Graph-based machine learning in healthcare: examples from clinical trials and drug-discovery

Sohrab Ferdowsi

HES-SO Switzerland, University of Geneva

GESAN Journal Club, Campus Biotech, Geneva
December 1st 2021

Haute école de gestion de Genève
Geneva School of Business Administration

UNIVERSITÉ DE GENÈVE

Overview

1 Introductions

2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

Introducing myself..

Background

- MSc. in Biomedical Eng. from SUT, Tehran,
  - Biomedical signal processing,
- PhD in CS from Unige, Dept. Computer Science:
  - "Learning to compress and search visual data in large-scale systems",
  - SP + ML to study fundamental concepts like compression and similarity search.

Currently

- Scientific collaborator at HEG, and "DS4DH" lab of Unige's FacMed,
- Areas of interest:
  - ML on graph-based data,
  - Projects in drug discovery and medical text analysis,
  - Further interest in privacy preserving data sharing.

Machine Learning

Basic idea

- Instead of systematically understanding a system/phenomenon, pays attention to its exemplar behavior.
- At its extreme, considers things as black-box and only pays attention to input-output example pairs.

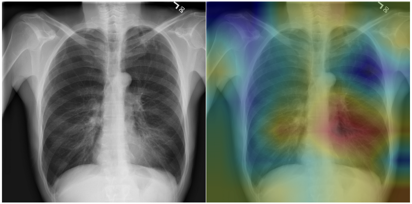
Machine Learning

Basic idea

- Instead of systematically understanding a system/phenomenon, pays attention to its exemplar behavior.
- At its extreme, considers things as black-box and only pays attention to input-output example pairs.

An example

- Pneumonia detection with CheXtNet [Rajpurkar et al, 2017]



	F1 Score (95% CI)
Radiologist 1	0.383 (0.309, 0.453)
Radiologist 2	0.356 (0.282, 0.428)
Radiologist 3	0.365 (0.291, 0.435)
Radiologist 4	0.442 (0.390, 0.492)
Radiologist Avg.	0.387 (0.330, 0.442)
CheXNet	0.435 (0.387, 0.481)

Machine Learning (contd.)

How it works

- Modeling every task as pre-defined parametric families (e.g. neural networks).
- For a task-related target objective, “learning” parameter values using mathematical optimization fed with “data” (input-output pairs).
- Uses high-dimensional “vectorial representations” for data and needs **algebra**.
- Is limited to “differentiability” of parameters and hence relies on **calculus**.
- Should ensure “generalization” and hence requires **probability theory**.

Machine Learning (contd.)

How it works

- Modeling every task as pre-defined parametric families (e.g. neural networks).
- For a task-related target objective, "learning" parameter values using mathematical optimization fed with "data" (input-output pairs).
- Uses high-dimensional "vectorial representations" for data and needs **algebra**.
- Is limited to "differentiability" of parameters and hence relies on **calculus**.
- Should ensure "generalization" and hence requires **probability theory**.

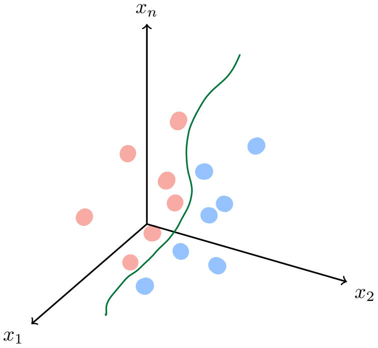
Modern ML practice

- The data to feed the learning procedures is usually abundant.
- The required compute power is relatively democratized.
- The optimization machinery is fully automated with prescribed recipes.
- The whole pipeline is highly transferable across problems and domains.
- Revolutionary breakthroughs observed across many domains since 2012.

