Graph Representation Learning and Application to Drug Discovery

Jian Tang

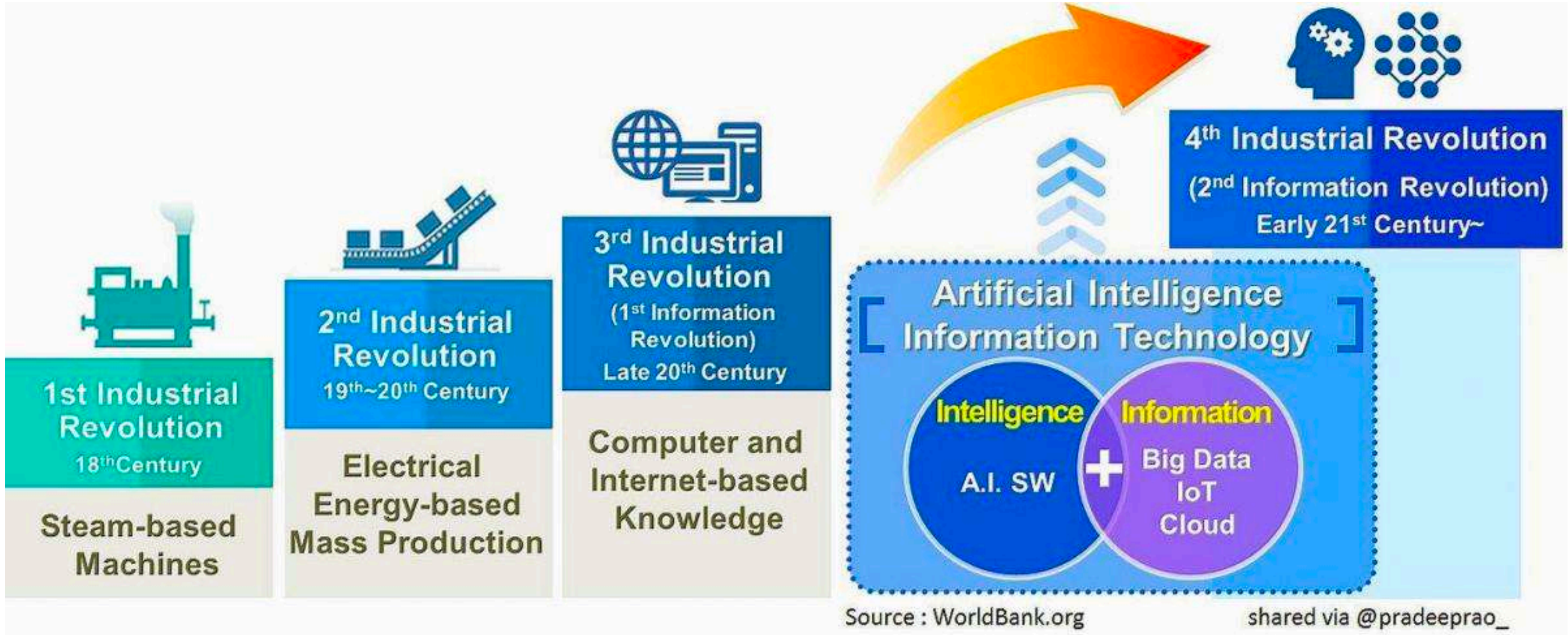
HEC Montreal

CIFAR AI Chair, Mila

Homepage: www.jian-tang.com

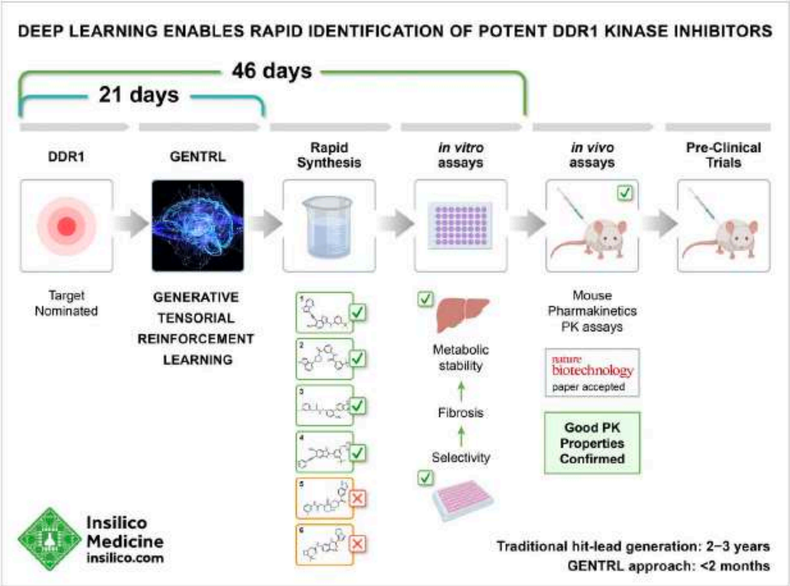
Artificial Intelligence: the Fourth Industrial Revolution

- Artificial Intelligence
  - “the term is often used to describe machines (or computers) that mimic ”cognitive“ functions that humans associate with the human mind, such as ”learning“ and ”problem solving“.” -- Wikipedia



-image from Internet

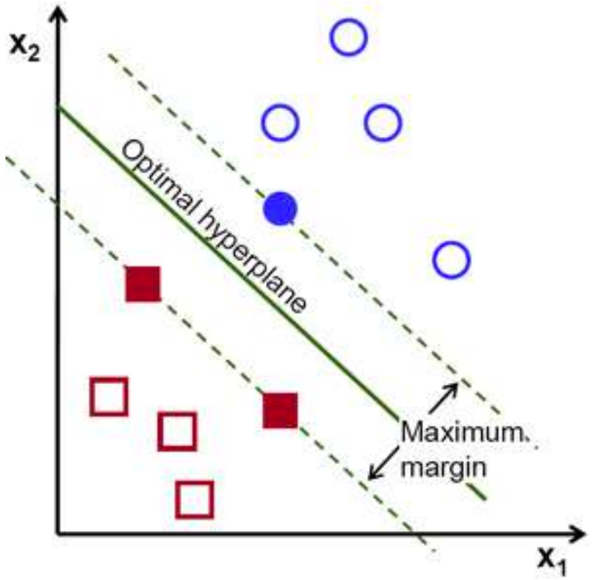
Applications



Machine Learning

- “**Machine learning** is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.”

-Wikipedia



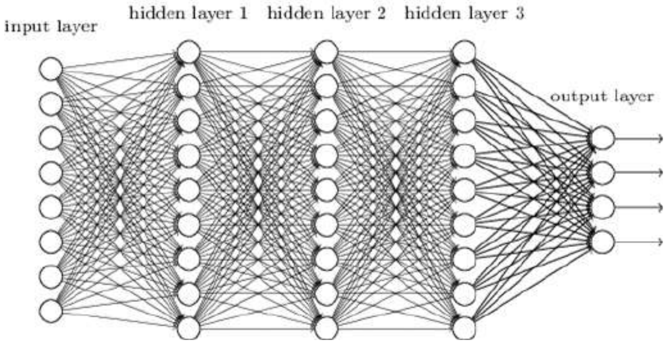
Support vector machines

Hand-crafted Feature Extractor → Simple Trainable Classifier e.g., SVM, LR

Domain experts

Deep Learning = Feature Representation Learning

- Algorithms that allow to learn from features from data (a.k.a, End-to-end learning)



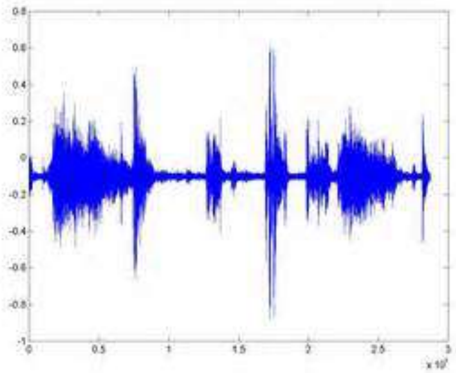
Deep Neural Networks

Trainable
Feature Extractor

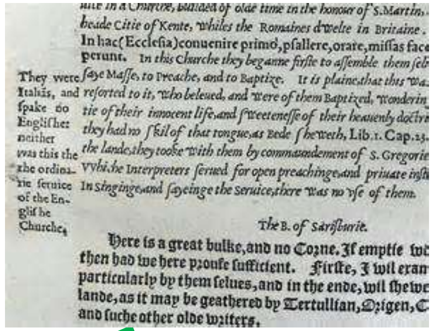
Simple Trainable Classifier
e.g., SVM, LR

Domain experts

Applications of Deep Learning



Speech


beside Citie of Kente, whiles the Romaines dwelte in Britaine.
In hac (Ecclesia) conuenire primo, psallere, orare, missas face
perunt. In this Churche they beganne firste to assemble them selu
They were Italians, and spake no Englishe: neither was this the ordinarie seruice of the Englishe Churche.
saye Masse, to Preache, and to Baptize. It is plaine, that this wa
resorted to it, who beleued, and were of them Baptized, wonderin
tie of their innocent life, and sweetenesse of their heauenly doctri
they had no skil of that tongue, as Bede sheweth, Lib.1. Cap.23.
the lande, they tooke with them by commaundement of S. Gregorie
Which Interpreters serued for open preaching, and priuate instr
In singinge, and sayinge the seruice, there was no vse of them.

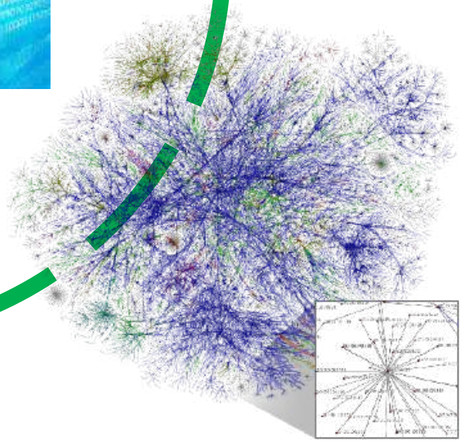
The B. of Sarisburie.

Here is a great bulke, and no Corne. If emptie wo
then had we here proufe sufficient. Firste, I wil exam
particularly by them selues, and in the ende, wil shewe
lande, as it may be geathered by Tertullian, Origen, C
and suche other olde writers.

