

# Exploring Expert Cognition for Attributed Network Embedding

Xiao Huang,<sup>†</sup> Qingquan Song,<sup>†</sup> Jundong Li,<sup>‡</sup> and Xia (Ben) Hu<sup>†</sup>

<sup>†</sup>Computer Science & Engineering, Texas A&M University, College Station, TX, USA

<sup>‡</sup>Computer Science & Engineering, Arizona State University, Tempe, AZ, USA

Emails: {xhuang, qqsong, xiahu}@tamu.edu, jundongl@asu.edu

# Expert Cognition Benefits Data Analysis

- **Definition:** Meaningful and Intelligence-related info that experts know beyond the data.



- Understanding of domain knowledge;
- Awareness of conventions;
- Perception of latent relations.

- **Example:** Human understand the sentiment in product reviews. This cognition could be applied to enhance the recommendations.



Happy



Sad



Angry



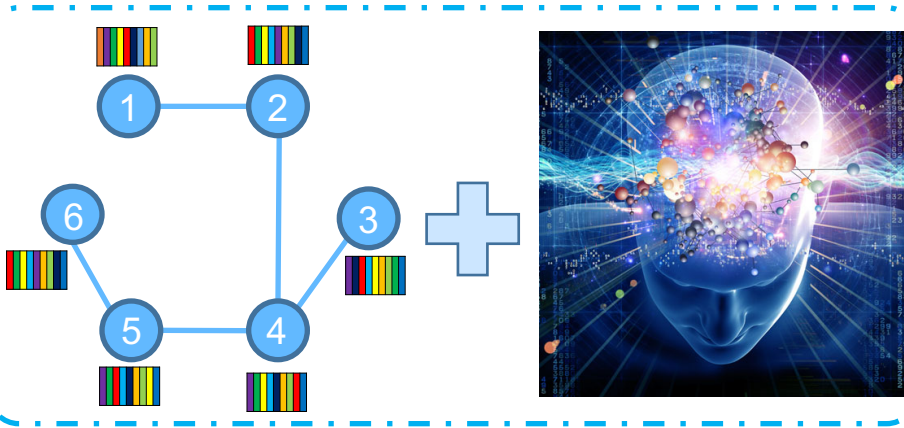
Surprised



Puzzled

# Model & Incorporate Expert Cognition into ANE

Initial Attributed Network



Embedding Representation

$$\mathbf{H} = \begin{bmatrix} 0.54 & 0.27 \\ 0.22 & 0.91 \\ 0.55 & 0.28 \\ 0.98 & 0.11 \\ 0.32 & 0.87 \\ 0.26 & 0.11 \end{bmatrix} \begin{matrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \end{matrix}$$

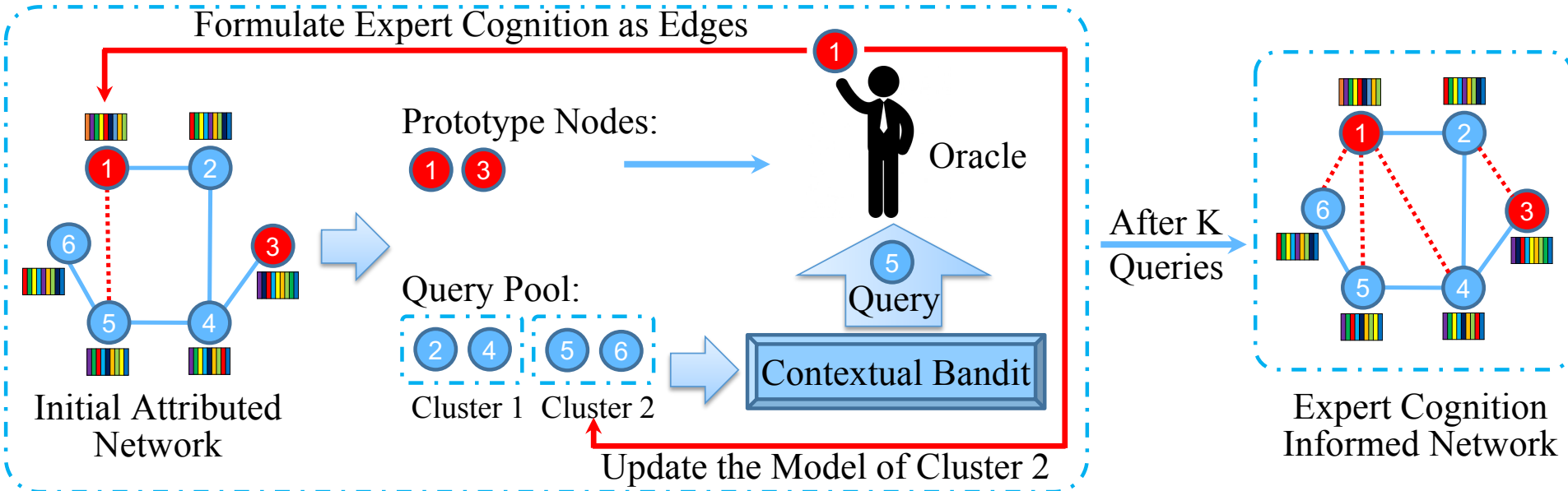
← Latent Space →

Tasks

- Classification
- Clustering
- Link Prediction
- Visualization
- ...