An important challenge

Strong bias towards Euclidean geometry

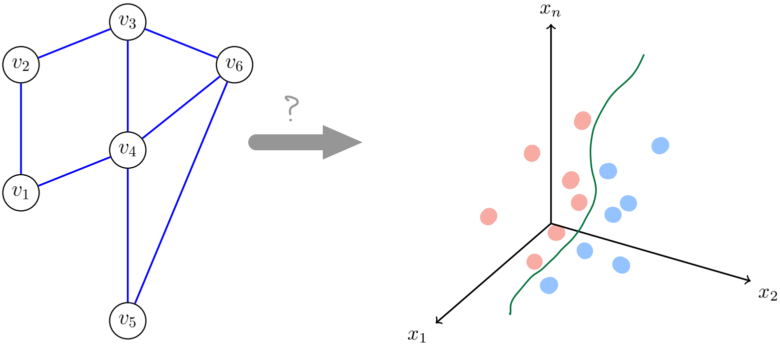
- Data should be modelled as fixed-size "vectors" in $\Re^n$ unified across all samples.



An important challenge

Strong bias towards Euclidean geometry

- Data should be modelled as fixed-size "vectors" in $\Re^n$ unified across all samples.
- Not obvious how to deal with non-vectorial data, notably "graphs".



Overview

1 Introductions

2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

Overview

1 Introductions

2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

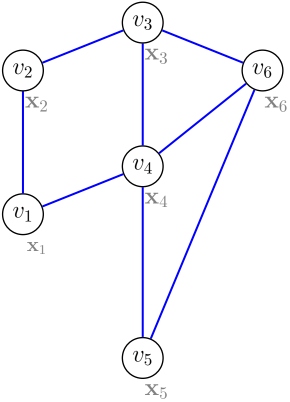
3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

What is a graph?

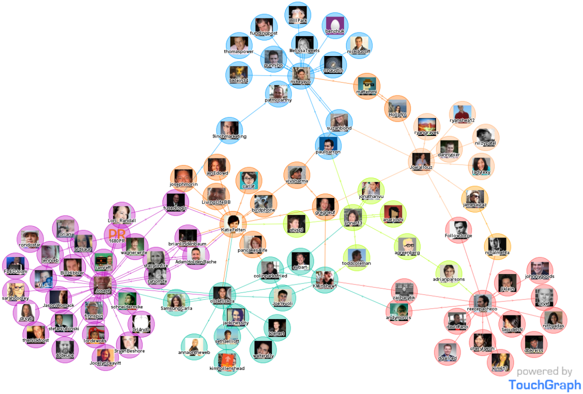
- Node sets $\mathcal{V} = \{v_1, \cdots, v_{|\mathcal{V}|}\}$,
- A set of edges $\mathcal{E}$ with pairs of nodes (u_i, v_i),
- $\mathcal{V}$ and $\mathcal{E}$ summarized in an adjacency matrix $\mathcal{A}$,
- A set of features $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_{|\mathcal{V}|}\}$ for each node.

$$\mathcal{A} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$



What is a graph?

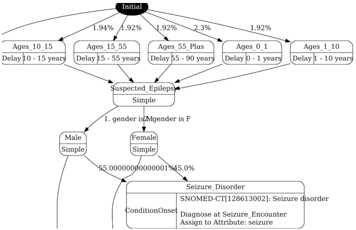
Example: Social network graph

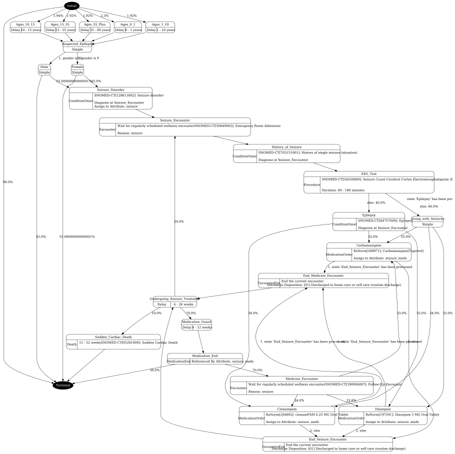


Graph-based data in healthcare

- Graph-based data potentially highly prevalent in healthcare and related fields,
- Usually highly heterogeneous and dynamic in nature,

Diseases (e.g., epilepsy):

