35th AAAI Conference on Artificial Intelligence (AAAI-21)

Neural Latent Space Model

for Dynamic Networks and Temporal Knowledge Graphs

Tony Gracious,[1] Shubham Gupta,[1] Arun Kanthali,[1]
Rui M. Castro,[2] Ambedkar Dukkipati[1]

[tonygracious, shubhamg, ambedkar]@iisc.ac.in

[1]Department of Computer Science and Automation,
Indian Institute of Science, Bangalore, India.

[2]Department of Mathematics,
Eindhoven University of Technology, Eindhoven, The Netherlands.

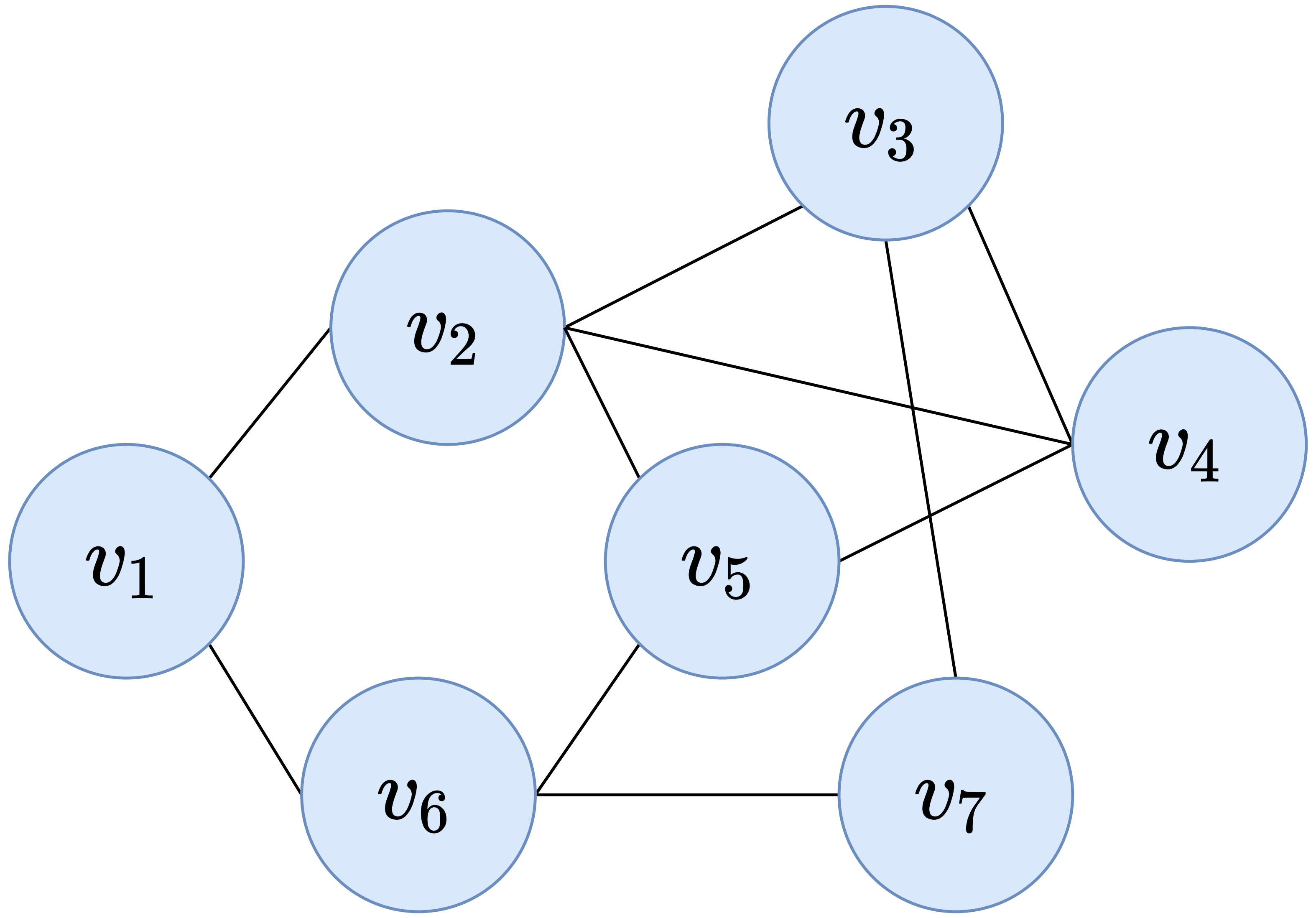
Outline

- Dynamic networks and temporal knowledge graphs
- Neural Latent Space Model (NLSM)
- Inference in NLSM
- Experimental validation
- Conclusion

Dynamic Networks and Temporal Knowledge Graphs

Networks (or Graphs)

- $G = (V, E)$
- V : Set of vertices
- E : Set of edges, $E \subseteq V \times V$



Example	Vertices	Edges	Directed
Facebook network	People	Friendship	No
Citation network	Papers	Citation	Yes
Knowledge graphs	People, movies, places, and so on	Likes, lives in, is father of, and so on	Yes

Dynamic Networks

- Social networks: New people join, others leave, friendships are made/broken
- Citation networks: New papers get added over time
- Knowledge graphs: New tourist destinations are added, people become parents, countries see leadership changes

- $G^{(t)} = (V^{(t)}, E^{(t)})$, for $t \in [T] := \{1,2,\ldots,T\}$
- Questions:
  - How communities evolve over time?
  - Which links will be formed in future?