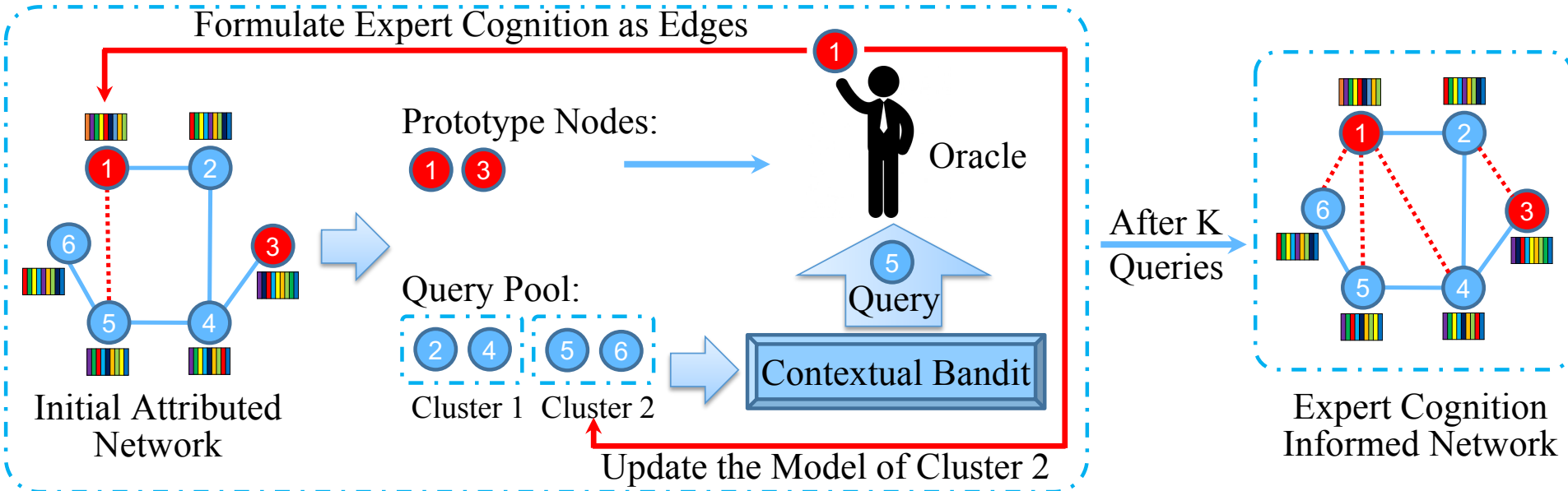
- ANE serves as infrastructures of various real-world applications.
- We aim to
  - Design a general and concise form of queries to learn expert cognition from the oracle while greatly save his/her effort.
  - Learn cognition from experts and incorporate it into ANE to advance downstream analysis algorithms.

# Strategies of NEEC



- 1) Two steps to find the top  $K$  meaningful queries.
  - a. Find few representative and distinct nodes (in red) as prototypes
  - b. Iteratively selects  $K$  nodes from the remaining nodes (in blue) with the largest amount of expected learned expert cognition.
- 2) Oracle needs to indicate a node from the prototypes (e.g.,  $j = 1$ ) that is the most similar to the queried node  $i = 5$ .

# Strategies of NEEC



- 3) Answers will be added into the network structure in the form of weighted edges, named as cognition edges (red dotted lines).
- 4) With these cognition edges, different ANE methods can be directly applied to the expert cognition informed network to learn  $\mathbf{H}$ .

# Experimental Settings

- To create the oracle, we randomly sample a certain percentage of the entire edges and attributes as the initial attributed network.
- The remaining data is considered as the cognition of the oracle, so he/she answers the queries based on the entire original dataset.
- Classification on three real-world network.
  - BlogCatalog (5,196 nodes)
  - Flickr (7,575 nodes)
  - ACM (16,484 nodes)

# Experimental Results

Training	BlogCatalog				Flickr			
	10%	25%	50%	100%	10%	25%	50%	100%
Node Num	1,455	2,078	3,118	5,196	2,121	3,030	4,545	7,575
K	242	495	1,114	3,092	338	691	1,556	4,316
Initial	0.287	0.359	0.430	0.520	0.261	0.304	0.358	0.432
RandomPair	0.286	0.358	0.431	0.520	0.261	0.306	0.365	0.435
LinkPredict	0.287	0.360	0.430	0.523	0.261	0.303	0.359	0.433
AddKEdges	0.312	0.379	0.462	0.547	0.255	0.312	0.363	0.457
HiddenEdge	0.298	0.370	0.437	0.529	0.265	0.301	0.363	0.436
w/o_Bandit	0.295	0.366	0.447	0.556	0.259	0.326	0.368	0.439
NEEC_Rand	<b>0.317</b>	<b>0.393</b>	0.463	0.558	0.277	<b>0.367</b>	0.433	0.508
NEEC_Kmed	0.313	0.389	<b>0.469</b>	<b>0.571</b>	<b>0.281</b>	0.364	<b>0.443</b>	<b>0.515</b>
RandomPair	-0.47%	-0.27%	0.07%	-0.07%	0.04%	0.87%	1.87%	0.64%
LinkPredict	-0.03%	0.35%	-0.07%	0.41%	0.00%	-0.36%	0.15%	0.15%
AddKEdges	8.43%	5.55%	7.36%	5.14%	-2.28%	2.79%	1.20%	5.83%
HiddenEdge	3.61%	3.06%	1.52%	1.68%	1.60%	-0.80%	1.41%	0.81%
w/o_Bandit	2.64%	2.06%	3.82%	6.75%	-0.80%	7.43%	2.61%	1.58%
NEEC_Rand	<b>10.41%</b>	<b>9.63%</b>	7.69%	7.27%	6.24%	<b>20.87%</b>	20.78%	17.43%
NEEC_Kmed	9.03%	8.34%	<b>9.01%</b>	<b>9.67%</b>	<b>7.64%</b>	19.78%	<b>23.60%</b>	<b>19.22%</b>

➤ NEEC achieves the largest amount of improvements after querying the oracle K simple questions.