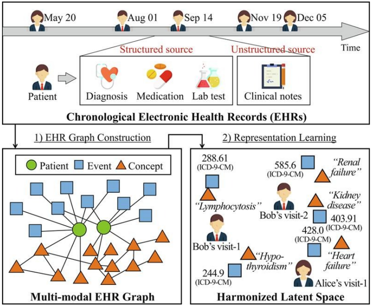




Graph-based data in healthcare

- Graph-based data potentially highly prevalent in healthcare and related fields,
- Usually highly heterogeneous and dynamic in nature,

Patient data ([Dongha Lee et al., 2020]):



Overview

1 Introductions

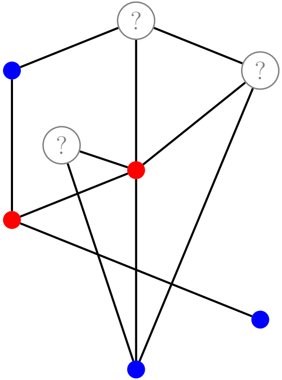
2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

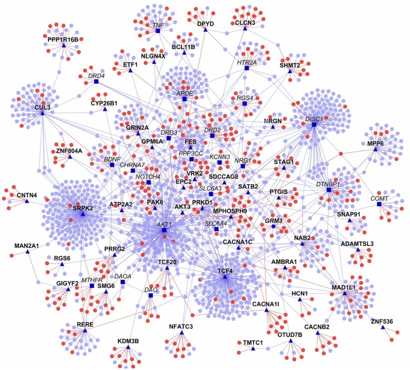
Node classification/regression

- One graph, potentially many nodes,
- Given some supervision (e.g., categorical labels) for some nodes, make inference for other nodes.



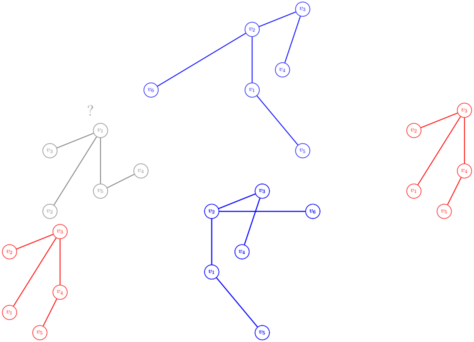
Node classification/regression

- One graph, potentially many nodes,
- Given some supervision (e.g., categorical labels) for some nodes, make inference for other nodes.
- Example in protein-protein interaction to predict Schizophrenia-related activity ([Ganapathiraju et al., 2016, Nature]):



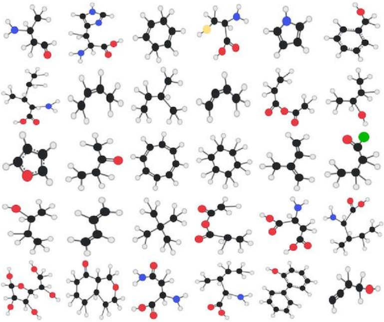
Graph classification/regression

- Multiple graphs treated independently,
- Given some supervision for some graphs, make inference for other graphs.



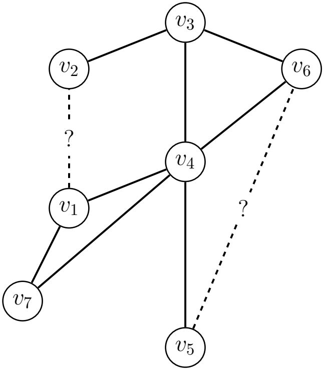
Graph classification/regression

- Multiple graphs treated independently,
- Given some supervision for some graphs, make inference for other graphs.
- Example in toxicity prediction of molecules:



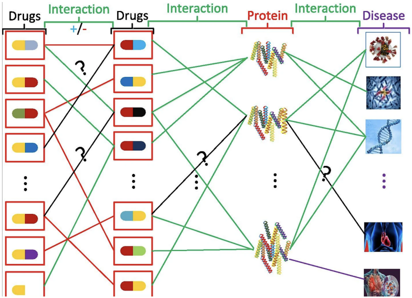
Link prediction

- One graph many edges (and nodes),
- Given partial topology, infer whether some more edges may exist,
- Example in drug discovery [Khushnood et al., 2021, BMC Bioinformatics]:



