

EVENTO
2018
BIGDATATECH

DATA FOR HUMAN



DAVIDE BACCIU



UNIVERSITÀ DI PISA

Bioinformatica intelligente - Il deep learning per grafi e le sue applicazioni biomediche e farmaceutiche

MILANO | 25 OTTOBRE 2018

www.bnova.it



Grafi: perché?





Grafi: perché?

Il contesto può
essere
fondamentale nella
valutazione
dell'informazione

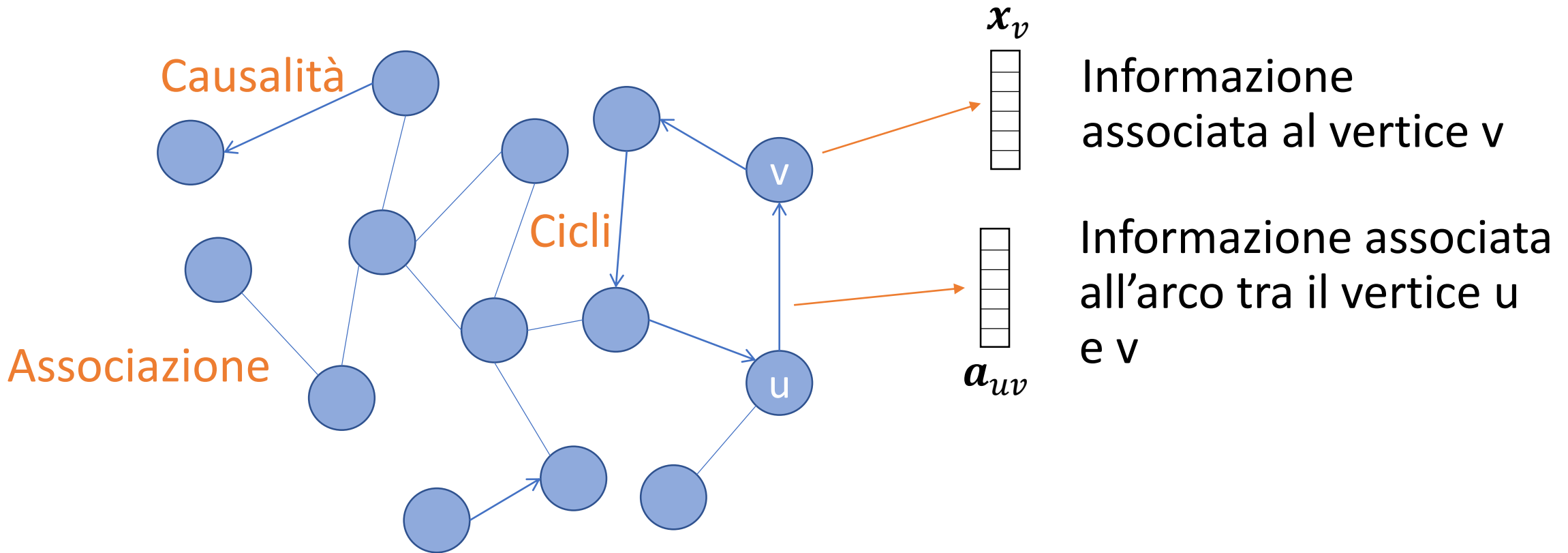




I DATI STRUTTURATI



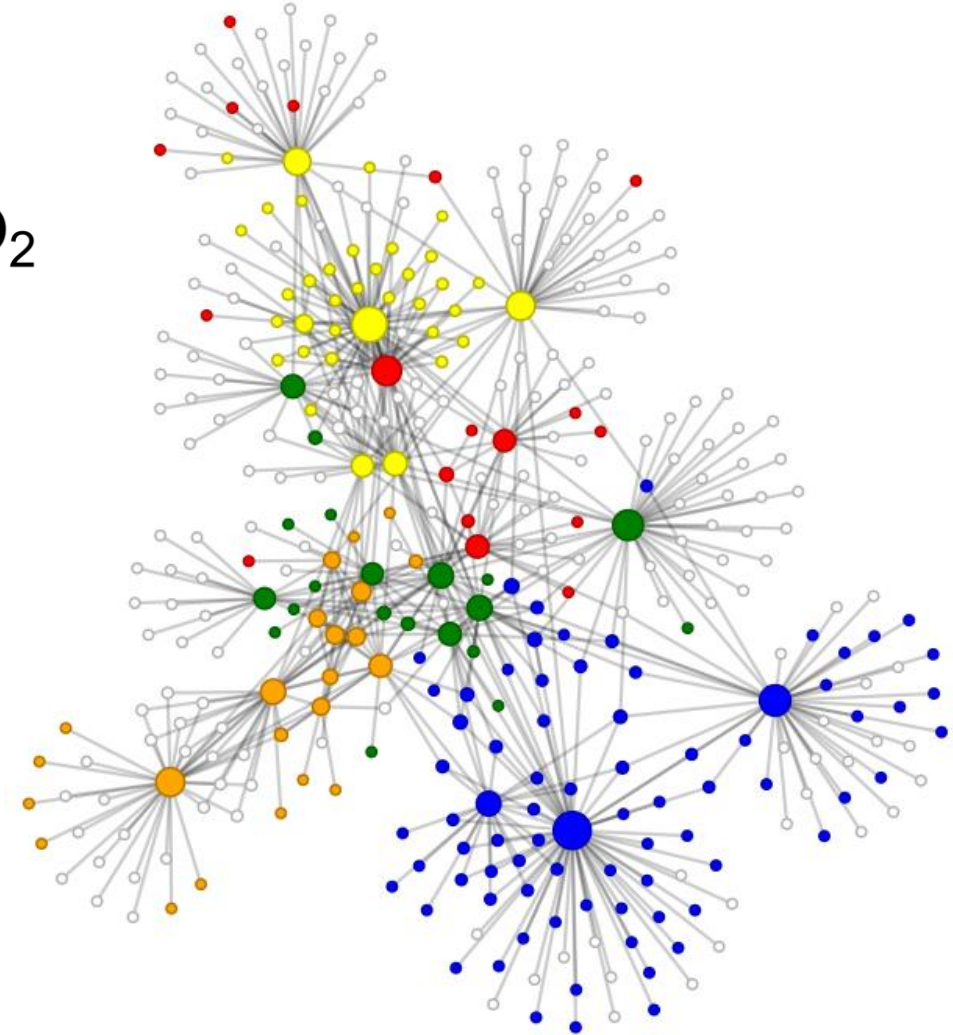
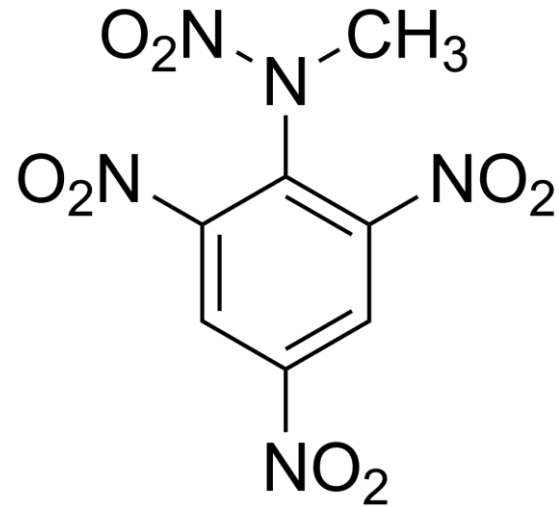
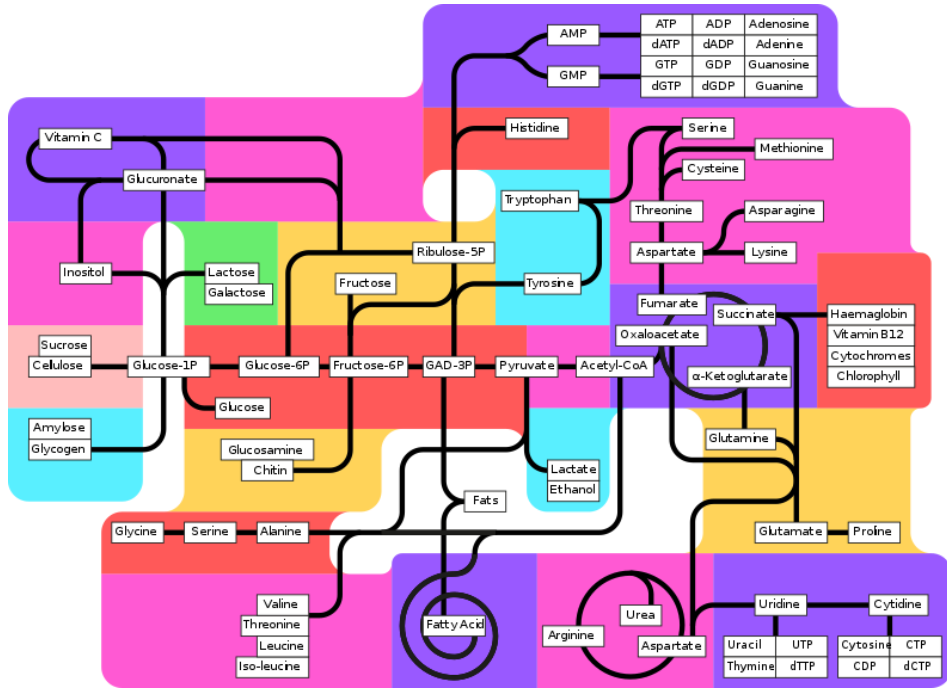
Grafi



