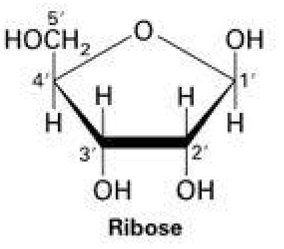
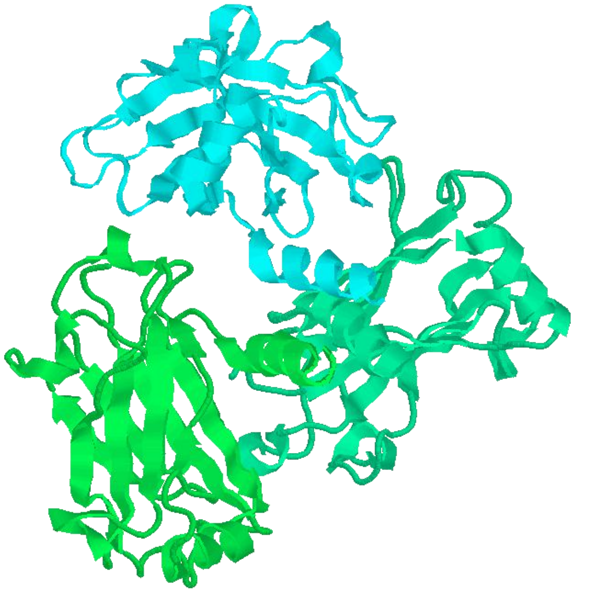
Traitement des graphes et réseaux biologiques

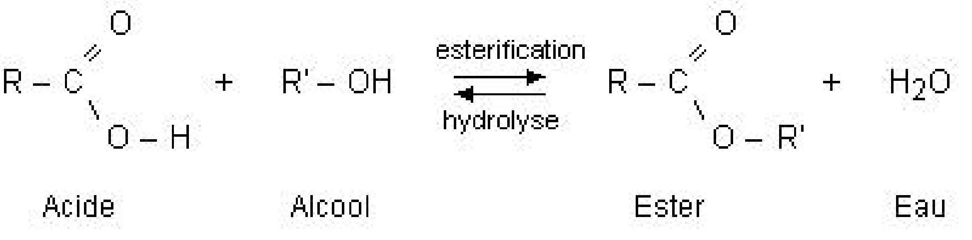
Master 1

Bioinformatique

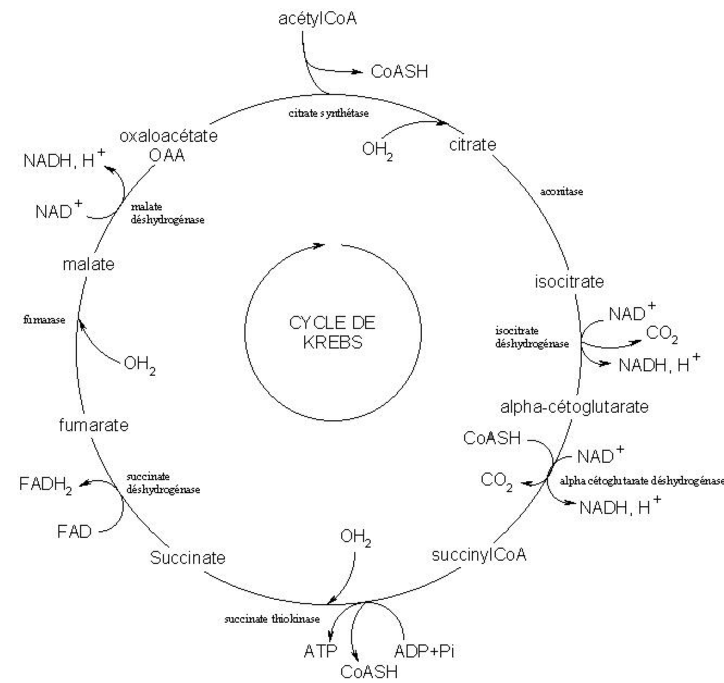
- Structures moléculaires et réactions
  - formats : SMILES, ...
  - banques : PDB, ...