Natural language

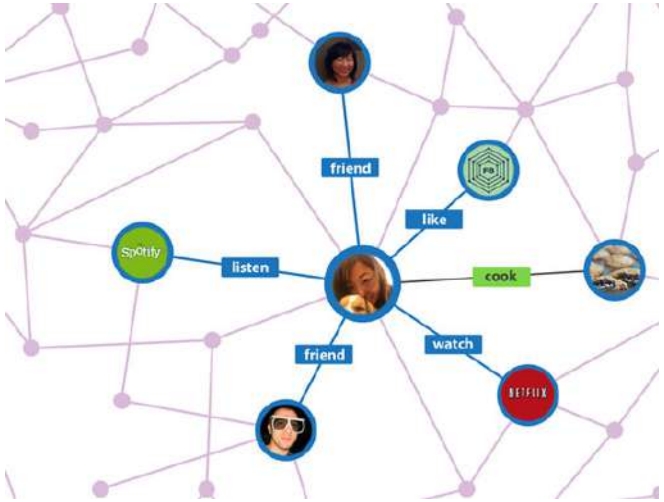
Deep Learning

Image

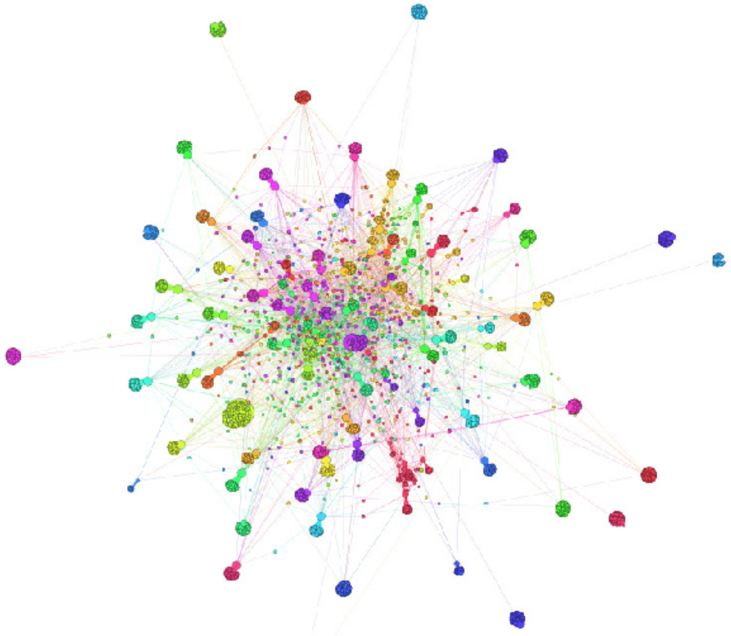


Graphs

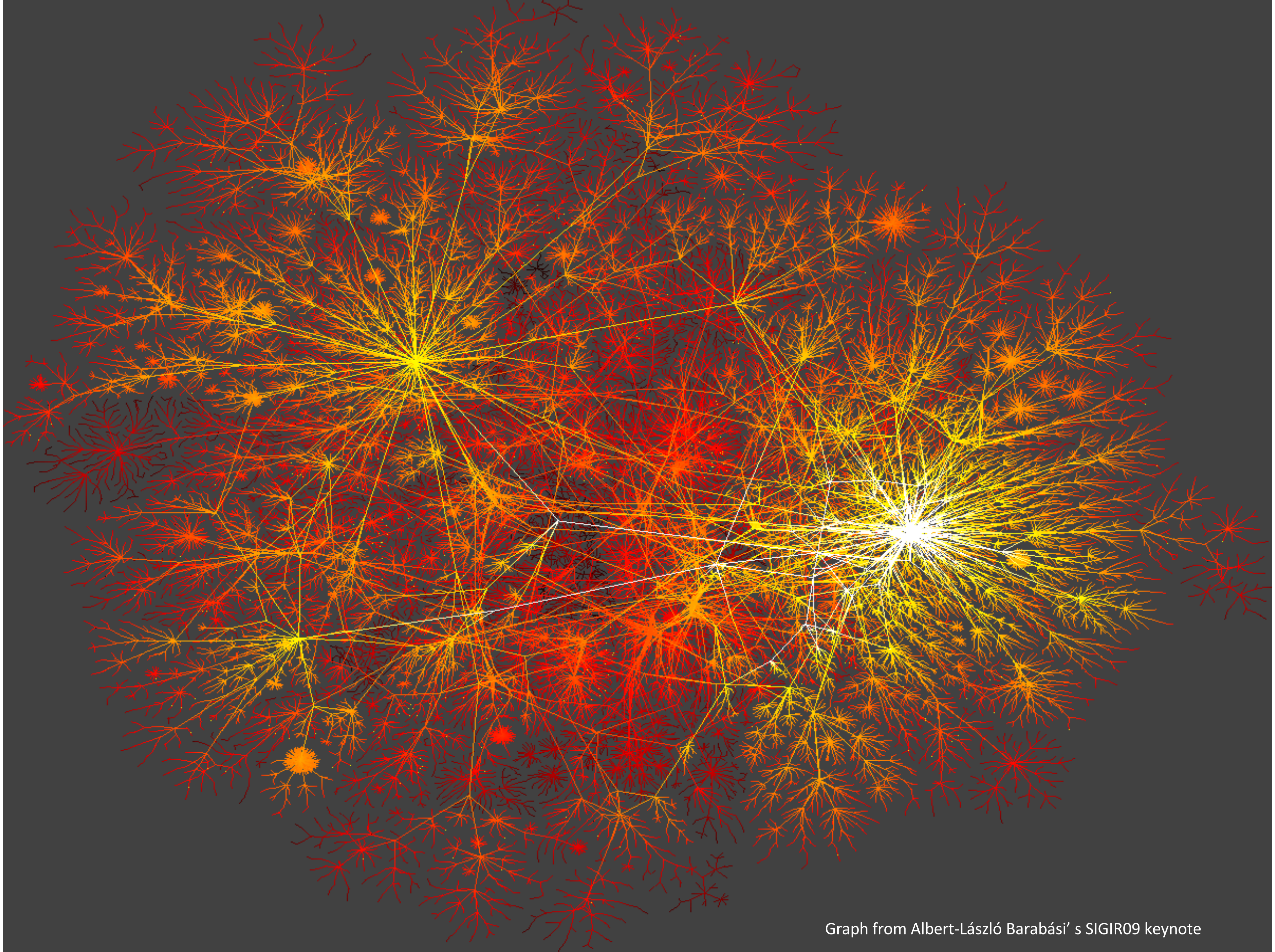
Social Networks



Facebook

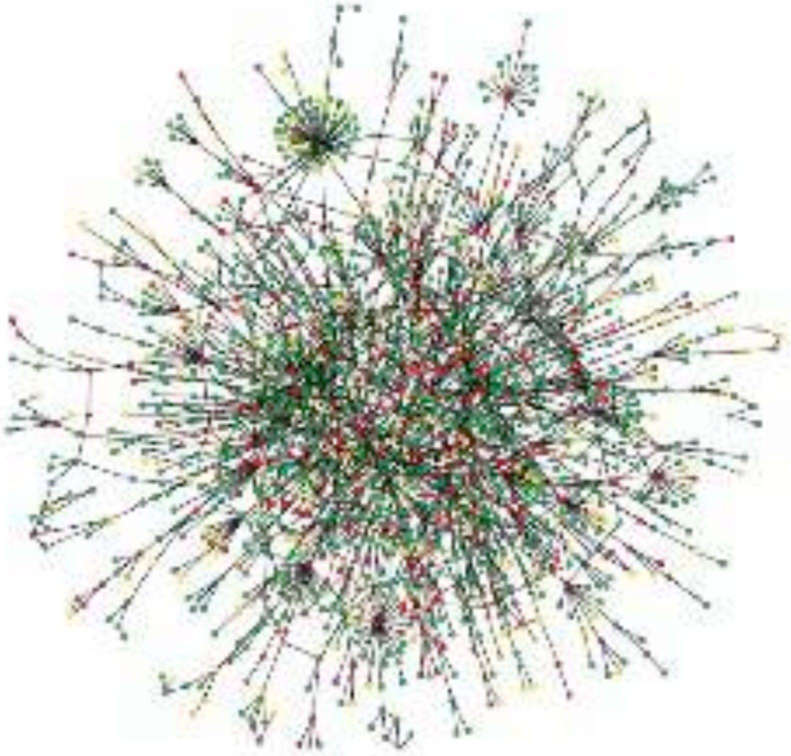


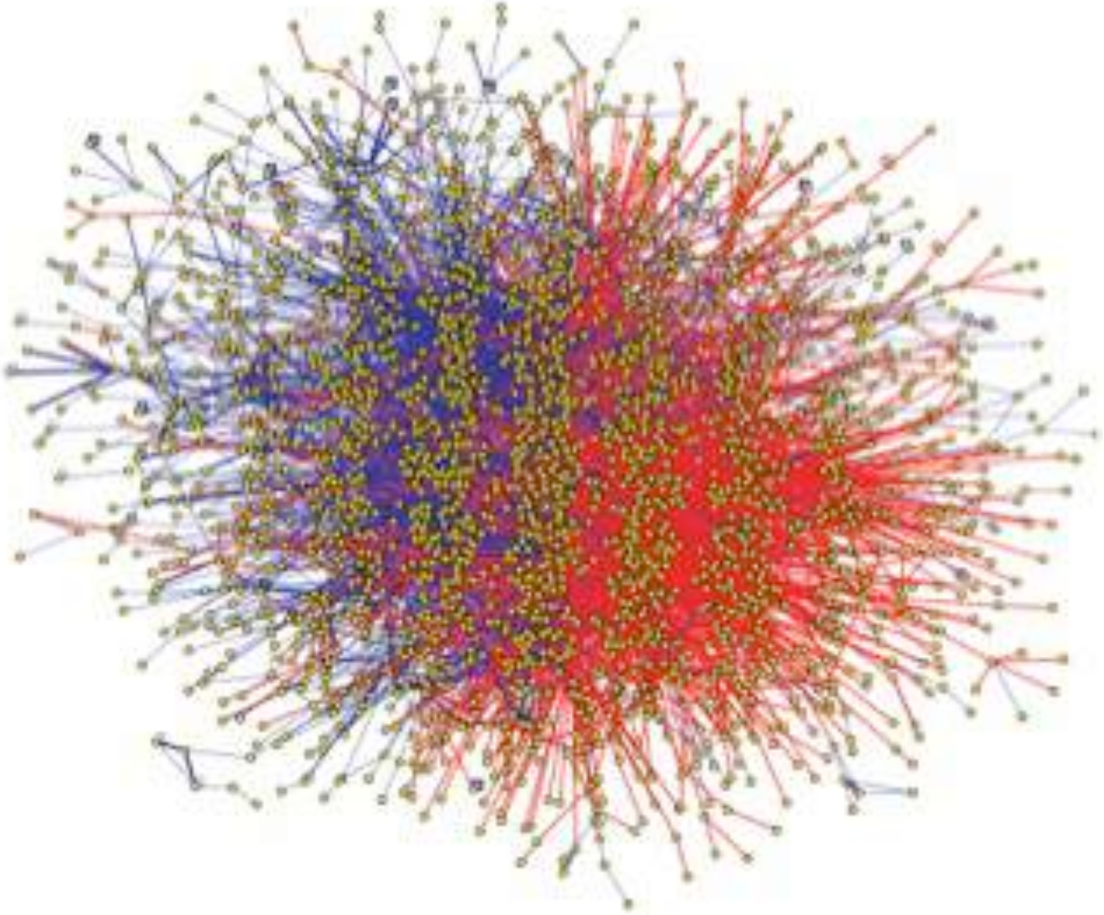
Twitter



Graph from Albert-László Barabási’ s SIGIR09 keynote

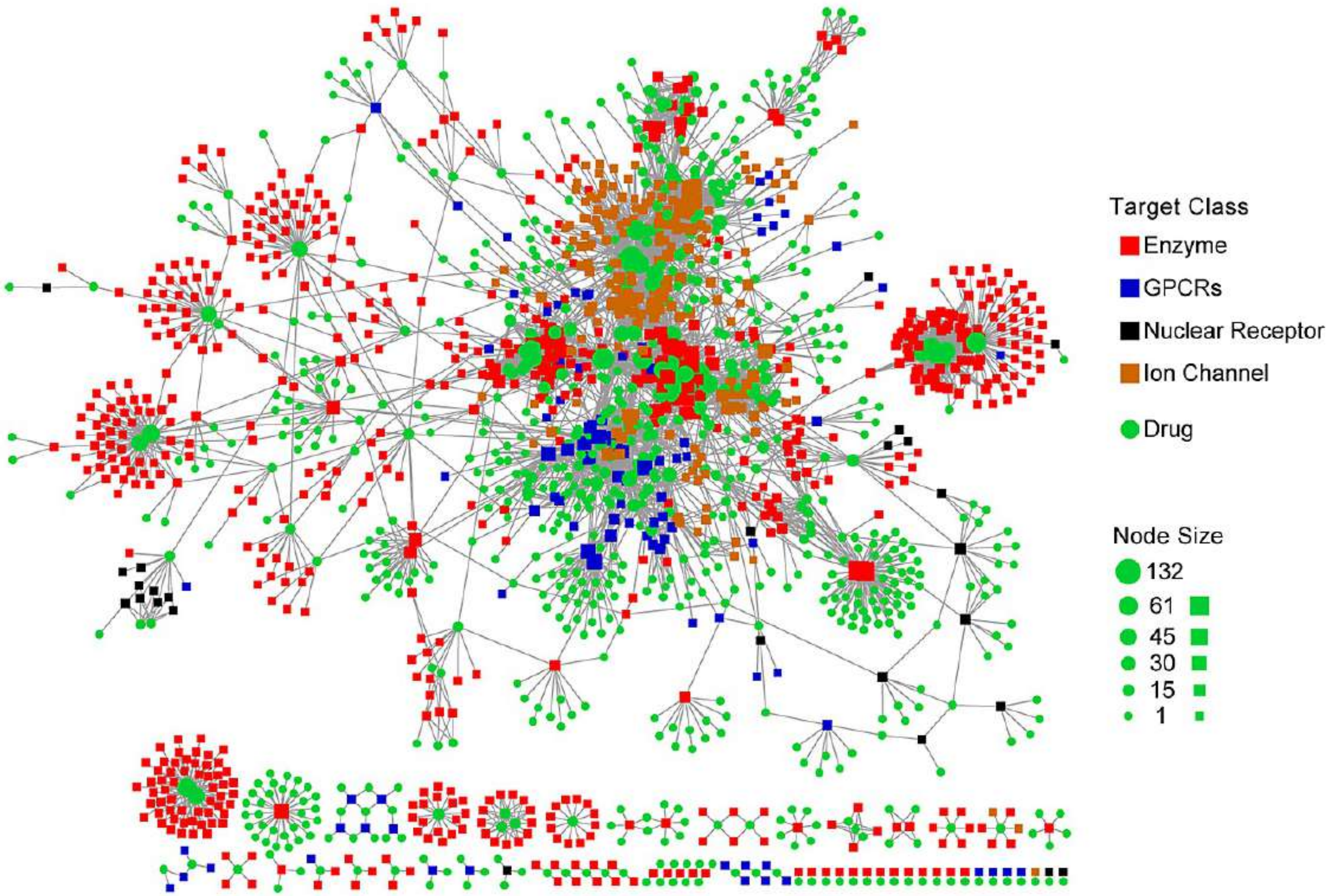
Protein-Protein Interaction Graph





https://www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction/protein-protein-interaction-networks

