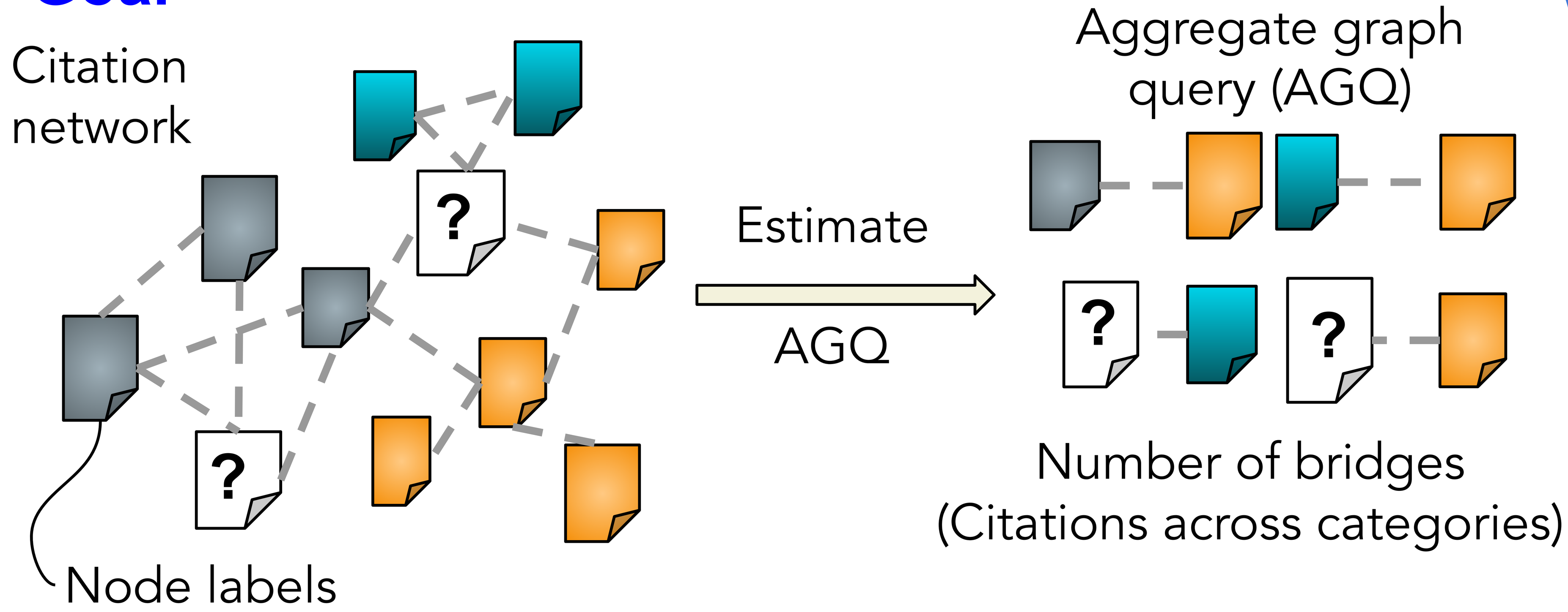
A Comparison of Statistical Relational Learning and Graph Neural Networks for Aggregate Graph Queries

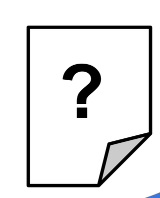
Varun Embar *, Sriram Srinivasan *, and Lise Getoor
University of California, Santa Cruz
*Equal Contribution

Goal



Challenge

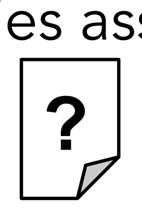
- Estimating aggregate graph query when network is not fully observed (E.g. missing node labels)



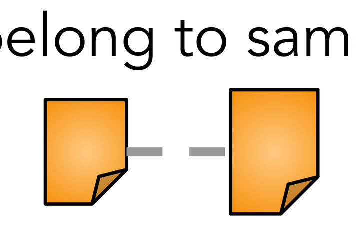
Aggregate graph queries

- Aggregate graph query (Q): Aggregate function computed on a set of subgraphs that satisfy given conditions (Q: graph → R)
  - Properties involving multiple nodes, edges and labels

