Propagation on Multi-relational Graphs for Node Regression [Accepted to Complex Networks 2021]

Eda Bayram

EPFL

IJCLR 2021 — Statistical Relational AI

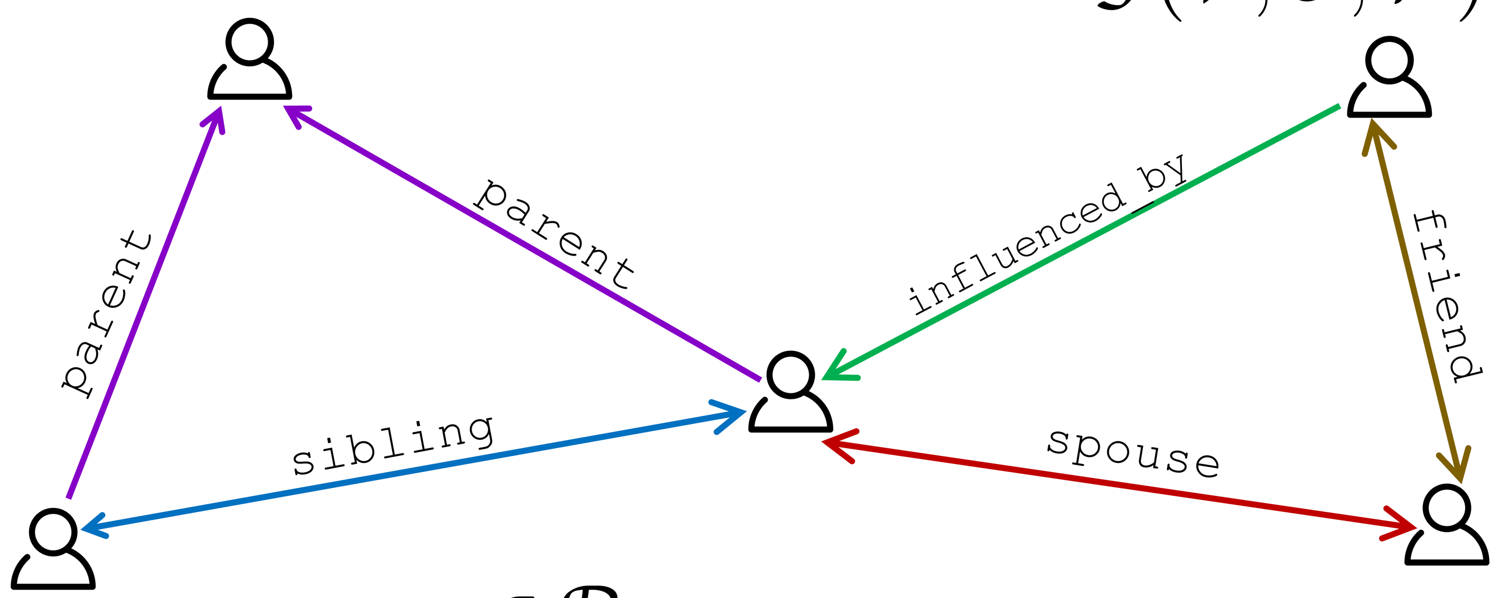
full paper @arXiv

Motivation

- Rise in complex network-structured data storing different level of interactions between data entities e.g., social networks, biological networks, transportation networks, etc..
- Better encoded by multi-relational graphs

Objective

- Given multi-relational structure of data $\mathcal{G}(\mathcal{V},\mathcal{E},\mathcal{P})$



- A relation type $\mathtt{p} \in \mathcal{P}$

$$\mathtt{r}(i,j) = \mathtt{p} \quad \text{if} \quad (i) \xleftarrow{\mathtt{p}} (j)$$

- Partially observed continuous features belonging to data entities $\{x_i \in \mathbb{R} \quad \forall i \in \mathcal{U} \subset \mathcal{V}\}$

How can we achieve node-value imputation on a multi-relational, directed graph?

Methodology

Propagation

- Simple and effective technique to complete missing information in case of few labeled data
  - **Semi-supervised learning on graphs**, e.g., Label Propagation [Zhu & Ghahramani, 2002]
    Completes categorical features across simple and weighted graphs
  - **Graph representation learning methods**
    Propagate node representations across the graph
- Iterative neighborhood aggregation technique with Smoothness prior
  "Neighboring nodes should have similar representations."