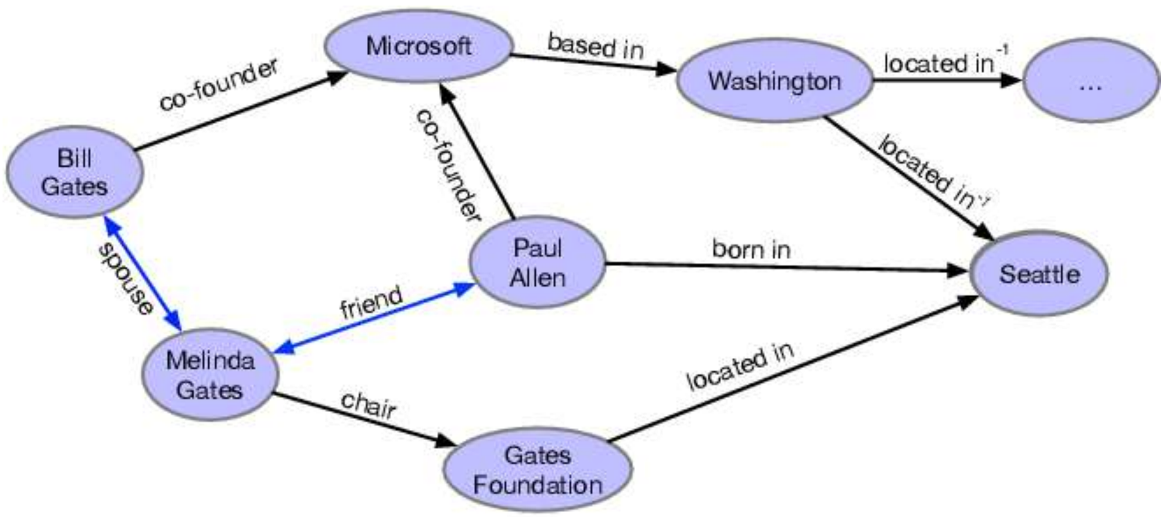
Drug-Protein Interaction Graph

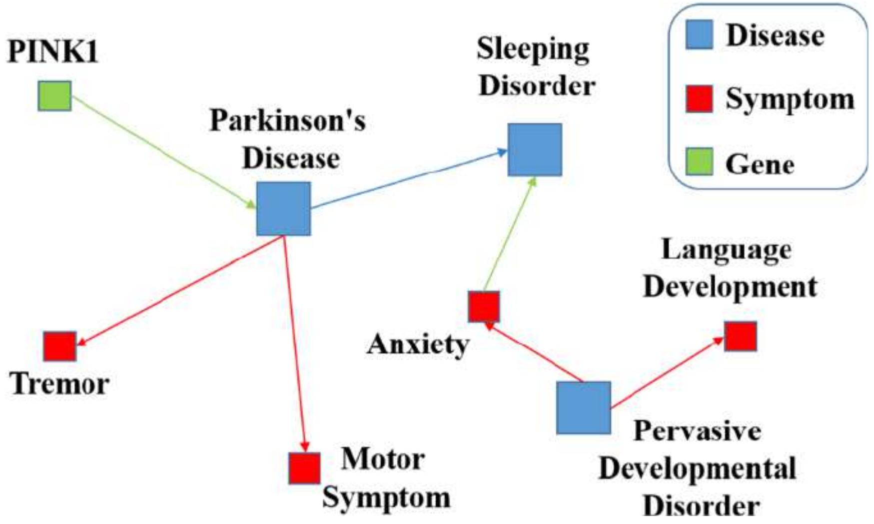


https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002503

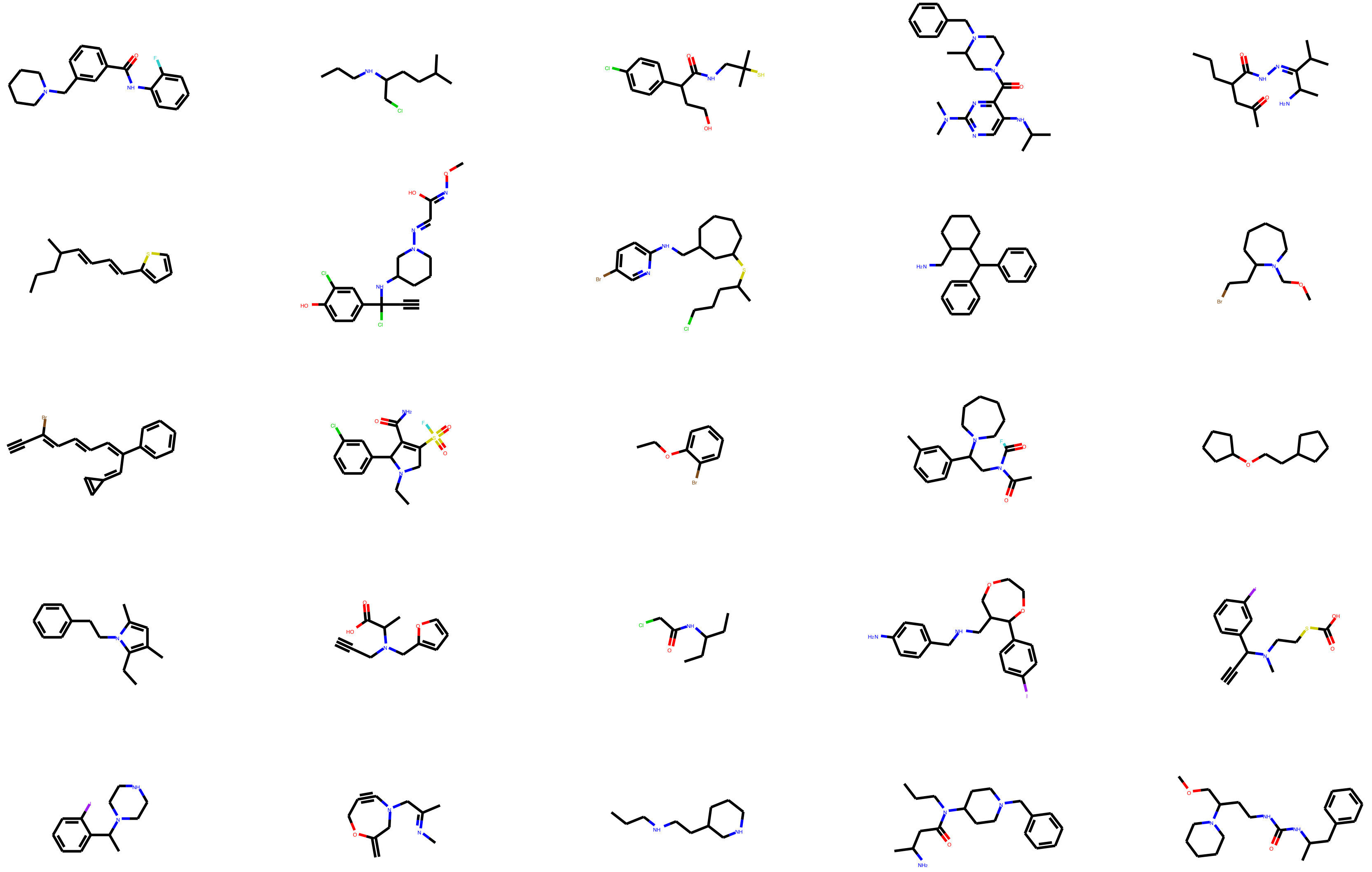
Knowledge Graphs

- Multiple types of edges
  - E.g. Co_founder, Based_in, Located_In
- A set of facts represented as triplets
  - (Bill_Gates, Co_founder, Microsoft)





Molecules

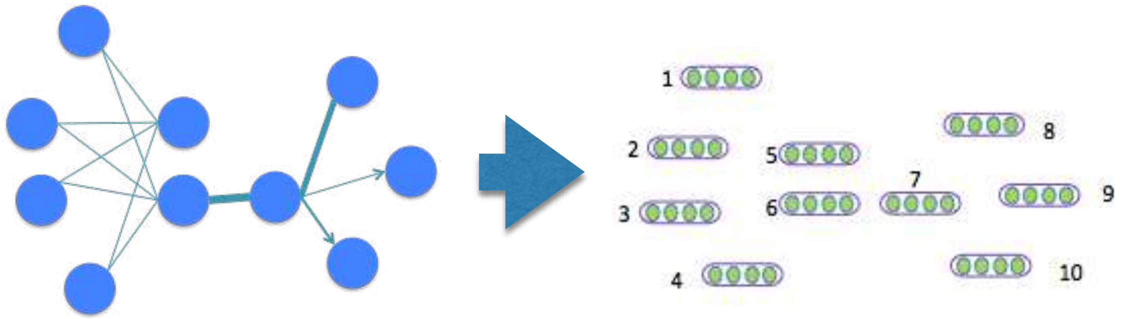


Various Applications on Graphs

- Predicting political preference of Facebook users?
- Recommending friends in social networks
- Predicting the roles of proteins in a protein-protein interaction graphs
- Predicting the effective drugs for a target disease in a biomedical knowledge graph, a.k.a. drug repurposing
- Predicting the chemical properties of molecules
- …
- **Most of these applications require good feature representation of graphs!!**

Research Problem (1): Node Representation Learning

- Represent each node as a low-dimensional vector
  - E.g. social networks
  - Biomedical knowledge graphs (relationships between diseases, proteins, drugs and symptoms)

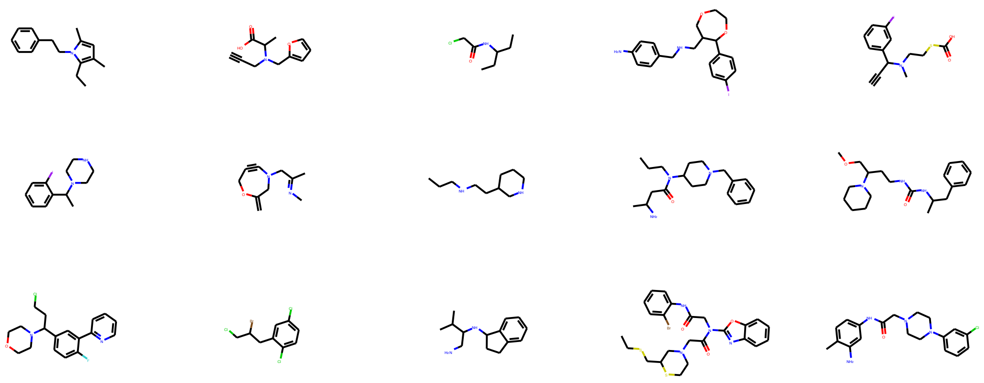


Network Node representations

Research Problem (2): Whole-Graph Representation Learning and Generation

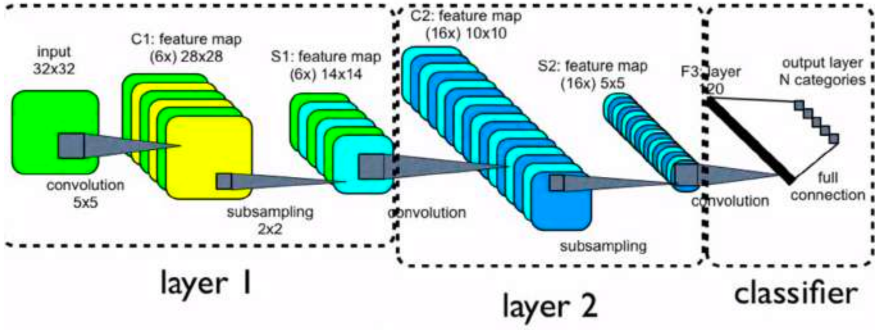
- Represent the whole graph as a low-dimensional vector
  - Predicting the chemical properties of molecules
- Generate graphs (e.g., molecular structures)
  - e.g., molecule design

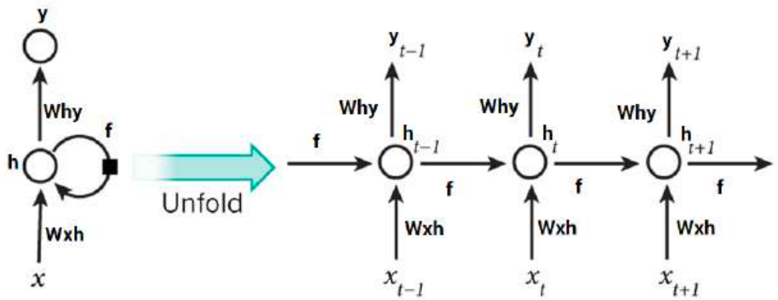
molecular graph



Challenges of Graph Representation Learning

- Existing deep neural networks are designed for data with regular-structure
    - images, text, and speech

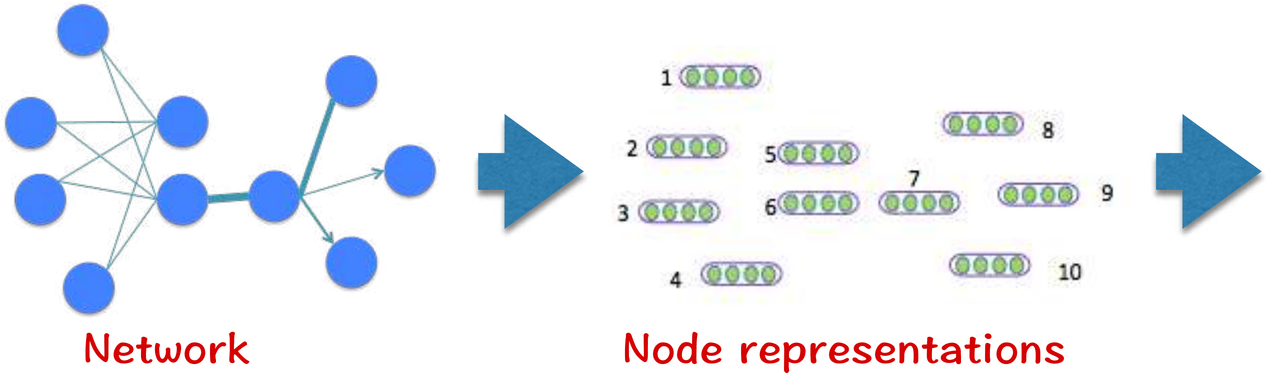




- Graphs are very complex
    - Arbitrary structures
    - Large-scale: more than millions of nodes and billions of edges
    - Heterogeneous: directed/undirected, binary/weighted/typed

Part I: Graph Representation Learning

Learning Node Representations (LINE, Tang et al. 2015)



- Node Classification
- Node Clustering
- Link Prediction
- Recommendation
- …

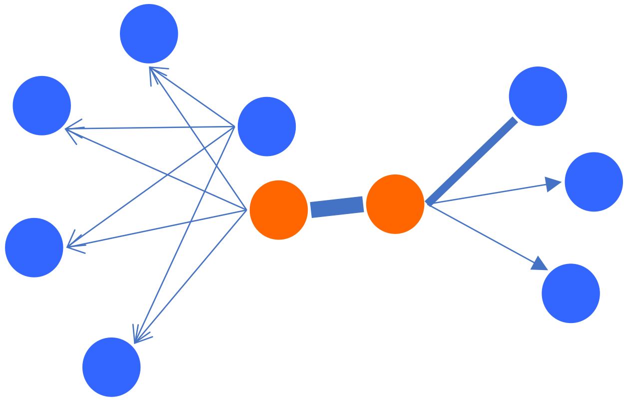
- E.g., Facebook social network -> user representations (features)-> friend recommendation
- Protein-Protein network/Gene-gene network -> protein/gene representations

Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan and Qiaozhu Mei. LINE: Large-scale Information Network Embedding. WWW'15.

