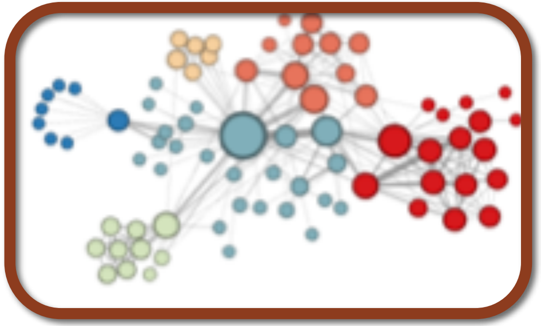
Graphite: Iterative Generative Modeling of Graphs

Aditya Grover, Aaron Zweig, Stefano Ermon
Stanford University

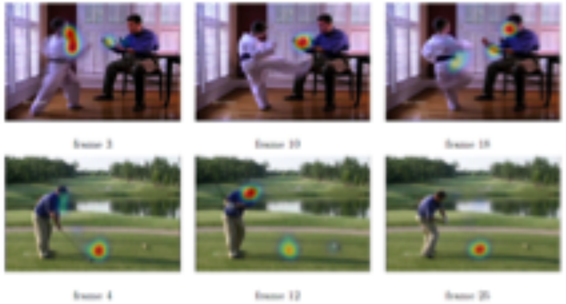
Different **modalities** of structured data


Images

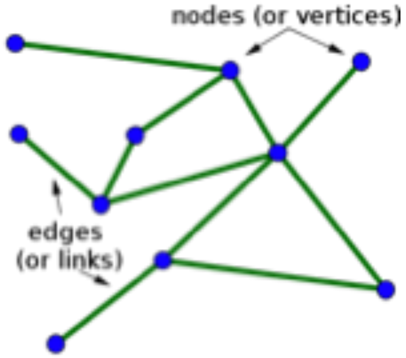

Audio


Graphs

Text


Video

Graphs are **ubiquitous**



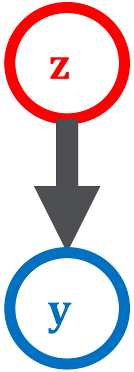
$G = (\mathbf{A}, \mathbf{X})$

Adjacency matrix $\mathbf{A} \in \{0,1\}^{n \times n}$
Feature matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$

Ecology: Food web networks
Biology: Brain networks, Protein-protein interaction networks
Chemistry: Molecules, materials
…

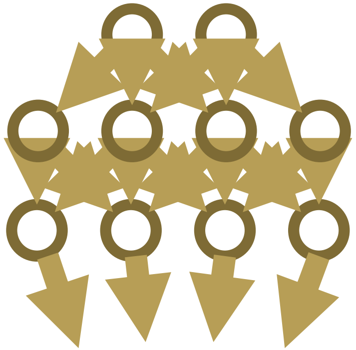
Learning **deep latent variable models** of graphs

$p(\mathbf{z})$



$p(\mathbf{y}|\mathbf{z})$

$p(\mathbf{Z})$

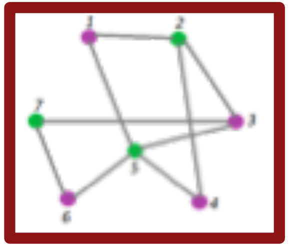


$\mathbf{Z} \in \mathbb{R}^{n \times k}$

H_1

H_2

$p_\theta(\mathbf{A}|\mathbf{X}, \mathbf{Z})$



$G = (\mathbf{A}, \mathbf{X})$

- **Z** is now a matrix with *n* rows
- Every row is a *k*-dimensional vector of node features

What is the right **network architecture** for graphs?

Images – Spatial structure – Convolutional Neural Networks (CNN)

Text, Audio – Temporal structure – Recurrent Neural Networks (RNN)

Inductive biases and invariances for graphs?

