Non-Parametric Learning of Embeddings for Relational Data using Gaifman Locality Theorem

Devendra Singh Dhami, Siwen Yan, Gautam Kunapuli & Sriraam Natarajan

Why?

Motivation

Key Question: Can graph neighborhood information be used to learn better representations?

Potential Benefits: Allows exploiting structural features of local neighborhood; aggregated from locally sampled neighborhoods (Gaifman locality theorem)

Niepert introduced Discriminative Gaifman models[$]
Basic Idea: nodes with shared neighbors more likely to be similar and thus more likely to have a link **(link prediction)**

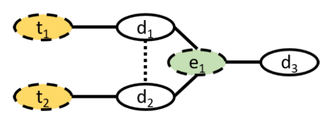
Issues: *Existing work focuses on node classification; No structure learning (learning first order rules)*

Background

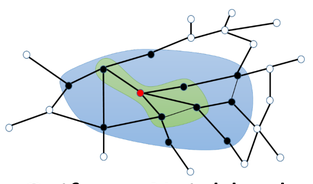
Gaifman Locality Theorem

- Every first-order sentence is logically equivalent to a **boolean combination** of basic local sentences
  - ***only a small part of structure relevant for query evaluation; global structure search is not required.***

Gaifman (Primal) graph: contains edges joining two nodes only if the entities corresponding to those nodes are present in a relation together




```
TransporterSubstrate(Pravastatin, Bile salt export pump)
TransporterInhibitor(Simvastatin, Multidrug resistance protein 1)
EnzymeInhibitor(Pravastatin, Cytochrome P450 2C9)
EnzymeSubstrate(Acetaminophen, Cytochrome P450 2C9)
EnzymeInhibitor(Simvastatin, Cytochrome P450 2C9)
```

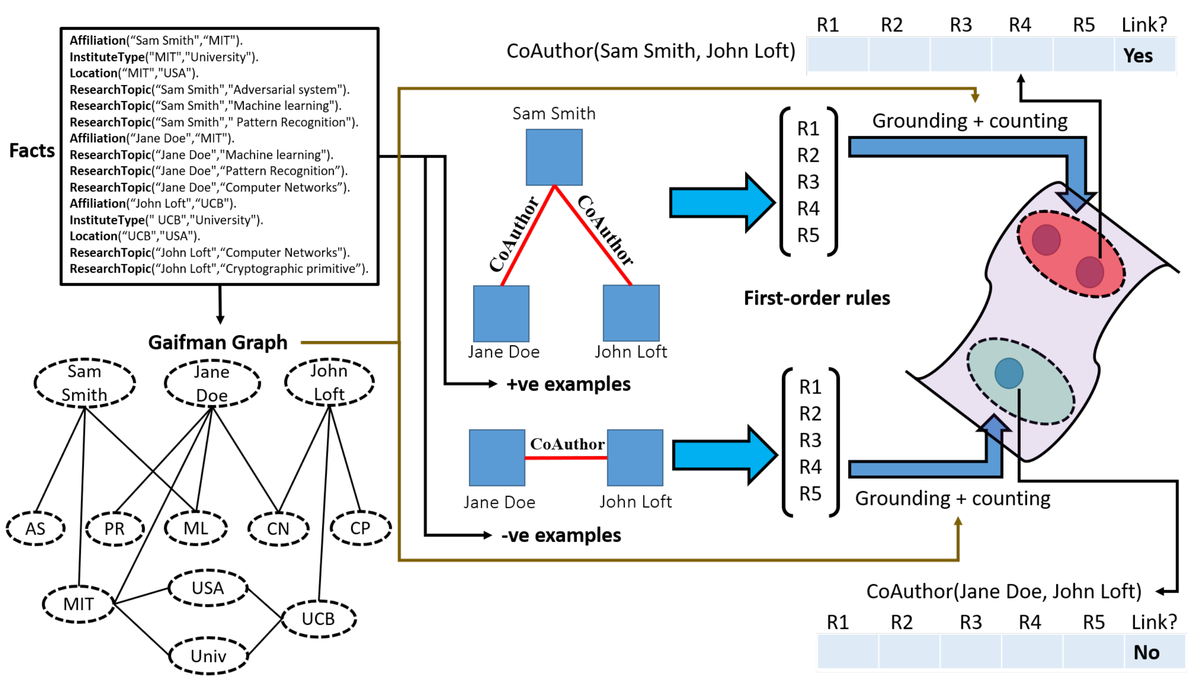


Gaifman Neighborhood

Directly adopted to relational data

How?