LINE: Large-scale Information Network Embedding (Tang et al. 2015, >2,600 citations)

- Arbitrary types of networks
  - Directed, undirected, and/or weighted
- Clear objective function
  - Preserve the first-order and second-order proximity
- Scalable
  - Asynchronous stochastic gradient descent
  - Millions of nodes and billions of edges: a coupe of hours on a single machine

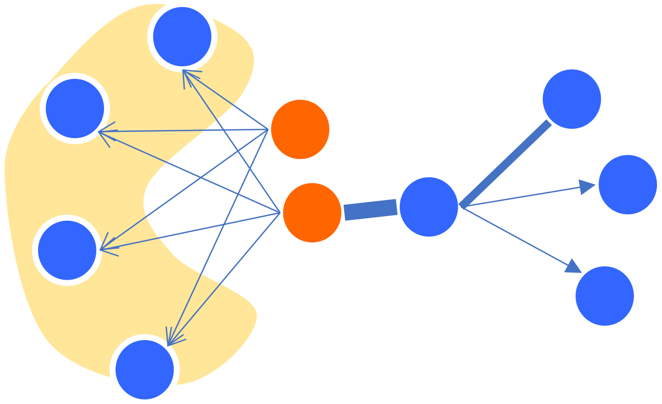
First-order Proximity



- The local pairwise proximity between the nodes
- However, many links between the nodes are not observed
  - Not sufficient for preserving the entire network structure

Second-order Proximity

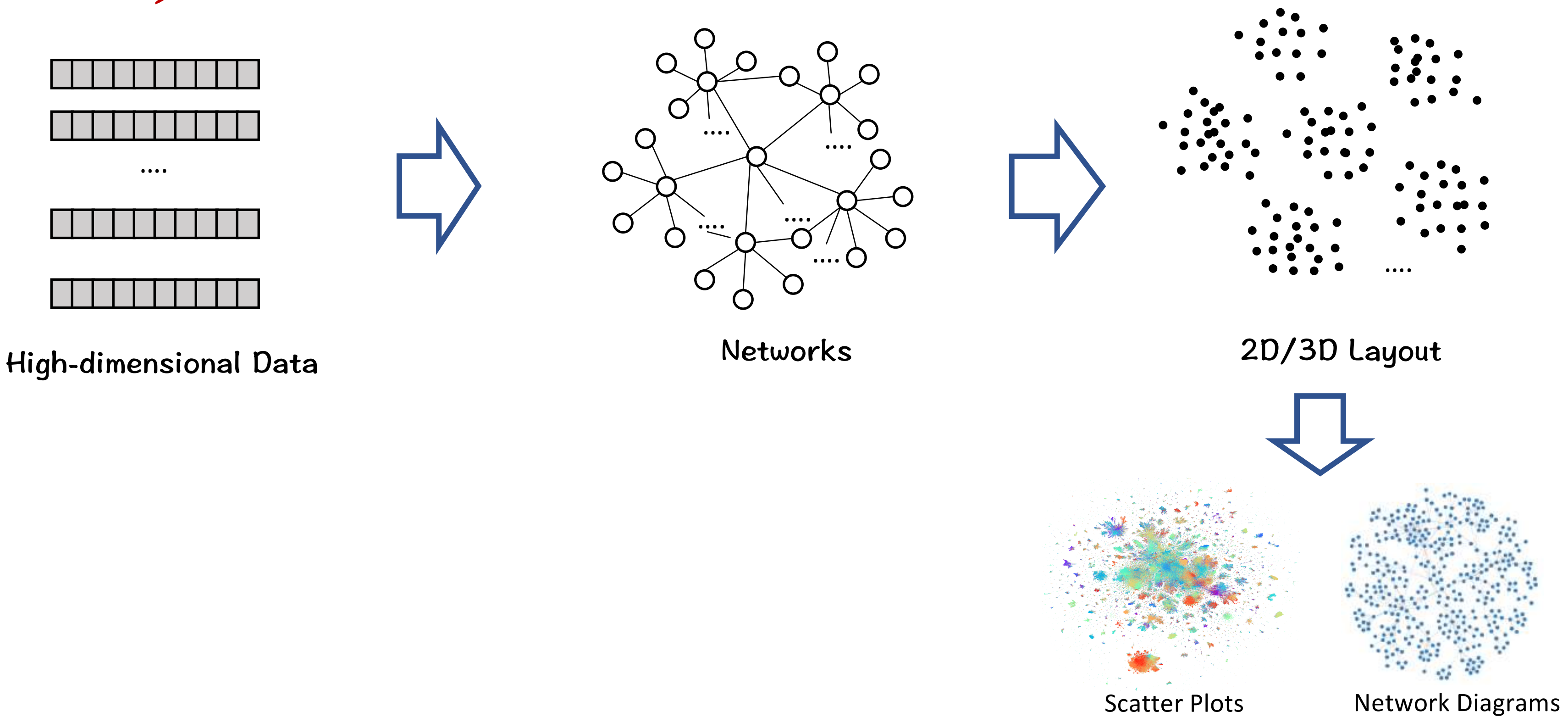
"The degree of overlap of two people's friendship networks correlates with the strength of ties between them" --Mark Granovetter

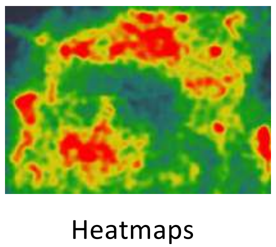


"You shall know a word by the company it keeps" --John Rupert Firth

- Proximity between the neighborhood structures of the nodes

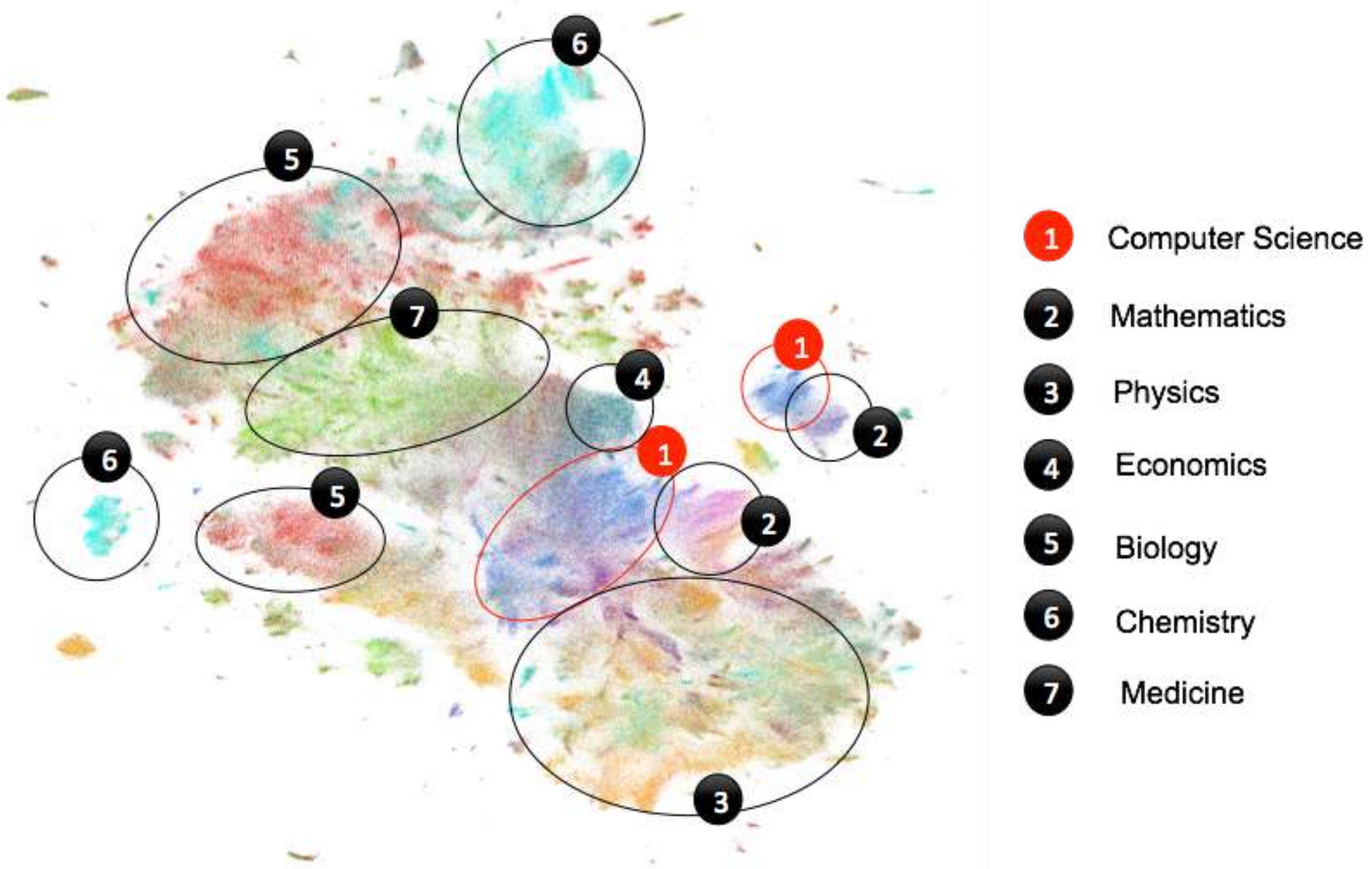
Extremely Low-dimensional Representations: 2D/3D for Visualizing Graphs (LargeVis, Tang et al. 2016)





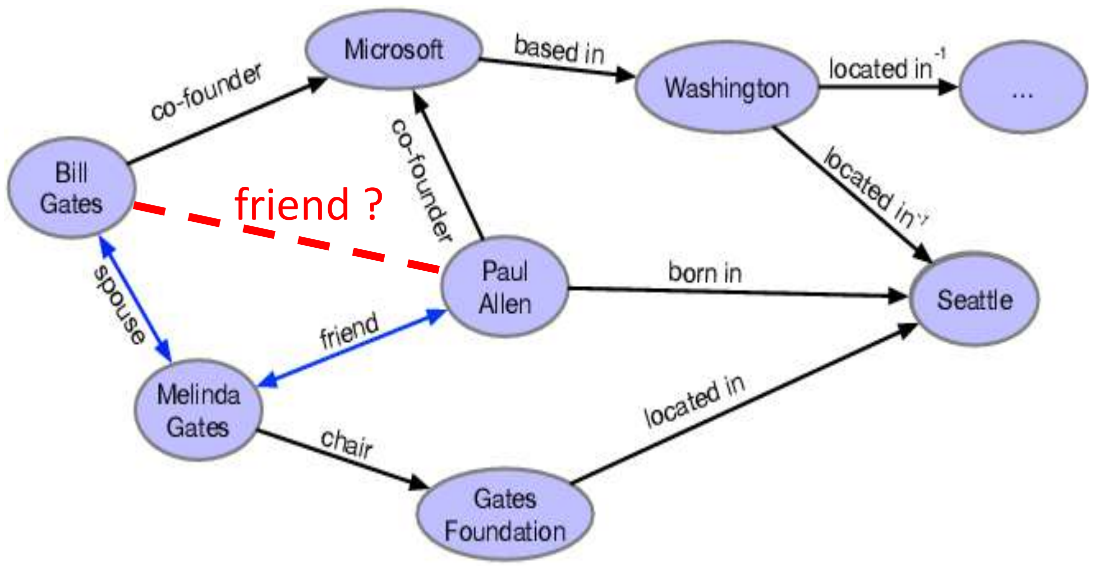
Jian Tang, Jingzhou Liu, Ming Zhang and Qiaozhu Mei. Visualizing Large-scale and High-dimensional Data. WWW’16.

