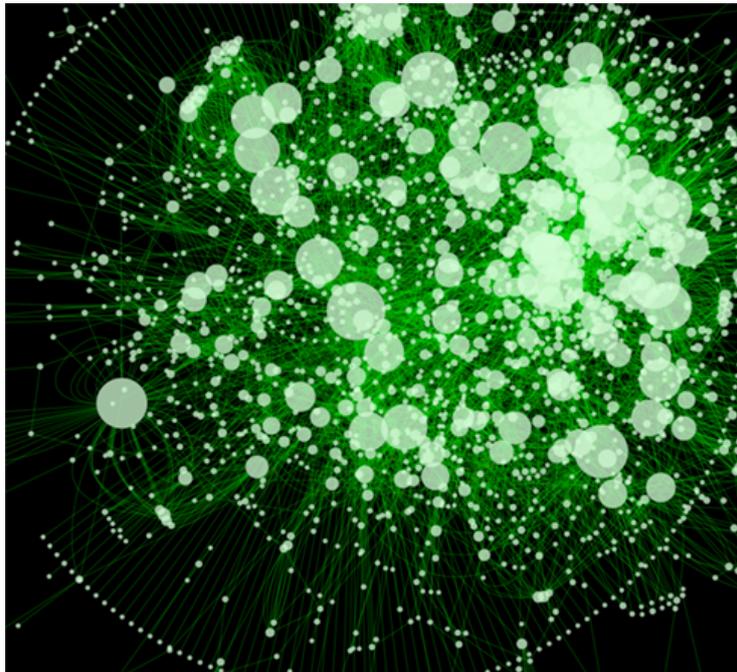


Analisi di reti biologiche con Cytoscape

Università degli Studi di Udine

20 dicembre 2016

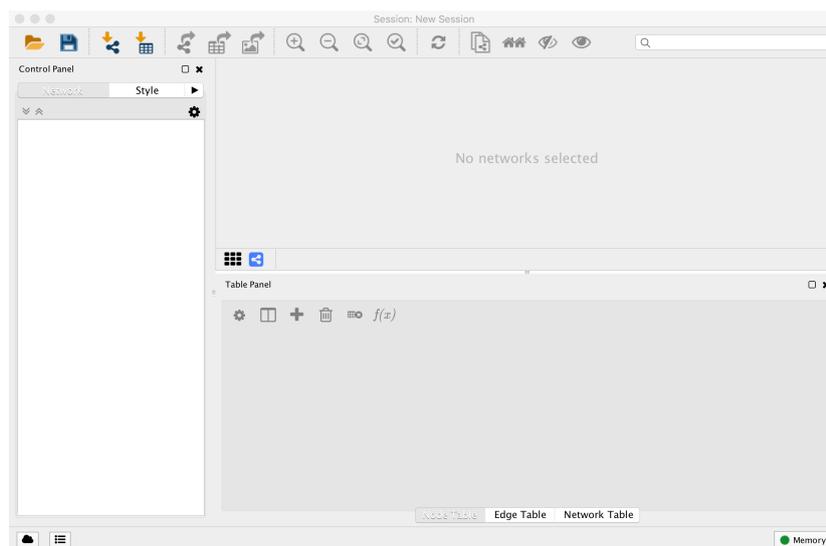


1 Introduzione

Cytoscape (<http://www.cytoscape.org/>) è un software *open-source* per la visualizzazione e l'analisi di reti d'interazione molecolare. Tra le caratteristiche più interessanti, mette conto di menzionare almeno le seguenti:

- possibilità di integrare fonti di dati diverse (ad esempio, informazioni su interazioni dna-proteina e profili di espressione genica), sia da file locali sia provenienti da ricerche online;
- capacità di associare una qualunque informazione a un attributo visivo della rete (come vedremo, si può ad esempio stabilire che la dimensione di un nodo che rappresenta un gene sia proporzionale al livello di espressione del gene stesso);
- estensibilità delle funzioni del programma mediante *app*. L'obiettivo di questa esercitazione è apprendere le funzionalità di base del software e di alcune sue app, al fine di effettuare semplici analisi.

All'avvio del programma, si presenta una finestra simile alla seguente:



La finestra principale è suddivisa in tre parti: un pannello di gestione delle reti (a sinistra), una finestra per la visualizzazione delle reti (a destra in alto) e un pannello per l'ispezione degli attributi di una rete (a destra in basso). I pannelli possono essere separati dalla finestra principale cliccando sull'icona in alto a destra nel corrispondente pannello (ciò può essere utile per aumentare lo spazio disponibile per la finestra di visualizzazione).

