Graph Representation Learning: Algorithms and Applications

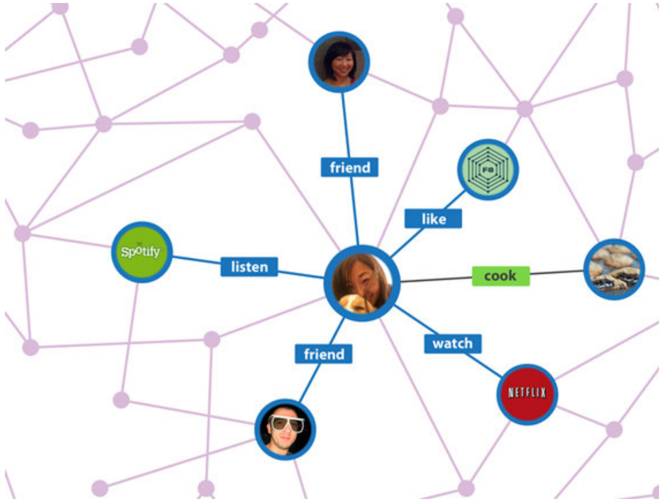
Jian Tang

HEC Montreal

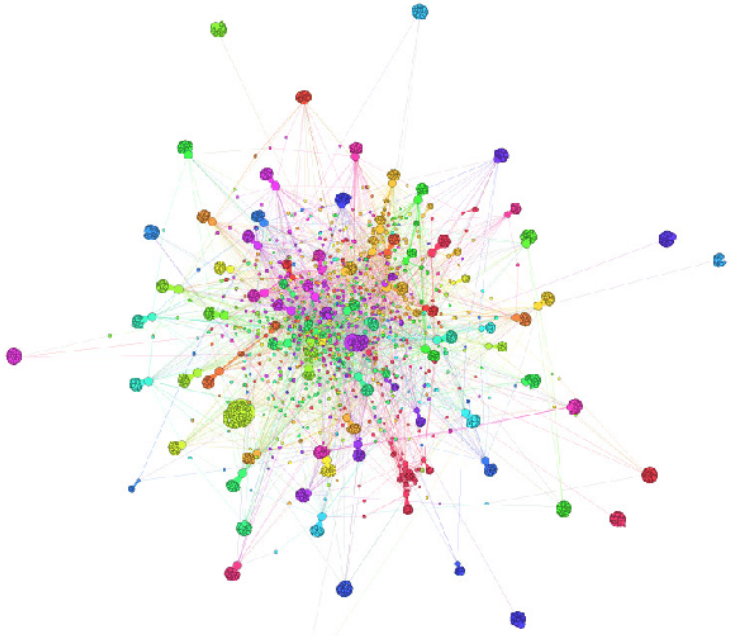
CIFAR AI Chair, Mila

Email: jian.tang@hec.ca

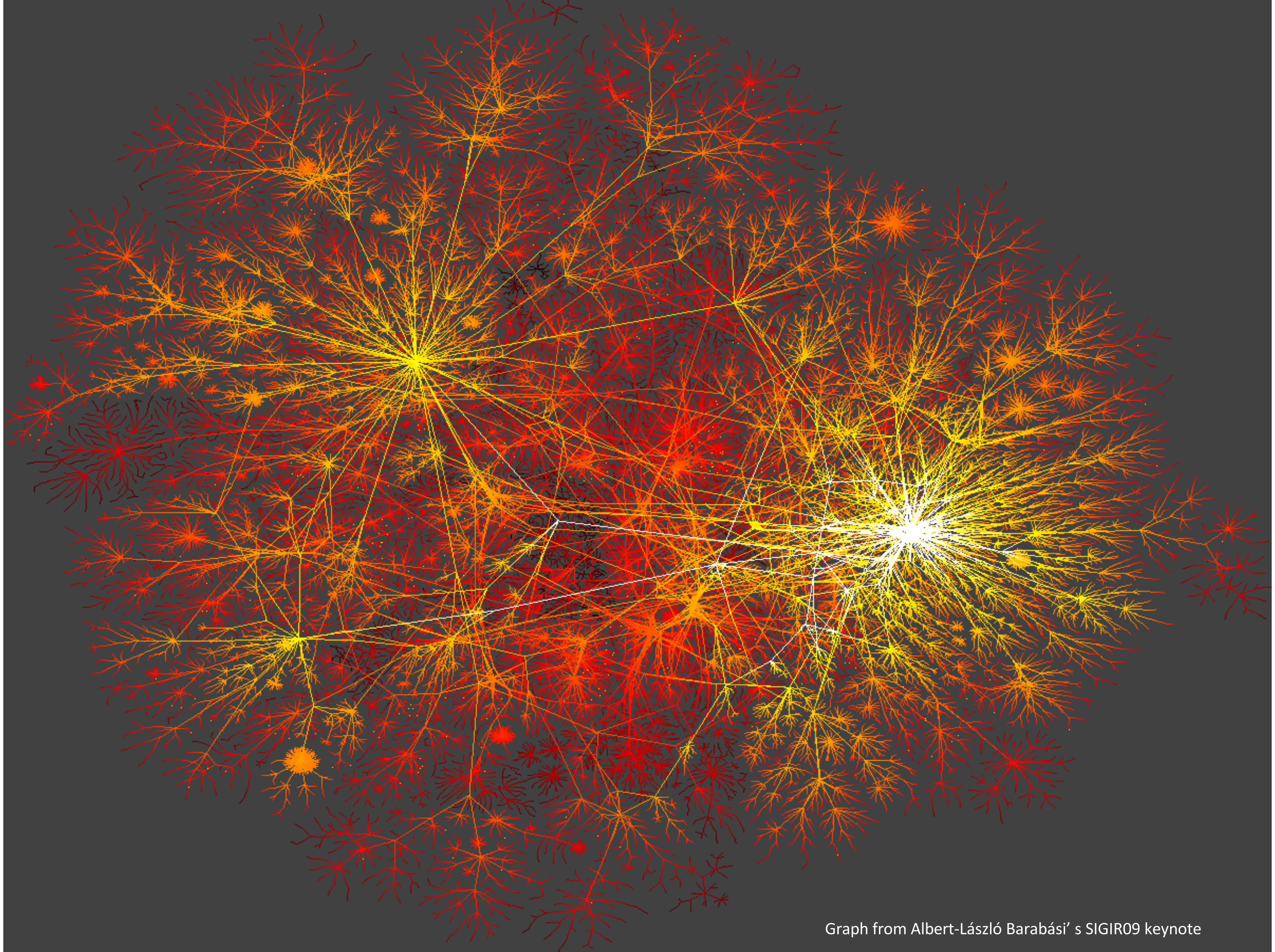
Social Networks



Facebook

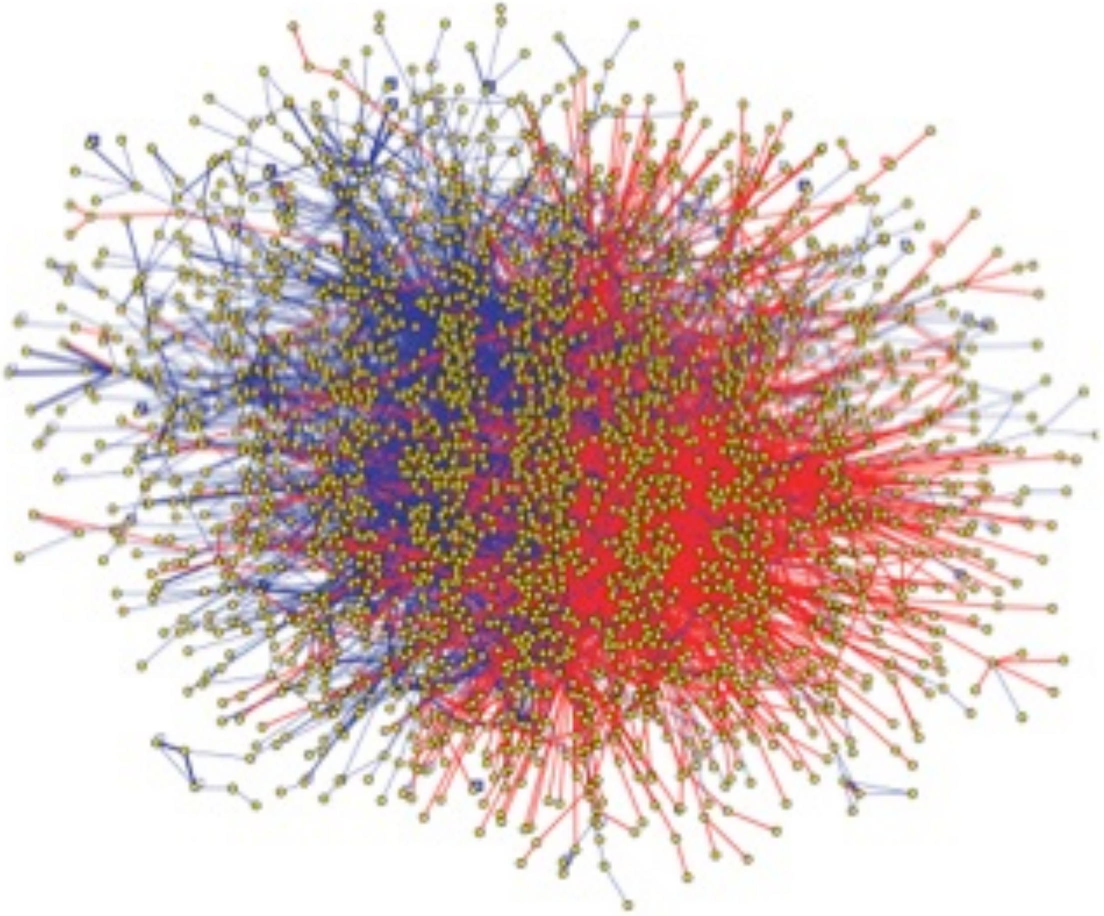


Twitter



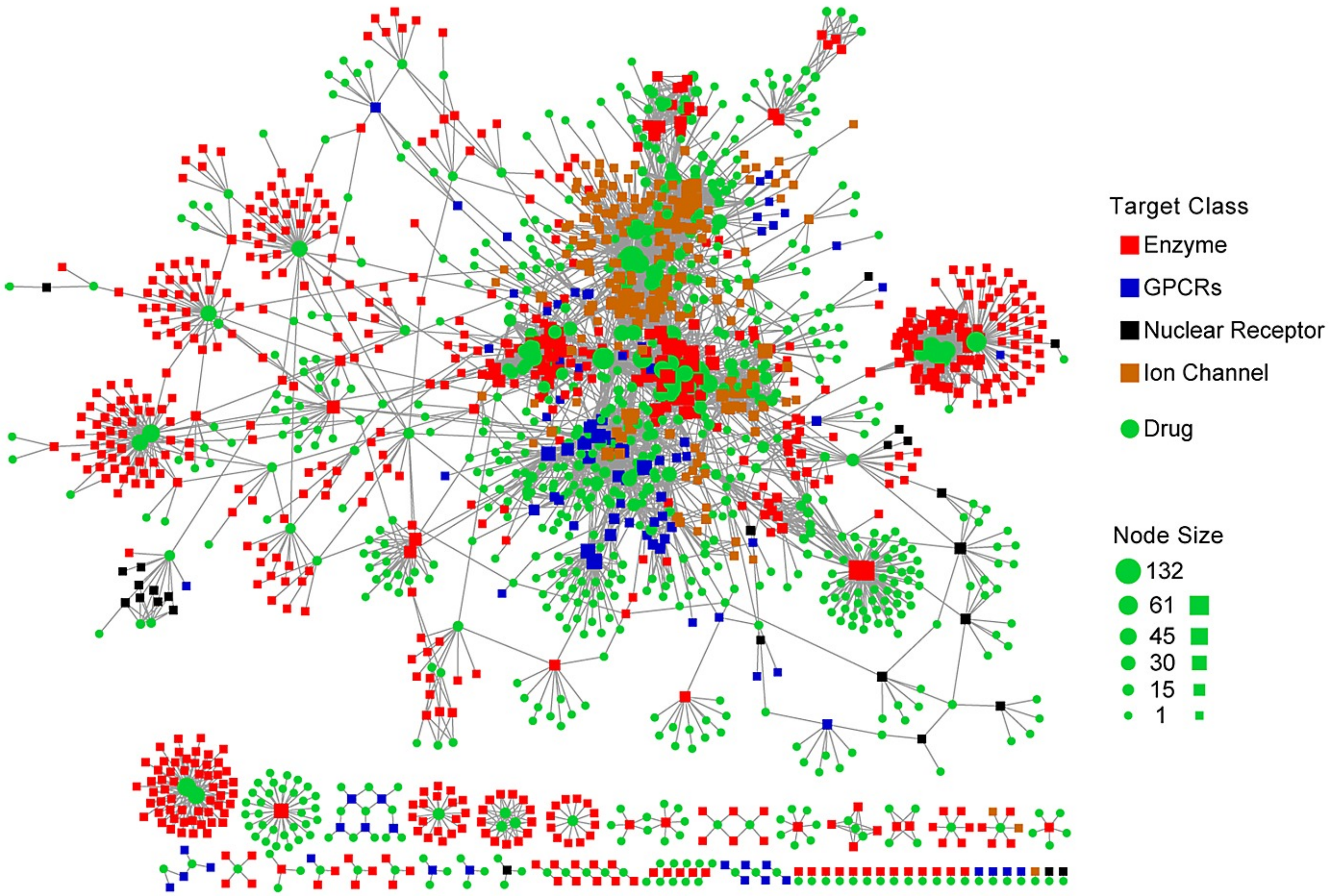
Graph from Albert-László Barabási’ s SIGIR09 keynote

