

Docenti: **Giorgio Valentini**  
**Matteo Re**

UNIVERSITÀ DEGLI  
STUDI DI MILANO



C.d.I. Informatica

# Bioinformatica

A.A. 2012-2013 semestre I

## modulo 2 – parte B.2

**1**

**Inferenze in reti biologiche**

---

Rappresentazioni basate su reti sono in grado di modellare molti tipi di oggetti/fenomeni osservabili nel mondo reale :

- reti **tecnologiche** : www, internet, circuiti elettrici,...
  - reti **sociali** : amicizie, collaborazioni, diffusione malattie,...
  - reti **biologiche**: struttura proteine, regolazione trascrizionale geni, reti metaboliche, protein-protein interaction (PPI), ...
-

Sono parte integrante di una disciplina denominata “systems biology” che si occupa, come suggerisce il nome, di studiare la vita non dal punto di vista della caratterizzazione (anche a livello molecolare) dei suoi singoli componenti ma dal punto di vista delle **interazioni** tra questi componenti, in modo da ottenere, appunto, una visione “di sistema”.

Esistono diverse aree di ricerca che si avvalgono di tecniche di analisi di reti biologiche. Esse si basano sull'utilizzo di dati “omici” :

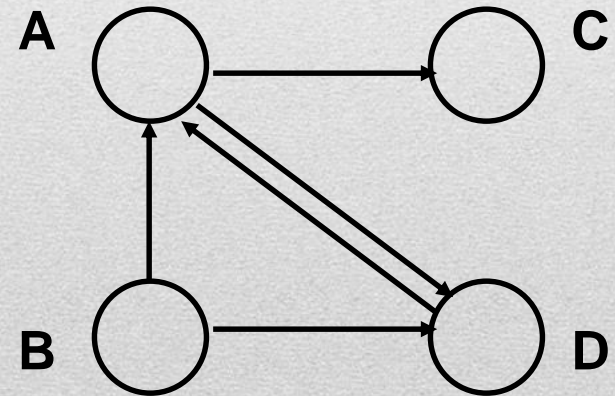
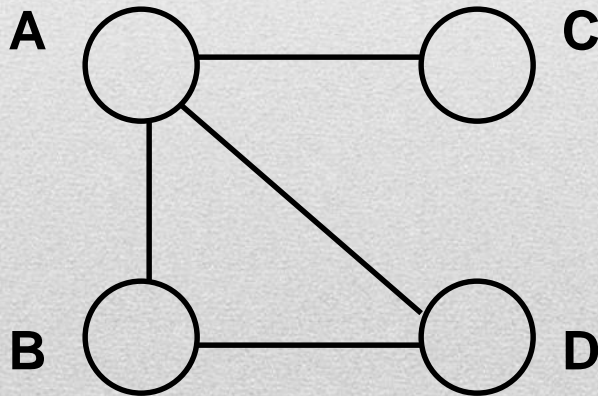
- **genOMA** ... genomica
  - **proteOMA** ... proteomica
  - **metabolOMA** ... metabolomica
-

Obiettivi systems biology:

- Comprensione a livello di sistema dei sistemi biologici (organismi)
  - Studio delle interazioni tra componenti in modo da individuare proprietà emergenti (proprietà che esistono solo considerando insiemi di componenti e che non sono riscontrabili in nessun singolo componente)
  - **Obiettivo delle inferenze in reti biologiche:** estrazione di nuova conoscenza biologica dalle caratteristiche topologiche delle reti biologiche.
-

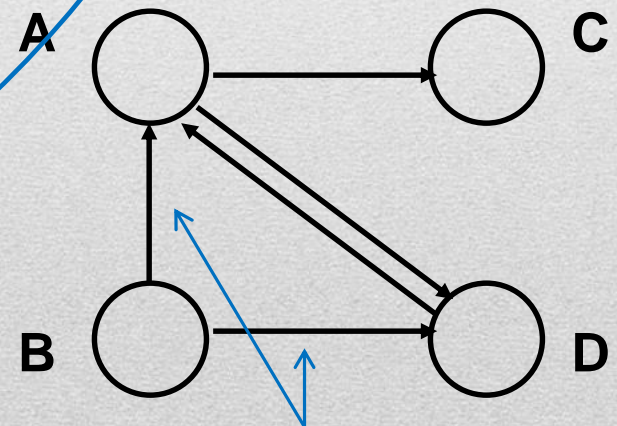
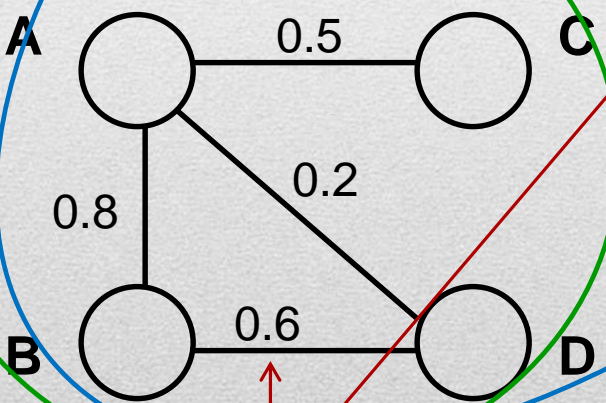
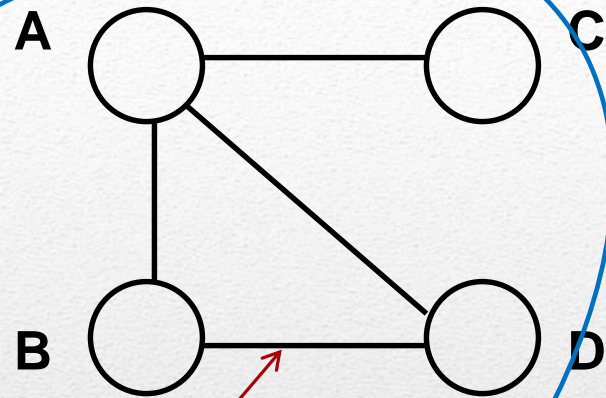
Una rete (descritta matematicamente mediante un **grafo**) è un oggetto composto da un set di **nodi** (o vertici) e da un set di **archi** :  $G=(V,E)$

Gli **archi** descrivono relazioni tra i nodi presenti nella rete.



Esistono diversi tipi di grafi/reti :

