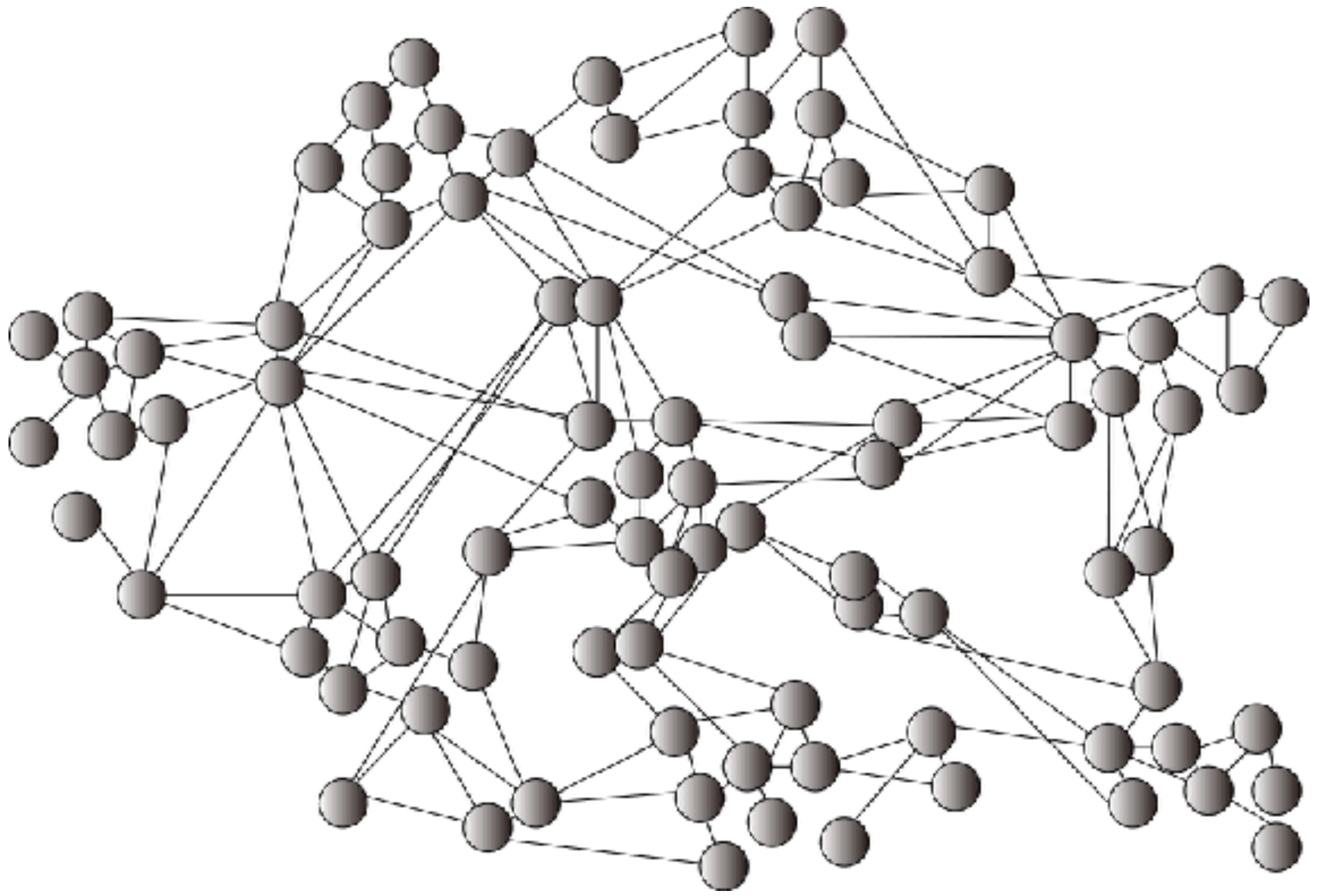
NavWalker: Information Augmented Network Embedding

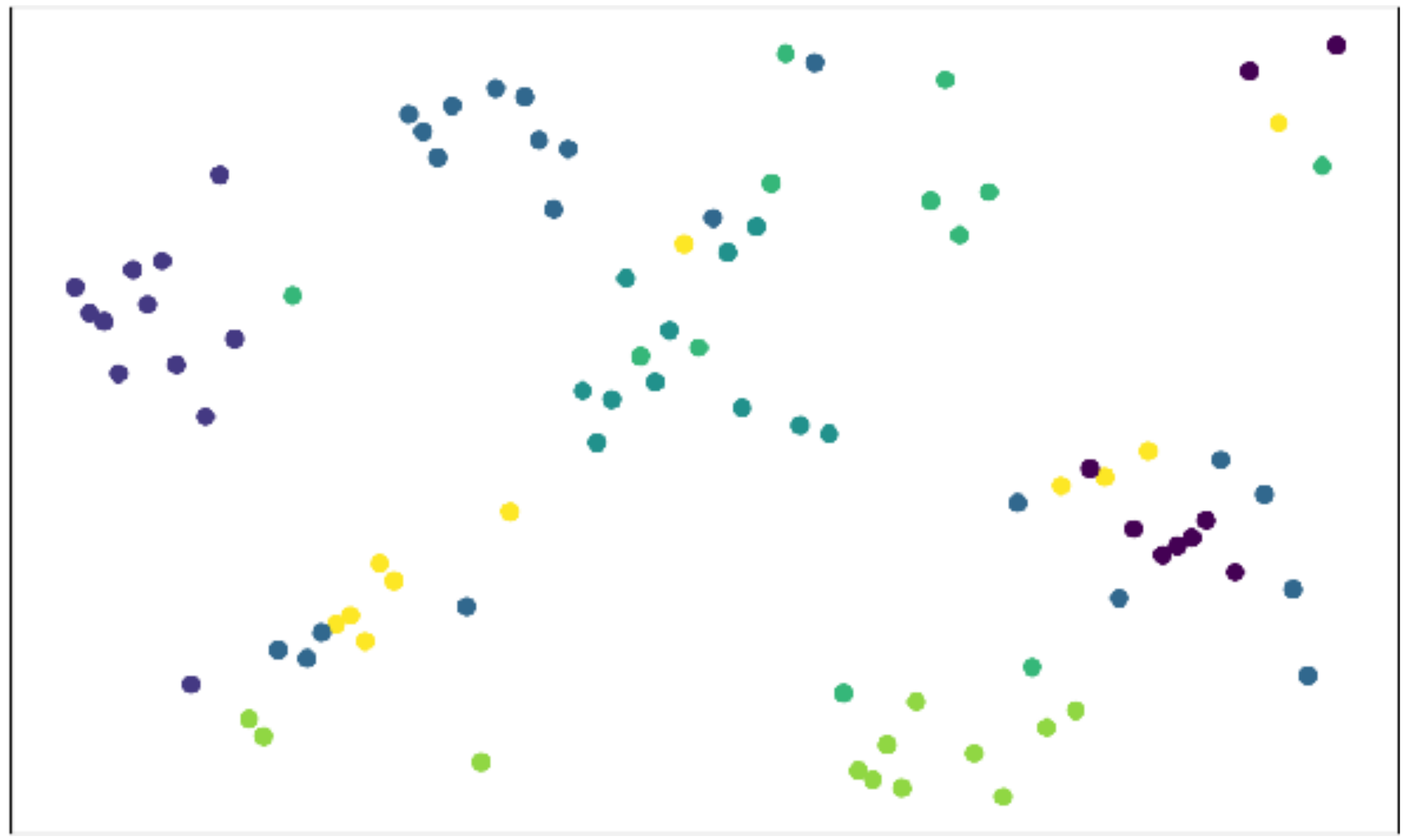
Kwei-Herng Lai[*], Chih-Ming Chen[§], Ming-Feng Tsai[§], Chuan-Ju Wang[*]
[*]Research Center for Information Technology Innovation, Academia Sinica
[§]Department of Computer Science, National Chengchi University
Email: *{khlai,cjwang}@citi.sinica.edu.tw, §{104761501,mftsai}@nccu.edu.tw

Brief Introduction

Toward understanding network embedding and its applications

"Automatically discover and map a network's structural properties into a latent space."



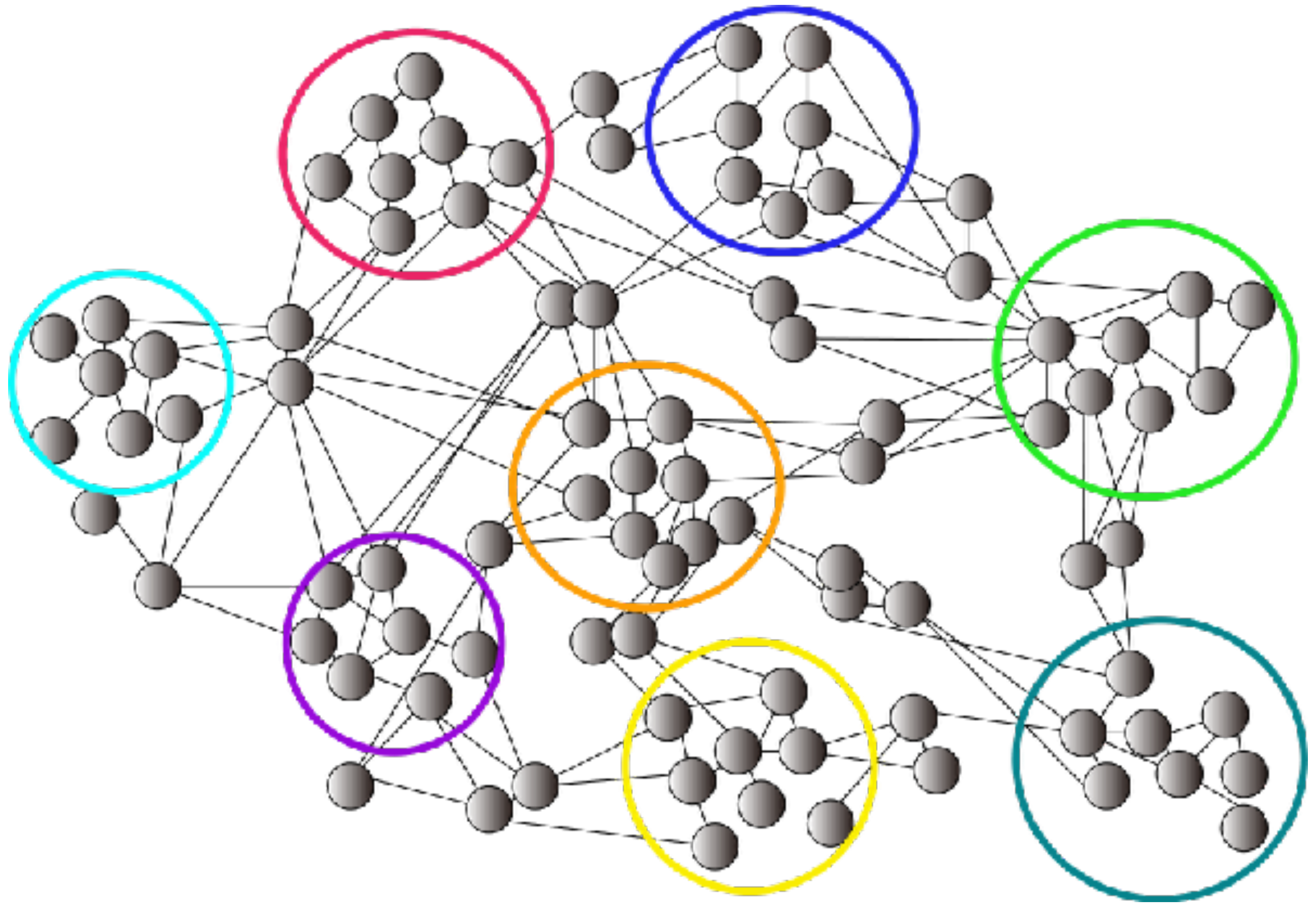


Ref: Network Embedding as Matrix Factorization: Unifying DeepWalk, LINE, PTE, and node2vec, Proceedings of WSDM'18

Applications

Community Detection

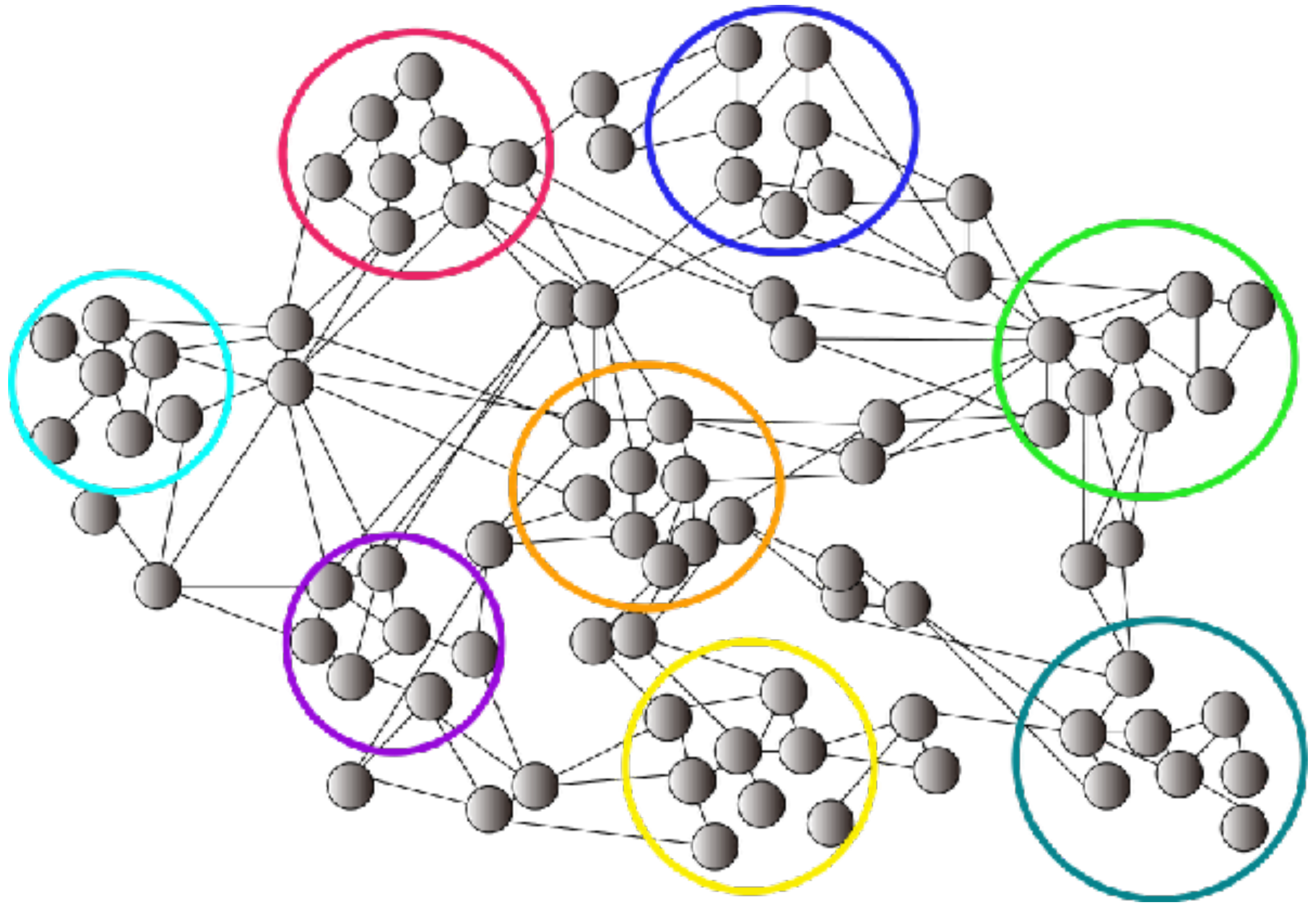
- Clustering
  - KNN
- Visualization
  - t-SNE, PCA

