Deep Learning for Graphs - Advanced

INTELLIGENT SYSTEMS FOR PATTERN RECOGNITION (ISPR)

DAVIDE BACCIU – DIPARTIMENTO DI INFORMATICA - UNIVERSITA' DI PISA

DAVIDE.BACCIU@UNIPI.IT

Lectures Outline

- ✤ Generative graph learning
 - Probabilistic models on graphs
 - Graph VAE, graph language models and graph diffusion models
- Issues with information propagation on graphs
 - Oversmoothing, oversquashing and undereaching
 - Topological approaches
 - Dynamical systems approaches
- Spatio-temporal and dynamic graphs
- Neural Algorithmic Reasoning
- Applications



Probabilistic Graph Models

Unsupervised Graph Embeddings

- Learn unsupervised node and graph embeddings
 - Requiring less supervised labelling
 - Reusing embeddings across multiple tasks
- Mix supervised and unsupervised modules



Contextual Graph Markov Model (CGMM)



DAVIDE BACCIU - ISPR COURSE

Incremental Construction

- 1. Map the graph to the model (base case)
- 2. Perform inference and freeze states
- 3. Add a new layer and use frozen states as observed variables in the graphical model

Go back to step 2



Computing embedding

Finding the most likely state assignment

$$\max_{i} P(y_u | Q_u = i) P(Q_u = i | \mathbf{q}_{\mathcal{N}(u)})$$

The inferred latent states are used as observable variables in subsequent layers

 A fixed-size vector of states frequencies as graph encoding



CGMM Layer Training



8

DAVIDE BACCIU - ISPR COURSE

CGMM – Depth Matters...

...possibly more than width



Bacciu, Errica, Micheli, JMLR 2020

Interpreting CGMM

Thanks to the probabilistic approach





Bacciu, Errica, Micheli, JMLR 2020

To infinity and beyond



The Infinite CGMM

- Hierarchical Dirichlet process to sample (potentially) infinitely many hidden states
- Automatically learn the size of node embedding space from data
- Choice of observations' groups determined by neighbors' states
- ✓Batch version for larger datasets



ICGMM – Finer grained control on hidden space



Dealing with Multimodal Graph Distributions



Deep Generative Models for Graphs

Graph Generation

Generate a prediction that is itself a graph





UNIVERSITÀ DI PISA

Graph Variational Autoencoder



Simonovsky, Komodakis, ICLR-WS 2018

I Iniversità di Pisa

Language-Based Graph Generation

Sample



Generate a graph node-by-node and edge-by-edge through a sequential approach

Bacciu, Micheli, Podda, Neurocomputing 2020



Generate Molecules by Fragmentation

- Molecule is scanned in SMILES order
- Find first breakable bond
- Break the molecule at that bond, set aside leftmost fragment
- Proceed recursively on rightmost fragment



- ✓ Order is **deterministic** and the molecule can be reconstructed
- Keep a vocabulary of all possible fragments found in a dataset
- Graphs are transformed into fragment sequences





BAVIBE BAEEIU = ISPR EOURSE

Information Propagation in Graphs

DAVIDE BACCIU - ISPR COURSE

(Catastrophic) Issues when learning node/graph embeddings

Many-layer networks are needed to capture long range node interactions into representative embeddings passing through topological bottlenecks



That is where our troubles begin

- Under-reaching
- Over-smoothing
- Over-squashing



Oversquashing



Occurs when an exponentiallygrowing amount of information is squashed into a fixed-size vector



Oversmoothing

Occurs when irrespectively of the propagation length required, the model cannot learn distinctive embeddings to solve the task



Rewiring Approaches



Analysing graph message passing as a diffusion process

Topping et al, ICLR 2022



DAVIDE BACCIU - ISPR COURSE

Rewiring Approaches



Reduce the bottleneck (e.g. by a targeted increase in connectivity) to solve over-squashing



DAVIDE BACCIU - ISPR COURSE

Topping et al, ICLR 2022

A Topological Perspective



Connecting messagepassing issues with the topological properties (curvature) of graphs

Negative curvature R
maybe one of the
causes of
oversquashing



Topping et al, ICLR 2022

A Dynamical Systems View on Deep Graph Networks



A Dynamical Systems View on Deep Graph Networks



- Node message passing can also be seen as a discretization of a continuous dynamical process
- The graph neural network has as many layers as the length of the unfolded ODE

• Neural (Graph) ODE



Non-Dissipative Propagation – Addressing the Problem through the Dynamical System

Leverage the ODE formulation of DGNs to optimize forward and backward message propagation



Chase optimal propagation by enforcing a stable dynamics + nondissipation of the input over time by looking into the properties of Jacobian $\mathbf{J} \Longrightarrow \forall i: \left(Re\left(\lambda_i(\mathbf{J}(t))\right) \right) \approx 0$ Haber & Ruthotto, 2017 Gravina et al, ICLR 2023



Non-Dissipative Propagation by Anti-Symmetry

Anti-symmetric weight matrix with a Euler discretization of the Neural ODE



Local Vs Global Non-Dissipation

So far we have achieved local conservation





UNIVERSITÀ DI PISA

Gravina et al, ICLR 2023

Local Vs Global Non-Dissipation

But we want to achieve it globally

 $\left\| \frac{\partial vec(\boldsymbol{X}(t))}{\partial vec(\boldsymbol{X}(0))} \right\| \approx c$