Protein-Protein Interaction Graph



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

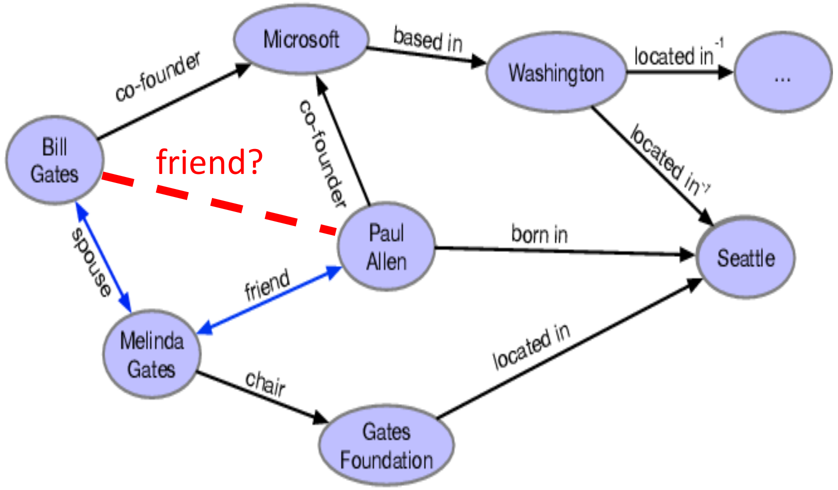
Drug-Protein Interaction Graph



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

Knowledge Graphs

- Multiple types of edges
  - Each corresponds to a relation type
- A set of facts, each of which is represented as a triplet
  - (Bill_Gates, CoFounder, Microsoft)



Google Knowledge Graph

Microsoft SATORI

Browse the Knowledge Base!

NELL: Never-Ending Language Learning

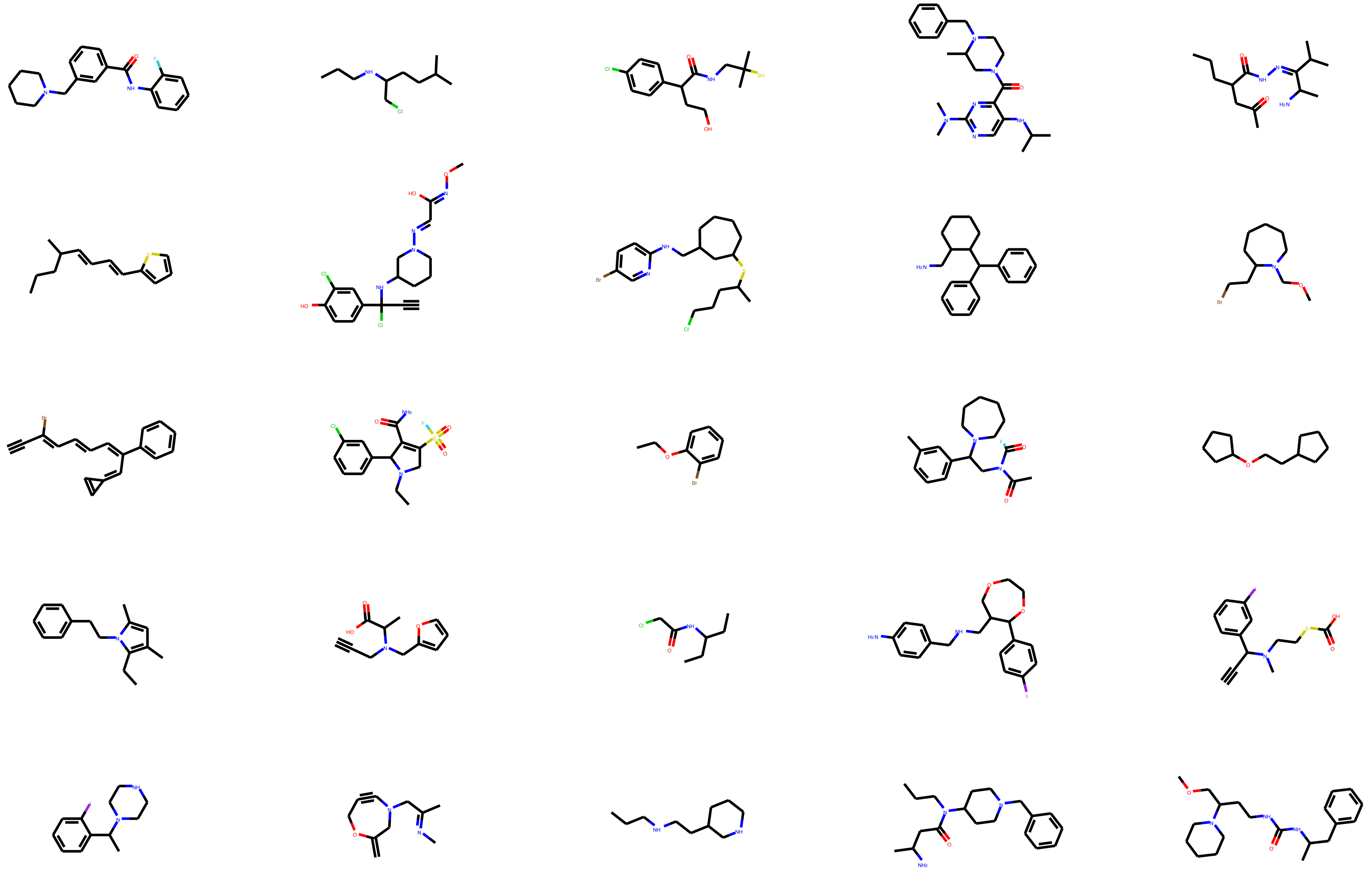
Freebase

DBpedia

yago select knowledge

OpenIE
(Reverb, OLLIE)

Molecules

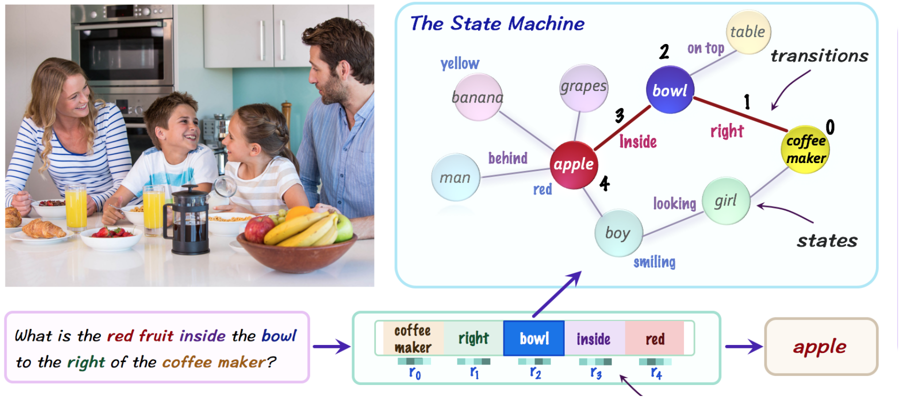


Various Applications on Graphs

- Predicting whether a user is a democratic or republican in Facebook?
- Recommending friends in social networks
- Predicting missing facts on knowledge 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!!**

Graphs as Bridges Between System I and System II Reasoning

- Existing deep learning systems are mainly good at **perception**
  - E.g., recognize the objects in images
  - **System I Reasoning**
- Real-world problems are very complex
  - Need to understand the relationships between different facts or high-level semantic variables
  - **System II Reasoning**
- Using graphs to capture the relationship between high-level semantic variables



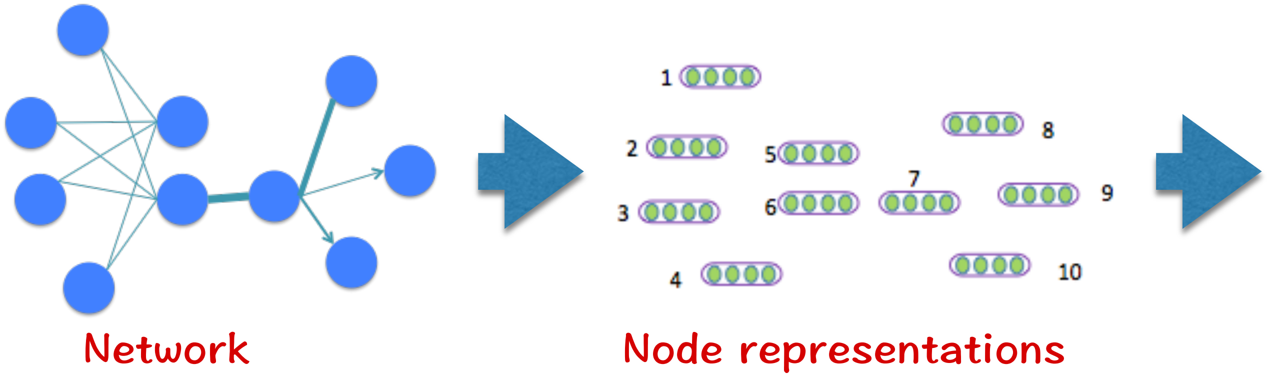
Outline

- Unsupervised Graph Representation Learning
- Relational Reasoning with Graph Representation Learning
- Graph Representation Learning for Drug Discovery

Outline

- **Unsupervised Graph Representation Learning**
- Relational Reasoning with Graph Representation Learning
- Graph Representation Learning for Drug Discovery