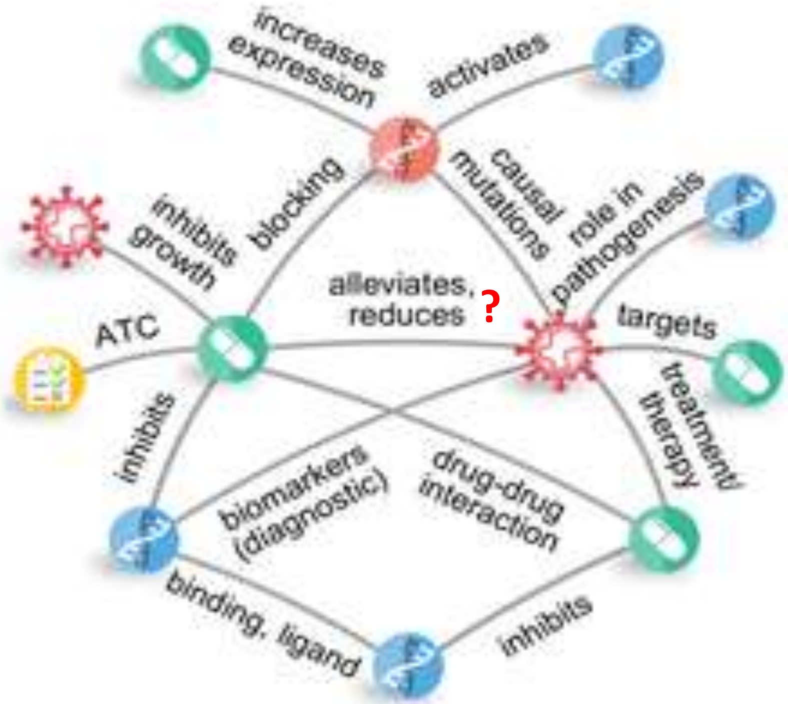
10M Scientific Papers on One Slide



Knowledge Graph Embedding (Sun et al. 2019)

- Learning low-dimensional representations of **entities** and **relations**
- Preserve the relationships between entities in low-dimensional spaces
- Task: Link prediction on knowledge graphs





Knowledge Graph Completion

- A fundamental task: **predicting missing links**
- Key Idea: model and infer the **logical rules** in knowledge graphs according to observed knowledge facts.
- Example:
    - Parents of Parents are Grandparents
    - Husband and wife are inverse to each other
    - A compound treats disease1, disease1 resembles disease 2
      => the compound treats disease 2

Relation Patterns

- **Symmetric/Antisymmetric** Relations
  - Symmetric: e.g., Marriage
  - Antisymmetric: e.g., Filiation
- **Inverse** Relations
  - Husband and wife
- **Composition** Relations
  - My mother's husband is my father

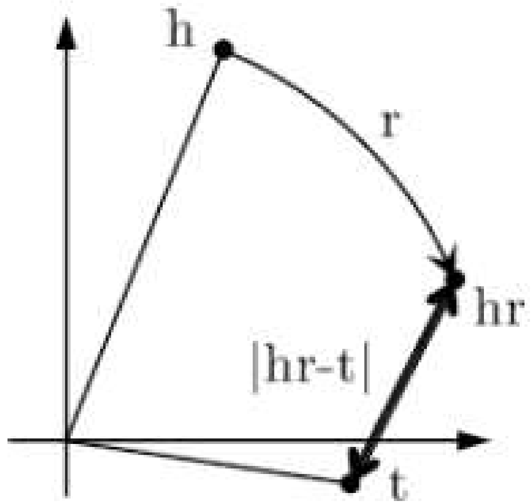
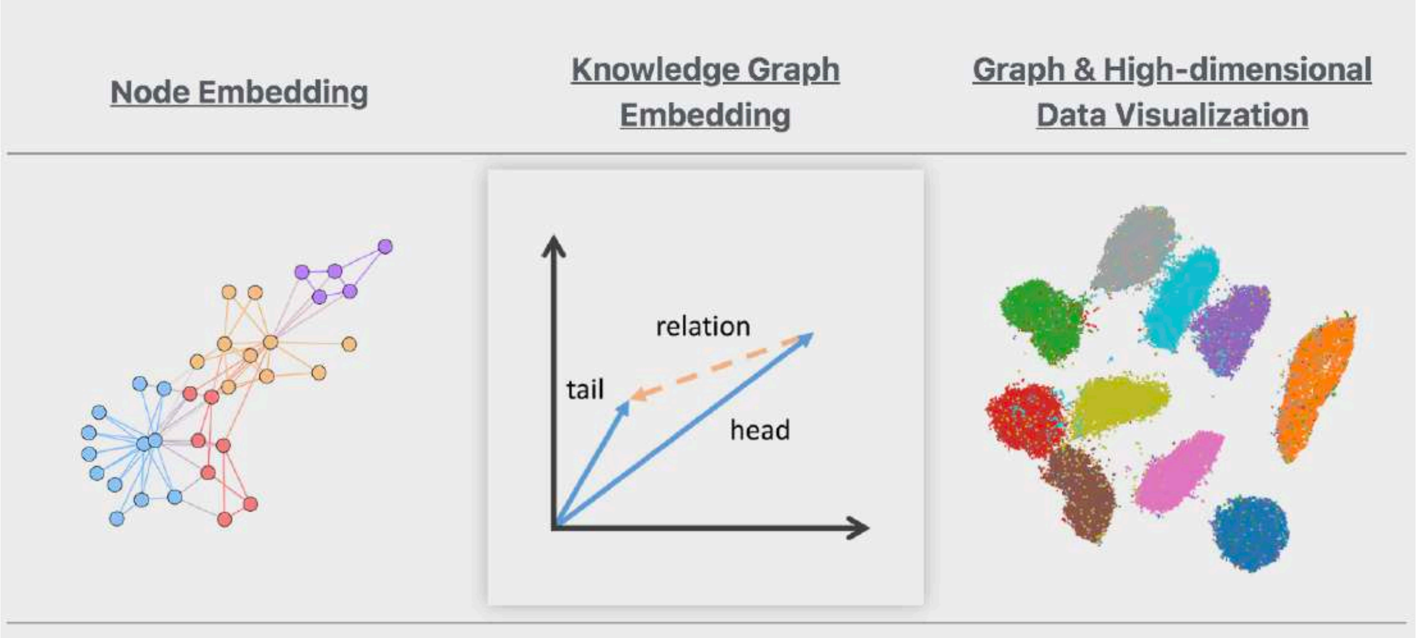


Figure: RotatE (Sun et al. 2019)

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, Jian Tang. "RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space." ICLR'19.

GraphVite: A High-performance and General Graph Embedding System (Zhu et al. 2019)

- A system specifically designed for learning graph embeddings with GPUs
- Super efficient!! Take only one minute for learning node representations of a graph with one-million nodes
- https://graphvite.io



Zhaocheng Zhu, Shizhen Xu, Meng Qu, **Jian Tang**. "GraphVite: A High-Performance CPU-GPU Hybrid System for Node Embedding". WWW'19.

Part II: Graph Representation Learning for Drug Discovery

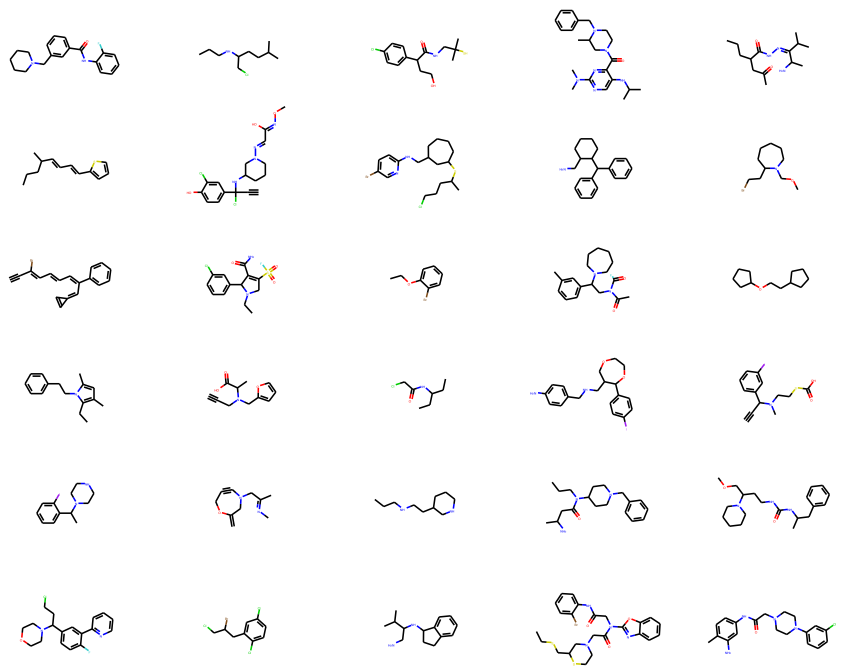
The Process of Drug Discovery

- A very long and costly process
  - On average takes more than 10 years and $2.5B to get a drug approved
- Big opportunities for AI to accelerate this process

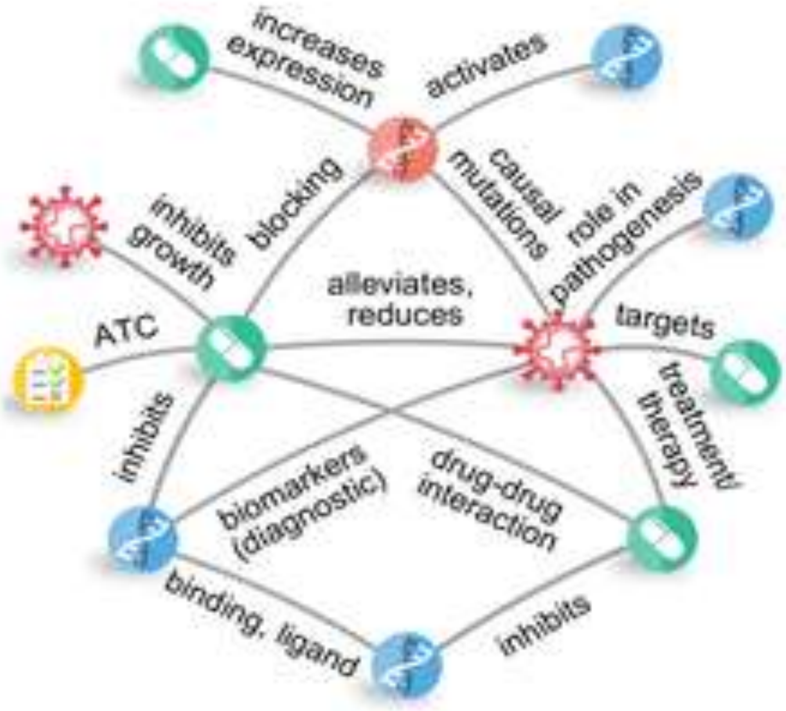
Target → Lead Discovery (2 years) → Lead Optimization (3 years) → Preclinical Study (2 years) → Clinical Trial

Lead Discovery 2 years	Lead Optimization 3 years	Preclinical Study 2 years	Clinical Trial
Screen millions of functional molecules; Found by serendipity: Penicillin	Modify the molecule to improve specific properties. e.g. toxicity, SA	In-vitro and in-vivo experiments; synthesis	Multiple Phases

Graphs in Biomedical Domains



Molecules



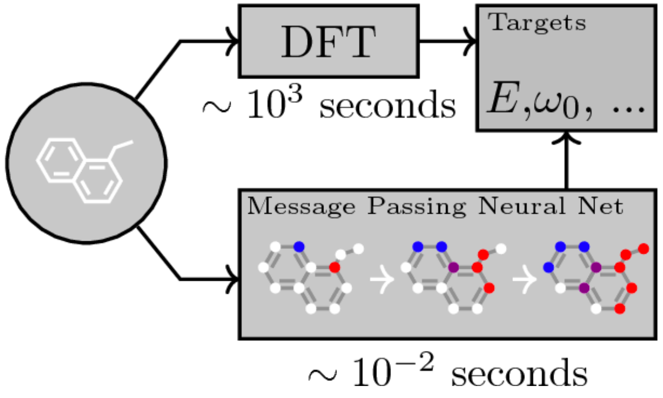
Biomedical Knowledge Graphs

Molecule Properties Prediction

- Predicting the properties of molecules is very important in many stages of drug discovery
  - Virtual screening
- Represent the whole molecule (graph) as a feature vector

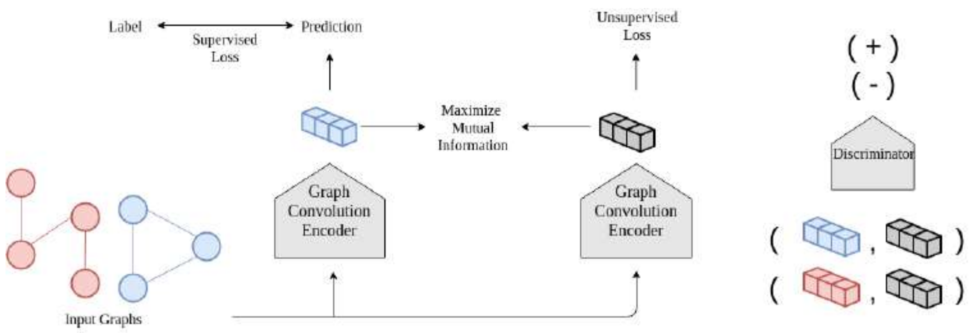
Unsupervised and Semi-supervised Learning for Molecular Graph Representation (Sun et al. ICLR 20)

- Most existing work on molecular representation are based on supervised learning with graph neural networks
  - Require a large number of labeled data
- However, the number of labeled data is very limited
- Leverage the unlabeled data!!