Rete pesata

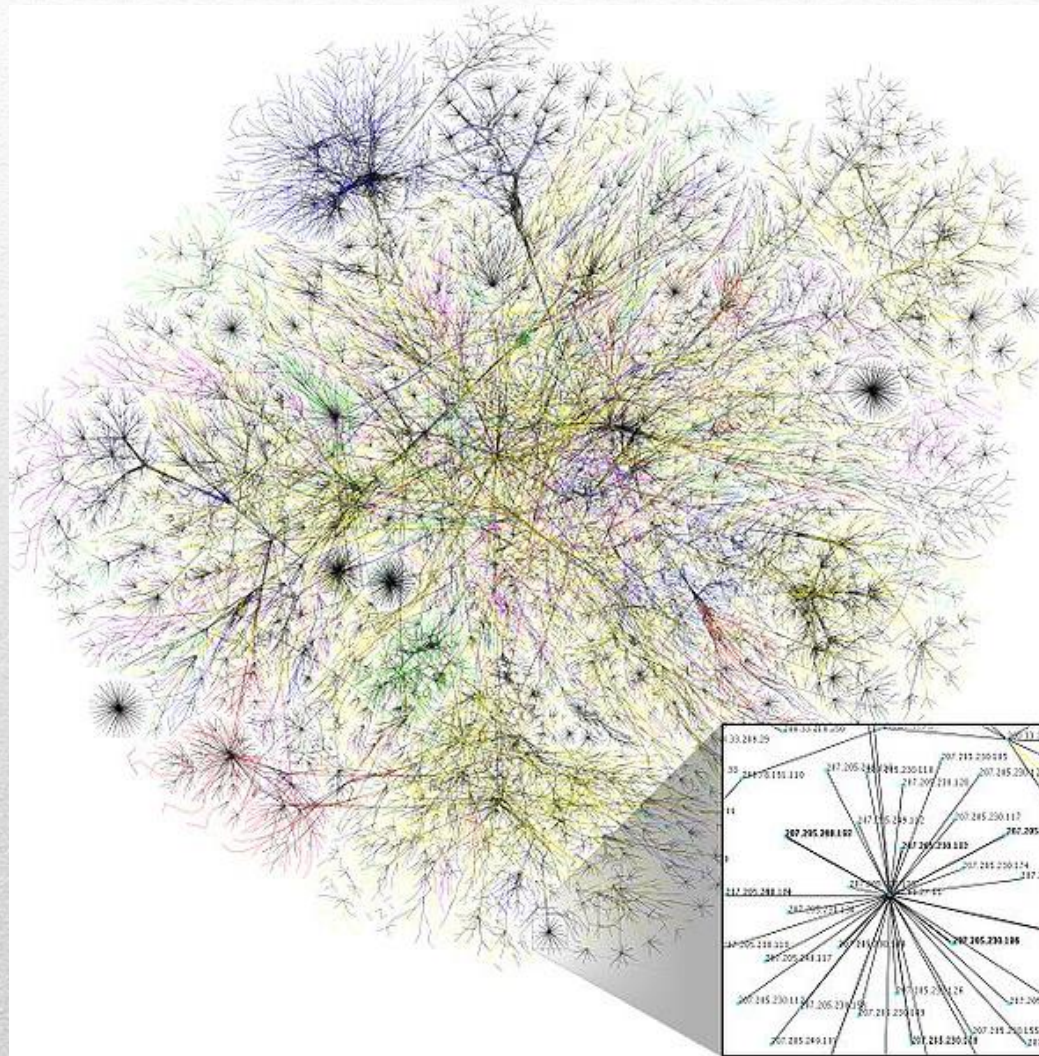


Grafi NON diretti

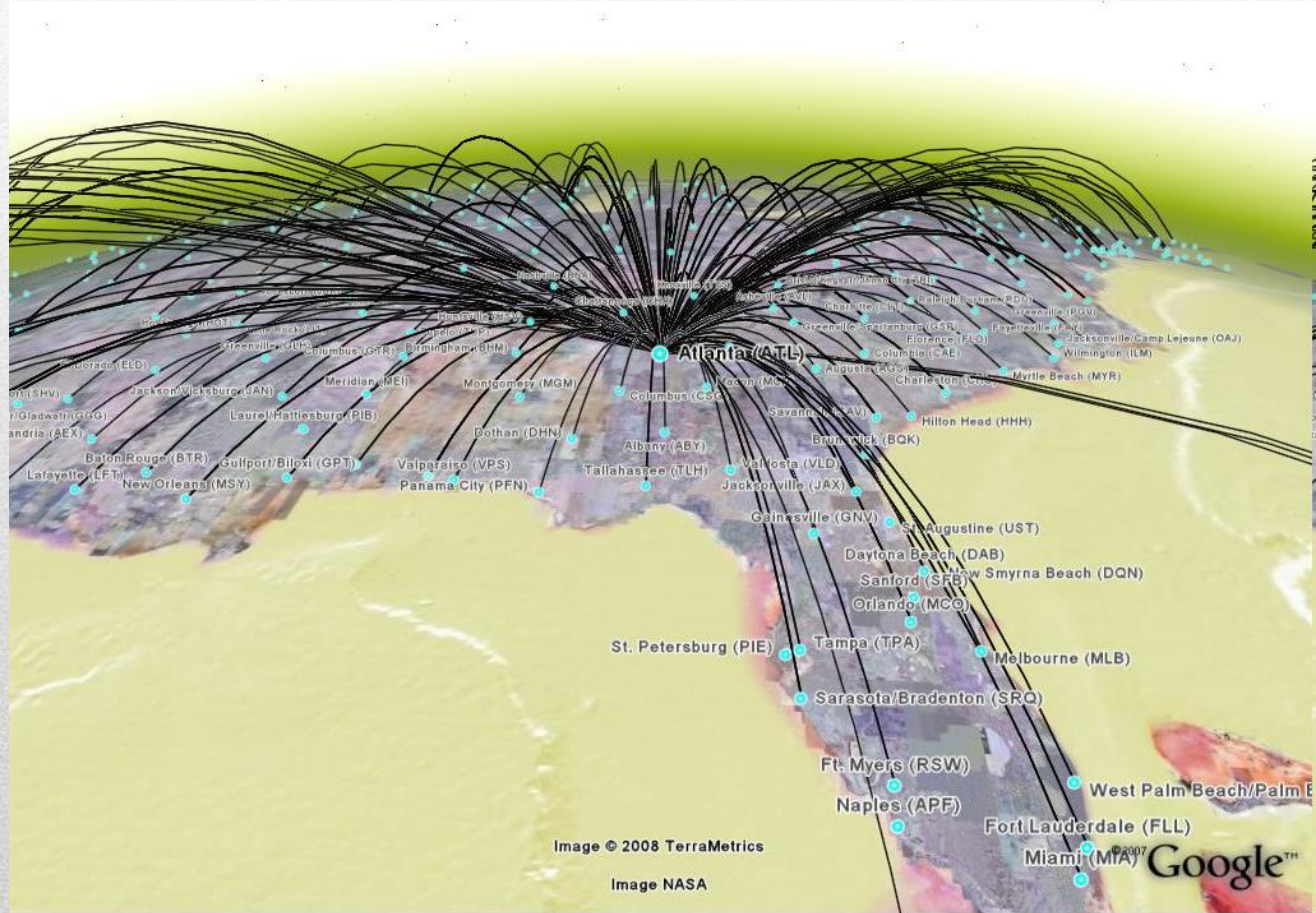
Grafi DIRETTI

# QUANTI E QUALI TIPI DI RETI ESISTONO?

Internet



## Linee aeree





## RETI INTRA-CELLULARI

- Reti metaboliche
- Reti di regolazione trascrizionale
- Reti di propagazione dei segnali
- Reti di interazione tra proteine (PPI)
- Reti di struttura delle proteine

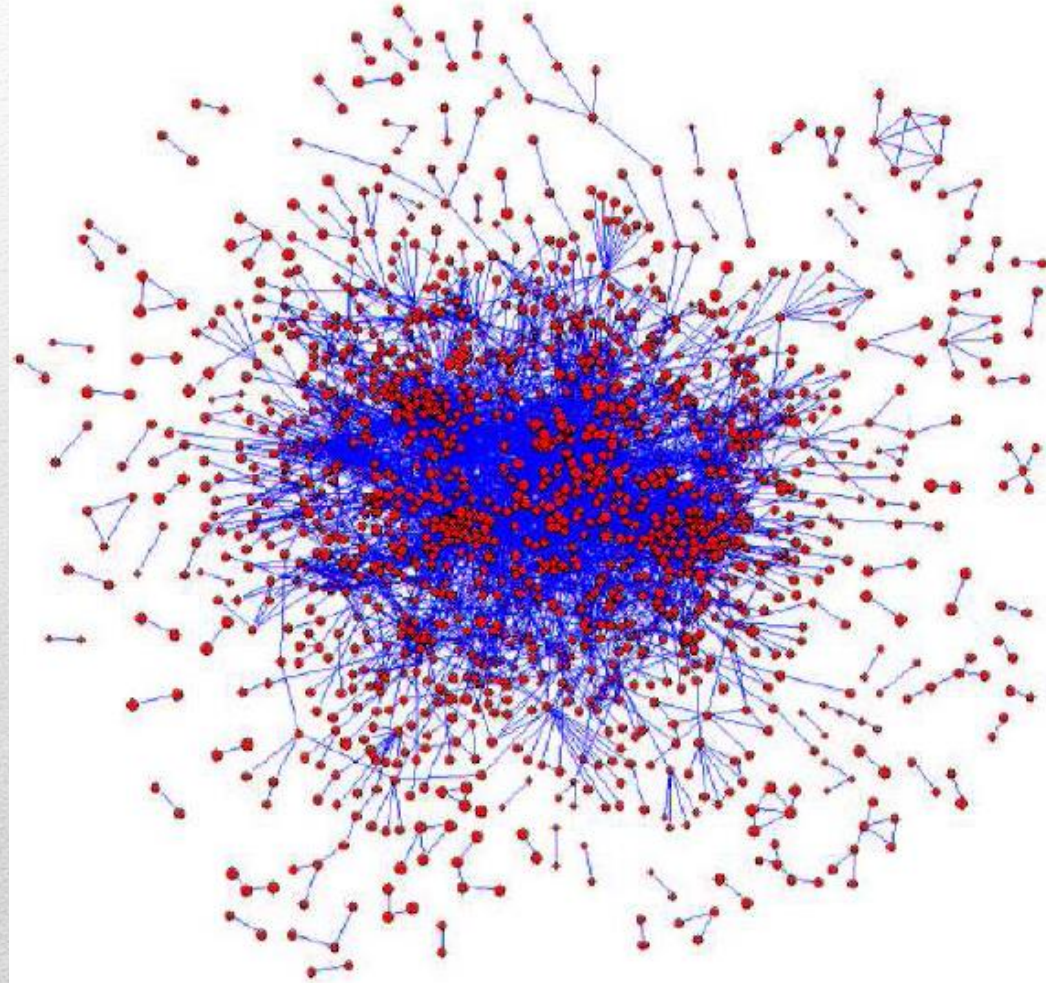
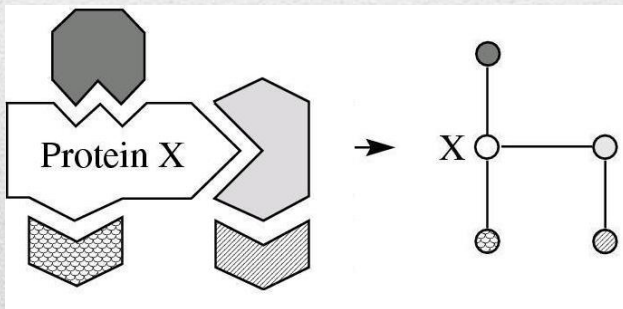
## ALTRI TIPI DI RETI BIOLOGICHE