Our Approach



Rule Learning (Density Estimation)

Rule Learning:
1. Relational Random Walks[%]
2. Inductive Logic Programming[#]
3. **Density estimation (relational one class classification[*])**

1. Partial Grounding using query variables
2. Count satisfied groundings **based on Gaifman neighborhood**

Intuition : Learning first order rules for positive and negative examples independently can result in better utilization of the search space thereby learning more discriminative features (both relational and thus propositional)

1. $\text{EnzymeInducer}(A, C), \text{EnzymeSubstrate}(B, C), \text{EnzymeInducer}(B, D), \text{EnzymeInducer}(A, D) \implies \text{Interacts}(A,B)$
2. $\text{EnzymeSubstrate}(A, C), \text{EnzymeSubstrate}(B, C), \text{TransporterInducer}(A, D) \implies \text{Interacts}(A,B)$
3. $\text{EnzymeInhibitor}(A, C), \text{Enzyme}(C, B), \text{TransporterInhibitor}(A, D) \implies \text{Interacts}(A,B)$
4. $\text{TargetAgonist}(B, C), \text{TransporterSubstrate}(A, D), \text{EnzymeSubstrate}(B, E), \text{EnzymeSubstrate}(A, E) \implies \text{Interacts}(A,B)$

What We Get?

Data Sets and Experiments

5 relational data sets

Data set	#Entities	#Relations	#Pos	#Neg	#RW rules	#ILP rules	#relOCC rules
DDI	355	15	2832	3188	68	36	25
PPI	797	7	1915	1915	42	5	15
NELL Sports	4147	6	300	600	36	15	13
Financial NLP	650	7	186	1029	222	6	25
ICML CoAuthor	558	5	155	6498	7	15	7

Comparison with Rule Learning Methods

Data set	Methods	Accuracy	Recall	F1	AUC-ROC	AUC-PR
DDI	Gaifman	0.771	0.658	0.691	0.697	0.710
	Neural LP	0.632	0.777	0.470	0.741	0.404
	metapath2vec	0.717	0.767	0.717	0.768	0.696
	relOCC	**0.897**	**0.991**	**0.901**	**0.902**	**0.853**
PPI	Gaifman	0.652	0.524	0.606	0.654	0.619
	Neural LP	0.395	0.336	0.357	0.345	0.440
	metapath2vec	0.642	0.767	0.715	0.660	**0.729**
	relOCC	**0.733**	**0.999**	**0.789**	**0.733**	0.652
NELL Sports	Gaifman	0.780	0.485	0.597	0.707	0.549
	Neural LP	0.663	0.400	0.442	0.583	0.412
	metapath2vec	0.778	**0.867**	**0.765**	**0.875**	**0.850**
	relOCC	**0.833**	0.600	0.731	0.778	0.643
Financial NLP	Gaifman	0.914	0.590	0.705	0.787	0.587
	Neural LP	0.705	0.745	0.434	0.768	0.314
	metapath2vec	0.699	**0.982**	0.568	**0.927**	0.675
	relOCC	**0.967**	0.800	**0.889**	0.879	**0.833**
ICML CoAuthor	Gaifman	0.984	0.327	0.493	0.664	0.343
	Neural LP	0.718	**0.800**	0.045	0.846	0.179
	metapath2vec	0.912	**0.800**	0.333	**0.922**	0.350
	relOCC	**0.997**	0.386	**0.557**	0.693	**0.400**

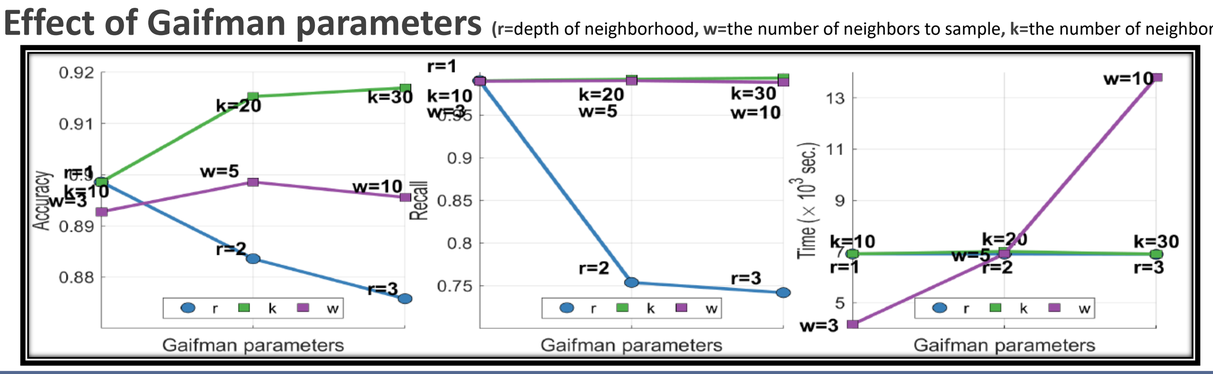
Comparison with Statistical Relational Learning Methods