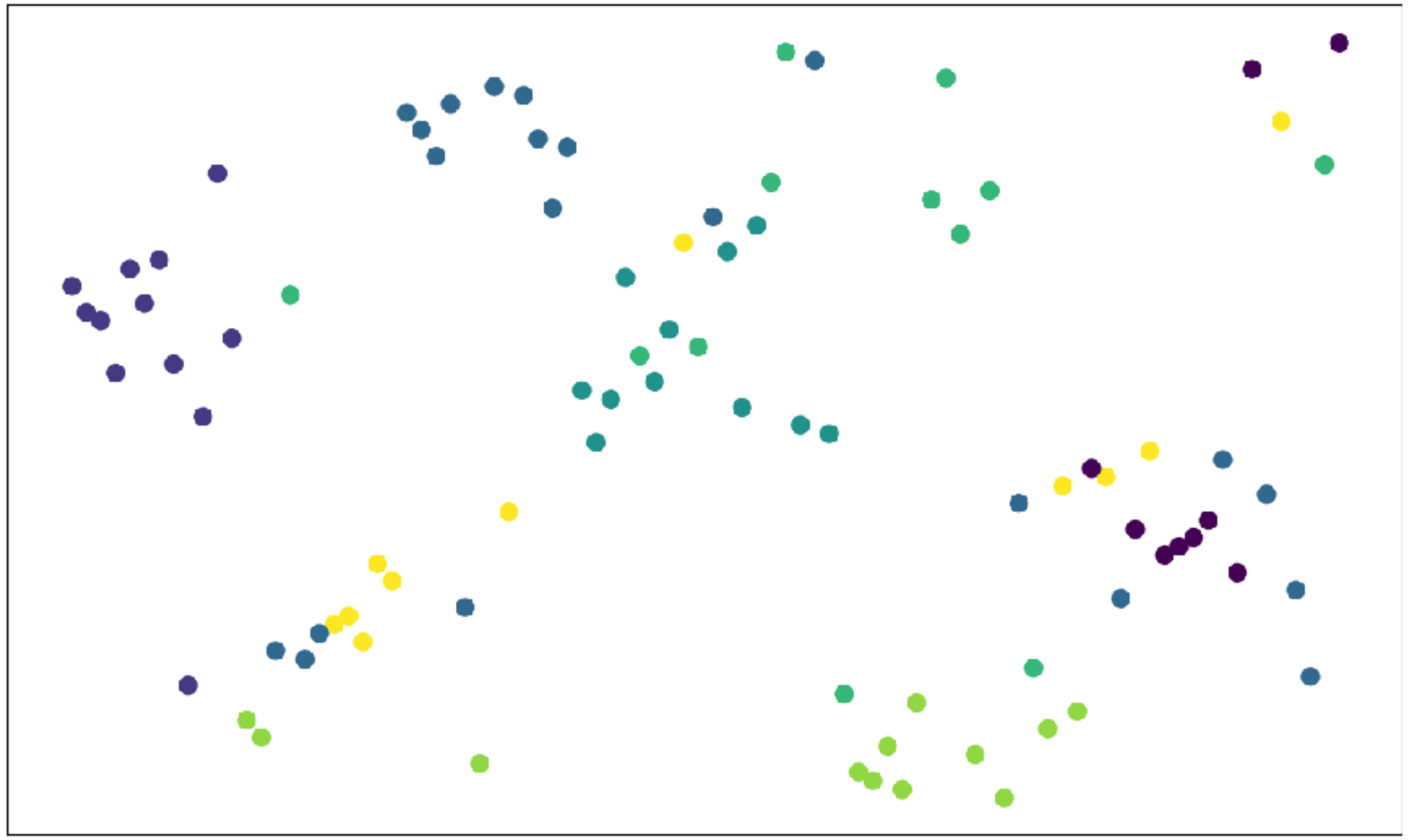


Applications

Community Detection

- Clustering
    - KNN
- Visualization
    - t-SNE, PCA

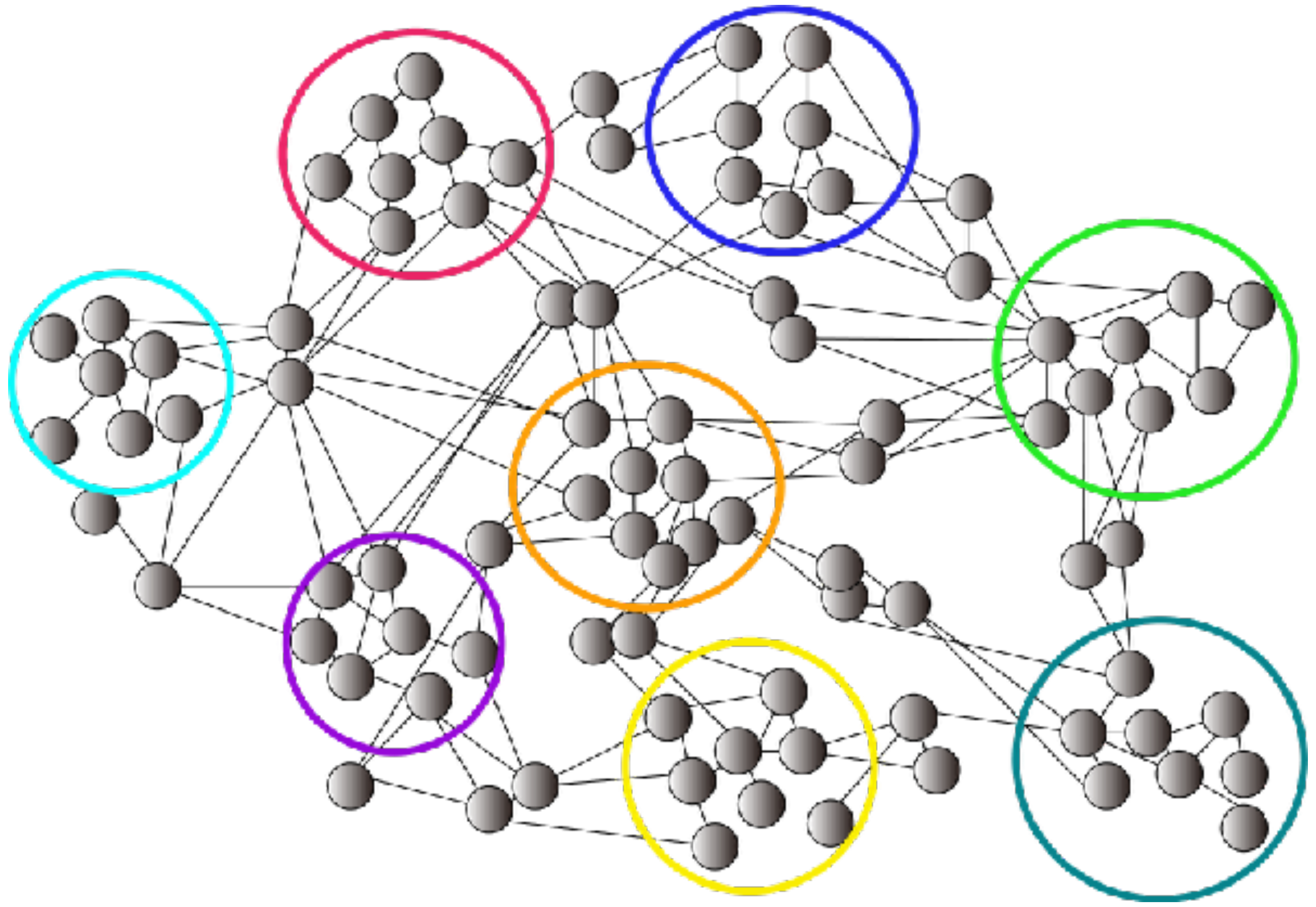




Applications

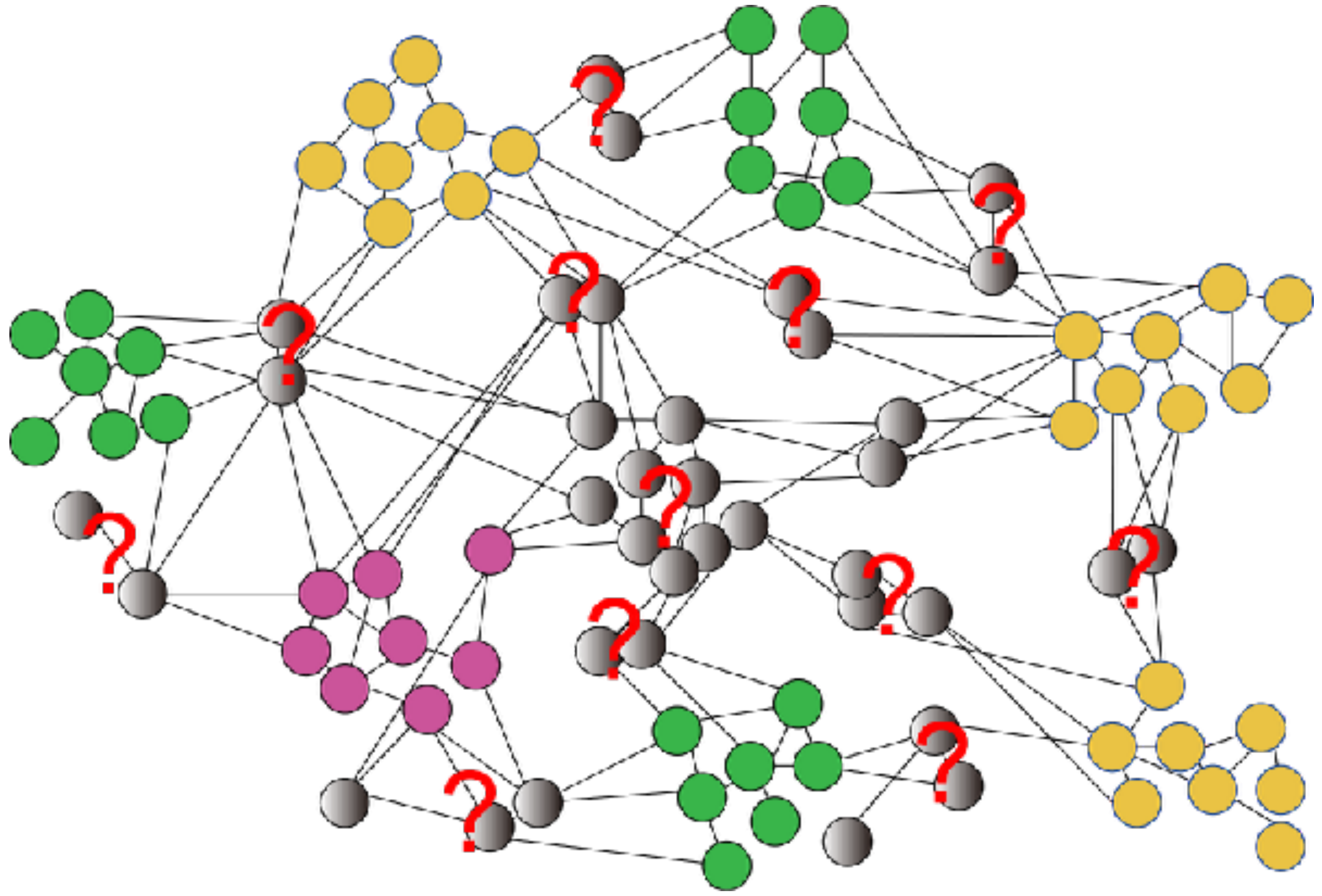
Community Detection

- Clustering
  - KNN
- Visualization
  - t-SNE, PCA



Node Classification

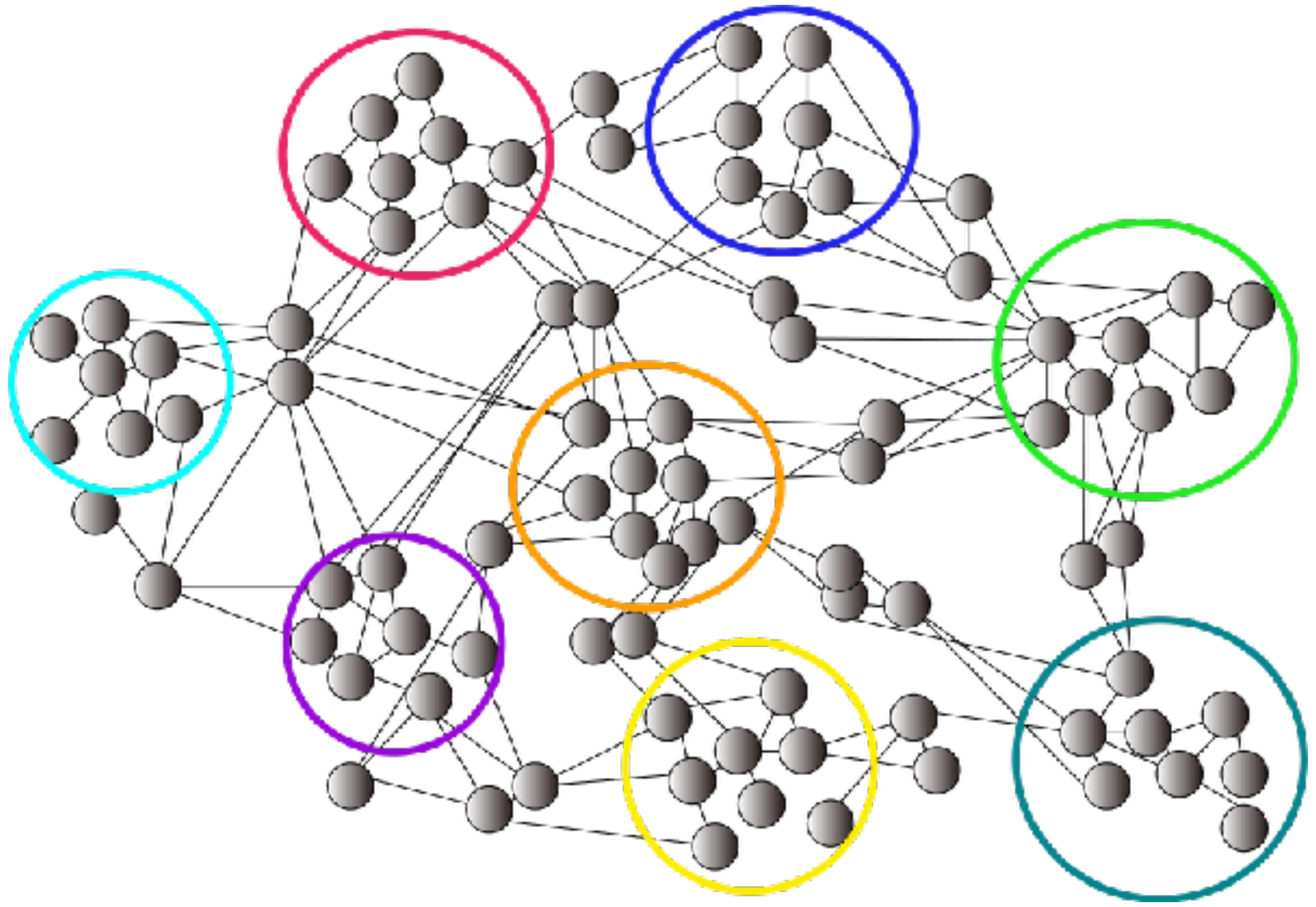
- Labeling full graph
  - Document Classification
  - Protein Classification