- Reti di connessione sinaptica tra neuroni
  - Reti funzionali cerebrali
  - Reti alimentari (ecologia)
  - Reti filogenetiche
  - Reti di correlazione (es, espressione genica)
  - Reti di associazione tra geni e malattie
  - Interazione farmaco-bersaglio (“drug target” networks)
-

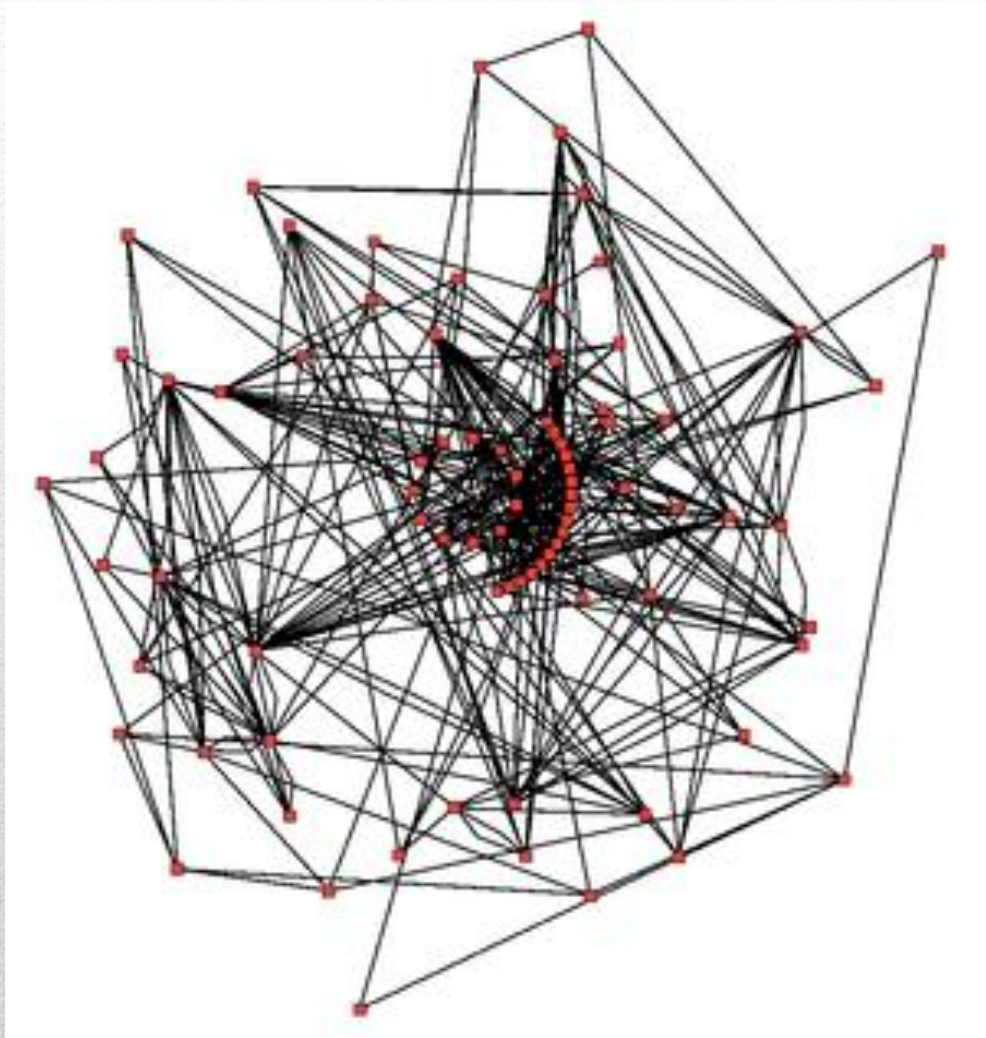
**Protein structure network  
(a.k.a. Residue Interactions graphs) :**



## Protein-protein Interaction networks (PPI) :

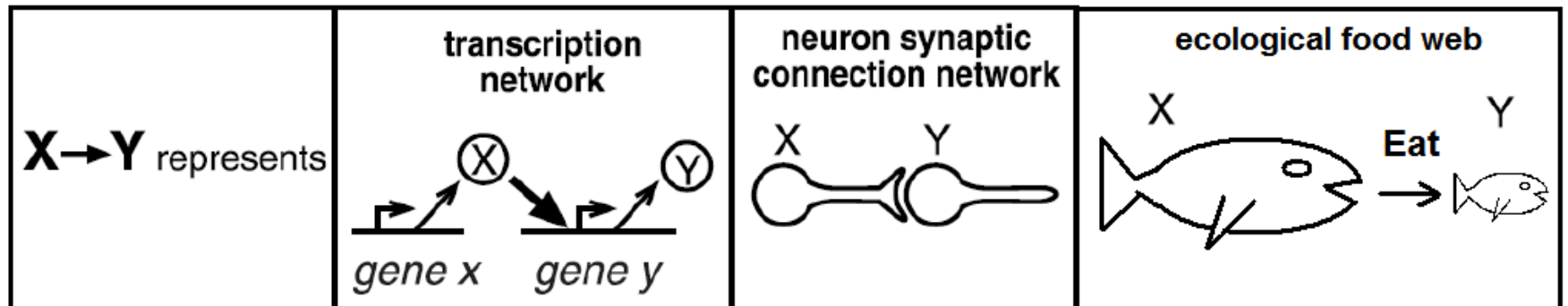


Metabolic networks :



Metabolic network of *A. thaliana*

E molti altri ...



## RETI INTRA-CELLULARI

- Reti metaboliche
- Reti di regolazione trascrizionale
- Reti di propagazione dei segnali
- Reti di interazione tra proteine (PPI)
- Reti di struttura delle proteine

Descrivono il funzionamento della cellula a **vari livelli** e spesso sono **“sovrapposte”**

---

## RETI INTRA-CELLULARI

- In funzionamento delle cellule è basato su un set molto ampio di **interazioni altamente interconnesse tra diversi tipi di molecole** (ad es. proteine, DNA, RNA, metaboliti).
  - Svariate attività cellulari sono controllate dall'azione di molecole su altre molecole.
  - Gli elementi fondamentali di queste reti di interazione molecolari sono indubbiamente le **proteine**.
-

## Reti metaboliche :

Vengono utilizzate per studiare e modellare il metabolismo.

Permettono di studiare l'insieme delle reazioni biochimiche che permettono agli organismi di :

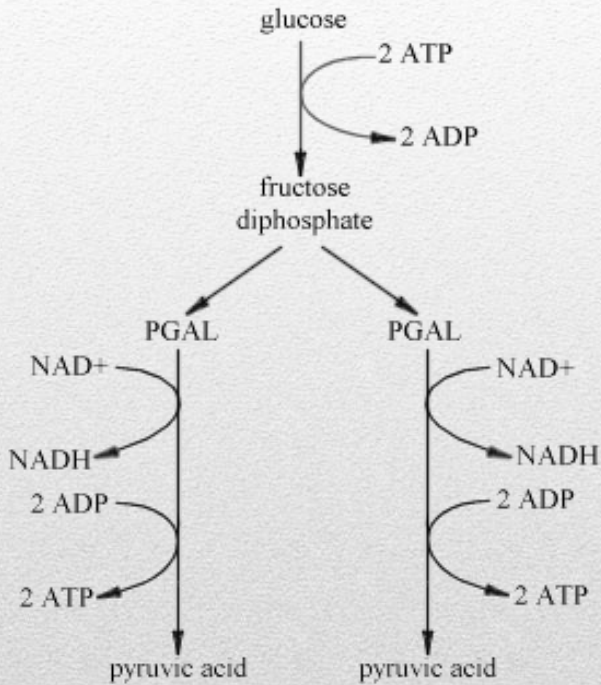
- Rispondere a stimoli dell'ambiente esterno (“esterno” ha senso generale ... ad esempio “esterno della cellula”)
  - Crescere
  - Mantenere la loro struttura
  - Mantenere uno stato di equilibrio dinamico lontano da quello (statico) dell'ambiente esterno e necessario per la loro sopravvivenza . Questo processo è detto **omeostasi**.
-



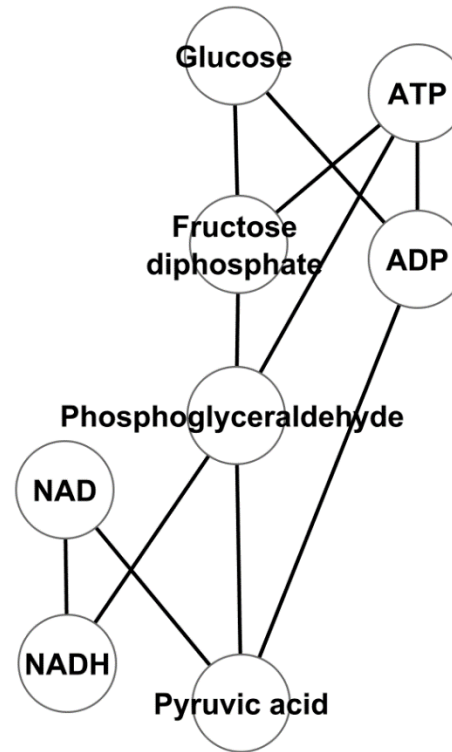
**Reti metaboliche** : elementi presenti nelle reti metaboliche