Link prediction

- One graph many edges (and nodes),
- Given partial topology, infer whether some more edges may exist,
- Example in drug discovery [Khushnood et al., 2021, BMC Bioinformatics]:



Overview

1 Introductions

2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

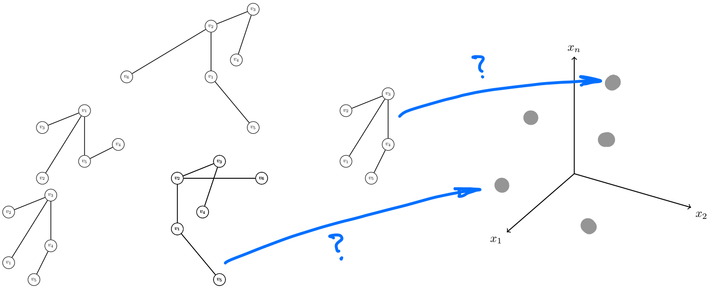
3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

Main challenge of ML on Graphs: Permutation ambiguity

Problem formulation (Graph-level example):

- Finding a vectorial representation ($\mathbf{z}_j$) for every $\mathcal{G}_j \in \{\mathcal{G}_1, \cdots, \mathcal{G}_N\}$,
- $\mathbf{z}_j$ incorporating both features and topology,

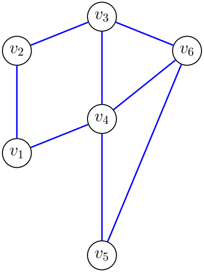


Main challenge of ML on Graphs: Permutation ambiguity

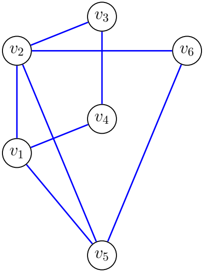
Problem formulation (Graph-level example):

- Finding a vectorial representation ($\mathbf{z}_j$) for every $\mathcal{G}_j \in \{\mathcal{G}_1, \cdots, \mathcal{G}_N\}$,
- $\mathbf{z}_j$ incorporating both features and topology,
- **Main challenge:** Node orderings arbitrary and variable across $\{\mathcal{G}_1, \cdots, \mathcal{G}_N\}$.

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$



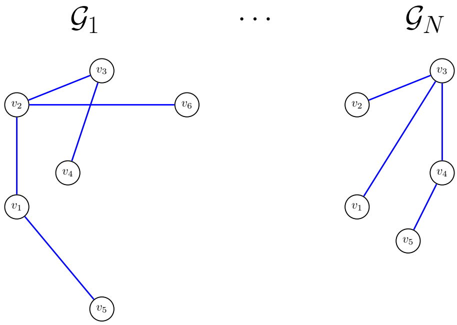
$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$



Main challenge of ML on Graphs: Permutation ambiguity

Problem formulation (Graph-level example):

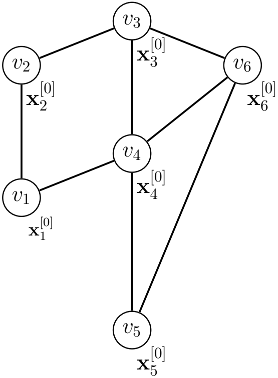
- Finding a vectorial representation ($\mathbf{z}_j$) for every $\mathcal{G}_j \in \{\mathcal{G}_1, \cdots, \mathcal{G}_N\}$,
- $\mathbf{z}_j$ incorporating both features and topology,
- **Main challenge:** Node orderings arbitrary and variable across $\{\mathcal{G}_1, \cdots, \mathcal{G}_N\}$.