- Local structure in terms of graph neighborhoods
- Permutation invariance to node reorderings
- Dynamic can work with graphs of different sizes

Graph Neural Networks

Graph Neural Networks

- Every node passes "messages" (hidden unit activations) to its neighbors
- E.g., Forward pass from $\mathbf{H}^{(l-1)}$ to $\mathbf{H}^{(l)}$:

$$\mathbf{H}^{(l)} = \eta(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{H}^{(l-1)}\Theta^{(l)})$$

  with non-linearity η, degree matrix $\mathbf{D}$, and parameters $\theta^{(l)}$.
- Many variants for message passing possible

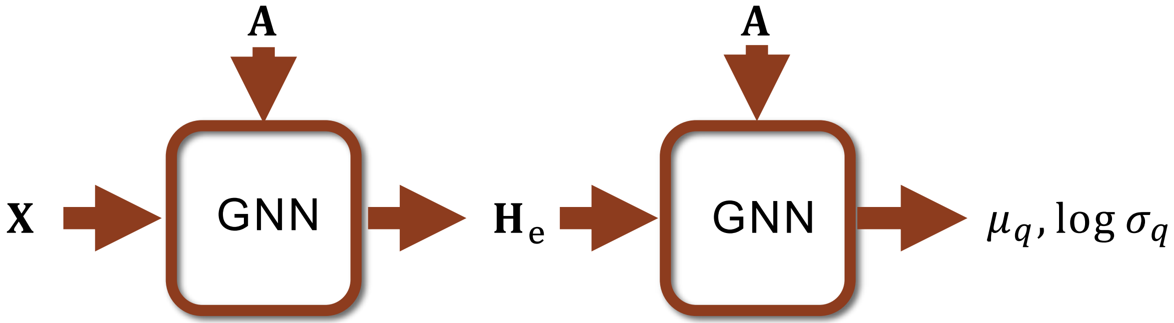
Variational Autoencoding using **Graphite**

- Maximizing the marginal log-likelihood $\log p_\theta(\mathbf{A}|\mathbf{X})$ is intractable
- Introduce a variational posterior $q_\phi(\mathbf{Z}|\mathbf{A},\mathbf{X})$ parameterized by ϕ
- Maximize an evidence lower bound (ELBO) to the log-likelihood

$$\log p_\theta(\mathbf{A}|\mathbf{X}) \geq \underbrace{\mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{A},\mathbf{X})}\left[\log \frac{p_\theta(\mathbf{A},\mathbf{Z}|\mathbf{X})}{q_\phi(\mathbf{Z}|\mathbf{A},\mathbf{X})}\right]}_{ELBO(\theta,\phi)}$$

Graphite **Encoder**

- Variational posterior $q_\phi(\mathbf{Z}|\mathbf{A},\mathbf{X})$ is a multivariate Gaussian with diagonal covariance
- Encoder parameterized by a graph neural network



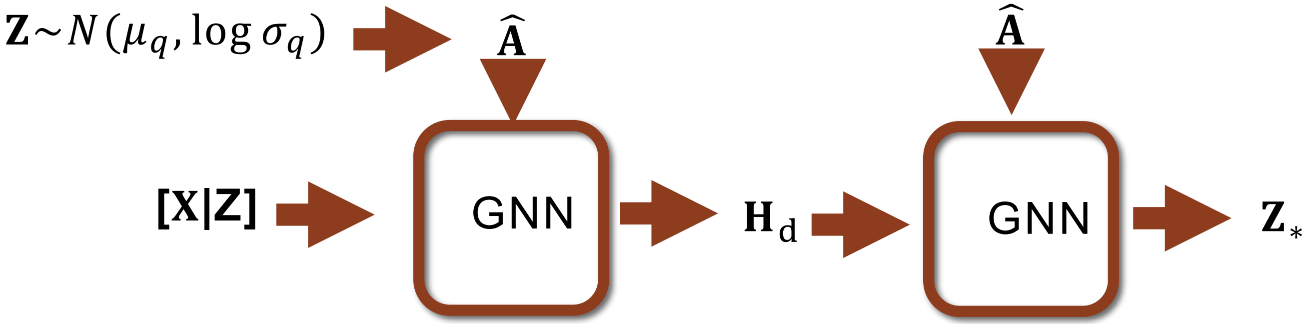
Forward pass of a two layer encoding GNN

