

# Random Walk in Network Medicine: come navigare tra i componenti cellulari

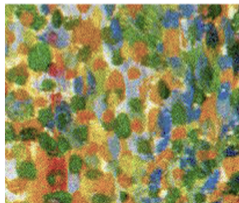
ELICSIR - SCUOLA ORTOGONALE

15 maggio 2026

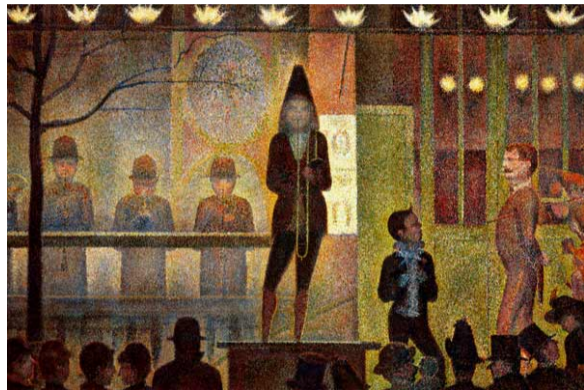
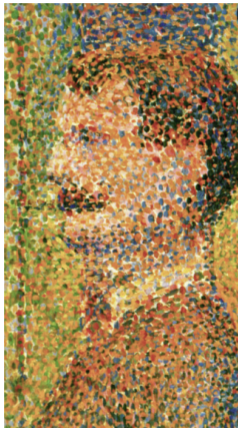
Studente: Sara Ferraro

Mentore: Professoressa Simona Ester Rombo

# La Metafora del Puntinismo



*Il punto osservato isolatamente*



*La visione complessiva dell'opera  
"Parade de Cirque" — G. Seurat, 1889*

## Riduzionismo



George Wells  
Beadle  
(1903 - 1989)

Edward Lawrie  
Tatum  
(1909 - 1975)

«Un gene, un enzima, una funzione»

**Beadle & Tatum**

Nobel 1958

## Network Medicine



«La malattia come perturbazione dell'interattoma»

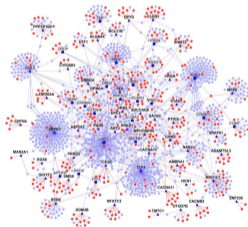
**A.-L. Barabási**

Network Science

L'interattoma rappresenta la **totalità delle interazioni** all'interno di una cellula.

## Complessità

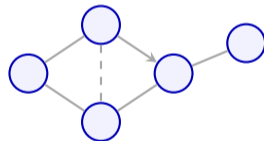
- > 25.000 geni
- Migliaia di metaboliti e RNA



1.087.908 interazioni documentate.

## Astrazione

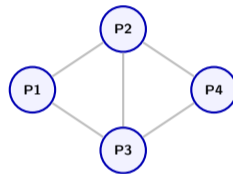
- **Nodi:** Macromolecole
- **Archi:** Interazioni



Modello matematico a **Grafo**.

## Classificazione

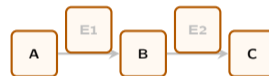
- **PPI** (Interazione Proteica)



**Proteine + Interazioni Fisiche**

## Classificazione

- **PPI** (Interazione Proteica)
- **Reti Metaboliche**

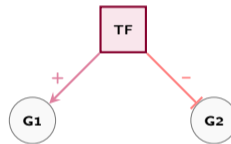


**Metaboliti + Reazioni Enzimatiche**

# Principali Tipi di Reti Biologiche

## Classificazione

- **PPI** (Interazione Proteica)
- **Reti Metaboliche**
- **Reti Regolatorie**



Fattori Trascrizione → Geni

## Classificazione

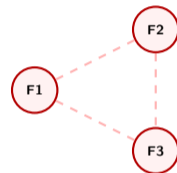
- **PPI** (Interazione Proteica)
- **Reti Metaboliche**
- **Reti Regolatorie**
- **Reti a RNA**



Regolazione Post-Trascrizionale

## Classificazione

- **PPI** (Interazione Proteica)
- **Reti Metaboliche**
- **Reti Regolatorie**
- **Reti a RNA**
- **Reti Fenotipiche**



Correlazione tra Patologie

## Classificazione

- **PPI** (Interazione Proteica)
- **Reti Metaboliche**
- **Reti Regolatorie**
- **Reti a RNA**
- **Reti Fenotipiche**
- **Reti di Co-espressione**



Correlazione Trascrizionale

L'interattoma segue precise **leggi universali**:

## Topologia Scale-Free

$P(k) \sim k^{-\gamma}$ : la rete è dominata da pochi **Hub** altamente connessi e molti nodi periferici.