Applications

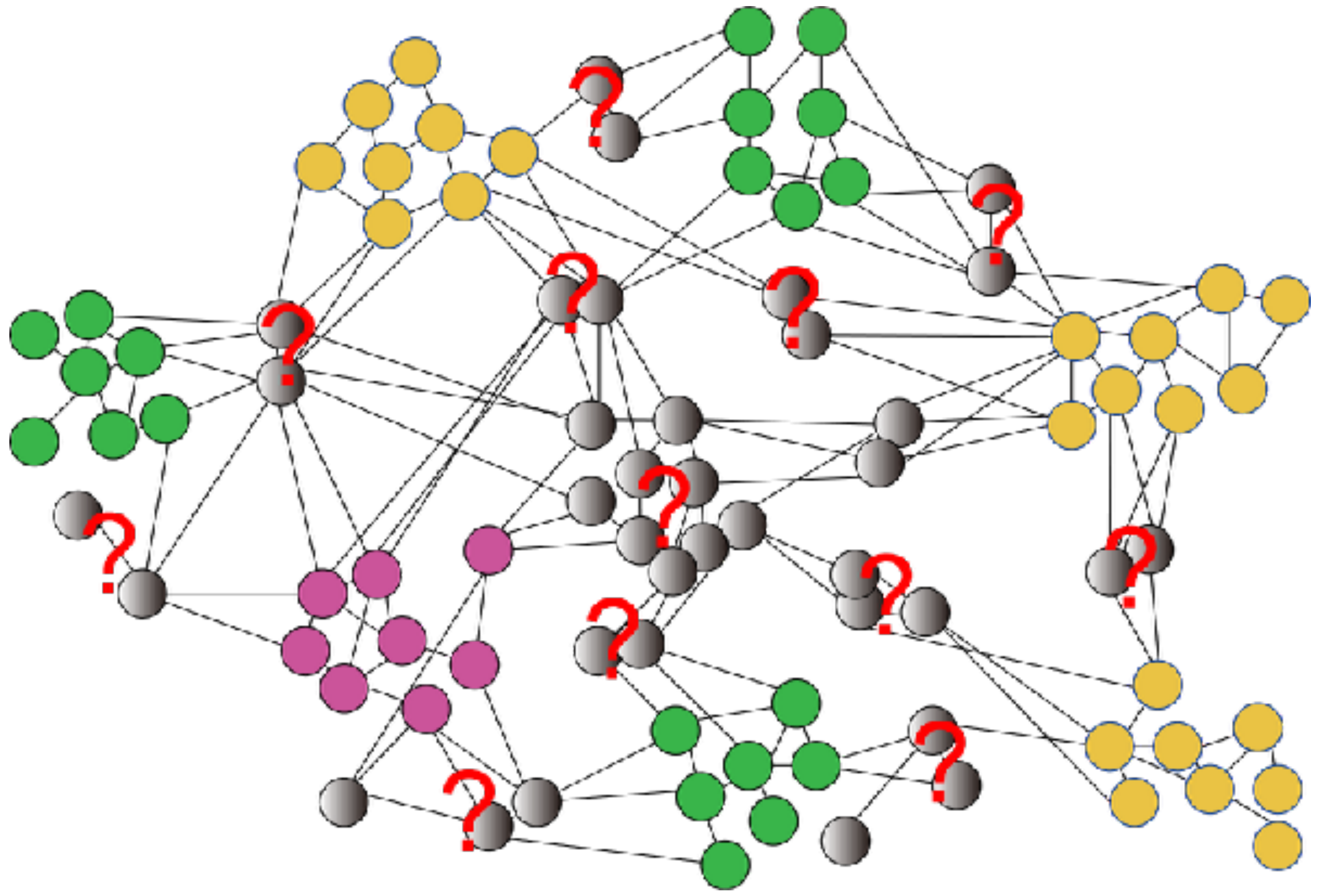
Community Detection

- Clustering
  - KNN
- Visualization
  - t-SNE, PCA



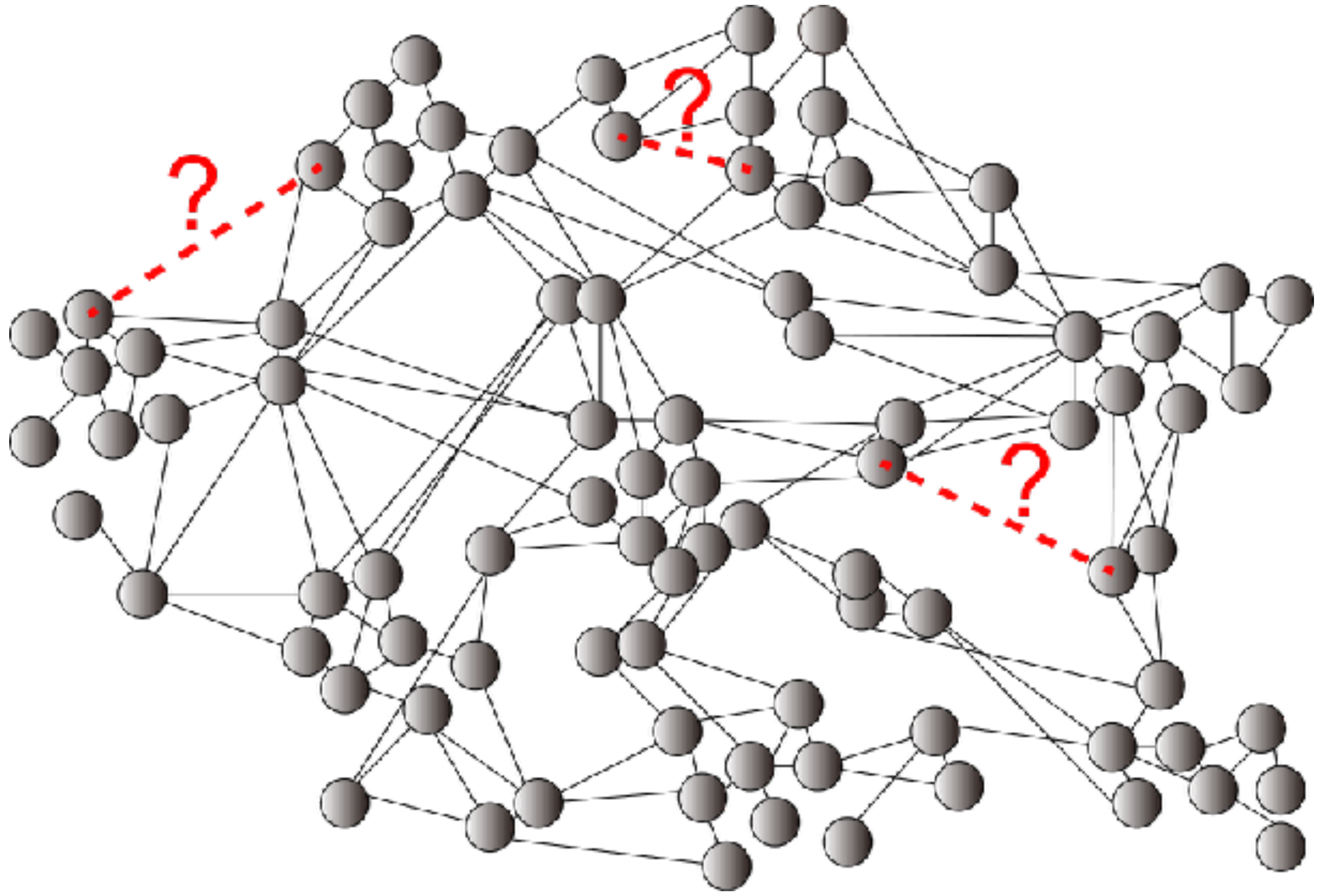
Node Classification

- Labeling full graph
  - Document Classification
  - Protein Classification



Link Prediction

- Predicting missing relations
  - Recommender system
  - Latent structure prediction



Related Works

[A bank is a financial institution] that accepts deposits from the public and creates credit.

Distributional Hypothesis

Word2vec

- Distributional Hypothesis
- Skip-gram/CBOW

Related Works

[A bank is a financial institution that] accepts deposits from the public and creates credit.

Distributional Hypothesis

Word2vec
- Distributional Hypothesis
- Skip-gram/CBOW

LINE
- 1st/2nd proximity
- Edge sampling

DeepWalk
- Skip-gram
- Random walk

Related Works

[A bank is a financial institution that accepts] deposits from the public and creates credit.

Distributional Hypothesis

Word2vec
- Distributional Hypothesis
- Skip-gram/CBOW

LINE
- 1st/2nd proximity
- Edge sampling

DeepWalk
- Skip-gram
- Random walk

HPE
- LINE + DeepWalk

Related Works

[A bank is a financial institution that accepts deposits] from the public and creates credit.

Distributional Hypothesis

Word2vec
- Distributional Hypothesis
- Skip-gram/CBOW

LINE
- 1st/2nd proximity
- Edge sampling

DeepWalk
- Skip-gram
- Random walk

HPE
- LINE + DeepWalk

node2vec
- Alias method (Biased RW)
- Homophily hypothesis (BFS)
- Structural equivalence (DFS)

Related Works

A [bank is a financial institution that accepts deposits from] the public and creates credit.

Distributional Hypothesis

Word2vec
- Distributional Hypothesis
- Skip-gram/CBOW

LINE
- 1st/2nd proximity
- Edge sampling

DeepWalk
- Skip-gram
- Random walk

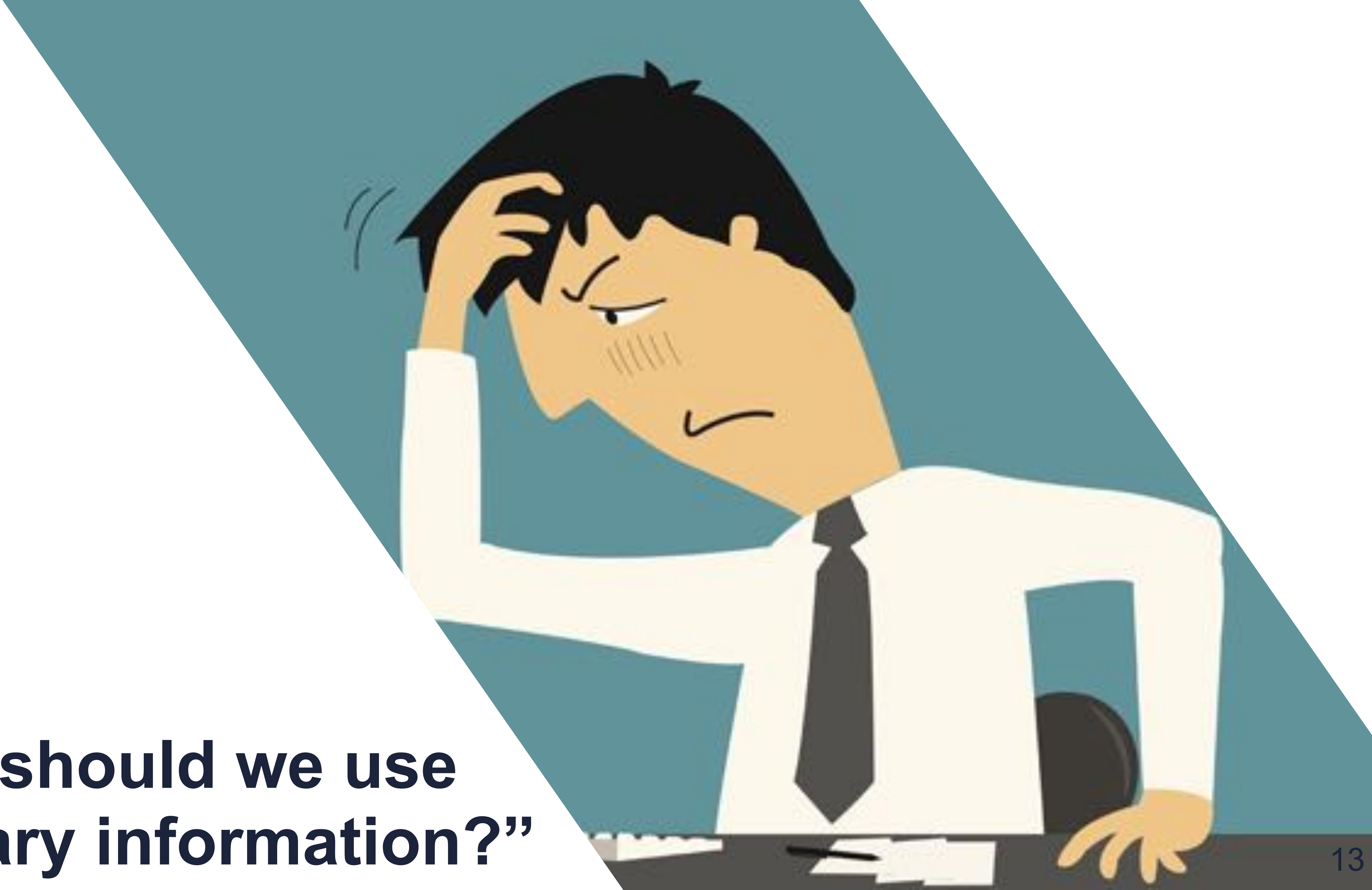
HPE
- LINE + DeepWalk

PTE
- Text -Text
- Text - Doc
- Text - Label

ICE
- Entity - Text Net.
- Text - Text Net.
- Concept Expansion

TADW
- Text Feature Matrix

node2vec
- Alias method (Biased RW)
- Homophily hypothesis (BFS)
- Structural equivalence (DFS)

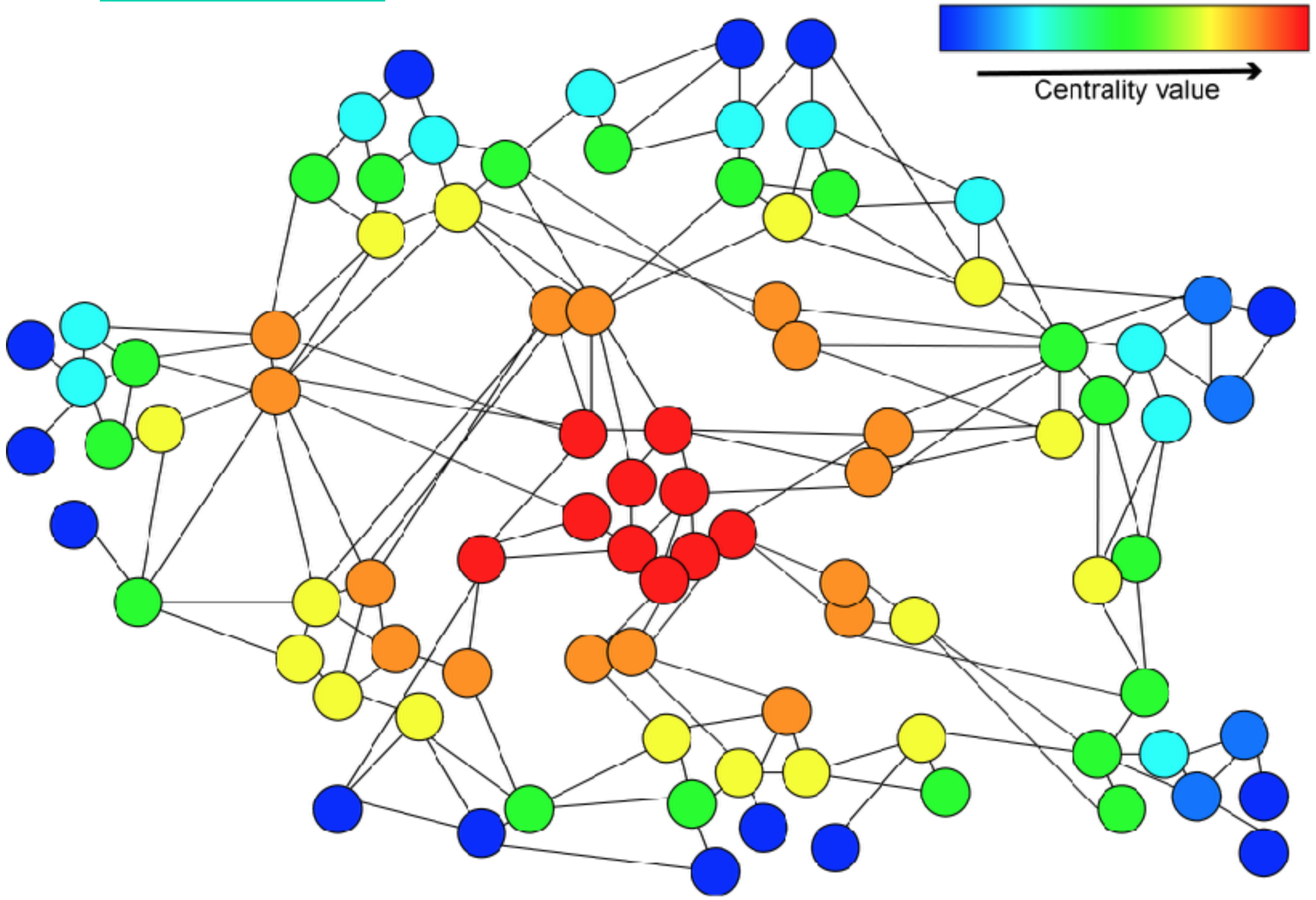


“Why should we use auxiliary information?”

