CERIAS

The Center for Education and Research in Information Assurance and Security

Subgraph Pattern Neural Network for High-order Graph Evolution Prediction

C. Meng, S. Mouli, B.Ribeiro, J. Neville

{meng40, chandr}@purdue.edu, {ribeiro, neville}@cs.purdue.edu

Problem Definition

Prediction task Predict Subgraph Evolution. From observed subgraphs in G_1 to G_2, predict their evolution in G_3

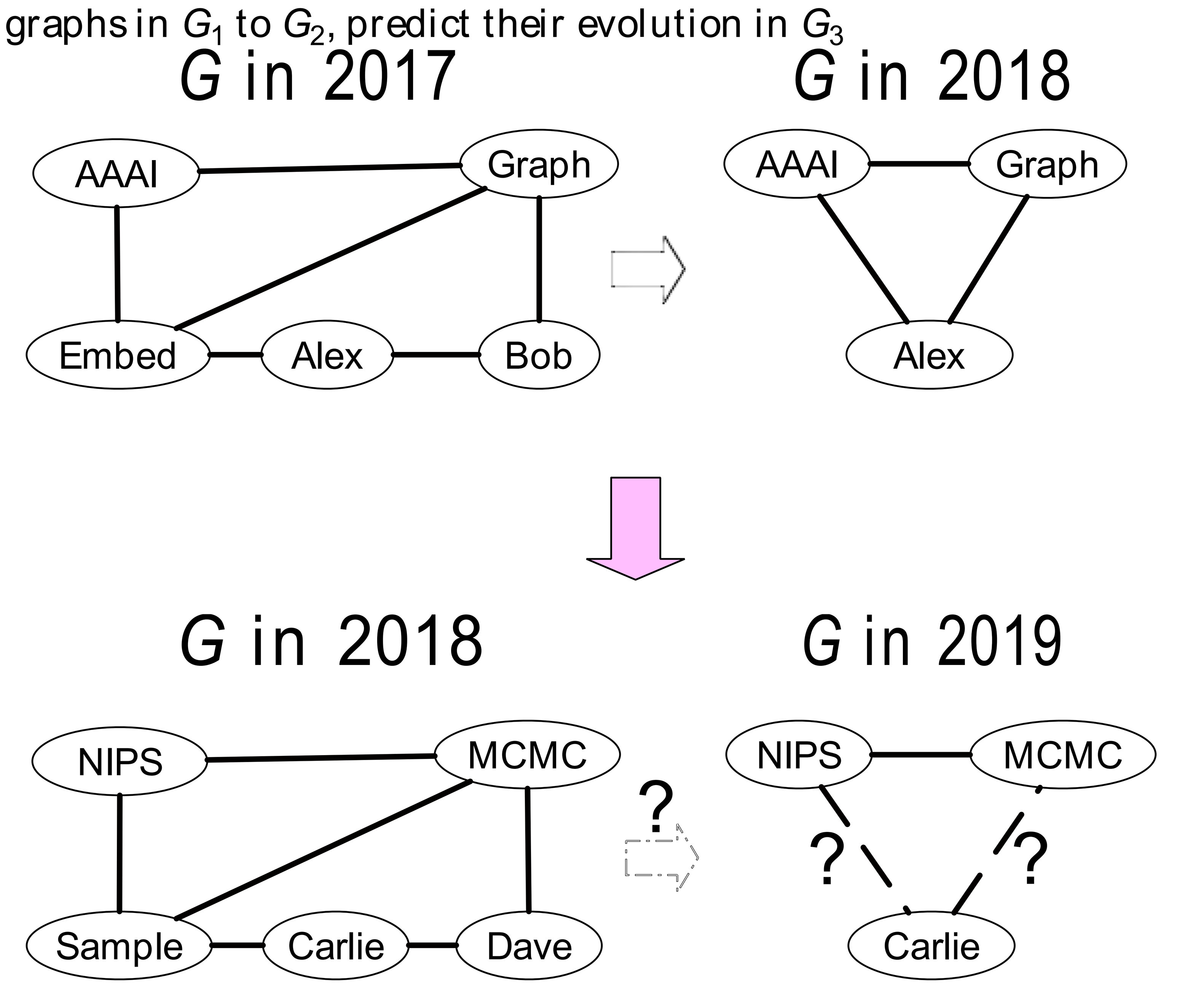

Figure 1: Example

Applications It can be used to predict topology or label evolution of subgraphs.