Standard solution: Message Passing (MP)

- Start with initial features $\mathbf{x}_{v_1}^{[0]}, \cdots \mathbf{x}_{v_{|\mathcal{V}|}}^{[0]}$, for every node u,
- "Aggregate" features from neighbors to create a "message" using $\mathbb{A}\left\{\cdot\right\}$,
- "Update" the features of u with the received message using $\mathbb{U}\left\{\cdot\right\}$:

$$\mathbf{x}_u^{[l+1]} = \mathbb{U}\left[\mathbf{x}_u^{[l]}; \mathbb{A}\left\{\mathbf{x}_v^{[l]}, \forall v \in \mathcal{N}(u)\right\}\right].$$



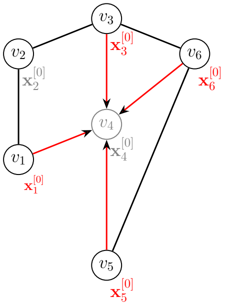
Standard solution: Message Passing (MP)

- Start with initial features $\mathbf{x}_{v_1}^{[0]}, \cdots \mathbf{x}_{v_{|\mathcal{V}|}}^{[0]}$, for every node u,
- "Aggregate" features from neighbors to create a "message" using $\mathbb{A}\{\cdot\}$,
- "Update" the features of u with the received message using $\mathbb{U}\{\cdot\}$:

$$\mathbf{x}_u^{[l+1]} = \mathbb{U}\left[\mathbf{x}_u^{[l]}; \mathbb{A}\left\{\mathbf{x}_v^{[l]}, \forall v \in \mathcal{N}(u)\right\}\right].$$

Aggregate:

$$\mathbf{m}_4^{[0]} = \mathbb{A}\left\{\mathbf{x}_1^{[0]}, \mathbf{x}_3^{[0]}, \mathbf{x}_5^{[0]}, \mathbf{x}_6^{[0]}\right\}$$



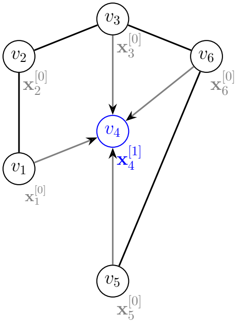
Standard solution: Message Passing (MP)

- Start with initial features $\mathbf{x}_{v_1}^{[0]}, \cdots \mathbf{x}_{v_{|\mathcal{V}|}}^{[0]}$, for every node u,
- "Aggregate" features from neighbors to create a "message" using $\mathbb{A}\{\cdot\}$,
- "Update" the features of u with the received message using $\mathbb{U}\{\cdot\}$:

$$\mathbf{x}_u^{[l+1]} = \mathbb{U}\left[\mathbf{x}_u^{[l]}; \mathbb{A}\left\{\mathbf{x}_v^{[l]}, \forall v \in \mathcal{N}(u)\right\}\right].$$

Update:

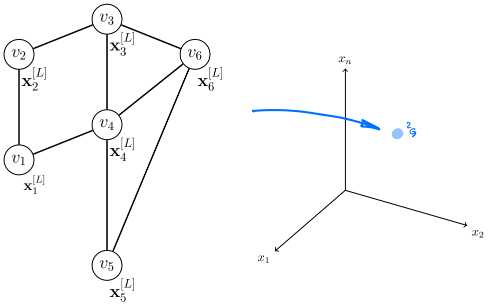
$$\mathbf{x}_4^{[1]} = \mathbb{U}\left\{\mathbf{x}_4^{[0]}, \mathbf{m}_4^{[0]}\right\}$$



Multi-layer Message Passing (MP)

- The operation is repeated L times (layers).
- $\mathbf{x}_u^{[L]}$'s contain both feature information and connectivity (L-hop)
- A global "pooling" $\mathbb{P}_G\{\cdots\}$, provides the final representation $\mathbf{z}_{\mathcal{G}_j}$ for each $\mathcal{G}_j$:

$$\mathbf{z}_{\mathcal{G}_j} = \mathbb{P}_G\left\{\mathbf{x}_v^{[L]}, v \in \mathcal{V}\right\}.$$