Supervised Methods (Gilmer et al. 17)

Gilmer et al. Neural Message Passing for Quantum Chemistry. ICML'17.



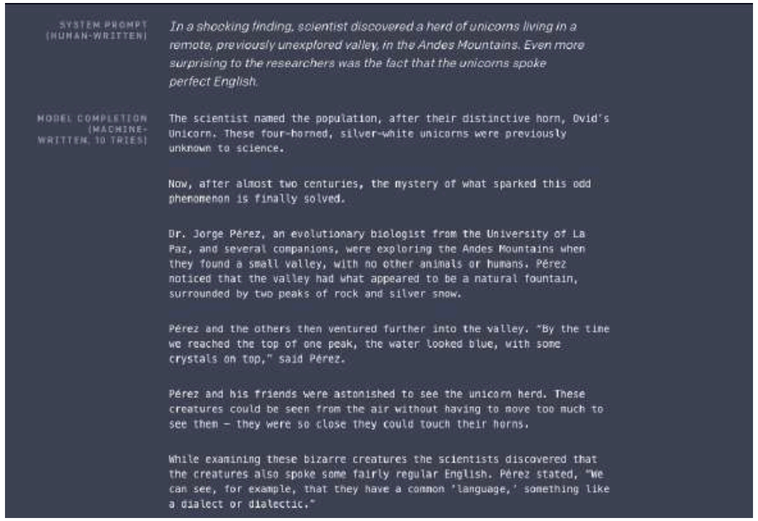
Unsupervised and semi-supervised methods (Sun et al. 19)

Sun et al. InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization. ICLR'20

De Novo Molecule Design and Optimization

- Deep generative models for data generation

Image generation
(by StyleGAN, From Internet)



SYSTEM PROMPT (HUMAN-WRITTEN)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

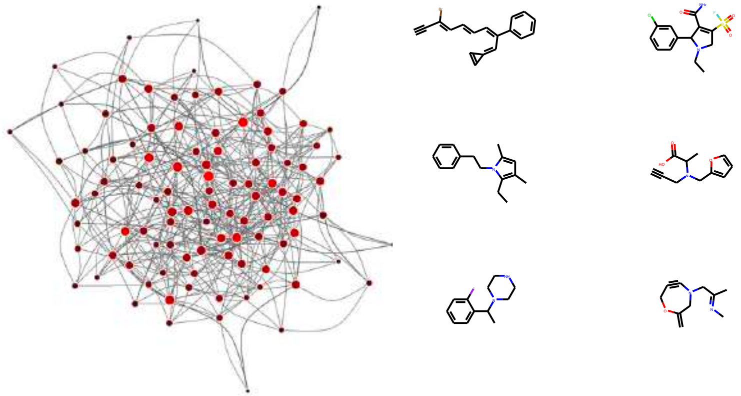
Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.

Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.

While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic."

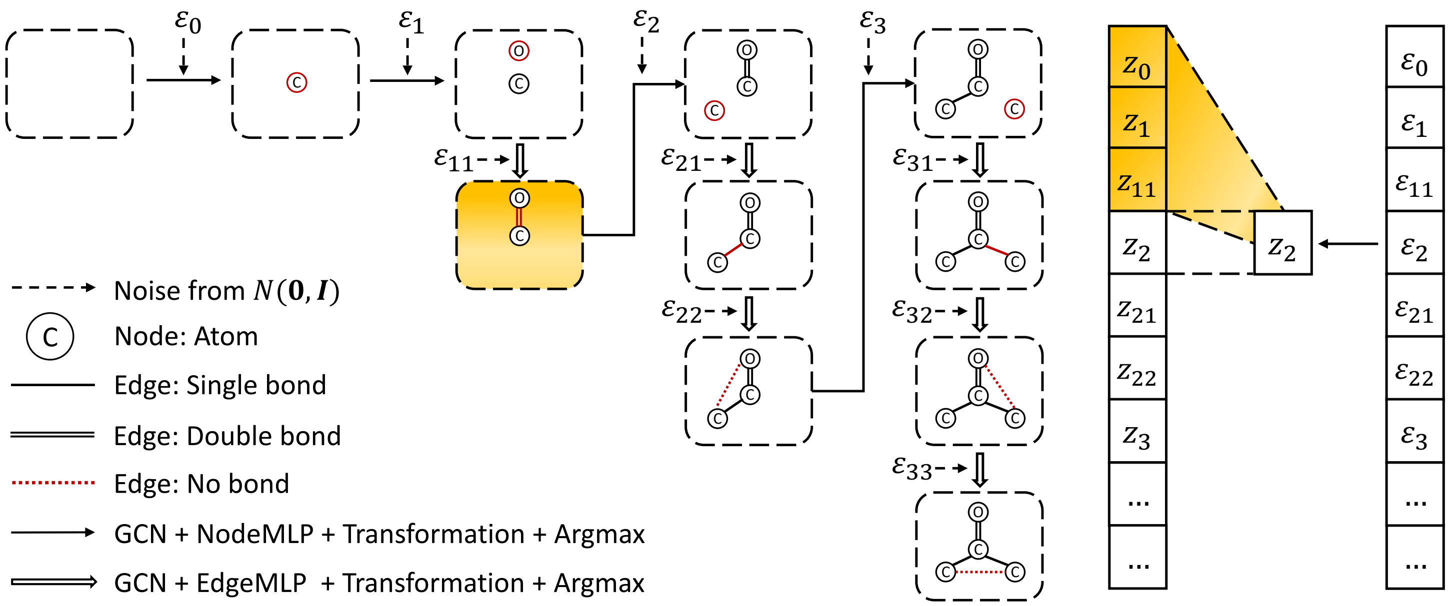
Text generated by by GPT-2,
Examples from Internet



Graphs?

GraphAF: a Flow-based Autoregressive Model for Molecular Graph Generation (Shi & Xu et al. ICLR'20)

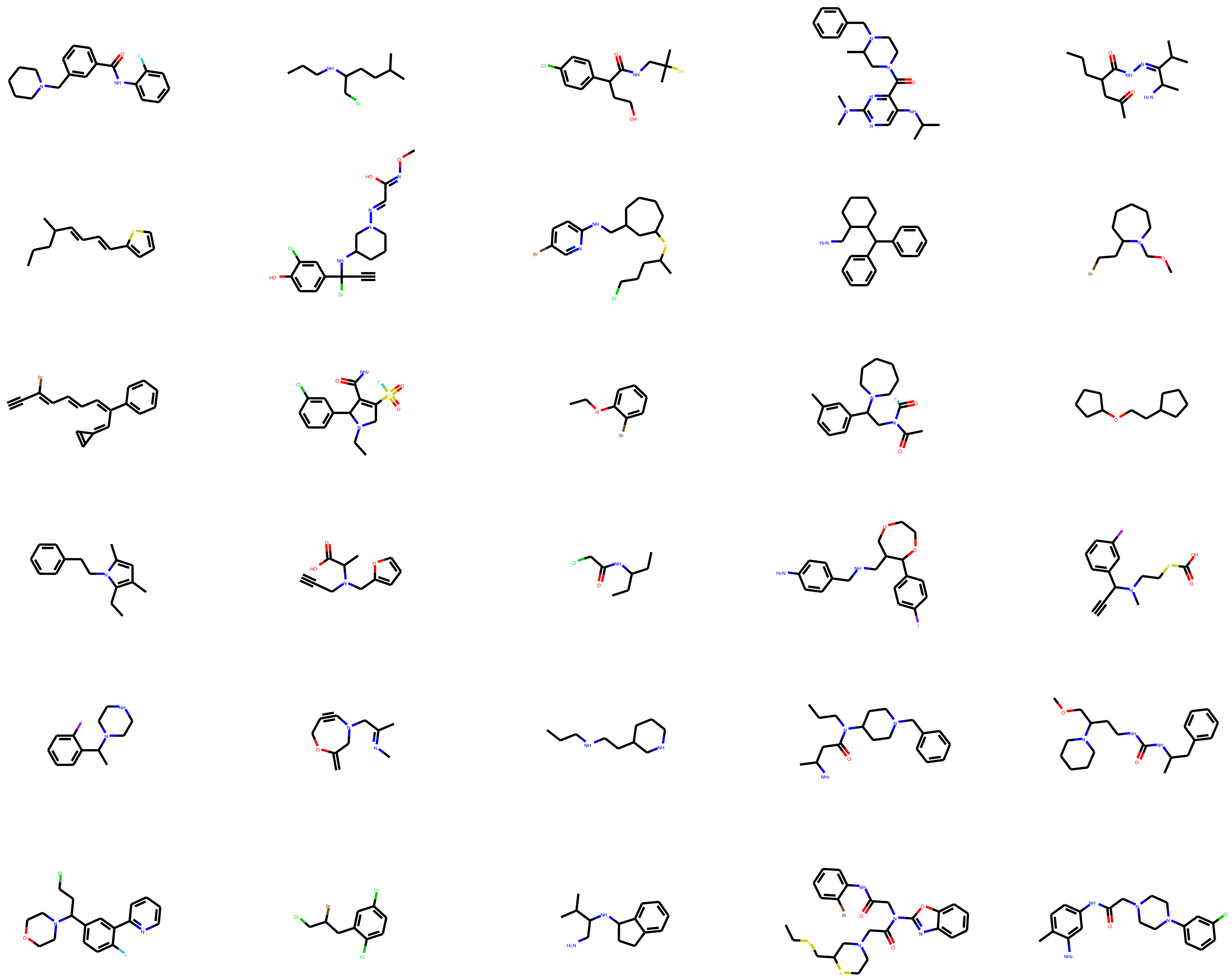
- Formulate graph generation as a sequential decision process
  - In each step, generate a new atom
  - Determine the bonds between the new atoms and existing atoms



(a) Sampling Framework (b) Autoregressive Flow

Molecule Generation

Method	Validity	Validity w/o check	Uniqueness	Novelty	Reconstruction
JT-VAE	100%	—	100%‡	100%‡	76.7%
GCPN	100%	20%†	99.97%‡	100%‡	—
MRNN	100%	65%	99.89%	100%	—
GraphNVP	42.60%	—	94.80%	100%	100%
GraphAF	100%	68%	99.10%	100%	100%



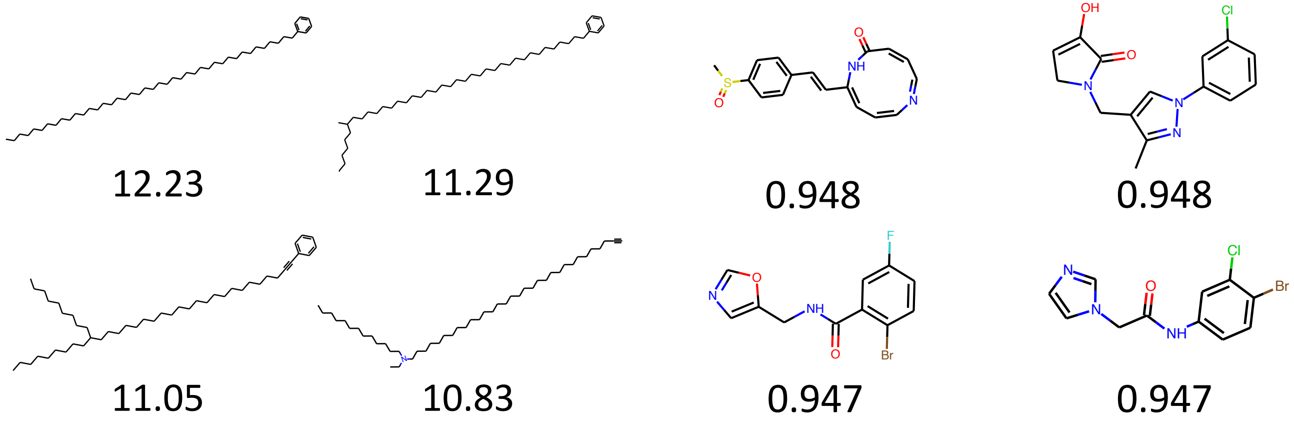
Goal-Directed Molecule Generation with Reinforcement Learning

- Fine tune the generation policy with reinforcement learning to optimize the properties of generated molecules
- **State**: current subgraph G_i
- **Action**: generating a new atom (i.e. $\mathrm{p}(X_i|G_i)$) or a new edge ($\mathrm{p}(A_{ij}|G_i, X_i, A_{i,1:j-1})$).
- **Reward Design**: the properties of molecules (final reward) and chemical validity (intermediate and final reward)