Smoothness prior →(1) Local generative model →(2) First order Bayesian estimate of node's value →(3) Neighborhood aggregation →(4, iterative application) Propagation algorithm

Simple Weighted Graph

Simple Local Generative Model

Notion of similarity between neighbors
- Uniform
- Symmetric

$$x_i = x_j + \epsilon \quad \forall (i,j) \in \mathcal{E} \quad \epsilon \sim \mathcal{N}(0, 1/\omega_{ij})$$

$$\underset{x_i}{\operatorname{argmin}} \sum_{(i,j)\in\mathcal{E}} \omega_{ij}(x_i - x_j)^2$$

Minimize Euclidean Distance between neighbors

$$\frac{\sum_{(i,j)\in\mathcal{E}} \omega_{ij} x_j}{\sum_{(i,j)\in\mathcal{E}} \omega_{ij}}$$

Simple Weighted Average

Multi-relational Directed Graph

Relational Local Generative Model

Notion of similarity conforms
- Different types of relationships
- Asymmetric relationships

$$x_i = \eta_{\mathtt{p}} x_j + \tau_{\mathtt{p}} + \epsilon \quad \forall \mathtt{r}(i,j) = \mathtt{p} \quad \epsilon \sim \mathcal{N}(0, 1/\omega_{\mathtt{p}})$$

$$\underset{x_i}{\operatorname{argmin}} \sum_{\mathtt{p}\in\mathcal{P}\cup\mathcal{P}^{-1}} \sum_{\mathtt{r}(i,j)=\mathtt{p}} \omega_{\mathtt{p}}(x_i - \eta_{\mathtt{p}} x_j - \tau_{\mathtt{p}})^2$$

$$\frac{\sum_{\mathtt{p}\in\mathcal{P}\cup\mathcal{P}^{-1}} \sum_{\mathtt{r}(i,j)=\mathtt{p}} \omega_{\mathtt{p}} f_{\mathtt{p}}(x_j)}{\sum_{\mathtt{p}\in\mathcal{P}\cup\mathcal{P}^{-1}} \sum_{\mathtt{r}(i,j)=\mathtt{p}} \omega_{\mathtt{p}}}$$

Relational Transformation $f_{\mathtt{p}}(x) = \eta_{\mathtt{p}} x + \tau_{\mathtt{p}}$

Multi-Relational Propagation Algorithm (MRP)

- **Aggregation** — Relational transformation applied to neighbors
- **Normalization** — Weighted average of aggregated values
- **Update** — Linearly combine with the previous state with a damping factor $\xi \in [0,1]$
- **Clamp** — Keep labeled nodes as they are

Indicator vector keeps track of labeled and propagated nodes throughout the iterations

Iterates until the stopping criterion is met.

Parameters of MRP

- Parameters of a certain relation type $\{\tau_{\mathtt{p}}, \eta_{\mathtt{p}}, \omega_{\mathtt{p}}\}$ estimated over initially labeled node pairs
  $$\{(x_i, x_j) | i,j \in \mathcal{U}, \mathtt{r}(i,j) = \mathtt{p}\}$$
  similar to parameter estimation of a linear regression model.
- Weight parameter $\omega_{\mathtt{p}}$ is equal to the uncertainty of the linear regression model of relation type $\mathtt{p}$
- MRP drops down to standard propagation by default parameter set $\{\tau_{\mathtt{p}} = 0,\ \eta_{\mathtt{p}} = 1,\ \omega_{\mathtt{p}} = 1\ \forall \mathtt{p} \in \mathcal{P}\}$
- Implemented with PyTorch-scatter package, linearly scales with $|\mathcal{E}|$

Experiments [Labeled set $\mathcal{U} \subset \mathcal{V}$ randomly sampled with sparsity, evaluation metrics are averaged on a series of experiments.]

Prediction of People's Date of Birth

- A subset of relational database FreeBase
- People $|\mathcal{V}| = 830$ connected via
- Different relationships $|\mathcal{P}| = 8$

	Relation type	RMSE	MAPE	nRMSE
LP	award_nomination	32.43	0.011	0.115
	friendship	31.92	0.011	0.113
	influenced_by	30.29	0.012	0.108
	sibling	32.69	0.012	0.116
	parent	33.62	0.013	0.119
	spouse	31.45	0.011	0.112
	dated	31.70	0.011	0.113
	awards_won	33.04	0.012	0.117
	union	24.22	0.008	0.086
MRP		15.62	0.005	0.055

Prediction of Weather Measurements

- Weather stations $|\mathcal{V}| = 86$ connected via $|\mathcal{P}| = 2$ relationships:
  - Geographical proximity (symmetric relationship)
  - Altitude proximity (edge directed to altitude ascend)

		RMSE	MAPE	nRMSE
Temperature	LP	1.120	0.155	0.050
	MRP	1.040	0.147	0.045
Snowfall	LP	194.49	0.405	0.112
	MRP	180.10	0.357	0.105

Precipitation

	RMSE	MAPE	nRMSE
LP-altitude	381.86	0.261	0.174
LP-gps	374.38	0.242	0.168
MRP	347.98	0.238	0.157

- MRP outperforms standard propagation (LP) by
  - Regarding asymmetric relations with relational transformation in aggregation
  - Assigning different level of importance to different types of relations