Overview

1 Introductions

2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

Drug development cycle

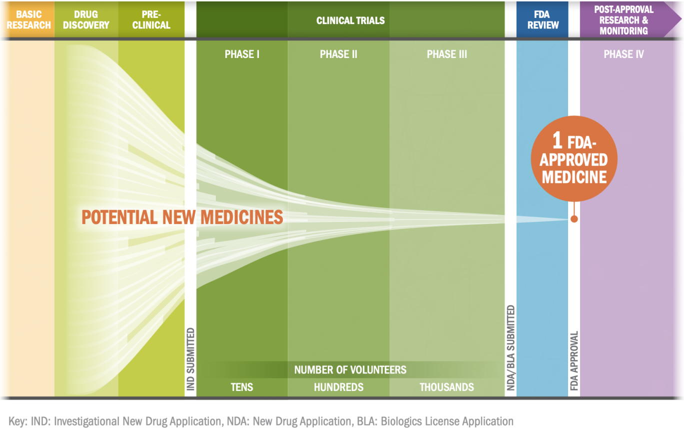


Photo credit: http://phrma-docs.phrma.org/

Overview

1 Introductions

2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

CHEM::AI project with Spirochem AG

Background:

- The chemical space is vast ($\simeq 10^{60}$), far from being randomly explorable,
- ML can help learn useful insights from existing molecules and already known chemistry rules.
- The aim is to reduce the search space in-silico using ML to in-vitro scales ($\simeq 10^{3}$)

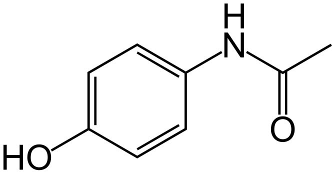
Problem formulation:

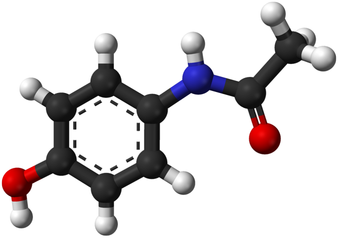
- Biochemists search for novel molecules with:
- Structural similarity to some known molecules,
- Certain chemical properties altered (increased solubility, decreased toxicity, ..)

Molecules as graphs

- Atoms as nodes,
- Bonds as edges,
- Features specify atom types, bond types, 3D coordinates, ..

Paracetamol ($C_8H_9NO_2$):

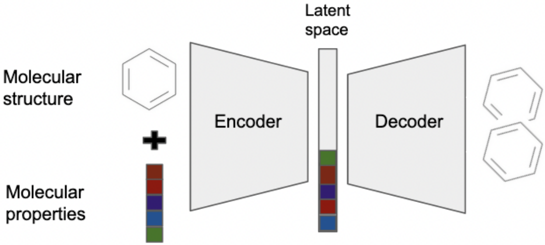




Conditional molecule generation

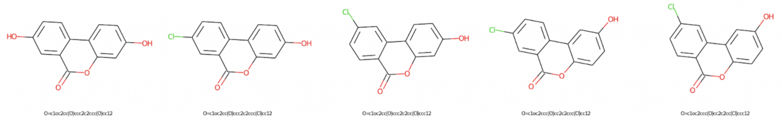
Basic principle:

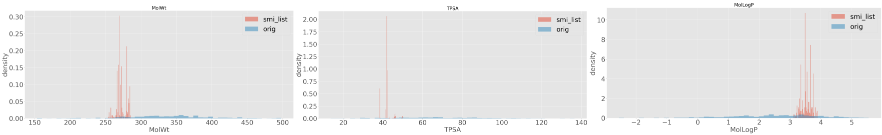
- Training a parametric model on known examples such that:
  - Encodes molecular graphs and their chemical properties in a vectorial space,
  - Disentangles structural and chemical information,
  - Enforces sampling properties to the vectorial space
- Using the model:
  - Get the vectorial equivalent of the example molecule.
  - Instead of its chemical properties, inject the desired ones.
  - Sample around that space and reconstruct the molecular graph equivalents.



Conditional molecule generation (Example results)

- Example molecule was Urolithin A ($C_{13}H_8O_4$),
- Target properties were:
  - Molecular Weight (MW) $= 270$ g/mol,
  - Topological Polar Surface Area (TPSA) $= 45\mathring{A}^2$,
  - Log-Partition (MolLogP) $= 3.5$





Overview

1 Introductions

2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

Risklick Project with Risklick AG

Project objective:

- Risk assessment of Clinical Trials (CT's),
- Identification and mitigation of risk-factors before execution.