Molecule Optimization

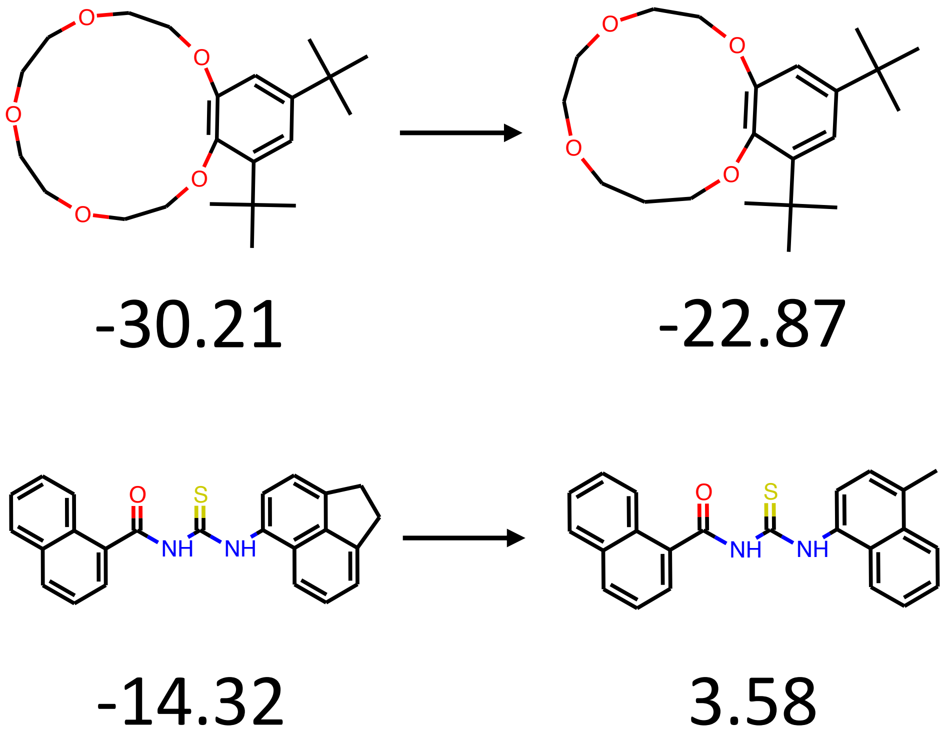
- Properties
  - Penalized logP
  - QED (druglikeness)

Method	Penalized logP				QED			
	1st	2nd	3rd	Validity	1st	2nd	3rd	Validity
ZINC (Dataset)	4.52	4.30	4.23	100.0%	0.948	0.948	0.948	100.0%
JT-VAE (Jin et al., 2018)	5.30	4.93	4.49	100.0%	0.925	0.911	0.910	100.0%
GCPN (You et al., 2018a)	7.98	7.85	7.80	100.0%	**0.948**	0.947	0.946	100.0%
MRNN[1] (Popova et al., 2019)	8.63	6.08	4.73	100.0%	0.844	0.796	0.736	100.0%
GraphAF	**12.23**	**11.29**	**11.05**	100.0%	**0.948**	**0.948**	**0.947**	100.0%



(a) Penalized logP optimization (b) QED optimization

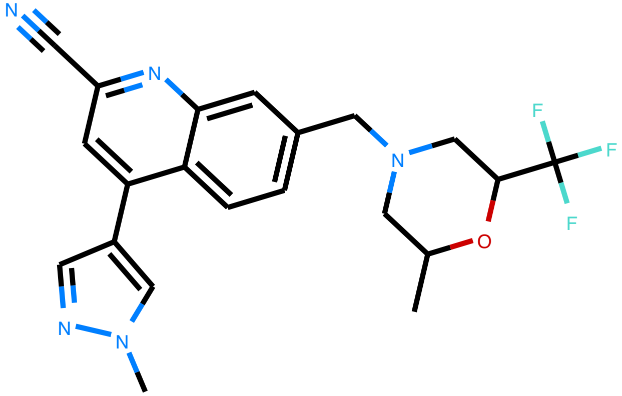
Constrained Optimization



(c) Constrained optimization

Retrosynthesis Prediction

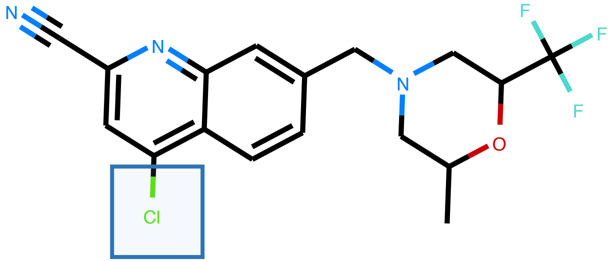
- Once a molecular structure is designed, how to synthesize it?
- Retrosynthesis planning/prediction
  - Identify a set of reactants to synthesize a target molecule



Product (Given)

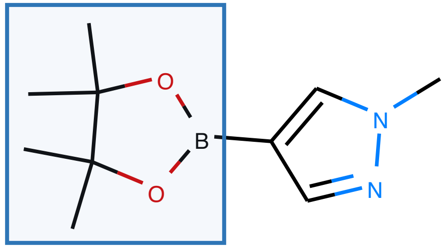
Predict Reactants

Reaction Type (optional)



Reactant A

→ …



Reactant B

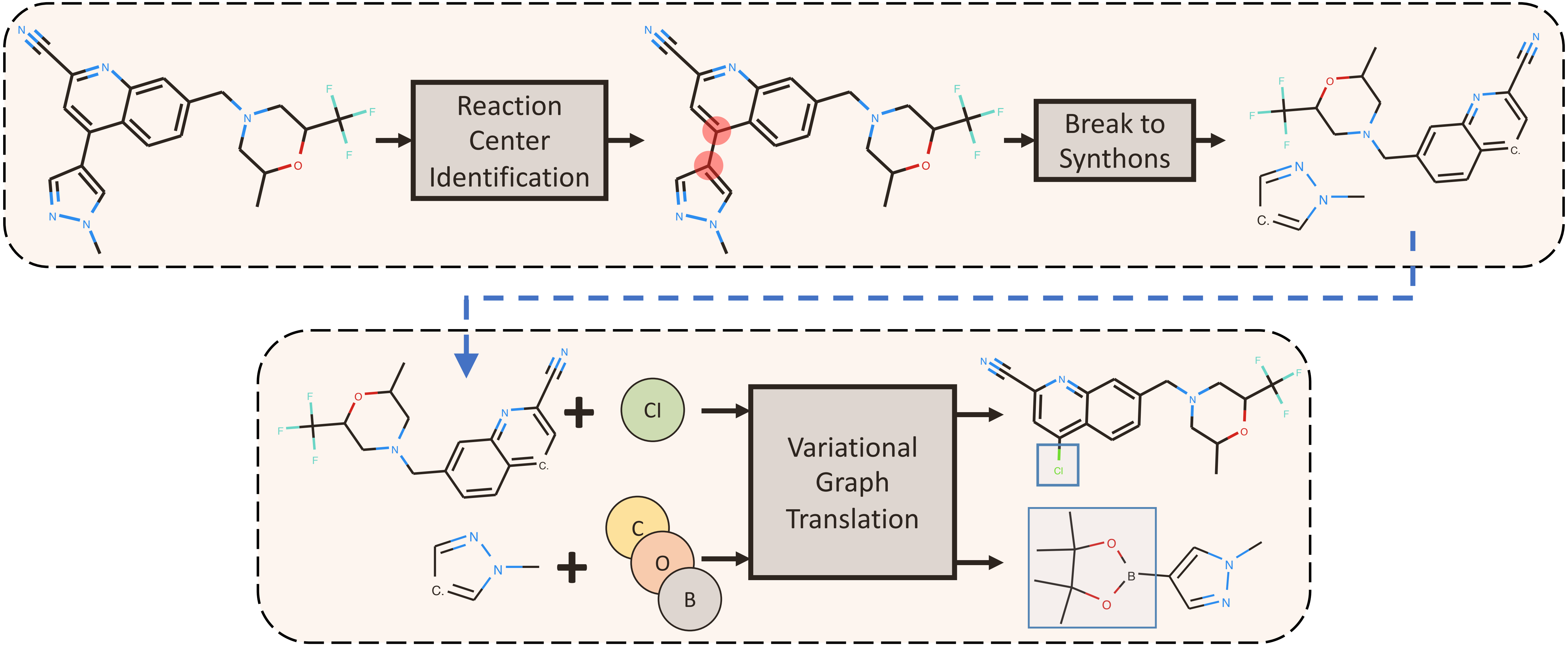
→ …

A Graph to Graphs Framework for Retrosynthesis Prediction (Shi et al. 2020)

- Each molecule is represented as a molecular graph
- Formulate the problem as a graph (product molecule) to a set of graphs (reactants)
- The whole framework are divided into two stages
  - Reaction center identification
  - Graph Translation

Chence Shi, Minkai Xu, Hongyu Guo, Ming Zhang and Jian Tang. **A Graph to Graphs Framework for Retrosynthesis Prediction**. In Submission, 2020.

The G2Gs Framework (Shi et al. 2020)

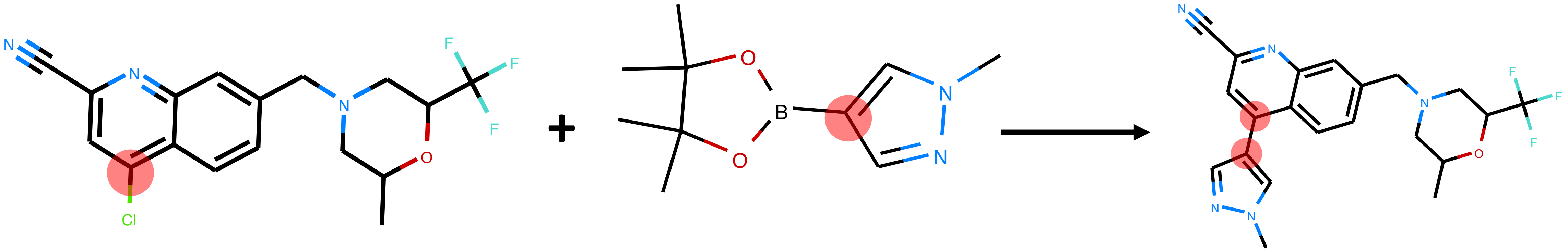


Shi et al., 2020, A Graph to Graphs Framework for Retrosynthesis Prediction

Reaction Center Prediction

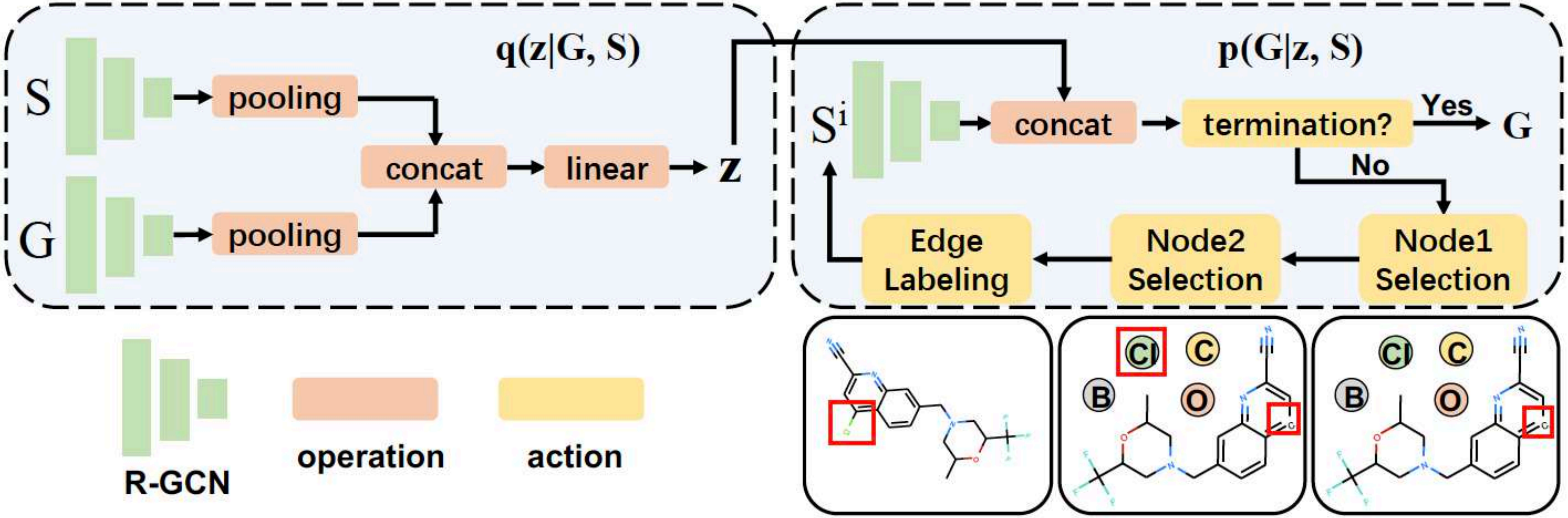
An atom pair *(i, j)* is a reaction center if:

- There is a bond between atom *i* and atom *j* in product
- There is no bond between atom *i* and atom *j* in reactants



Graph Translation

- Translate the incomplete synthon to the final reactant
- A variational graph to graph framework
  - A latent variable z is introduced to capture the uncertainty during translation



Experiments

- Experiment Setup
  - Benchmark data set USPTO-50K, containing 50k atom-mapped reactions
  - Evaluation metrics: top-k exact match (based on canonical SMILES) accuracy

Table 1. Top-k exact match accuracy when reaction class is given. Results of all baselines are directly taken from (Dai et al., 2019).

