

# Analisi di reti metaboliche basata su proprietà di connessione

Massimo Nocentini

`massimo.nocentini@gmail.com`

Università degli Studi di Firenze

Firenze, 22 febbraio 2012

# Contenuti

## 1 Motivazioni

- Analisi di reti metaboliche
- Il problema delle storie

## 2 Nostro contributo

- Obiettivi
- Metodologia
- Risultati



# Contenuti

## 1 Motivazioni

- Analisi di reti metaboliche
- Il problema delle storie

## 2 Nostro contributo

- Obiettivi
- Metodologia
- Risultati



# Contenuti

## 1 Motivazioni

- Analisi di reti metaboliche
- Il problema delle storie

## 2 Nostro contributo

- Obiettivi
- Metodologia
- Risultati



# Definizione di rete metabolica

## Definizione

Una rete metabolica è un insieme di reazioni chimiche che si verificano all'interno di una cellula

È interessante studiarle per:

- capire le proprietà fisiologiche e biochimiche delle cellule
- ricostruire le reazioni che avvengono all'interno di organismi, sia batteri che esseri umani
- studiare il comportamento della cellula in relazione al contesto che la ospita



# Codifica delle reti

**SBML (Systems Biology Markup Language)** è un linguaggio per descrivere sistemi oggetto di trasformazioni fisiche

```
<sbml>
  <model id="ACYPI" name="Acyrtosiphon pisum">
    <listOfCompartments>
      <compartment id="CCO_45_IN" name="CCO-IN" />
      <compartment id="CCO_45_OUT" name="CCO-OUT" />
    </listOfCompartments>
    <listOfSpecies>
      <species id="ADENYLOSUCC_IN_NIL"
        name="adenylo-succinate"
        compartment="CCO-IN"/>
      ...
    </listOfSpecies>
    <listOfReactions>
      <reaction id="CITRAMALYL_45_COA_45_LYASE_45_RXN"
        name="Citramalyl-CoA lyase"
        reversible="false">
        <listOfReactants>
          <speciesReference species="CPD_45_627_IN_NIL"/>
        </listOfReactants>
        <listOfProducts>
          <speciesReference species="PYRUVATE_IN_NIL"/>
          <speciesReference species="ACETYL_45_COA_IN_NIL"/>
        </listOfProducts>
      </reaction>
      ...
    </listOfReactions>
  </model>
</sbml>
```



# Contenuti

## 1 Motivazioni

- Analisi di reti metaboliche
- Il problema delle storie

## 2 Nostro contributo

- Obiettivi
- Metodologia
- Risultati

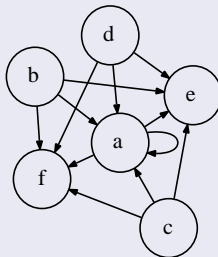


# Dalla rete al grafo

Data una reazione *non reversibile*  $r$  tale che  
 $reagenti(r) = \{r_1, \dots, r_n\} \wedge prodotti(r) = \{p_1, \dots, p_m\}$ ,  
costruiamo il sotto grafo  $reagenti(r) \times prodotti(r)$

## Esempio

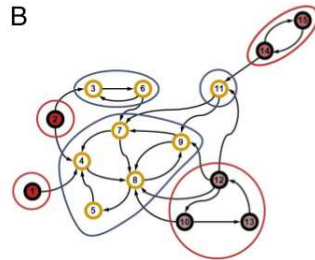
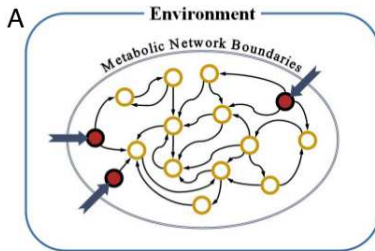
Se  $reagenti(r) = \{a, b, c, d\} \wedge prodotti(r) = \{a, e, f\}$  allora





# Seed Compounds

- cercare le componenti fortemente connesse
- selezionare vertici in componenti sorgenti come *seed compounds*