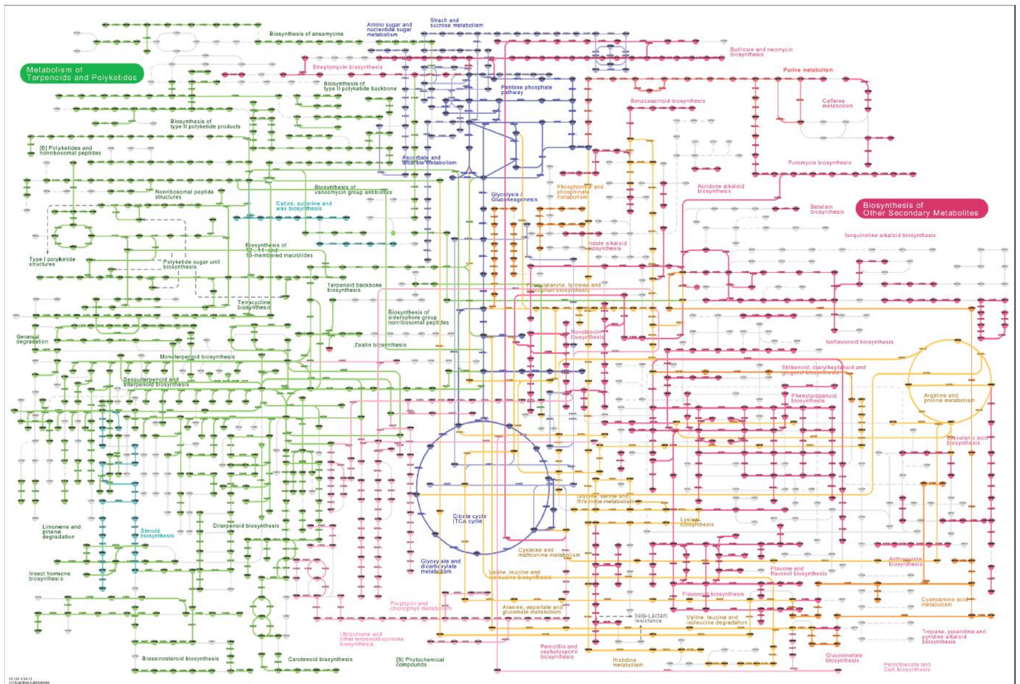


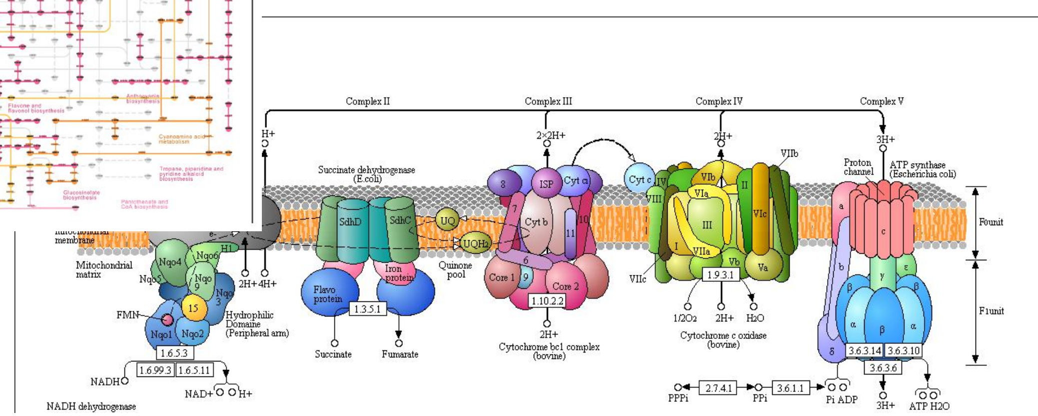




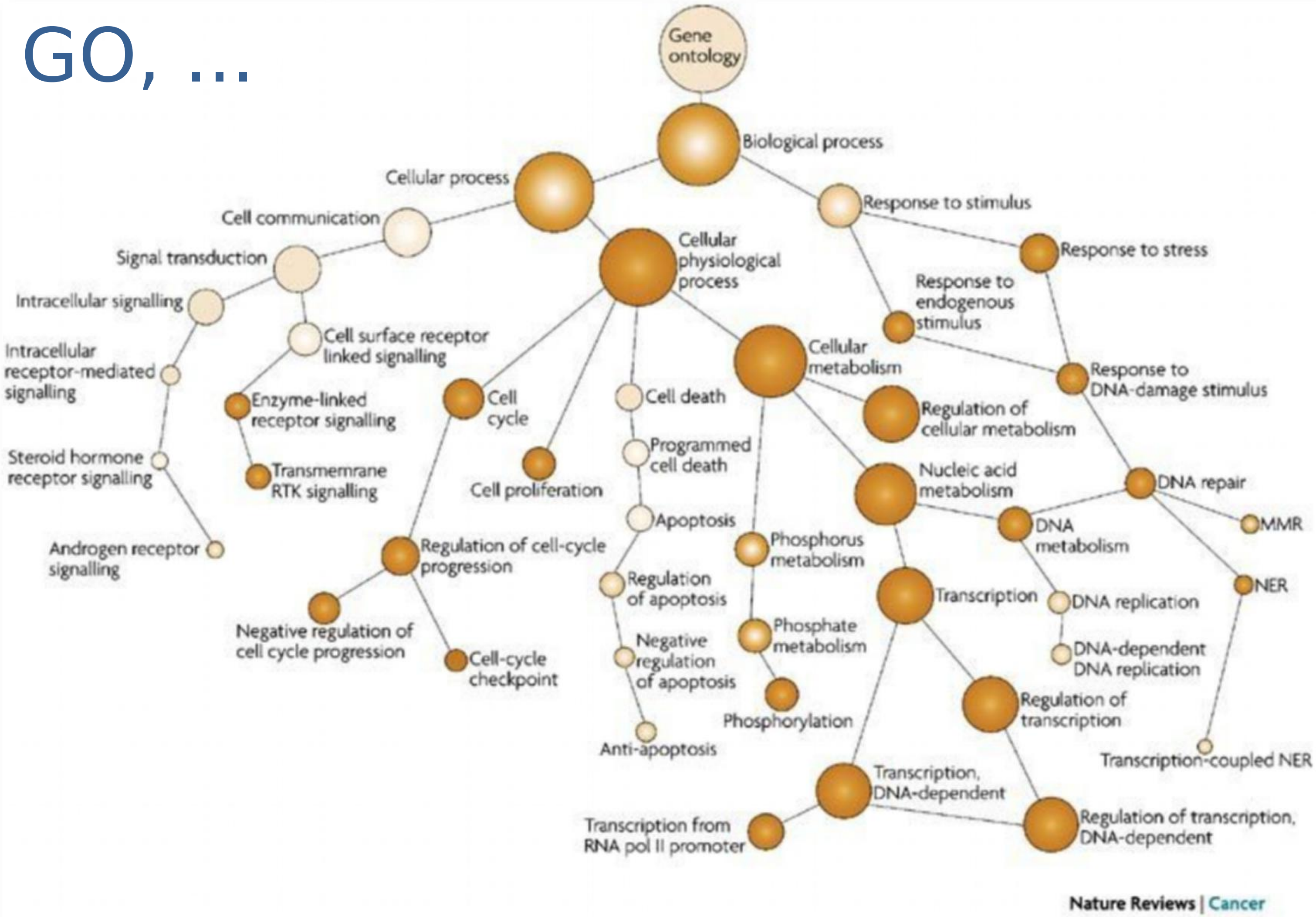
- Métabolisme
  - banques
    - KEGG
    - BioCyc, MetaCyc



