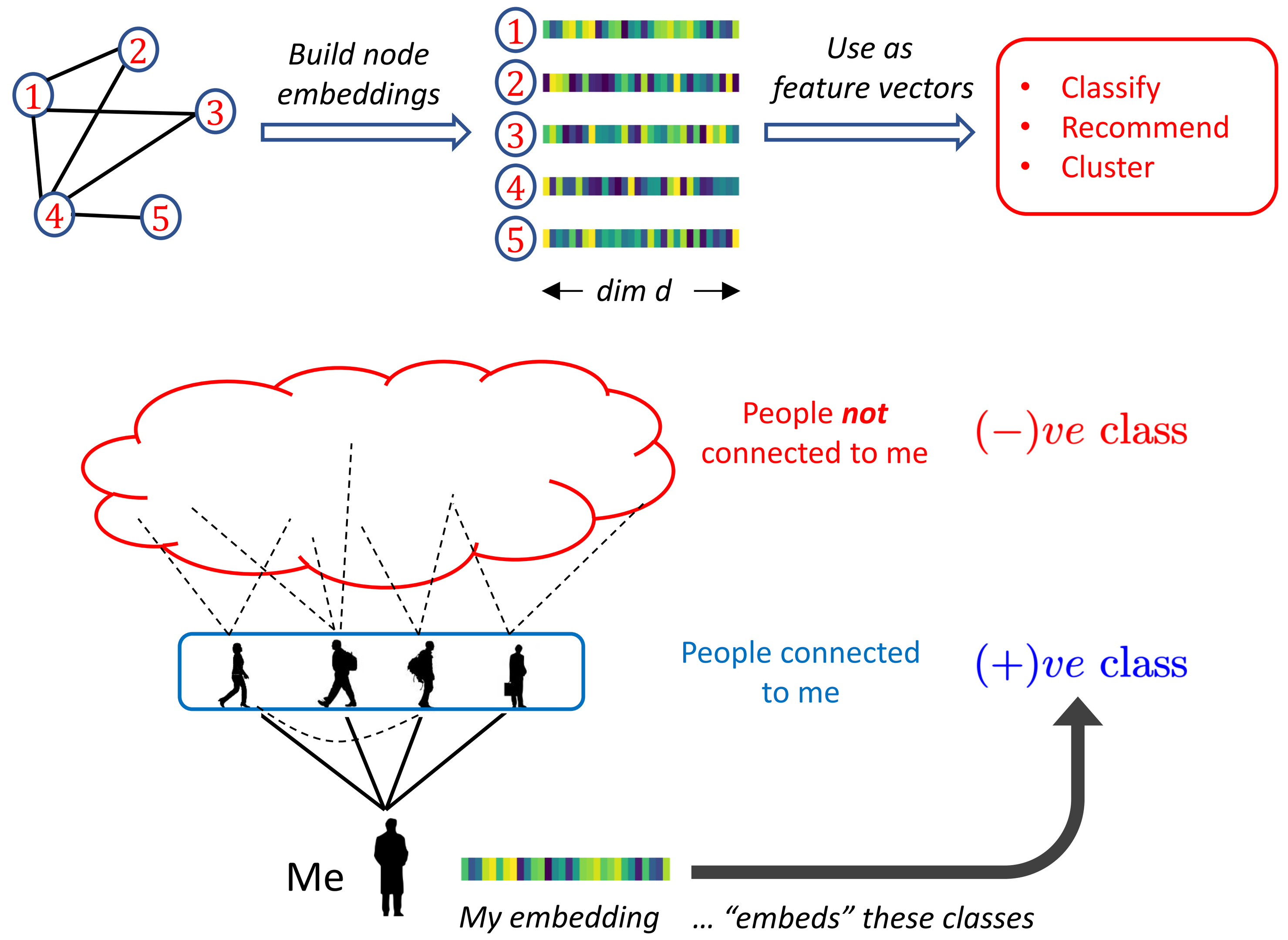
AVOIDING BIASES DUE TO SIMILARITY ASSUMPTIONS IN NODE EMBEDDINGS

Deepayan Chakrabarti (deepay@utexas.edu)

