A L, Mongru, D A, Oltvai, Z N, and Barabási, A L. Hierarchical organization of modularity in metabolic networks. *science*, 297(5586):1551–1555, 2002.

Redner, S. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems*, 4(2): 131–134, 1998.

Rosvall, Martin and Bergstrom, Carl T. Mapping change in large networks. *PloS one*, 5(1):e8694, January 2010. ISSN 1932-6203.

Roweis, Sam T. NIPS Dataset, 2002. URL http://www.cs.nyu.edu/~roweis/.

Roy, D M and Teh, Y W. The Mondrian Process. In *Advances in Neural Information Processing Systems*, volume 21, 2009.

Roy, D.M., Kemp, C., Mansinghka, V.K., and Tenenbaum, J.B. Learning annotated hierarchies from relational data. *Advances in neural information processing systems*, 19:1185, 2007.

Sales-Pardo, M, Guimera, R, Moreira, A A, and Amaral, L A N. Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences*, 104(39):15224, 2007.

Schmidt, M. N., Herlau, T., and Mørup, M. Nonparametric bayesian models of hierarchical structure in complex networks. *ArXiv*, 2012.

Simon, H A. The architecture of complexity. *Proceedings of the American philosophical society*, 106(6):467–482, 1962.

Teh, Yee Whye, Blundell, Charles, and Elliott, Lloyd. Modelling Genetic Variations using Fragmentation-Coagulation Processes, 2011.

Vu, Duy, Asuncion, Arthur, Hunter, David, and Smyth, Padhraic. Dynamic Egocentric Models for Citation Networks. In *ICML*, 2011.

Watts, Duncan J, Dodds, Peter Sheridan, and Newman, M E J. Identity and search in social networks. *Science (New York, N.Y.)*, 296(5571):1302–5, May 2002. ISSN 1095-9203.