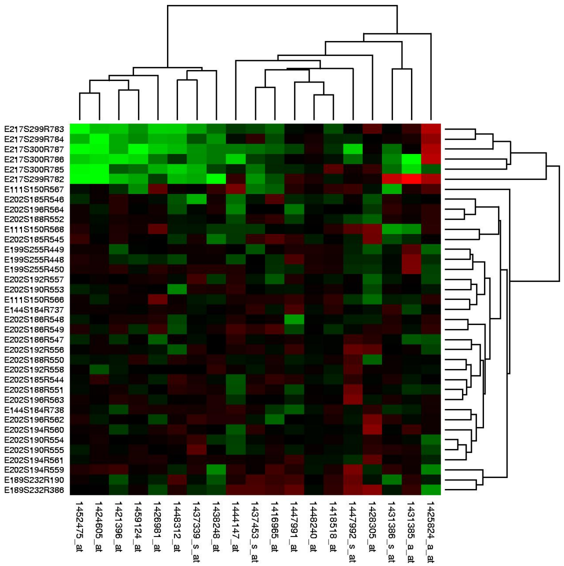


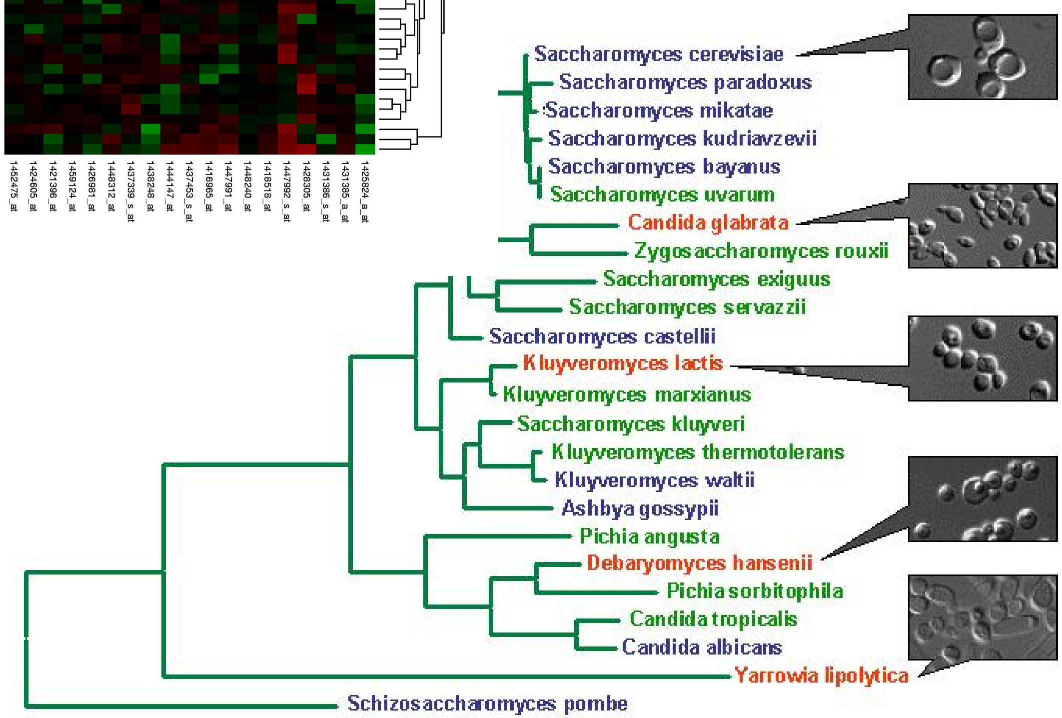


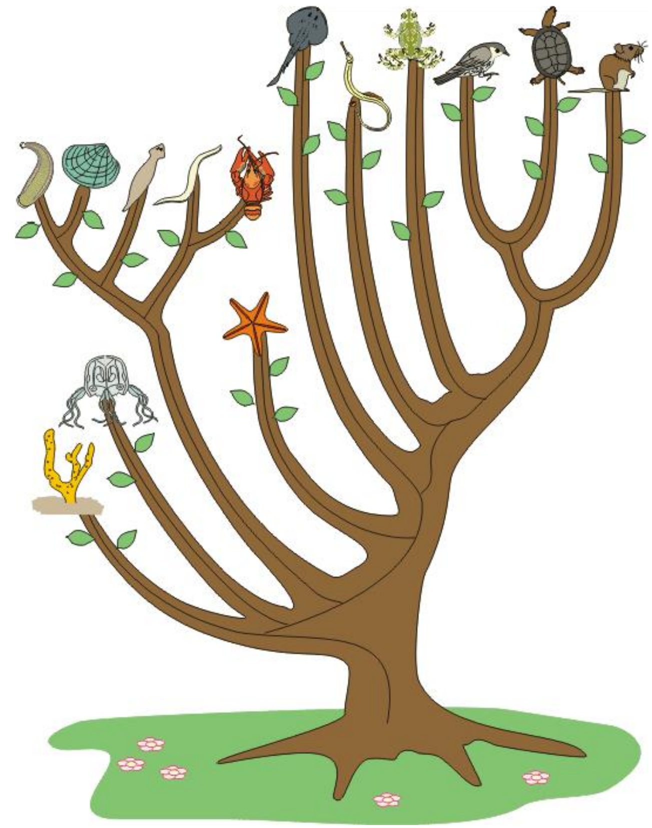
- Ontologies
  - GO, ...



- Arbres : clustering hiérarchique, phylogénie, ...