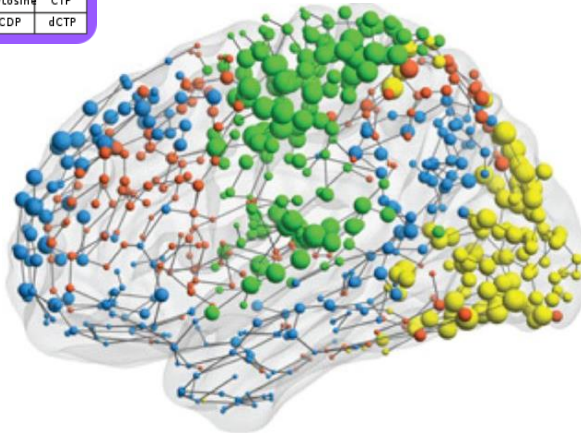
Astrazione che rappresenta pezzi di informazione (i **vertici**) e le loro relazioni (gli **archi**)



L'informazione strutturata e dove trovarla



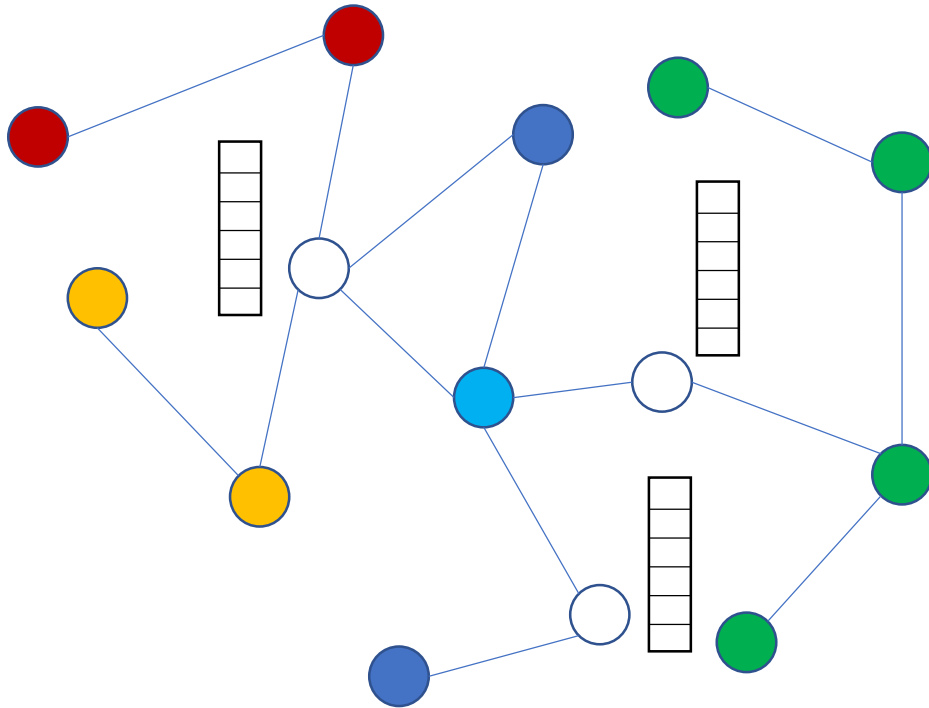
Prodotta da processi
naturali e artificiali