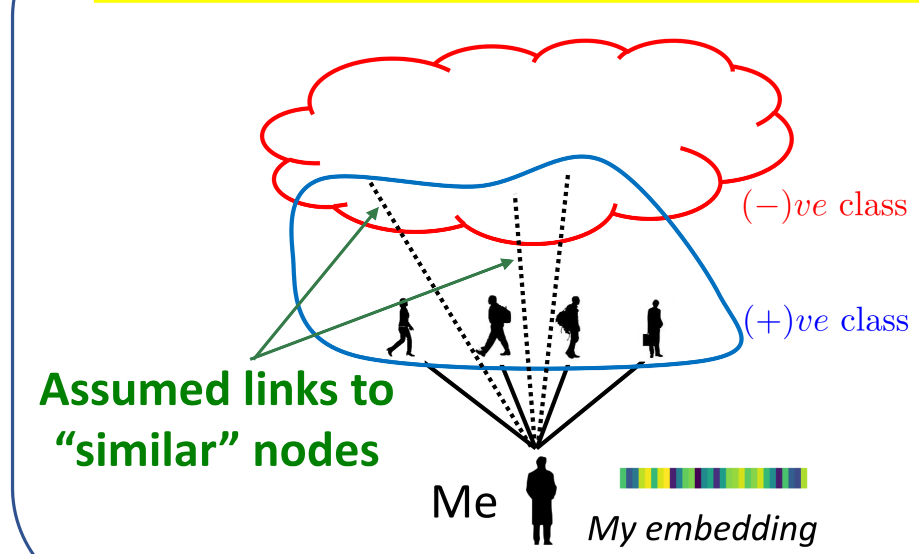
PROBLEM



Most nodes have low degrees ⇒

- **Very few positive examples**
- **Extreme class imbalance**

Existing Methods Assume A Similarity Measure



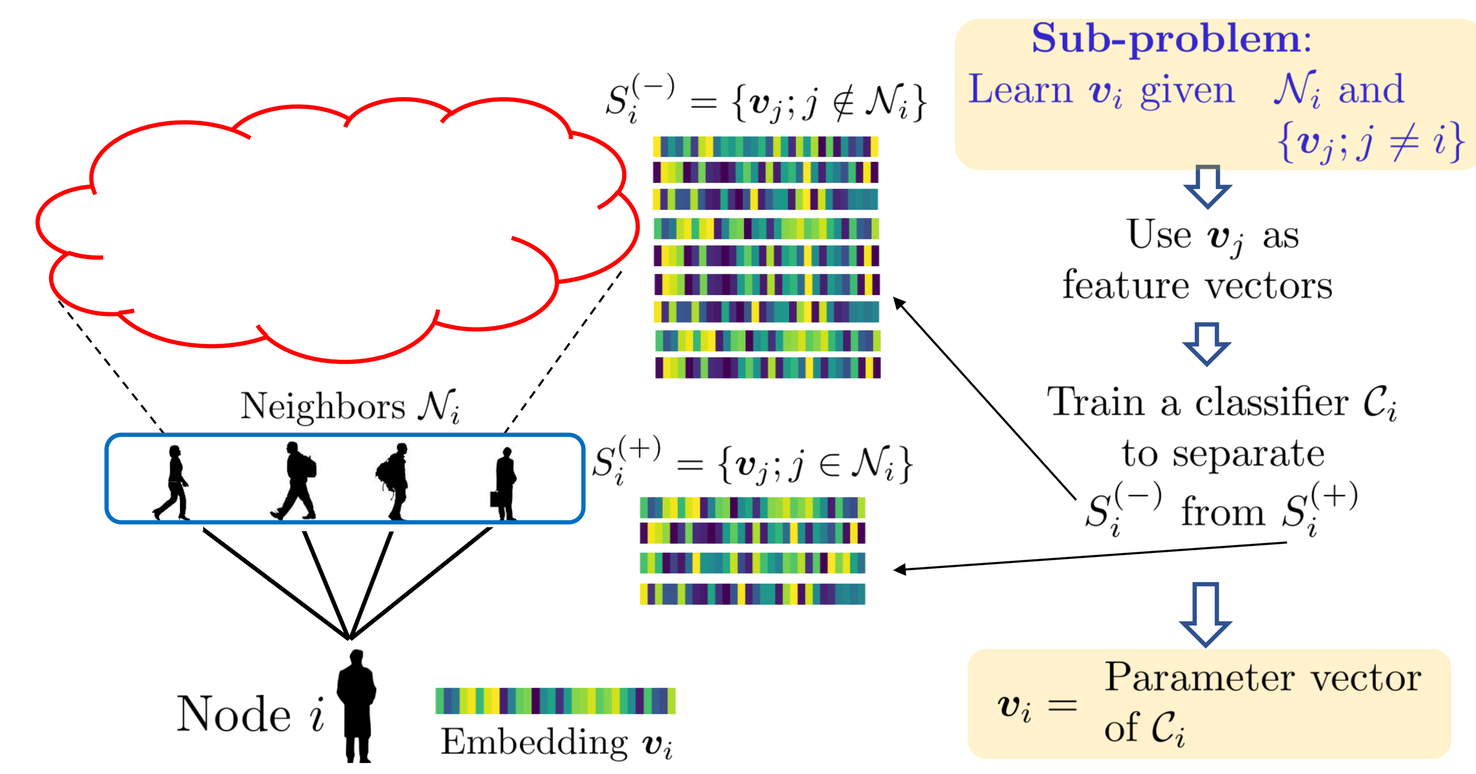