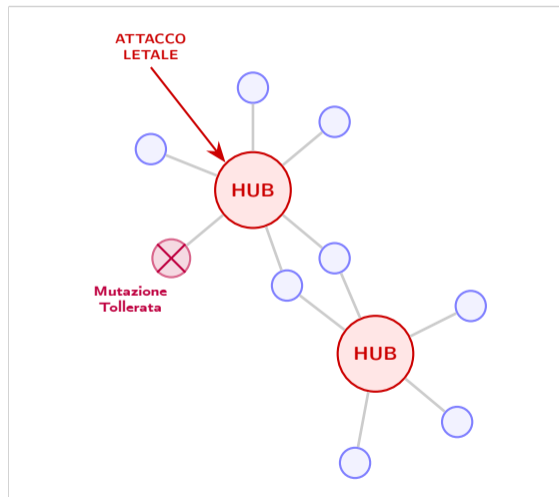
## Robustezza e Vulnerabilità

Il sistema è resiliente ai guasti *casuali*, ma vulnerabile ad attacchi *mirati* sugli Hub.

---

## Il Paradosso Medico:

- **Hub**: Geni essenziali (mutazione letale).
- **Periferia**: Geni malattia (mutazione tollerata).



## Proprietà Small-World

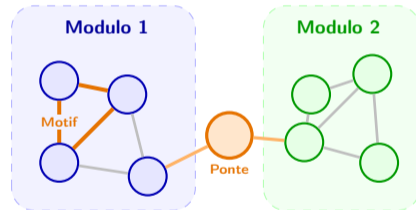
Distanza media ridotta tra i nodi: le perturbazioni biologiche si propagano rapidamente in tutta la rete.

## Modularità

Moduli topologici densamente connessi, associati a specifiche funzioni biologiche e collegati da nodi *ponte*.

## Motifs

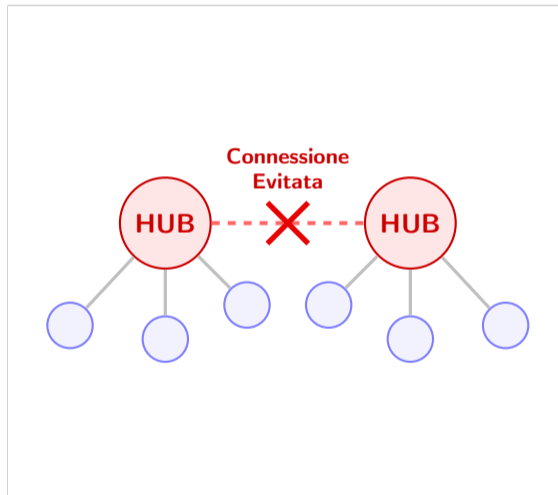
Sottografi ricorrenti.



## Disassortatività

Nelle reti biologiche, gli **Hub** evitano **connessioni reciproche**.

Essi tendono a legarsi preferenzialmente a nodi periferici a bassa connettività, piuttosto che ad altri Hub.



# Postulati della Network Medicine

La Network Medicine interpreta la malattia come una perturbazione della rete biologica, formulando principi che collegano **struttura topologica** e **fenotipo patologico**.

## Ipotesi Locale

Le proteine associate alla stessa patologia tendono a interagire in modo **diretto**.

## Modulo di Malattia

I geni malattia si concentrano in **specifiche regioni** dell'interattoma.

## Parsimonia di Rete

I percorsi causali coincidono spesso con i **cammini minimi** tra i nodi malattia.

## Componenti Condivise

Malattie che condividono componenti molecolari mostrano **similarità fenotipica e comorbidità**.

## i. Metodi Locali (Linkage)

Si basano sull'**interazione fisica diretta** con le proteine di malattia note.

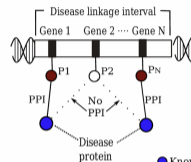
## ii. Metodi Modulari (Clustering)

Identificano **moduli funzionali** densamente interconnessi intorno ai geni seed.

## iii. Metodi di Diffusione (Globali)

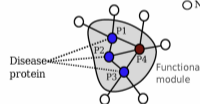
Sfruttano la **topologia dell'intera rete** tramite processi di Random Walk.

### i. Linkage methods

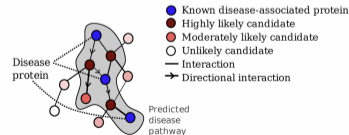


- Known disease-associated protein
- Candidate protein
- Not a candidate

### ii. Clustering methods



### iii. Diffusion-based methods



# L'Algoritmo Random Walk

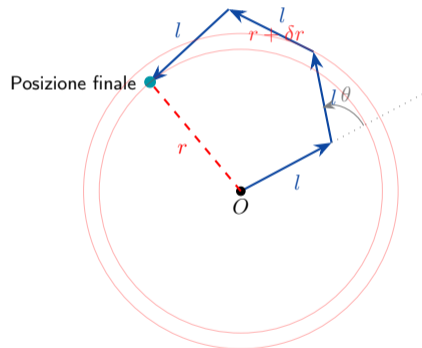
# Il problema del Random Walk

Nel 1905, **Karl Pearson** introdusse il problema del *Random Walk* ponendo questo quesito ai lettori della rivista *Nature*:

## La Formulazione Originale (27 Luglio 1905)

*“Un uomo parte da un punto  $O$  e cammina per  $l$  yards in linea retta; dopodiché ruota di un angolo qualsiasi e cammina per altre  $l$  yards lungo una seconda linea retta. Ripete questo processo  $n$  volte.*

*Richiedo la probabilità che, dopo questi  $n$  tratti, egli si trovi a una distanza compresa tra  $r$  e  $r + \delta r$  dal suo punto di partenza,  $O$ .”*



## L'analogia con l'acustica (3 Agosto 1905)

Il problema è matematicamente identico alla **composizione di  $n$  vibrazioni isoperiodiche** con fasi casuali, un modello già risolto da Rayleigh nel 1880 per i suoi studi sul suono.

Per un numero  $n$  di passi molto grande, la probabilità cercata si riduce all'espressione asintotica:

$$\frac{2}{n} e^{-\frac{r^2}{n}} r dr$$

*\*Nota:  $dr$  è il differenziale di  $r$ , ovvero l'equivalente matematico continuo dell'intervallo  $\delta r$  richiesto da Pearson. Assumiamo per semplicità un passo unitario  $l = 1$ .*

## La conclusione di Pearson (10 Agosto 1905):

*"La lezione della soluzione di Lord Rayleigh è che in aperta campagna il luogo più probabile in cui trovare un ubriaco [...] è da qualche parte vicino al suo punto di partenza!"*

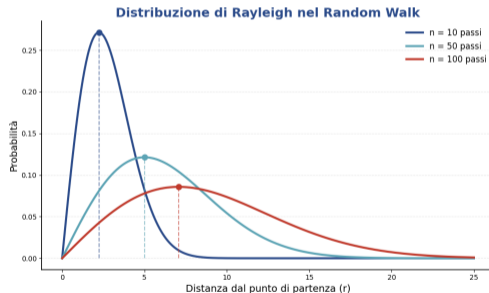
# Interpretazione della Soluzione di Rayleigh

Rayleigh ottiene per la distanza dal punto di partenza la seguente distribuzione radiale:

$$p(r) dr \propto r e^{-r^2/n} dr$$

## Un bilancio tra due fattori

- **Lo spazio disponibile ( $r$ ):** spinge la probabilità a crescere allontanandosi dal centro.
- **Il decadimento Gaussiano ( $e^{-r^2/n}$ ):** penalizza fortemente le grandi distanze.



## Il Teorema del Ritorno di Pólya

Una domanda fondamentale guida lo studio di Pólya:

*“Lasciato vagare all’infinito, il camminatore tornerà prima o poi con certezza al suo punto di partenza?”*

La risposta dipende dalla dimensionalità dello spazio ( $d$ ):

- **Spazi 1D e 2D (Stato Ricorrente):** La probabilità di ritorno all’origine è  $P = 1$ .
- **Spazi 3D o superiori (Stato Transiente):** Lo spazio disponibile cresce troppo velocemente. In 3D, la probabilità di ritorno è del **34%** circa.

### Osservazione di Shizuo Kakutani

*“Un ubriaco ritroverà sempre la strada di casa, ma un uccellino ubriaco potrebbe perdersi per sempre.”*

# Random Walk su Grafi

Dato un grafo non orientato  $G = (V, E)$  e un vertice di partenza  $s$ , un **Random Walk** (cammino casuale) di lunghezza  $t$  è un processo stocastico che modella l'esplorazione sequenziale della rete.

## Spostamento

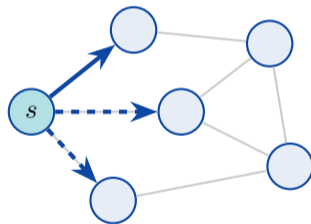
Da un nodo  $u$ , il camminatore sceglie casualmente un nodo adiacente  $v$  e vi si trasferisce.

## Propagazione

Il processo viene ripetuto per  $t$  passi.

## Generalità

Il modello si adatta a grafi pesati (scelta basata sui pesi) o orientati (vincoli di direzione).



# La Dinamica del Random Walk

Per analizzare l'evoluzione del processo, passiamo a una formulazione algebrica. Sia  $A$  la matrice di adiacenza e  $D$  la matrice dei gradi.