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

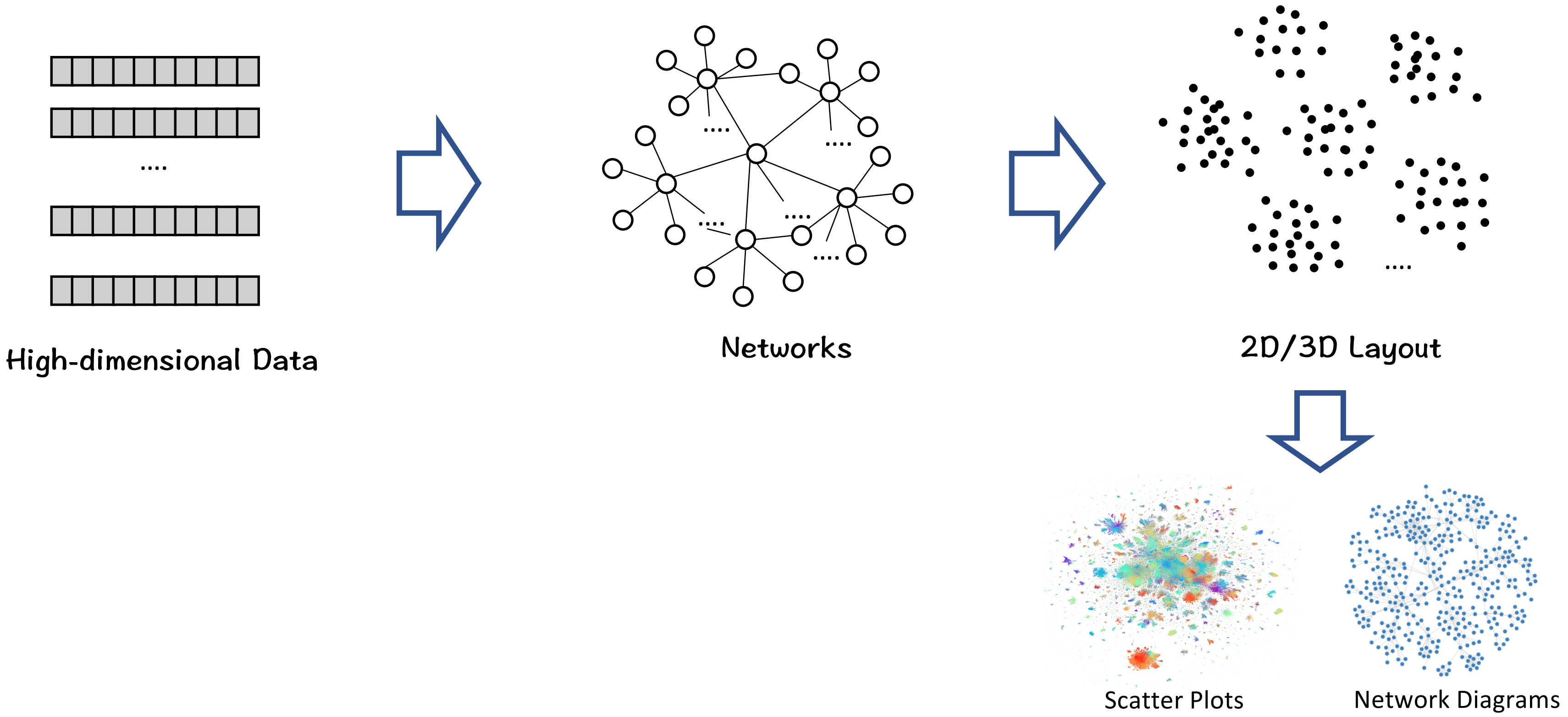


- Node Classification
- Node Clustering
- Link Prediction
- Recommendation
- …

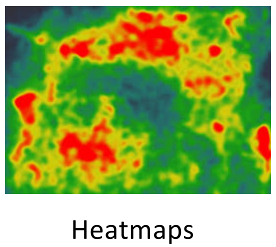
- E.g., Facebook social network -> user representations (features)-> friend recommendation

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

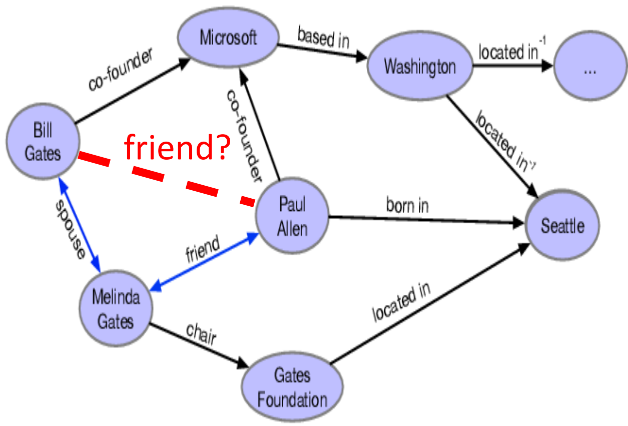
Visualizing Graphs and High-Dimensional Data (LargeVis, Tang et al. 2016)



Jian Tang, Jingzhou Liu, Ming Zhang and Qiaozhu Mei. [Visualizing Large-scale and High-dimensional Data](). WWW'16.



Knowledge Graph Embedding



- Knowledge graphs: a set of facts represented as triplets
  - (head entity, relation, tail entity) or (h,r,t)
- Knowledge graph embedding: learning low-dimensional representations of **entities** and **relations**
- A fundamental task on knowledge graphs**: predicting missing links**
- Key Idea: implicitely model and infer **logical rules**

Wife (X,Y) => Husband (Y,X)

Father (X,Y) ⋀ Father (Y,Z) => GrandFather (X,Z)

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

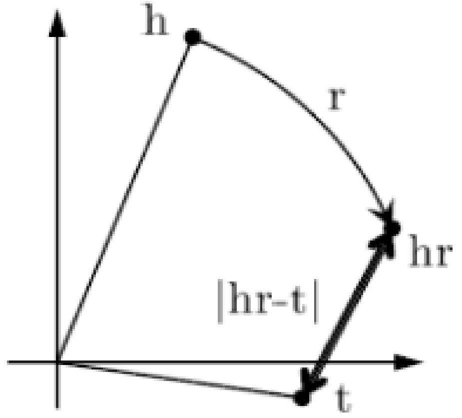
RotatE: Relation as Elementwise Rotation in Complex Space (Sun et al. 2019)

- Representing entities in complex vector space, i.e., $\mathbf{h}, \mathbf{t} \in \mathbb{C}^k$
- Each relation **r** as an element-wise rotation from the head entity **h** to the tail entity **t,** i.e.,

$$\mathrm{t_i} = \mathrm{h_i r_i}, \quad \text{where } |r_i|=1$$

- Modeling different logical rules
    - Symmetric
    - Inverse
    - Composition

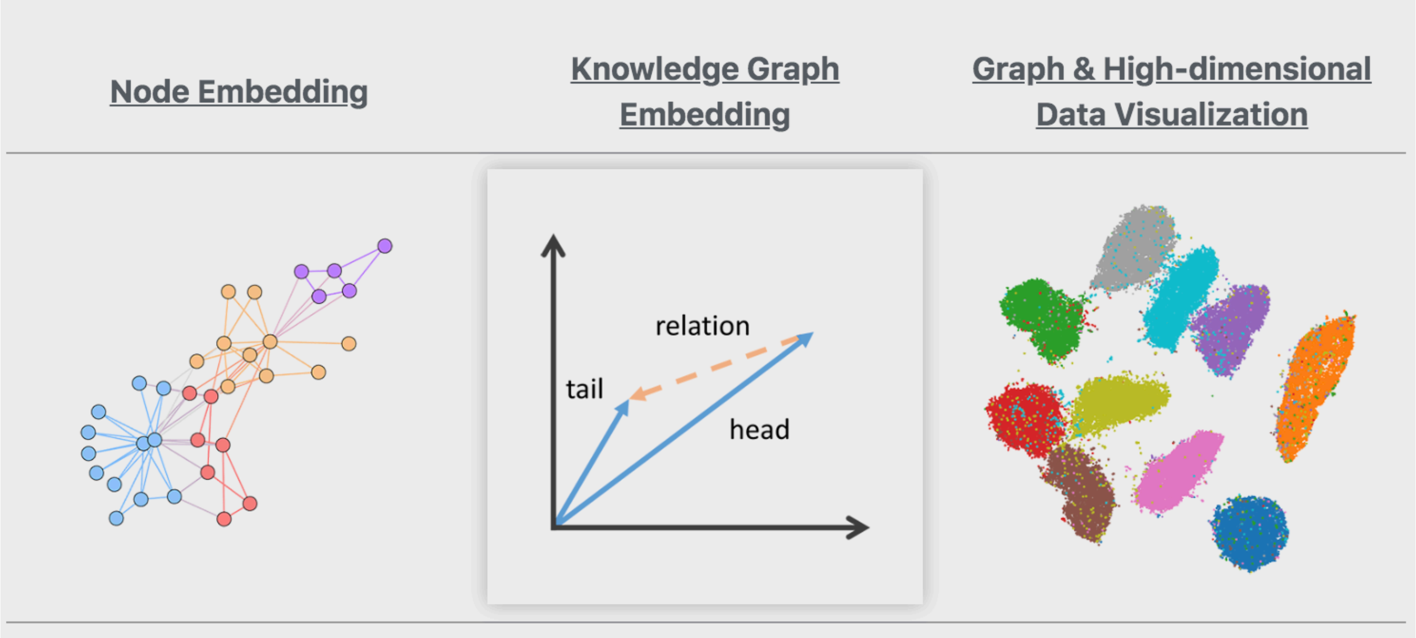
(h, r, t)



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)

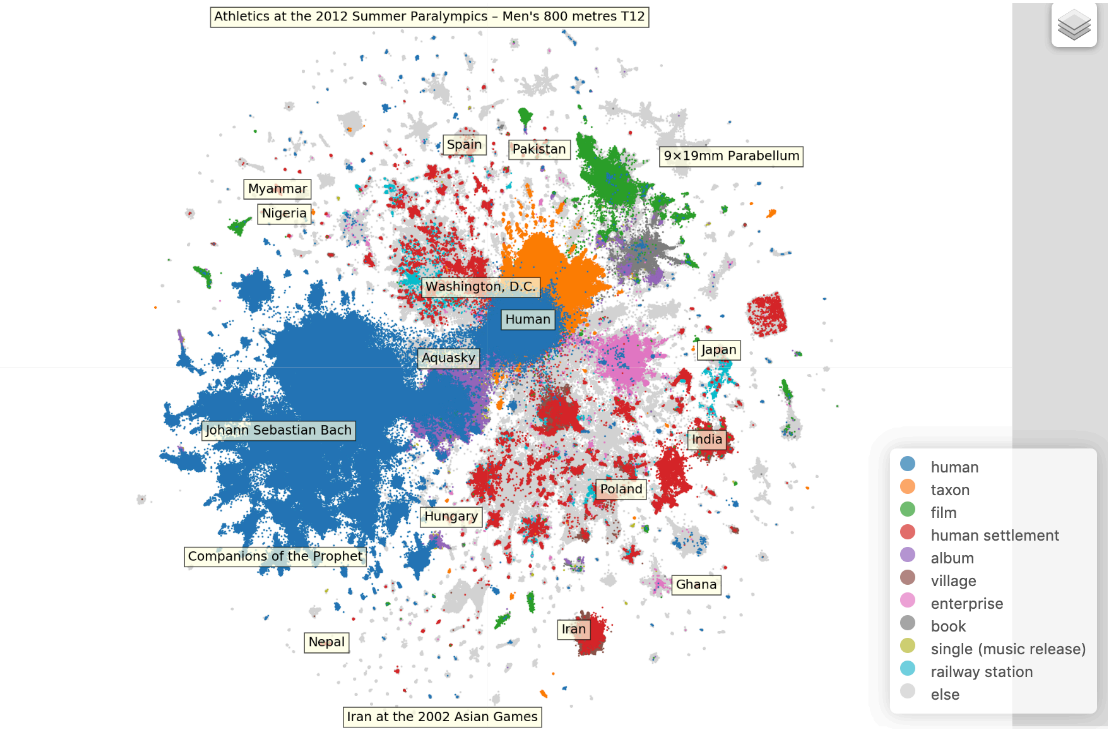
- A system specifically designed for learning graph embeddings with GPUs
- Super efficient!! Take only one minute for learning node representations of a graph with one-million nodes
- https://graphvite.io



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

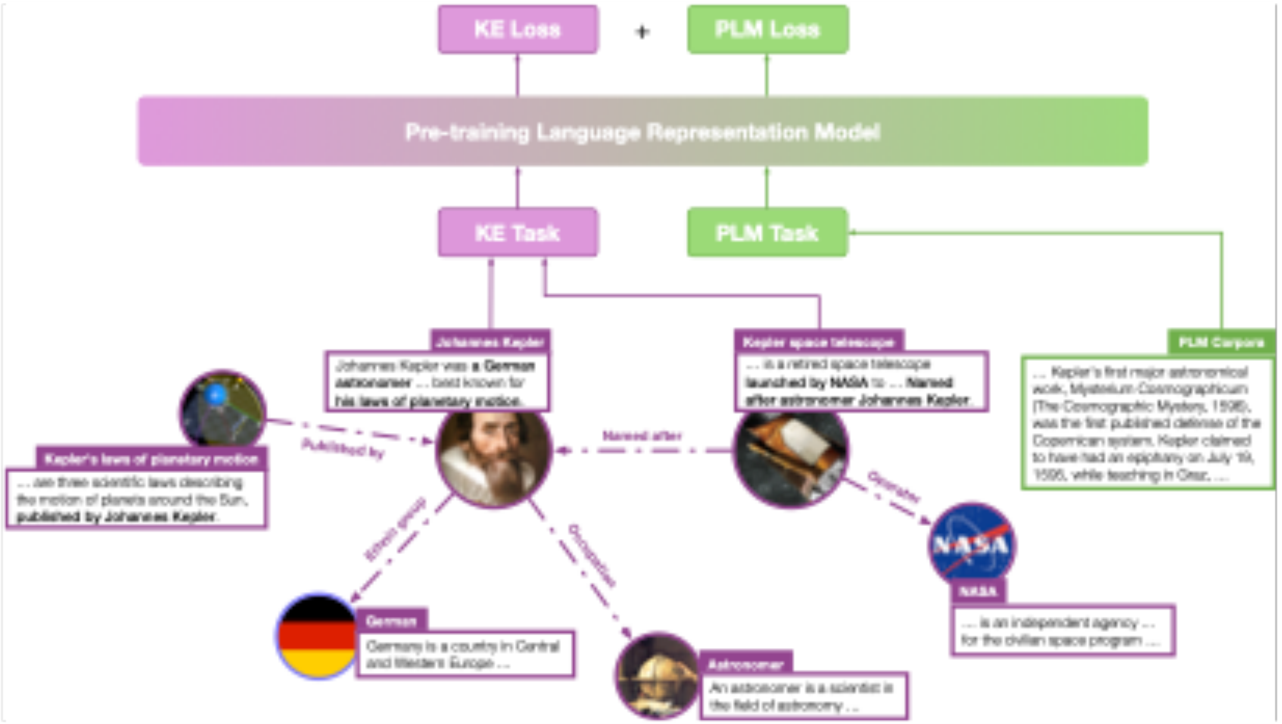
Wikidata5M: a Large-scale Knowledge Graph

