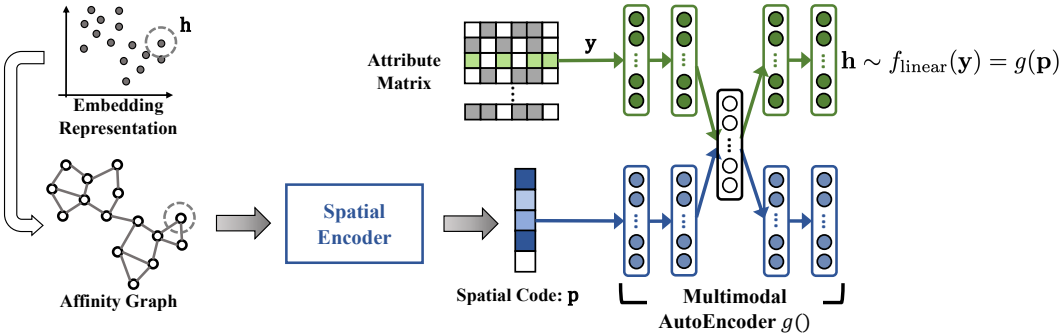
Attributed Network Embedding

- **Motivations & challenges**
- **Mining attributed networks with shallow embedding**
  - Coupled spectral embedding
  - Coupled matrix & tri-factorization
  - Random walk based embedding
- **Mining attributed networks with deep embedding**
  - Objective function based deep embedding
  - Graph neural networks
- **Human-centric network analysis**
  - Interpretable node representation learning
  - Attributed network analysis with humans in the loop

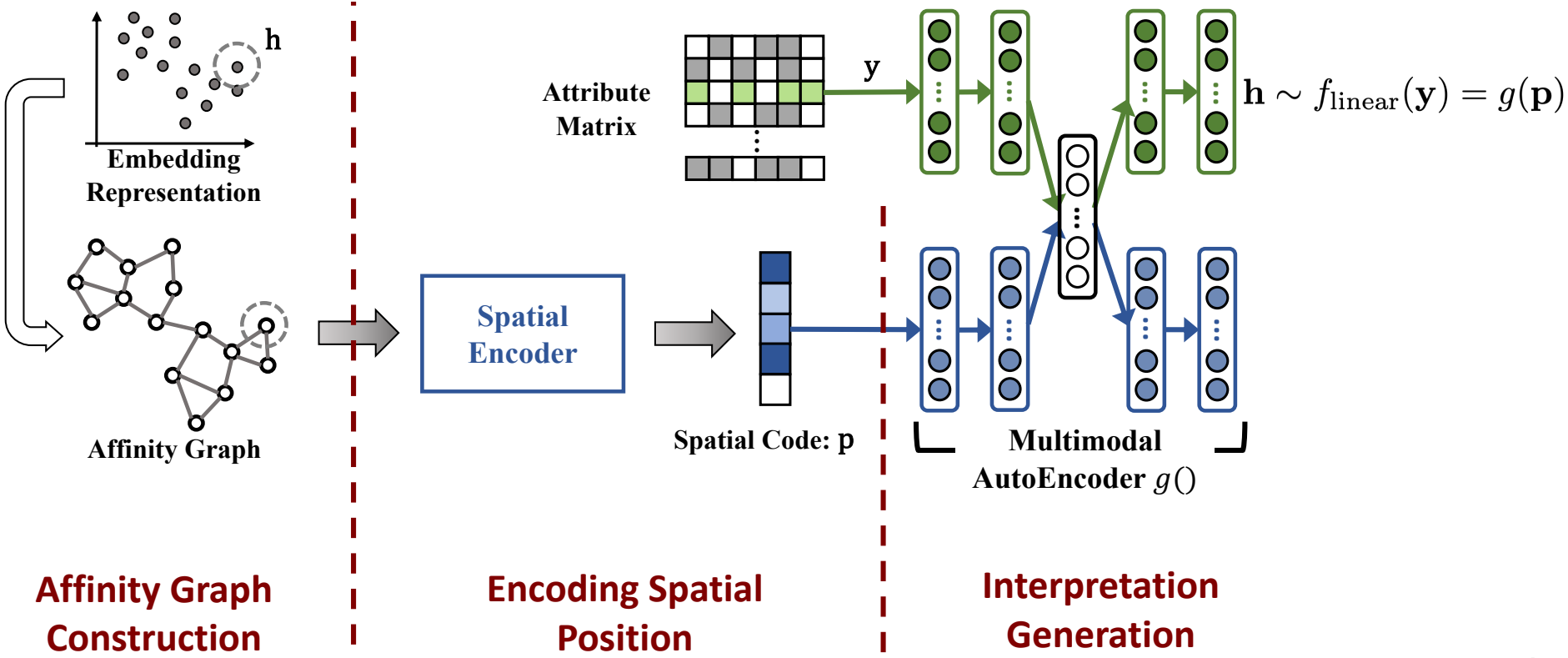
Interpretable node representation learning