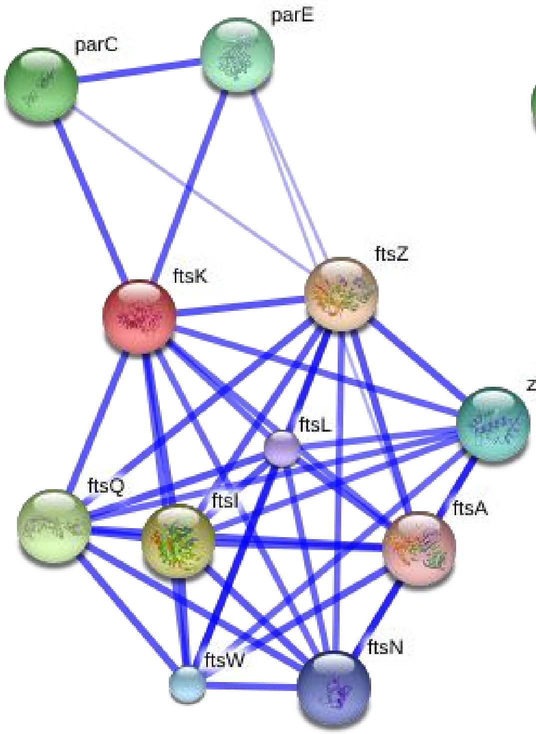


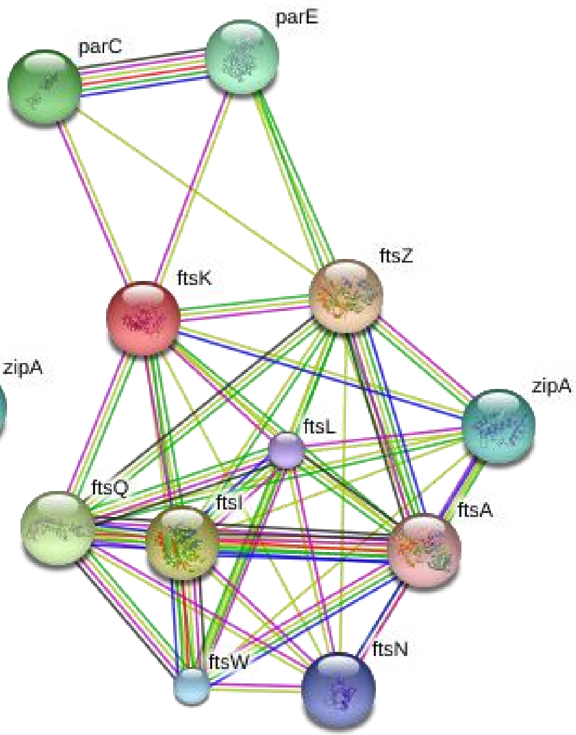


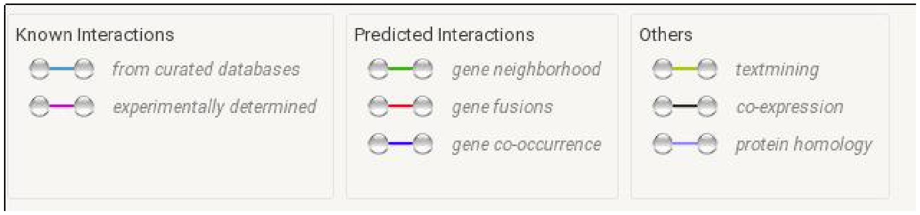


- Intéractions physiques, génétiques, fonctionnelles, ...
  - STRING, DIP, BIND, IntAct, BioGRID, iHOP, ...

Yeast protein-protein interaction network
[Jeong *et al.*, 2001]

