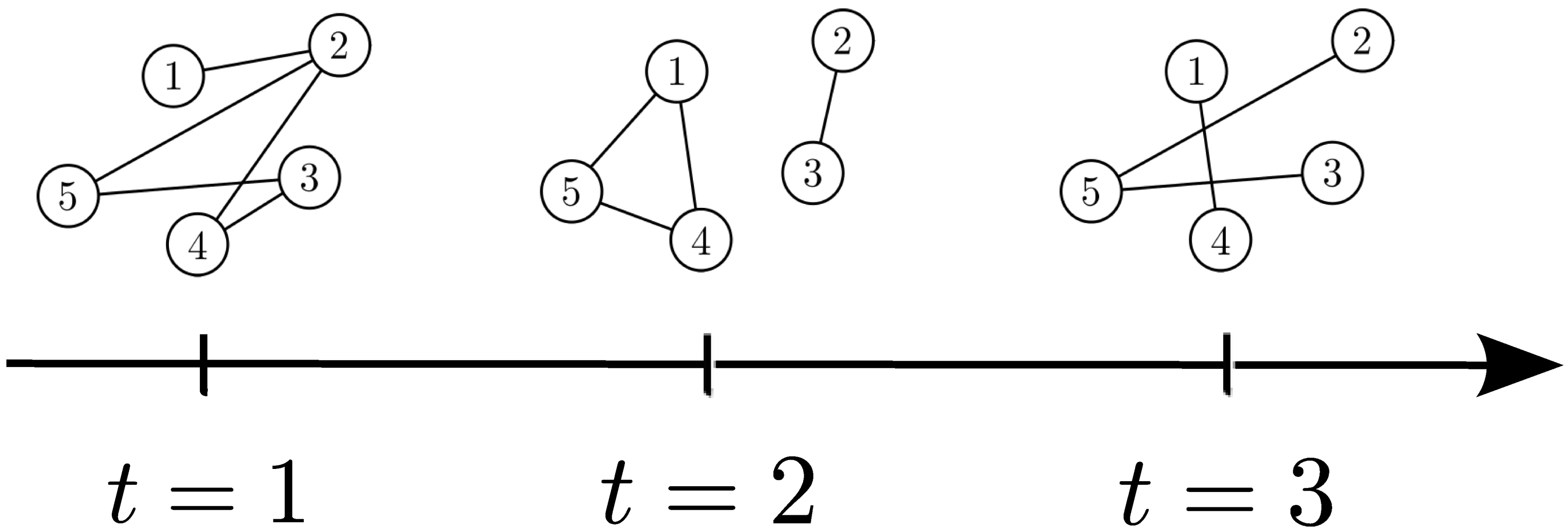
Dynamic Probabilistic Models for Latent Feature Propagation in Social Networks

Creighton Heaukulani and Zoubin Ghahramani
University of Cambridge

TU Denmark, June 2013

A Network

Dynamic network data record the link statuses in the network at T time points:



Generative Models of Networks

... encoded as link adjacency matrices $\{\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \ldots, \boldsymbol{Y}^{(T)}\}$:

$$\boldsymbol{Y}^{(t)} = \begin{pmatrix} \cdot & 1 & 0 & 0 & \\ 1 & \cdot & 1 & 1 & \\ 0 & 1 & \cdot & 0 & \\ 0 & 1 & 0 & \cdot & \\ & & & & \cdot \end{pmatrix}$$

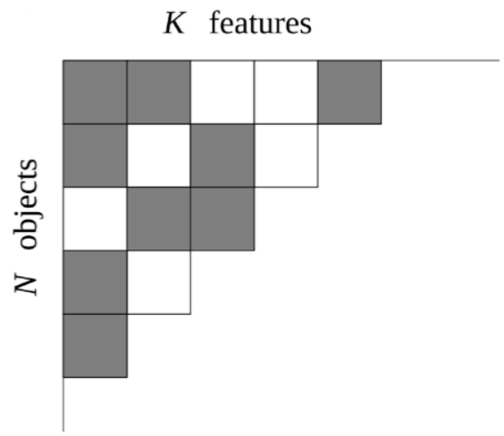
Example: *friendship statuses* in a social network

- $y_{ij}^{(t)} = 1$: actors i and j are friends at time t;
- $y_{ij}^{(t)} = 0$: actors i and j are not friends at time t.