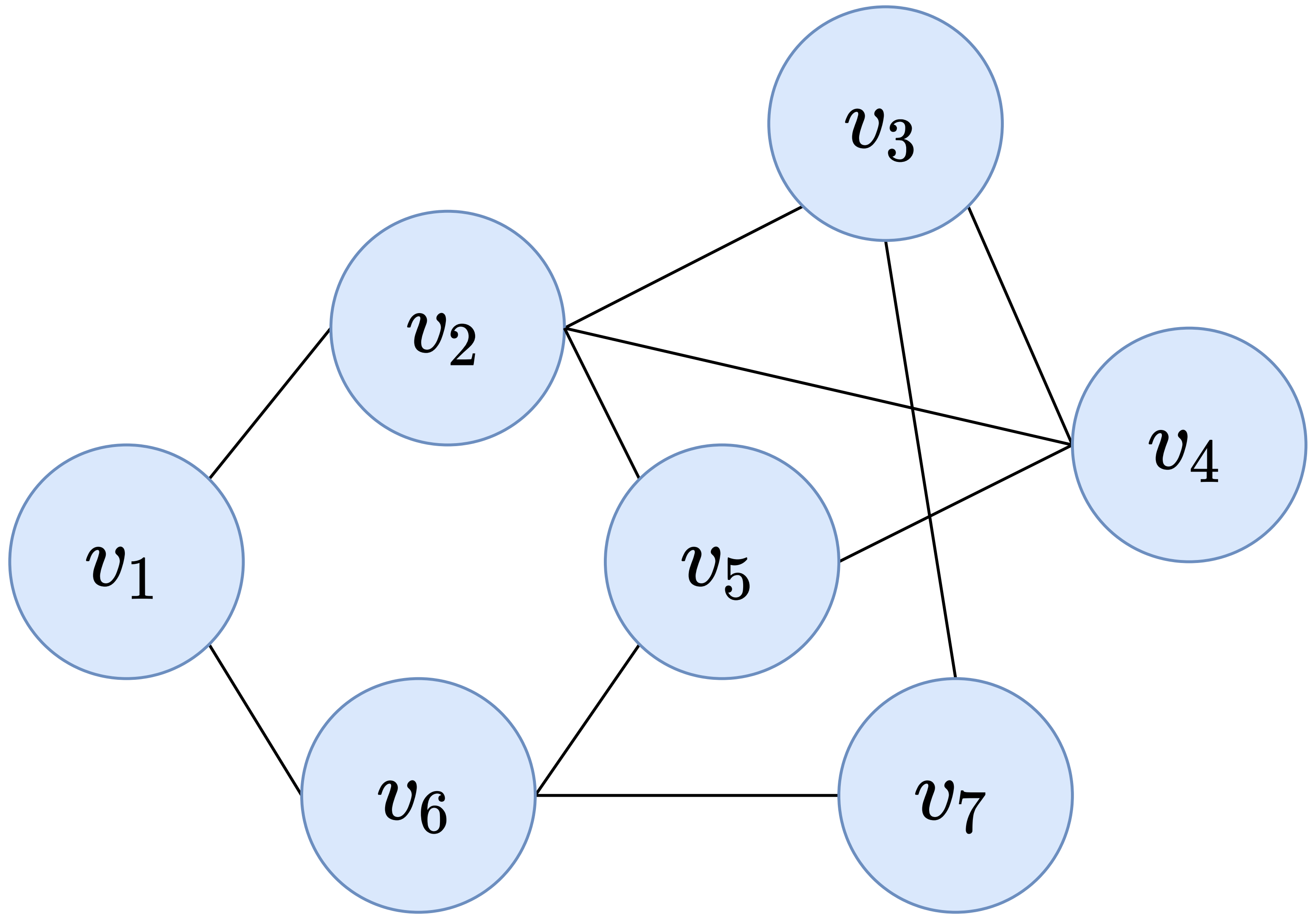
Assumption

$$V^{(1)} = V^{(2)} = \ldots = V^{(T)}$$

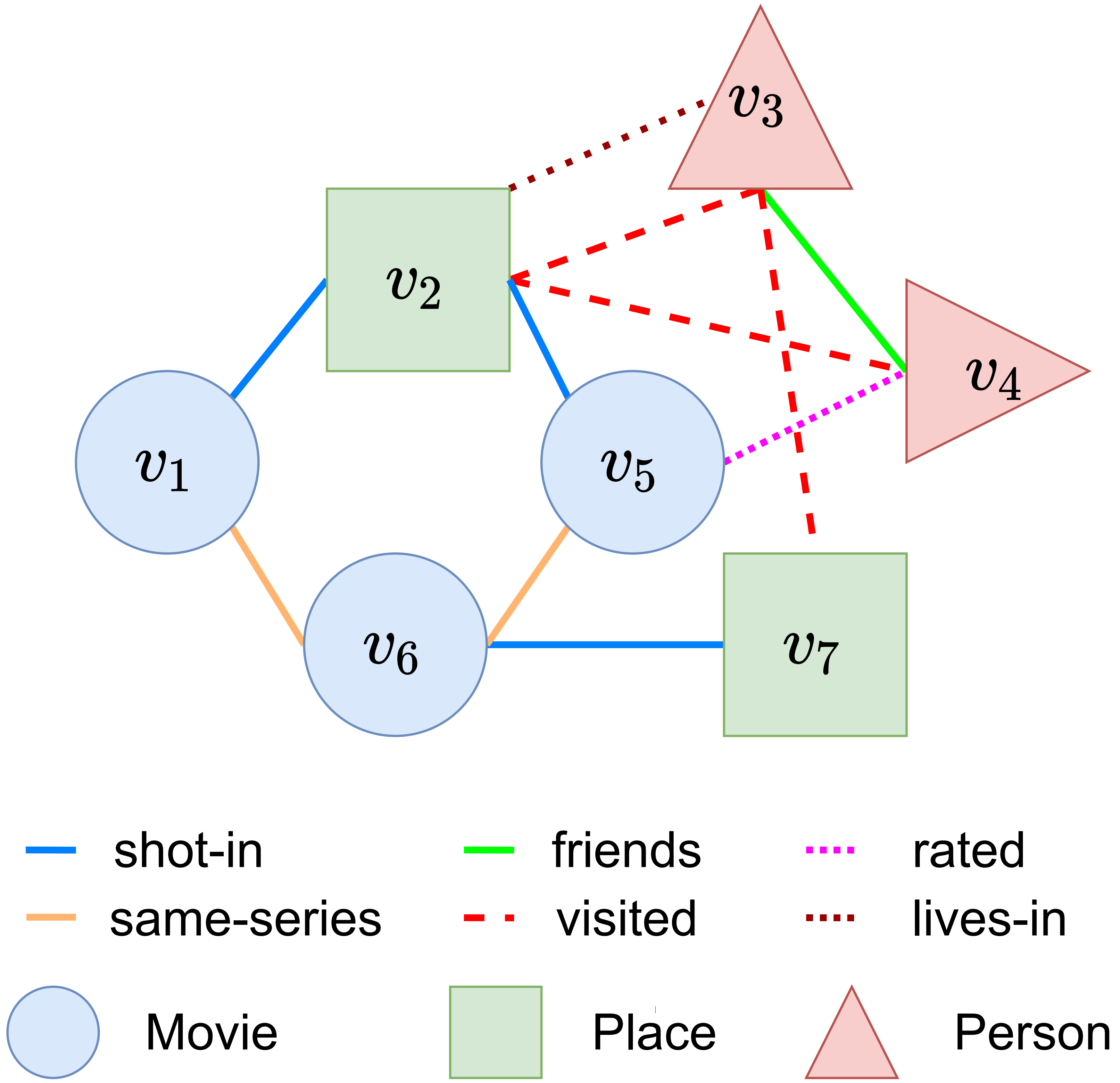
Notation

$A^{(t)}$: Adjacency matrix at time t

Heterogeneous Networks



Homogeneous Network



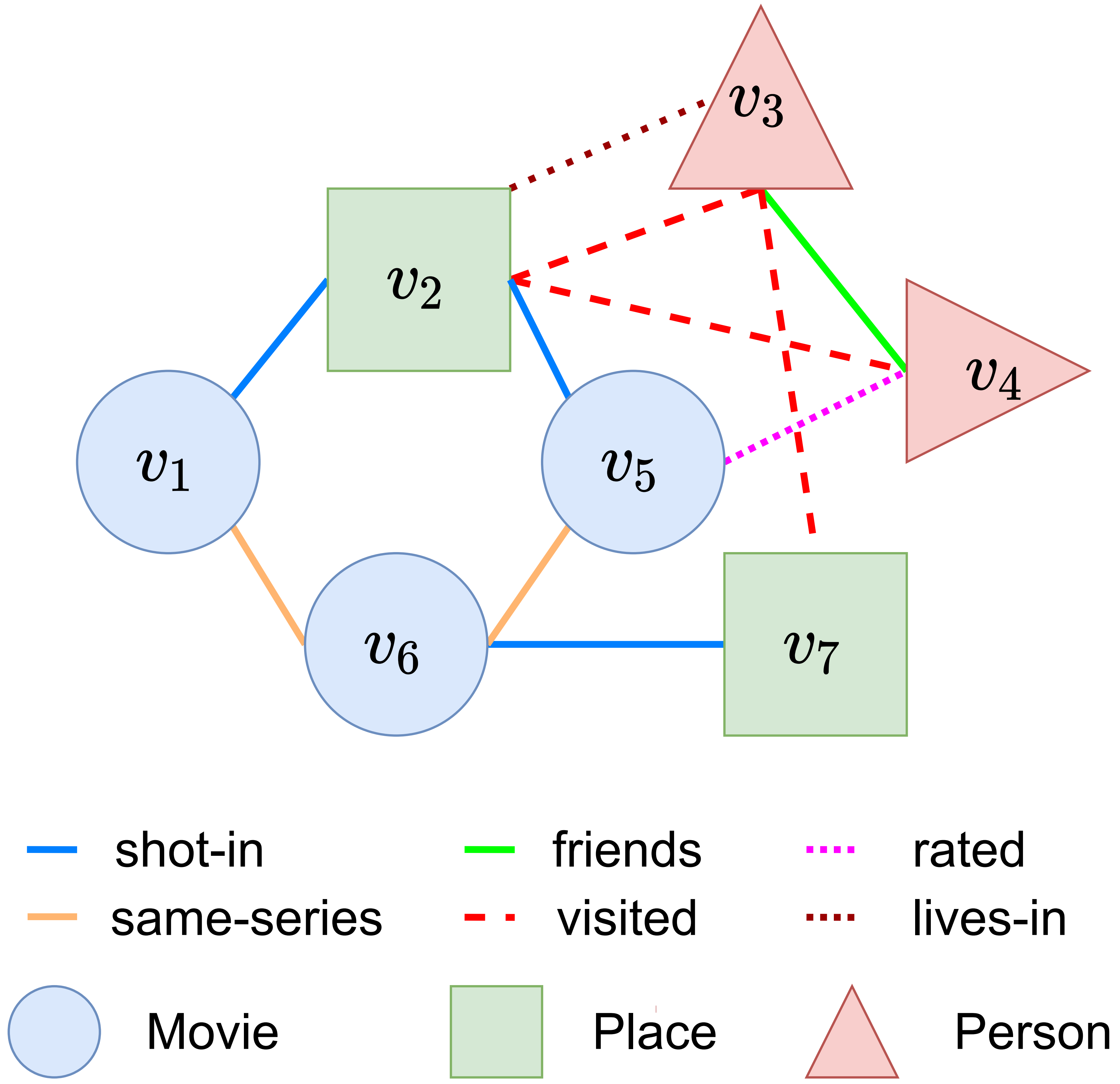
Heterogeneous Network

Temporal Knowledge Graphs

- A dynamic heterogeneous network
- Set of (subject, relation, object, time) tuples
  - (Alex, rated, Toy Story, 2013)
  - (Emily, visited, India, 2019)

Notation

$A_r^{(t)}$: Adjacency matrix for relation r at time t



Existing Literature

Homogeneous Networks

Static Networks

- Stochastic Block-model and its variants (Holland et al., 1983; Airoldi et al., 2008)
- node2vec (Grover & Leskovec, 2016) and similar methods (Perozzi et al., 2014)

Dynamic Networks

- Dynamic variants of SBM (Xing et al., 2010; Xu 2015)
- Models that use MCMC inference (Foulds et al., 2011; Kim et al., 2013)
- Neural network based models (Zhou et al., 2018; Gupta et al., 2019; Goyal et al., 2020)