- **Metaboliti** : piccole molecole come glucosio (zucchero) o aminoacidi
  - **Vie metaboliche** (o pathways) : serie ordinate di reazioni biochimiche che realizzano una specifica funzione. Ad ogni passo un metabolita viene convertito in un altro metabolita. Ognuno di questi passaggi è realizzato da una specifica proteina che serve unicamente a facilitare questa trasformazione chimica. Tale proteina è detta **enzima**.
  - **In una rete metabolica.:** I metaboliti sono **NODI**, le reazioni sono **ARCHI** (caso semplice) oppure abbiamo due tipi di nodi, I metaboliti e gli enzimi (caso più complesso).
-

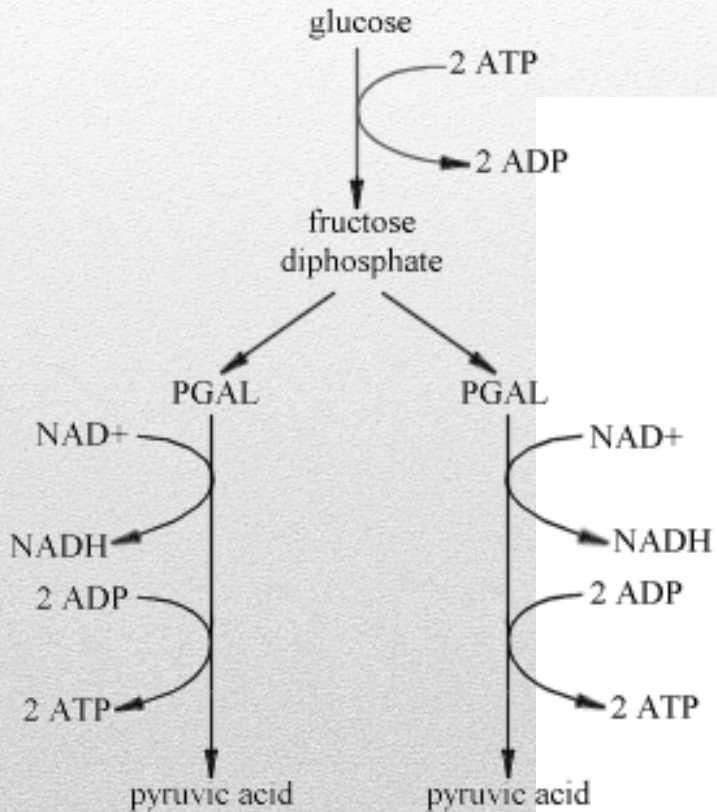
# Glycolysis



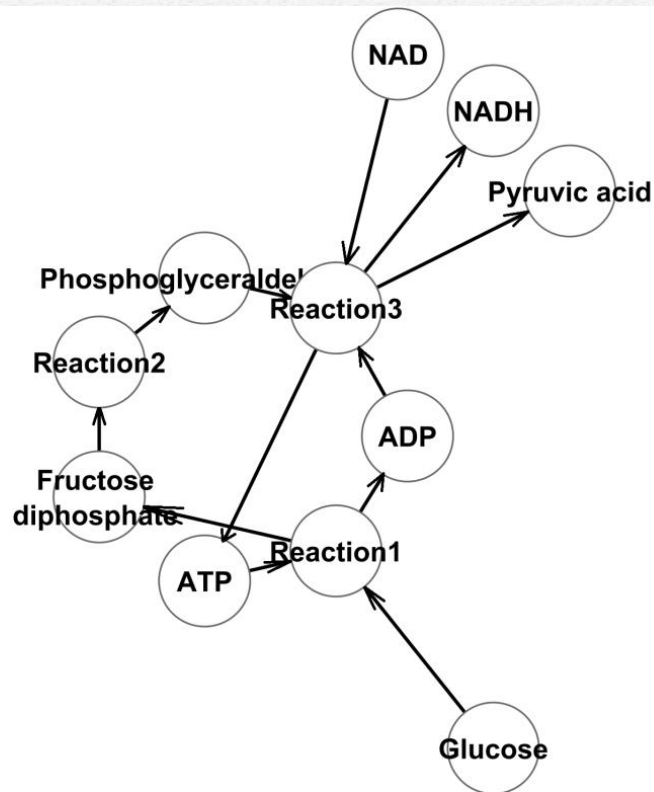
## Metabolite-centric representation:



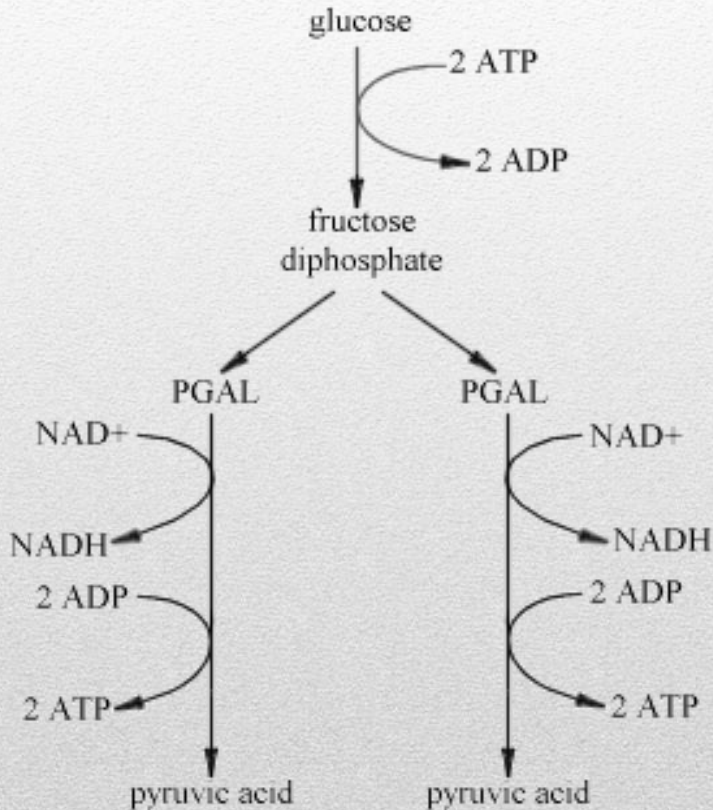
# Glycolysis



## Reactions + metabolites:



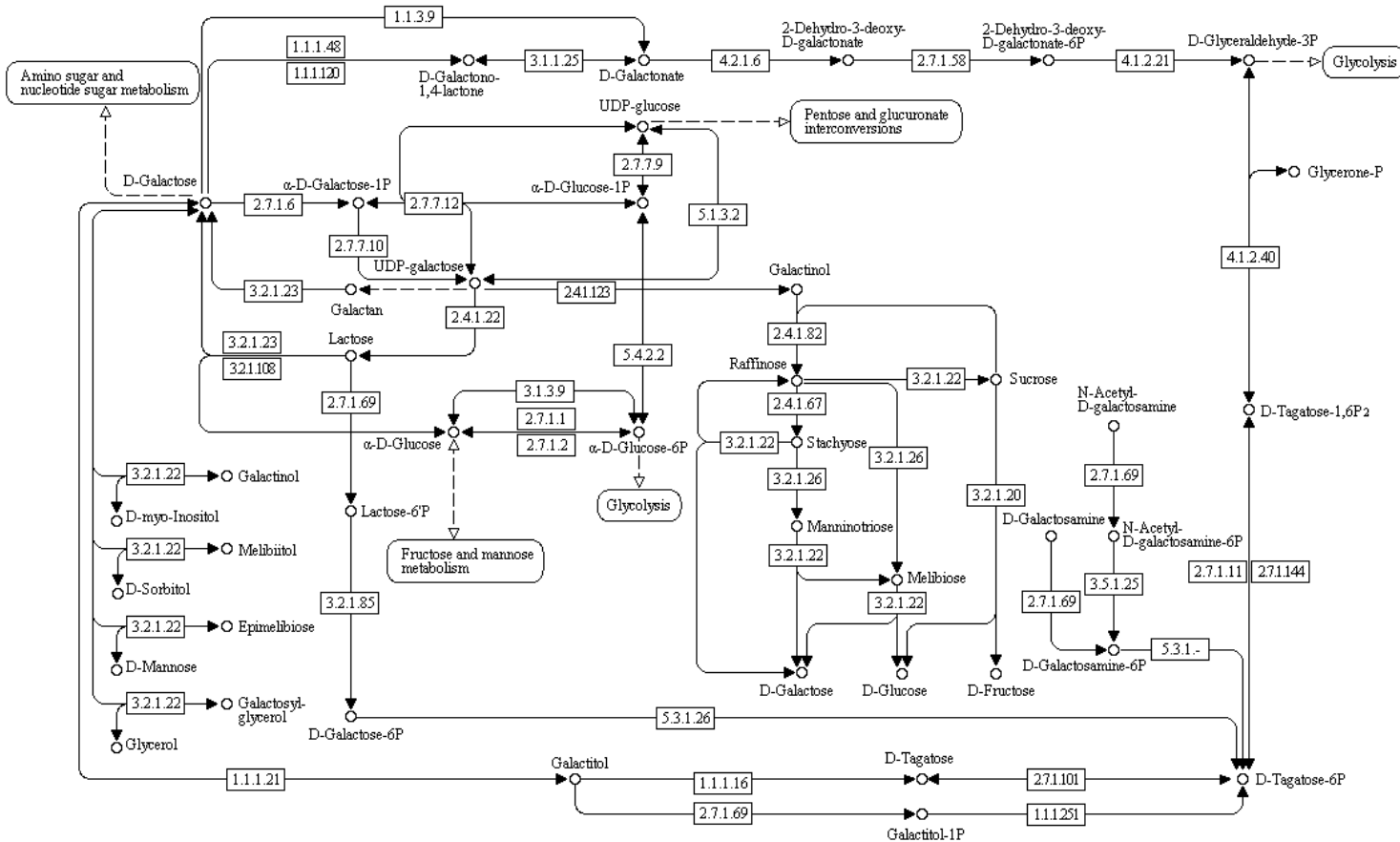
# Glycolysis



**Glicolisi non è una via metabolica molto estesa ...**  
ne esistono altre più complesse  
(con più reazioni/passi)

metabolismo del galattosio (un altro zucchero)

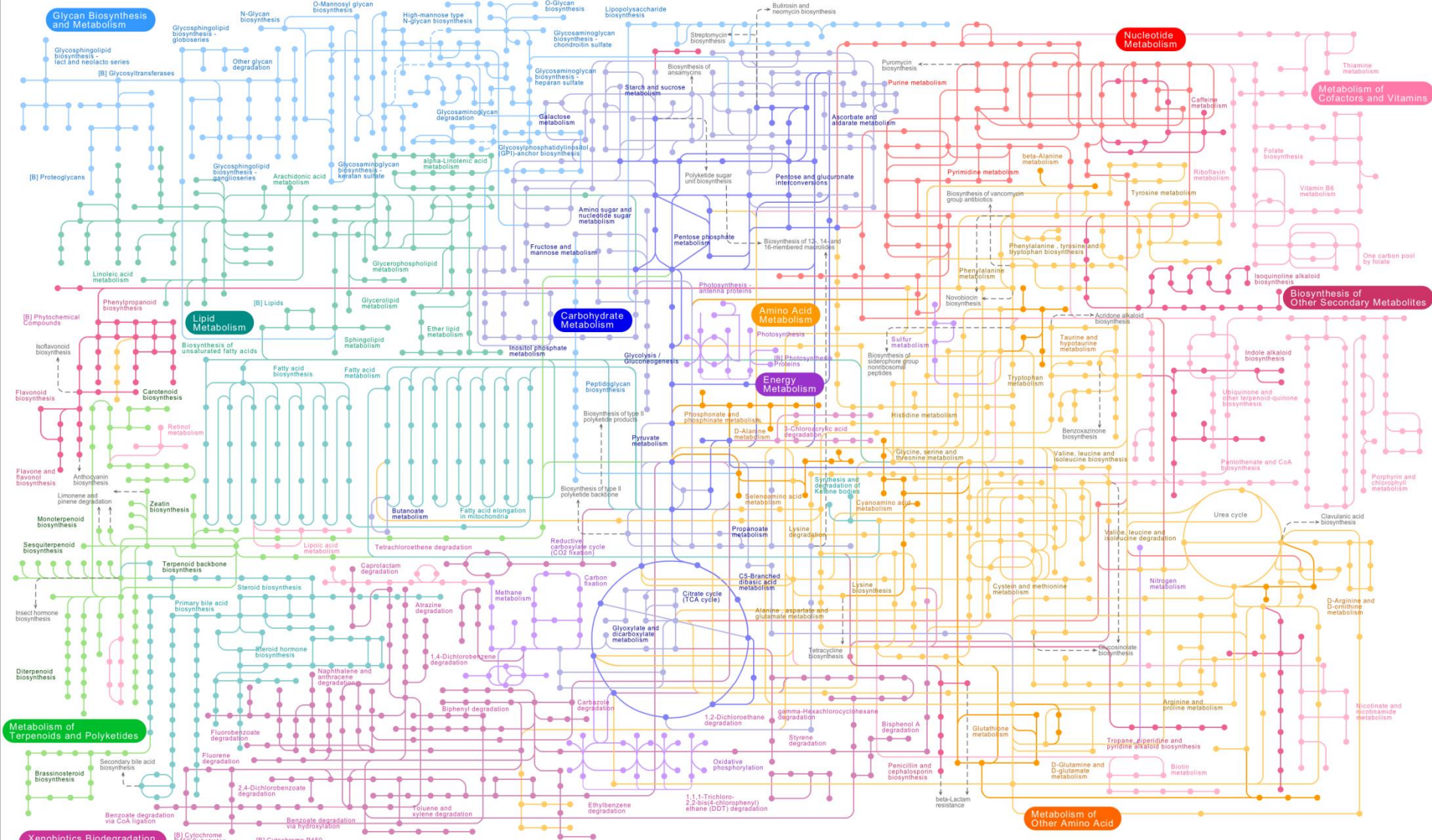
GALACTOSE METABOLISM



Ma quante vie metaboliche esistono in un organismo (ad esempio *H.sapiens*)?

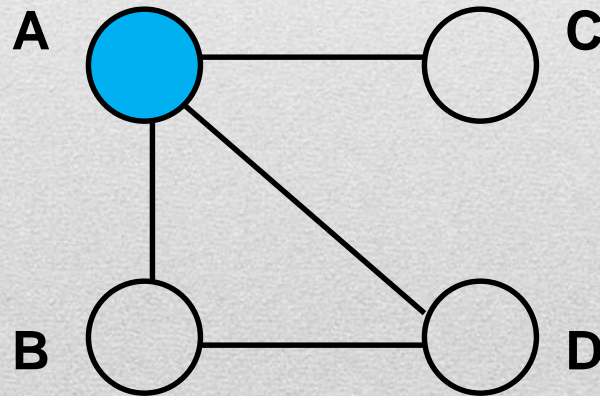
# Bio

## Reazioni metabolismo umano (REACTOME)



Una rete (descritta matematicamente mediante un **grafo**) è un oggetto composto da un set di **nodi** (o vertici) e da un set di **archi** :  $G=(V,E)$

Gli **archi** descrivono relazioni tra i nodi presenti nella rete.



1. Dato un grafo ed un punto di inizio (nodo) scegliamo in modo casuale uno dei vicini (nodi **direttamente** connessi) del nodo corrente e ci muoviamo su di esso. Il nodo di arrivo diventa il nodo corrente.

2. Scegliamo in modo casuale uno dei vicini (nodi **direttamente** connessi) del nodo corrente e ci muoviamo su di esso. Il nodo di arrivo diventa il nodo corrente.

La sequenza di nodi che ci ha portato al passo  $t$  nel nodo  $v_i$  è detta **RANDOM WALK** (cammino casuale) sul grafo.

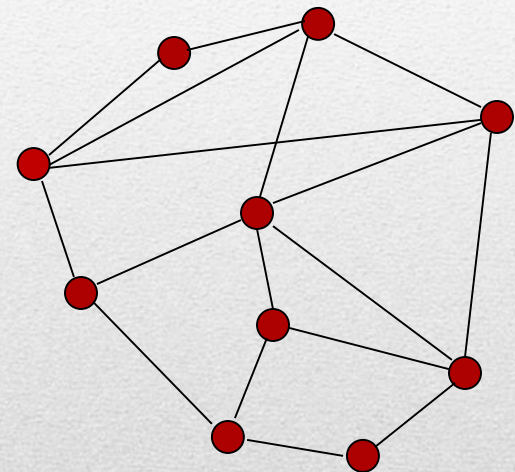
ripeti fino al raggiungimento di una condizione di arresto

---



In un grafo non diretto  $G$ :

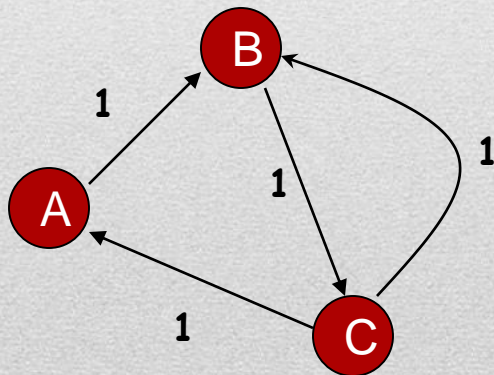
- Iniziando da un nodo  $v_0$
- Ripeti molte volte:
  - Spostati su un vicino (scelto a **caso**) del nodo corrente.