Latent feature representations

Assume there are K latent features underlying the population.
Associate actor n with feature indicators $\boldsymbol{h}_n^{(t)} \in \{0,1\}^K$ at time t:

$\boldsymbol{H}^{(t)} =$



Interpretation: features represent unobserved *hobbies/interests*, e.g., if feature k represents "plays tennis", then

- $h_{nk}^{(t)} = 1$: actor n plays tennis at time t;
- $h_{nk}^{(t)} = 0$: actor n doesn't play tennis at time t.

Hidden Markov models

Hidden Markov models assume features evolve independently

$$h_{ik}^{(t)} \mid h_{ik}^{(t-1)} \sim \boldsymbol{Q}\big(\, h_{ik}^{(t-1)},\, h_{ik}^{(t)}\,\big)$$

where $\boldsymbol{Q}$ is a Markov transition matrix. Then the edges are conditionally independent given the latent features

$$y_{ij}^{(t)} \mid \boldsymbol{h}_i^{(t)}, \boldsymbol{h}_j^{(t)} \sim \mathrm{Bernoulli}\left[\sigma\left(\sum_{k\ell} h_{ik}^{(t)}\, h_{j\ell}^{(t)}\, v_{k\ell}\right)\right]$$

where $\boldsymbol{V}$ is a feature-interaction weight matrix

$$v_{k\ell} \sim \mathcal{N}(0, \sigma_H^2)$$

An Example

Consider the following example: feature $h_{ik}^{(t)}$ evolves as[a]

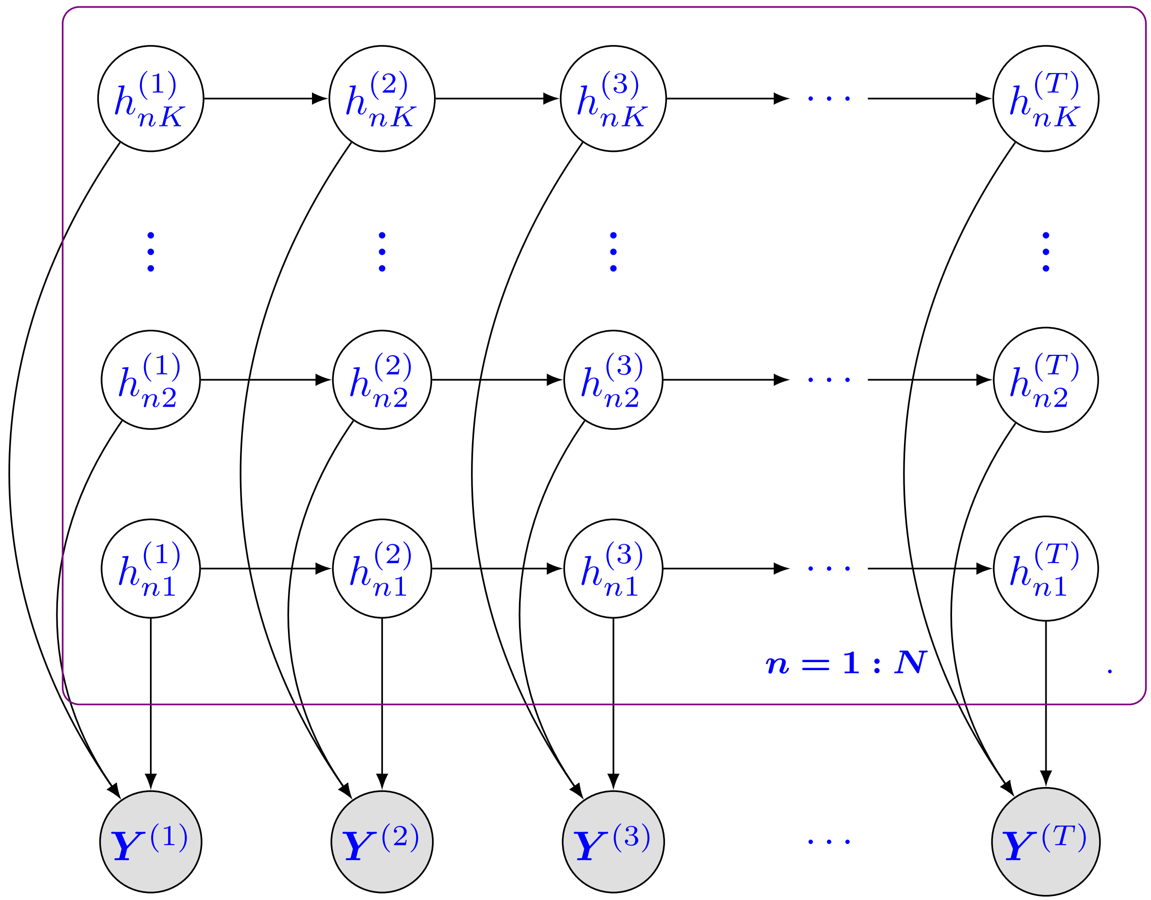
$$h_{ik}^{(t)} \mid h_{ik}^{(t-1)} \sim \text{Bernoulli}\left(a_k^{1-h_{ik}^{(t-1)}} b_k^{h_{ik}^{(t-1)}} \right)$$