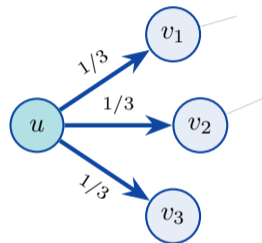
## Matrice di Transizione ( $W$ )

Definiamo la **Walk Matrix**:

$$W = AD^{-1}$$

La probabilità di passare da  $u$  a  $v$  è:

$$P(u \rightarrow v) = 1/d(u)$$



## Evoluzione Temporale della Distribuzione

Dato un vettore di probabilità iniziale  $p^0$ , l'evoluzione del sistema al tempo  $t$  è descritta come:

$$p^{t+1} = Wp^t \quad \implies \quad p^t = W^t p^0$$

Dopo un numero elevato di passi ( $t \rightarrow \infty$ ), la distribuzione di probabilità del camminatore raggiunge una distribuzione limite che riflette la topologia locale della rete.

## Distribuzione Stazionaria

Per un grafo connesso e non bipartito, la probabilità limite per ogni nodo  $v$  converge verso il valore  $\pi_v$ :

$$\pi_v = \frac{d(v)}{\sum_{u \in V} d(u)}$$

## L'Invarianza del Sistema

$\pi$  rappresenta lo **stato stazionario** del processo. Una volta raggiunto questo equilibrio, le successive transizioni non alterano più la distribuzione:

$$W\pi = \pi$$

---

**Algoritmo 1:** Random Walk su Grafo

---

**Input:** Grafo  $G = (V, E)$ , tolleranza  $\epsilon$ , distribuzione iniziale  $P^0$

**Output:** Distribuzione stazionaria  $P^* \approx \pi$

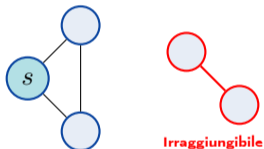
```
// Inizializzazione
1 foreach  $u \in V$  do
2   |  $P^0(u) \leftarrow$  valore iniziale;
3 end

// Iterazione fino alla convergenza
4  $t \leftarrow 0$ ;
5 repeat
6   | foreach  $u \in V$  do
7     | |  $P^{t+1}(u) \leftarrow \sum_{v \in N(u)} \frac{P^t(v)}{\deg(v)}$ ;
8     | end
9     |  $t \leftarrow t + 1$ ;
10 until  $\|P^{t+1} - P^t\| \leq \epsilon$ ;
11 return  $P^* \leftarrow P^t$ ;
```

---

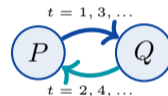
Il cammino casuale standard può fallire nel raggiungere l'equilibrio per due ragioni strutturali:

## 1. Disconnessione "Trappola Topologica"



Il walker resta confinato nella componente di partenza.

## 2. Bipartizione "Oscillazione Periodica"



La probabilità "rimbalza" in base alla parità del tempo  $t$ .

Come forzare la convergenza in ogni grafo connesso? → *Lazy Random Walk*

Per garantire la convergenza su ogni grafo connesso, introduciamo un fattore di "pigrizia".

## Lazy Walk

Ad ogni passo, il walker ha probabilità  $1/2$  di restare fermo e  $1/2$  di muoversi normalmente.

$$\hat{W} = \frac{1}{2}I + \frac{1}{2}W$$

### Vantaggio:

- Elimina le oscillazioni.

---

**Algoritmo 2:** Lazy Random Walk su Grafo

---

**Input:** Grafo  $G = (V, E)$ , tolleranza  $\epsilon$ , distribuzione iniziale  $P^0$ , parametro di pigrizia  $\alpha \in (0, 1)$

**Output:** Distribuzione stazionaria  $P^* \approx \pi$

```
// Inizializzazione
1 foreach  $u \in V$  do
2   |  $P^0(u) \leftarrow$  valore iniziale;
3 end

// Iterazione fino alla convergenza
4  $t \leftarrow 0$ ;
5 repeat
6   | foreach  $u \in V$  do
7     |  $P^{t+1}(u) \leftarrow \alpha \cdot P^t(u) + (1 - \alpha) \sum_{v \in N(u)} \frac{P^t(v)}{\deg(v)}$ ;
8   | end
9   |  $t \leftarrow t + 1$ ;
10 until  $\|P^{t+1} - P^t\| \leq \epsilon$ ;
11 return  $P^* \leftarrow P^t$ ;
```

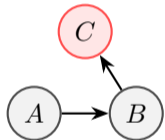
---

# Dal Random Walk al PageRank

Il Random Walk standard garantisce la convergenza a una distribuzione stazionaria  $\pi$  solo su grafi non orientati e connessi. In **grafi orientati**, come il Web, il modello fallisce a causa di due anomalie topologiche:

## 1. Dangling Nodes

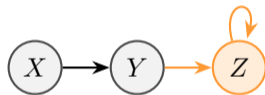
Nessun arco uscente



Il walker raggiunge  $C$  e non può più uscire.