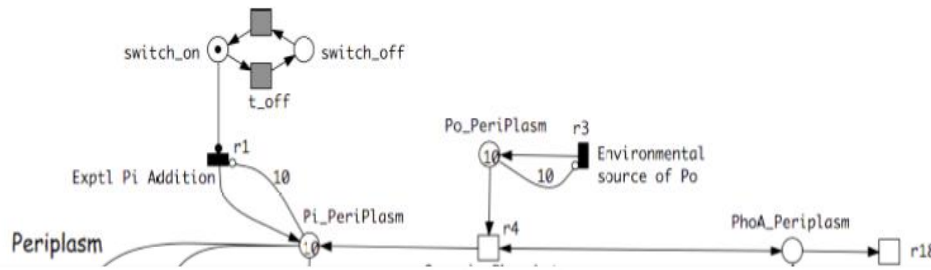


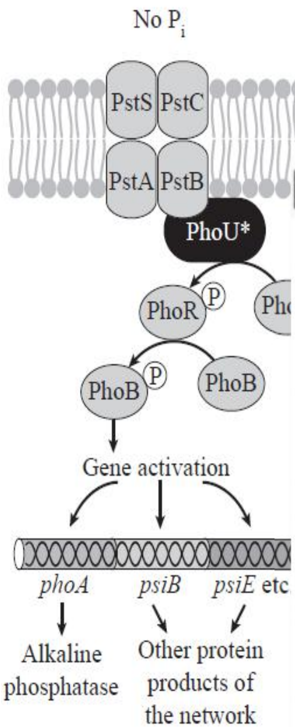




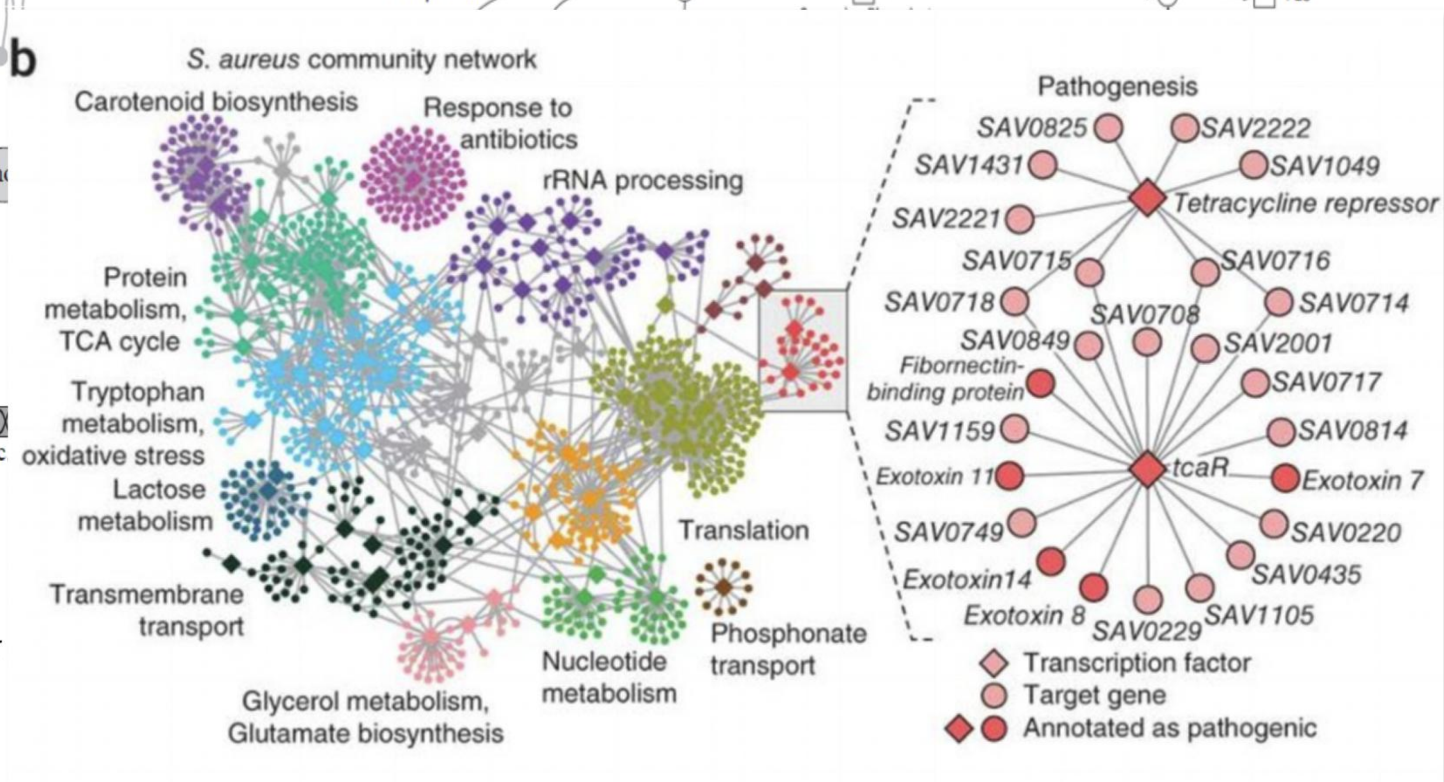
from string-db.org

- Réseaux de régulation
  - BioCyc, KEGG, TransFac, RegulonDB, ...
  - SBML, ...

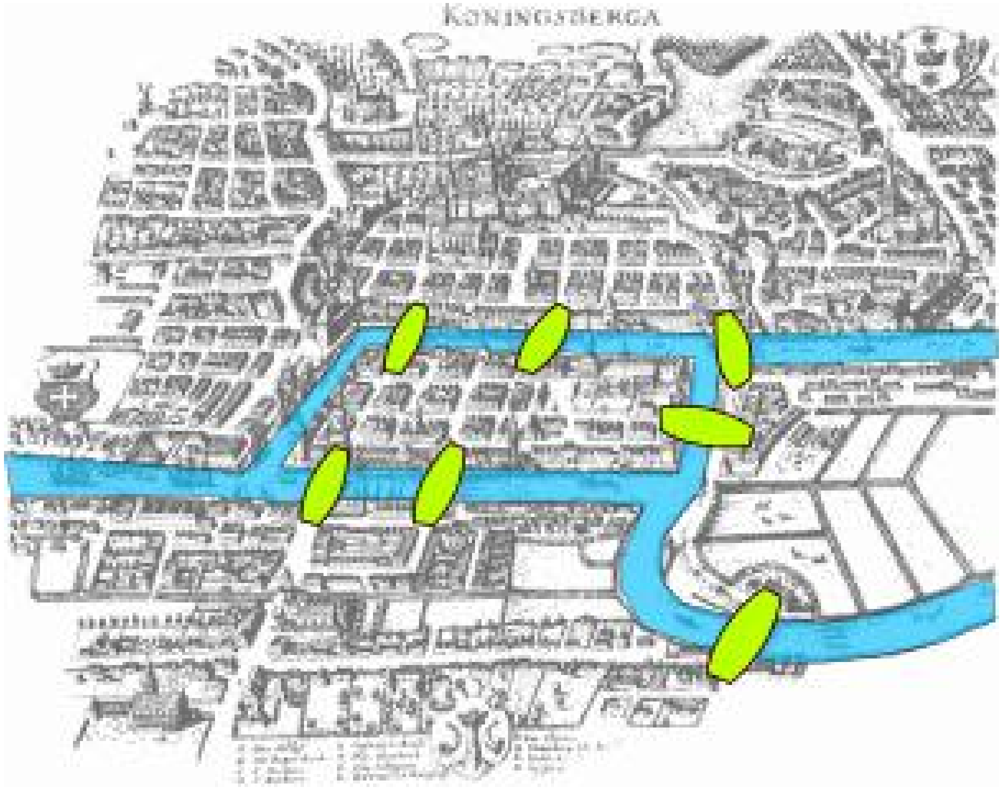


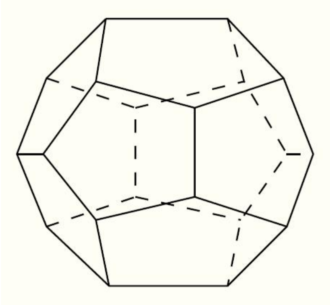


Neidhardt *et al.* 1990



- Historique
  - 1736 Euler
  - 1847 Kirchhof
    théorie des arbres
    analyse de circuits
    électriques
  - 1850 - 1880
    Cayley, arbres : isomères
    hydrocarbures C_nH_{2n+2}
    Conjecture des 4 couleurs
    (Möbius, Morgan, Cayley)
    (théorème de Appel et
    Haken en 1976)
    Chemin Hamiltonien
  - 1936 – König 1er ouvrage sur
    Théorie des graphes
  - 1946 –
    Kuhn, Ford et Fulkerson, Roy, Claude Berge, …





Applications

- tournées de distribution, ramassage d'ordure, inspection de réseaux de distribution
- Recherche de chemins, GPS
- tracé automatique (voies métaboliques, circuit imprimé)
- ordonnancement (chantier de construction, processus)
- maximisation de flux (FBA, routier, internet)
- minimisation du coût d'un flot (routage, métabolisme)

- Réseau métabolique, de régulation de l'expression des gènes
- Réseaux d'intéractions
- Arbre phylogénétique, de parenté
- Clustering (hiérarchique)
- Ontologies, ex : Gene Ontology
- …