Attenzione: al fine di evitare la perdita di dati, si consiglia di salvare con frequenza la propria sessione di lavoro scegliendo **File** → **Save**. Così facendo, Cytoscape memorizza tutte le informazioni correntemente presenti nella sessione di lavoro in un file con suffisso **.cys**; in tal modo la sessione può essere ripresa in un momento successivo dallo stato in cui è stata lasciata.

1.1 Installazione di app

In questa esercitazione faremo uso di alcune app di Cytoscape. Una *app* è una componente software esterna che estende le funzionalità di Cytoscape. Le app sono accessibili dal menù **Apps**. Nel caso non fossero presenti in tale menù, è necessario installare i seguenti plugin:

- MCODE;
- NetworkAnalyzer;
- jActiveModules;
- BiNGO.

Per installare un'app, seleziona **Apps** → **App Manager...**, quindi scegli l'app desiderata nella finestra che appare e clicca su **Install**. Nota che per installare un'app in questo modo è necessario avere una connessione di rete attiva perché le app sono scaricate da un sito remoto.

1.2 Importare una rete

In questa esercitazione lavoreremo con i dati relativi a una rete metabolica di consumo del galattosio in lievito. I dati provengono da uno studio descritto in [1] e sono disponibili nel sito del corso. Per caricare tali dati, scegli **File** → **Import** → **Network** → **File...** e apri il file `galFiltered.sif`. Cytoscape visualizza la rete usando uno tra diversi possibili *layout*, che possono essere selezionati dal menù **Layout**. Cerca un layout che permetta una chiara visualizzazione della rete.

1. Prova a navigare all'interno della rete usando la visione "a volo d'uccello" in basso a destra e zoomando con il tasto destro o la rotella del mouse (o mediante le apposite icone nella barra degli strumenti).
2. Selezionare un nodo: è possibile selezionare un nodo cliccandoci sopra (i nodi selezionati diventano gialli) o effettuando una ricerca per nome: seleziona il nodo YPL248C scrivendo il nome nel campo di ricerca che si trova nella barra degli strumenti.
3. Una volta selezionati uno o più nodi, è possibile visualizzare i nodi che interagiscono direttamente con quelli selezionati, scegliendo **Select** → **Nodes** → **First Neighbors of Selected Nodes**. Con quali proteine interagisce YPL248C?

2 Analisi di una rete

Cytoscape consente di calcolare diversi parametri legati alla topologia di una rete, tra i quali:

- diametro della rete;
- grado medio dei nodi;
- distribuzione dei gradi;

- coefficiente di clustering medio;
- distribuzione dei coefficienti di clustering;
- lunghezza media dei cammini minimi;
- etc...

Questo tipo di analisi si effettua in modo molto semplice:

1. seleziona Tools → NetworkAnalyzer → Network Analysis → Analyze Network...

2. Scegli di trattare il grafo come un grafo non orientato e clicca su Ok.

Prova a rispondere alle seguenti domande:

1. Qual è il coefficiente di clustering?
2. Qual è il grado medio di un nodo?
3. Qual è la lunghezza media di un cammino minimo?
4. Qual è la lunghezza massima di un cammino minimo? Che cosa si può dedurre a proposito della *small-world property*?
5. Studia la definizione di *coefficiente topologico* nel manuale del *NetworkAnalyzer* (<http://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.7/>). Che cosa si può dedurre dal grafico dei coefficienti topologici per la rete considerata?
6. Di che tipo di rete si tratta (random, scale-free, hierarchical)?

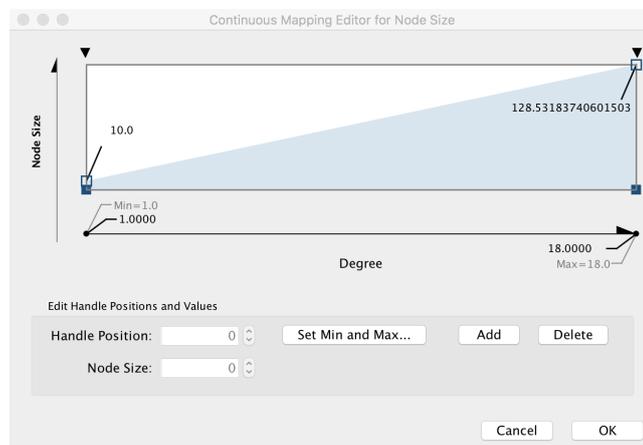
2.1 Rappresentare attributi in modo visivo

In Cytoscape è possibile associare a una qualunque proprietà della rete un corrispondente attributo grafico. Nella rete considerata, ad esempio, sono presenti interazioni proteina-proteina (due proteine che interagiscono fisicamente) e interazioni dna-proteina (ad esempio, un fattore di trascrizione che regola un gene). Possiamo distinguere tali tipi di interazione come segue:

1. Seleziona il pannello *Style*.
2. Dal menù a tendina in alto a destra, scegli la voce *Create New Style...* e dà un nome al nuovo stile.
3. Seleziona il sotto-pannello *Edge* (il pulsante si trova in basso).
4. Cambia lo spessore degli archi (attributo *Width*) impostandolo al valore 3.0 (clicca sul valore corrente per modificarlo). Questo è un esempio di associazione di un valore uguale per tutti gli archi della rete.
5. Cerca l'attributo *Stroke Color (unselected)* (se non lo vedi, apri il menù a tendina *Properties*) e clicca per espanderlo. In corrispondenza della voce *Column*, scegli l'attributo *interaction*, e come *Mapping Type* seleziona *Discrete Mapping*. L'attributo *interaction* ha due possibili valori: *pd*

(protein-dna) e **pp** (protein-protein). Associa un colore al valore **pd** cliccandovi sopra. In modo, analogo scegli un colore per il valore **pp**. In tal modo, diventa possibile distinguere archi che corrispondono a interazioni dna-proteina da archi che corrispondono a interazioni proteina-proteina. Questo è un esempio di visualizzazione di una proprietà discreta (con un numero finito di valori distinti).

6. Per distinguere ancor meglio tra i due tipi d'interazione, associa alle interazioni proteina-proteina i seguenti attributi visuali: linea tratteggiata di spessore 4.0.
7. In ciascuna interazione dna-proteina, un nodo rappresenta una proteina e l'altro rappresenta un gene: esiste quindi una direzionalità nell'interazione. Per visualizzare tale direzionalità, rendi gli archi corrispondenti a interazioni dna-proteina orientati (usa l'attributo Target Arrow Shape).
8. Etichetta i nodi della rete con il nome corrispondente: seleziona l'attributo **Label** e usa **Passthrough Mapping** come **Mapping Type**. Questo è un esempio di come si possa associare un valore distinto (in questo caso, un'etichetta) a ciascun nodo della rete.
9. Prova a individuare gli *hub* della rete (gli hub sono nodi con molte interazioni): per fare ciò, assicurati di aver analizzato la rete con il **NetworkAnalyzer** e associa il grado di ciascun nodo alla dimensione del nodo stesso usando **Continuous Mapping** come **Mapping Type**. Cerca di rendere ben distinta la dimensione di nodi di grado elevato da quella dei nodi con grado basso usando il **Continuous Mapping Editor**:



Quali sono gli hub? Qual è il loro nome? Il numero di hub corrisponde a quanto si può dedurre dal grafico della distribuzione dei gradi dei nodi (Tools → NetworkANalyzer → Network Analysis → Plot Parameters...)?

2.2 Associare attributi alla rete

Selezionando un nodo si possono esaminare i suoi attributi nel Table Panel. Gli attributi disponibili sono accessibili cliccando su . Inizialmente, i nodi hanno solo un “id” e un “canonicalName” (che coincide con l’id). Se avete eseguito l’analisi della rete, vi saranno diversi altri attributi creati dal plugin di analisi.

Proviamo ad aggiungere alla rete le informazioni relative a un sottoinsieme dei risultati degli esperimenti di microarray descritti in [1]. In tali esperimenti, un gene alla volta è stato silenziato e le variazioni d’espressione genica dell’organismo mutante sono state confrontate con i livelli di un organismo di riferimento (lievito *wild-type*).

Scegli `File → Import → Table → File...`, e apri il file `galExpData.pvals`, quindi premi OK. I valori vengono aggiunti nel Table Panel (se non risultano visibili, selezionali da ). I livelli di espressione sono dati da `gal1RGexp`, `gal4RGexp` e `gal80Rexp` e sono calcolati come il logaritmo in base 10 del rapporto tra il livello d’espressione in un mutante in cui un gene (rispettivamente, Gal1, Gal4 e Gal80) è stato silenziato e il livello d’espressione nel wild-type. I livelli di espressione genica hanno anche associato un valore di significatività statistica: `gal1Rgsig`, `gal4Rgsig` e `gal80Rsig` rappresentano il grado di significatività di un cambiamento d’espressione (sono p-value).

1. Seleziona il nodo YBR020W. Come s’interpreta il valore negativo che si trova in `gal1RGexp`?
2. Per avere un riscontro visivo delle variazioni di espressione, associa al colore di un nodo (Fill Color) il valore dell’attributo `gal80Rexp`, in modo che sfumature più scure corrispondano a espressioni negative più intense e sfumature verso il bianco corrispondano a espressioni positive più intense. Quali sono i geni il cui livello d’espressione è influenzato maggiormente dal silenziamento di Gal80?
3. Associa la dimensione dell’etichetta di un nodo (Label Font Size) al livello di significatività dei valori `gal80Rsig` in modo che etichette più grandi corrispondano a un p-value più basso. Regola la visualizzazione in modo che solo le etichette corrispondenti a nodi con p-value $< 10^{-5}$ siano visibili. I geni individuati al punto precedente hanno un p-value significativo?