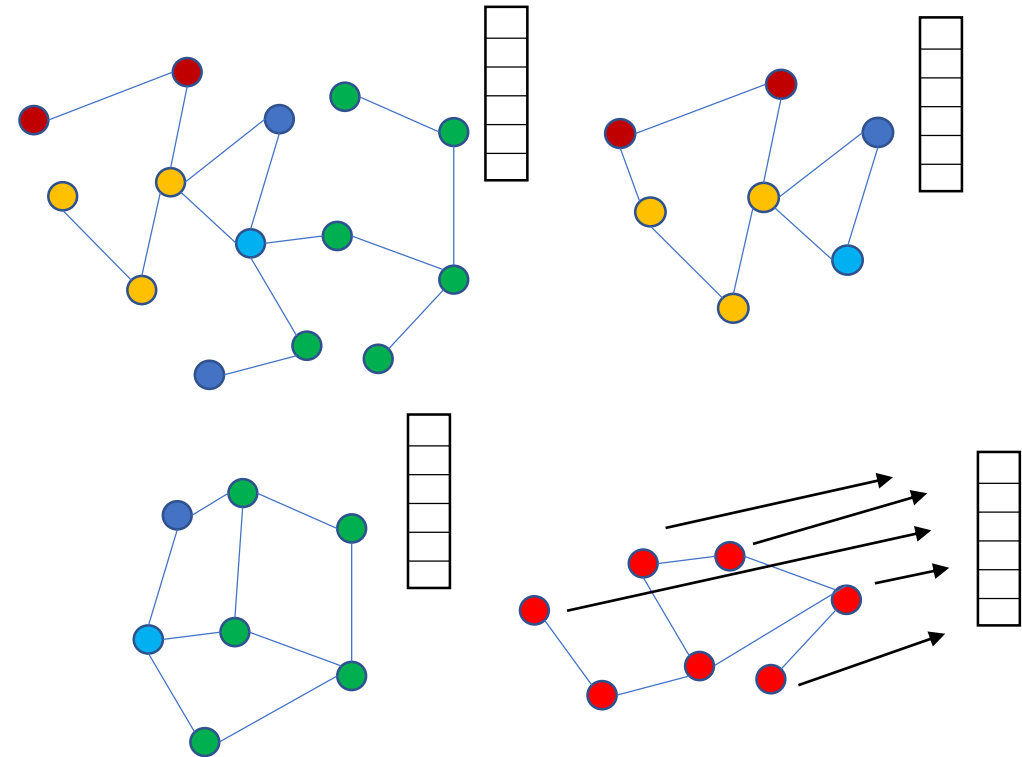


Grafi e il loro uso predittivo

Un **unico grafo**, con informazioni mancanti o da inferire



Una **dataset di grafi** con struttura variabile, per cui generare predizioni

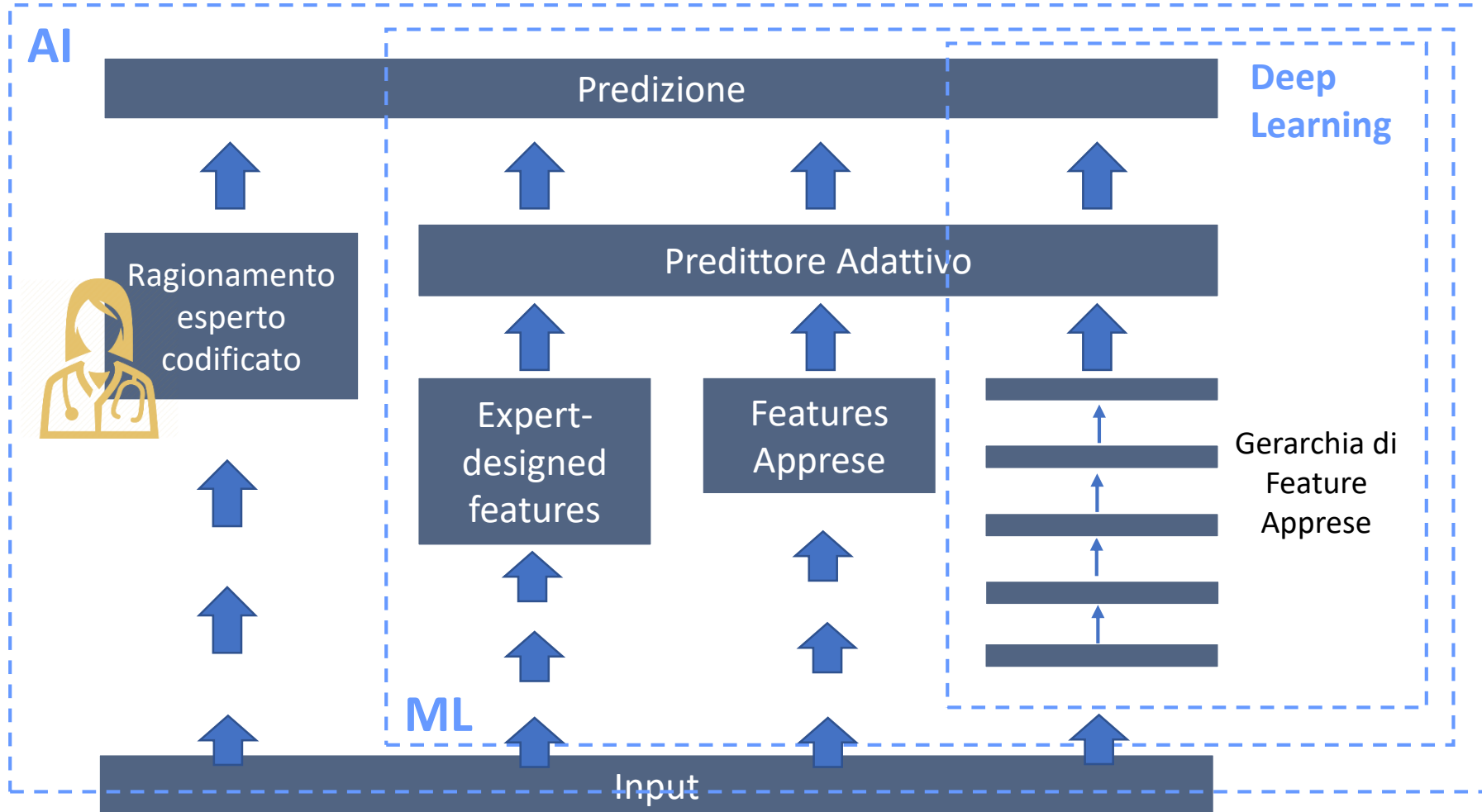




DEEP LEARNING

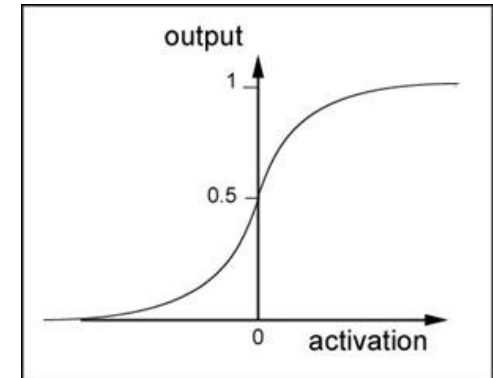
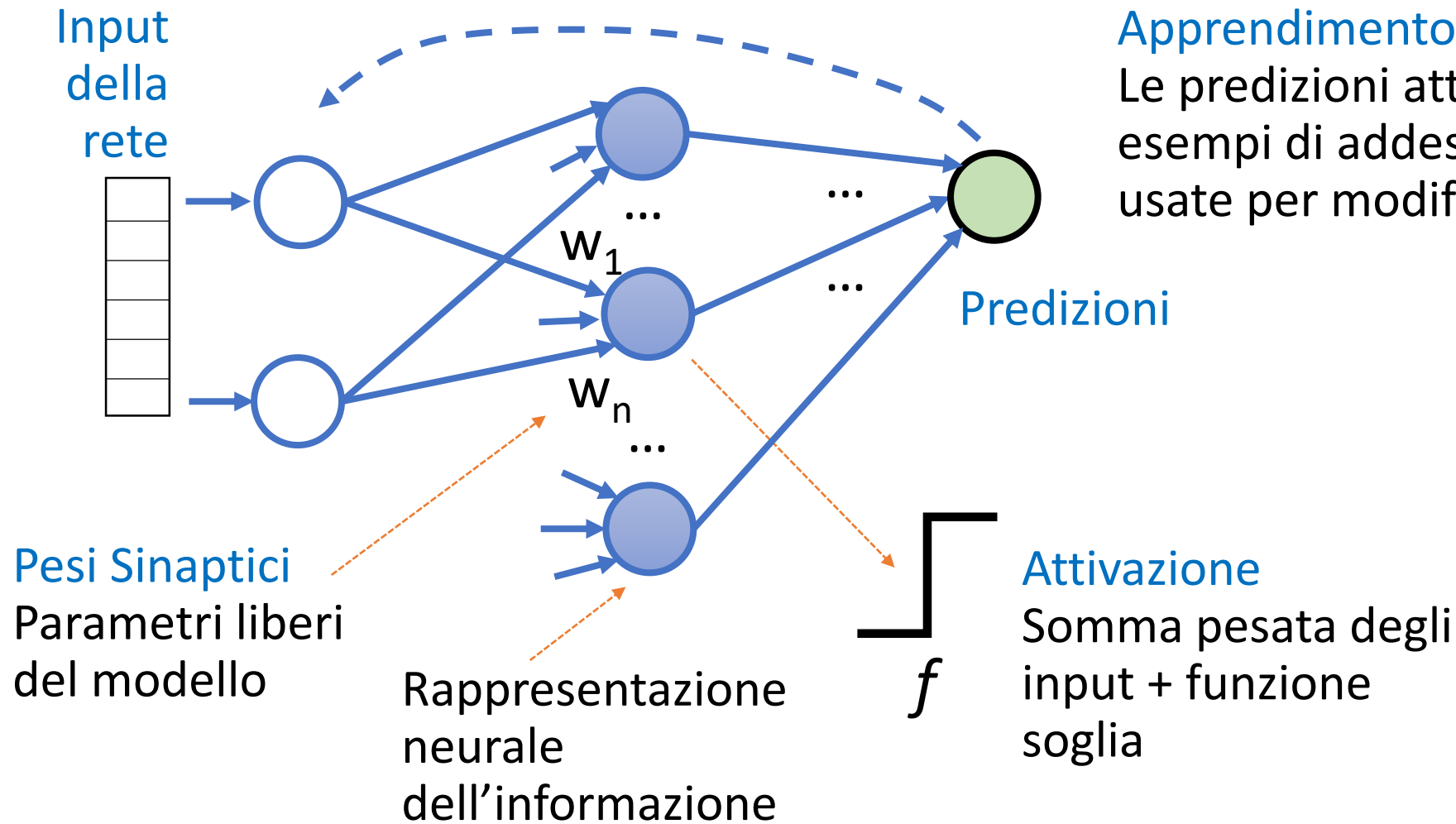