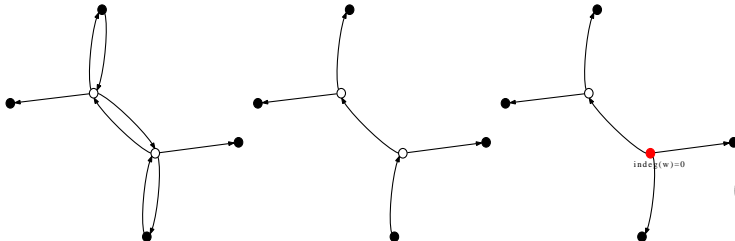
# Storie

## Definizione

Dato un grafo orientato  $G = (\mathbb{B} \cup \mathbb{W}, E)$ , una *storia* è un sotto grafo aciclico  $G' = (\mathbb{B} \cup \mathbb{W}', E')$  di  $G$  tale che  $E' \subseteq E$  e

$$\mathbb{W}' = \{w \in \mathbb{W} : \text{indeg}(w) > 0 \wedge \text{outdeg}(w) > 0\}$$

$\mathbb{B}$ : vertici a cui è permesso essere sorgenti o pozzi



# Contenuti

## 1 Motivazioni

- Analisi di reti metaboliche
- Il problema delle storie

## 2 Nostro contributo

- Obiettivi
- Metodologia
- Risultati



# Analizzare reti, costruire $\mathbb{B}$ e verificare il metodo

Gli obiettivi del nostro lavoro sono:

- rappresentare una rete mediante un grafo  
astraendo dai molti dettagli di *SBML*
- fornire strumenti per analizzare insiemi di reti  
per avere informazioni sui metaboliti che appaiono in più di una rete
- costruire in modo automatico l'insieme  $\mathbb{B}$   
sfruttando le informazioni di tutte le reti studiate
- verificare se il metodo è accettabile  
per misurare la validità di  $\mathbb{B}$



# Analizzare reti, costruire $\mathbb{B}$ e verificare il metodo

Gli obiettivi del nostro lavoro sono:

- rappresentare una rete mediante un grafo  
astraendo dai molti dettagli di *SBML*
- fornire strumenti per analizzare insiemi di reti  
per avere informazioni sui metaboliti che appaiono in più di una rete
- costruire in modo automatico l'insieme  $\mathbb{B}$   
sfruttando le informazioni di tutte le reti studiate
- verificare se il metodo è accettabile  
per misurare la validità di  $\mathbb{B}$



# Analizzare reti, costruire $\mathbb{B}$ e verificare il metodo

Gli obiettivi del nostro lavoro sono:

- rappresentare una rete mediante un grafo  
astraendo dai molti dettagli di *SBML*
- fornire strumenti per analizzare insiemi di reti  
per avere informazioni sui metaboliti che appaiono in più di una rete
- costruire in modo automatico l'insieme  $\mathbb{B}$   
sfruttando le informazioni di tutte le reti studiate
- verificare se il metodo è accettabile  
per misurare la validità di  $\mathbb{B}$



# Analizzare reti, costruire $\mathbb{B}$ e verificare il metodo

Gli obiettivi del nostro lavoro sono:

- rappresentare una rete mediante un grafo  
astraendo dai molti dettagli di *SBML*
- fornire strumenti per analizzare insiemi di reti  
per avere informazioni sui metaboliti che appaiono in più di una rete
- costruire in modo automatico l'insieme  $\mathbb{B}$   
sfruttando le informazioni di tutte le reti studiate
- verificare se il metodo è accettabile  
per misurare la validità di  $\mathbb{B}$



# Contenuti

## 1 Motivazioni

- Analisi di reti metaboliche
- Il problema delle storie

## 2 Nostro contributo

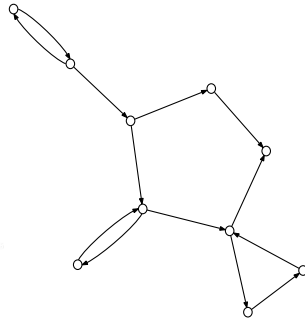
- Obiettivi
- **Metodologia**
- Risultati



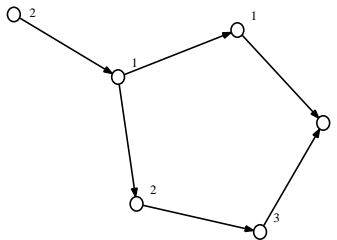
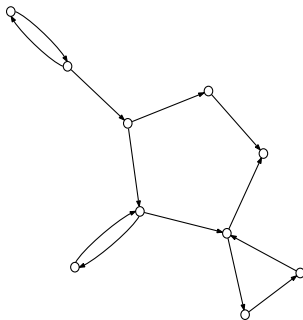