## 2. Spider Traps

Ciclo chiuso



**Effetto:** Il grafo non è fortemente connesso. Il walker entra in un ciclo chiuso (es.  $Z \rightarrow Z$ ).

## Il Teletrasporto (Damping Factor $\alpha$ )

L'algoritmo PageRank modifica la matrice di transizione: ad ogni passo, con probabilità  $\alpha$  il walker segue un arco uscente, e con probabilità  $(1 - \alpha)$  si teletrasporta su un nodo casuale dell'intera rete.

---

**Algoritmo 3:** PageRank su Grafo

---

**Input:** Grafo  $G = (V, E)$ , tolleranza  $\epsilon$ , fattore di damping  $\alpha \in (0, 1)$ , distribuzione iniziale  $P^0$

**Output:** Vettore PageRank  $P^* \approx \pi$

```
// Inizializzazione
1 foreach  $u \in V$  do
2    $P^0(u) \leftarrow \frac{1}{|V|}$ ;
3 end

// Iterazione fino alla convergenza
4  $t \leftarrow 0$ ;
5 repeat
6   foreach  $u \in V$  do
7      $P^{t+1}(u) \leftarrow \frac{1-\alpha}{|V|} + \alpha \sum_{v \in \mathcal{B}(u)} \frac{P^t(v)}{\deg^+(v)}$ ;
8   end
9    $t \leftarrow t + 1$ ;
10 until  $\|P^{t+1} - P^t\| \leq \epsilon$ ;
11 return  $P^* \leftarrow P^t$ ;
```

---

# Personalized PageRank (PPR)

Il **Personalized PageRank** introduce un bias strutturale: il camminatore non si teletrasporta in modo uniforme, ma segue una distribuzione di probabilità a priori  $p$ , che favorisce specifici nodi target.

---

---

// PageRank: teletrasporto uniforme

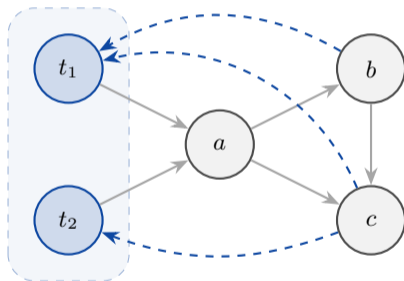
$$1 \quad P^{t+1}(u) \leftarrow \frac{1 - \alpha}{|V|} + \alpha \sum_{v \in \mathcal{B}(u)} \frac{P^t(v)}{\deg^+(v)}$$

// PPR: teletrasporto personalizzato

$$2 \quad P^{t+1}(u) \leftarrow (1 - \alpha)p(u) + \alpha \sum_{v \in \mathcal{B}(u)} \frac{P^t(v)}{\deg^+(v)}$$

---

Teletrasporto verso Nodi Target



$$p(t_1) = p(t_2) = 1/2 \\ p(a) = p(b) = p(c) = 0$$

— Transizione topologica ( $\alpha$ )

-- Teletrasporto  $(1 - \alpha)$  distribuito su  $t_1, t_2$

# Random Walk con Restart (RWR)

Il **Random Walk con Restart** è un caso speciale del PPR in cui il vettore di personalizzazione è interamente concentrato su un *unico* nodo sorgente  $s$ :  $p = e_s$

---

---

// RWR su grafo non orientato

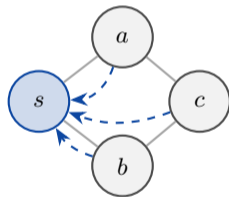
$$1 \quad P^{t+1}(u) \leftarrow (1 - \alpha) \mathbf{1}[u = s] + \alpha \sum_{v \in N(u)} \frac{P^t(v)}{\deg(v)}$$

// RWR su grafo orientato

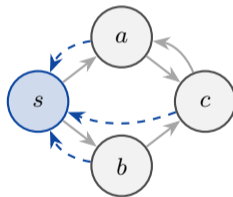
$$2 \quad P^{t+1}(u) \leftarrow (1 - \alpha) \mathbf{1}[u = s] + \alpha \sum_{v \in \mathcal{B}(u)} \frac{P^t(v)}{\deg^+(v)}$$

---

Non orientato



Orientato



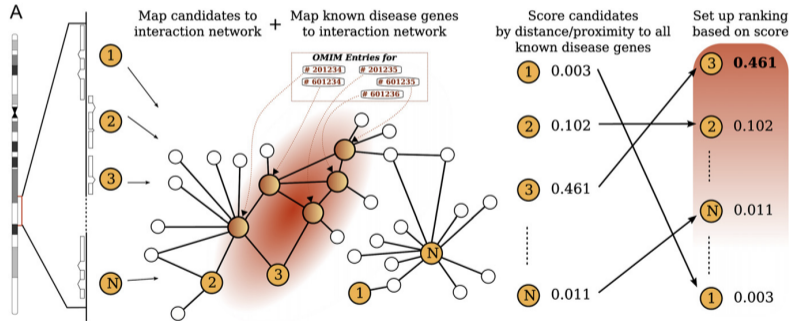
— topologia ( $\alpha$ )

- - restart  $(1 - \alpha)$  verso  $s$

Torniamo al problema biomedico. Tecniche tradizionali come l'analisi di linkage non identificano un singolo gene malattia, ma un'intera regione cromosomica che può contenere centinaia di candidati.