This work:

- Classifying CT's into low-risk and high-risk categories,
- Data: US-based *ClinicalTrials.gov* with around 360K samples.
- Title: "Classification of hierarchical text using geometric deep learning: the case of clinical trials corpus",
- Addressing the NLP community: [Ferdowsi et al., EMNLP - November 2021]

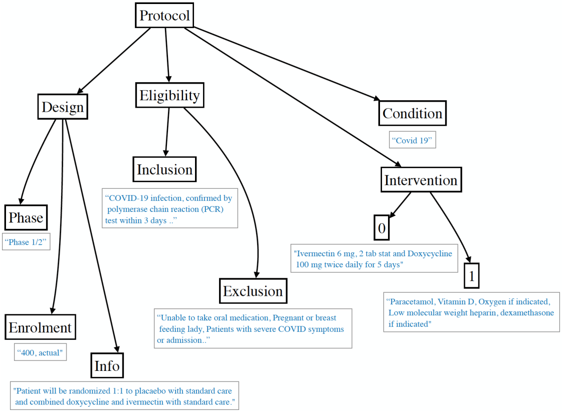
Clinical Trials

Basic facts:

- Clinical Trials (CT's): The ultimate way to assess the safety and efficacy of clinical interventions.
- Highly regulated CT protocol design, indicating all implementation details prior to the study start.
- CT execution takes around 70% of the 13.8 yrs-long drug development cycle and cost $\simeq 40$Mio $.
- Most CT's ($\simeq \%86$) fail to go from phase I to approval, due to various reasons.
- Risk identification highly desirable during protocol design using past records.
- WHO recognizes 18 CT registries, most of them available for download.

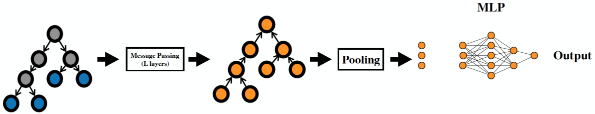
Clinical Trials

A simplified CT example:



From node features to global representation

- Hierarchical text usually appears as a tree.
- $\mathcal{N}(u)$ reduces simply to the set of its children nodes ($\mathcal{C}(u)$).
- Non-leaf nodes $\{u \in \mathcal{V} | \mathcal{C}(u) \neq \emptyset\}$ initialized with zero features.
- Aggregate features from their children during MP iterations.



Selective pooling

- Usually, some of the general structure is invariant across all instances.
- Instead of the global pooling ($\mathbf{z}_{\mathcal{G}_j} = \mathbb{P}_G\{\mathbf{x}_v^{[L]}, v \in \mathcal{V}\}$), we can incorporate this extra knowledge.
- **Selective pooling:**
  - A set of nodes $\bar{\mathcal{V}} \in \mathcal{V}$, with fixed enumeration across all $\mathcal{G}_1, \cdots, \mathcal{G}_N$,
  - Concatenation:
  $$\mathbb{P}_S\left[\mathbf{x}_v^{[L]}, v \in \bar{\mathcal{V}}\right] = \mathop{\Vert}_{v \in \bar{\mathcal{V}}} \mathbf{x}_v^{[L]},$$
  - Final representation:
  $$\mathbf{z}_{\mathcal{G}_j} = \mathbb{P}_G\left\{\mathbf{x}_v^{[L]}, v \in \mathcal{V}\right\} \Vert \mathbb{P}_S\left[\mathbf{x}_v^{[L]}, v \in \bar{\mathcal{V}}\right].$$

CT classification

Method	Precision	Recall	F1-score macro	F1-score micro	AUC
(Elkin and Zhu, 2021)	-	-	-	-	0.7281
Fast-text **(Joulin et al., 2017)**	0.8489	0.7205	0.7531	0.8402	0.8456
BOW-500000-RP768-flat-1	0.6145	0.6300	0.6146	0.6453	0.7034
PubMedBERT-pretrain-768-flat-1	0.6512	0.6763	0.6260	0.6346	0.7246
BOW-1000-flat-9	0.6489	0.6713	0.6488	0.6346	0.7369
BOW-500000-RP768-flat-9	0.7572	0.7793	0.7652	0.7906	0.8701
PubMedBERT-pretrain-768-flat-9	0.8144	0.8144	0.8144	0.8419	0.8911
BOW-500000-RP768-GCN-global	0.8185	0.8233	0.8208	0.8462	0.9116
PubMedBERT-pretrain-768-GCN-global	0.8426	0.8503	0.8463	0.8675	0.8881
BOW-500000-RP768-GCN-selective-9	0.8419	0.8337	0.8376	0.8632	0.9082
PubMedBERT-pretrain-768-GCN-selective-9	0.8454	0.8519	0.8485	0.8697	0.9267