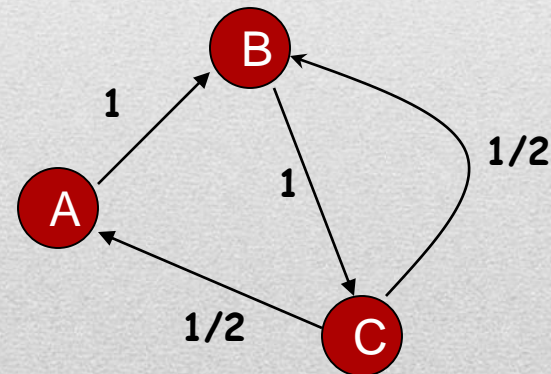
La sequenza di nodi che ci ha portato al passo  $t$  nel nodo  $v_i$  è detta RANDOM WALK (cammino casuale) sul grafo.

	A	B	C
A	0	1	0
B	0	0	1
C	1	1	0

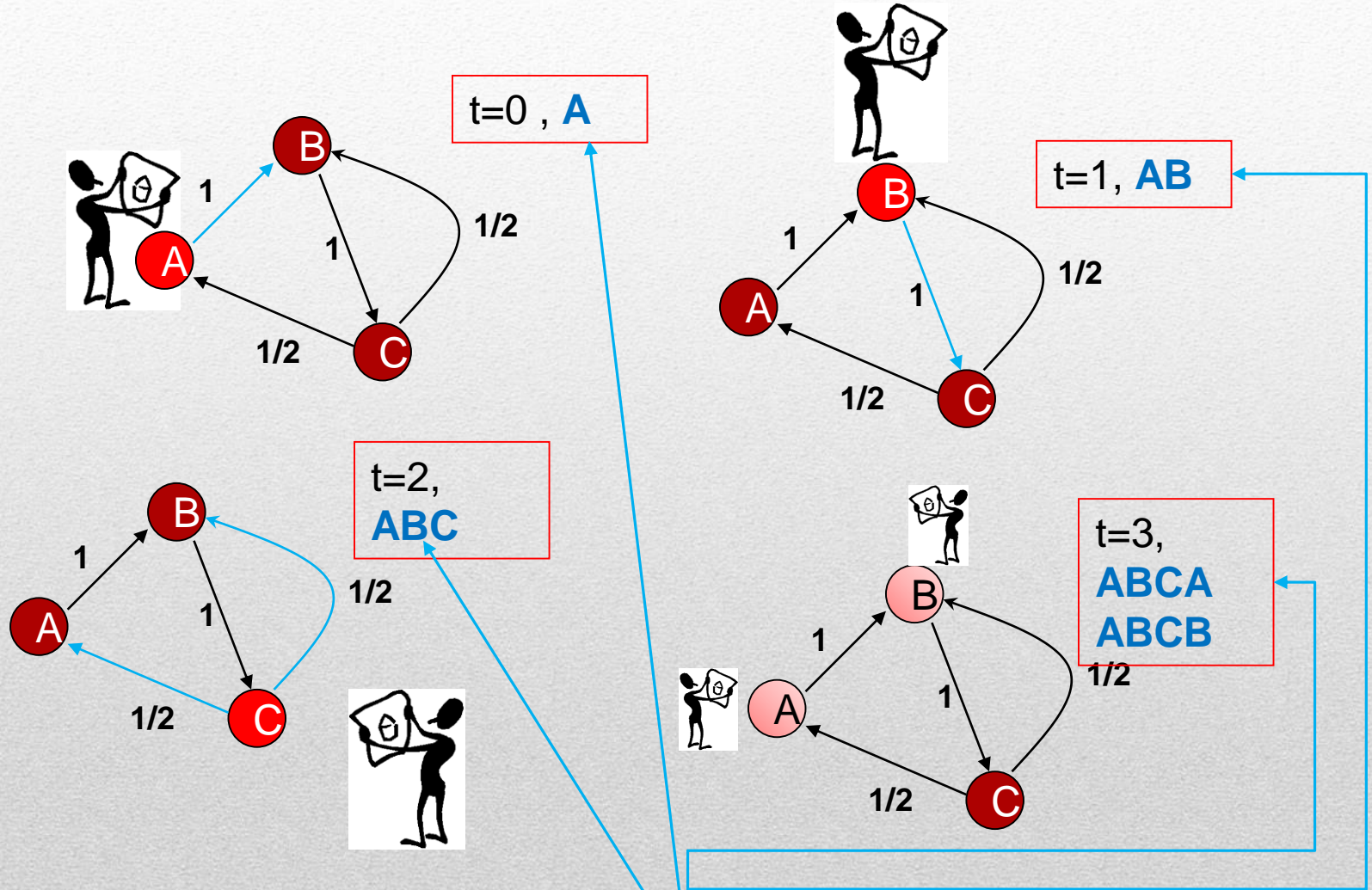
Matrice di adiacenza A



	A	B	C
A	0	1	0
B	0	0	1
C	1/2	1/2	0

Matrice di transizione P

**NB: somma probabilità archi (transizioni) uscenti deve essere uguale a 1 !**



Cammini casuali

- $x_t(i) = \text{Prob}(v_t = i) =$  **probabilità di essere al nodo  $i$  al tempo  $t$**
- $x_{t+1}(i) = \sum_j (\text{probabilità di essere al nodo } j) * \text{Pr}(j \rightarrow i)$   
 $= \sum_j x_t(j) * P(j, i)$
- $x_{t+1} = x_t P = x_{t-1} * P * P = x_{t-2} * P * P * P = \dots = \mathbf{x_0 P^t}$

Cosa succede alla distribuzione delle probabilità associate ai nodi del grafo se continuiamo a camminare per un numero **consistente** di passi? .

**I cammini casuali sul grafo sono**  
**catene di Markov !!!**

- Per definizione se siamo in stato di equilibrio  $\mathbf{x}_{t+1} = \mathbf{x}_t$
  - Nel tipo di grafi che consideriamo in queste lezioni la distribuzione stazionaria associata allo stato di equilibrio **non dipende dalla distribuzione delle probabilità iniziale ( $\mathbf{x}_0$ ).**
-

Input:

- the adjacency matrix  $\mathbf{W}$  of a graph  $G=\langle V,E\rangle$
- A subset of nodes  $V_C$  having property  $C$

• Initialization of nodes:

if  $v \in V_C$  then  $p_0(v) = 1 / |V_C|$  else  $p_0(v)=0$

• Set transition matrix:  $\mathbf{Q} = \mathbf{D}^{-1}\mathbf{W}$

where  $\mathbf{D}$  is a diagonal matrix with

$$d_{ii} = \sum_j w_{ij}$$

• Iteratively update until convergence or until  $t=k$

$$\mathbf{p}_t = \mathbf{Q}^T \mathbf{p}_{t-1}$$

Output:  $\mathbf{p}_t$

---

- Classificazione funzionale di geni
- Predizione di associazione tra geni e malattie
- Riposizionamento di farmaci
- ...

Nella prossima lezione utilizzeremo il random walk per realizzare un esempio pratico di riposizionamento di farmaci.

---

Docenti: **Giorgio Valentini**  
**Matteo Re**

UNIVERSITÀ DEGLI  
STUDI DI MILANO



C.d.I. Informatica

# Bioinformatica

A.A. 2012-2013 semestre I

## modulo 2 – parte B.2

**2**

**Applicazione random walk:  
riposizionamento farmaci**

---





Il ciclo di sviluppo di un nuovo farmaco pone diversi tipi di problemi :

- L'azione terapeutica di una molecola dipende dalla sua capacità di interagire con una o più molecole presenti in cellula il cui comportamento anomalo determina uno stato patologico.
  - Per progettare un farmaco, quindi, è necessario comprendere la patologia ad un livello di dettaglio molto alto (livello molecolare)
  - Per molte malattie questo livello di conoscenza non è attualmente disponibile.
-

Il primo passo nel ciclo di sviluppo di un farmaco consiste nell'identificare una **molecola** in cellula che abbia queste caratteristiche:

- Il suo coinvolgimento nella patologia che si vuole curare è noto a priori o, almeno, è supportato da dati preliminari pubblicati in letteratura (o osservazioni sperimentali ottenute in laboratorio).
- Interferenze nel suo comportamento dovute ad un agente esterno (il farmaco che si vuole progettare) devono avere effetti SOLO sullo stato patologico che si vuole curare ma **NON DEVONO INTERFERIRE** con le funzioni metaboliche che permettono la vita dell'organismo (altrimenti otterremmo un **veleno**, e non un farmaco)