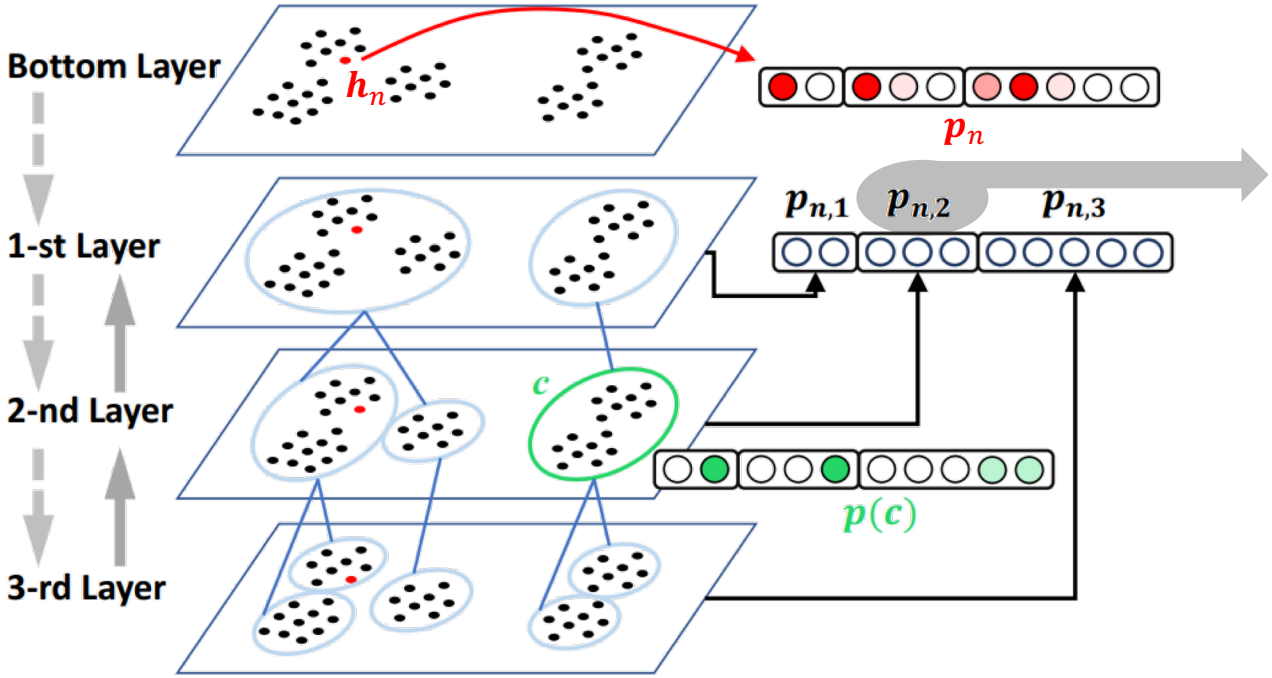
[Liu et al. WSDM, 2019]

- Opacity of embedding space
  - How representation vectors distribute in the embedding space?
  - What information is encoded in different embedding space regions?
  - Existing methods for explaining classifiers are not directly applicable
- Comprehensible node attributes are available
- Goal: Mining explainable structures and identifying characteristic factors from the mass of representation vectors

Spatial encoding and multimodal analytics



Spatial encoding



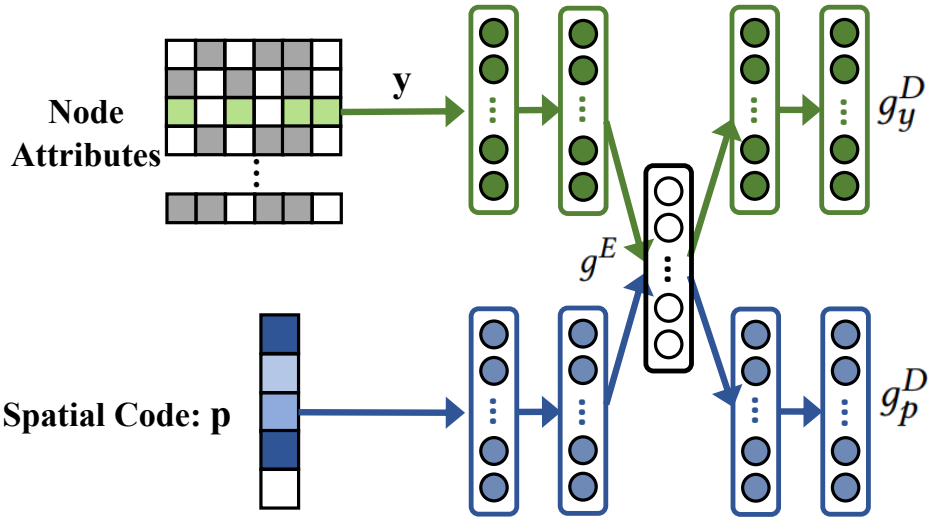
$$\mathbf{p}_{n,l} = \mathbf{e}_n \mathbf{W}_L \mathbf{W}_{L-1} \ldots \mathbf{W}_l$$