Data set	Methods	Accuracy		Recall		F1		AUC-ROC		AUC-PR	
		LR	GB	LR	GB	LR	GB	LR	GB	LR	GB
DDI	RW	0.657	0.669	0.469	0.530	0.564	0.602	0.647	0.662	0.581	0.593
	ILP	0.696	0.774	0.467	0.674	0.592	0.729	0.684	0.767	0.710	0.765
	relOCC	0.860	**0.897**	0.939	**0.991**	0.864	**0.901**	0.864	**0.902**	0.797	**0.853**
	Niepert [2016]	0.534	0.534	0.0	0.0	0.0	0.0	0.5	0.5	0.466	0.466
	MLN-Boost	0.638		0.504		0.618		0.798		0.784	
	RDN-Boost	0.755		0.662		0.718		0.828		0.831	
PPI	RW	0.700	**0.785**	0.586	0.707	0.661	0.767	0.699	**0.785**	0.651	**0.740**
	ILP	0.613	0.661	0.397	0.553	0.506	0.620	0.613	0.661	0.579	0.614
	relOCC	0.727	0.733	0.996	**0.999**	0.785	**0.789**	0.727	0.733	0.647	0.652
	Niepert [2016]	0.608	0.652	0.382	0.524	0.499	0.606	0.613	0.654	0.591	0.619
	MLN-Boost	0.548		0.453		0.571		0.743		0.733	
	RDN-Boost	0.671		0.615		0.652		0.728		**0.740**	
NELL Sports	RW	0.783	0.822	0.414	0.569	0.569	0.689	0.696	0.762	0.565	0.594
	ILP	0.782	0.824	0.431	0.590	0.578	0.699	0.699	0.769	0.530	0.564
	relOCC	0.793	**0.833**	0.431	0.6	0.59	**0.731**	0.708	0.778	0.574	0.643
	Niepert [2016]	0.756	0.780	0.314	0.485	0.465	0.597	0.648	0.707	0.512	0.549
	MLN-Boost	0.605		0.533		0.667		**0.894**		**0.853**	
	RDN-Boost	0.812		**0.756**		0.714		0.884		0.834	
Financial NLP	RW	0.833	0.833	0.0	0.0	0.0	0.0	0.5	0.5	0.168	0.168
	ILP	0.838	0.921	0.068	0.633	0.112	0.727	0.530	0.806	0.200	0.6023
	relOCC	0.965	0.967	0.788	0.800	0.882	0.889	0.867	0.879	0.826	0.833
	Niepert [2016]	0.827	0.914	0.0	0.59	0.0	0.705	0.5	0.787	0.173	0.587
	MLN-Boost	0.928		0.764		0.757		**0.989**		0.807	
	RDN-Boost	**0.975**		**0.963**		**0.929**		**0.989**		**0.901**	
ICML CoAuthor	RW	0.977	0.977	0.0	0.0	0.0	0.0	0.5	0.5	0.023	0.023
	ILP	0.983	0.985	0.272	0.339	0.427	0.506	0.636	0.669	0.289	0.356
	relOCC	0.986	**0.997**	0.346	0.386	0.517	**0.557**	0.653	**0.693**	0.370	**0.40**
	Niepert [2016]	0.981	0.984	0.195	0.327	0.326	0.493	0.597	0.664	0.214	0.343
	MLN-Boost	0.938		0.326		0.214		0.294		0.210	
	RDN-Boost	0.940		**0.434**		0.231		0.153		0.157	

