

Big Network Visualizzazione Tool for iNSIdEnano

Luigi Giugliano

Università degli studi di Salerno

28 gennaio 2016

OVERVIEW

iNSIdEnano

Dati

Generazione Network

Implmentazione

Problema

Soluzione

OVERVIEW

iNSIdEnano

Dati

Generazione Network

Implmentazione

Problema

Soluzione

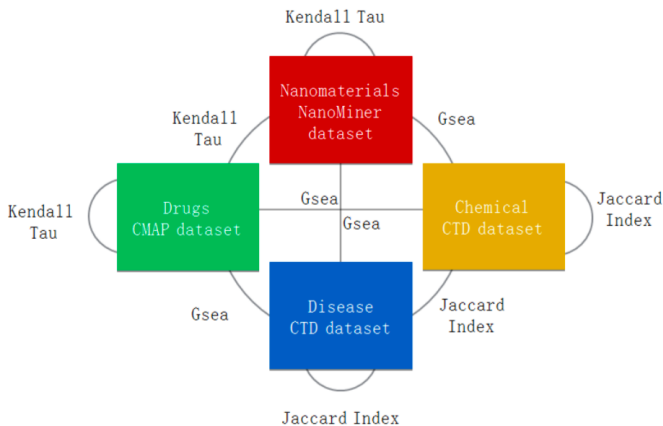


iNSIdENANO

iNSIdEnano è un tool grafico che mette in evidenza le connessioni tra entità fenotipiche del tipo:

- ▶ Esposizione ai nanomateriali
- ▶ Trattamenti farmaceutici
- ▶ Esposizione ad agenti chimici
- ▶ Malattie

L' interazione tra queste entità è valutata in base al loro effetto sull'espressione dei geni.



E' stata calcolata la distanza per ogni coppia di entità. Sono poi state normalizzate tra -1 e 1 per renderle confrontabili.



Per ogni entità fenotipica nel dataset, è assegnata una lista di geni. In particolare un'insieme di geni è associato a ogni malattia e ogni agente chimico, invece per ogni farmaco e per ogni nanomateriale è associata una lista ordinata di geni.

Quindi per costruire una network di similarità tra entità fenotipiche è stato necessario calcolare la similarità a coppie per ogni entità.



INSIEME DI GENI VS INSIEME DI GENI

Il Jaccard index è stato utilizzato per calcolare la similarità tra due malattie, tra due agenti chimici o tra un agente chimico e una malattia.

Dati due insiemi A e B l'indice di Jaccard è dato dalla dimensione della loro intersezione diviso la dimensione della loro unione.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Questa misura è zero se i due insiemi non condividono neanche un gene, mentre 1 se sono esattamente uguali.

Per ogni agente chimico vengono considerati due set di geni: quelli che sono up-regolati da quell'agente chimico e quelli che sono down-regolati. Per quelli down-regolati il Jaccard index è calcolato con il segno negativo.



GENI ORDINATI VS GENI ORDINATI

La distanza Kendall Tau è stata utilizzata per calcolare la similarità tra nanomateriali e nanomateriali, tra farmaci e farmaci e tra nanomateriali e farmaci, basata sulla lista ordinata dei geni. La distanza Kendall Tau tra due liste T_1 e T_2 è definita come segue:

$$K(T_1, T_2) = |(i, j) : i < j, (T_1(i) < T_1(j) \wedge T_2(i) > T_2(j)) \vee (T_1(i) > T_1(j) \wedge T_2(i) < T_2(j))| \quad (2)$$

questa distanza è compresa tra 0 e $n(n-1)$, dove n è la lunghezza della lista. Il valore 0 significa che gli elementi nella lista sono nello stesso ordine, mentre il valore $n(n-1)$, indica che gli elementi sono in ordine opposto



GENIO ORDINATI VS INSIEME DI GENI

La Gen Set Enrichment Analysis (GSEA), basata sul test di Kolmogorov-Smirnov, è stata usata per calcolare la similarità a coppie tra nanomateriali e malattie, tra nanomateriali e agenti chimici, tra farmaci e malattie ed infine tra farmaci e agenti chimici. Il test di KolmogorovSmirnov può essere usato per confrontare elementi con una distribuzione di probabilità. La distribuzione empirica F_n per osservazioni *iid*, è definito:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I[-\inf, x](x_i) \quad (3)$$



dove:

$$I[-\inf, x](x_i)$$

è la funzione definita su X che indica l'appartenenza di un elemento in un sottoinsieme A di X che ha valore 1 per tutti gli elementi di A e 0 per tutti gli elementi di X non in A . La statistica KolmogorovSmirnov per una distribuzione cumulativa $F(x)$ è

$$D_n = \sup_x [F_n(x) - F(x)]$$

La statistica KolmogorovSmirnov è stata usata non in valore assoluto per preservare il segno. Ciò aiuta a capire se un gene è up o down-regolato, ovviamente anche questi valori sono stati normalizzati tra $[-1 : 1]$



IMPLMENTAZIONE

INSIdEnano è stato implemetato in *R* usando Shiny come libreria per l'interfaccia grafica. Il sistema è stato implementato in una struttura client-server: il client è responsabile per la gestione dell'interfaccia, mentre il serve processa i dati dal database in base agli input dell'utente, e restituisce il risultato di tale computazione al client.

OVERVIEW

iNSIdEnano

Dati

Generazione Network

Implmentazione

Problema

Soluzione



Parliamo ora della mole di questi dati,
i nodi presenti in questo grafo sono: 3686 per quanto riguarda
gli archi la situazione risulta essere molto più complessa,
perché come spiegato precedente sono state calcolate le
distanze tra tutte le coppie. e quindi il grafo senza sogliature ha
~15 000 000 archi. Il che rende questo grafo:

INVISUALIZZABILE

METTERE IMMAGINE GRAFICO INVISUALIZZABILE

OVERVIEW

iNSIdEnano

Dati

Generazione Network

Implmentazione

Problema

Soluzione

.