NavWalker

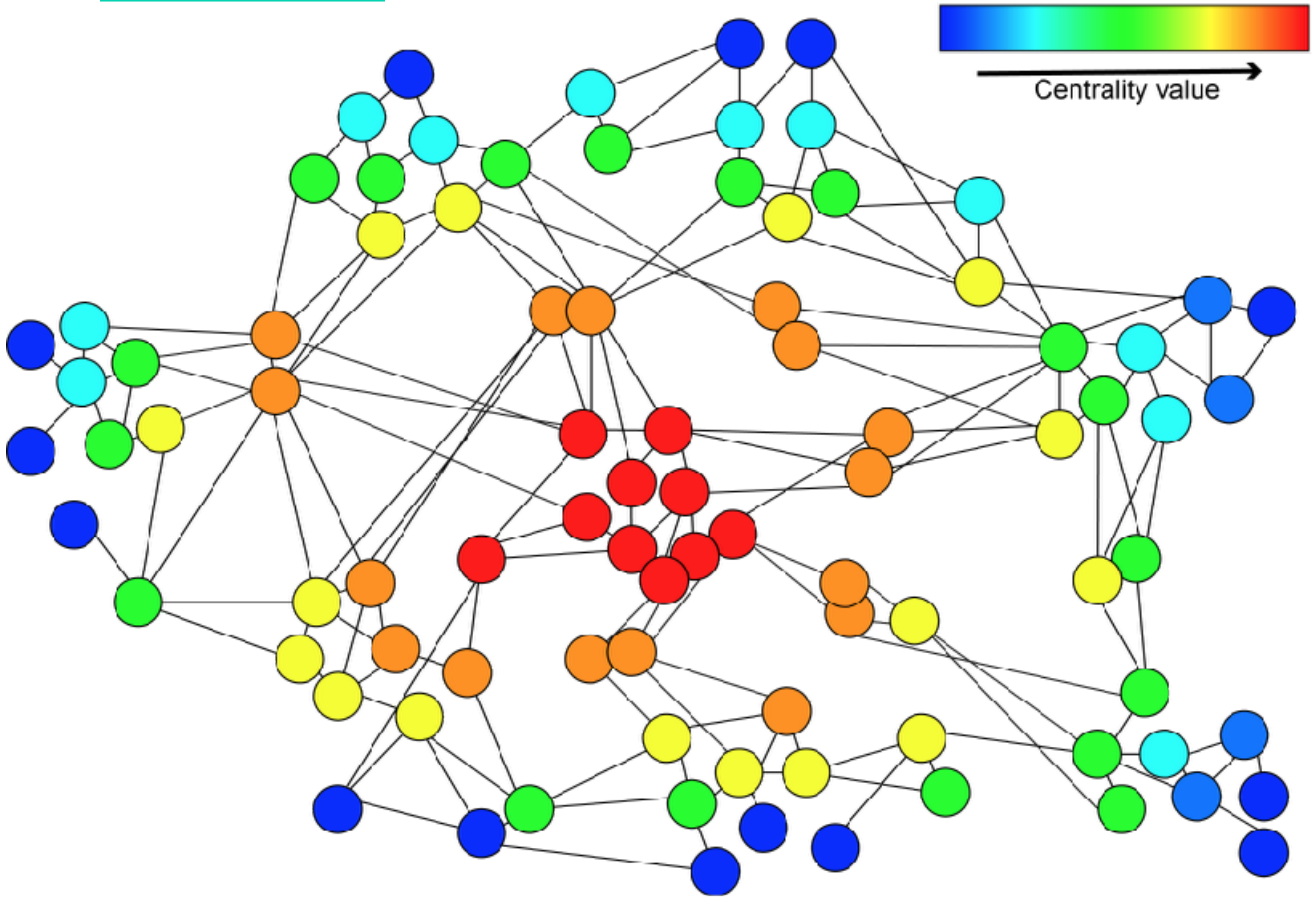
Information Augmentation

Information Augmentation - Closeness Centrality



$$C(i) = \frac{n-1}{\sum_{v_i, v_j \in V} dist\,(i,j)}$$

Information Augmentation - Closeness Centrality



$$C(i) = \frac{n-1}{\sum_{v_i, v_j \in V} dist(i, j)}$$

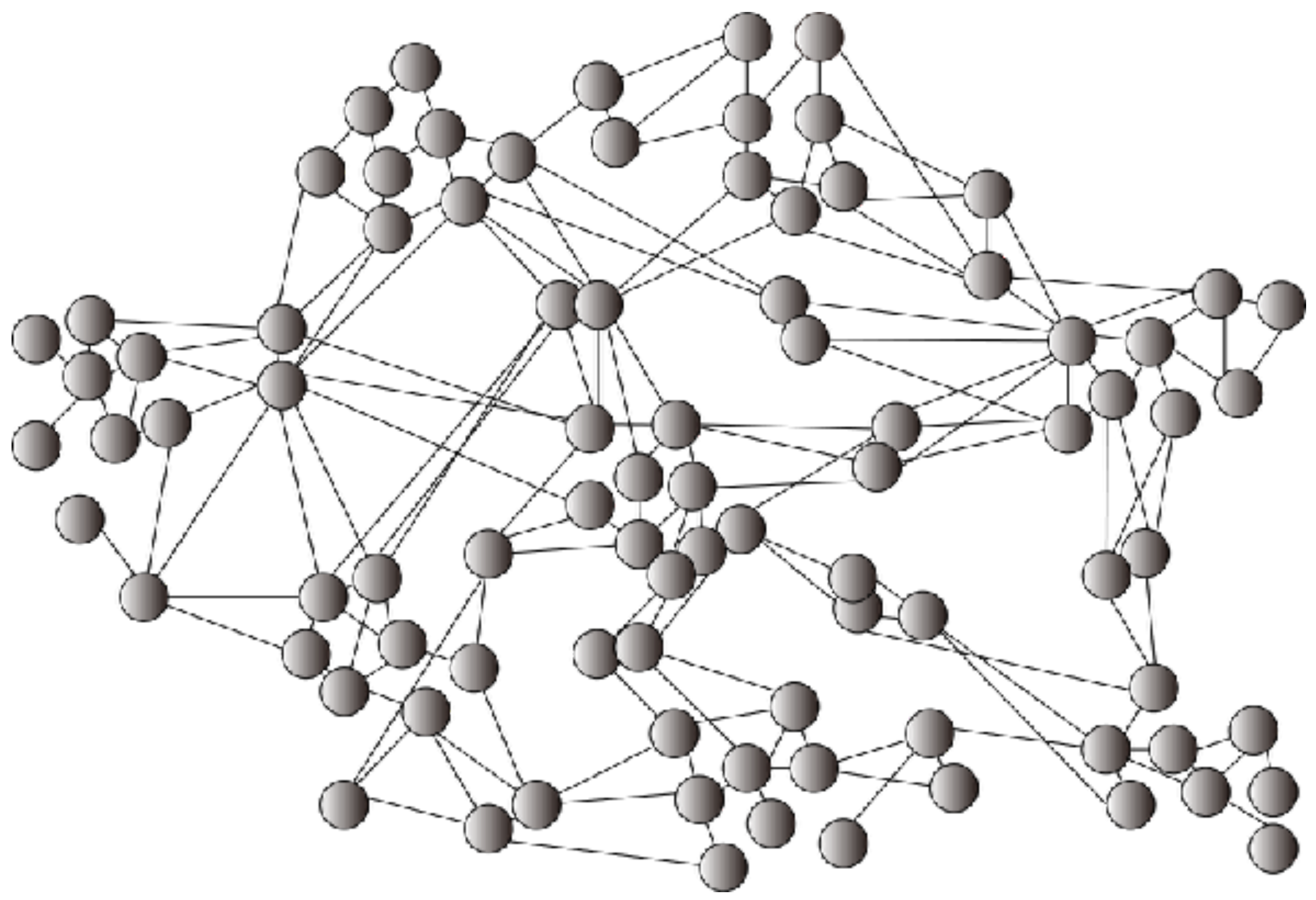
$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{bmatrix}$$

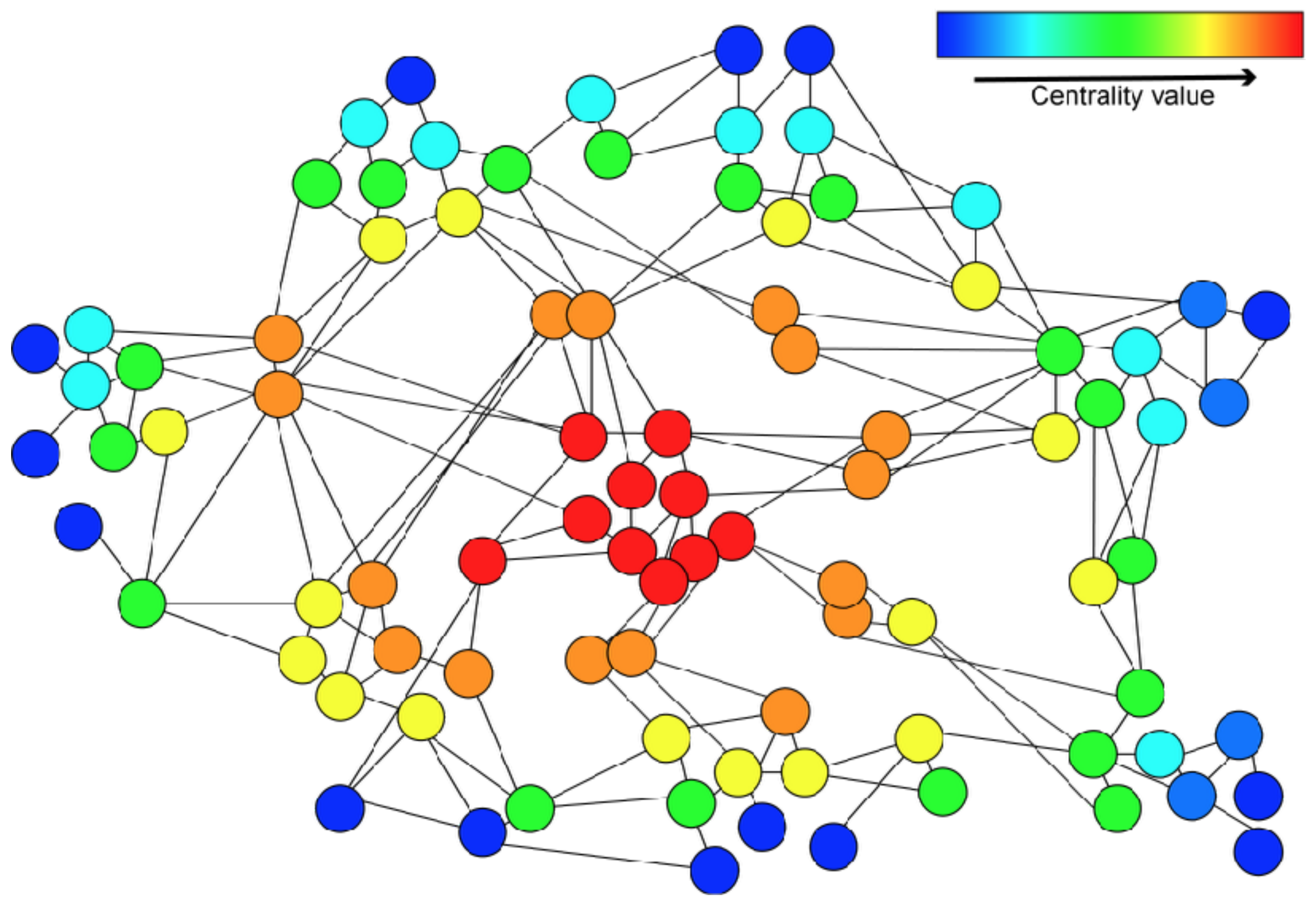
1. $t_{ij} = C(j)$
2. $t_{ij} = C^{-1}(j)$

Information Augmentation

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad , \quad n = |V|$$

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{bmatrix} \quad , \quad n = |V|$$

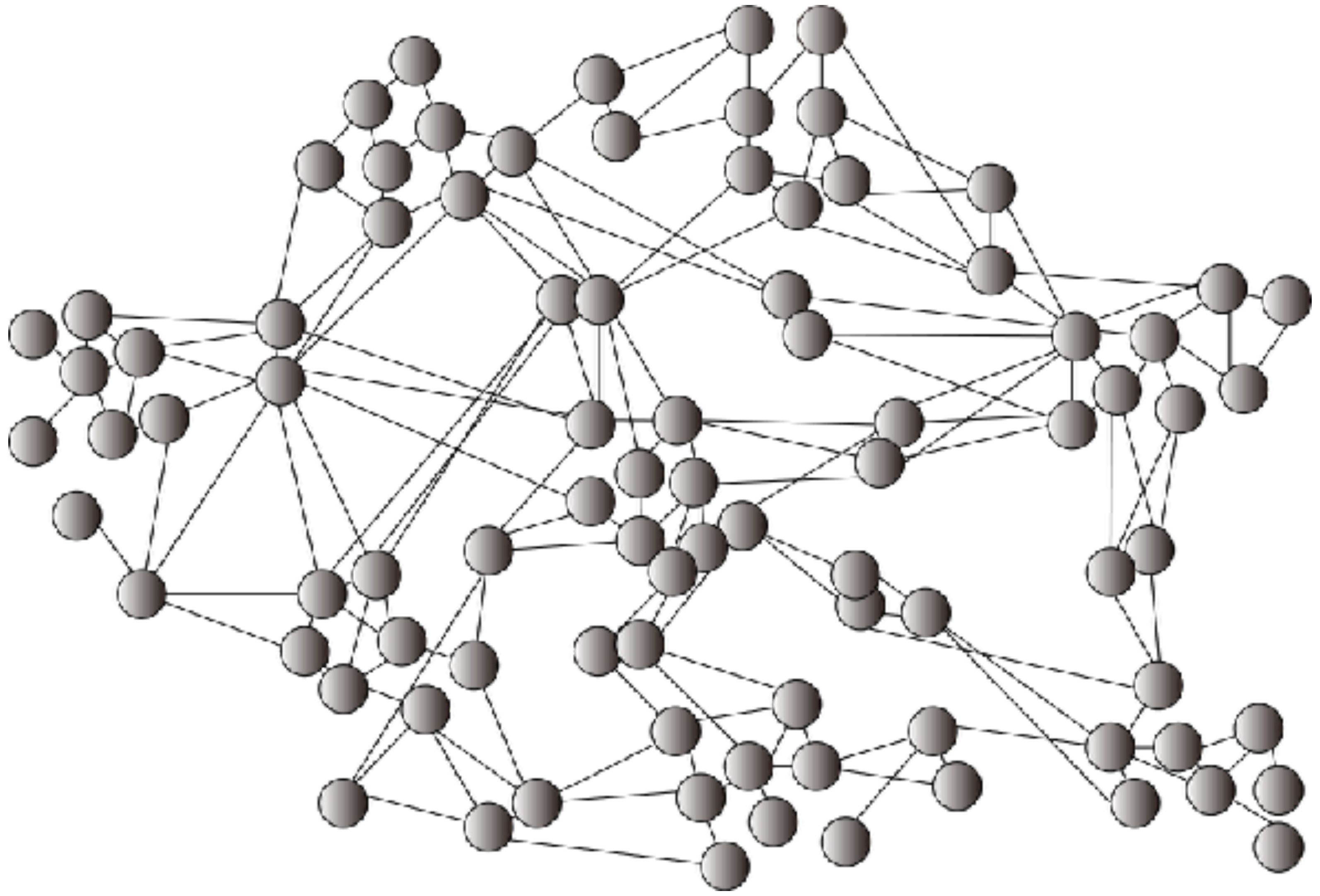




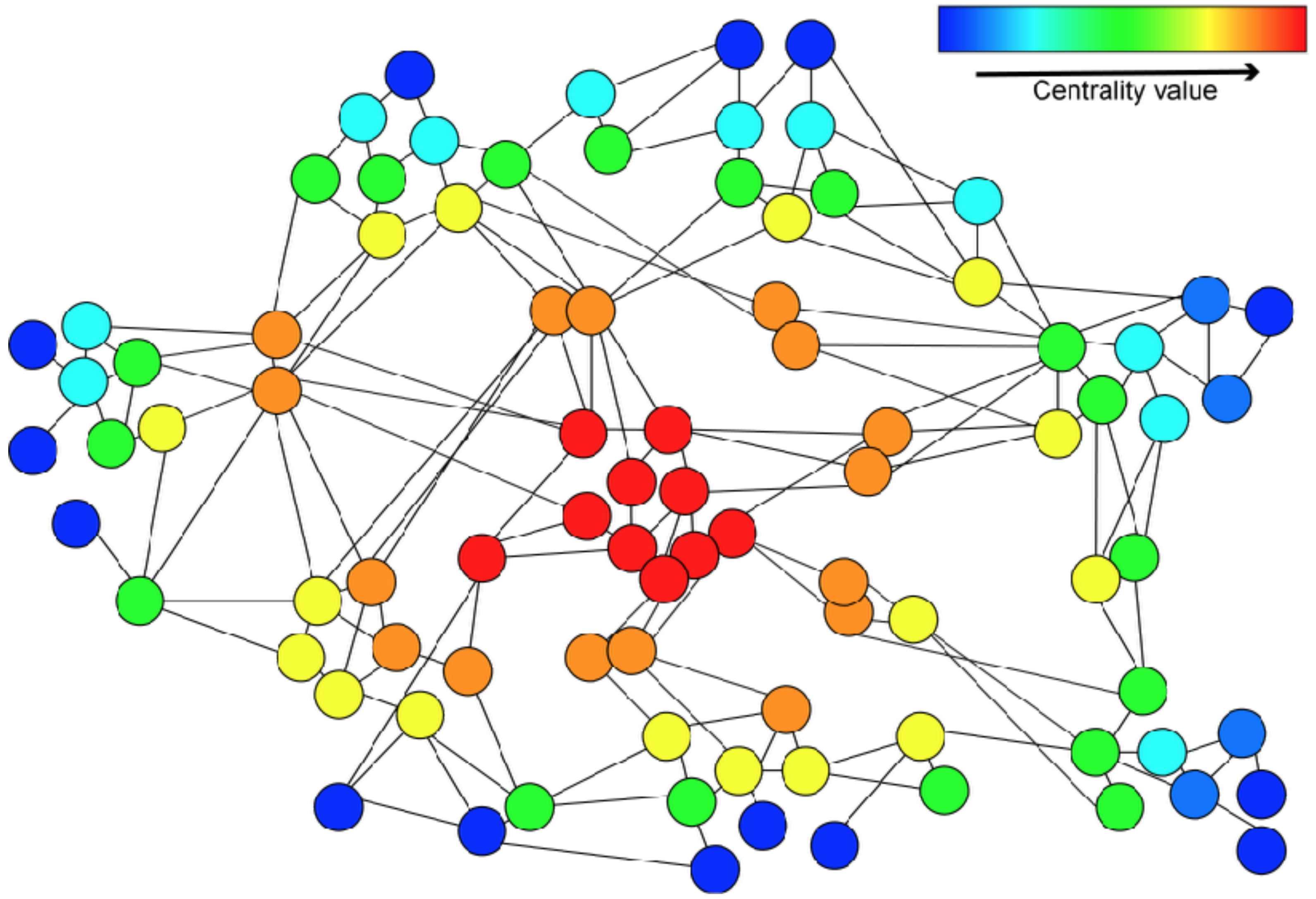
Information Augmentation

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad , \quad n = |V|$$

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{bmatrix} \quad , \quad n = |V|$$