$\mathbf{e}_n \in \mathbb{R}^N$ is a one-hot vector, where $\mathbf{e}_n(i) = 1$ for $i = n$

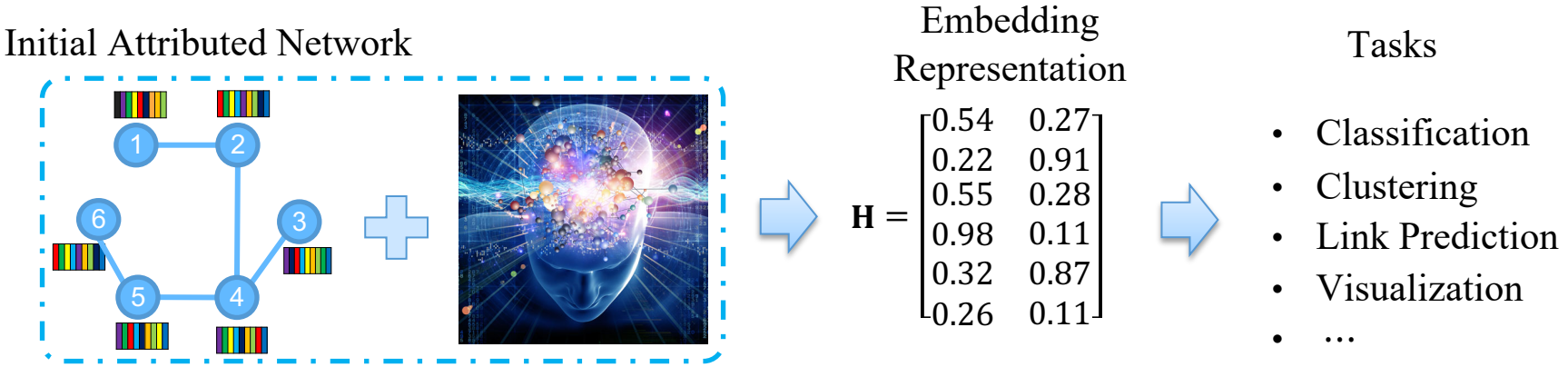
The spatial code for node n is $\mathbf{p}_n = [\hat{\mathbf{p}}_{n,1}, \hat{\mathbf{p}}_{n,2}, \ldots, \hat{\mathbf{p}}_{n,L-1}, \hat{\mathbf{p}}_{n,L}]$

Multimodal autoencoder

- $\boldsymbol{y}$ are comprehensible node attributes
- Variational autoencoder is used to reconstruct $\boldsymbol{y}$ and $\boldsymbol{p}$
- After training the autoencoder, the interpretation for embedding representation $\boldsymbol{h}$ is,
  - $\mathbf{h} \sim f_{\text{linear}}(\mathbf{y}) = g(\mathbf{p}) = g_y^D \circ g^E(\mathbf{p}, \mathbf{0})$
  - The input to the node attribute side is set to be absent
  - The output from node attribute decoder is used as the interpretation



Attributed network analysis with humans in the loop



[Huang et al. WSDM, 2018]

- Attributed network embedding (ANE) serves as infrastructures of various real-world applications
- We aim to learn cognition from experts and incorporate it into ANE to advance downstream analysis algorithms