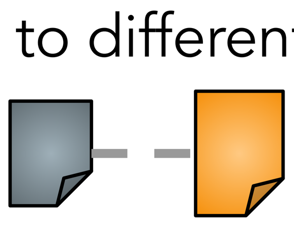
Q0: Accuracy:
# of documents with the correct categories assigned to them



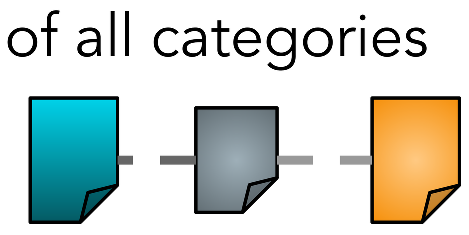
Q1: Edge cohesion:
# of links across documents that belong to same category



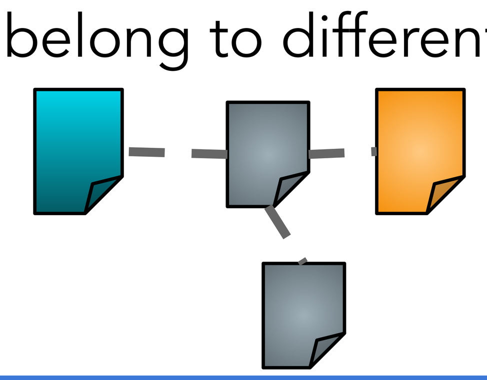
Q2: Edge separation:
# of links across documents that belong to different category



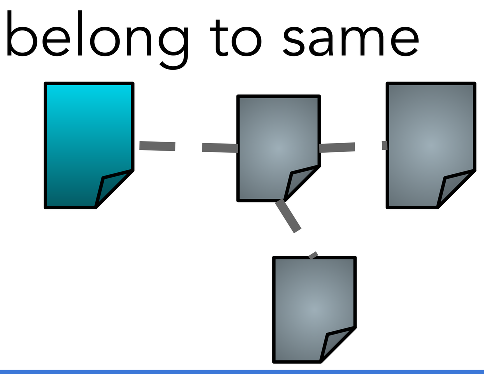
Q3: Diversity of influence:
# of nodes linked to at least half of all categories