$$A' = A \circ T$$

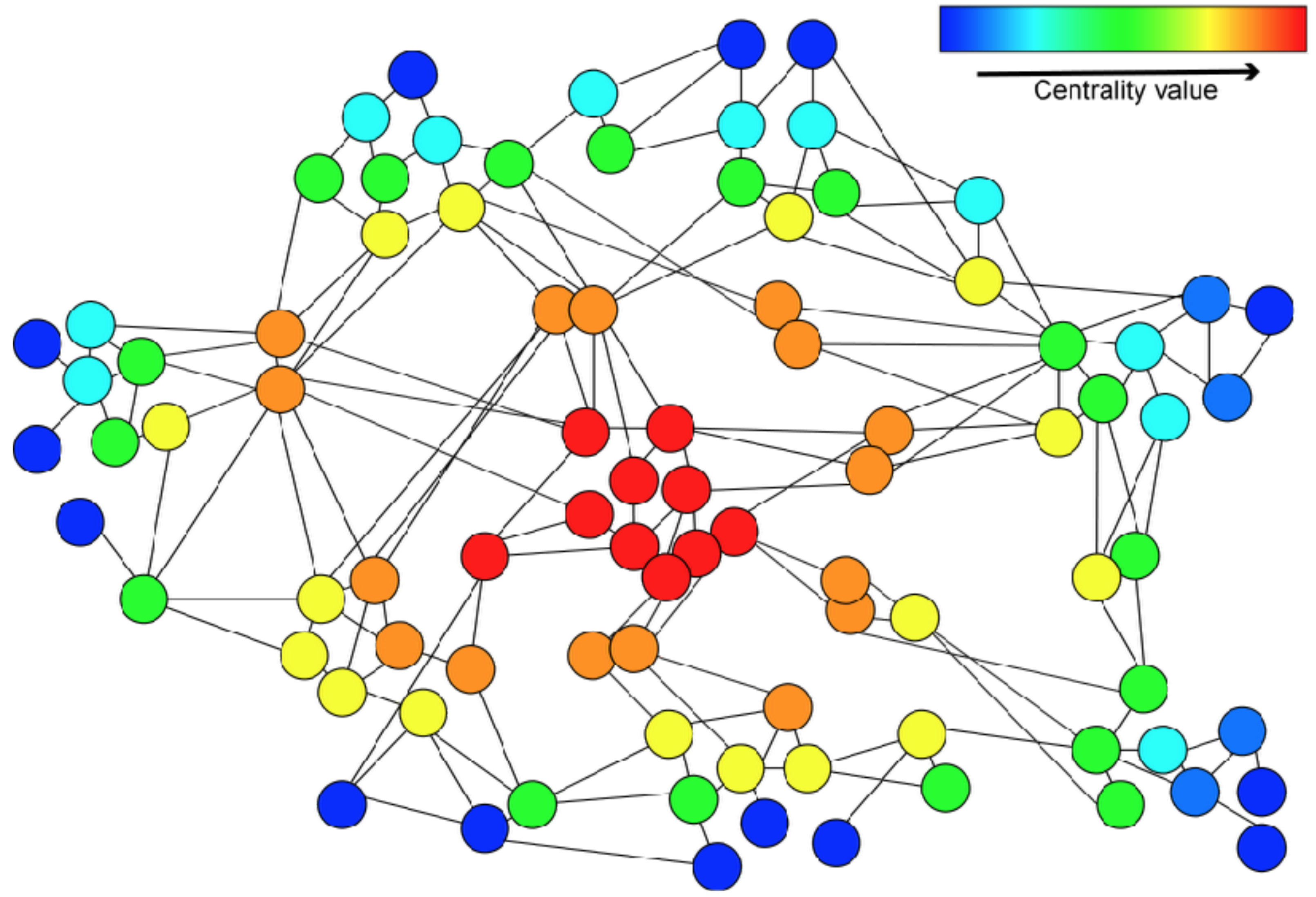


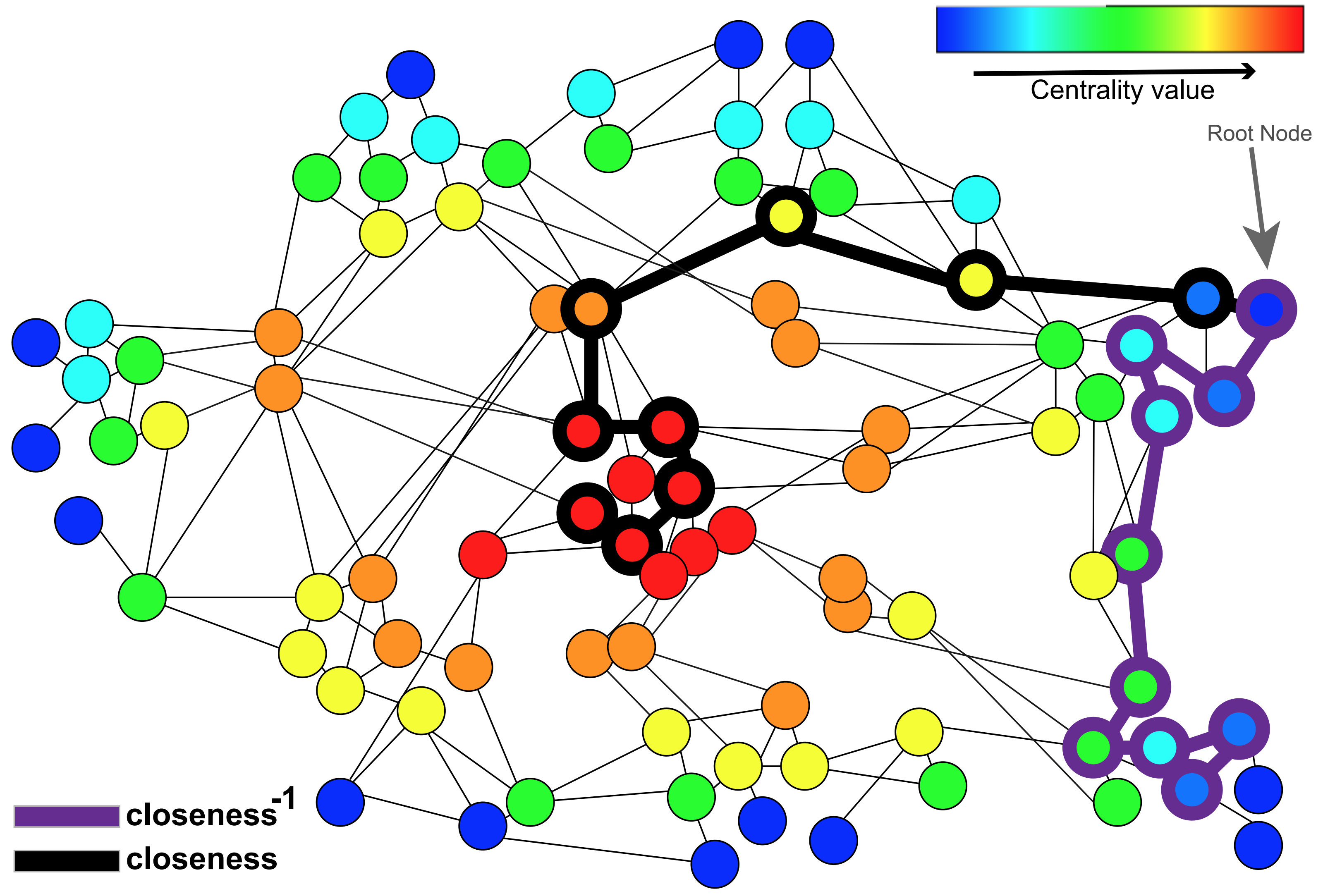
Information Augmentation

$$A' = \begin{bmatrix} a'_{11} & a'_{12} & \cdots & a'_{1n} \\ a'_{21} & a'_{22} & \cdots & a'_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a'_{n1} & a'_{n2} & \cdots & a'_{nn} \end{bmatrix} \quad , \quad n = |V|$$

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & pt_{n2} & \cdots & p_{nn} \end{bmatrix} \quad , \quad n = |V|$$

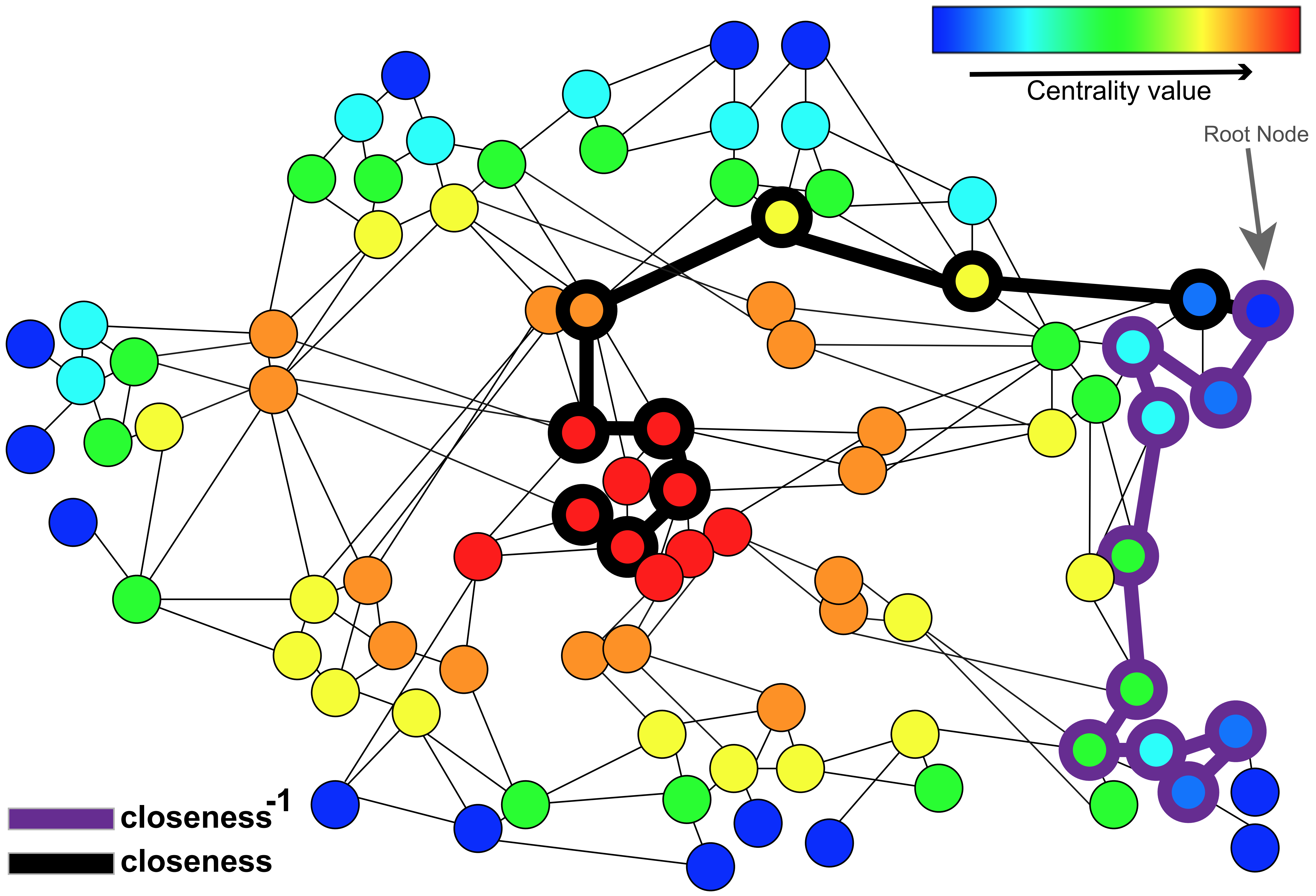
$$\sum_{j=1}^{n} p_{ij} = 1 \quad \text{for } i = 1, \cdots, n$$





Information Augmentation

$$O = -\sum_{v_i \in V} \log P(N(v_i))|\Phi(v_i))$$



“How to use auxiliary information PROPERLY?”

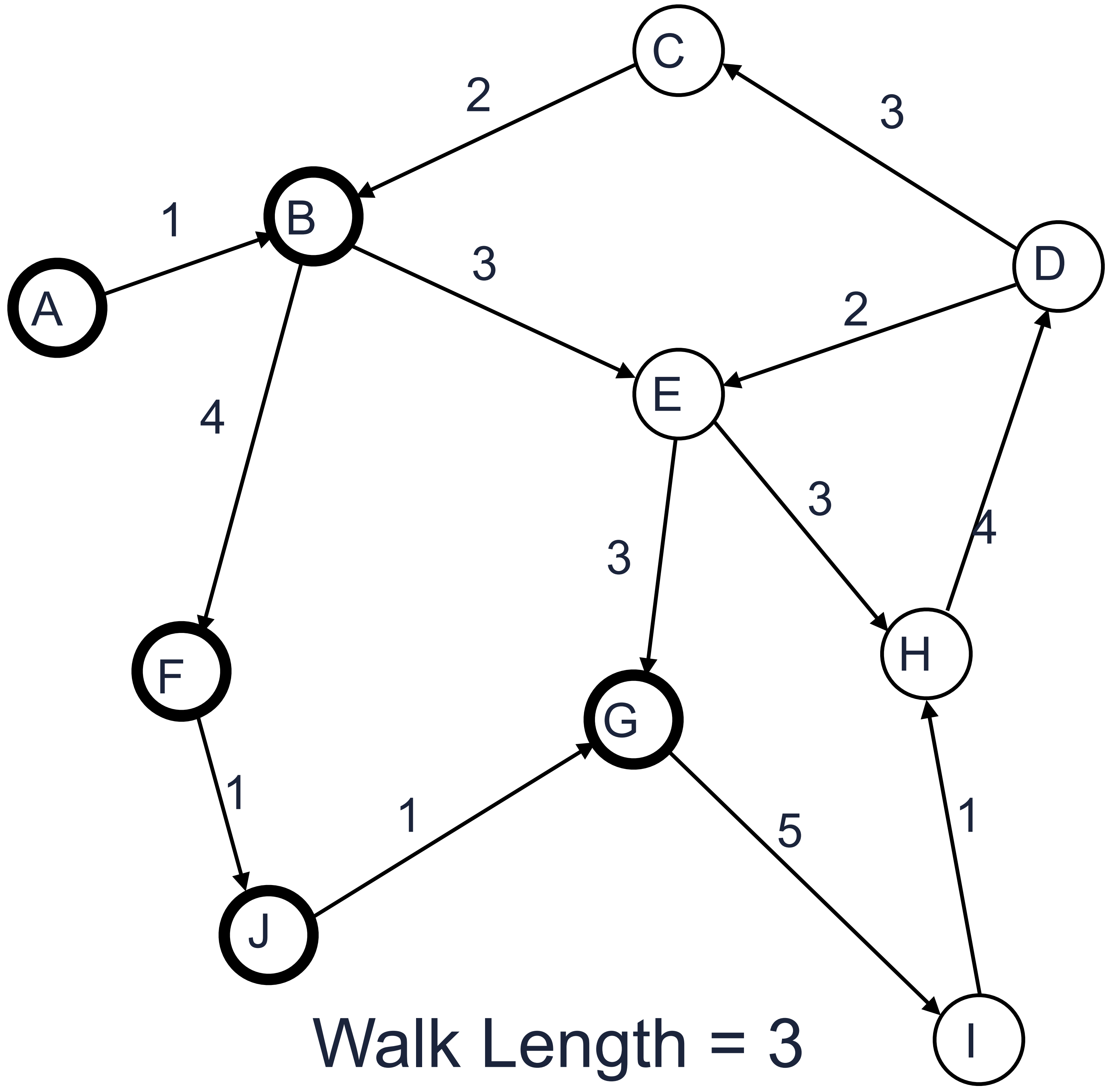
NavWalker

Rooted Random Walk Sampling

Random Walk Sampling

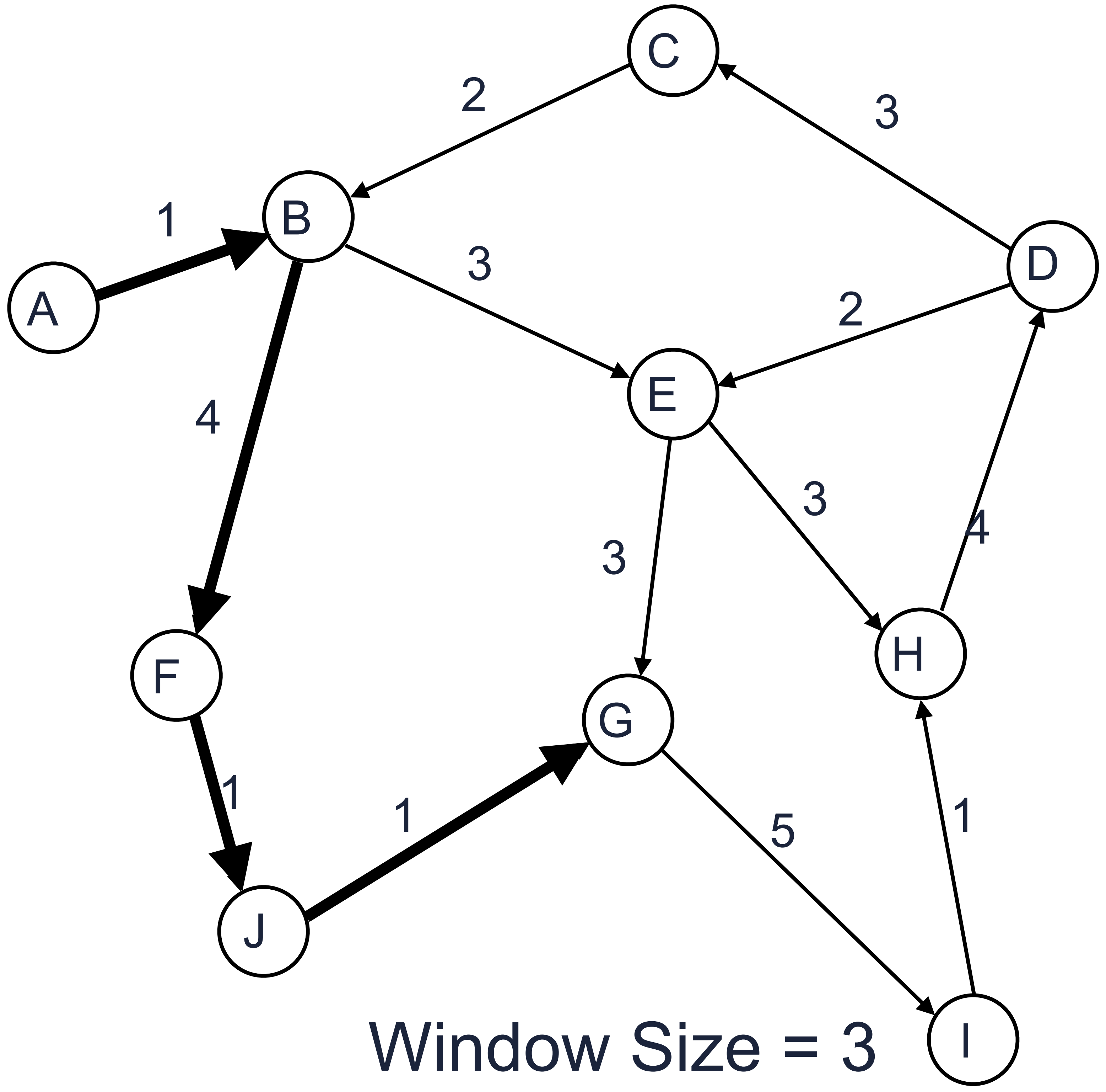
From node A to node G
Neighborhood order: 3