## La Sfida

Sequenziare in laboratorio tutti i geni di un "intervallo di linkage" è costoso e lento. Serve un metodo computazionale per stabilire una **lista di priorità**



Nel 2008, Köhler et al. hanno applicato per la prima volta il **Random Walk with Restart (RWR)** all'interattoma umano per risolvere questo problema.

## Il Modello

- **Il Grafo ( $G$ ):** È la rete delle interazioni proteina-proteina (PPI).
- **I Nodi Seed ( $S$ ):** L'insieme dei geni già noti per essere associati a quella malattia. Il teletrasporto si distribuisce equamente su di essi.
- **Il Teletrasporto ( $\alpha$ ):** Evita che il camminatore si perda nell'intera rete, forzandolo a esplorare il *modulo di malattia* locale.
- **I Punteggi ( $P^*$ ):** La probabilità stazionaria finale di ogni nodo candidato diventa il suo punteggio di priorità.

Sebbene efficace, il modello di Köhler si limitava alla sola rete PPI.  
Tuttavia, la biologia cellulare è intrinsecamente **multidimensionale**.

## Pluralismo dei Dati

Per una visione completa del funzionamento cellulare, è necessario integrare:

- Reti di co-espressione
- Reti metaboliche e regolatorie
- Reti fenotipiche

**Aggregare le reti o mantenerle separate?** → *Reti Multiplex ed Eterogenee.*

Valdeolivas et al. (2019) estendono il RWR a strutture grafiche complesse che preservano la topologia di ogni sorgente dati.

## Rete Multiplex (RWR-M) [Fig. A]

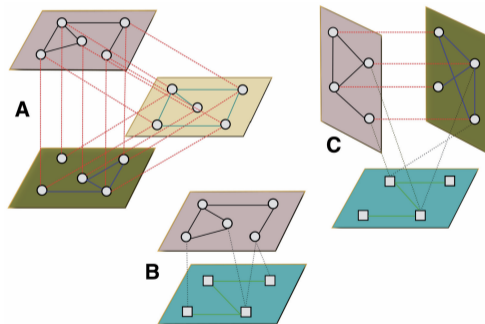
Collezione di strati (Layers) con gli stessi nodi ma diverse tipologie di archi.

## Rete Eterogenea (RWR-H)[Fig. B]

Unione di due domini diversi (es. geni e malattie) collegati da archi bipartiti.

## Rete Multiplex-Eterogenea (RWR-MH) [Fig. C]

Combinazione delle due.



In una rete multiplex-eterogenea, il camminatore dispone di tre tipologie di "salti" ad ogni passo:

- 1 **Salto Intra-layer:** Spostamento standard tra vicini all'interno dello stesso strato (es. PPI).
- 2 **Salto Inter-layer:** "Cambio di piano" verso lo stesso gene in un altro strato (es. Co-espressione).
- 3 **Salto Bipartito:** Transizione tra i diversi domini (geni e malattie).

*Questo approccio permette di sfruttare il "pluralismo dei dati" senza perdere le proprietà topologiche individuali.*

Il framework **BRW** (Gentili et al., 2022) affronta la sfida dell'integrazione di sorgenti eterogenee non solo a livello strutturale, ma raffinando la dinamica stessa del cammino.

## Informazioni Multi-omiche

Per migliorare la predizione, BRW integra informazioni **multi-omiche**:

- **Annotazioni Funzionali:** GO, KEGG, Reactome (Conoscenza biologica).
- **Trascrittomica:** Dati di espressione genica da coorti di pazienti (Dati clinici reali).

*L'obiettivo è spostare il Random Walk da una navigazione "cieca" a una guidata dall'evidenza biologica specifica per la patologia.*

BRW ridefinisce i componenti fondamentali del Random Walk integrando pesi biologici tramite parametri di bilanciamento  $(\alpha, \beta)$ :

- **Teletrasporto Personalizzato ( $p$ ):**

$$p = \alpha p_{\text{annotazioni}} + (1 - \alpha) p_{\text{espressione}}$$

Il walker è forzato a saltare verso geni funzionalmente simili ai geni seed o clinicamente alterati.

- **Matrice di Transizione Pesata ( $W$ ):**

$$W = \beta W_{\text{annotazioni}} + (1 - \beta) W_{\text{espressione}}$$

Gli archi della rete PPI vengono "allargati" se le proteine condividono funzioni specifiche per la malattia.