- Contains 5 million entities and also the the descriptions of entities
- Pretrained knowledge graph embeddings with Wikidata5M: https://graphvite.io/pretrained_models



Joint Pretrained Language Representations and Knowledge Graph Embeddings (Wang et al. 2019)

- Entities are encoded with Language models on the entity descriptions
  - Knowledge graph embedding and language representation space are aligned.



$$\mathcal{L} = \mathcal{L}_{KE} + \mathcal{L}_{LM}$$

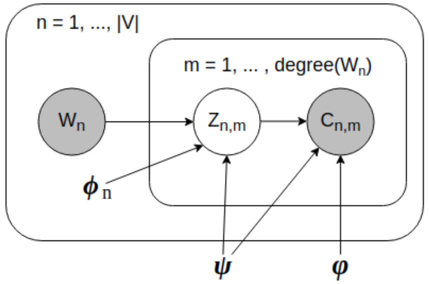
Knowledge graph embedding loss

Language model loss

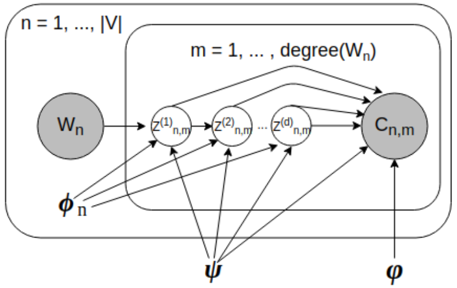
Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhiyuan Liu, Juanzi Li, Jian Tang. **KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation**. arXiv:1911.06136.

vGraph: Combining Community Detection and Node Representation Learning (Sun et al. 2019)

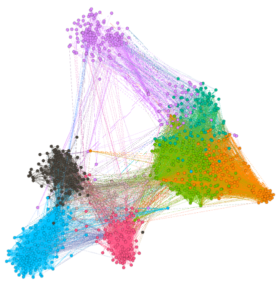
- Two classical tasks on graphs:
  - Community detection
  - Node representation Learning
- vGraph: a generative model for joint community detection and node representation learning

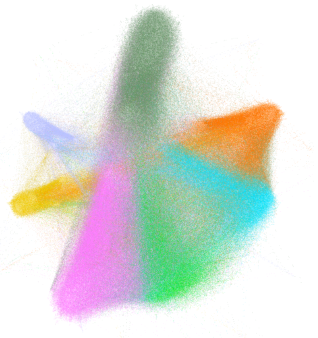


(a) vGraph



(b) Hierarchical vGraph



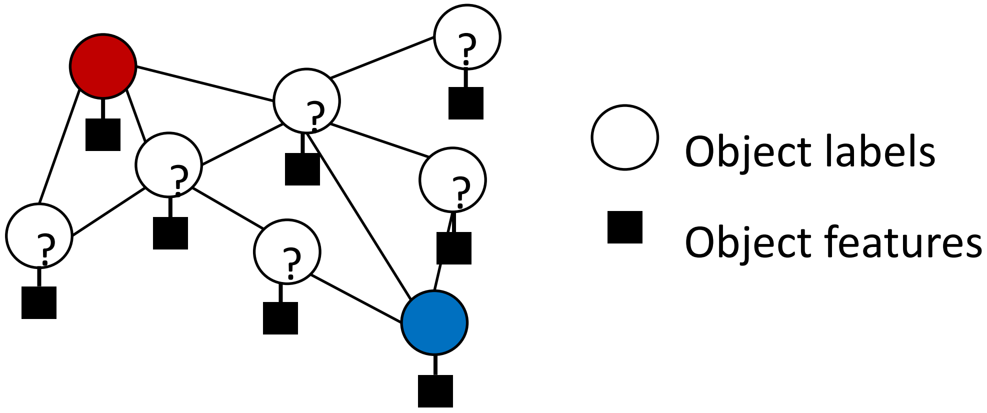


Fan-Yun Sun, Meng Qu, Jordan Hoffmann, Chin-Wei Huang, Jian Tang. vGraph: **A Generative Model for Joint Community Detection and Node Representation Learning**. NeurIPS'19.

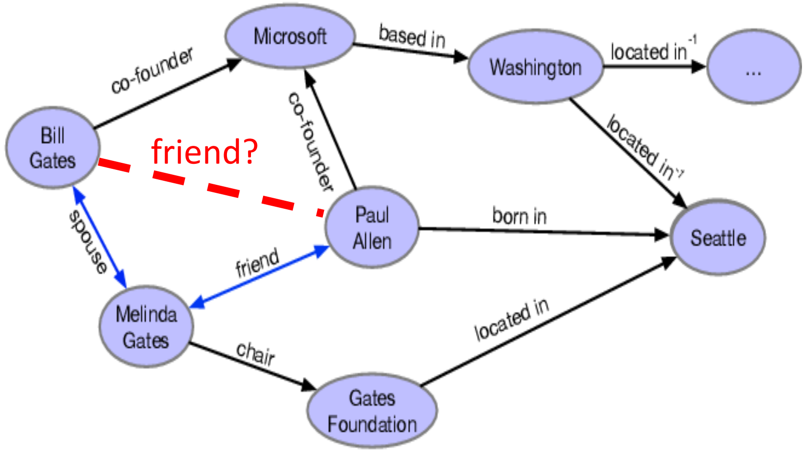
Outline

- Unsupervised Graph Representation Learning
- **Relational Reasoning with Graph Representation Learning**
- Graph Representation Learning for Drug Discovery

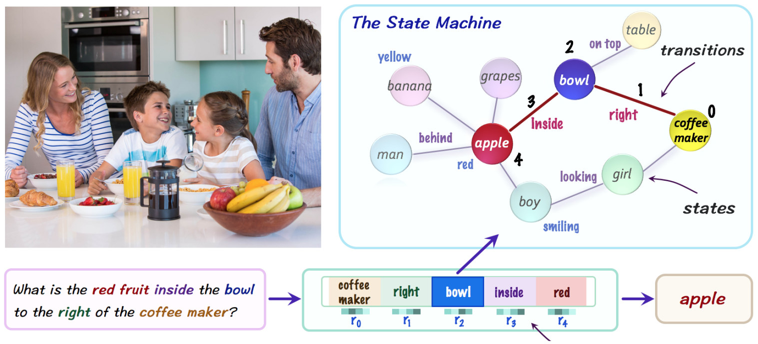
Relational Prediction and Reasoning



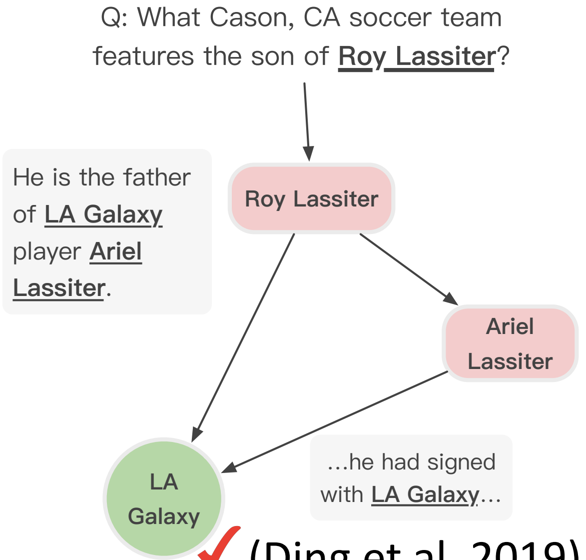
Node classification



Reasoning on knowledge graphs



Visual relational reasoning
(Hudson et al. 2019)

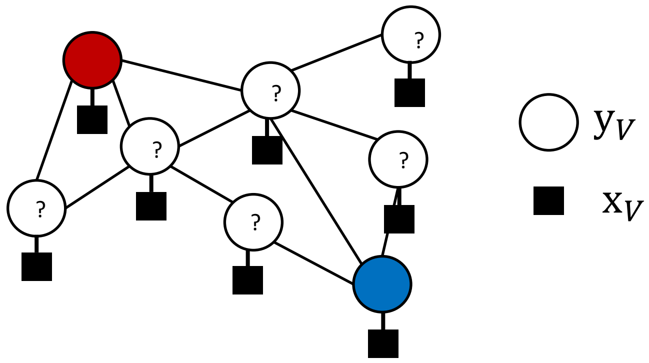


Multi-hop Question answering

(Ding et al. 2019)

Statistical Relational Learning

- Probabilistic graphical models for relational data
  - Markov Networks (Ross et al. 1980)
  - Conditional Random Fields (Lafferty et al. 2001)
  - Markov Logic Networks (Richardson and Domingos, 2006)
- Pros:
  - Captures uncertainty and domain knowledge
  - Collective inference
- Cons:
  - Limited model capacity
  - Inference is difficult

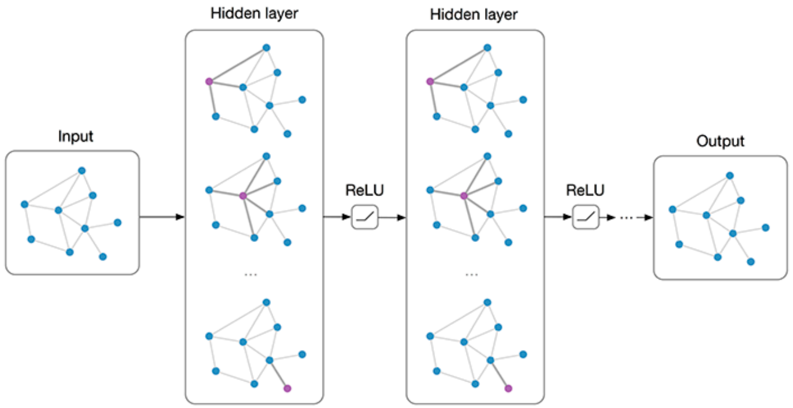


$$p(\mathbf{y}_V|\mathbf{x}_V) = \frac{1}{Z(\mathbf{x}_V)} \prod_{(i,j)\in E} \psi_{i,j}(\mathbf{y}_i, \mathbf{y}_j, \mathbf{x}_V)$$

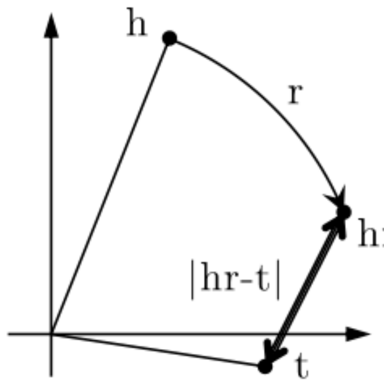
Figure: Conditional Random Fields

Graph Representation Learning

- Graph Neural Networks
  - Graph convolutional Networks (Kipf et al. 2016)
  - Graph attention networks (Veličković et al. 2017)
  - Neural message passing (Gilmer et al. 2017)
- Node Embedding and Knowledge Graph Embedding
  - DeepWalk, LINE, TransE, RotatE (Sun et al. 2019)
- Pros:
  - Learning effective node (and relation) representations
  - High model capacity
- Cons
  - Independent inference