- $a_k \in [0,1]$ controls the probability of feature k switching from off to on

- $b_k \in [0,1]$ controls the persistency of feature k in the off state

[a]Finite version of the DRIFT model from **?**.

Factorial HMM



Each feature evolves over time independently. The latent feature configuration at a given time step produces the observed network. See **?**.

Latent Feature Evolution

But consider the following:

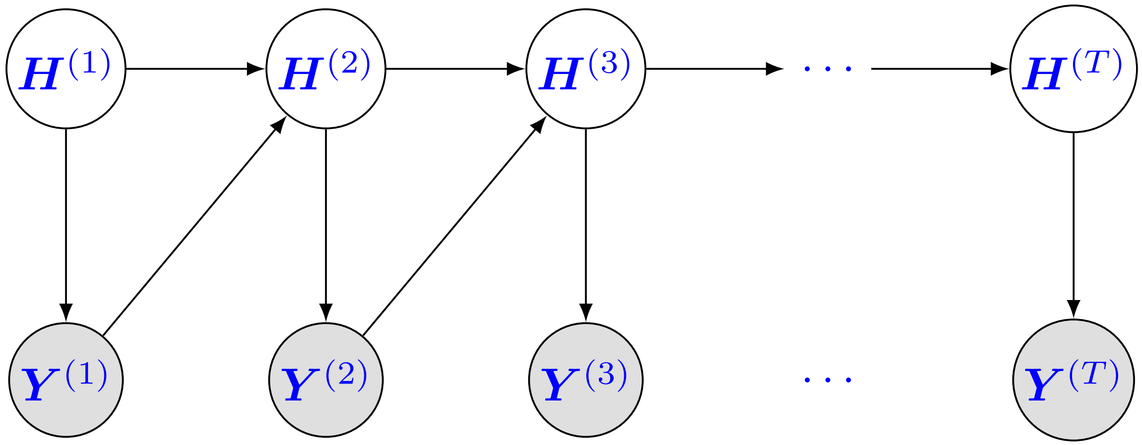
- If my friends enjoy playing tennis, I am likely to start playing tennis

- If a friend gets me to join the tennis team, then I am more likely to befriend other tennis players

We call this phenomenon *latent feature propagation*

Latent Feature Propagation

Want something more like this:



Network observations influence future latent features; information propagates between the observed and latent structures throughout the network over time

Latent Feature Propagation

We use the following model (**?**)

$$h_{ik}^{(t+1)} \mid \mu_{ik}^{(t+1)} \sim \text{Bernoulli}\left[\sigma\left(c_k\left[\mu_{ik}^{(t+1)} - b_k\right]\right)\right]$$