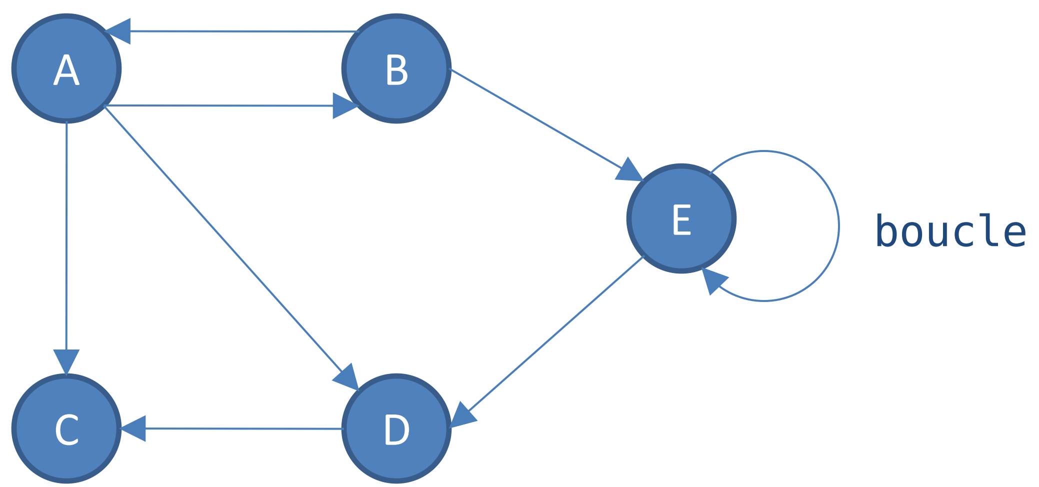
Terminologie, définitions et notations

- Définition
  - G = (*V*, *E*)
  - Ensemble *V* de **sommets**. Ils peuvent porter des **étiquettes** (nombre entier, couleur, …, peu importe)
  - Ensemble *E* d'**arête**s ou **arcs** de la forme (*v1*,*v2*). Les arêtes peuvent être étiquetées (nombre, mot, …).
  - Le nombre de sommets est appelé **ordre** du graphe

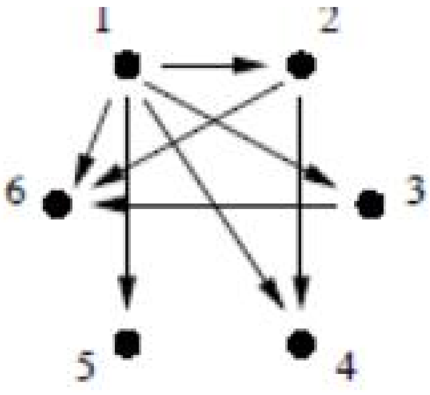
- Types de graphes
  - Graphe **non orienté**
    - arêtes : (*v1*,*v2*) $\leftrightarrow$ (*v2*,*v1*)
  - Graphe **orienté**
    - arcs : (*v1*,*v2*) ≠ (*v2*,*v1*)
  - **multigraphe** : possibilité d'avoir plusieurs arêtes reliant les 2 mêmes sommets
  - **hypergraphe** : une hyperarête peut relier plus de 2 sommets

Graphe orienté ou digraphe

- $G = (V,E)$
- $V = \{A, B, C, D, E\}$
- $E = \{ (A,B), (B,A), (A,C), (A,D), (B,E), (D,C), (E,D), (E,E) \}$
- Représentation **sagittale** :



- $G = (V,E)$
- un arc $a = (x,y) \in E$ peut être noté $x \rightarrow y$
- x et y sont les **extrémités** de a
- x est le **début** ou **origine** ou **extrémité initiale** de a. y est la **fin** ou l'**extrémité finale** de a.
- a est **sortant** en x et **incident** en y.
- x et y sont **adjacents**.
- y est un **successeur** de x.
  x est un **prédécesseur** de y.
- propriété caractéristique d'un graphe
  - ex : $\forall x, y \in V / (x,y) \in E \leftrightarrow x$ divise y



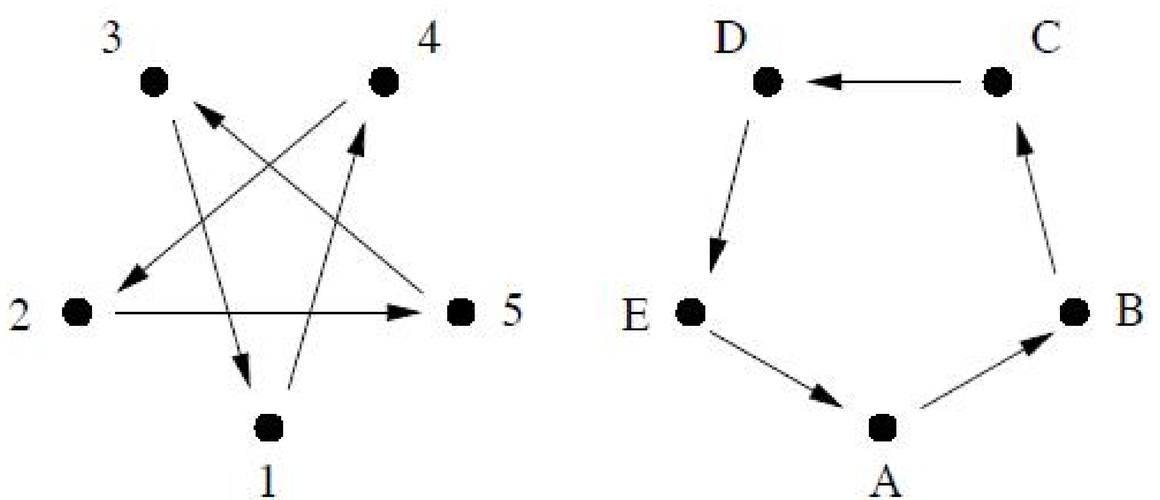
Isomorphisme de graphes

- 2 graphes orientés $G_1 = (V_1,E_1)$ et $G_2 = (V_2,E_2)$ sont **isomorphes** s'il existe une application bijective

  $f : V_1 \to V_2$ telle que

$$\forall x, y \in V_1 / (x,y) \in E_1 \Leftrightarrow (f(x), f(y)) \in E_2$$

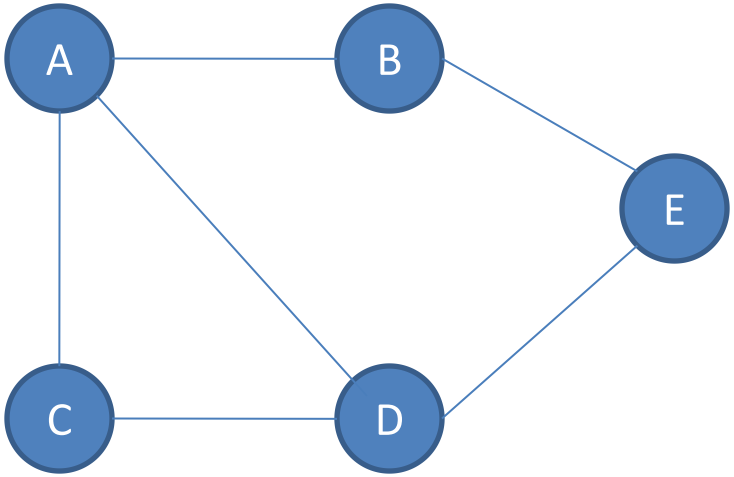
- Exemple :



$$f : \begin{cases} 1 \mapsto A \\ 2 \mapsto C \\ 3 \mapsto E \\ 4 \mapsto B \\ 5 \mapsto D \end{cases}$$