Graph convolutional Networks
(Kipf et al. 2016)

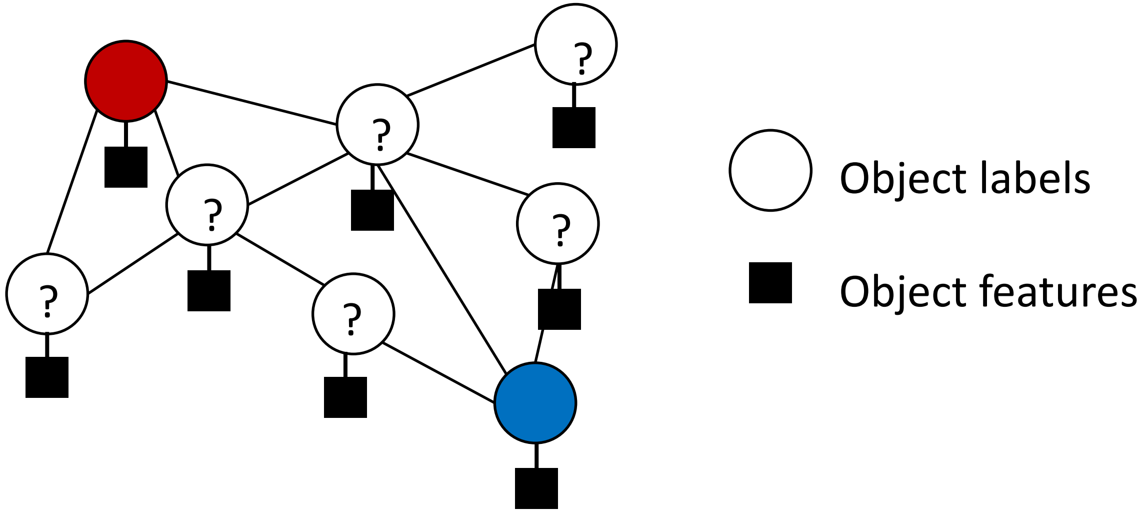


RotatE
(Sun et al. 2019)

Can we combine the two learning frameworks?

Semi-supervised Object Classification

- Given G= (V, E, $\mathbf{x}_V$)
  - $V = V_L \cup V_U$: objects/nodes
  - E : edges
  - $\mathbf{x}_V$: object features



- Give some labeled objects V_L, we want to infer the labels of the rest of objects V_U
- Many other tasks on graphs can be formulated as object classification
  - E.g., link classification

GMNN: Graph Markov Neural Networks (Qu, Bengio, and Tang, ICML'19)

- Towards combining statistical relational learning and graph neural networks
  - Combining CRFs + GNNs
- Learning effective node representations for predicting the node labels
- Modeling the label dependencies of nodes
- State-of-the-art performance
  - semi-supervised node classification
  - unsupervised node representation
  - link classification

Meng Qu, Yoshua Bengio, Jian Tang. "GMNN: Graph Markov Neural Networks". In ICML'19.

GMNN: Graph Markov Neural Networks

- Model the joint distribution of object labels $\mathbf{y}_V$ conditioned on object attributes $\mathbf{x}_V$, i.e., $\mathrm{p}_\phi(\mathbf{y}_V|\mathbf{x}_V)$, with a conditional random field
- Learning the model parameters ϕ by maximizing the lower-bound of log-likelihood of the observed data, $\log \mathrm{p}_\phi(\mathbf{y}_L|\mathbf{x}_V)$

$$\log p_\phi(\mathbf{y}_L|\mathbf{x}_V) \geq$$
$$\mathbb{E}_{q_\theta(\mathbf{y}_U|\mathbf{x}_V)}[\log p_\phi(\mathbf{y}_L, \mathbf{y}_U|\mathbf{x}_V) - \log q_\theta(\mathbf{y}_U|\mathbf{x}_V)]$$

Optimization with Pseudolikelihood Variational-EM

- E-step: fix p_ϕ and update the variational distribution $q_\theta(\mathbf{y}_U|\mathbf{x}_V)$ to approximate the true posterior distribution $p_\phi(\mathbf{y}_U|\mathbf{y}_L,\mathbf{x}_V)$.
- M-step: fix q_θ and update p_ϕ to maximize the lower bound

$$\ell(\phi) = \mathbb{E}_{q_\theta(\mathbf{y}_U|\mathbf{x}_V)}[\log p_\phi(\mathbf{y}_L,\mathbf{y}_U|\mathbf{x}_V)]$$

- Directly optimize the joint likelihood is difficult due to the partition function in p_ϕ, instead we optimize the pseudolikelihood function

$$\begin{aligned}\ell_{PL}(\phi) &\triangleq \mathbb{E}_{q_\theta(\mathbf{y}_U|\mathbf{x}_V)}[\sum_{n\in V}\log p_\phi(\mathbf{y}_n|\mathbf{y}_{V\backslash n},\mathbf{x}_V)] \\ &= \mathbb{E}_{q_\theta(\mathbf{y}_U|\mathbf{x}_V)}[\sum_{n\in V}\log p_\phi(\mathbf{y}_n|\mathbf{y}_{\mathrm{NB(n)}},\mathbf{x}_V)]\end{aligned}$$

Inference/E-step: approximate $p_\phi(\mathbf{y}_U|\mathbf{y}_L, \mathbf{x}_V)$

- Approximate it with variational distribution $q_\theta(\mathbf{y}_U|\mathbf{x}_V)$. Specifically we use mean-field method:

$$q_\theta(\mathbf{y}_U|\mathbf{x}_V) = \prod_{n \in U} q_\theta(\mathbf{y}_n|\mathbf{x}_V)$$

- We parametrize each variational distribution with a Graph Neural Network

$$q_\theta(\mathbf{y}_n|\mathbf{x}_V) = \mathrm{Cat}(\mathbf{y}_n|\mathrm{softmax}(W_\theta \mathbf{h}_{\theta,n}))$$

Object representations learned by GNN

Learning/M-step:

- The log-pseudo likelihood:

$$\ell_{PL}(\phi) \triangleq \mathbb{E}_{q_\theta(\mathbf{y}_U|\mathbf{x}_V)}[\sum_{n\in V} \log p_\phi(\mathbf{y}_n|\mathbf{y}_{V\backslash n}, \mathbf{x}_V)]$$

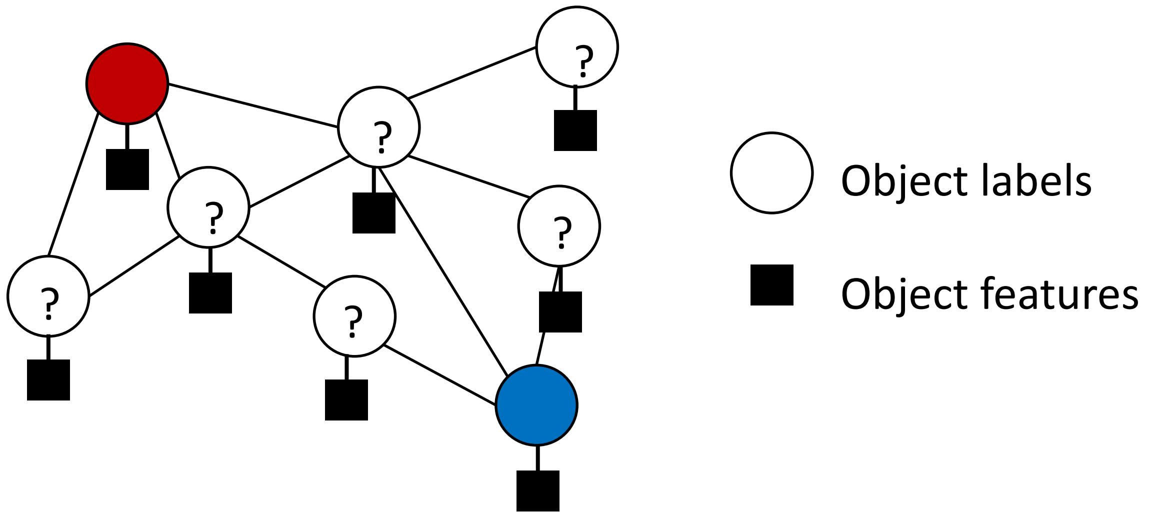
$$= \mathbb{E}_{q_\theta(\mathbf{y}_U|\mathbf{x}_V)}[\sum_{n\in V} \log p_\phi(\mathbf{y}_n|\mathbf{y}_{\mathrm{NB(n)}}, \mathbf{x}_V)]$$

- According to the inference, only the $p_\phi(\mathbf{y}_\mathbf{n}|\mathbf{y}_{\mathrm{NB(n)}}, \mathbf{x}_\mathrm{V})$ is required
- Parametrize $p_\phi(\mathbf{y}_\mathbf{n}|\mathbf{y}_{\mathrm{NB(n)}}, \mathbf{x}_\mathrm{V})$ with another GCN

$$p_\phi(\mathbf{y}_n|\mathbf{y}_{\mathrm{NB}(n)}, \mathbf{x}_V) = \mathrm{Cat}(\mathbf{y}_n|\mathrm{softmax}(W_\phi \mathbf{h}_{\phi,n}))$$

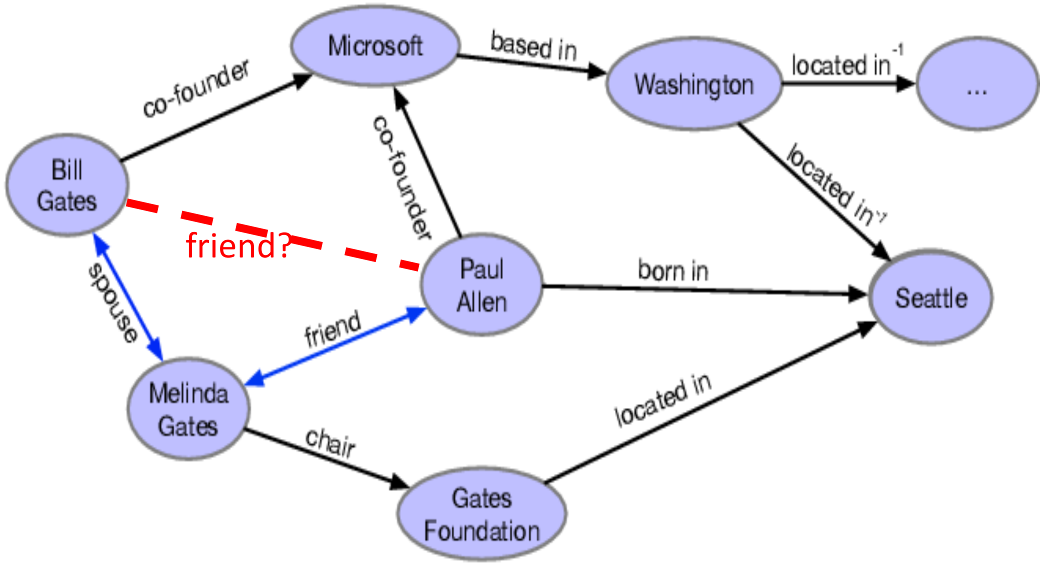
Overall Optimization Procedure

- Two Graph Neural Networks Collaborate with each other
  - p_ϕ: learning network, modeling the label dependency
  - q_θ: inference network, learning the object representations
- q_θ infer the labels of unlabeled objects trained with supervision from p_ϕ and labeled objects
- p_ϕ is trained with a fully labeled graph, where the unlabeled objects are labeled by q_θ



Reasoning on Knowledge Graphs

- A set of facts $\text{K}G = \{(h, r, t)\}$ represented as triplets
  - (Bill_Gates, Co_Founder, Microsoft)
- A fundamental problem: **predicting the missing facts by reasoning with existing facts**



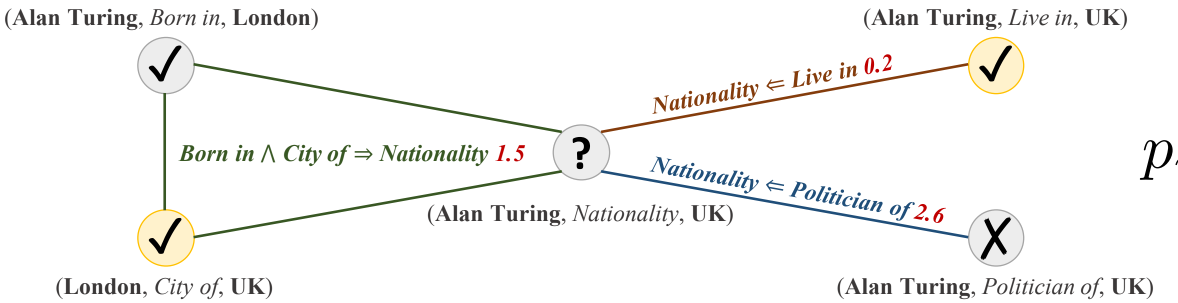
Probabilistic Logic Neural Networks for Reasoning (Qu and Tang, NeurIPS'19.)