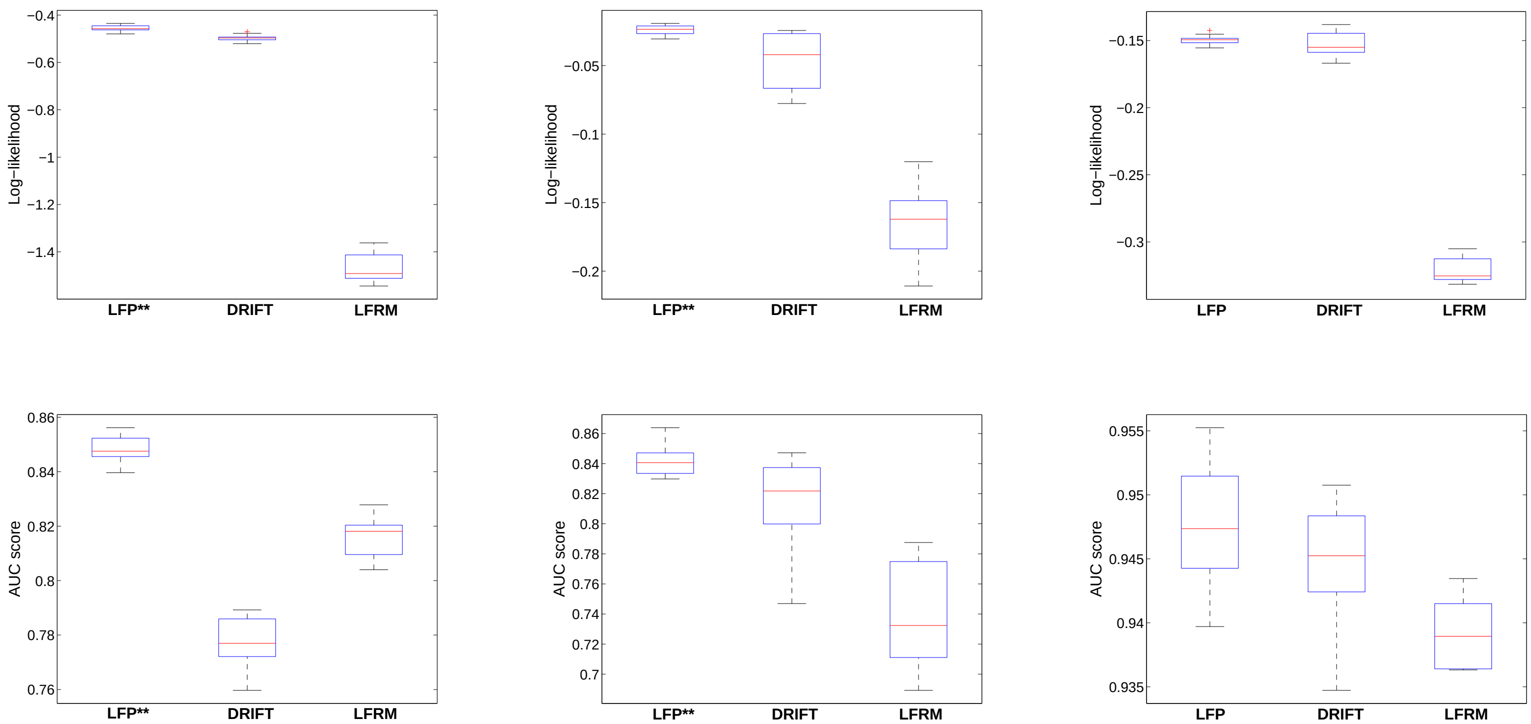
$$\mu_{ik}^{(t+1)} = (1-\lambda_i)h_{ik}^{(t)} + \lambda_i \frac{h_{ik}^{(t)} + \sum_{j \in \varepsilon(i,t)} w_j h_{jk}^{(t)}}{1 + \sum_{j' \in \varepsilon(i,t)} w_{j'}}$$

1. $\lambda_i \in (0,1)$: actor i's susceptibility to the influence of friends; $(1-\lambda_i)$ is the corresponding measure of social independence;
2. $w_i \in \mathbb{R}_+$: the weight of influence of person i;
3. $c_k \in \mathbb{R}_+$: a scale parameter for the persistence of feature k;
4. $b_k \in \mathbb{R}_+$: a bias parameter for feature k.

Latent Feature Propagation

- Inference by MCMC; Forward-Backwards Algorithm from **?**
- Datasets:
  - Simulated: $N = 50, T = 100, K = 10$.
  - NIPS: $N = 110, T = 17, K = 15$.
  - INFOCOM (**?**): $N = 78, T = 50, K = 10$.