Weaknesses of Assumptions

- Hidden biases
- May not match intuition
- Costly re-computations for dynamic networks

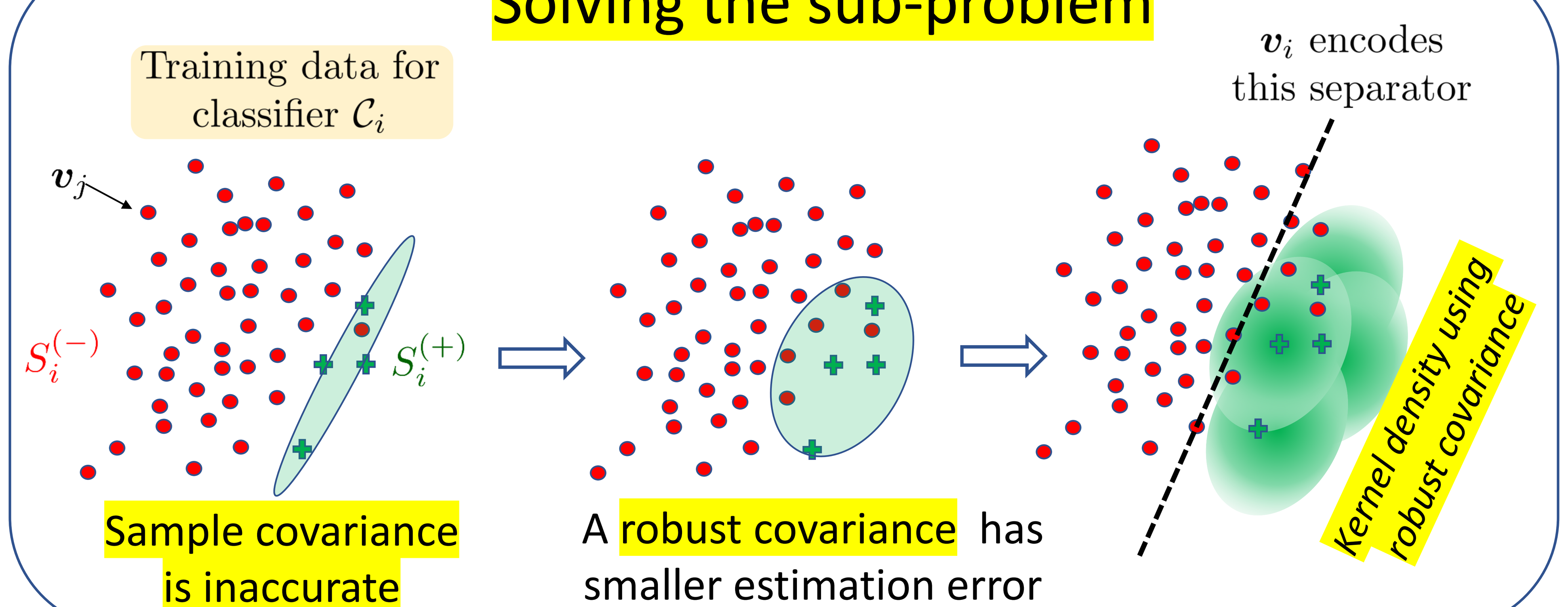
Goal: build personalized node embeddings without similarity assumptions