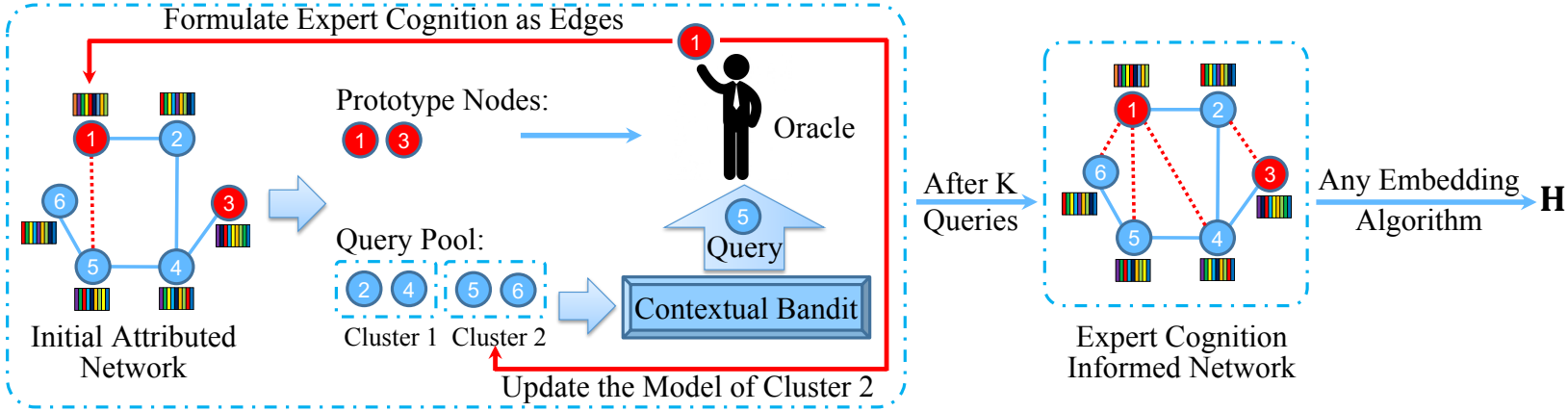
Expert cognition benefits data analysis

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

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

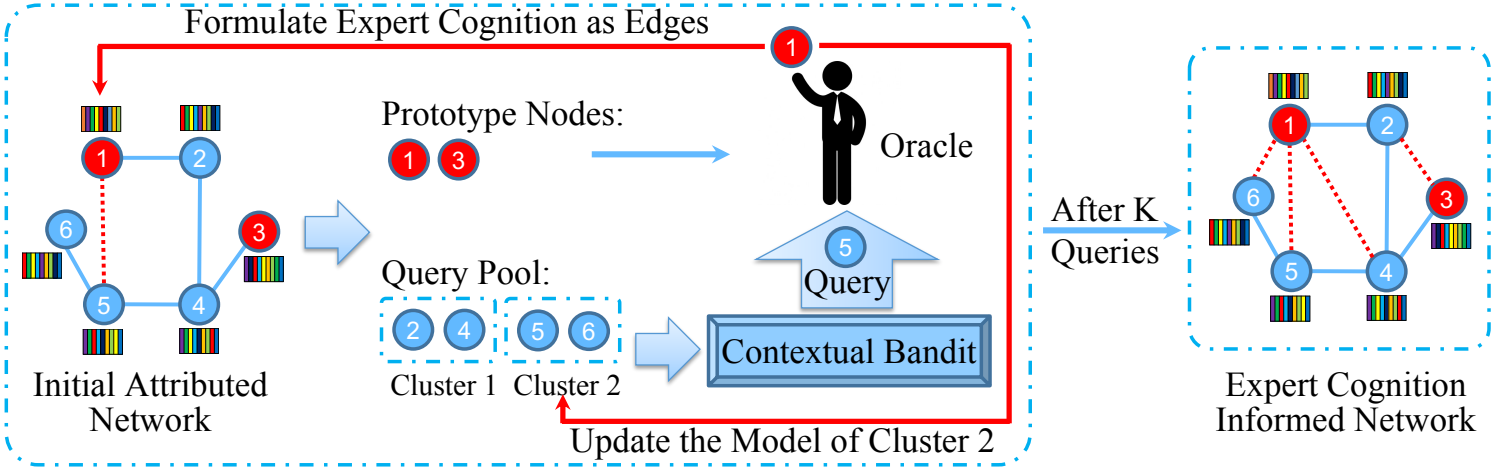
- Example: Human understand the sentiment in product reviews. This cognition could be applied to enhance the recommendations

Network embedding with expert cognition - NEEC



- Convert the abstract and meaningful cognition of domain experts into concrete answers
- Incorporate answers into ANE towards a more informative $\mathbf{H}$
- Employ a general and concise form of queries to learn expert cognition from the oracle while greatly saving his/her effort