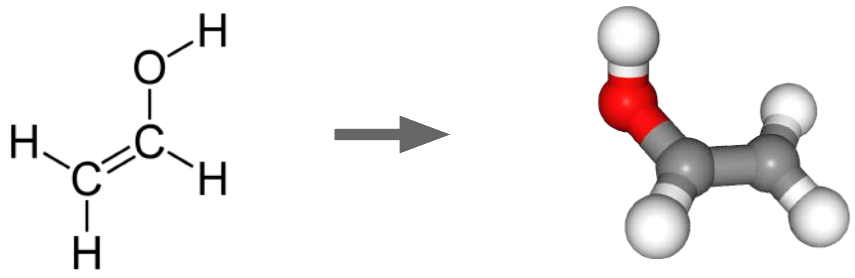
Methods	Top-k accuracy %			
	1	3	5	10
Template-free				
Seq2seq	37.4	52.4	57.0	61.7
G2Gs	**61.0**	**81.3**	**86.0**	**88.7**
Template-based				
Retrosim	52.9	73.8	81.2	88.1
Neuralsym	55.3	76.0	81.4	85.1
GLN	**64.2**	**79.1**	**85.2**	**90.0**

Table 2. Top-k exact match accuracy when reaction class is unknown. Results of all baselines are taken from (Dai et al., 2019).

Methods	Top-k accuracy %			
	1	3	5	10
Template-free				
Transformer	37.9	57.3	62.7	/
G2Gs	**48.9**	**67.6**	**72.5**	**75.5**
Template-based				
Retrosim	37.3	54.7	63.3	74.1
Neuralsym	44.4	65.3	72.4	78.9
GLN	**52.5**	**69.0**	**75.6**	**83.7**

Going Beyond 2D Graphs: 3D Structures

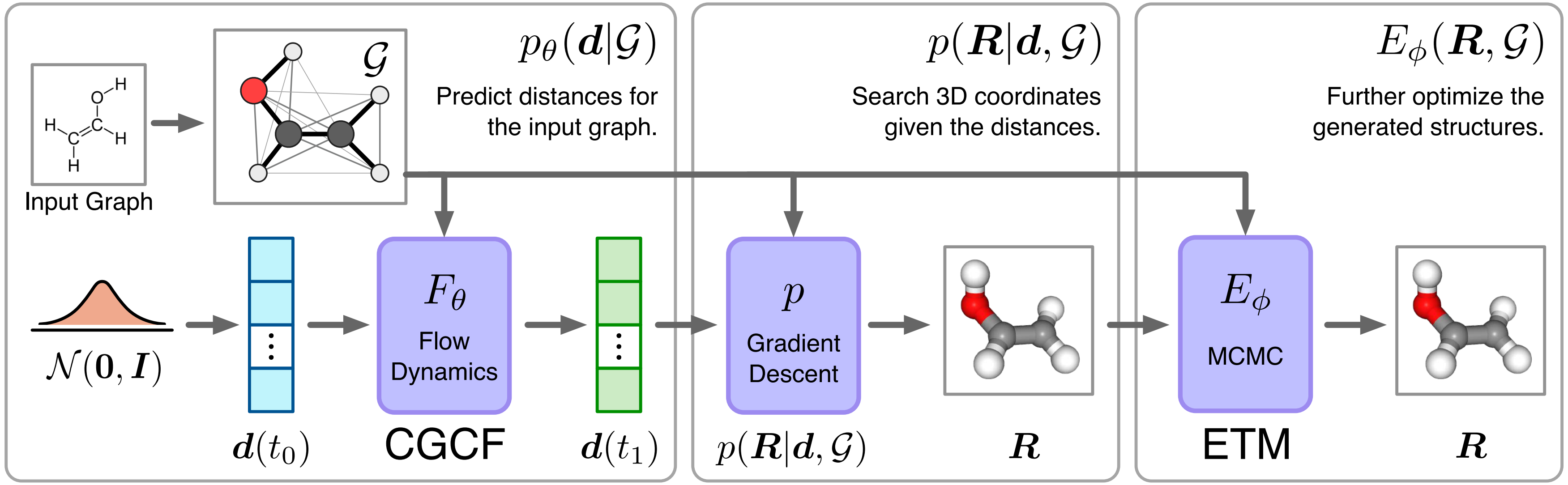
- The chemical/physical properties of molecules are determined by their 3D structures (a.k.a. conformations)
- Predicting stable 3D conformations given a molecular approach



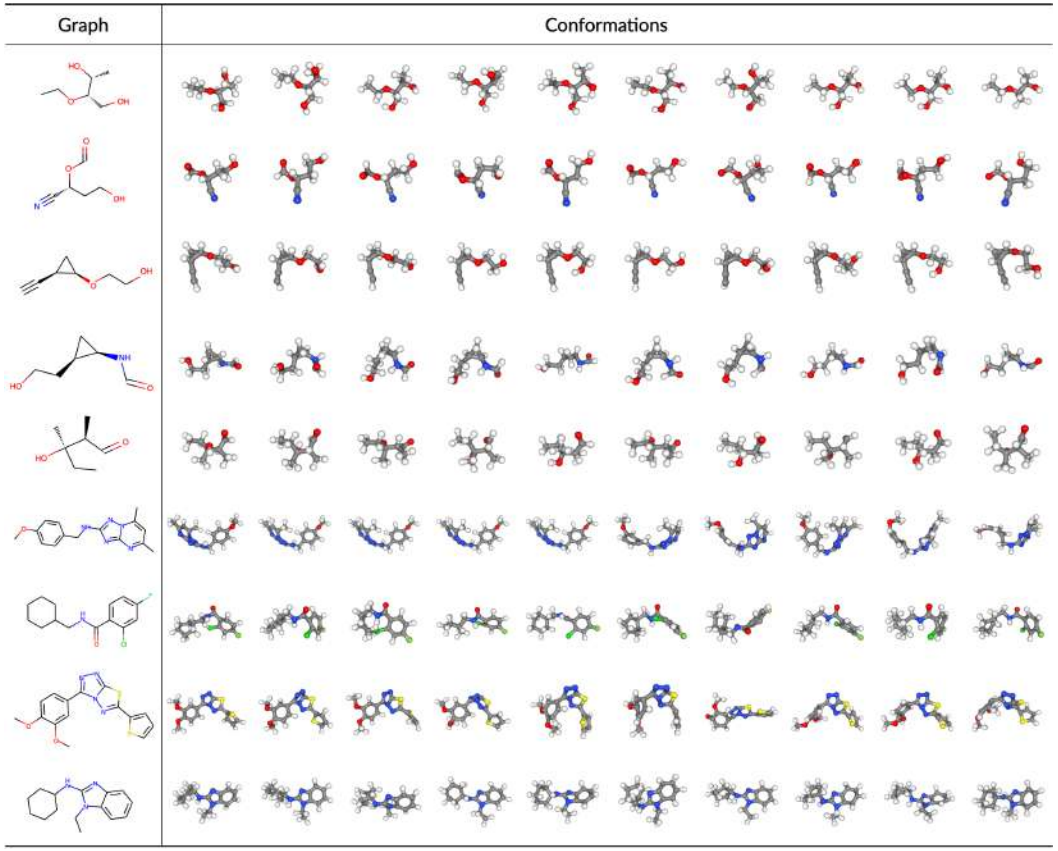
- Traditional approaches
  - Molecular dynamics, Markov chain Monte Carlo
  - Very computational Expensive

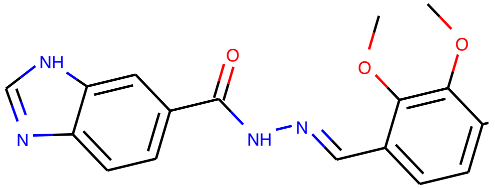
Our Solution: Data-Driven Approaches with Deep Generative Models

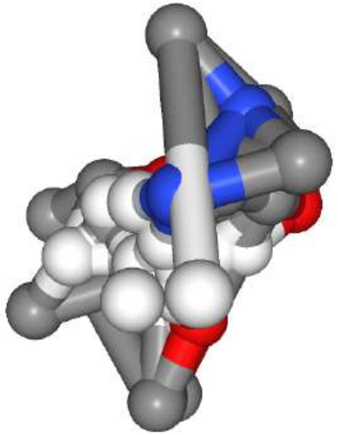
- Train a probabilistic model over conformations R given a molecular graph G, i.e., **P(R|G)**

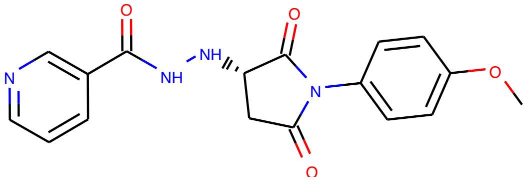


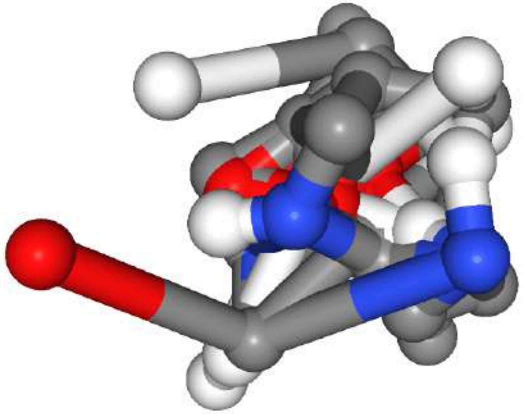
Examples











Medical Knowledge Graph Construction (Ongoing)

- >7M Entities, ~300M facts
  - Disease
  - Drug
  - Phenotype
  - Gene
  - Protein
  - Side effect
- Biomedical literature

PubMed

DrugBank

ctd

Comparative Toxicogenomics Database

STITCH

SIDER
Side Effect Resource

DISEASE ONTOLOGY

OMIM

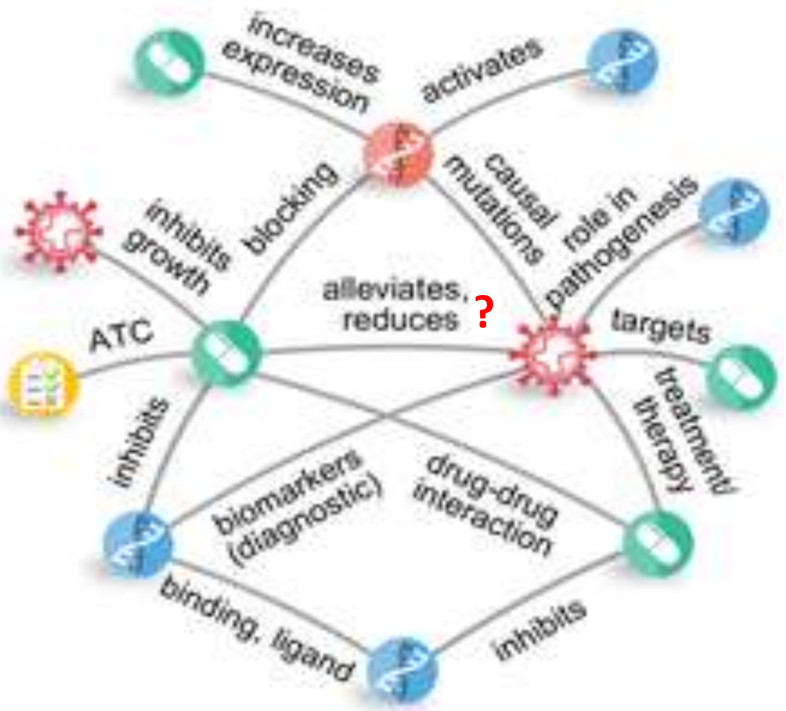
GENEONTOLOGY
Unifying Biology

e!Ensembl

BRENDA
The Comprehensive Enzyme Information System

Drug Repurposing with Biomedical Knowledge Graph (Ongoing)

- Represent each entity with a feature vector
  - Diseases, drugs, genes, …



Take Away

- Graph representation learning
  - A growing research topic in machine learning focusing on deep learning for graph-structured data
- Graph representation learning for drug discovery
  - Unsupervised and semi-supervised molecule properties prediction
  - De novo drug design and optimization
  - Retrosynthesis prediction
  - Drug repurposing based on medical knowledge graph
- A huge opportunity for biomedical applications
  - Looking forward to collaborating with you!

Thanks!

- **Current Students**
  - Meng Qu
  - Zhaocheng Zhu
  - Andreea Deac
  - Louis-Pascal Xhonneux
  - Shengchao Liu
  - Chence Shi
  - Minkai Xu
- **Collaborators and previous students**: Qiaozhu Mei, Yoshua Bengio, Jian-Yun Nie, Pietro Liò, Zhiyuan Liu, Ming Zhang, Jingzhou Liu, Zhiqing Sun, Fanyun Sun, Weiping Song, Mingzhe Wang, Shizhen Xu, Xiaozhi Wang, Tianyu Gao, Hongyu Guo, Jordan Hoffmann, Vikas Verma,….