- Towards combining Markov Logic Networks and knowledge graph embedding
  - Leverage logic rules and handling their uncertainty
  - Effective and efficient inference
- Define the joint distribution of facts with Markov Logic Network
- Optimization with variational EM
  - Parametrize the variational distribution with knowledge graph embedding methods

Meng Qu and Jian Tang. "Probabilistic Logic Neural Networks for Reasoning." In NeurIPS'2019.

pLogicNet

- Define the joint distribution of facts with an MLN



$$p_w(\mathbf{v}_O, \mathbf{v}_H) = \frac{1}{Z} \exp\left(\sum_l w_l n_l(\mathbf{v}_O, \mathbf{v}_H)\right)$$

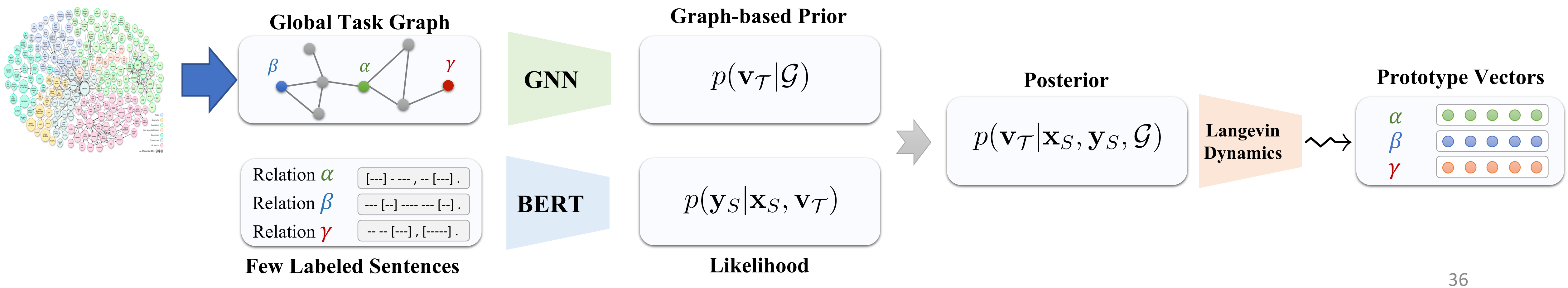
- Learning by maximizing the variational lower-bound of the log-likelihood of observed facts

$$\log p_w(\mathbf{v}_O) \geq \mathcal{L}(q_\theta, p_w) = \mathbb{E}_{q_\theta(\mathbf{v}_H)}[\log p_w(\mathbf{v}_O, \mathbf{v}_H) - \log q_\theta(\mathbf{v}_H)]$$

Other Projects Related to GNNs

Graph Neural Networks for Bayesian Meta-Learning (Qu et al. 2020)

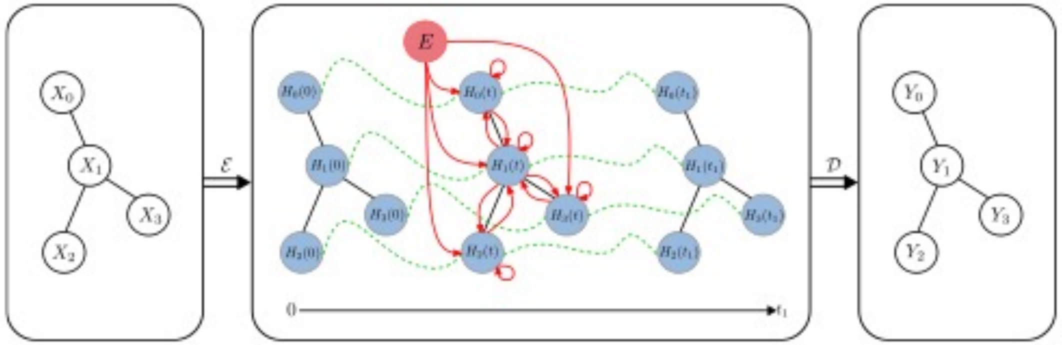
- The prototype vectors of tasks v_T are treated as random variables
- The prior of v_T are encoded with a GNN on the task graph, $\mathrm{p}(\mathrm{y}_\mathrm{T}|G)$.
- The likelihood is defined on the support set, $\mathrm{p}(\mathrm{y}_\mathrm{s}|\mathrm{x}_\mathrm{s}, \mathrm{v}_T)$.
- Sample from the posterior $\mathrm{p}(\mathrm{v}_\mathrm{T}|\mathrm{x}_\mathrm{s}, \mathrm{y}_\mathrm{s}, G)$ with Langevin Dynamics
  - Similar to MAML but can handle the uncertainty of v_T



Graph Neural Networks with Neural ODEs (Xhonneux and Qu et al. 2020)

- Graph Neural Networks: discrete dynamics of node representations with graph convolutional layers
  - Can we generalize it to continuous dynamics?
- Model the dynamics of node representations with Neural ODEs
- Inspired by epidemiological models, dynamic of node representations
  - Depending on the infection from neighbors: **AH(t)**
  - Nature recovery: **-H(t)**
  - Initial condition: **E**

$$\frac{\mathrm{d}\boldsymbol{H}(t)}{\mathrm{d}t} = (\boldsymbol{A} - \boldsymbol{I})\boldsymbol{H}(t) + \boldsymbol{E}$$



Louis-Pascal A. C. Xhonneux, Meng Qu, Jian Tang. **Continuous Graph Neural Networks**. arXiv:1912.00967

Outline

- Unsupervised Graph Representation Learning
- Relational Reasoning with Graph Representation Learning
- **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

Medical Knowledge Graph Construction

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

DrugBank

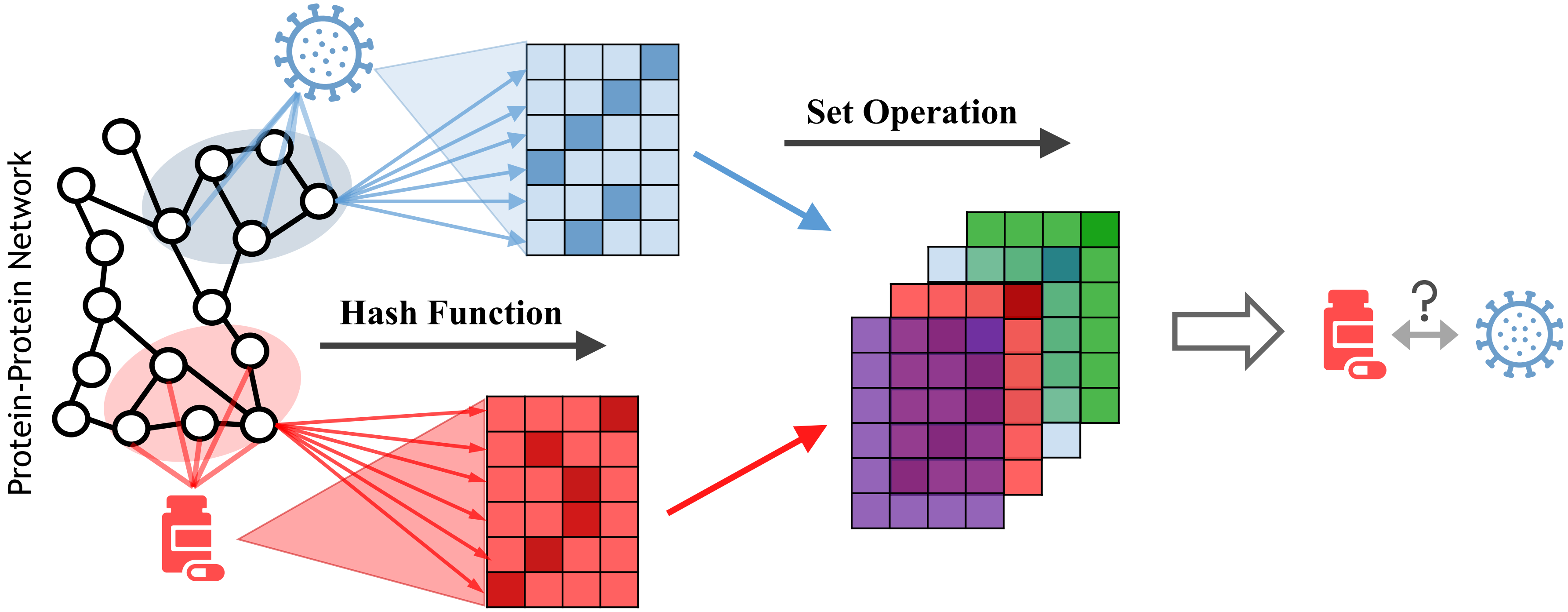
Comparative Toxicogenomics Database

Medical Knowledge Graph Representation Representation and Applications

- Each entity is a represented as a vector
  - Disease, Drug, Phenotype, Gene, Protein, Side effect
- Applications
  - Drug repurposing
  - Drug side effect prediction

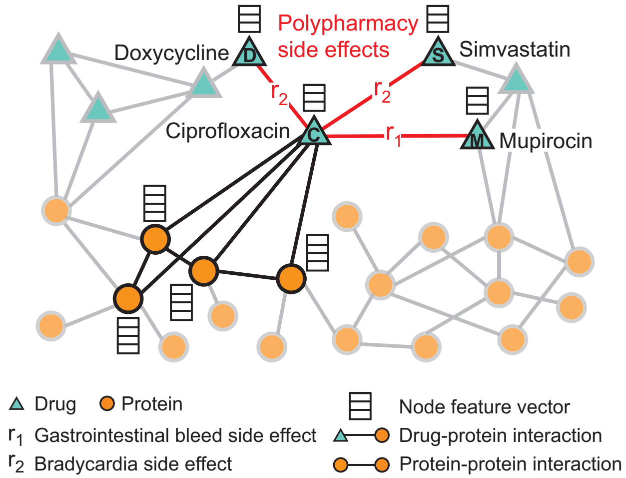
Drug Repurposing with Medical Knowledge Graph for COVID-19

- Predict the disease-drug relationships on the medical knowledge graph
  - Diseases, proteins, drugs

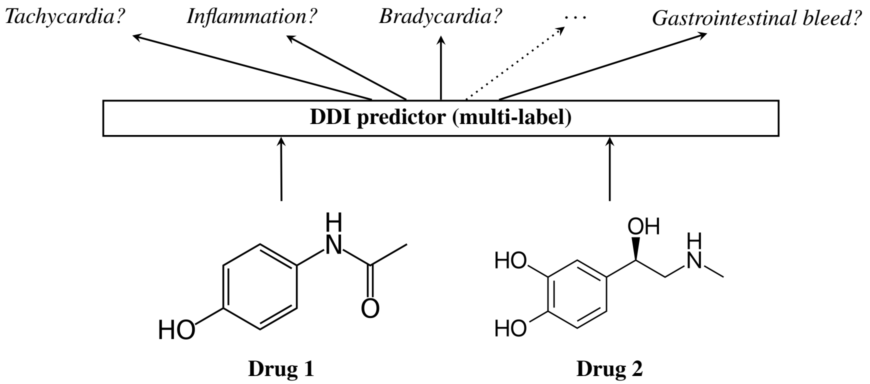


Drug-Drug Adverse Effect Prediction (Deac et al. 2019)

- Predicting the side effects of two drugs
  - Based on medical knowledge graph
  - Based on molecular graph structures



Knowledge graph based (Zitnik et al. 2018)