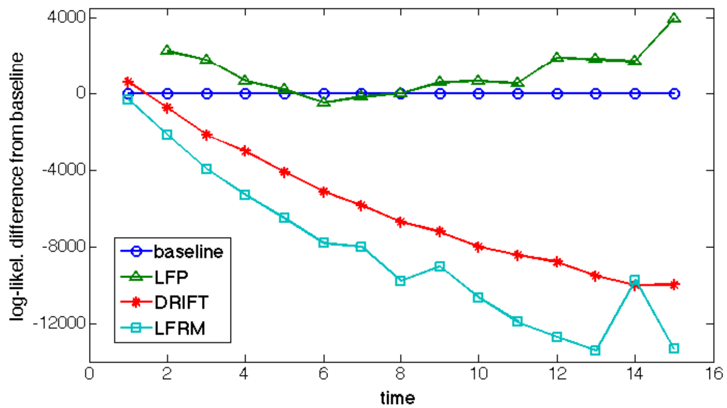
Prediction of Missing Links



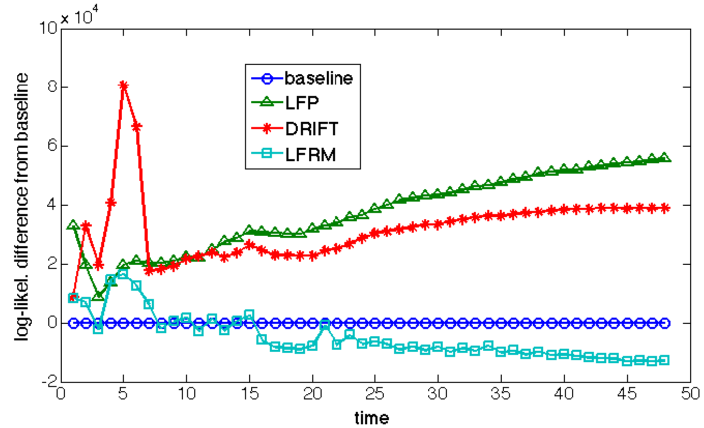
(d) Simulated data (e) NIPS dataset (f) INFOCOM dataset

Top: log-likel. of test edges. Bottom: AUC scores for classifying test edges. 10 repeats on different 20% hold-outs. Averaged over 300 samples. Significance indicated by $(\star\star)$ $(\alpha = 0.05)$.