## Limiti

- **Dipendenza dai Parametri:**  $\alpha$  (restart) ed  $\varepsilon$  (convergenza) sono scelti in modo euristico e problem-dependent.
- **Scalabilità:** Le reti genomiche massive causano colli di bottiglia computazionali.
- **Visione Statica e Piatta:** L'algoritmo opera su un'istantanea temporale fissa e tratta i layer in modo omogeneo.

## Sviluppi Futuri

- **Ottimizzazione Teorica:** Metodi automatici (data-driven) per la selezione ottimale di  $\alpha$  ed  $\varepsilon$ .
- **Calcolo Parallelo e Approssimato:** Algoritmi approssimati e calcolo parallelo per task su larga scala.
- **Modelli Dinamici e Asimmetrici:** Reti *context-aware* (tempo-varianti) e pesi di integrazione differenziati.

## Fondamenti di Network Science e Biologia

### **Network Medicine: A Network-Based Approach to Human Disease**

Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). *Nature Reviews Gen.*, 12(1), 56–68.

### **Network Science**

Barabási, A. L. (2016). *Cambridge Univ. Press.*

### **Interactome Networks and Human Disease**

Vidal, M., Cusick, M. E., & Barabási, A. L. (2011). *Cell*, 144(6), 986–998.

### **Network Biology: Understanding the Cell's Functional Organization**

Barabási, A. L., & Oltvai, Z. N. (2004). *Nature Reviews Gen.*, 5(2), 101–113.

## Random Walk: Origini e Fondamenti

### **The Problem of the Random Walk**

Pearson, K. (1905); Rayleigh, Lord (1905). *Nature*, 72.

### **Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz**

Pólya, G. (1921). *Math. Annalen*, 84, 149–160.

## Teoria dei Random Walk su Grafi

### **Random Walks on Graphs: A Survey**

Lovász, L. (1993). *Combinatorics, Paul Erdős is Eighty*, 2(1), 1–46.

## Applicazioni

### **A Review of Random Walk-Based Method for the Identification of Disease Genes and Disease Modules**

Tay Xin Hui, S., Kasim, S., et al. (2019). *IEEE Access*.

### **Biological Random Walks: Multi-Omics Integration for Disease Gene Prioritization**

Gentili, M., Martini, L., et al. (2020). *Systems Biology*.

### **Walking the Interactome for Prioritization of Candidate Disease Genes**

Köhler, S., et al. (2008). *Am. Journal of Human Genetics*.

### **Random Walk with Restart on Multiplex and Heterogeneous Biological Networks**

Valdeolivas, A., et al. (2019). *Bioinformatics*.

### **DiCE: differential centrality-ensemble analysis based on gene expression profiles and protein–protein interaction network**

Pashaei, E., Liu, S., et al. (2025). *Nucleic Acids Research*, 53.

Grazie

Abbandoniamo lo spazio continuo per vincolare il movimento ad un percorso discreto: immaginiamo un camminatore che si muove lungo uno stretto marciapiede, facendo passi casuali avanti o indietro.

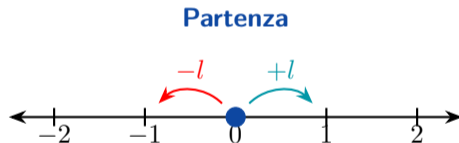
## Il Modello Matematico

A ogni passo  $i$ , la variabile aleatoria  $X_i$  può assumere solo due valori con uguale probabilità (avanti o indietro):

$$P(X_i = +l) = P(X_i = -l) = \frac{1}{2}$$

Lo spostamento totale dopo  $n$  passi è la somma dei singoli contributi discreti:

$$S_n = \sum_{i=1}^n X_i$$



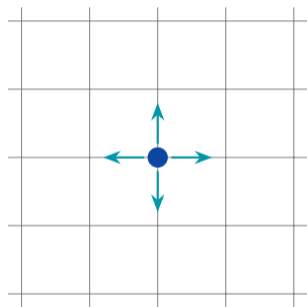
Nel 1921, **George Pólya** generalizzò il Random Walk su griglie multidimensionali  $\mathbb{Z}^d$ . Il camminatore si sposta tra i nodi (gli incroci) di un reticolo regolare.

## Transizioni Uniformi

In un reticolo 2D ( $\mathbb{Z}^2$ ), a ogni passo si sceglie uniformemente una delle 4 direzioni cardinali:

$$P(\text{Nord}) = P(\text{Sud}) = \dots = \frac{1}{4}$$

Questo approccio trasforma il problema da puramente geometrico a combinatorio e topologico. L'analisi si sposta su percorsi e nodi.