Heterogeneous Networks

Dynamic Networks

- Know-Evolve (Trivedi et al., 2017)
- DyRep (Trivedi et al., 2019)
- RE-NET (Jin et al., 2019)

Static Networks

- TransE (Bordes et al., 2013), TransH (Wang et al., 2014), TransD (Lin et al., 2015), TransR (Ji et al., 2015)
- DistMult (Yang et al., 2015)
- ComplEx (Trouillon et al., 2016)

NLSM: A statistical model for homogeneous and heterogeneous dynamic networks

Contributions

- Neural Latent Space Model (NLSM) - For homogeneous and heterogeneous dynamic networks
- An efficient neural network based variational inference procedure
- Model validation using link forecasting experiments on homogeneous dynamic networks and temporal knowledge graphs

Contributions

Neural network based inference procedure

- Neural Latent Space Model (NLSM) - For homogeneous and heterogeneous dynamic networks
- An efficient neural network based variational inference procedure
- Model validation using link forecasting experiments on homogeneous dynamic networks and temporal knowledge graphs

Contributions

Nodes and edges are represented via latent node vectors and interaction matrices

- Neural Latent Space Model (NLSM) - For homogeneous and heterogeneous dynamic networks

- An efficient neural network based variational inference procedure

- Model validation using link forecasting experiments on homogeneous dynamic networks and temporal knowledge graphs

Neural Latent Space Model

NLSM: Individual Snapshots

- N : #Nodes, K : Embedding dimension, and T : #Time-steps
- Node attributes: $\mathbf{z}_n^{(t)} \in \mathbb{R}^K$, $t = 1,2,\ldots,T$, and $n = 1,2,\ldots,N$
- Interaction matrices: $\Theta_k^{(t)} \in \mathbb{R}^{2\times 2}$, $t = 1,2,\ldots,T$, and $k = 1,2,\ldots,K$

$$P(A_{ij}^{(t)} = 1 \mid \mathbf{z}_i^{(t)}, \mathbf{z}_j^{(t)}, \{\Theta_k^{(t)}\}_{k=1}^K) = \frac{1}{1 + \exp\left(-\sum_{k=1}^K \tilde{\theta}_k^{(t)}(i,j)\right)}$$

$$\tilde{\theta}_k^{(t)}(i,j) = \mathrm{E}_{x\sim B(z_{ik}^{(t)}),\ y\sim B(z_{jk}^{(t)})}\left[\Theta_k^{(t)}(x,y)\right]$$