---

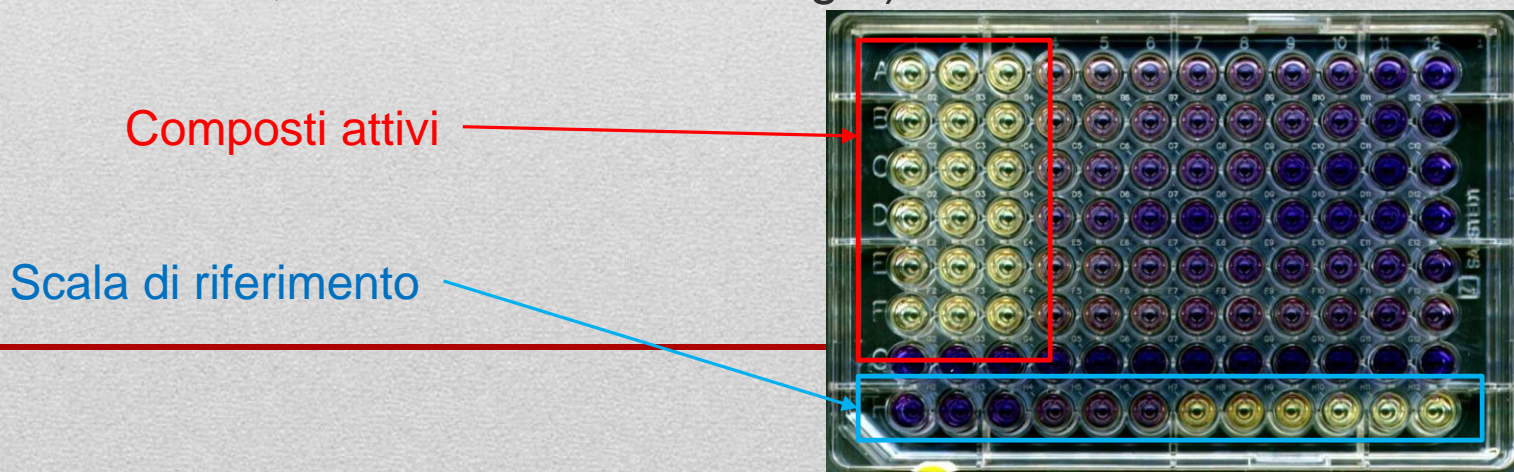
**QUESTA MOLECOLA VIENE DETTA TARGET**

## FASI DEL CICLO DI SVILUPPO: saggi di attività sul target di librerie di molecole

Una volta identificato un potenziale target su cui lavorare inizia una serie di esperimenti su vasta scala che hanno come scopo quello di identificare il maggior numero di piccole molecole (potenziali farmaci) che hanno un effetto (quantificabile) su di esso.

Per ottenere questo risultato vengono utilizzati delle librerie di piccole molecole costruite mediante tecniche di chimica combinatoria (altissimo numero di potenziali farmaci).

Ognuno di essi viene inserito in un piccolo contenitore contenente il target (in soluzione acquosa di solito). Tutto il processo (aggiunta potenziale farmaco , test dell'attività sul target) è robotizzato.



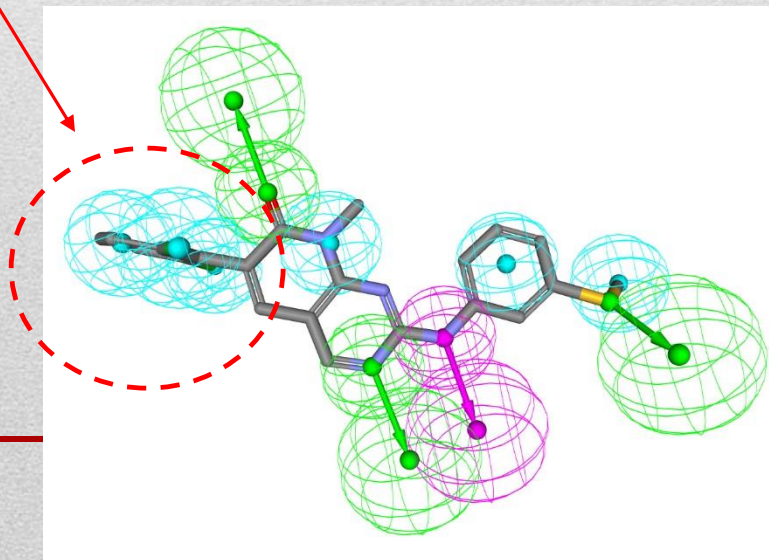
Vengono esaminati tutti i potenziali farmaci per cui è possibile rilevare attività nei confronti del target (tali molecole si indicano come **hits**).

Si cercano i migliori hits :  ( i composti «attivi» )

- Attivi a **bassa** concentrazione
  - Caratteristiche chimico/fisiche che permettano una **facile somministrazione**
  - Processo di sintesi chimica **poco oneroso**
-

Tra tutti gli hit disponibili la molecola con caratteristiche migliori diventa il **LEAD** ossia la molecola che entrerà negli stadi successivi del processo di sviluppo.

Tutte le **parti della struttura chimica** del lead vengono modificate per identificare quale parte della molecola è la **responsabile** dell'effetto terapeutico (questa parte di molecola è detta **farmacoforo**) e quali parti possono essere modificate per abolire eventuali effetti indesiderati.

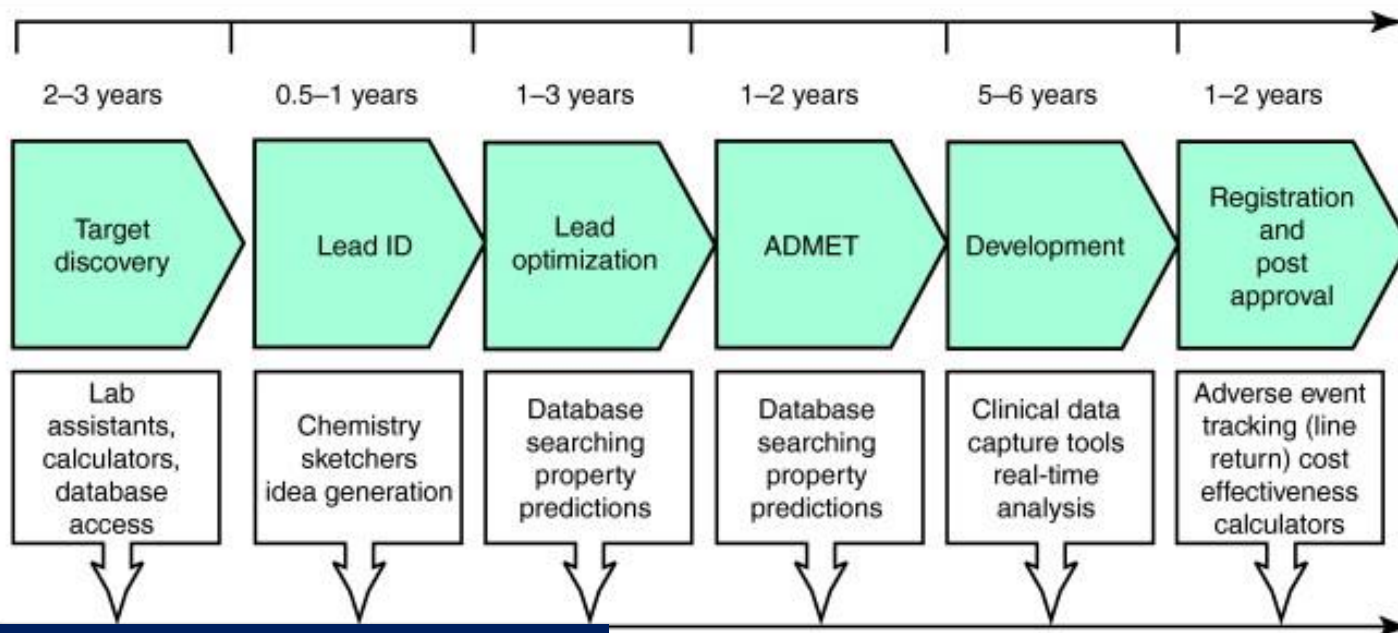


Per ogni versione «ottimizzata» del lead è necessario effettuare test in vivo tesi a determinare alcune caratteristiche del composto:

- **A**ssorbimento
- **D**istribuzione
- **M**etabolismo
- **E**scresione
- **T**ossicità

Da qui il nome di test «ADMET»

Le fasi di ottimizzazione del lead e i test ADMET vengono ripetute molte volte ciclicamente (a volte si cambia anche lead) fino ad ottenere un composto che può essere utilizzato in sperimentazione clinica e portare, eventualmente, all'approvazione della commercializzazione come farmaco.