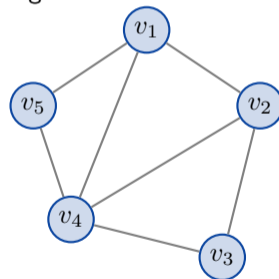
I reticoli  $\mathbb{Z}^d$  possiedono una struttura altamente simmetrica: sono grafi regolari, in cui ogni vertice ha lo stesso grado.

Per descrivere sistemi più complessi, privi di simmetria spaziale e con connessioni eterogenee, è necessario introdurre una formalizzazione più generale: il concetto di grafo.

## Definizione di Grafo

Un grafo è definito matematicamente dalla coppia  $G = (V, E)$ , dove:

- $V$  (*Vertices*): l'insieme dei nodi (o vertici).
- $E$  (*Edges*): l'insieme degli archi che collegano i nodi.



## Convergenza: Riduzione Spettrale

**Idea chiave:** esprimere  $p^t = W^t p^0$  tramite gli autovalori di  $W$ .

Poiché  $W$  non è simmetrica, si introduce la matrice simmetrica ausiliaria:

$$S := D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = D^{-\frac{1}{2}} W D^{\frac{1}{2}}$$

con autovalori reali  $\omega_1 \geq \dots \geq \omega_n$  e autovettori ortonormali  $v^1, \dots, v^n$ .

### Trasferimento degli autovalori a $W$

$\omega_i$  è autovalore di  $W$  con autovettore  $D^{\frac{1}{2}} v^i$ :

$$W \left( D^{\frac{1}{2}} v^i \right) = D^{\frac{1}{2}} S v^i = \omega_i \left( D^{\frac{1}{2}} v^i \right)$$

Questo garantisce che  $W$  abbia  $n$  autovalori reali, e permette di scrivere:

$$p^t = W^t p^0 = D^{\frac{1}{2}} \left( \sum_{i=1}^n \omega_i^t v^i (v^i)^T D^{-\frac{1}{2}} p^0 \right)$$

I soli termini dipendenti da  $t$  sono le potenze  $\omega_i^t$ .

## Lemma (bound sugli autovalori)

Per ogni grafo connesso e non bipartito:  $|\omega_i| \leq 1$ ,  $\omega_1 = 1$  con autovettore  $\pi$ , e  $|\omega_i| < 1$  per ogni  $i > 1$ .

**Dimostrazione**  $|\omega_i| \leq 1$ : sia  $u$  il vertice che massimizza  $|y_u^i|/d(u)$ , con  $y^i = D^{\frac{1}{2}}v^i$ . Per definizione di autovalore:

$$|\omega_i| |y_u^i| = \left| \sum_{(w,u) \in E} \frac{y_w^i}{d(w)} \right| \leq \sum_{(w,u) \in E} \frac{|y_u^i|}{d(u)} = |y_u^i|$$

**Conclusione:** per  $t \rightarrow \infty$ , tutti i termini con  $|\omega_i| < 1$  si annullano nella sommatoria, e rimane solo il termine  $i = 1$ :

$$p^t \xrightarrow{t \rightarrow \infty} D^{\frac{1}{2}} v^1 (v^1)^T D^{-\frac{1}{2}} p^0 = C\pi = \pi$$

poiché ogni autovettore associato a  $\omega_1 = 1$  è della forma  $C\pi$ , e  $W^t p^0$  deve essere una distribuzione di probabilità ( $\Rightarrow C = 1$ ). □

# Teoria Spettrale e Velocità di Convergenza

La velocità con cui il cammino casuale esplora la rete e raggiunge l'equilibrio è determinata dalla struttura algebrica della matrice di transizione.

## Teorema della Convergenza

L'errore rispetto alla distribuzione stazionaria al tempo  $t$  decade secondo la potenza degli autovalori:

$$|p_t(v) - \pi(v)| \leq \sqrt{\frac{d(v)}{d(u)}} \omega_2^t$$

*Il secondo autovalore  $\omega_2$  è il parametro critico che governa l'intero processo.*

### Spectral Gap

Definito come  $\lambda = 1 - \omega_2$ .

Rappresenta la "distanza" tra lo stato stazionario e il resto dello spettro.

### Mixing Time

Il tempo  $t_\epsilon$  necessario affinché la distribuzione sia "vicina" a  $\pi$ .

È inversamente proporzionale al Gap.

I metodi precedenti misuravano distanze **locali** (es. il "Cammino Minimo" o l'interazione diretta). Il RWR introduce una misura di similarità **globale**.

## Limite del Cammino Minimo

Due nodi connessi da un singolo percorso hanno la stessa "distanza" di due nodi connessi da decine di percorsi paralleli.

## Vantaggio del RWR

Il Random Walk calcola la probabilità di transizione considerando **tutti i percorsi possibili**.

*Risultato: Su 110 famiglie di malattie genetiche, il RWR ha superato i metodi locali raggiungendo un'accuratezza fino al 98%.*