	$1 - z_{jk}^{(t)}$	$z_{jk}^{(t)}$
$1 - z_{ik}^{(t)}$	$\Theta_k^{(t)}(0,0)$	$\Theta_k^{(t)}(0,1)$
$z_{ik}^{(t)}$	$\Theta_k^{(t)}(1,0)$	$\Theta_k^{(t)}(1,1)$

NLSM: Network Evolution

Notation

$$\Theta^{(t)} = \{\Theta_k^{(t)}\}_{k=1}^{K}, \quad \mathbf{Z}^{(t)} = \{\mathbf{z}_n^{(t)}\}_{n=1}^{N}, \quad \Psi^{(t)} = \{\psi_n^{(t)}\}_{n=1}^{N}$$

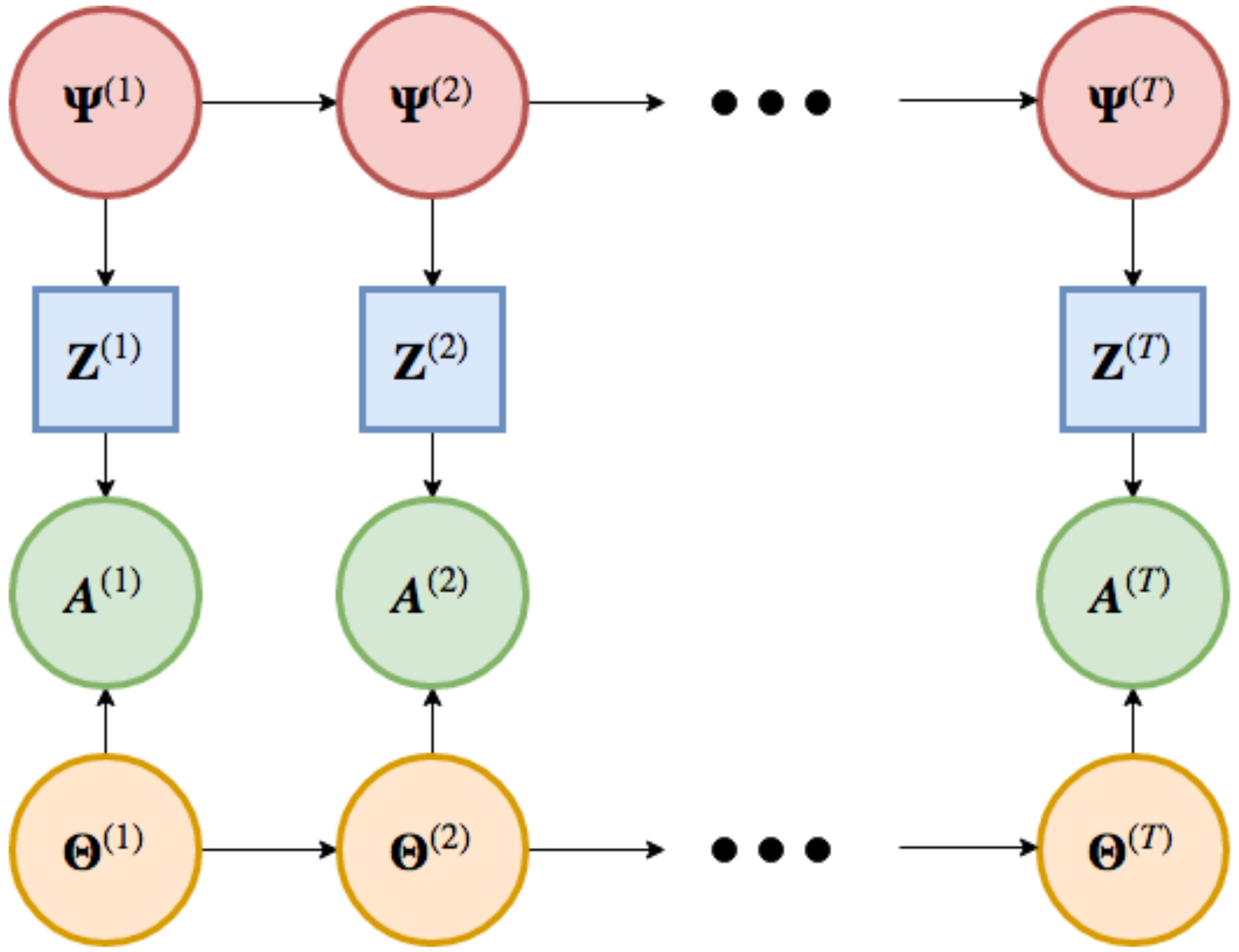
Gaussian random walk on vectorized $\Theta_k^{(t)}$

$$\bar{\theta}_k^{(t)} \sim \mathcal{N}(\bar{\theta}_k^{(t-1)}, s_\theta^2 I), \text{ for } k = 1,\ldots,K, \text{ and } t = 2,\ldots,T$$