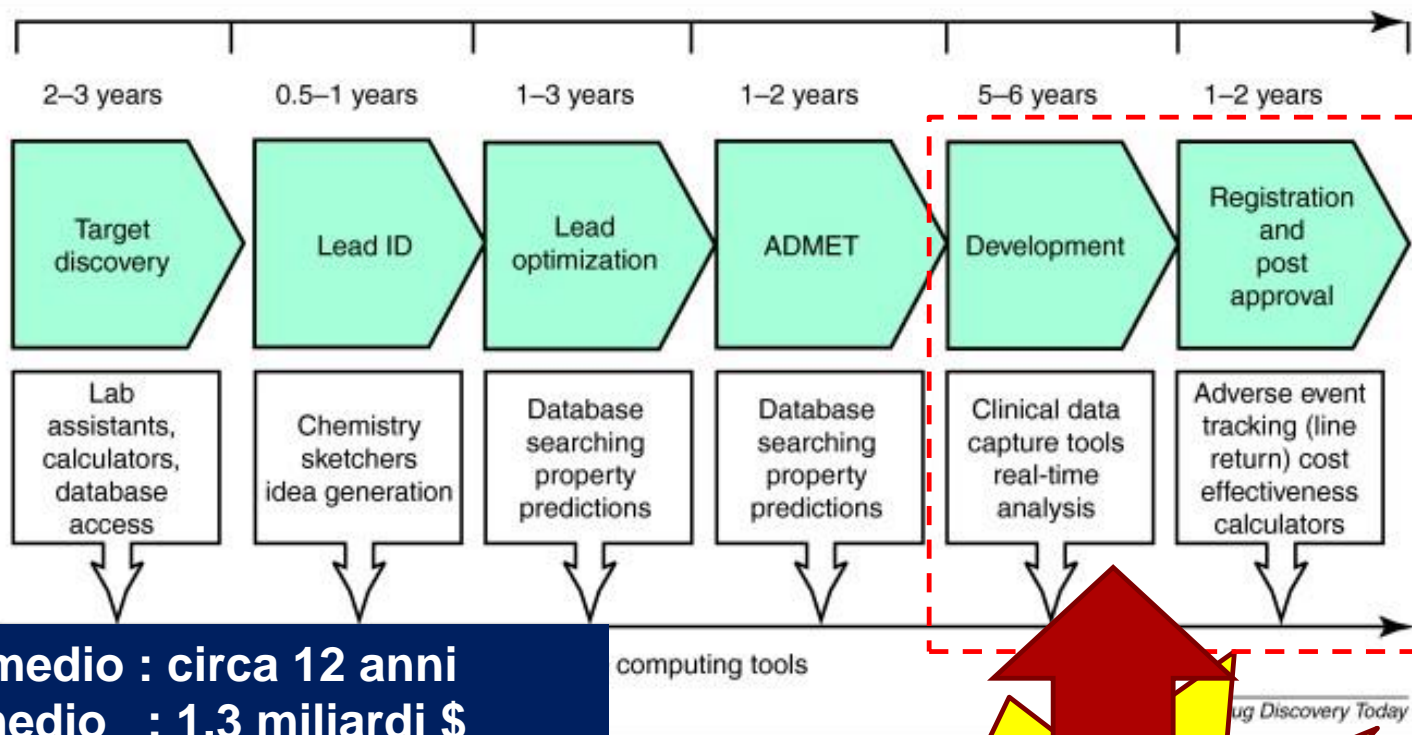


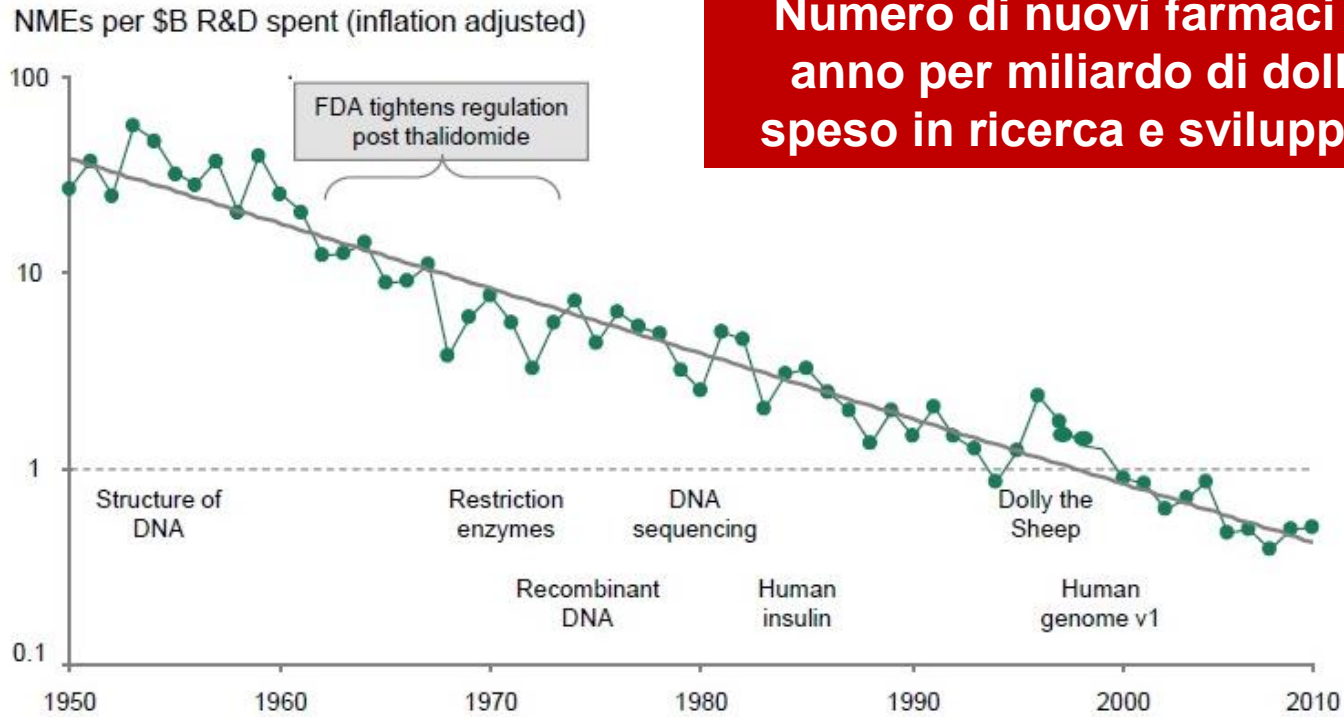
**Tempo medio : circa 12 anni**  
**Costo medio : 1.3 miliardi \$**

computing tools

*Drug Discovery Today*







Numero di nuovi farmaci per anno per miliardo di dollari speso in ricerca e sviluppo ...

Note: R&D costs are estimated from PhRMA annual survey 2009; NMEs are the total number of small molecule and biologic approvals by the FDA  
 Source: Bernstein Research "The Long View - R&D Productivity" (September 30, 2010)

Motivi che possono portare al fallimento del ciclo in fase avanzata:

- Scoperta di **effetti collaterali** indesiderati
- Scoperta di **interazioni** del farmaco con altre molecole che portano ad una drastica diminuzione della sua efficacia
- Scoperta di **interazioni con altri farmaci** che incrementano la tossicità oltre i livelli accettabili
- Scoperta di effetti deleteri associati all'utilizzo del farmaco **per lunghi periodi** (o nel trattamento di condizioni croniche)

Gli investimenti per lo sviluppo del farmaco sono già stati spesi...  
che fare in caso di fallimento?

---

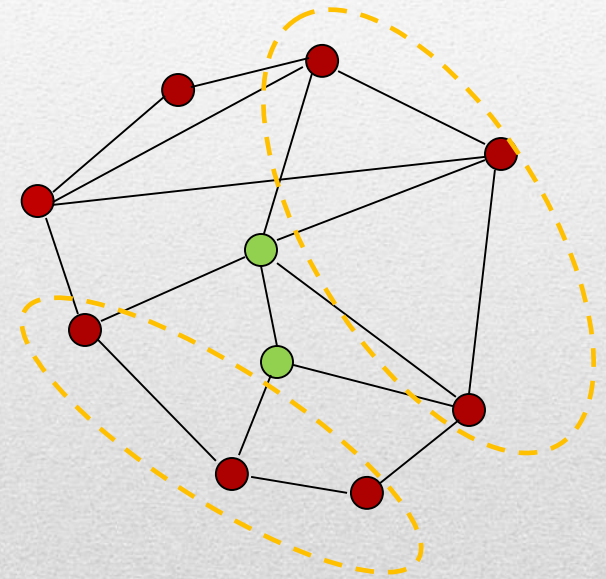
- Paradigma di sviluppo che sta prendendo piede negli ultimi anni.
- Si basa sulla ricerca di effetti collaterali di un farmaco già approvato ed immesso sul mercato che siano “potenzialmente utili” nel trattamento di una malattia diversa da quella per cui il farmaco era stato progettato.
- Si basa sulla similarità tra il farmaco da **riposizionare** e molecole che sono già utilizzate per il trattamento di una malattia.

**VANTAGGI:** il farmaco di partenza (quello che si vuole riposizionare) è già stato approvato per il commercio. Su di esso sono stati già effettuati molti test per garantire l'assenza di effetti indesiderati gravi. Inoltre sono disponibili per esso una quantità di informazioni molto maggiore rispetto ad un generic compost estratto da una libreria di molecole.

---

**POSITIVI:** farmaci già utilizzati per trattare una malattia

**POTENZIALI CANDIDATI** per un riposizionamento



**NB:** è possibile identificarli mediante un random walk su grafo !!!

---

- Permette di ridurre in modo consistente (5-6 anni) il ciclo di sviluppo di un nuovo farmaco
  - Permette di produrre farmaci più sicuri
  - Permette di ridurre drasticamente i costi associati allo sviluppo
  - **SI BASA PREVALENTEMENTE SU TECNICHE COMPUTAZIONALI** (almeno nelle prime fasi ... poi devono comunque essere effettuati test di sicurezza ma sono molto meno stringenti poichè, di solito, ne sono già stati effettuati a monte della prima approvazione del farmaco per l'immissione sul mercato).
-