Intelligenza Artificiale, Machine Learning e Deep Learning





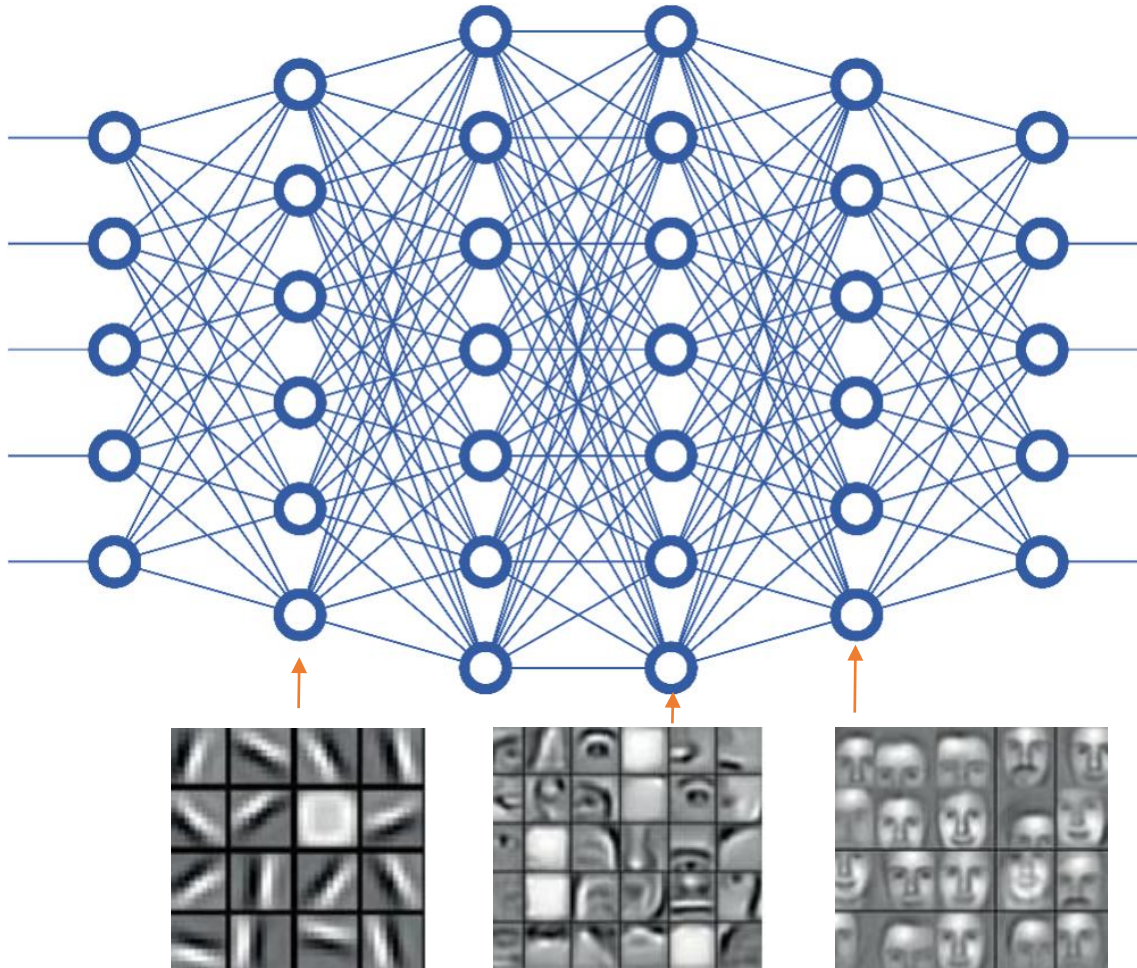
Reti neurali in 1 slide





Deep Learning

Reti Neurali Profonde



Identità 1

Identità 2

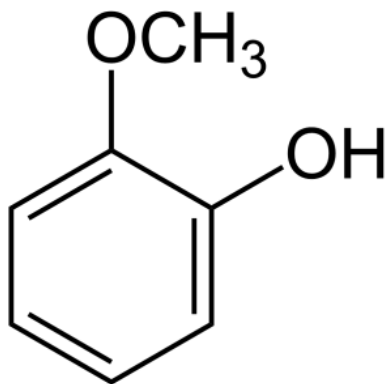
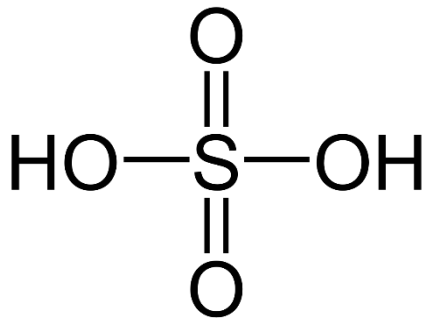
...

Un insieme di tecniche e metodologie per **estrarre automaticamente** dai dati una **rappresentazione efficace e gerarchica** dell'informazione



Perchè il Deep Learning è interessante per trattare grafi?

Machine learning *classico*

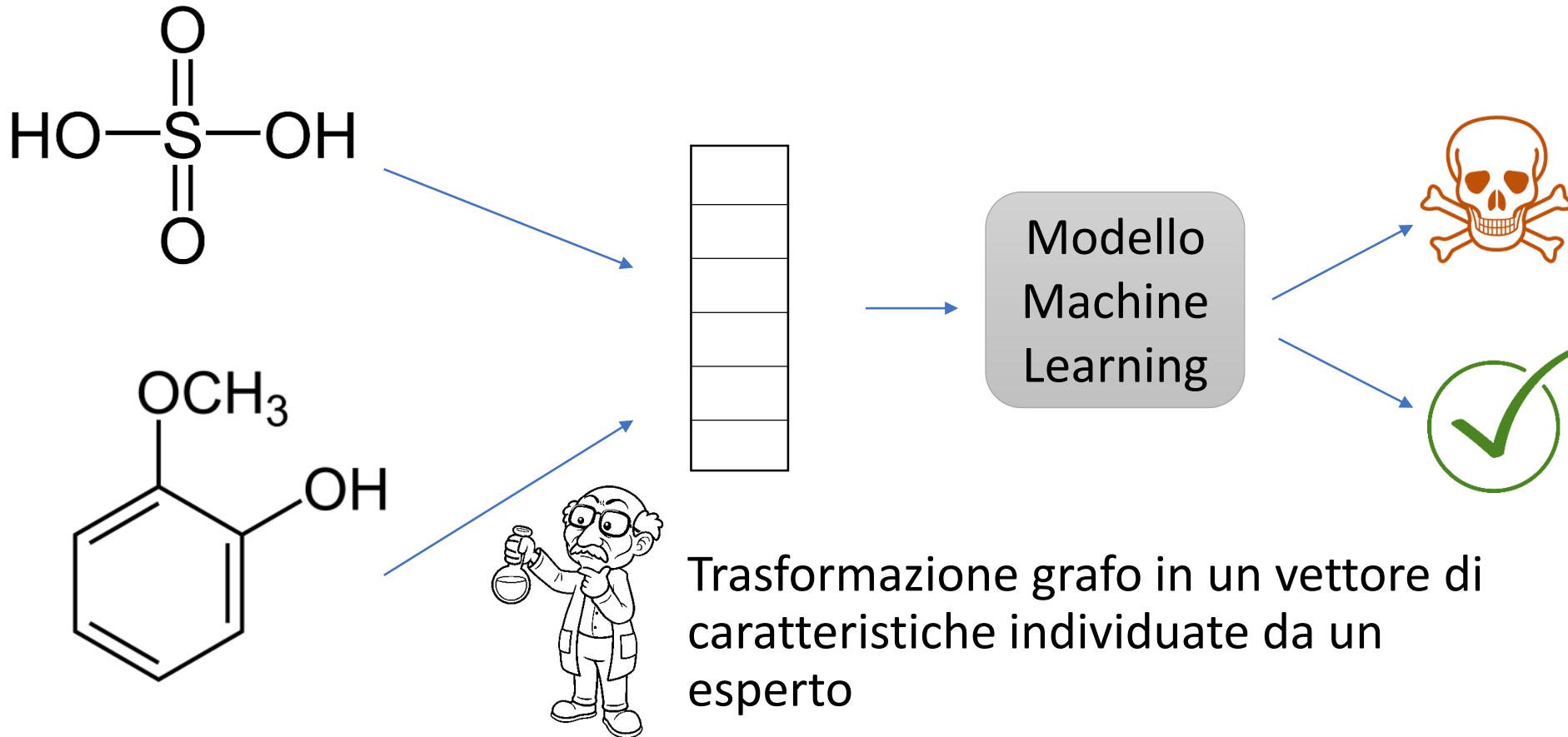


Modello
Machine
Learning

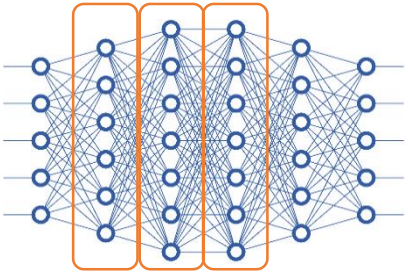


Perchè il Deep Learning è interessante per trattare grafi?