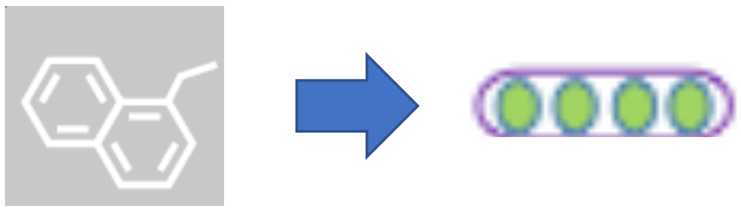


Molecular graph structures (Deac et al., 2019)

- Marinka Zitnik, Monica Agrawal, Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018.
- Andreea Deac, Yu-Hsiang Huang, Petar Veličković, Pietro Liò, Jian Tang. **Drug-Drug Adverse Effect Prediction with Graph Co-Attention**. arXiv:1905.00534

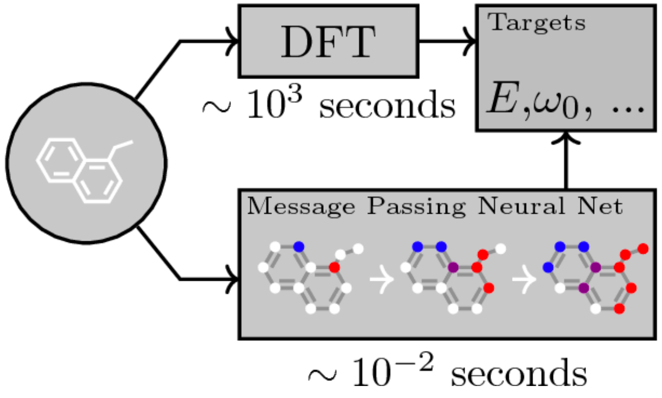
Molecule Properties Prediction

- 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



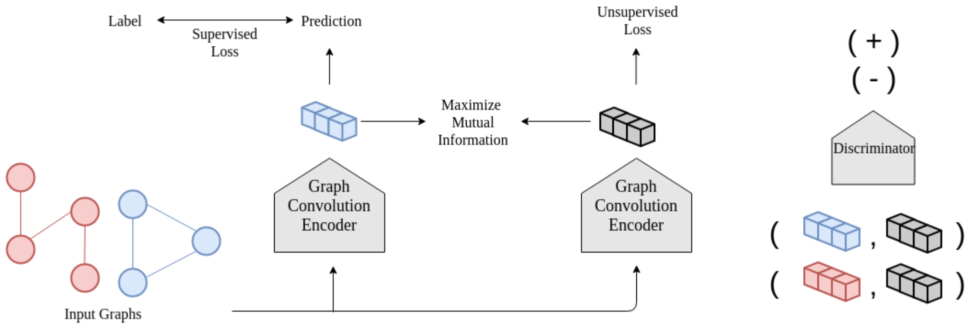
Unsupervised and Semi-supervised Learning for Molecular Graph Representation (Sun et al. ICLR 20)

- 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

Application: Finding Effective Antibiotics for Secondary Infections in COVID-19

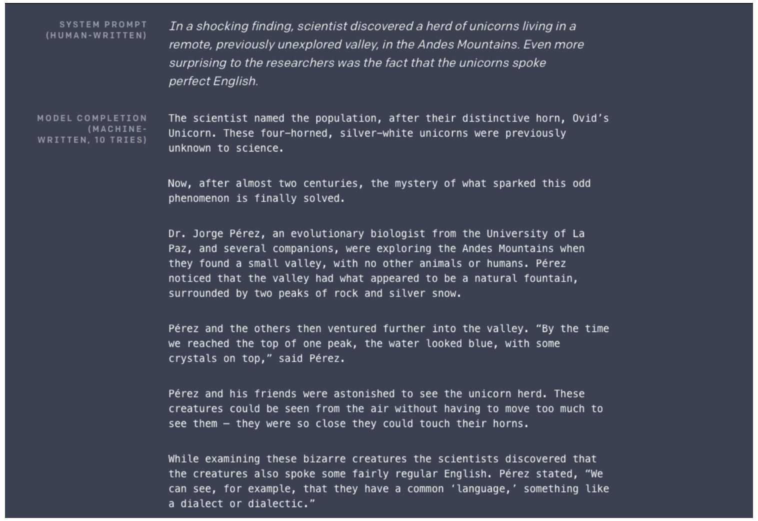
- A high proportion of non-surviving patients of COVID-19 developed a secondary infection (Zhou et al. 2020)
- Finding effective antibiotics
- Predict antibacterial properties
  - Collaboration with MIT
  - https://www.aicures.mit.edu/tasks

Rank	Model	Author	10-fold CV ROC-AUC	10-fold CV PRC-AUC	Test ROC-AUC	Test PRC-AUC
1	Pre-trained OGB-GIN (ensemble)	Weihua Hu@Stanford	0.905 +/- 0.133	0.494 +/- 0.333	0.837	0.651
2	Chemprop ++	AICures@MIT	0.810 +/- 0.160	0.423 +/- 0.332	0.891	0.641
3	Graph Self-supervised Learning	SJTU_NRC_Mila	0.825 +/- 0.210	0.530 +/- 0.342	0.800	0.622
4	Pre-trained OGB-GIN	Weihua Hu@Stanford	0.907 +/- 0.100	0.464 +/- 0.328	0.870	0.616
5		zhenxingwu	0.813 +/- 0.165	0.465 +/- 0.332	0.73	0.594
6	MLP	IITM			0.777	0.59
7		Gianluca Bontempi			0.848	0.584
8	Chemprop	IITKGP	0.820	0.602	0.818	0.58
9		Gianluca Bontempi			0.858	0.578
10		Gianluca Bontempi			0.789	0.569
11	Random forest	Gianluca Bontempi	0.852		0.832	0.561
12		Aymen Waheb			0.721	0.556
13		zhenxingwu			0.705	0.539
14		Apoorv Umang			0.84	0.513
15	XGBOOST	Andrea Loreggia	0.705 +/- 0.189		0.714	0.301
16	Chemprop	AICures@MIT	0.725 +/- 0.207	0.310 +/- 0.271	0.716	0.183

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.

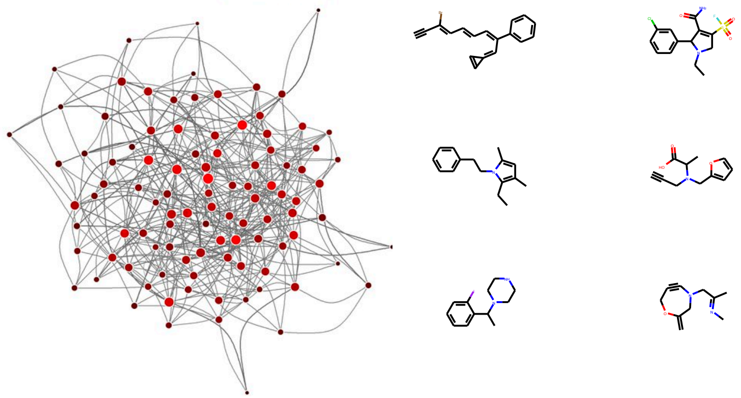
Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock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

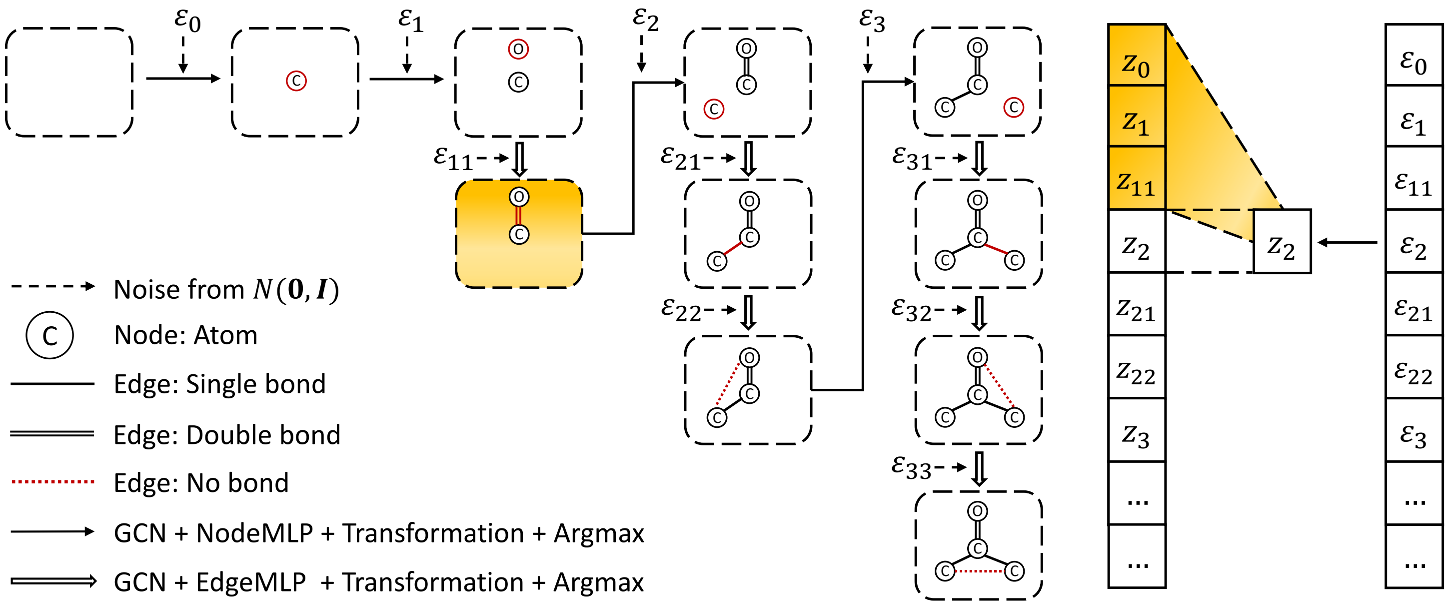
Text generated by by GPT-2,
Examples from Internet



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

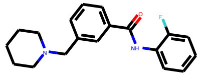
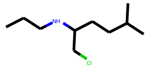
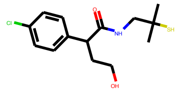
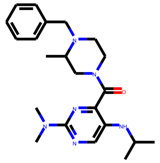
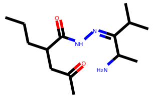
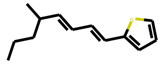
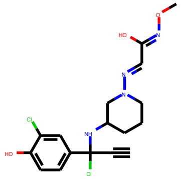
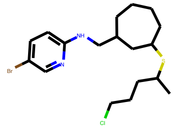
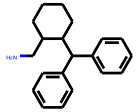
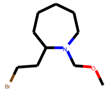
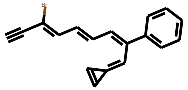
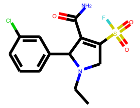
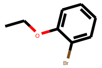
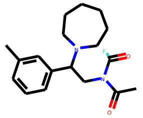
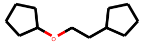
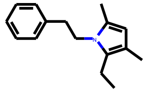
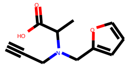
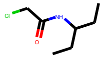
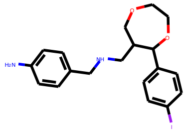
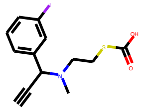
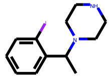
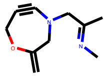
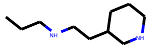
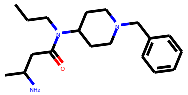
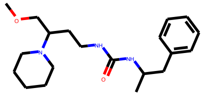
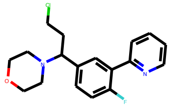
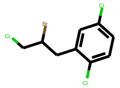
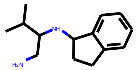
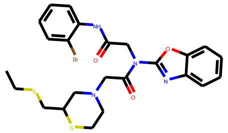
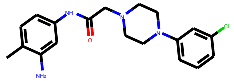


(a) Sampling Framework (b) Autoregressive Flow

Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, Jian Tang. **GraphAF: a Flow-based Autoregressive Model for Molecular Graph Generation**. ICLR'20

