Visualizing Graph Neural Networks

with 🐦 **CorGIE** : Corresponding a Graph to Its Embedding

Zipeng Liu
University of British Columbia (UBC)

Yang Wang
Facebook

Jürgen Bernard
University of Zurich

Tamara Munzner
UBC

Paper under review
ChinaVis 2021 at Wuhan
July 2021
Outline

• Introduction of GNN
• Visual evaluation of GNN
  • Previous
  • Ours: CorGIE
    • Overview
    • Data & tasks
    • CorGIE interface
    • Reflections
Graph neural network (GNN)

- Machine learning (ML) models for graph
  - Like CNN for images
  - Like Transformer for text

- Many real-world graph-related applications
  - Node classifications
    - e.g. fraud detection, disease classification
  - Link prediction
    - e.g. recommendation of products, protein interactions
Graph neural network (GNN)

- **Input graph**
- **Graph neural Network (GNN)**
- **Node embedding**
- **High Dim latent space**
- **Downstream ML applications**
- **Predictions (node labels, links)**

**Graph neural network (GNN)**

- **Input graph**
- **Node embedding**
- **High Dim latent space**
- **Downstream ML applications**
- **Predictions (node labels, links)**

**Movie – user graph**

A vector for each node

Node 0: Alice

Node 12: Lord of the Rings

Movie recommendation

Node 0: Alice

Movie recommendation

Node 12: Lord of the Rings

Movie recommendation
GNN: neighborhood aggregation

Node features are aggregated / passed through topological neighborhood

Evaluate GNN

Two big-picture questions:

- “Are we there yet?”: should we train / tune more?
- “Are we lost?”: does it behave as we expect?
Outline

• Introduction of GNN
• Visual evaluation of GNN
  • Previous
  • Ours: CorGIE
    • Overview
    • Data & tasks
    • CorGIE interface
    • Reflections
Evaluate GNN: previous approaches

Input graph

- Topology
  - Node features

Graph neural network (GNN)

- Node embedding
  - High Dim latent space
    - Dimensional reduction (simple inspection)

Downstream ML applications

- Predictions (node labels, links)
  - Quant metrics
    - e.g. accuracy, hit rate (how many predicted links are true)

Influential nodes

Evaluate GNN: previous approaches

Input graph

- Topology
- Node features

Graph neural Network (GNN)

- Node embedding
- High Dim latent space

Downstream ML applications

- Predictions (node labels, links)
- Quant metrics
- Influential nodes

Dimensional reduction (simple inspection)

Graph metrics

- Jin et al. GNNVis. Arxiv’20.
- Li et al. EmbeddingVis. VAST’18.
Evaluate GNN: previous approaches

Input graph

Graph neural Network (GNN)

High Dim latent space

Downstream ML applications

Predictions (node labels, links)

Quant metrics

E.g. accuracy, hit rate (how many predicted links are true)

Influential nodes

Node classification (predictions → input graph)

Input graph & node embedding under-used!

Ying et al. GNNExpaliner.

Jin et al. GNNVis. Arxiv’20.

Li et al. EmbeddingVis. VAST’18.
Outline

• Introduction of GNN
• Visual evaluation of GNN
  • Previous
  • Ours: CorGIE
    • Overview
    • Data & tasks
    • CorGIE interface
    • Reflections
Evaluate GNN: CorGIE idea

Input graph

Topology
Node features

Graph neural
Network (GNN)

High Dim latent space

Downstream
ML applications

Predictions
(node labels, links)

Shared topo neighbors
Similar node features

GNN

Explore
correspondences

Nearby positions
Evaluate GNN: CorGIE idea

**Input graph**
- **Topology**
- **Node features**

**Graph neural Network (GNN)**
- **Node embedding**
- **High Dim latent space**
- **Downstream ML applications**
- **Predictions** (node labels, links)

**Examples of correspondences:**

- Check [Similar topology? Similar node features?]
- Pick [a cluster]

**Shared** topo neighbors
**Similar** node features

**Explore correspondences**
**Nearby** positions
Evaluate GNN: CorGIE idea

- **Input graph**
  - Topology
  - Node features
  - **Shared** topo neighbors
  - **Similar** node features

- **Graph neural Network (GNN)**
  - Explore correspondences
  - Nearby positions

- **High Dim latent space**
  - **Correspondences**
  - Node embedding

- **Downstream ML applications**
  - Predictions (node labels, links)

**Examples of correspondences:**

- Check [Similar topology? Similar node features?]
  - Pick [a cluster]

- Check [Different topology? Different node features?]
  - Pick [two far-away clusters]
Evaluate GNN: CorGIE idea

Examples of correspondences:

- Check [Similar topology? Similar node features?] \(\rightarrow\) Pick [a cluster]
- Check [Different topology? Different node features?] \(\rightarrow\) Pick [two far-away clusters]
- Pick [two nodes sharing many topo neighbors] \(\rightarrow\) Check [how close the nodes are compared to others?]
Evaluate GNN: CorGIE idea

- **Input graph**
  - Topology
  - Node features

- **Graph neural Network (GNN)**
  - Node embedding

- **High Dim latent space**
- **Downstream ML applications**
  - Predictions (node labels, links)

**Explore correspondences**

- **Shared topo neighbors**
- **Similar node features**

- **K-hop layout**
  - Topo neighbors in hops
  - Clustering structure within hop
Data (sub-)spaces

- **Topology space**
  - Targets: Neighbors; connections

- **Latent space**
  - Clustering; relative positions

- **Feature space**
  - Feature distribution
Tasks

- **Specify** nodes in space
  - Properties of the targets
    - E.g. tight clusters in latent space, disconnected nodes in topology space

- **Correspond** them between spaces
  - Targets should tell the same story between spaces
    - E.g. nodes in tight clusters in latent space are expected to share neighbors in topo space

- **Iterative process**
  - Refine specification
  - Inspire new specification
Outline

• Introduction of GNN
• Visual evaluation of GNN
  • Previous
  • Ours: CorGIE
    • Overview
    • Data & tasks
    • CorGIE interface
    • Reflections
CorGIE interface: K-hop layout

- Show topo neighbors of user-specified node sets
  - Mimic how info is aggregated in the GNN
    - Boxes from left to right: Focal nodes, hop-1, hop-2, ...
  - Within box, cluster neighbors using their topo connections
Multiple views for data spaces and connecting them
Reflections

• Correspondences between input, output, middleware
  • Grey-box approach
    • Works for many GNN models
  • Generalizable to other types of models

• Data space notion

  Topology space
  Neighbors; connections

  Latent space
  Clustering; relative positions

  Feature space
  Feature distribution

• Useful to think about connecting data spaces
• New spaces for future extension
  • e.g., geospatial spaces for graphs dealing with traffic
Visualizing Graph Neural Networks with CorGIE: Corresponding a Graph to Its Embedding

Zipeng Liu, Yang Wang, Jürgen Bernard, Tamara Munzner
Paper under review
Presented at ChinaVis 2021

Take-away

• Evaluate GNN visually by exploring **correspondences** between graph & its embedding
  • Abstraction: connecting data spaces

• Reveal graph topology used in GNN with **K-hop layout**