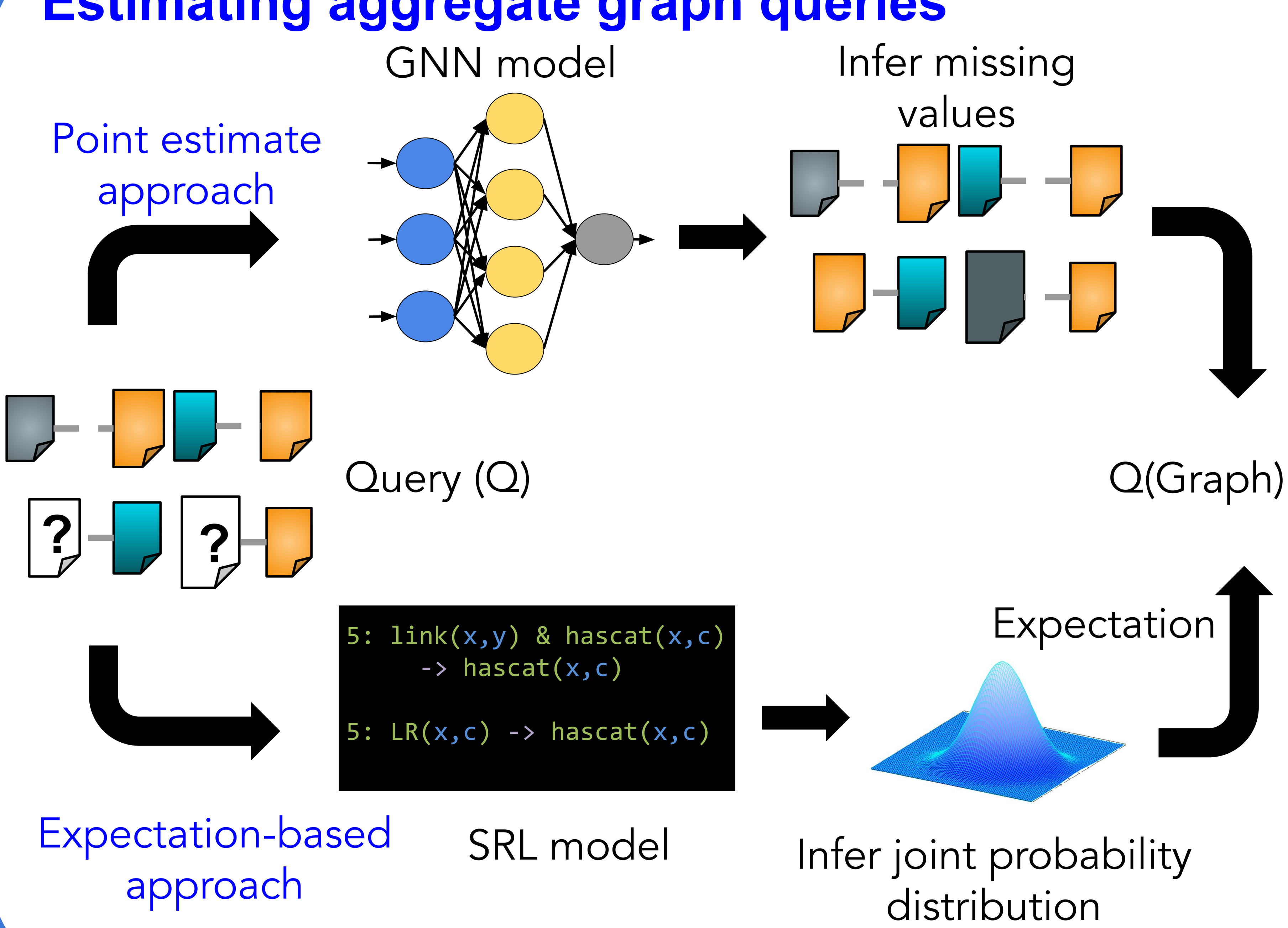
Q4: Exterior documents
# of nodes where half the neighbors belong to different categories



Q5: Interior documents
# of nodes where half the neighbors belong to same categories



Estimating aggregate graph queries



Tractable expectation computation for PSL

Theorem

For even simple graphs, generated using stochastic block models, with two nodes, point estimate approaches cannot minimize the expected mean squared error

Expectation-based approach

- Probabilistic soft logic[1] is a state-of-the-art SRL framework
- Computing expectation is intractable due to integration
- Monte Carlo approximation using samples from Gibbs sampler

Challenge

Conditional distribution for Gibbs sampler

$$p(y_i \mid X, Y_{-i}) \propto exp\{-\sum_{r=1}^{N_i} w_r \phi_r(y_i, X, Y_{-i})\}$$

Hard to sample from

- Single step of Metropolis sampler inside gibbs sampler

$$\alpha = \frac{exp\{-\sum_{r=1}^{N_i} w_r \phi_r(y_i', X, Y_{1:i-1}^{(t+1)}, Y_{i:n}^{(t)})\}}{exp\{-\sum_{r=1}^{N_i} w_r \phi_r(y_i, X, Y_{1:i-1}^{(t+1)}, Y_{i:n}^{(t)})\}}$$

Acceptance ratio

[1] http://psl.linqs.org

Experimental evaluation

Data: Pubmed, Citeseer and Cora

Graph Neural Networks : Graph Convolutional Networks (GCN), Graph Attention Network (GAT), Graph Markov Neural Networks (GMNN)
Statistical Relational Learning: Markov Logic Networks (MLN), Probabilistic Soft Logic (PSL)

Metric: Relative error

Aggregate property estimation:

Methods	Q0	Q1	Q2	Q3	Q4	Q5	AQE
GCN	**0.152**	0.129	0.524	2.732	0.737	0.126	0.733
GAT	0.168	0.144	0.583	2.364	0.764	0.132	0.692
GMNN	0.157	0.134	0.545	2.581	0.743	0.127	0.714
LR	0.219	0.126	0.513	6.342	0.712	0.120	1.33
MLN-MAP	0.205	0.075	0.362	3.613	0.435	0.080	0.795
PSL-MAP	0.170	0.016	0.064	4.259	**0.007**	**0.001**	0.752
MLN-SAM	0.223	0.037	0.070	0.057	0.051	0.007	0.073
PSL-SAM	0.171	**0.009**	**0.038**	**0.011**	0.022	0.003	**0.042**

Pubmed

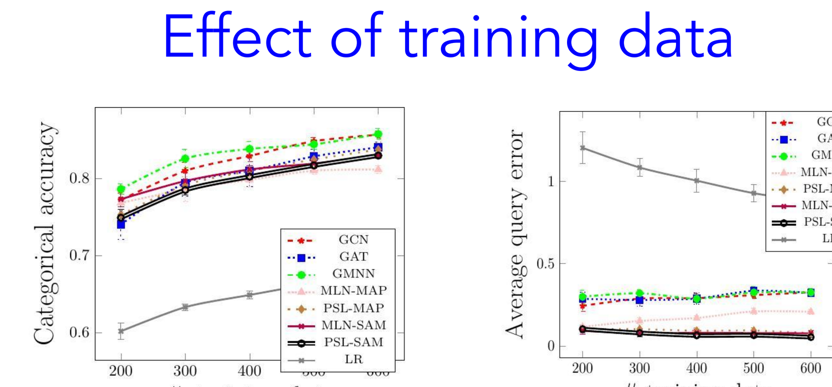
Methods	Q0	Q1	Q2	Q3	Q4	Q5	AQE
GCN	**0.263**	0.241	0.736	0.876	0.907	0.429	0.575
GAT	0.272	0.254	0.775	0.888	0.939	0.439	0.594
GMNN	0.268	0.251	0.766	0.867	0.905	0.412	0.578
LR	0.327	**0.134**	**0.408**	**0.378**	**0.382**	**0.176**	**0.300**
MLN-MAP	0.292	0.192	0.595	0.625	0.789	0.369	0.477
PSL-MAP	0.283	0.151	0.460	0.600	0.516	0.237	0.374
MLN-SAM	0.297	0.161	0.506	0.641	0.485	0.217	0.384
PSL-SAM	0.286	0.143	0.435	0.586	0.509	0.231	0.365

Citeseer

Methods	Q0	Q1	Q2	Q3	Q4	Q5	AQE
GCN	0.143	0.0756	0.323	0.281	0.768	0.363	0.325
GAT	0.159	0.076	0.326	0.281	0.729	0.361	0.322
GMNN	**0.142**	0.081	0.348	0.254	0.754	0.367	0.324
LR	0.324	0.320	1.371	1.854	0.993	0.401	0.709
MLN-MAP	0.188	**0.011**	0.110	0.136	0.529	0.268	0.207
PSL-MAP	0.162	0.027	0.116	0.063	0.060	0.034	0.077
MLN-SAM	0.170	0.021	0.092	0.068	0.074	0.035	0.076
PSL-SAM	0.170	0.015	**0.066**	**0.005**	**0.040**	**0.022**	**0.053**

Cora

Effect of training data



Cora

Runtime

Methods	Cora Time (sec)	Pubmed Time (sec)	Citeseer Time (sec)
GCN	24	59	29
GAT	142	138	122
GMNN	30	17	8
LR	2	5	2
PSL-MAP	14	124	37
MLN-MAP	65	368	36
PSL-SAM	105	638	124
MLN-SAM	270	1947	166

Conclusion

- Defined a suite of practical aggregate graph queries
- Proposed a novel sampling framework for PSL
- Extensive evaluation shows SRL approaches outperform GNNs when estimating aggregate graph queries

Supported by the NSF grants CCF-1740850,CCF-2023495, IIS-1703331, AFRL, IFDS DMS-2023495 and the DARPA.