Gaussian random walk on reparameterized $\mathbf{z}_n^{(t)}$

$$z_{nk}^{(t)} = \frac{1}{1+\exp(-\psi_{nk}^{(t)})}, \text{ for } n = 1,\ldots,N, \text{ and } t = 1,\ldots,T$$

$$\psi_n^{(t)} \sim \mathcal{N}(\psi_n^{(t-1)}, s_\psi^2 I), \text{ for } n = 1,\ldots,N, \text{ and } t = 2,\ldots,T$$



NLSM: Heterogeneous Networks

- R : #Relations
- Have a set of K interaction matrices for each relation: $\{\{\Theta^{(t)}_{k,r}\}_{k=1}^{K}\}_{r=1}^{R}$
- Edge probabilities for relation type r uses only $\{\Theta^{(t)}_{k,r}\}_{k=1}^{K}$

$$P((A_r^{(t)})_{ij} = 1 \mid \mathbf{z}_i^{(t)}, \mathbf{z}_j^{(t)}, \{\Theta^{(t)}_{k,r}\}_{k=1}^{K}) = \frac{1}{1 + \exp\left(- \sum_{k=1}^{K} \tilde{\theta}^{(t)}_{k,r}(i,j)\right)} \qquad \tilde{\theta}^{(t)}_{k,r}(i,j) = \mathrm{E}_{x \sim B(z_{ik}^{(t)}),\, y \sim B(z_{jk}^{(t)})}\left[\Theta^{(t)}_{k,r}(x,y)\right]$$

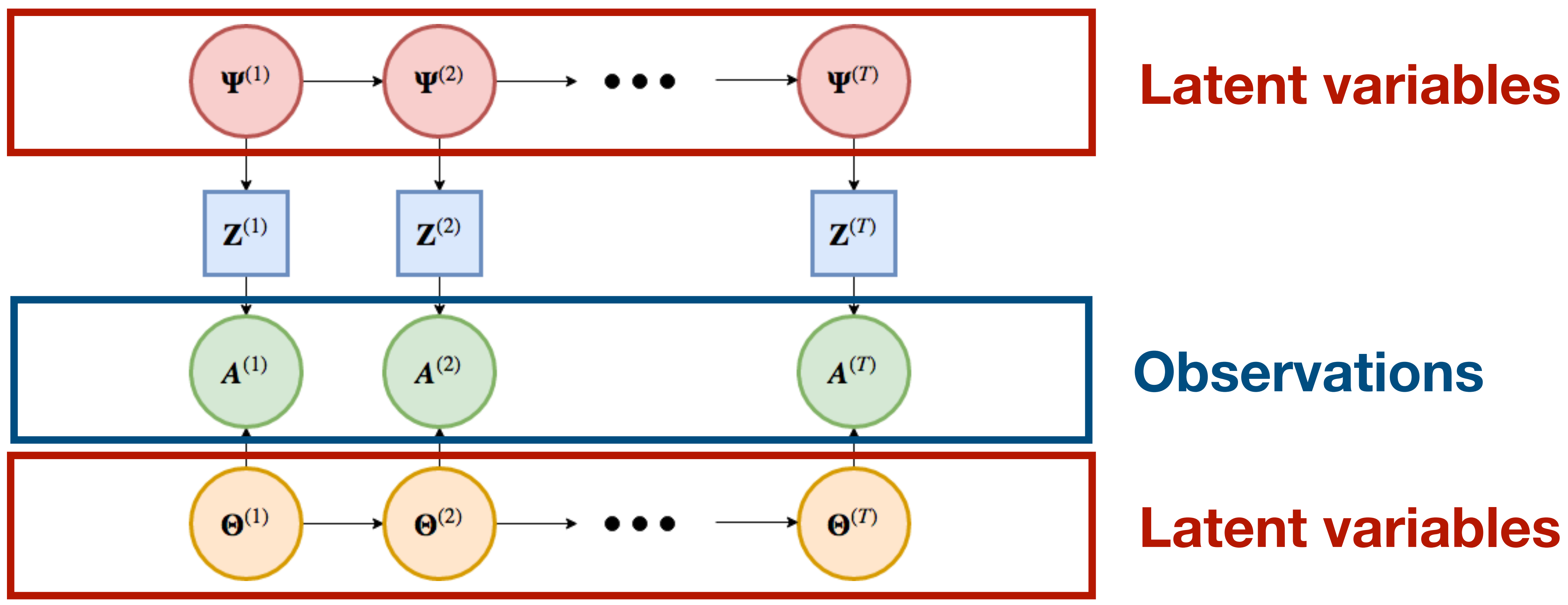
- Shared attribute vectors help in capturing information across relations

Inference in NLSM

Goal

$$\max \log \mathrm{P}(A^{(1)}, A^{(2)}, \ldots, A^{(T)})$$

Use variational inference!