1. Social Network Security: Anomaly detection of group users activity.
2. Network Security: Predict traffic flow and detect anomaly.
3. DDoS Detection: Distinguish normal and abnormal users.

Figure 2: Applications

Subgraph Pattern Neural Networks

SPNN is a 3-layer gated neural network with a sparse structure generated from the training data in a pre-processing step.

Consider a K-class classification task, $y_{t+1}(U)$ as a one-of-K encoding vector. The probability nodes $V(U)$ form an induced subgraph in G_{t+1} with a pattern of class i is

$$p(y_{t+1}(U))_i = \text{softmax}((\mathrm{W}^{(1)}h_t(U;\mathrm{W}^{(2)},\mathrm{b}^{(2)})+\mathrm{b}^{(1)})_i),$$

The pattern layer is a set of neurons. Each neuron corresponds to one subgraph pattern. The input to the pattern layer is

$$h_t(U;\mathrm{W}^{(2)},\mathrm{b}^{(2)}) = (\Delta(U,F_1^{\square},G_t)\cdot\sigma(\mathrm{b}_1^{(2)}+(\mathrm{W}_1^{(2)})^{\mathrm{T}}$$
$$\varphi(U,F_1^{\square},G_t)),\Delta(U,F_2^{\square},G_t)\cdot\sigma(\mathrm{b}_2^{(2)}+(\mathrm{W}_2^{(2)})^{\mathrm{T}}\varphi(U,F_2^{\square},G_t)),\ldots)$$

Each input feature nodes only connected to their corresponding neutron according to the graph topology.

Local induced isomorphism density Induced subgraph $\{A_2,V_2,T_3\}$ in G_1 in Figure 3a in contained in a 4-node subgraph with $V(R) = \{A_2,V_2,T_3,A_1\}$. The pattern $H^{\sharp}$ has 4-node subgraphs that are within a radius of $d = 1$ of the nodes $\{A_2,V_2,T_3\}$.