Comparison with Relational Embedding Methods

Data set	Metric	ConvE	ComplEx	SimplE	DistMult	HolE	TransE	TransH	TransR	TransD	relOCC
DDI	Accuracy	0.744	0.787	0.509	0.683	0.586	0.533	0.479	0.465	0.476	**0.897**
	Recall	0.931	0.832	0.051	0.988	0.922	0.522	0.662	0.802	0.793	**0.991**
	F1	0.544	0.618	0.030	0.567	0.483	0.320	0.348	0.387	0.389	**0.901**
	AUC-ROC	0.744	0.818	0.195	0.962	0.844	0.541	0.554	0.653	0.659	**0.902**
	AUC-PR	0.678	0.705	0.118	0.912	0.641	0.231	0.222	0.313	0.332	**0.853**
PPI	Accuracy	**0.747**	0.676	0.739	0.787	0.500	0.390	0.388	0.417	0.446	0.733
	Recall	0.685	0.603	0.793	0.707	0.0	0.401	0.408	0.449	0.512	**0.999**
	F1	0.729	0.650	0.752	0.768	0.0	0.397	0.400	0.435	0.480	**0.789**
	AUC-ROC	**0.829**	0.732	0.828	0.823	0.500	0.332	0.331	0.385	0.424	0.733
	AUC-PR	**0.855**	0.704	0.843	0.870	0.500	0.400	0.385	0.430	0.447	0.652
NELL Sports	Accuracy	0.667	0.629	0.548	0.607	0.756	0.544	0.530	0.470	0.448	**0.833**
	Recall	0.711	**0.733**	0.633	0.633	0.633	0.622	0.600	0.489	0.511	0.600
	F1	0.587	0.569	0.484	0.518	0.633	0.477	0.460	0.381	0.382	**0.731**
	AUC-ROC	0.743	0.762	0.620	0.694	0.745	0.589	0.571	0.456	0.489	**0.778**
	AUC-PR	0.517	0.628	0.437	0.645	0.730	0.452	0.423	0.332	0.3	**0.643**
Financial NLP	Accuracy	0.796	0.634	0.421	0.708	0.848	0.526	0.501	0.551	0.584	**0.967**
	Recall	0.963	0.472	**0.964**	0.982	0.0	0.673	0.527	0.691	0.691	0.800
	F1	0.589	0.281	0.335	0.505	0.0	0.301	0.243	0.318	0.335	**0.889**
	AUC-ROC	**0.953**	0.574	0.779	0.918	0.5	0.631	0.485	0.648	0.711	0.879
	AUC-PR	0.765	0.232	0.359	0.749	0.152	0.225	0.139	0.278	0.402	**0.833**
ICML CoAuthor	Accuracy	0.981	0.977	0.985	0.515	0.992	0.494	0.500	0.389	0.467	**0.997**
	Recall	0.636	0.85	0.200	0.964	0.0	0.909	0.836	0.727	**1.0**	0.386
	F1	0.020	0.030	0.007	0.032	0.0	0.029	0.027	0.020	0.031	**0.557**
	AUC-ROC	0.005	0.018	0.010	**0.921**	0.500	0.790	0.691	0.502	0.858	0.693
	AUC-PR	0.015	0.040	0.005	**0.640**	0.008	0.031	0.015	0.008	0.043	0.400

Effect of Gaifman parameters (r=depth of neighborhood, w=the number of neighbors to sample, k=the number of neighborhoods)



Conclusion

- Provide a method of constructing relational embeddings that can be combined with a scalable local model
- First algorithm for learning the structure of Gaifman models for relational data
- Learned rules are instantiated and aggregated over a Gaifman graph to produce raw features
- Density estimation-based structure learning outperforms classical rule learning techniques

Future Work

- Joint learning of Gaifman models (multi relation prediction)
- Generating explanations for a given prediction
- Extending Gaifman locality to hypergraphs

DSD, SY and SN acknowledge DARPA Minerva award FA9550-19-1-0391. DSD acknowledges ICT-48 Network of AI Research Excellence Center \TAILOR" (EU Horizon 2020, GA No 952215) and the Collaboration Lab \AI in Construction" (AICO). SN acknowledges AFOSR award FA9550-18-1-0462. Any opinions, findings, conclusion or recommendations expressed are those of the authors and do not necessarily represent the view of AFOSR, DARPA or the US government.

https://starling.utdallas.edu/
https://sites.google.com/view/devendradhami

[$]Niepert 16, *Khot et al. 14, [&] Khot et al. 11, [+] Natarajan et al. 12, [%]Lao & Cohen 10, [#]Muggleton 94, Natarajan 09, Srinivasan 01.