Graphite **Decoder**

- Decoder is a hybrid that iterates between:
  - intermediate graph construction using an inner product decoder on initial node representations **Z**

$$\mathbf{A} = \frac{\mathbf{Z}\,\mathbf{Z}^{\mathrm{T}}}{||\mathbf{Z}||_2^2} + \frac{\mathbf{1}\mathbf{1}^{T}}{n}$$

  - Pass graph $\mathbf{A}$ through graph neural network to obtain refined node representations $\mathbf{H}_d$
- Repeat alternative graph refinement steps for every layer to obtain final $\mathbf{Z}_*$



Graphite Decoder

- The final latent feature matrix is specified as a convex combination of the latent layers

$$\mathbf{Z}' = \lambda \mathbf{Z} + (1 - \lambda)\mathbf{Z}_*$$

  where $\lambda \in [0,1]$ is a tunable hyperparameter.

- Observation model $p_\theta(\mathbf{A}|\mathbf{X}, \mathbf{Z})$ is a factorized multivariate Bernoulli

$$p_\theta(\mathbf{A}|\mathbf{Z}, \mathbf{X}) = \Pi_{i=1}^{n} \Pi_{j=1}^{n} p_\theta(A_{ij}|\mathbf{Z}, \mathbf{X})$$
$$\text{where } p_\theta(A_{ij}|\mathbf{Z}, \mathbf{X}) = \sigma(\mathbf{Z}_i'\mathbf{Z}_j')$$

Link Prediction

- Given two nodes in a graph, does an edge exist between the nodes?
- **Baselines:**
  - Spectral Clustering (SC)
  - DeepWalk (DW): random walks + skipgram objective
  - (Variational) Graph Autoencoder (VGAE, GAE): GCN encoder but a single-step inner product decoder
- For Graphite, the task can be formulated as denoising.
- **Datasets:** Protein-protein Interaction, Cora, Citeseer, Pubmed
- **Evaluation metrics:** Area Under the ROC Curve and Average Precision

Evaluation for Link Prediction

Table 1: Area Under the ROC Curve (AUC) scores for link prediction

	Cora	Citeseer	Pubmed
SC	89.9 ± 0.20	91.5 ± 0.17	$\mathbf{94.9} \pm 0.04$
DeepWalk	85.0 ± 0.17	88.6 ± 0.15	91.5 ± 0.04
node2vec	85.6 ± 0.15	89.4 ± 0.14	91.9 ± 0.04
GAE	90.2 ± 0.16	92.0 ± 0.14	92.5 ± 0.06
VGAE	90.1 ± 0.15	92.0 ± 0.17	92.3 ± 0.06
Graphite-AE	91.0 ± 0.15	92.6 ± 0.16	94.5 ± 0.05
Graphite-VAE	$\mathbf{91.5} \pm 0.15$	$\mathbf{93.5} \pm 0.13$	94.6 ± 0.04

Evaluation for Link Prediction

Table 2: Average Precision (AP) scores for link prediction

	Cora	Citeseer	Pubmed
SC	92.8 ± 0.12	94.4 ± 0.11	$\mathbf{96.0} \pm 0.03$
DeepWalk	86.6 ± 0.17	90.3 ± 0.12	91.9 ± 0.05
node2vec	87.5 ± 0.14	91.3 ± 0.13	92.3 ± 0.05
GAE	92.4 ± 0.12	94.0 ± 0.12	94.3 ± 0.5
VGAE	92.3 ± 0.12	94.2 ± 0.12	94.2 ± 0.04
Graphite-AE	92.8 ± 0.13	94.1 ± 0.14	95.7 ± 0.06
Graphite-VAE	$\mathbf{93.2} \pm 0.13$	$\mathbf{95.0} \pm 0.10$	$\mathbf{96.0} \pm 0.03$