2.3 Un esempio di analisi biologica

Cerchiamo di trarre alcune conclusioni utili dai dati di cui disponiamo. A tal fine, è conveniente associare ai nodi i nomi comuni dei geni corrispondenti, che possono essere caricati dal file `ORF2name.csv` scegliendo `File → Import → Table → File...`

1. Cerca e seleziona il gene GAL4.
2. Seleziona i nodi adiacenti e crea una nuova sottorete con i nodi selezionati: `File → New → Network → From selected nodes, all edges`.

3. Per meglio visualizzare la nuova rete, elimina gli attributi visivi precedentemente creati.
4. Seleziona un layout gerarchico: Layout → yFiles Layouts → Hierarchic.
5. Distingui le interazioni proteina-proteina da quelle dna-proteina associando a queste ultime archi diretti di colore diverso.
6. Quali sono i tre geni la cui espressione è maggiore?
7. Quali sono i geni che interagiscono con tutti e tre?
8. Come varia il livello d'espressione genica di questi ultimi?
9. Che cosa riesci a dedurre sull'interazione tra tali geni?
10. Gal80 induce o inibisce l'attività di Gal4?

3 Ricerca di moduli

3.1 MCODE

Un *cluster* è un sottografo della rete i cui nodi sono densamente interconnessi. L'app MCODE di Cytoscape consente di individuare i cluster all'interno di una rete. Il motivo per cui è utile cercare cluster è che essi sono potenzialmente complessi molecolari con una ben definita funzione (ad esempio, in una rete d'interazioni proteiche, un cluster è un insieme putative di proteine che formano un complesso). Nella rete considerata, il coefficiente di clustering medio è relativamente alto (rispetto a una rete comparabile per numero di nodi e archi generata in modo casuale), perciò è ragionevole aspettarsi che vi siano alcuni cluster significativi.

1. Torna al network di partenza (galFiltered.sif) e seleziona Apps → MCODE → Open MCODE.
2. Clicca su Analyze Current Network.

Il risultato dell'analisi fatta da MCODE è riportato in nuovi attributi visibili nel Table Panel, nonché in una nuova finestra. I risultati più significativi (quelli con punteggio maggiore) dovrebbero corrispondere a geni coinvolti in un medesimo processo biologico. Per verificare tale ipotesi, recuperiamo le annotazioni funzionali associate usando l'app BiNGO:

1. Nella finestra dei risultati di MCODE seleziona il cluster 1.
2. Scegli Apps → BiNGO.
3. Assegna un nome arbitrario al cluster.
4. Nel menù a tendina Select reference set seleziona Use network as reference set.
5. In Select ontology file seleziona GO_biological_process.
6. In Select organism/annotation seleziona Saccharomyces cerevisiae.

7. **Start BiNGO**. Il risultato è un grafo i cui nodi sono etichettati con termini estratti dall'ontologia prescelta (processi biologici). I nodi colorati con un'arancione più intenso sono quelli maggiormente significativi.
8. Quale processo biologico è associato al network che abbiamo analizzato?

3.2 Ricerca di moduli

Il plugin *jActiveModules* consente di trovare moduli perturbati all'interno della rete sulla base di dati di espressione genica, ossia (nel caso specifico) sottoinsiemi di geni in cui (quasi) tutti i geni presentano una risposta significativa alle stesse condizioni sperimentali.

1. Torna a visualizzare il network `galFiltered.sif`.
2. Seleziona **Apps** → **jActiveModules**.
3. Nella finestra a sinistra, selezionare le righe corrispondenti alle tre condizioni sperimentali (`gal1RGsig`, `gal4RGsig`, `gal80Rsig`).
4. Premi il pulsante **Search**.

Il plugin trova cinque moduli putativi: maggiore è lo **Score**, più il modulo è significativo. Il primo modulo ha uno score elevato ed è composto da 17 nodi. Tale modulo risulta essere significativo in tutte e tre le condizioni sperimentali.

1. Crea una nuova sottorete a partire dai nodi del modulo 1: ritrovi elementi riconducibili a un pathway metabolico noto?
2. Conferma le tue deduzioni cercando su NCBI (<http://www.ncbi.nlm.nih.gov/>) i geni coinvolti (usa gli accession number dei nodi per la ricerca).

Riferimenti bibliografici

- [1] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518):929–934, 2001.