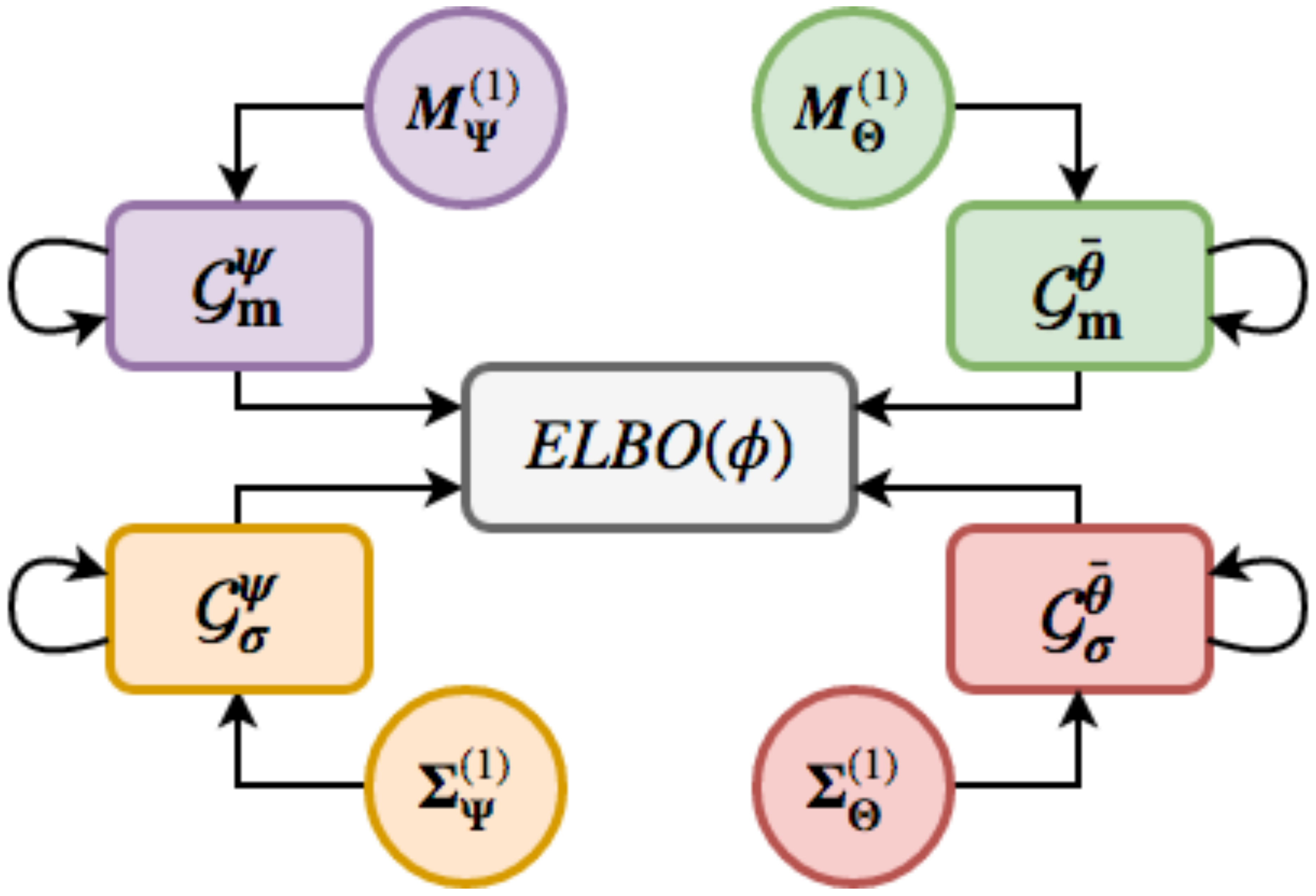


Variational Inference

$$\log \mathrm{P}(X) = \underbrace{\mathrm{E}_{H\sim Q}\Big[\log \mathrm{P}(X, H) - \log Q(H)\Big]}_{\mathrm{ELBO(Q)}} + \mathrm{KL}\Big(\mathrm{P}(H|X) \;||\; Q(H)\Big)$$