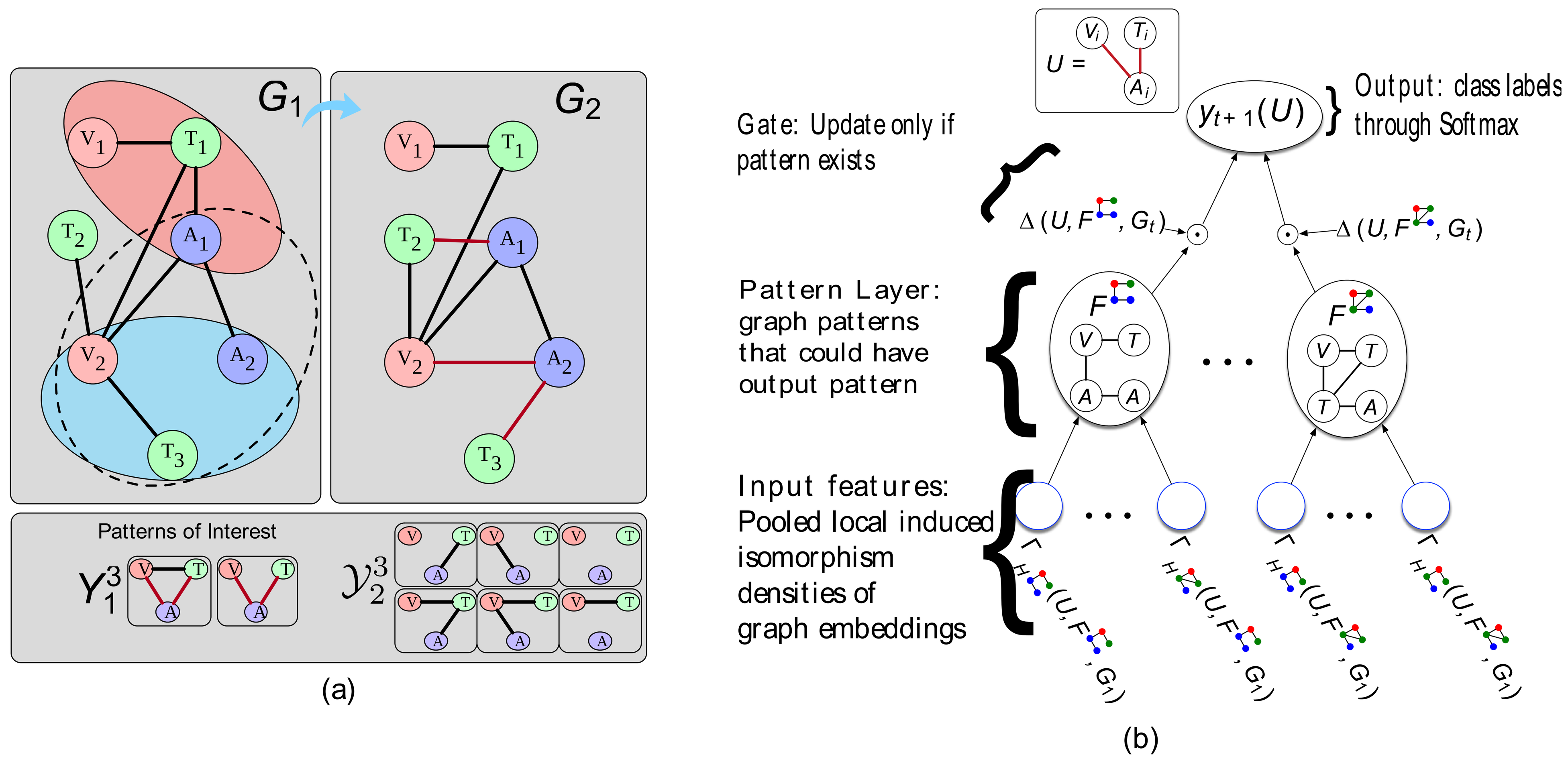

Figure 3: (a) Illustration of the training in a citation network with (A)uthors, (T)opics, (V)enues. Learning the evolution from G_1 to G_2, we can predict the subgraph patterns in G_3. (b) Subgraph-Pattern Neural Network (SPNN).