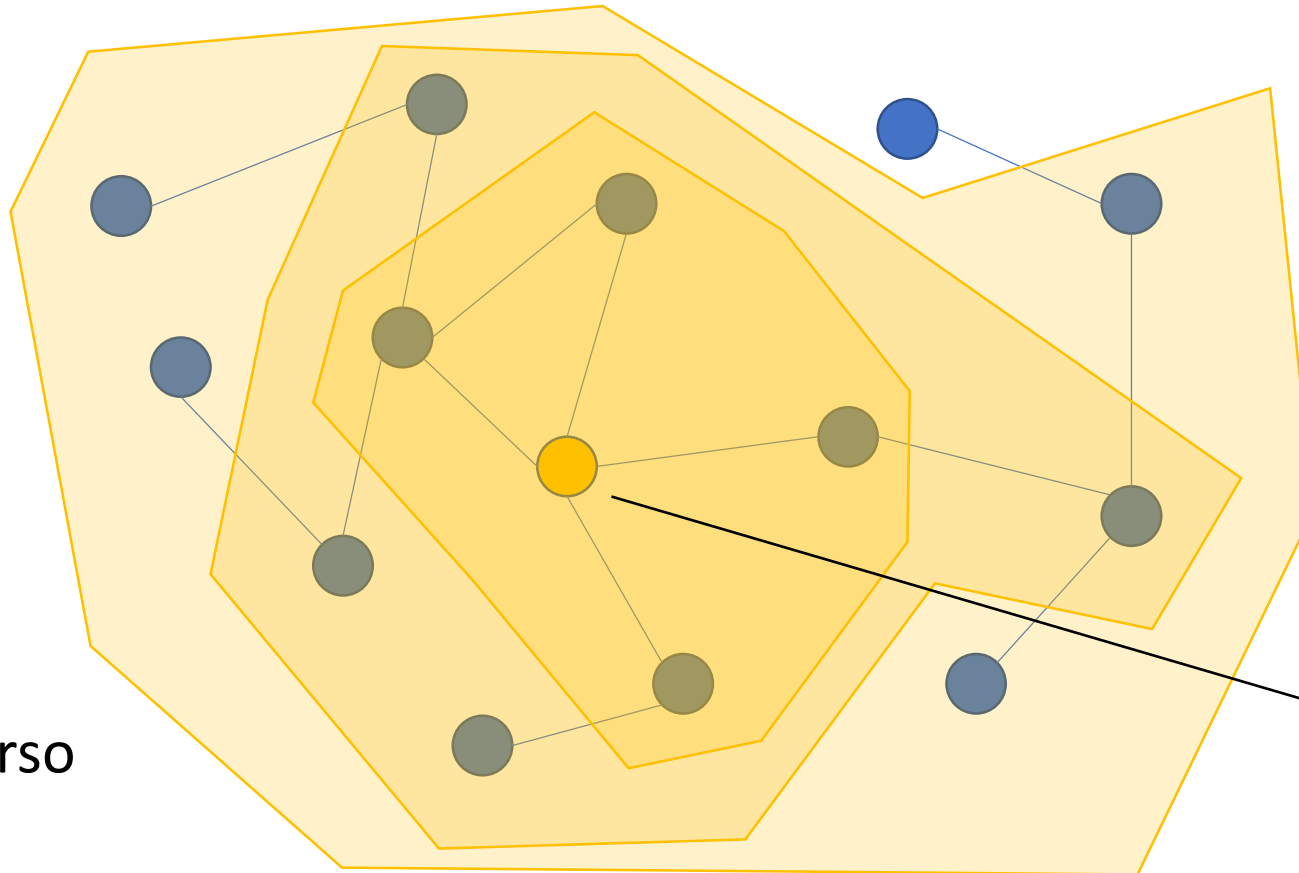
Machine learning *classico*



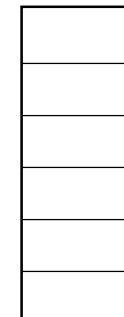
Perchè il Deep Learning è interessante per trattare grafi?



Approccio gerarchico permette la diffusione efficiente del contesto attraverso la struttura



Rappresentazione appresa automaticamente, tenendo in considerazione le relazioni strutturali (contesto)

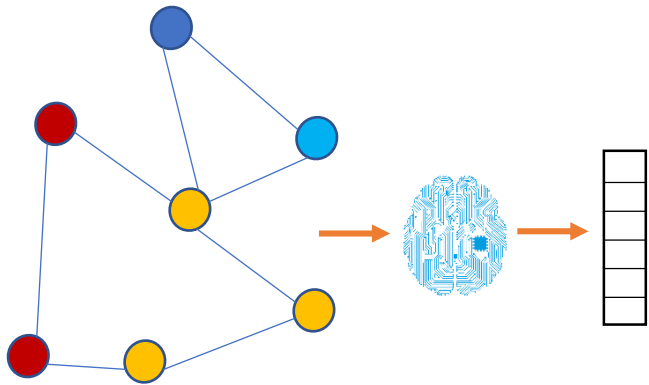




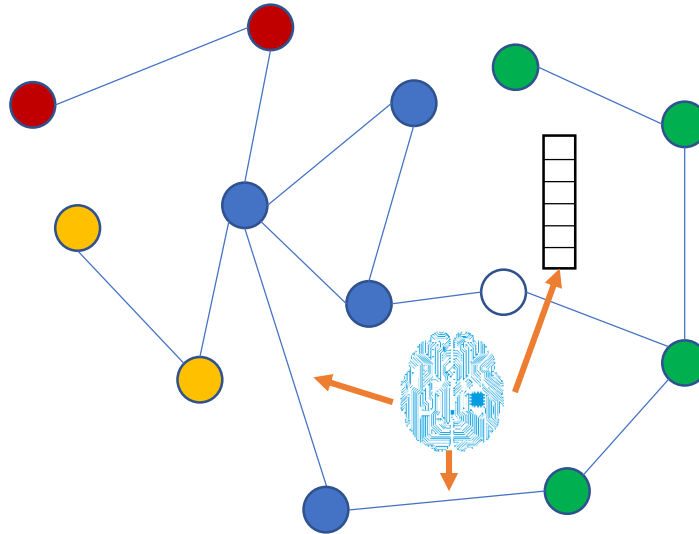
CASI STUDIO



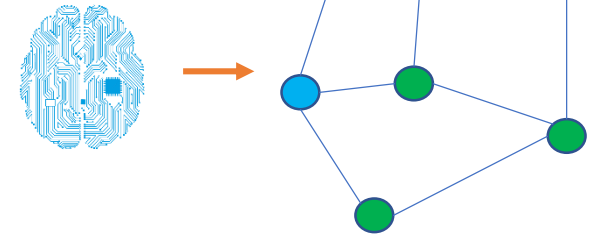
Tre task di apprendimento



(1) Classificazione o regressione su grafi



(2) Inferenza su nodi e archi

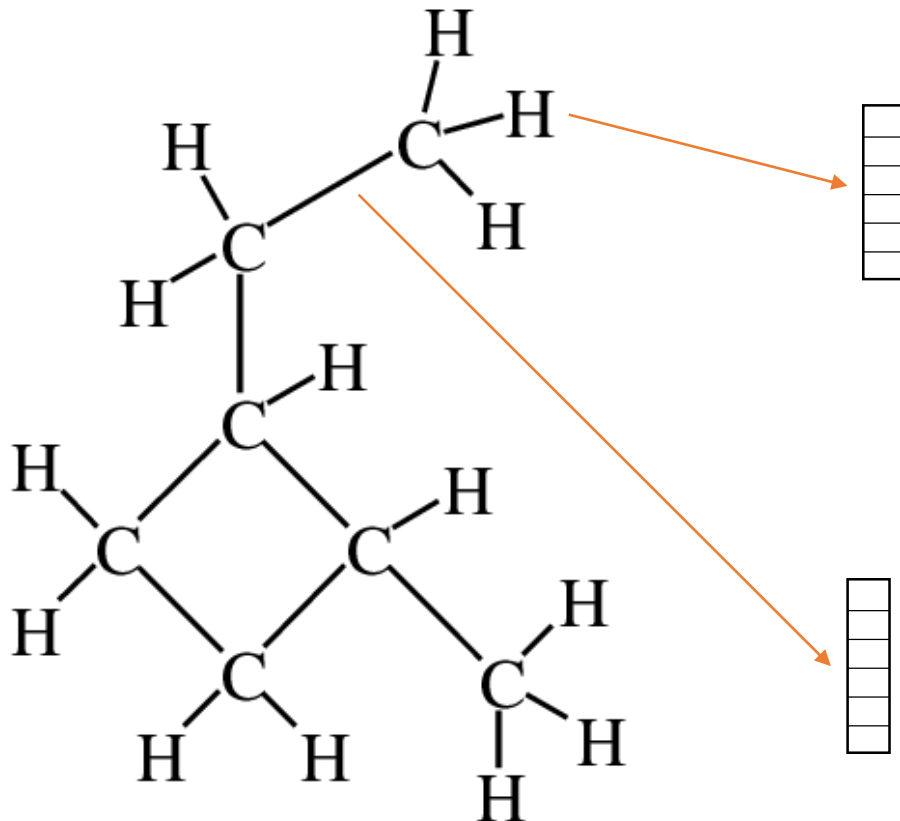


(3) Generazione di grafi



Grafi Molecolari

Rappresentazione ricca dei componenti e delle loro relazioni spaziali, essenziali nel determinarne la funzionalità