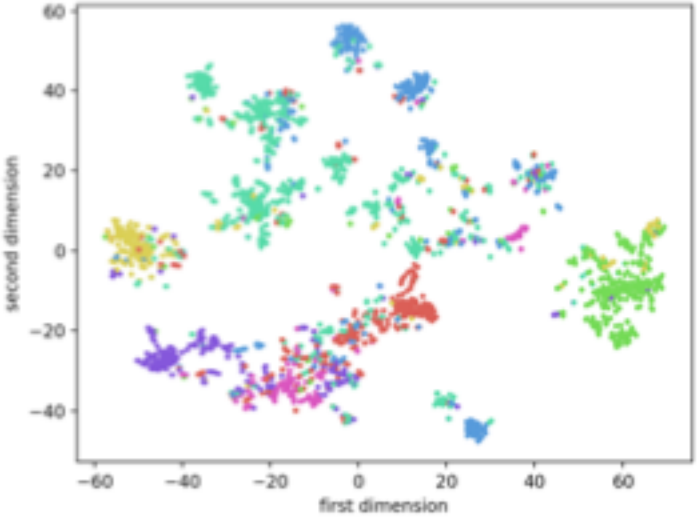
Graphite outperforms **competing methods** on both ROC and AP metrics!

Node Classification

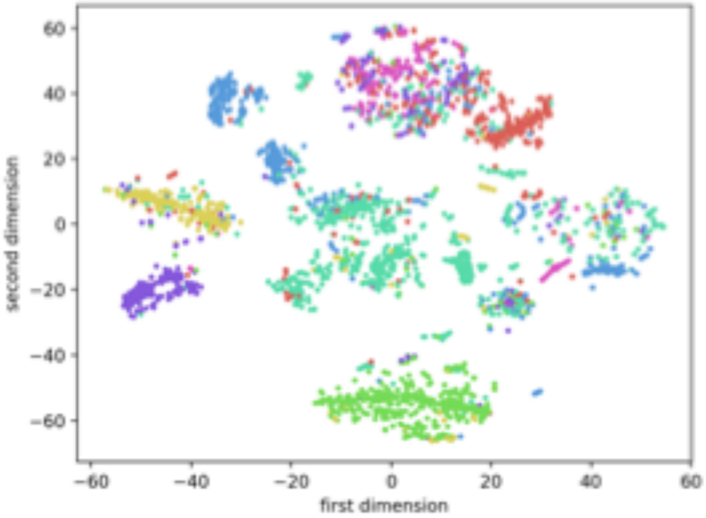
Given labels of few nodes in the graph, predict the labels of other nodes.

	Cora*	Citeseer*	Pubmed*
SemiEmb	59.0	59.6	71.1
DeepWalk	67.2	43.2	65.3
ICA	75.1	69.1	73.9
Planetoid	75.7	64.7	77.2
GCN	81.5	70.3	79.0
Graphite	82.1 ± 0.06	71.0 ± 0.07	$\mathbf{79.3} \pm 0.03$

Visualization of Latent Space



Graphite Autoencoder



Graphite Variational Autoencoder

Cora Dataset

Summary

- Proposed **Graphite**, an algorithmic framework for generative modeling of graphs using variational autoencoding
- Encoder and decoder are parameterized via a Graph Neural Network
- Decoder first creates an intermediate graph via inner products and then gradually refines this intermediate graph
- See paper for more details on a) connections of graph neural networks with mean-field variational inference b) scaling Graphite to large graphs
- Code available at: https://github.com/ermongroup/graphite