Molecule Generation

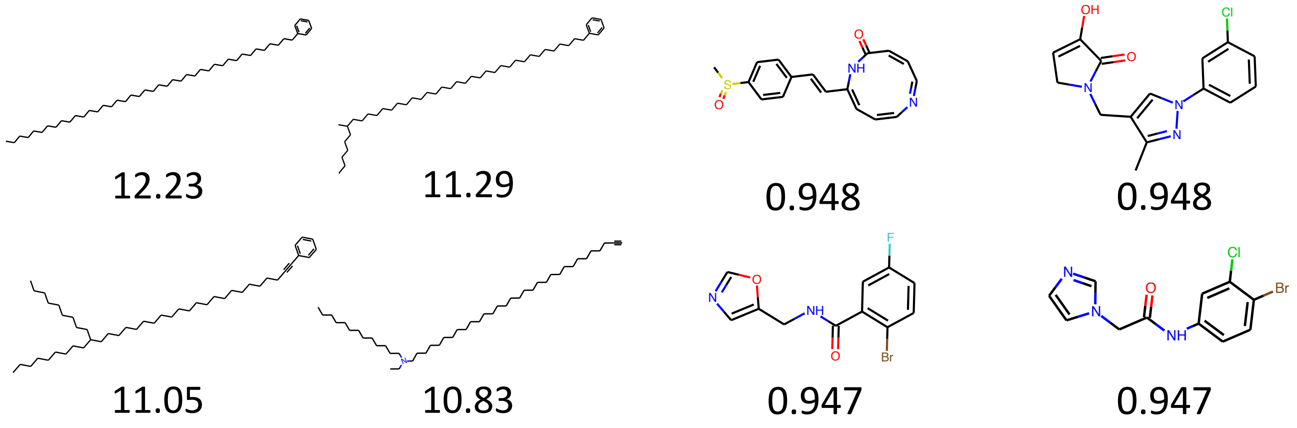
Method	Validity	Validity w/o check	Uniqueness	Novelty	Reconstruction
JT-VAE	100%	—	100%‡	100%‡	76.7%
GCPN	100%	20%†	99.97%‡	100%‡	—
MRNN	100%	65%	99.89%	100%	—
GraphNVP	42.60%	—	94.80%	100%	100%
GraphAF	100%	68%	99.10%	100%	100%

Molecule Optimization

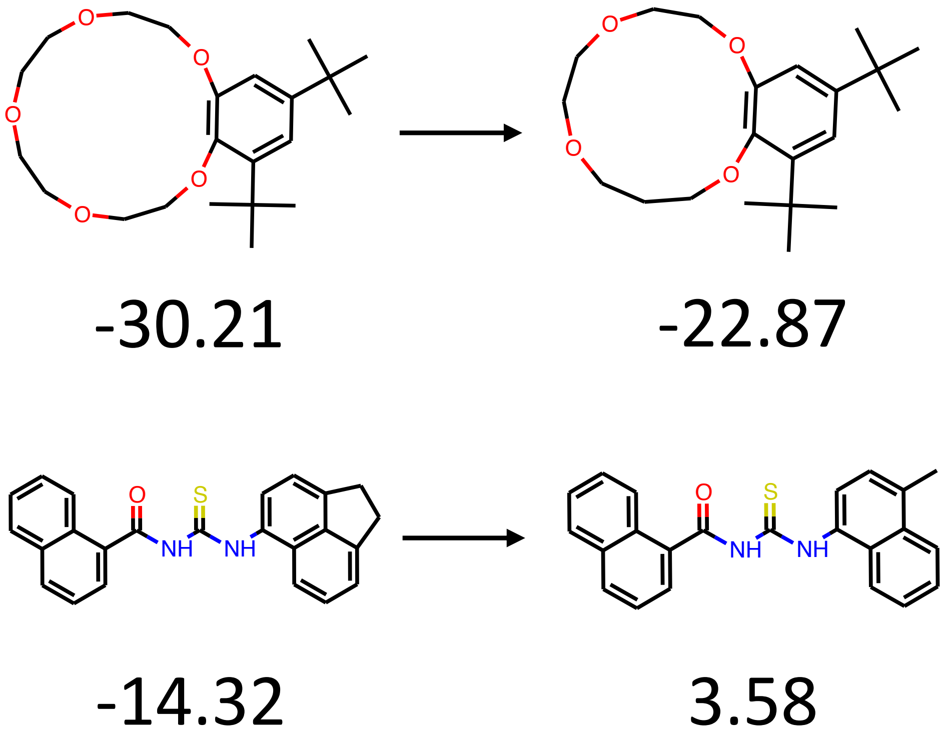
- Properties
  - Penalized logP
  - QED (druglikeness)

Method	Penalized logP 1st	2nd	3rd	Validity	QED 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

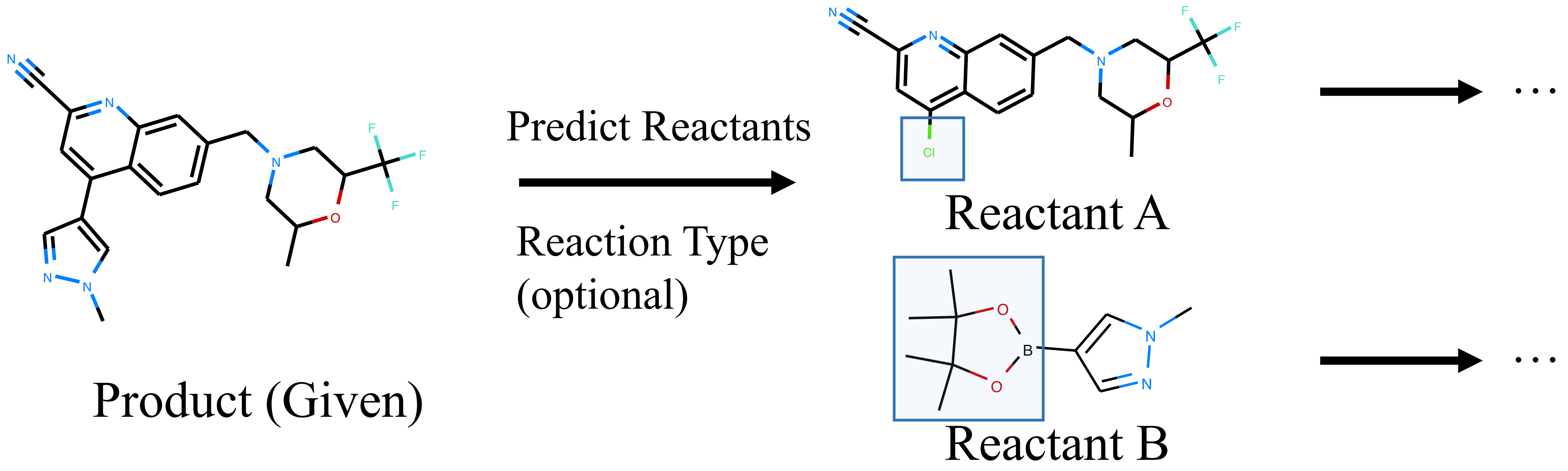
Constrained Optimization



(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

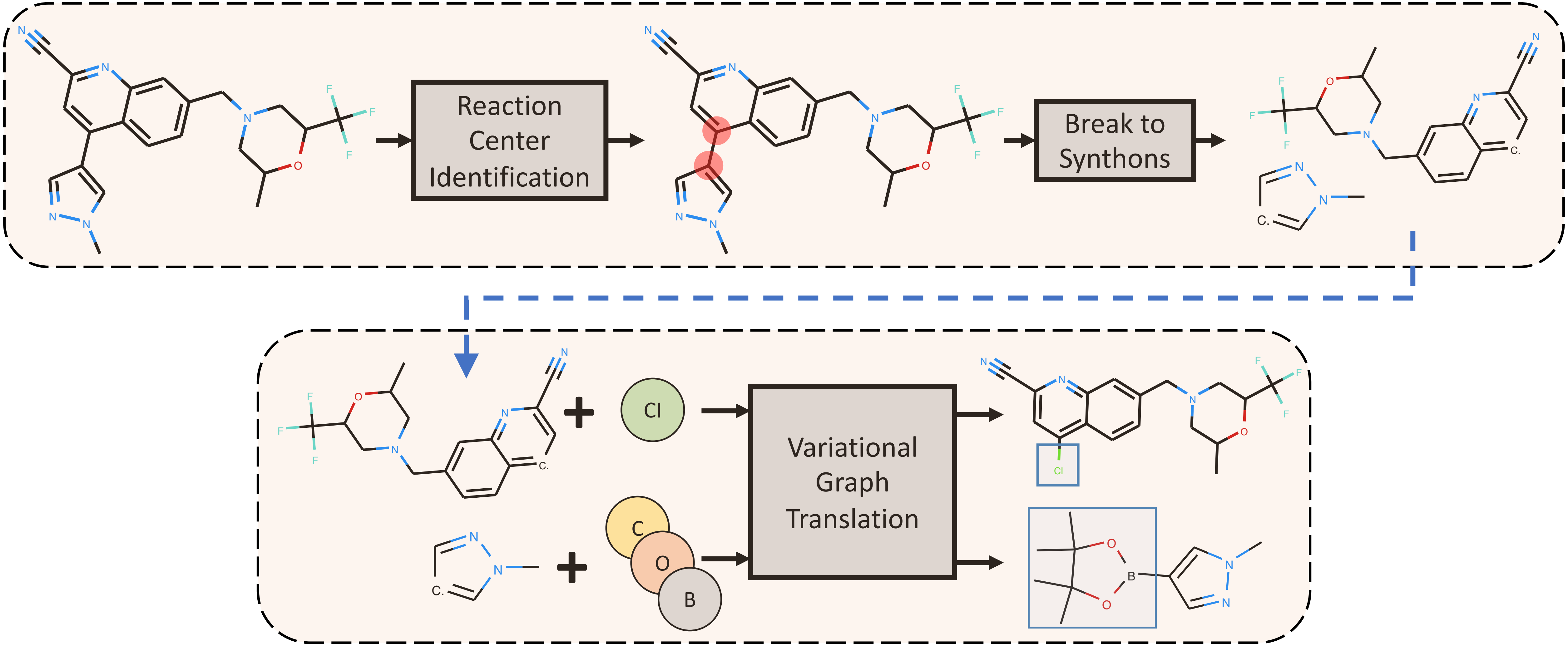


A Graph to Graphs Framework for Retrosynthesis Prediction (Shi et al. 2020)

- 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

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**

Take Away

- Graph representation learning
  - A growing research topic in machine learning focusing on deep learning for graph-structured data
- Graph Representation learning for relational/logical reasoning
  - Graph as bridges between system I and II reasoning
- Graph representation learning for drug discovery
  - Many data in this domain are graph-structured, e.g., molecules and medical knowledge graph

Selected Publications

- Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan and Qiaozhu Mei. LINE: Large-scale Information Network Embedding. WWW'15. (> 2000 citations)
- Jian Tang, Jingzhou Liu, Ming Zhang and Qiaozhu Mei. Visualizing Large-scale and High-dimensional Data. WWW'16. (Best paper nomination)
- Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, Jian Tang. **RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space.** ICLR'19.
- Zhaocheng Zhu, Shizhen Xu, Meng Qu, Jian Tang. **GraphVite: A High-Performance CPU-GPU Hybrid System for Node Embedding.** WWW'19.
- Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhiyuan Liu, Juanzi Li, Jian Tang. **KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation.** arXiv:1911.06136.
- Meng Qu, Yoshua Bengio, Jian Tang. **"GMNN: Graph Markov Neural Networks".** In ICML'19.
- Meng Qu, Jian Tang. **Probabilistic Logic Neural Networks for Reasoning.** NeurIPS'19.
- Fan-Yun Sun, Meng Qu, Jordan Hoffmann, Chin-Wei Huang, Jian Tang. vGraph: **A Generative Model for Joint Community Detection and Node Representation Learning**. NeurIPS'19.
- Meng Qu, Tianyu Gao, Louis-Pascal Xhonneux, Jian Tang. **Few-shot Relation Extraction via Bayesian Meta-learning on Task Graphs.** In Submission.
- Louis-Pascal A. C. Xhonneux, Meng Qu, Jian Tang. **Continuous Graph Neural Networks**. arXiv:1912.00967
- Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, Jian Tang. **GraphAF: a Flow-based Autoregressive Model for Molecular Graph Generation**. ICLR'20
- Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, Jian Tang. **InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization.** ICLR'20
- Andreea Deac, Yu-Hsiang Huang, Petar Veličković, Pietro Liò, Jian Tang. **Drug-Drug Adverse Effect Prediction with Graph Co-Attention.** arXiv:1905.00534
- Chence Shi, Minkai Xu, Hongyu Guo, Ming Zhang and Jian Tang. **A Graph to Graphs Framework for Retrosynthesis Prediction**. arXiv:2003.12725.

Thanks!

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