NavWalker



Walk Length = 3
Number of nodes = 5

SkipGram



Window Size = 3
Number of nodes = 5

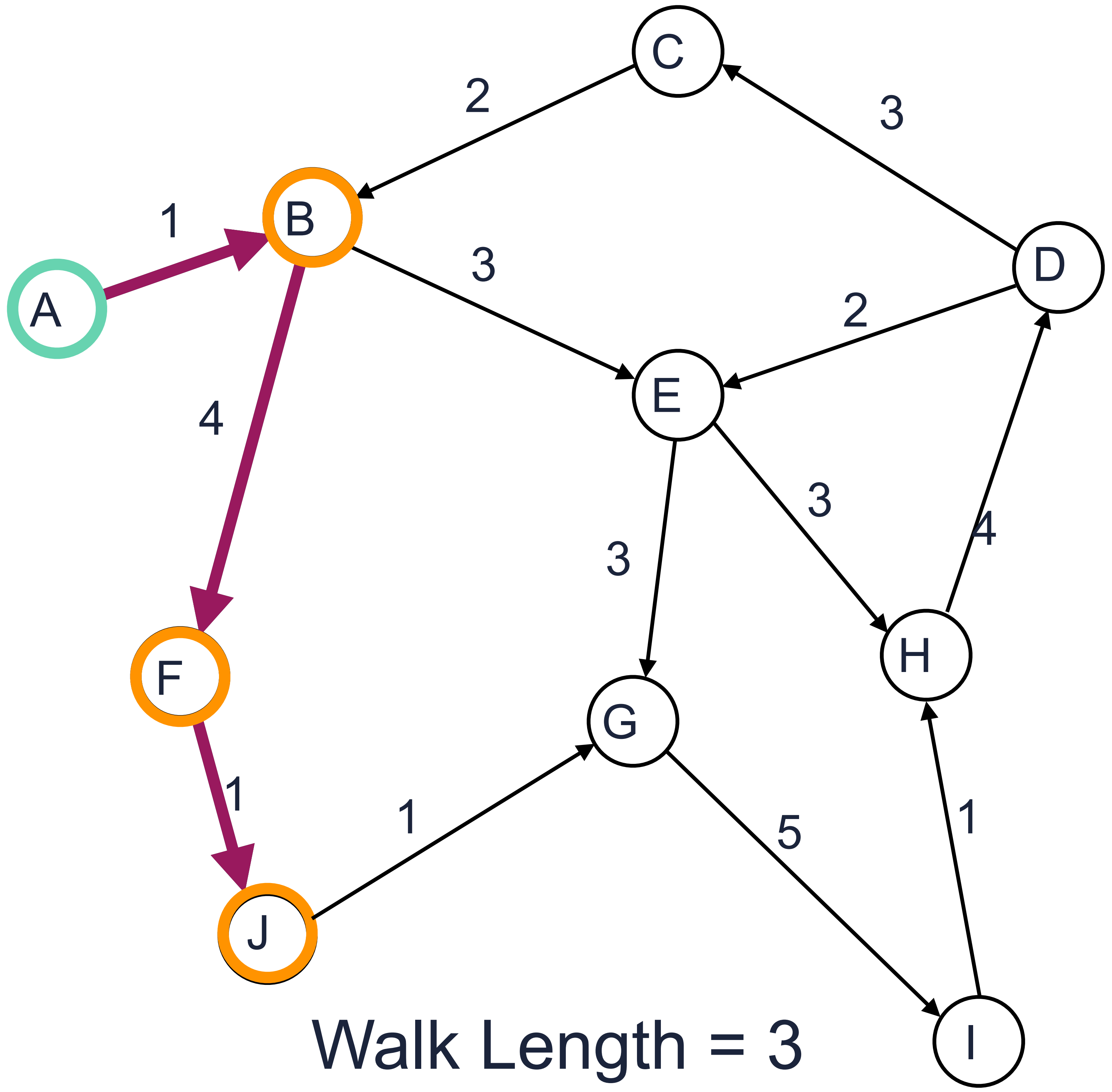
Random Walk Sampling

From node A to node G
Neighborhood order: 3

Root/Center Node
Neighbor Node

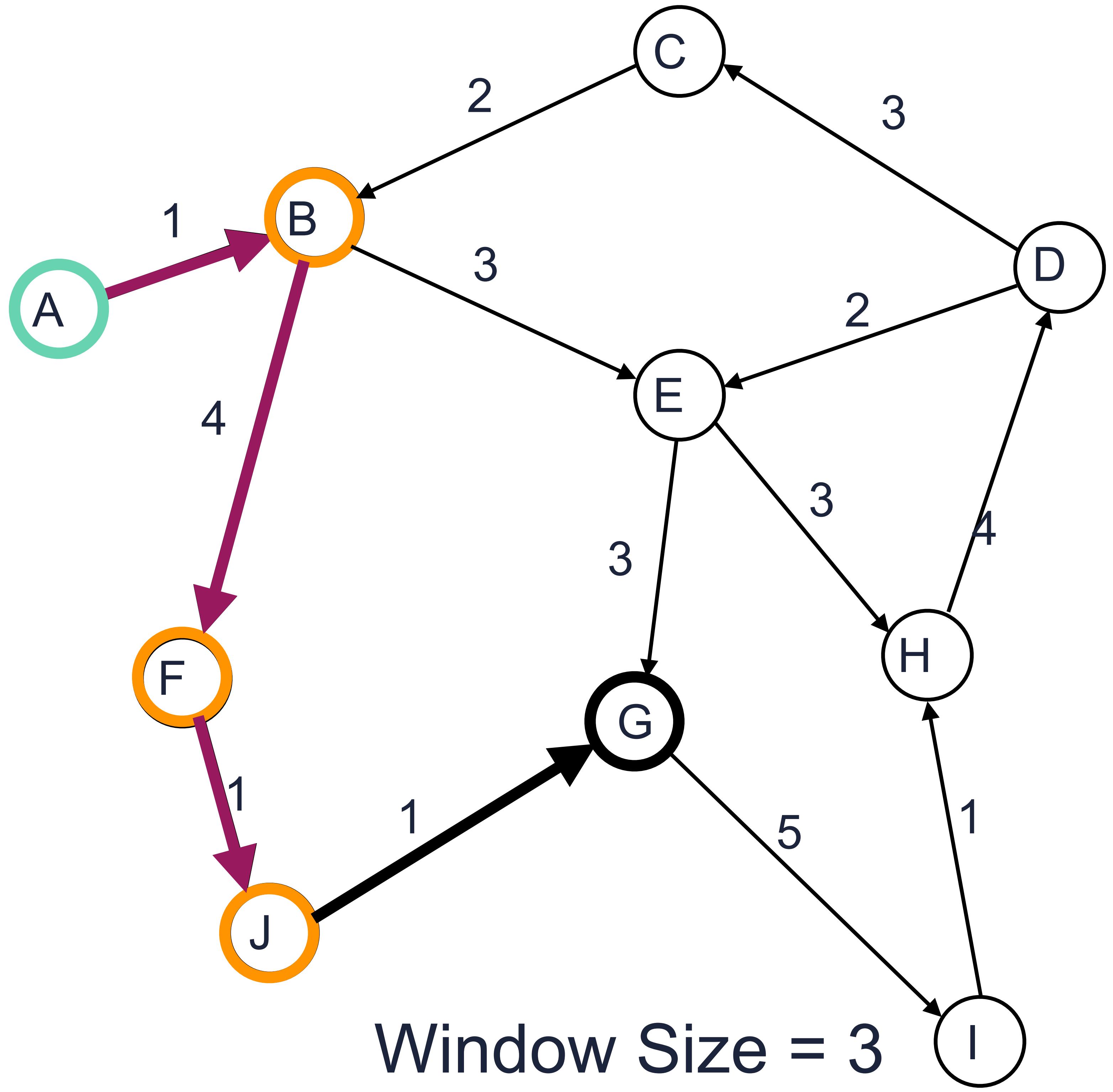
NavWalker

Path:
1. A-B-F-J



Walk Length = 3
Number of nodes = 5

SkipGram



Path:
1. A-B-F-J-G

Context Window:
- A: [A,B,F,J]

Window Size = 3
Number of nodes = 5

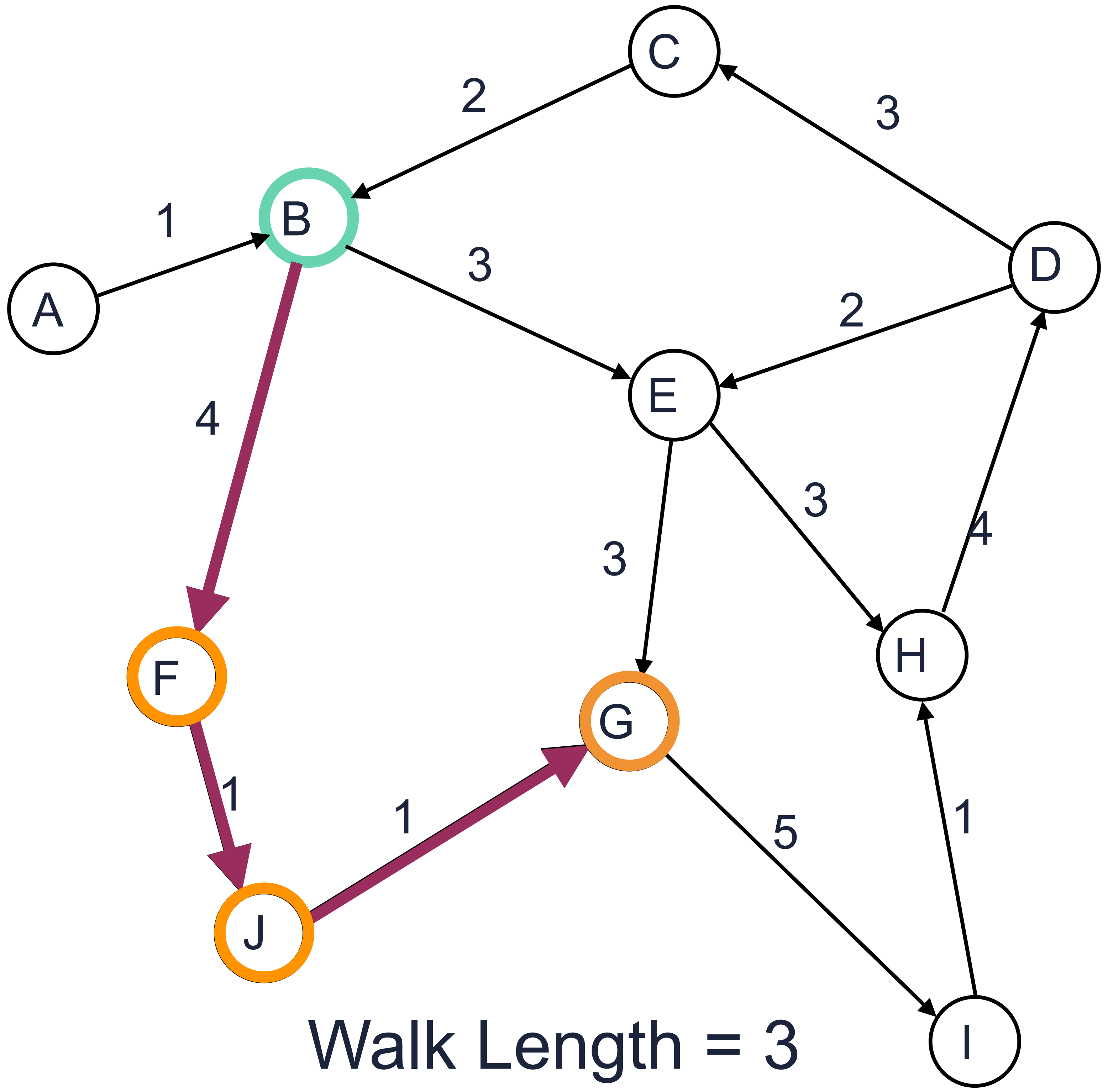
Random Walk Sampling

From node A to node G
Neighborhood order: 3

Root/Center Node
Neighbor Node

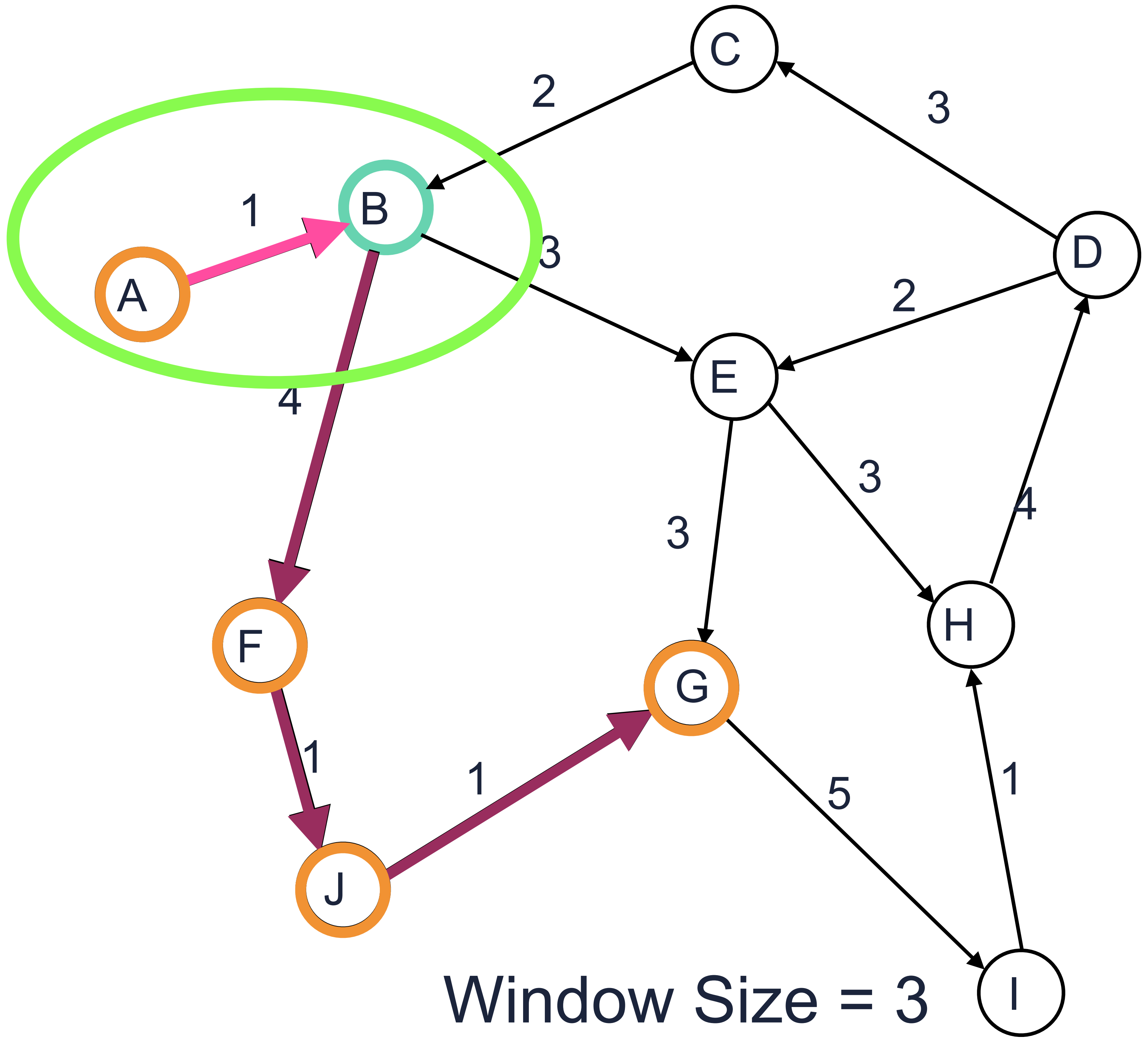
Path:
1. A-B-F-J
2. B-F-J-G

NavWalker



Walk Length = 3
Number of nodes = 5

SkipGram



Window Size = 3
Number of nodes = 5

Path:
1. A-B-F-J-G

Context Window:
- A: [A,B,F,J]
- B: [A,B,F,J,G]