UNIVERSITÀ DI PISA

Gravina et al, Arxiv 2024

$$\begin{aligned} & \frac{d\mathbf{x}_{u}(t)}{dt} = \sigma\left((\mathbf{W} - \mathbf{W}^{\top})\mathbf{x}_{u}(t) + \Phi(\{\mathbf{x}_{v}\}_{v \in \mathcal{N}_{u}}) + \beta\Psi(\{\mathbf{x}_{v}\}_{v \in \mathcal{N}_{u}})\right) \\ & \text{Node-wise Antisymmetric} \\ & \text{propagation} \end{aligned} \quad (any) \text{Neighbourhood} \\ & \text{aggregation} \end{aligned} \quad Antisymmetric \\ & \text{neighbourhood} \\ & \text{aggregation} \end{aligned} \quad \Psi = \sum_{v \in \mathcal{N}_{u}} (\tilde{\mathbf{A}}_{uv} - \tilde{\mathbf{A}}_{vu}) (\mathbf{Z} + \mathbf{Z}^{\top}) \mathbf{x}_{v}(t) \\ & \text{learnable weights} \end{aligned}$$

One Further Push

Can we allow to the model to trade between dissipative and nondissipative behaviors, adaptively and in a principled way?



UNIVERSITÀ DI PISA

One Further Push

Can we allow to the model to trade between dissipative and nondissipative behaviors, adaptively and in a principled way?



Dynamic Graphs

DAVIDE BACCIU - ISPR COURSE

Learning with Dynamic Graphs



Dynamic Graphs Vs Static DGNs

t = 0



- DGNs cannot be directly applied to all real-life graphs
 - Most real-life graphs are dynamic
 - Majority of DGN approaches assume that the input graph is static
- Ignoring temporal information can make the problem impossible to solve
- Objective: develop methods that are able to exploit both spatial and temporal information



Common Tasks with Dynamic Graphs

- Future link/node prediction
 - Predict at time t + k
- Path classification
 - E.g. predict path congestion
- Event time prediction
 - When an event will occur?
- Imputation





A Taxonomy of Approaches



Can again be tackled as a diffusion process using Graph ODEs

- Spatial and temporal diffusion
- Can be made non-dissipative
- Can naturally handle irregular sampling



A Graph ODE on Continuous Time

A neural ODE that propagates node signals between event occurrences



Integrating Algorithmic Knowledge

Neural Algorithmic Reasoning - Combining algorithms and neural networks



- + Reusable across tasks
- + Executing on noisy conditions
- Sensitive to shift-of-distribution
- No interpretable operations
- Requires lots of data



Veličković el al, ICRL 2020

- Sensitive to task variation
- Input must match pre-conditions
- + Inherent generalisation
- + Interpretable
- + Theoretical guarantees



Can we get the best of both worlds?

Learning Algorithmic Reasoning on Graphs



Example: Ford-Fulkerson, Max-Flow & Min-Cut





DAVIDE BACCIU - ISPR COURSE

Scaling up way out of distribution



Numeroso, Bacciu, Velickovic, ICLR 2023

Example: Ford-Fulkerson, Max-Flow & Min-Cut





DAVIDE BACCIU - ISPR COURSE

Applications

Predicting Properties of Chemical Compounds



DAVIDE BACCIU - ISPR COURSE

-OURS

0.00 0.25

0.50

QED

2

0



2



Generating Molecules

Podda, Bacciu, Micheli, AISTATS 2020

0.2

0.0

0.75 1.00

50

5

10

DI PISA

OURS

RINGS

Tri Quad Pent Hex

ZINC

-ZINC

-OURS

-5

0

LOGP

2

0.2

0.0

-OURS

6

4 SAS

Floorplans Generation



Università di Pisa

Knowledge graphs



A natural way of representing known entities and relationships in a domain

Node/link embeddings are numerical encodings of entities and relationships



Side Effects of Drug Combinations



Analyzing a multimodal graph of interactions

- Drug-drug
- Drug-protein
- Protein-protein





Zitnik, Agrawal, Leskovec, Bioinformatics 2018



Recommendation Systems

...and other kinds of social network analyses

Relational Stock Learning



Point Clouds – Semantic Segmentation



Build point cloud graphs and train semantic class predictors based on vertex embeddings

UNIVERSITÀ DI PISA

Analysis of ICT systems/Blockchains

UNIVERSITÀ DI PISA

Spatio-Temporal Transportation Networks

Forecasting arrival times

Identifying anomalies

 Route replanning

Wrap-Up

Conclusions

- Generative learning for graphs is receiving growing attention, especially in connection with bio-chemical applications
 - Defining continuous processes over combinatorial data is non-trivial
 - Need careful thinking about differentiability
- Current research is heavily focusing on
 - Dynamic graphs and spatio-temporal networks processing
 - DGNs as dynamical systems and their physical interpretation
 - Learning and aligning with (graph) algorithms
 - Oversmoothing, oversquashing and problems of the sort
- …in other words, plenty of opportunities for thesis work!

Advertisement time

Learning on Graph course

- Coming up on Semester 1, Year 2027
- 6 CFU Elective of the AI curriculum
- From foundations of learning on graphs to edge-of-research models

Next Lecture

No lecture on May 20th (Giro d'Italia)

A super-compressed introduction to reinforcement learning (May 21st)

- RL Fundamentals
- Model based RL
- Model free RL
- Hints of deep reinforcement learning