Graphe non orienté

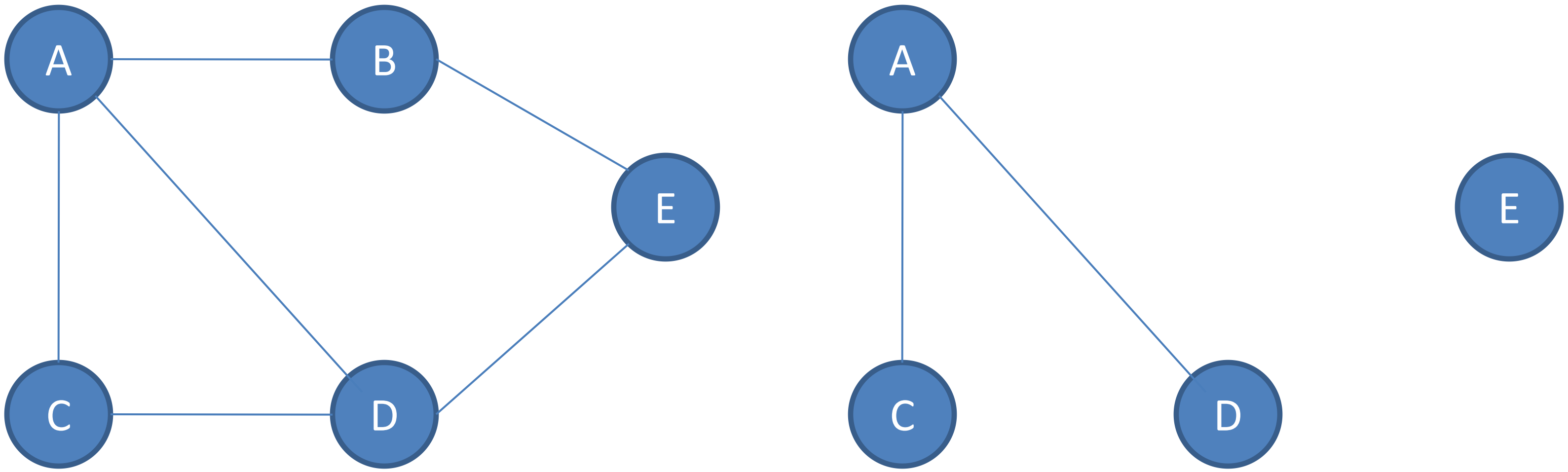
- $G = (V,E)$
- $V = \{A, B, C, D, E\}$
- $E = \{ (A,B), (A,C), (A,D), (B,E), (C,D), (D,E) \}$
- Représentation sagittale :



- On peut associer à un graphe orienté un graphe non orienté appelé **graphe non orienté associé** ou **sous-jacent**

- Soit $G = (V,E)$ un graphe (orienté ou non). Un **sous-graphe** de G est un graphe $G' = (V',E')$ tel que

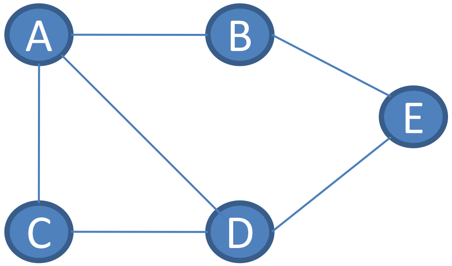
$$V' \subseteq V \text{ et } E' \subseteq E$$

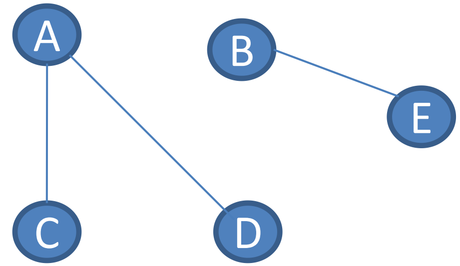


- Exemple : voie métabolique $\subset$ réseau métabolique

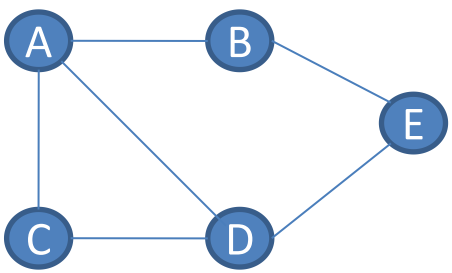
Sous-graphe induit, cliques et sous-graphe couvrant

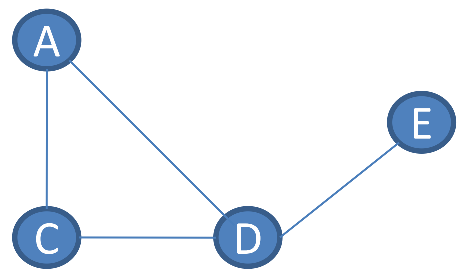
- Un sous-graphe *G'* d'un graphe *G* est **couvrant** si il contient tous les sommets de G



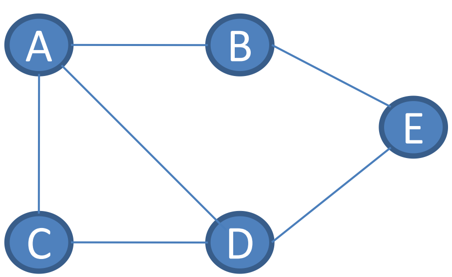


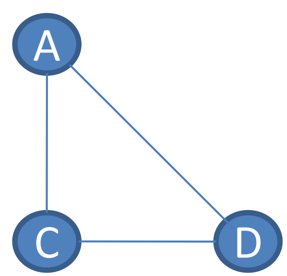
- Un sous-graphe *G'* d'un graphe *G* est un **sous-graphe induit** si *E'* est formé de tous les arcs (ou arêtes) de *G* ayant leurs extrémités dans *G'*