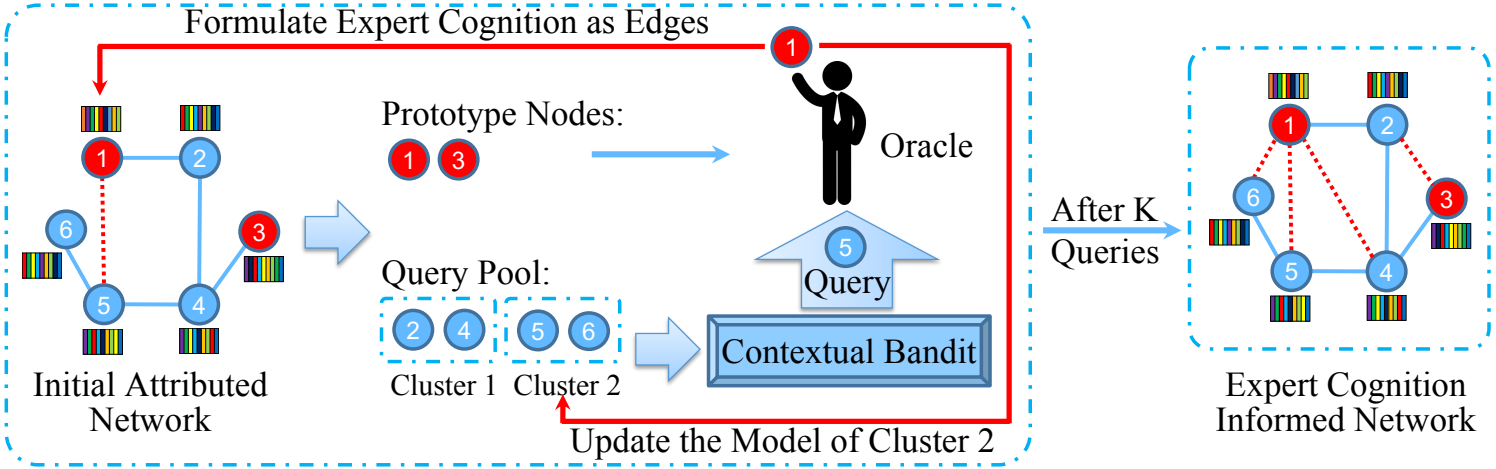
Strategies of framework NEEC



- Two steps to find the top K meaningful queries
  - Find few representative and distinct nodes (in red) as prototypes
  - Iteratively select K nodes from the remaining nodes (in blue) with the largest amount of expected learned expert cognition

- 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 framework NEEC



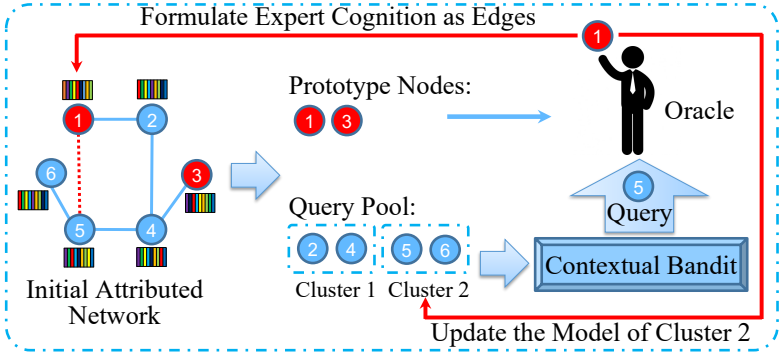
- Answers will be added into the network structure in the form of weighted edges, named as cognition edges (red dotted lines)
- With these cognition edges, different ANE methods can be directly applied to the expert cognition informed network to learn $\mathbf{H}$

Human-centric network analysis

- **Focuses:**
  Interpretable embedding, & utilizing network embedding to incorporate human knowledge

- **Methods:**
  Interpretable node representation learning
  Attributed network analysis with humans in the loop

- **Techniques:**
  Linking embedding with interpretable node attributes, converting knowledge into links, etc.



Summary of attributed network embedding

- ANE learns low-dimensional vectors to represent all nodes, bridging the gap between real-world systems & ML algorithms
- Challenges: Heterogeneity, large-scale, & Data Characteristics Vary Significantly
- Compare with other research topics
  - Multiview learning: Learn a unified representation of instances from multiple feature matrices observed from different aspects
  - Multimodal learning: Embed multiple sources with distinct modalities such as networks, images, and audio
  - Attributed network embedding: Preserve proximity information in networks and (one or multiple types of) node attributes

Summary of Attributed Network Embedding

- Shallow attributed network embedding:
  - Coupled spectral embedding
  - Coupled matrix & tri-factorization
  - Random walk based embedding
- Deep attributed network embedding:
  - Objective function based deep embedding
  - Graph neural networks
- Comprehensible node attributes help humans interact with systems.
  - Interpretable node representation learning
  - Attributed network analysis with humans in the loop