# Rappresentazione della rete con grafo

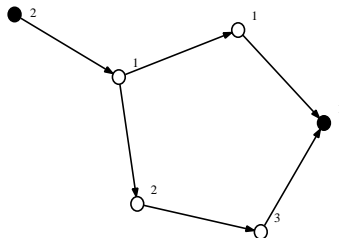
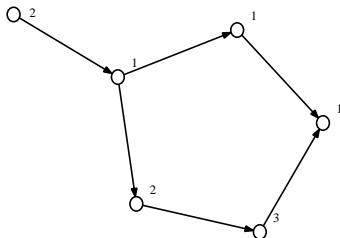
```
<sbml>
  <model id="ACYPI" name="Acyrtosiphon pisum">
    <listOfCompartments>
      <compartment id="CCO_45_IN" name="CCO-IN" />
      <compartment id="CCO_45_OUT" name="CCO-OUT" />
    </listOfCompartments>
    <listOfSpecies>
      <species id="ADENYLOSUCC_IN_NIL"
        name="adenylo-succinate"
        compartment="CCO-IN"/>
      ...
    </listOfSpecies>
    <listOfReactions>
      <reaction id="CITRAMALYL_45_COA_45_LYASE_45_RXN"
        name="Citramalyl-CoA lyase"
        reversible="false">
        <listOfReactants>
          <speciesReference species="CPD_45_627_IN_NIL"/>
        </listOfReactants>
        <listOfProducts>
          <speciesReference species="PYRUVATE_IN_NIL"/>
          <speciesReference species="ACETYL_45_COA_IN_NIL"/>
        </listOfProducts>
      </reaction>
      ...
    </listOfReactions>
  </model>
</sbml>
```



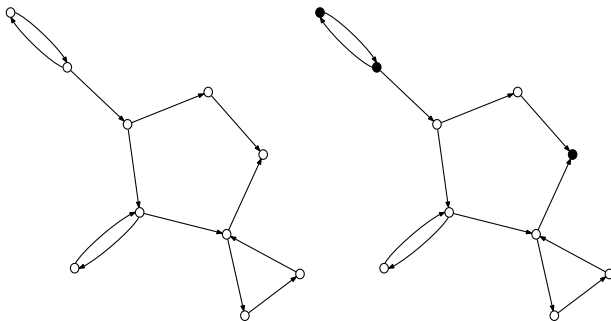
# Componenti fortemente connesse



# Componenti sorgenti e pozzo



# Selezione dell'insieme $\mathbb{B}$



# Contenuti

## 1 Motivazioni

- Analisi di reti metaboliche
- Il problema delle storie

## 2 Nostro contributo

- Obiettivi
- Metodologia
- Risultati



# Analisi delle reti

Models (165)	Sources	Whites	Sinks
Wigglesworthia gloss...	(C: 65, V: 65:6)	(C: 11, V: 535:250)	(C: 76, V: 94:21)
Wolbachia endosym...	(C: 61, V: 61:9)	(C: 14, V: 619:326)	(C: 73, V: 91:20)
Wolbachia pipientis...	(C: 71, V: 71:8)	(C: 14, V: 581:282)	(C: 75, V: 93:21)
Wolinella succinogen...	(C: 78, V: 78:14)	(C: 12, V: 848:420)	(C: 105, V: 136:43)
Xylella fastidiosa-XYL...	(C: 67, V: 68:14)	(C: 9, V: 996:546)	(C: 103, V: 125:34)
Yersinia pestis-YERPE3	(C: 132, V: 135:40)	(C: 19, V: 1348:711)	(C: 152, V: 186:83)
Yersinia pseudotube...	(C: 122, V: 122:36)	(C: 22, V: 1465:781)	(C: 156, V: 192:69)
Average pairs	(C: 165, V: 168:0)	(C: 17, V: 1167:0)	(C: 191, V: 220:0)

Tutte le reti hanno una struttura a “clessidra”:

- molte componenti *sorgenti* contenenti pochi vertici
- poche componenti *intermedie* contenenti molti vertici
- molte componenti *pozzo* contenenti pochi vertici

Troppi vertici sorgenti e pozzi

L'obiettivo sarebbe stato averne pochi



# Analisi delle reti

Models (165)	Sources	Whites	Sinks
Wigglesworthia gloss...	(C: 65, V: 65:6)	(C: 11, V: 535:250)	(C: 76, V: 94:21)
Wolbachia endosym...	(C: 61, V: 61:9)	(C: 14, V: 619:326)	(C: 73, V: 91:20)
Wolbachia pipientis...	(C: 71, V: 71:8)	(C: 14, V: 581:282)	(C: 75, V: 93:21)
Wolinella succinogen...	(C: 78, V: 78:14)	(C: 12, V: 848:420)	(C: 105, V: 136:43)
Xylella fastidiosa-XYL...	(C: 67, V: 68:14)	(C: 9, V: 996:546)	(C: 103, V: 125:34)
Yersinia pestis-YERPE3	(C: 132, V: 135:40)	(C: 19, V: 1348:711)	(C: 152, V: 186:83)
Yersinia pseudotube...	(C: 122, V: 122:36)	(C: 22, V: 1465:781)	(C: 156, V: 192:69)
Average pairs	(C: 165, V: 168:0)	(C: 17, V: 1167:0)	(C: 191, V: 220:0)

Tutte le reti hanno una struttura a “clessidra”:

- molte componenti *sorgenti* contenenti pochi vertici
- poche componenti *intermedie* contenenti molti vertici
- molte componenti *pozzo* contenenti pochi vertici

Troppi vertici sorgenti e pozzi

L'obiettivo sarebbe stato averne pochi



# Analisi dei vertici

```
(types: [Sources], count: 1586, distribution: 12.791%)  
(types: [Sinks], count: 1448, distribution: 11.678%)  
(types: [Whites], count: 7705, distribution: 62.142%)  
(types: [Sinks, Sources], count: 38, distribution: 0.306%)  
(types: [Whites, Sources], count: 538, distribution: 4.339%)  
(types: [Whites, Sinks], count: 847, distribution: 6.831%)  
(types: [Whites, Sinks, Sources], count: 237, distribution: 1.911%)
```

Studiando un insieme di reti, più del 10% dei vertici non hanno un ruolo univoco

Incertezza sulla costruzione di  $\mathbb{B}$

Vertici con più di un ruolo inducono incertezza sul decidere la loro appartenenza all'insieme  $\mathbb{B}$





# Analisi dei vertici

```
(types: [Sources], count: 1586, distribution: 12.791%)
(types: [Sinks], count: 1448, distribution: 11.678%)
(types: [Whites], count: 7705, distribution: 62.142%)
(types: [Sinks, Sources], count: 38, distribution: 0.306%)
(types: [Whites, Sources], count: 538, distribution: 4.339%)
(types: [Whites, Sinks], count: 847, distribution: 6.831%)
(types: [Whites, Sinks, Sources], count: 237, distribution: 1.911%)
```

Studiando un insieme di reti, più del 10% dei vertici non hanno un ruolo univoco

## Incertezza sulla costruzione di $\mathbb{B}$

Vertici con più di un ruolo inducono incertezza sul decidere la loro appartenenza all'insieme  $\mathbb{B}$



# Combinazione delle due analisi

Models (165)	Sources	Whites	Sinks
Crithidia deanei-CDE...	(C: 153, V: 153:6)	(C: 18, V: 472:264)	(C: 152, V: 158:14)
Cupriavidus taiwane...	(C: 127, V: 129:38)	(C: 14, V: 1484:843)	(C: 164, V: 217:82)
Desulfotalea psychr...	(C: 75, V: 75:12)	(C: 14, V: 1118:606)	(C: 108, V: 131:31)
Drosophila melanog...	(C: 87, V: 87:11)	(C: 5, V: 978:559)	(C: 114, V: 138:37)
Erwinia carotovora s...	(C: 126, V: 127:46)	(C: 18, V: 1428:766)	(C: 152, V: 191:75)
Escherichia coli-EC4...	(C: 315, V: 320:220)	(C: 22, V: 1559:810)	(C: 337, V: 374:267)