Main observations:

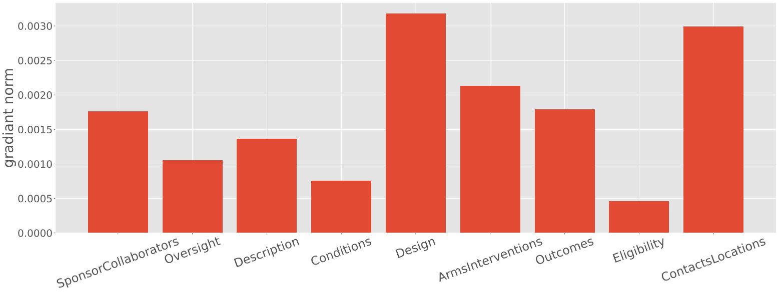
- A net increase of performance from "flat" methods to "graph" methods.
- BOW-based methods perform slightly worse, yet around 50 times faster in inference time.
- Selective pooling successful in incorporating our knowledge of the fixed structure.

Explainability: Identifying the risk factors

General questions:

- Explainability in GNN's: A growing but challenging research area.
- Extension of grid-based techniques to graphs is not trivial (due e.g., to non-differentiability of adjacency matrix, lack of locality).
- A useful workaround: Gradient norms of the selectively-pooled nodes.
- Quantitative results correspond to empirical studies from the CT literature.

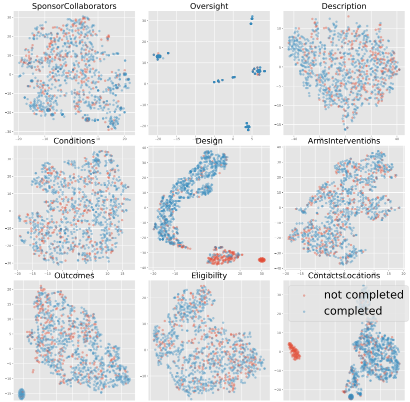
$$\alpha_v = \left|\left|\frac{\partial y^c}{\partial \mathbf{x}_v^{[L]}}\right|\right|_2, \quad v \in \bar{\mathcal{V}}$$



Explainability: Identifying the risk factors

General questions:

- Explainability in GNN's: A growing but challenging research area.
- Extension of grid-based techniques to graphs is not trivial (due e.g., to non-differentiability of adjacency matrix, lack of locality).
- A useful workaround: Gradient norms of the selectively-pooled nodes.
- Quantitative results correspond to empirical studies from the CT literature.



Overview

1 Introductions

2 Graph-based Machine Learning
- Basic definitions and examples
- Common tasks on graphs and applications in life-sciences
- ML on Graphs and GNN's

3 Graphs in our work
- Conditional molecule generation
- Hierarchical text and Clinical Trials data

4 Conclusions

Conclusions

Conceptual:

- Machine Learning focuses on exemplar behavior.
- It requires vectorial representations to encode semantics into the neighborhood.
- Graphs intrinsically have node-ordering ambiguity,
- Yet can (somehow) be cast into vectorial representations thanks to the Message Passing framework (among others).

Applications

- Graph-based data everywhere, particularly within life sciences!
- Many ways how to formulate a problem using graphs.
- Successful examples of the graph approach do exist in many fields.

Graph-based machine learning in healthcare: examples from clinical trials and drug-discovery

Sohrab Ferdowsi

HES-SO Switzerland, University of Geneva

GESAN Journal Club, Campus Biotech, Geneva
December 1st 2021

h e g
Haute école de gestion de Genève
Geneva School of Business Administration

UNIVERSITÉ DE GENÈVE