Forecasting Future Networks



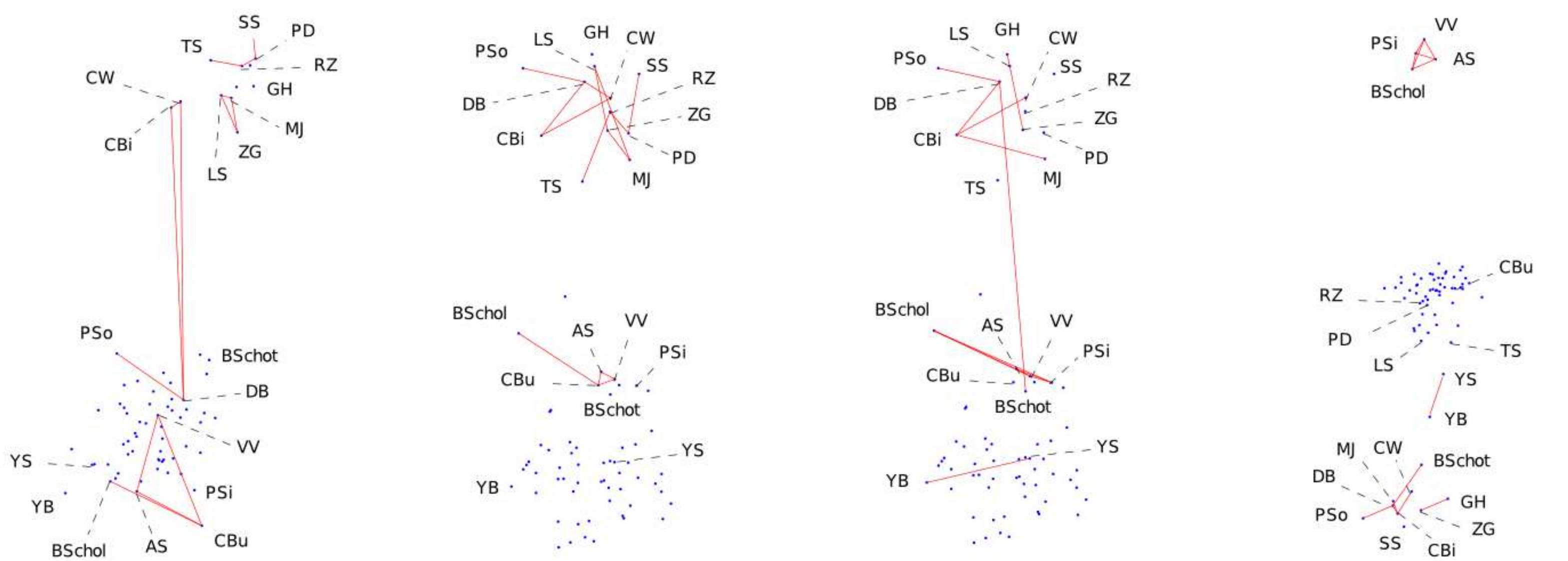
(g) NIPS dataset, $K = 15$



(h) INFOCOM dataset, $K = 10$

Forecasting a future unseen network. Differences from a naive baseline of the log-likelihoods of $\boldsymbol{Y}^{(t)}$ after training on $\boldsymbol{Y}^{(1:t-1)}$

Visualising Feature Propagation



(i) $\boldsymbol{H}_{1997}, \boldsymbol{Y}_{1997}$ (j) $\boldsymbol{H}_{1998}, \boldsymbol{Y}_{1997}$ (k) $\boldsymbol{H}_{1998}, \boldsymbol{Y}_{1998}$ (l) $\boldsymbol{H}_{1999}, \boldsymbol{Y}_{1998}$

Visualising feature propagation in the NIPS dataset ($K = 15$).

Visualising Feature Propagation

We can look at the research interests of a trio of sparsely linked authors between 1997 and 1998 who are nearby in latent space:

Author & year	Topics
Top cluster	"Prior Knowledge in Support Vector Kernels"
Bengio_Y, Singer_Y, *et al.*	"Shared Context Probabilistic Transducers"
Hinton_G, Ghahramani_Z	"Hierarchical Non-linear Factor Analysis and Topographic Maps"
Bishop_C, Williams_C, *et al.*	"Regression w/ Input-depend. Noise: A Gaussian Process Treatment"
Sollich_P, Barber_D	"On-line Learning from Finite Training Sets in Nonlinear Networks"
Barber_D, Bishop_C	"Ensemble Learning for Multi-Layer Networks"
Bishop_C, Jordan_M	"Approx. Posterior Distributions in Belief Networks Using Mixtures"

Visualising Feature Propagation

We can also examine the five authors with the largest inferred weights w_n and some of their research interests (1995 - 1999):

Author	w_n (relative)	Topics
Barto_A	1.55	reinforcement learning, planning algorithms
Rasmussen_C	1.32	Gaussian processes, Bayesian methods
Vapnik_V	1.29	SVMs, learning theory
Scholkopf_B	1.28	SVMs, kernel methods
Tresp_V	1.26	neural networks, Bayesian methods

References

Foulds, J., DuBois, C., Asuncion, A., Butts, C., and Smyth, P. (2011). A dynamic relational infinite feature model for longitudinal social networks. In *Proc. AISTATS*.

Ghahramani, Z. and Jordan, M. (1997). Factorial hidden Markov models. *Machine Learning*, 29:245–273.

Heaukulani, C. and Ghahramani, Z. (2013). Dynamic probabilistic models for latent feature propagation in social networks. In *Proc. ICML*.

Scott, J., Gass, R., Crowcroft, J., Hui, P., Diot, C., and Chaintreau, A. (2009). CRAWDAD data set cambridge/haggle (v. 2009-05-29). Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle.

Scott, S. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *J. Am. Statist. Assoc.*, 97(457):337–351.