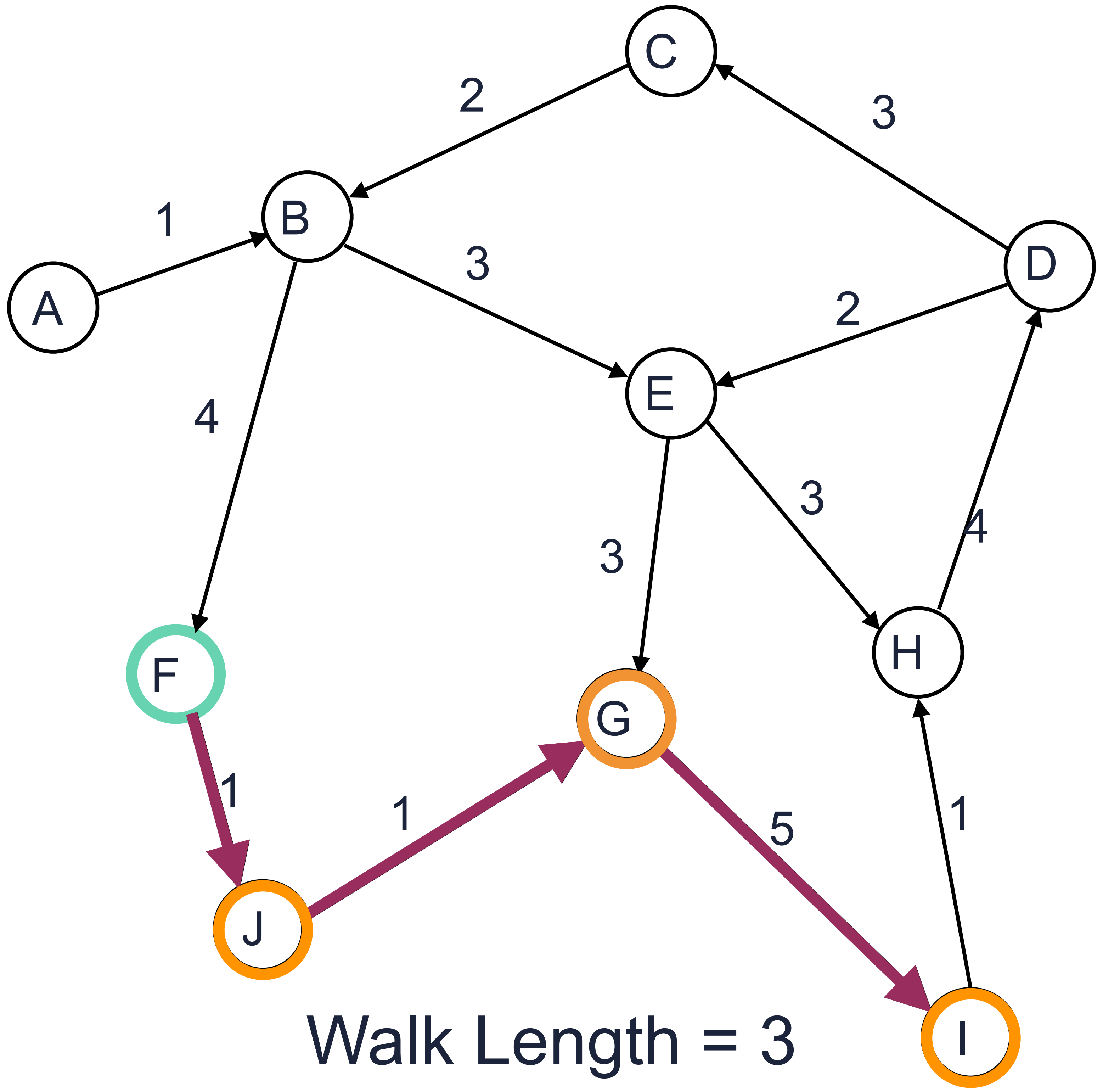
Random Walk Sampling

From node A to node G
Neighborhood order: 3

Root/Center Node
Neighbor Node

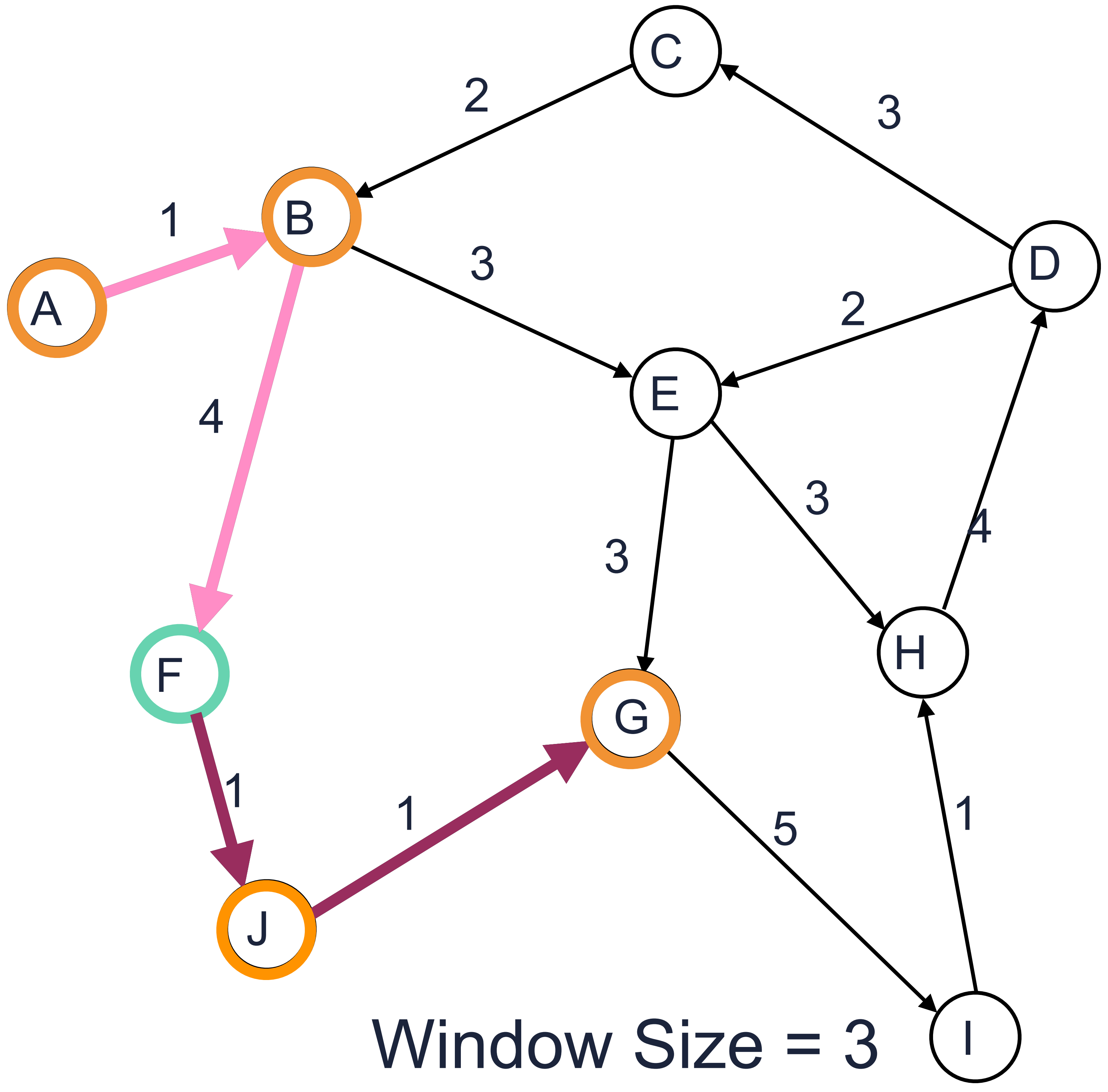
Path:
1. A-B-F-J
2. B-F-J-G
3. F-J-G-I

NavWalker



Walk Length = 3
Number of nodes = 5

SkipGram



Window Size = 3
Number of nodes = 5

Path:
1. A-B-F-J-G

Context Window:
- A: [A,B,F,J]
- B: [A,B,F,J,G]
- F: [A,B,F,J,G]

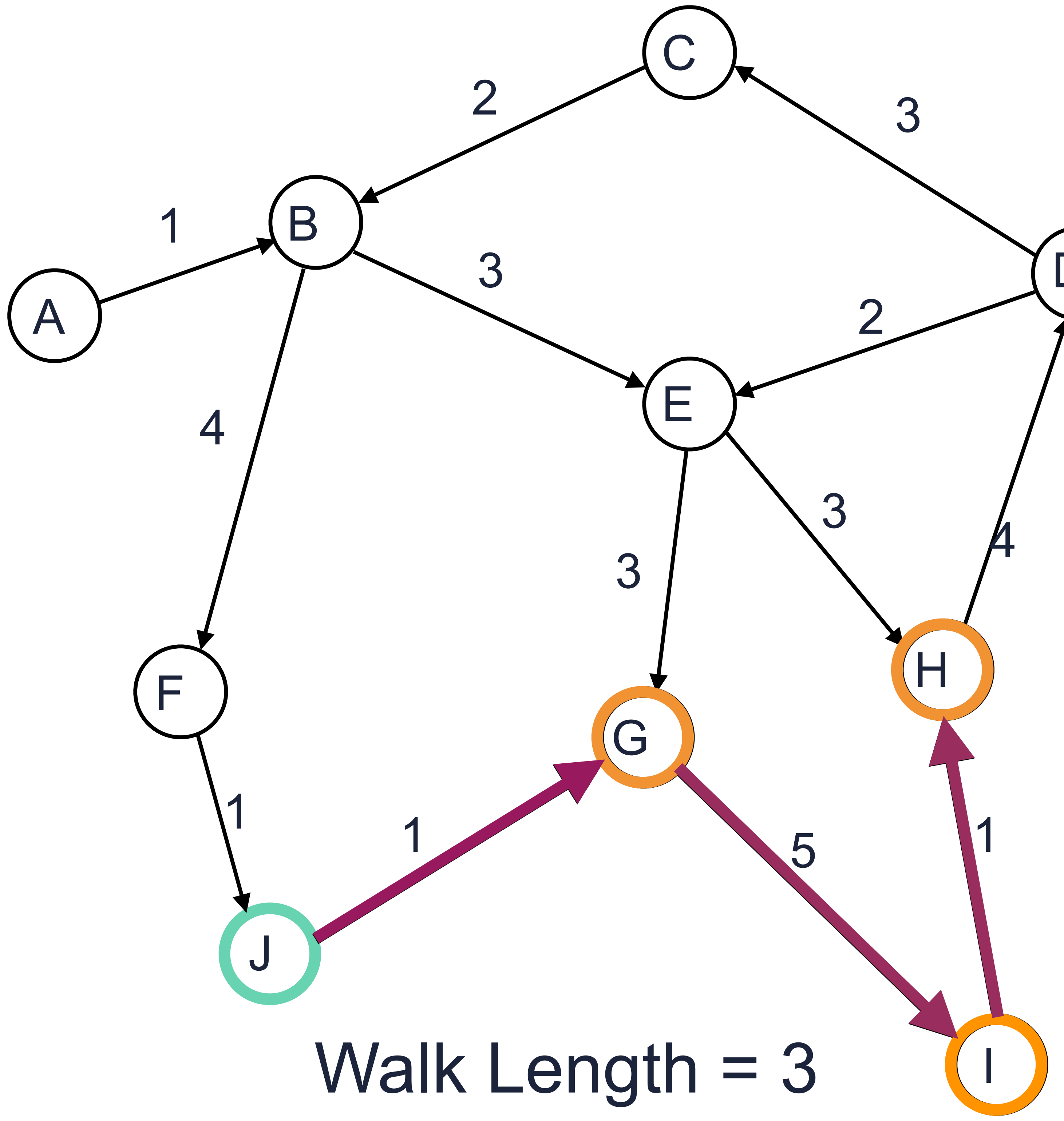
Random Walk Sampling

From node A to node G
Neighborhood order: 3

Root/Center Node
Neighbor Node

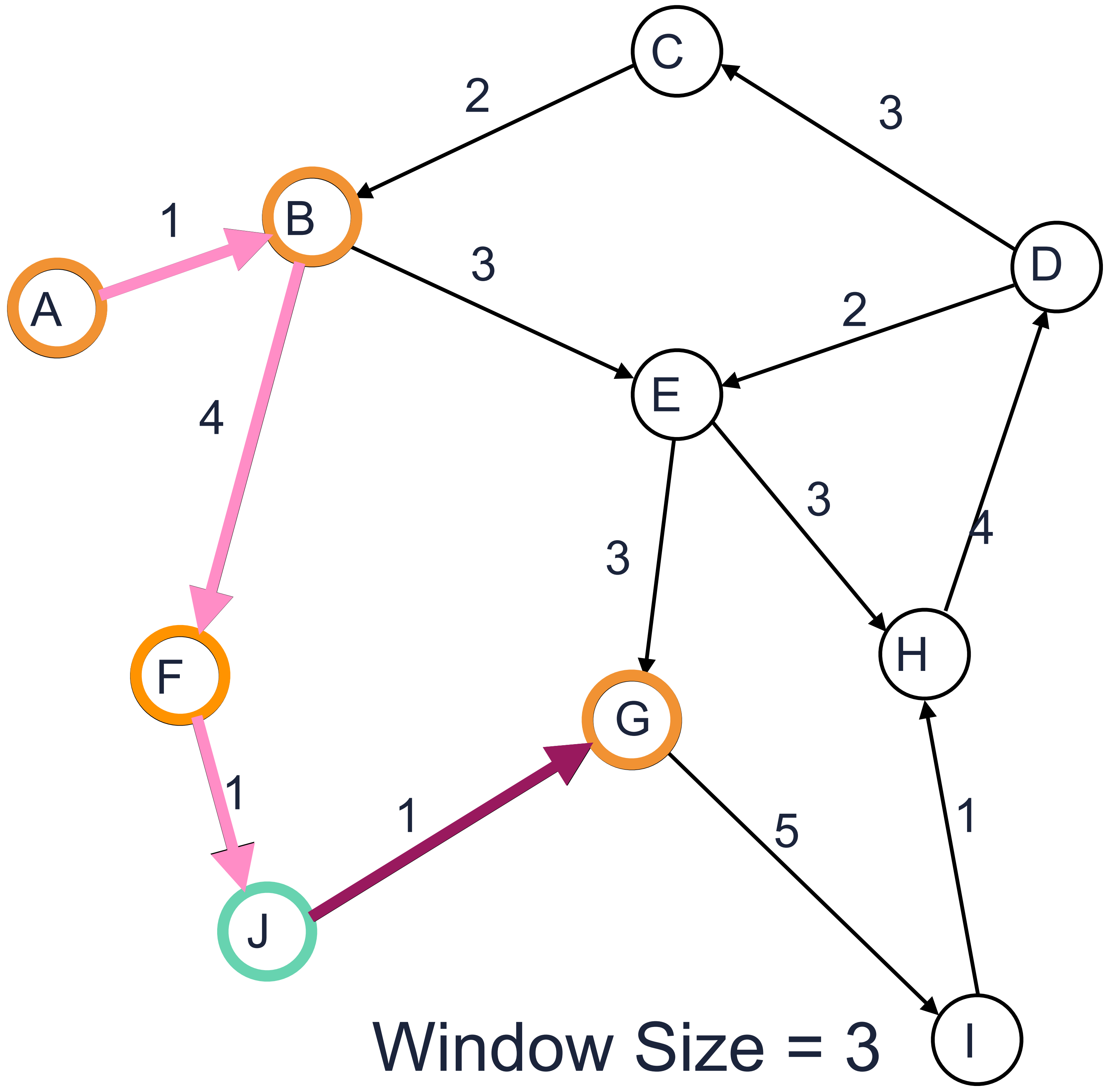
Path:
1. A-B-F-J
2. B-F-J-G
3. F-J-G-I
4. J-G-I-H

NavWalker



Walk Length = 3
Number of nodes = 5

SkipGram



Window Size = 3
Number of nodes = 5

Path:
1. A-B-F-J-G

Context Window:
- A: [A,B,F,J]
- B: [A,B,F,J,G]
- F: [A,B,F,J,G]
- J: [A,B,F,J,G]