APPROACH

NEWS: Node Embeddings Without Similarity Assumptions



Solving the sub-problem



NEWS is robust, personalized, and parameter-free

https://github.com/deepayan12/news

RESULTS

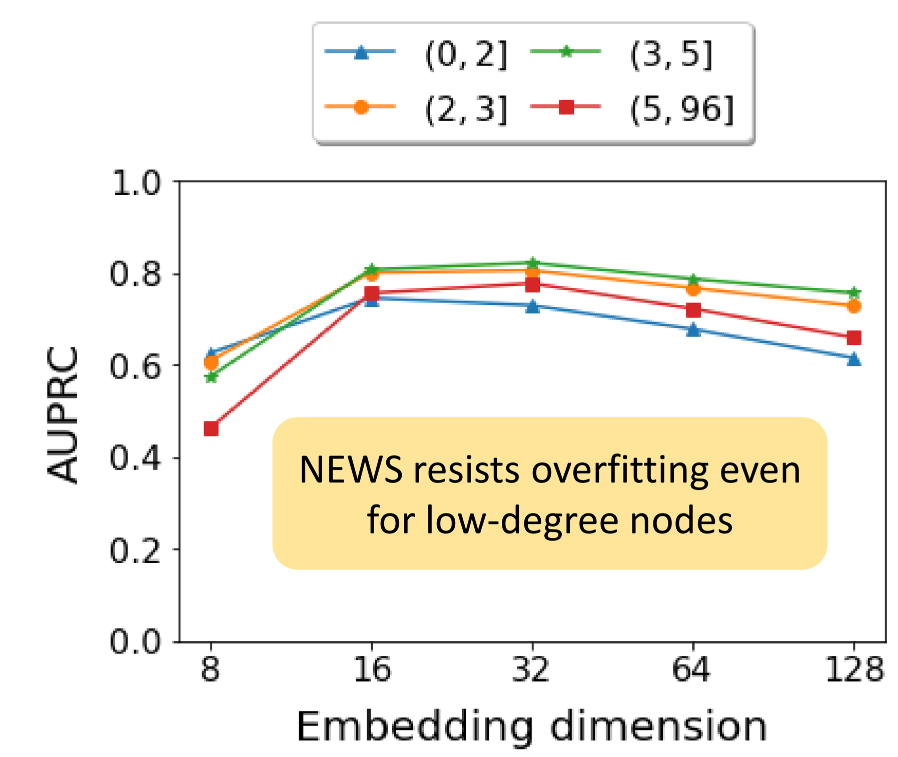
21 real-world datasets
Social, Collaboration, Citation, Product, Financial, Transportation, Biological, Location-based, …

Metric
Area under the Precision-Recall curve (AUPRC)

Link Prediction Accuracy

Energy-based: G2G; Matrix-based: HOPE; Random walks: Node2Vec; Auto-encoder: SDNE

Degree	G2G	GraRep	HOPE	LINE	Node2Vec	ProNE	SDNE	VERSE	NEWS
Cora (1, 434 **nodes,** 4, 256 **edges**)									
(0, 2]	0.04	0.48	0.10	0.17	0.69	0.48	0.25	0.70	0.76
(2, 3]	0.04	0.47	0.31	0.34	0.76	0.52	0.31	0.67	0.79
(3, 5]	0.04	0.44	0.26	0.23	0.66	0.46	0.28	0.63	0.70
(5, 10]	0.05	0.36	0.28	0.23	0.59	0.37	0.30	0.54	0.69
(10, 140]	0.09	0.37	0.36	0.27	0.55	0.29	0.25	0.49	0.54
Gowalla (196, 591 **nodes,** 950, 327 **edges**)									
(0, 4]	0.03	0.12	0.09	0.42	0.56	0.36	0.09	0.69	0.65
(4, 9]	0.03	0.20	0.10	0.54	0.72	0.48	0.13	0.77	0.79
(9, 17]	0.03	0.26	0.13	0.59	0.77	0.53	0.20	0.78	0.84
(17, 35]	0.04	0.34	0.19	0.64	0.80	0.59	0.34	0.79	0.86
(35, 14118]	0.11	0.54	0.41	0.74	0.88	0.68	0.58	0.81	0.90



Arxiv (Condensed Matter) network