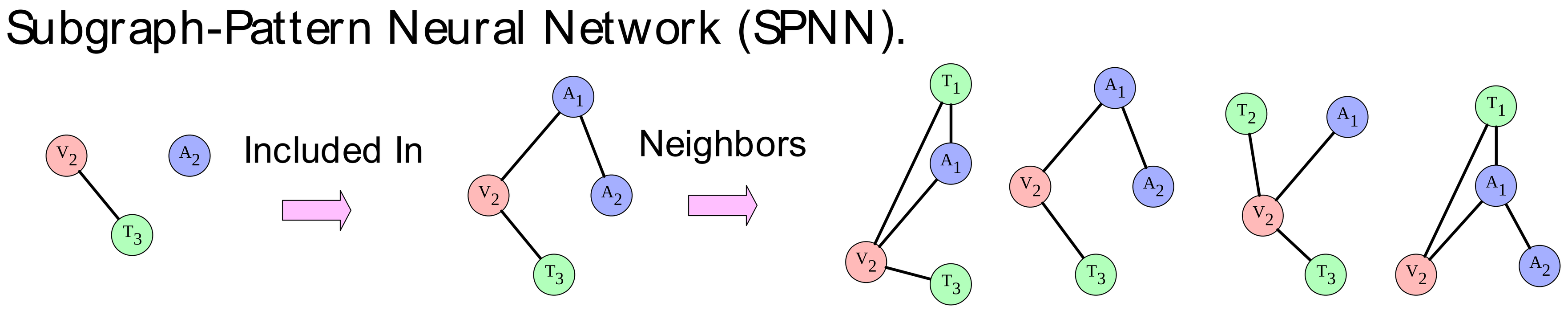

Figure 4: Local induced isomorphism density example

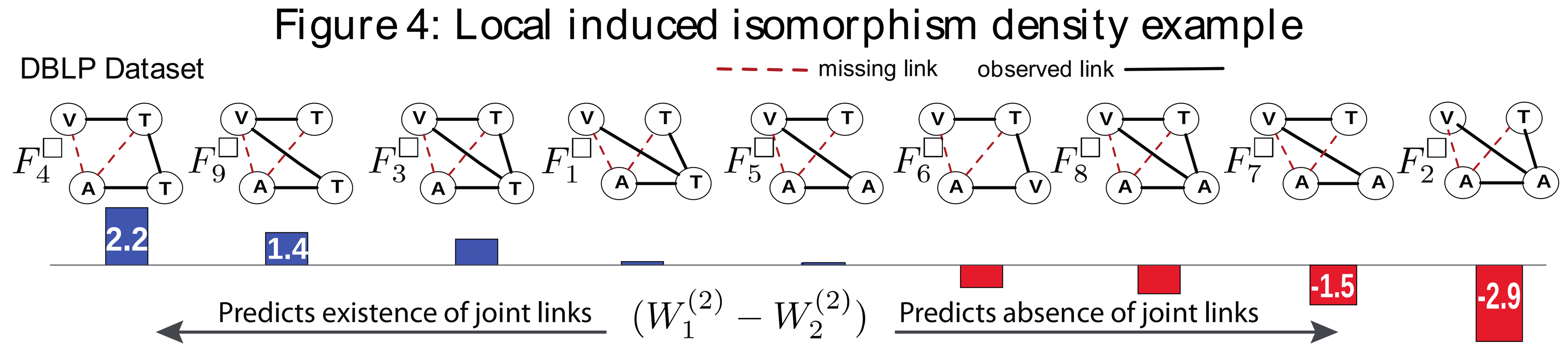

Figure 5: (DBLP task) Pattern layer $F_1^{\square},\ldots,F_8^{\square}$, representing connected subgraphs patterns that appear in the training data. Bars show the di□erence between learned weights of Class 1 and Class 2 for each pattern.

Results

Datasets. DBLP contains scientific papers in four related areas with 14k papers, 14k authors, 8k topics, and 20 venues.
Friendster contains 14 millions of nodes and 75 million messages includes hometown, college, interests, and messages sent between users.

Subgraph Pattern Prediction Tasks.

1. DBLP task is to predict whether an author will publish in a venue and a topic.
2. Friendster Activity task predicts whether the total number of messages sent between 4 users will increase.
3. Friendster Structure task predicts whether four friends who were weakly connected by three edges will not contact in the future.

Baselines. (i) AA: Adamic–Adar ; (ii) EdgeInfo: ; (iii) PC: Path counts ; (iv) PCRW: Path constrained random walk; (v) Node2Vec: Node embedding. (vi) Rescal: Rescal embedding; (vii) HolE: Holographic embedding. (viii) Patchy: Patchy CNN graph kernel; (ix) GraphNN: Embedding Mean-Field Inference.

	Independently Trained (Single Link Predictions)						Jointly Trained Multi-Link Task								
	EdgeInfo	PCRW	PC	N2V	Rescal	HolE	EdgeInfo	PCRW	PC	N2V	Rescal	HolE	Patchy	GraphNN	SPNN
DBLP	0.811	0.786	0.783	0.567	0.611	0.681	0.830	0.782	0.788	0.582	0.611	0.690	0.627	0.571	0.846
	±0.012	±0.007	±0.012	±0.008	±0.025	±0.024	±0.007	±0.007	±0.014	±0.007	±0.025	±0.024	±0.003	±0.021	±0.011
Friendster	0.530	0.516	0.509	0.512	0.521	0.513	0.502	0.516	0.515	0.524	0.502	0.506	0.519	0.521	0.690
(Activity)	±0.088	±0.007	±0.006	±0.011	±0.031	±0.006	±0.007	±0.012	±0.012	±0.018	±0.012	±0.013	±0.010	±0.023	±0.008
Friendster	0.568	0.501	0.501	0.501	0.558	0.501	0.501	0.502	0.552	0.540	0.521	0.530	0.547	0.523	0.607
(Structure)	±0.011	±0.002	±0.002	±0.003	±0.009	±0.002	±0.004	±0.002	±0.019	±0.017	±0.017	± 0.021	± 0.025	±0.019	±0.017

Table 1: Max Area Under Curve (AUC) scores of SPNN against baselines.

* Code and data are available at https://github.com/PurdueMINDS/SPNN

PURDUE UNIVERSITY Discovery Park

CERIAS