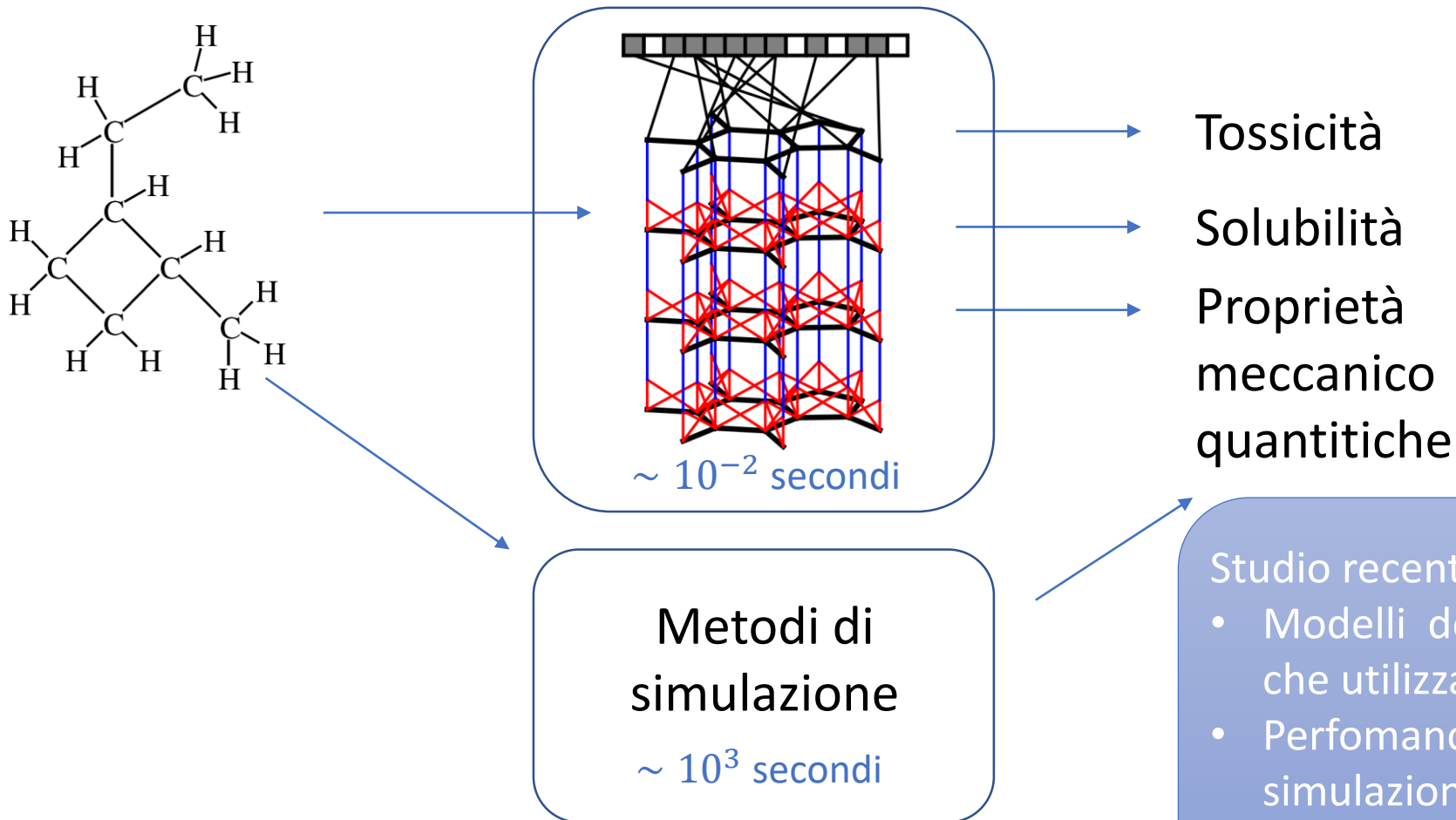


- Tipo atomo
- Peso atomico
- Accetta/dona elettroni
- Aromatico
- ...

- Tipo legame
- Distanza tra atomi
- ...



Predizione Proprietà dei Composti Chimici (1)

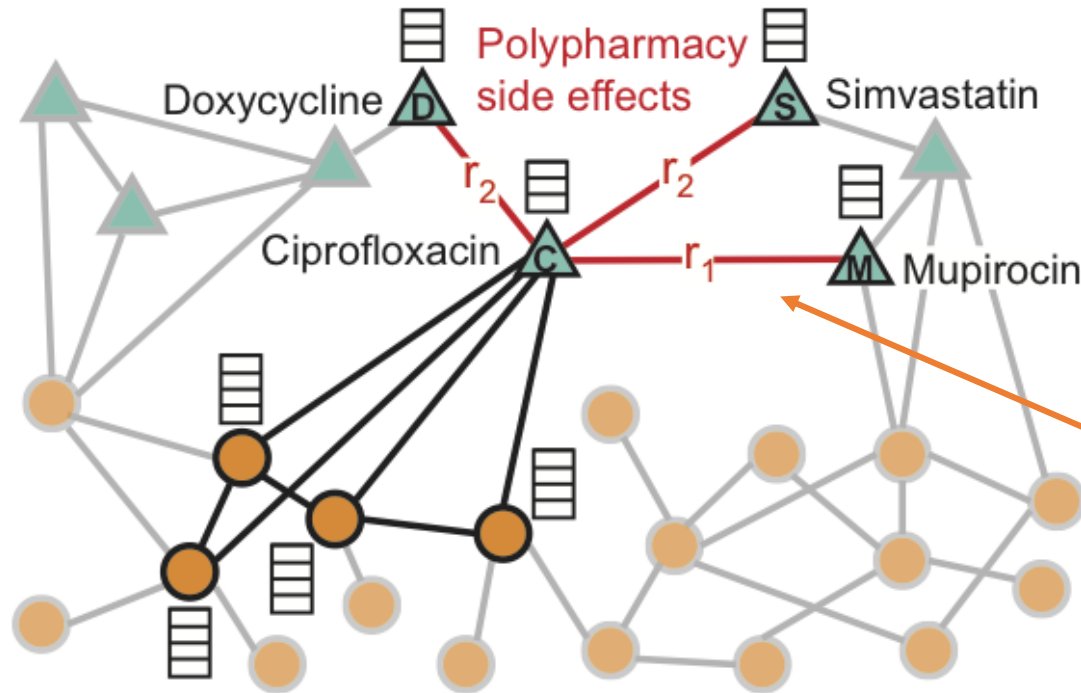


- Studio recente su 134K molecole
- Modelli deep learning superiori a modelli che utilizzano conoscenza di esperti
 - Performance identica ai modelli di simulazione in 11 su 13 proprietà

Gilmer et al, ICML 2017



Effetti collaterali da interazione tra farmaci (2)



Grafo **multimodale** che **rappresenta interazioni**

- Farmaco-farmaco
- Farmaco-proteina
- Proteina-proteina

Inferire la presenza di interazioni sconosciute (archi)