Parameterize Q using a neural network

Inference Network



Notation

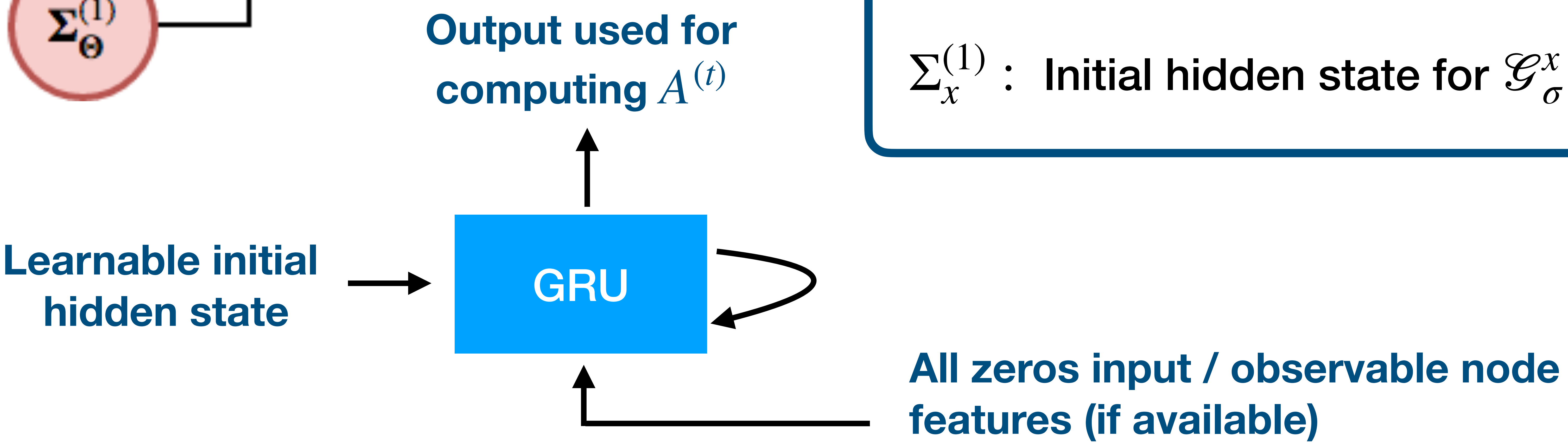
ϕ : Parameters of the inference network

$\mathcal{G}_m^x$: GRU computing mean for x

$\mathcal{G}_\sigma^x$: GRU computing standard deviation for x

$M_x^{(1)}$: Initial hidden state for $\mathcal{G}_m^x$

$\Sigma_x^{(1)}$: Initial hidden state for $\mathcal{G}_\sigma^x$



Trained to maximise $\mathrm{ELBO}(\phi)$

Experiments

Link Forecasting

- **Input:** $A^{(1)}, A^{(2)}, \ldots, A^{(t)}$
- **Goal:**
  - Single-step link forecasting: Predict links in $A^{(t+1)}$
  - Multi-step link forecasting: Predict links in $A^{(t+1)}, A^{(t+2)}, \ldots, A^{(t+p)}$, where p is the number of lookahead steps

Link Forecasting - Single Step

Dataset	AUC	node2vec [1]	DynamicTriad [2]	DynGEM [3]	DynAERNN [4]	DySAT [5]	NLSM
Enron	Micro	83.72 ± 0.7	80.26 ± 0.8	67.83 ± 0.6	72.02 ± 0.7	85.71 ± 0.3	87.05 ± 0.3
	Macro	83.05 ± 1.2	78.98 ± 0.9	69.72 ± 1.3	72.01 ± 0.7	86.60 ± 0.2	86.24 ± 0.4
UCI	Micro	79.99 ± 0.4	77.59 ± 0.6	77.49 ± 0.4	79.95 ± 0.4	81.03 ± 0.2	86.24 ± 0.4
	Macro	80.49 ± 0.6	80.28 ± 0.5	79.82 ± 0.5	83.52 ± 0.4	85.81 ± 0.1	88.90 ± 0.3
Yelp	Micro	67.86 ± 0.2	63.53 ± 0.3	66.02 ± 0.2	69.54 ± 0.2	70.15 ± 0.1	81.38 ± 0.2
	Macro	65.34 ± 0.2	62.69 ± 0.3	65.94 ± 0.2	68.91 ± 0.2	69.87 ± 0.1	80.12 ± 0.3
ML-10M	Micro	87.74 ± 0.2	88.71 ± 0.2	73.69 ± 1.2	87.73 ± 0.2	90.82 ± 0.3	92.21 ± 0.4
	Macro	87.52 ± 0.3	88.43 ± 0.1	85.96 ± 0.3	89.47 ± 0.1	93.68 ± 0.1	92.39 ± 0.3

[1](Grover & Leskovec, 2016) [2](Zhou et al., 2018) [3](Goyal et al., 2018) [4](Goyal et al., 2020) [5](Sankar et al., 2020)

Link Forecasting - Multi-Step

Approach	WIKI-Filtered			WIKI-Raw			YAGO-Filtered			YAGO-Raw		
	MRR	H@3	H@10	MRR	H@3	H@10	MRR	H@3	H@10	MRR	H@3	H@10
Know-Evolve [1]	12.64	14.33	21.57	10.54	13.08	20.21	6.19	6.59	11.48	5.23	5.63	10.23
DyRep [2]	11.60	12.74	21.65	10.41	12.06	20.93	5.87	6.54	11.98	4.98	5.54	10.19
RE-NET [3]	53.57	54.10	55.72	32.44	35.42	43.16	66.80	67.23	69.77	48.60	54.20	63.59
NLSM	56.70	57.80	61.10	35.25	38.60	47.55	69.40	71.25	73.90	52.50	59.20	68.40

[1](Trivedi et al., 2017) [2](Trivedi et al., 2019) [3](Jin et al., 2019)

Conclusion

Conclusion

- Proposed a statistical model that applies to both homogeneous and heterogeneous dynamic networks
- Ours is the first method that deals with dynamic networks and temporal knowledge graphs in a unified way
- Developed an efficient neural network based approximate inference procedure
- State-of-art link forecasting results on several datasets including temporal knowledge graphs

Additional Details in the Paper

- Justification for modelling choices
- Inference procedure
- Experimental setup and link forecasting results on smaller datasets
- Ablation studies
- Qualitative case studies

Thank you!