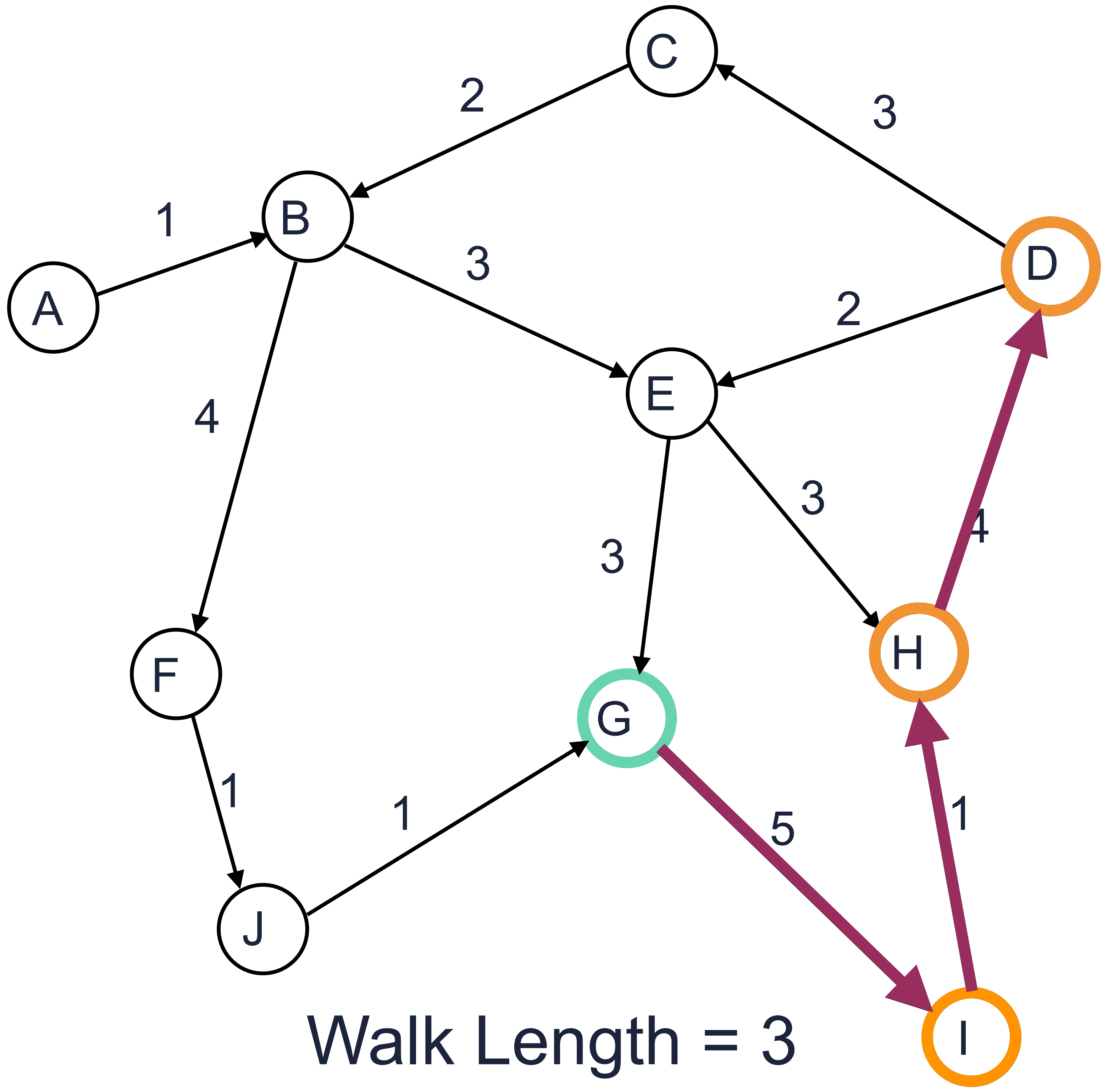
Random Walk Sampling

From node A to node G
Neighborhood order: 3

Root/Center Node
Neighbor Node

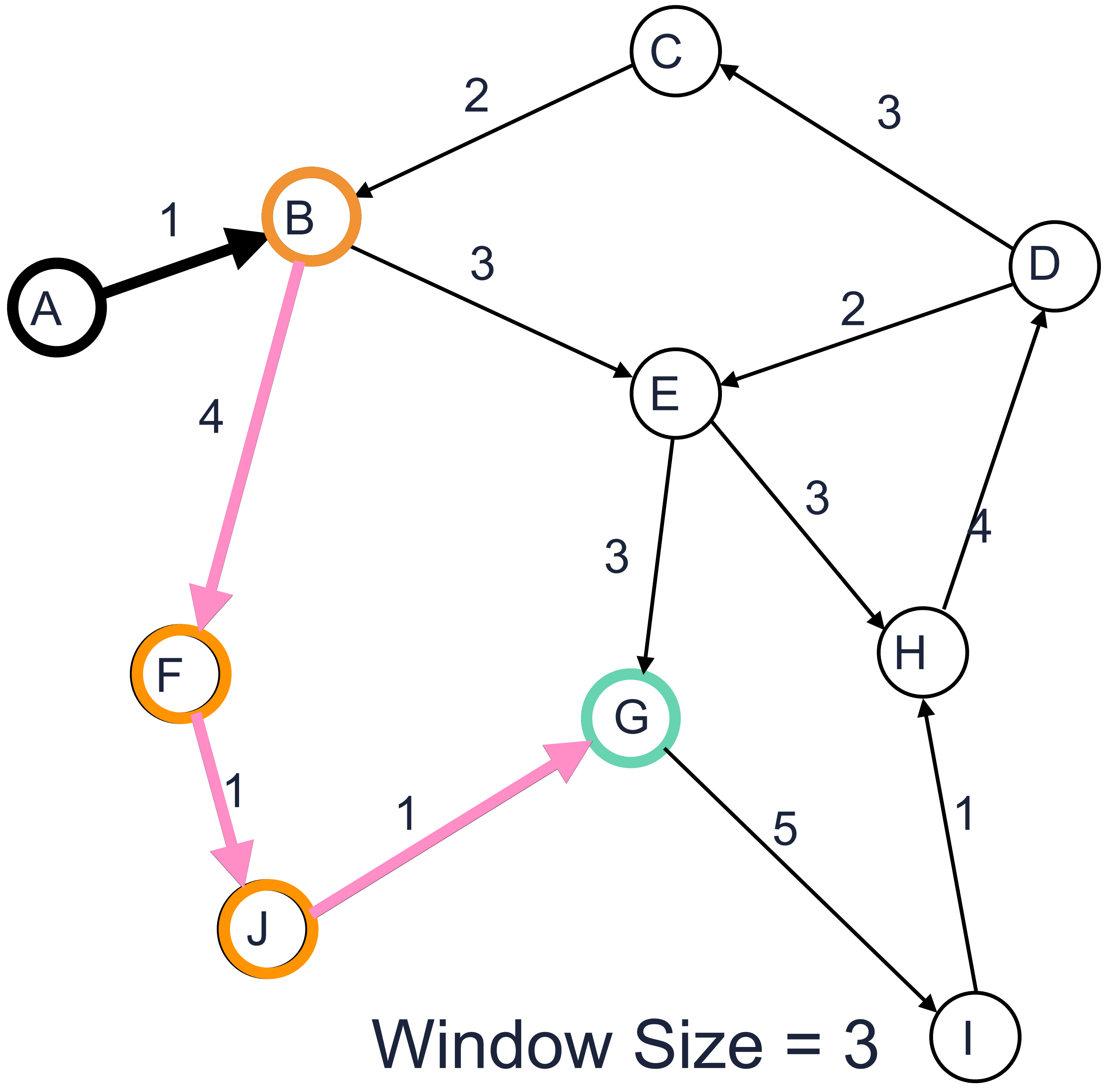
Path:
1. A-B-F-J
2. B-F-J-G
3. F-J-G-I
4. J-G-I-H
5. G-I-H-D

NavWalker



Walk Length = 3
Number of nodes = 5

SkipGram

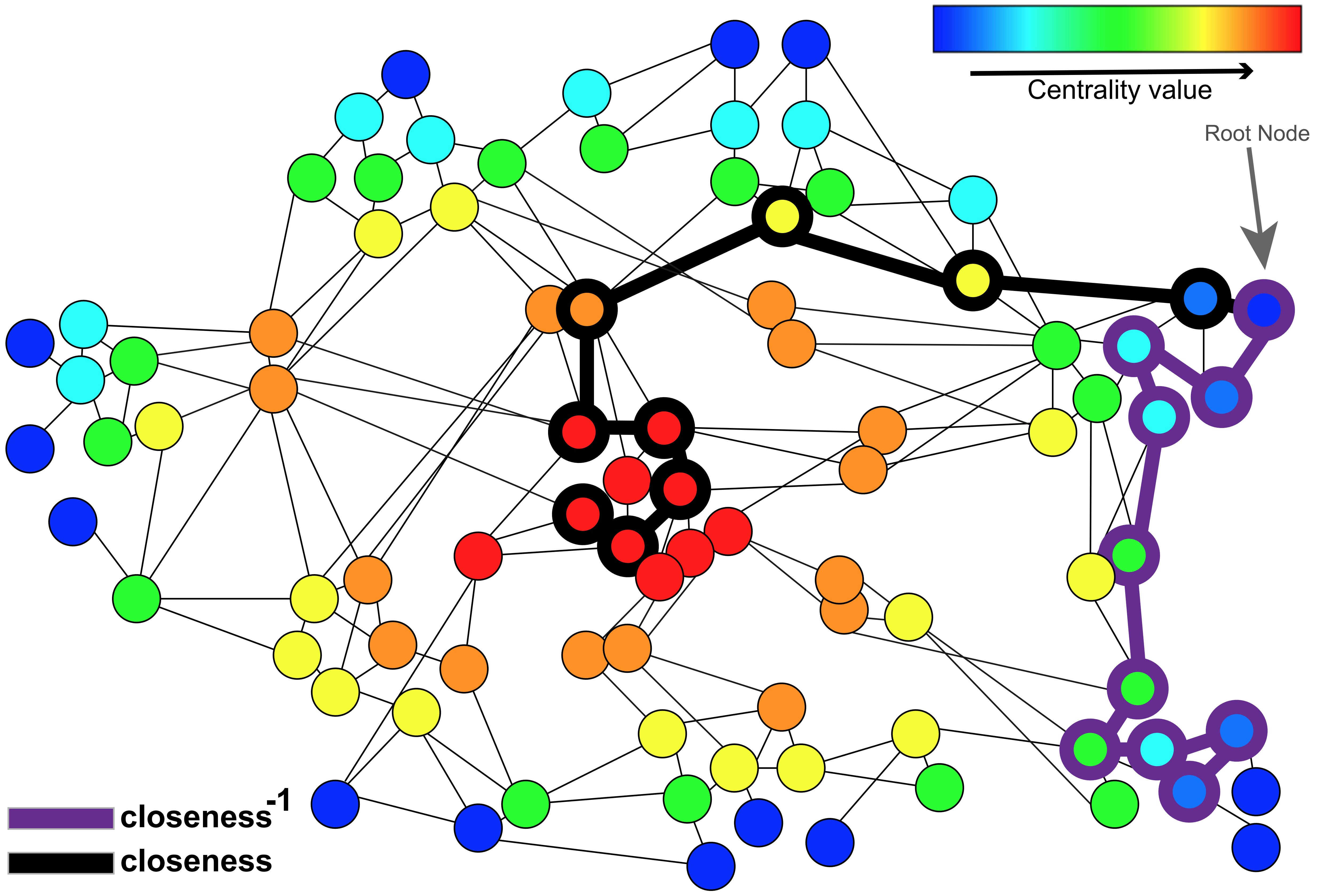


Window Size = 3
Number of nodes = 5

Path:
1. A-B-F-J-G

Context Window:
- A: [A,B,F,J]
- B: [A,B,F,J,G]
- F: [A,B,F,J,G]
- J: [A,B,F,J,G]
- G: [B,F,J,G]

Information Augmentation



$$p(v_j|v_i) = (p_j) \in \left[\sum_{k=1}^{n}(P)^k\right]_i$$

Experiment

Classification & Recommendation

Experiment — Settings

- Multi-label Classification:
  - Label:
    - Blogger classification
    - Biological state classification
    - POS tag classification
  - Classifier:
    - Logistic regression (L2)

Classification	|V|	|E|	#(labels)
BlogCatalog	10,312	333,983	39
PPI	3,890	76,584	50
Wikipedia	4,777	184,812	40

Experiment — Settings

- Multi-label Classification:
  - Label:
    - Blogger classification
    - Biological state classification
    - POS tag classification
  - Classifier:
    - Logistic regression (L2)
- Query-Based (Item) Recommendation:
  - Recommending items to user based on user's accessed items
  - Train:Test = 8:2

Classification	|V|	|E|	#(labels)
BlogCatalog	10,312	333,983	39
PPI	3,890	76,584	50
Wikipedia	4,777	184,812	40

Recommendation	#(users)	#(items)	#(edges)
Movielens-1M	6,070	3,706	1,000,209
Movielens-10M	69,878	10,677	10,000,054
Last.fm-2k	1,892	92,800	92,834

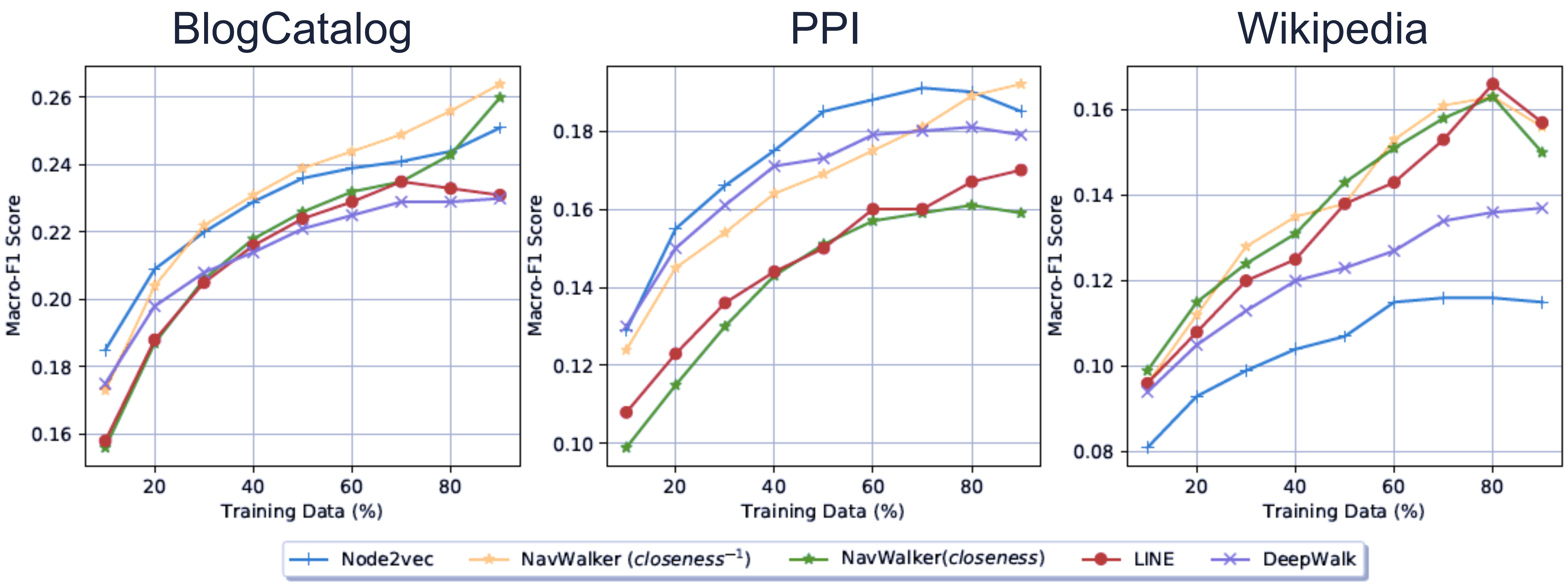
Experiment — Settings

- Multi-label Classification:
  - Label:
    - Blogger classification
    - Biological state classification
    - POS tag classification
  - Classifier:
    - Logistic regression (L2)
- Query-Based (Item) Recommendation:
  - Recommending items to user based on user's accessed items
  - Train:Test = 8:2
- Baseline:
  - SkipGram Based Method
    - DeepWalk
    - node2vec
  - Edge Sampling Based Method
    - LINE (2nd)
    - HPE

Classification	|V|	|E|	#(labels)
BlogCatalog	10,312	333,983	39
PPI	3,890	76,584	50
Wikipedia	4,777	184,812	40

Recommendation	#(users)	#(items)	#(edges)
Movielens-1M	6,070	3,706	1,000,209
Movielens-10M	69,878	10,677	10,000,054
Last.fm-2k	1,892	92,800	92,834

Experiment — Multi-Label Classification



	BlogCatalog	PPI (Homo Sapiens)	Wikipedia (POS)
DeepWalk	0.230	0.179	0.137
Node2vec (p, q)	0.251 (0.25, 0.25)	0.185 (4, 1)	0.115 (4, 0.5)
LINE	0.231	0.170	**0.157**
NavWalker-clossness^{-1} (Improvement)	**0.265 (5.5%)**	**0.192 (3.7%)**	0.156

Experiment — Query-based Recommendation

MovieLens1M

Algorithm	MAP@10	Recall@10
LINE	0.303*	0.377*
DeepWalk	0.270*	0.333*
Node2vec ($p = 1$, $q = 0.5$)	0.269*	0.331*
HPE	0.218*	0.283*
NavWalker (closeness)	0.333	0.411*
NavWalker (closeness^{-1})	**0.338**	**0.417**

MovieLens10M

Algorithm	MAP@10	Recall@10
LINE	0.308*	0.375*
DeepWalk	0.275*	0.334*
Node2vec	-	-
HPE	0.164*	0.202*
NavWalker (closeness)	0.281	0.348
NavWalker (closeness^{-1})	**0.315**	**0.391**

Last.fm

Algorithm	MAP@10	Recall@10
LINE	0.270*	0.295*
DeepWalk	0.276*	0.300*
Node2vec ($p = 0.5, q = 4$)	0.283*	0.295*
HPE	0.319*	0.361*
NavWalker (closeness)	0.169	0.190
NavWalker (closeness^{-1})	**0.343**	**0.396**

Take Home Message & Future Work

Take Home Messages

- Rooted based random walk sampling is suitable for learning network representation.
- The flexibility of proposed framework facilitates the learning of information-enhanced representation.
- Centrality information strengthens the quality of learned representations.

Future Works

- Exploring other auxiliary information.
- Investigation for information enhancement methods.

Any Questions?

I am Kwei-Herng Lai

Research Assistant@AS-CITI

Contact: khlai@citi.sinica.edu.tw

Thank You!