▲ Drug ● Protein

≡ Node feature vector

r_1 Gastrointestinal bleed side effect

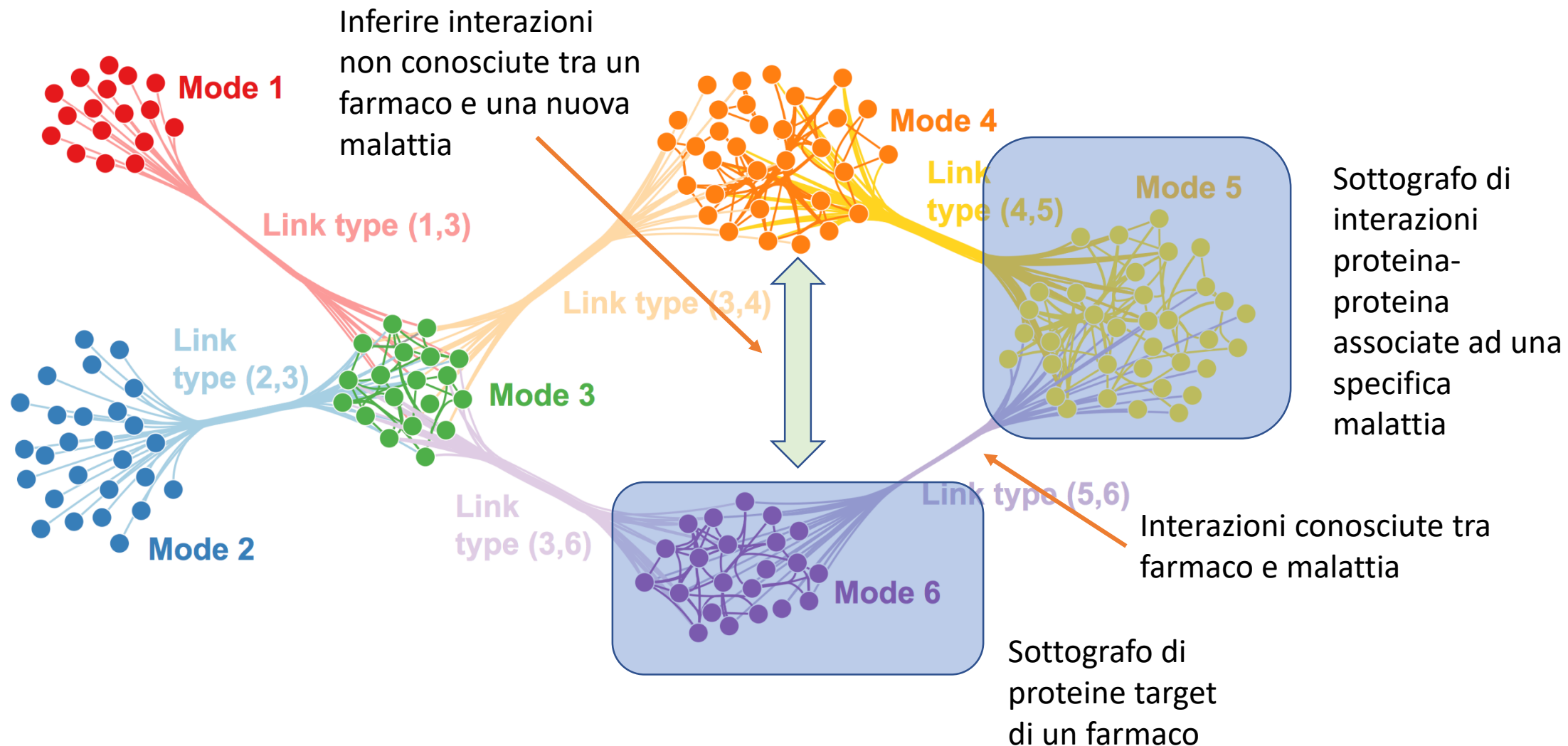
r_2 Bradycardia side effect

▲—● Drug-protein interaction

●—● Protein-protein interaction

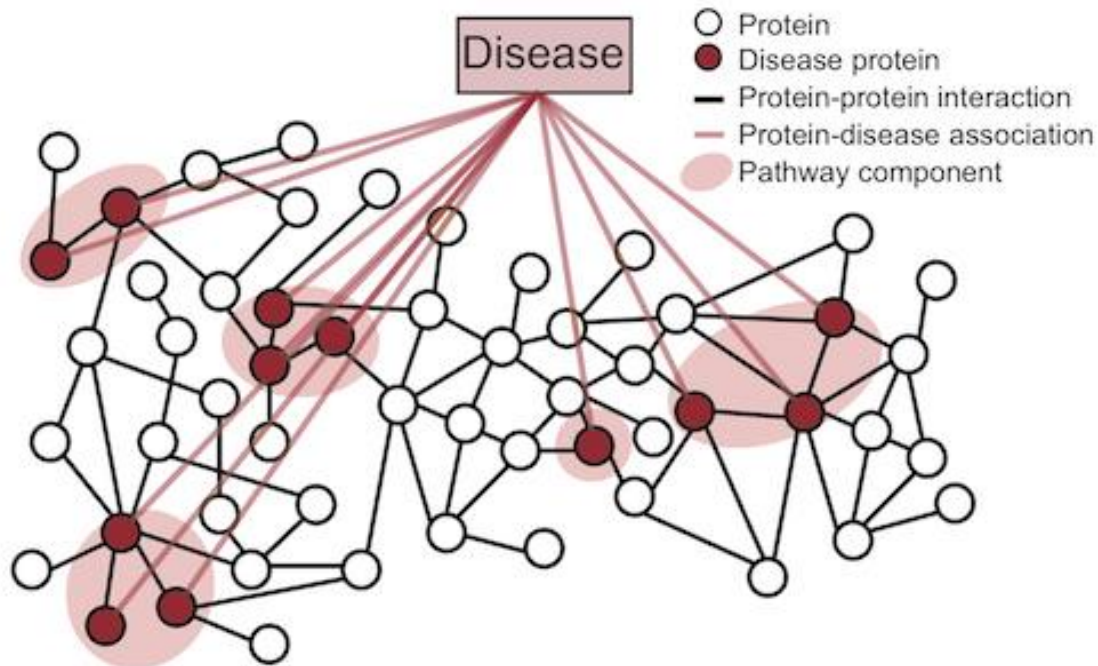
Zitnik, Agrawal, Leskovec, *Bioinformatics* 2018

Riposizionamento di Farmaci (2)



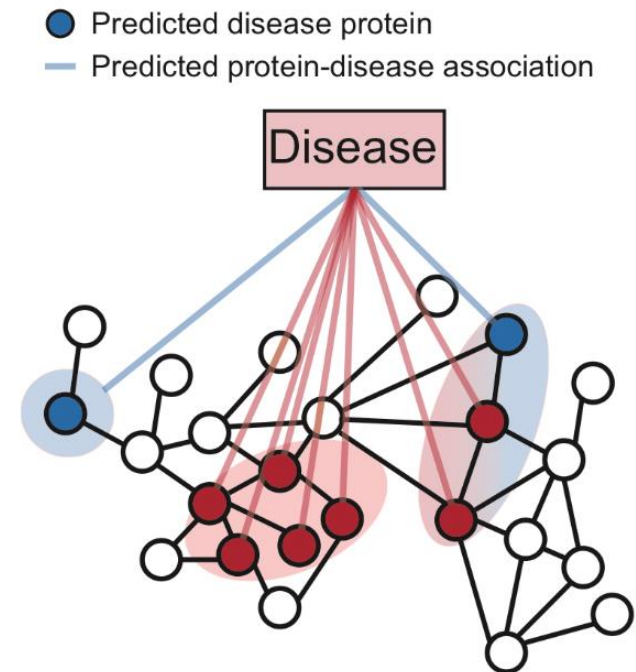
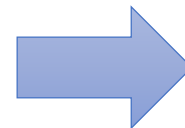


Predizione Percorsi di Co-Regolazione (2)



Rete delle interazioni conosciute

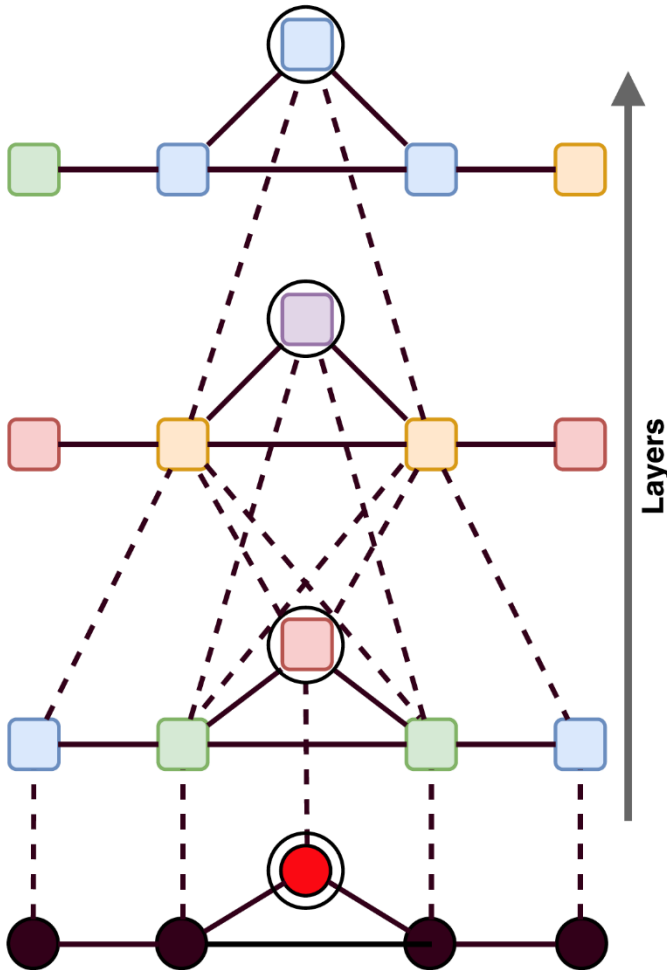
- proteine-proteine
- proteine-malattia



Predizione interazioni sconosciute tra
proteine e malattia



Approccio Contestuale e Neuro-Probabilistico (1,2)



Contestuale

- Efficacia e scalabilità da piccole molecole a grandi reti di interazione
- Addestramento: 25 vertici per millisecondo
- Predizione: 230 vertici per millisecondo