Se consideriamo tutti i vertici che hanno un ruolo univoco rispetto ad un insieme di reti, è possibile raffinare ogni  $\mathbb{B}$  escludendo i vertici che non hanno un ruolo conforme



# Riepilogo

- costruire l'insieme  $\mathbb{B}$  in modo automatico attraverso le componenti fortemente connesse *non* sembra produrre risultati soddisfacenti
- l'unico caso utile riguarda l'analisi di modelli singoli, anche se in molti casi rimangono molti vertici in  $\mathbb{B}$



# Ricerca delle componenti fortemente connesse

Attualmente l'insieme  $\mathbb{B}$  viene costruito in base a osservazioni e studi *empirici*, costituito da metaboliti con proprietà particolari. Per costruirlo in modo automatico partizioniamo la rete in componenti connesse in quanto:

- è difficile assegnare il ruolo ad ogni vertice studiando l'intera rete date le sue dimensioni
- è possibile astrarre dai cicli ed identificare classi di metaboliti equivalenti
- due metaboliti equivalenti si producono a vicenda, pertanto gli associamo il ruolo della componente che li contiene
- se una componente è *sorgente* o *pozzo* nel “meta grafo” allora aggiungiamo i vertici che la compongono in  $\mathbb{B}$



# Ricerca delle componenti fortemente connesse

Attualmente l'insieme  $\mathbb{B}$  viene costruito in base a osservazioni e studi *empirici*, costituito da metaboliti con proprietà particolari. Per costruirlo in modo automatico partizioniamo la rete in componenti connesse in quanto:

- è difficile assegnare il ruolo ad ogni vertice studiando l'intera rete date le sue dimensioni
- è possibile astrarre dai cicli ed identificare classi di metaboliti equivalenti
- due metaboliti equivalenti si producono a vicenda, pertanto gli associamo il ruolo della componente che li contiene
- se una componente è *sorgente* o *pozzo* nel “meta grafo” allora aggiungiamo i vertici che la compongono in  $\mathbb{B}$



# Ricerca delle componenti fortemente connesse

Attualmente l'insieme  $\mathbb{B}$  viene costruito in base a osservazioni e studi *empirici*, costituito da metaboliti con proprietà particolari. Per costruirlo in modo automatico partizioniamo la rete in componenti connesse in quanto:

- è difficile assegnare il ruolo ad ogni vertice studiando l'intera rete date le sue dimensioni
- è possibile astrarre dai cicli ed identificare classi di metaboliti equivalenti
- due metaboliti equivalenti si producono a vicenda, pertanto gli associamo il ruolo della componente che li contiene
- se una componente è *sorgente* o *pozzo* nel “meta grafo” allora aggiungiamo i vertici che la compongono in  $\mathbb{B}$



# Ricerca delle componenti fortemente connesse

Attualmente l'insieme  $\mathbb{B}$  viene costruito in base a osservazioni e studi *empirici*, costituito da metaboliti con proprietà particolari. Per costruirlo in modo automatico partizioniamo la rete in componenti connesse in quanto:

- è difficile assegnare il ruolo ad ogni vertice studiando l'intera rete date le sue dimensioni
- è possibile astrarre dai cicli ed identificare classi di metaboliti equivalenti
- due metaboliti equivalenti si producono a vicenda, pertanto gli associamo il ruolo della componente che li contiene
- se una componente è *sorgente* o *pozzo* nel “meta grafo” allora aggiungiamo i vertici che la compongono in  $\mathbb{B}$



# Ricerca delle componenti fortemente connesse

Attualmente l'insieme  $\mathbb{B}$  viene costruito in base a osservazioni e studi *empirici*, costituito da metaboliti con proprietà particolari. Per costruirlo in modo automatico partizioniamo la rete in componenti connesse in quanto:

- è difficile assegnare il ruolo ad ogni vertice studiando l'intera rete date le sue dimensioni
- è possibile astrarre dai cicli ed identificare classi di metaboliti equivalenti
- due metaboliti equivalenti si producono a vicenda, pertanto gli associamo il ruolo della componente che li contiene
- se una componente è *sorgente* o *pozzo* nel “meta grafo” allora aggiungiamo i vertici che la compongono in  $\mathbb{B}$