Neuro probabilistico

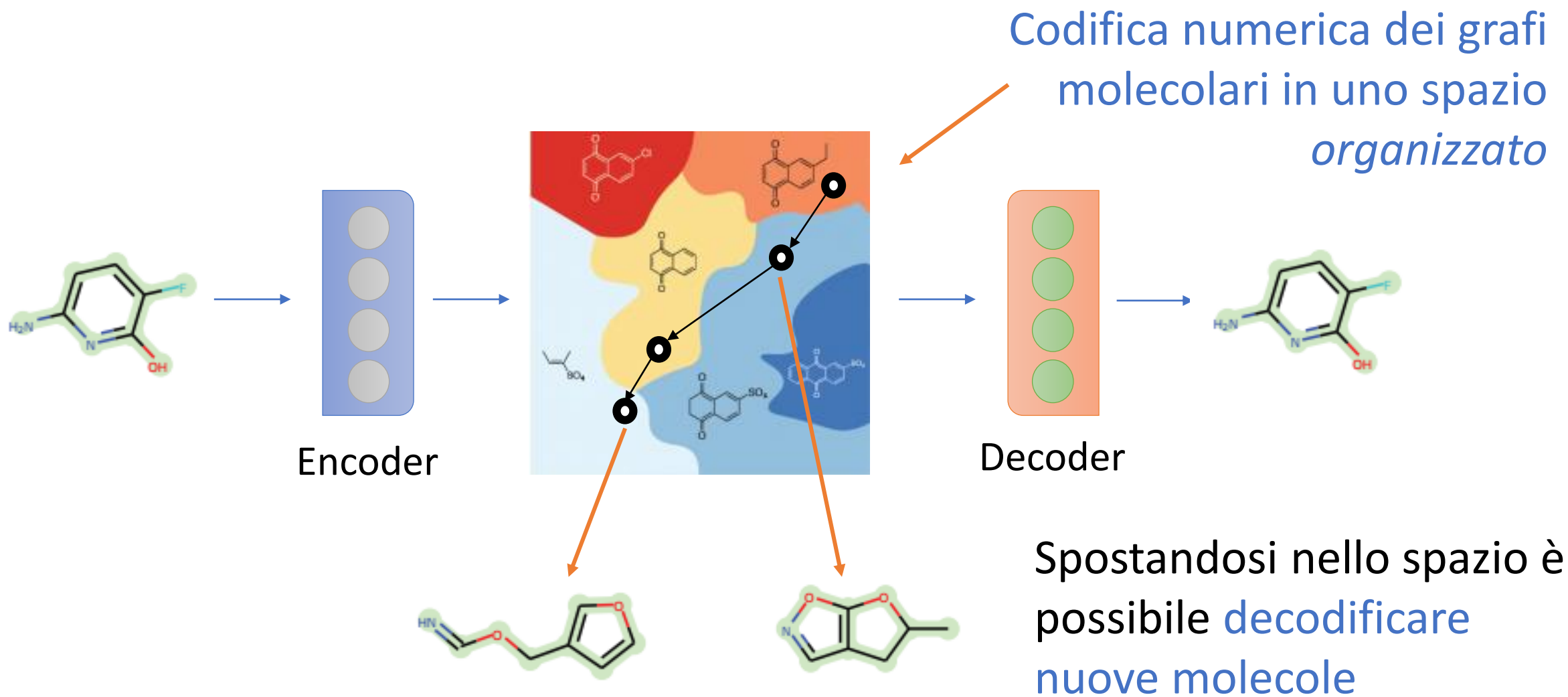
- Approccio deep learning con formalizzazione probabilistica
- Interpretabilità del modello
- Inferenza diffusa e built-in



UNIVERSITÀ DI PISA

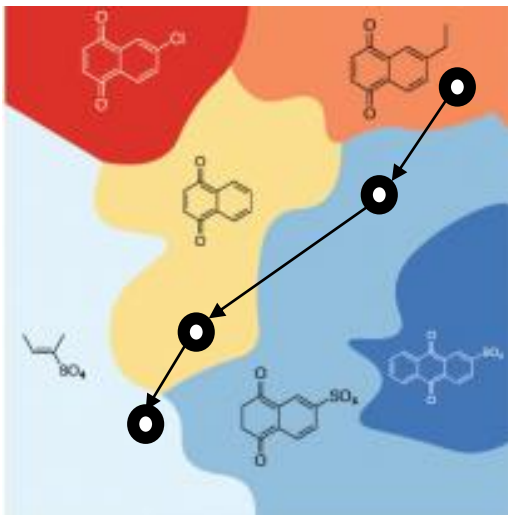


Generazione Molecole e Materiali (3)



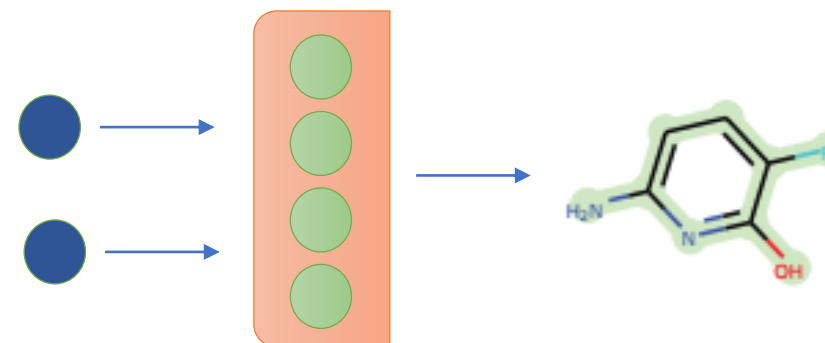


Generazione Condizionale e Controllata (3)



- Spostandosi nello spazio della rappresentazione numerica non sempre si ottengono grafi “buoni”
- Modelli che controllano la **bontà dei grafi durante il processo di generazione**

Generazione grafi **condizionata** da informazioni di input





CONCLUSIONI



Conclusioni

- I grafi non sono solo una struttura dati ma un dato che **cattura informazione relazionale**
- Processare l'informazione relazionale **migliora la capacità predittiva** rispetto ad un approccio non-strutturato
- Applicazioni sostanziali alle **scienze della vita**
 - Design molecole e farmaci, inferenza di vie metaboliche, interazioni (proteine, farmaci,...)
 - Brain networks, genomica, epidemiologia, ...
- Uso **generativo** del Deep learning



Risorse & Riferimenti

- Tutorial @UNIFI
 - <https://sites.google.com/view/dl4sd>
- Contextual graph Markov model
 - <https://github.com/diningphil/CGMM>
 - <http://arxiv.org/abs/1805.10636>

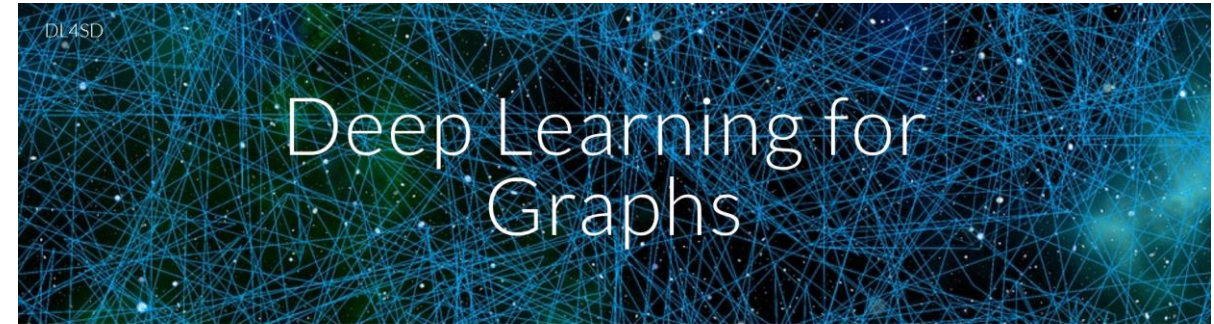
(Piccola) Bibliografia

1. Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, George E. Dahl, Neural Message Passing for Quantum Chemistry, ICML 2017
2. Marinka Zitnik, Monica Agrawal and Jure Leskovec, Modeling polypharmacy side effects with graph convolutional networks, Bioinformatics 2018
3. Martin Simonovsky, Nikos Komodakis, GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders, NIPS Workshop, 2017

INNS BIG DATA AND DEEP LEARNING 2019

April 16 – 18, Sestri Levante, Italy

<https://innsbddl2019.org/>



ECML-PKDD 2018 Tutorial - 14th September 2018

Organized by:

[Davide Bacciu](#) - Università di Pisa (bacciu@di.unipi.it) - Contact person

[Alessio Micheli](#) - Università di Pisa (micheli@di.unipi.it)



GRAZIE PER L'ATTENZIONE



 EVENTO **DATA FOR**
2018
BIGDATATECH **HUMAN**



www.bnova.it