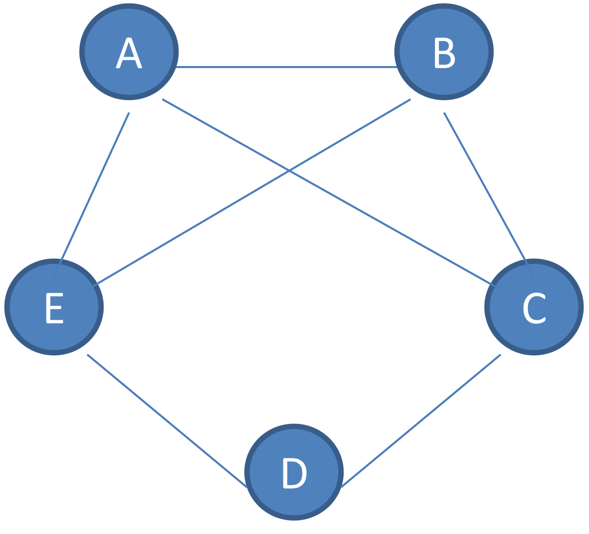
- Une **clique** est un sous-graphe induit complet



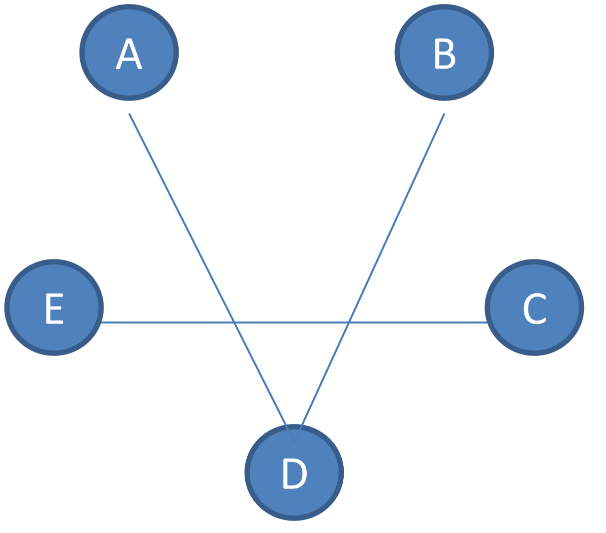


Complémentaire d'un graphe

- Le **complémentaire** ou **complément** ou **inverse** d'un graphe *G* est noté $\overline{G}$, il a les mêmes sommets qui sont reliés si et seulement si ils ne sont pas reliés dans *G*

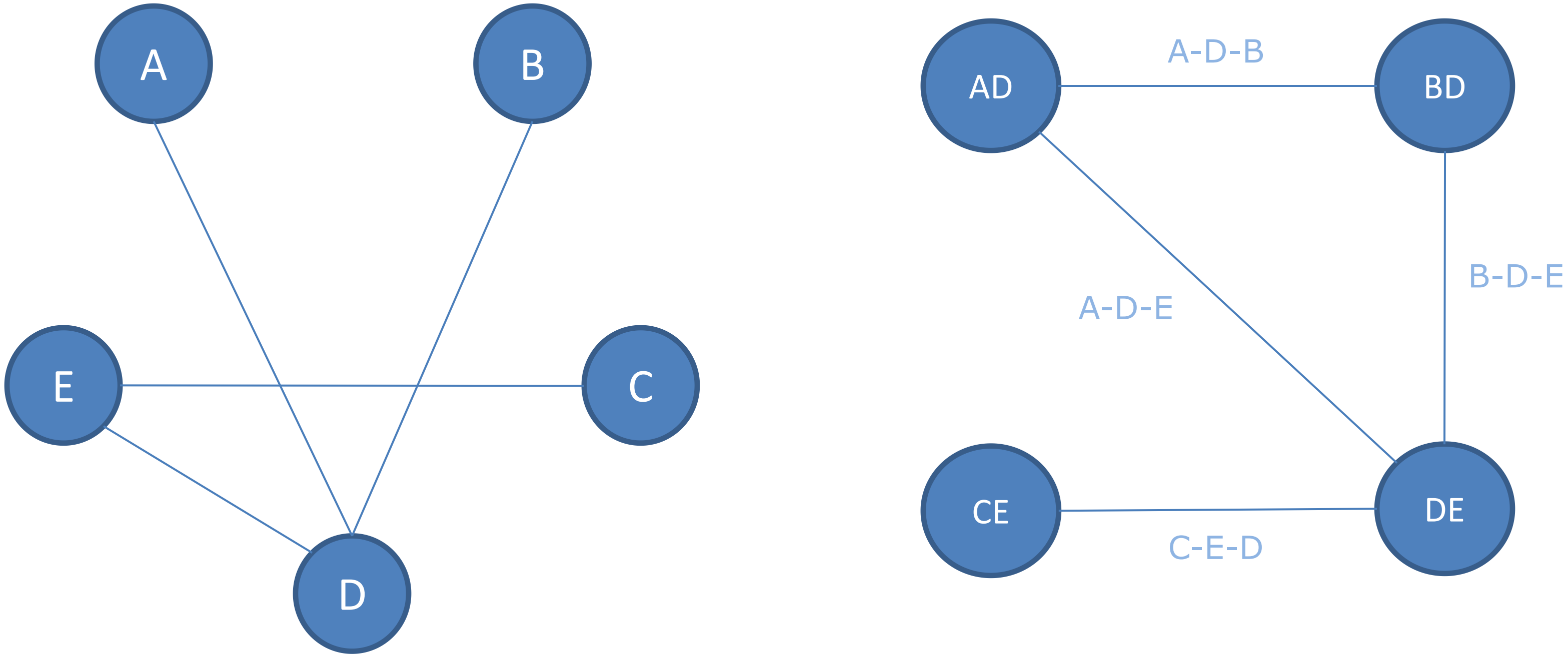


G



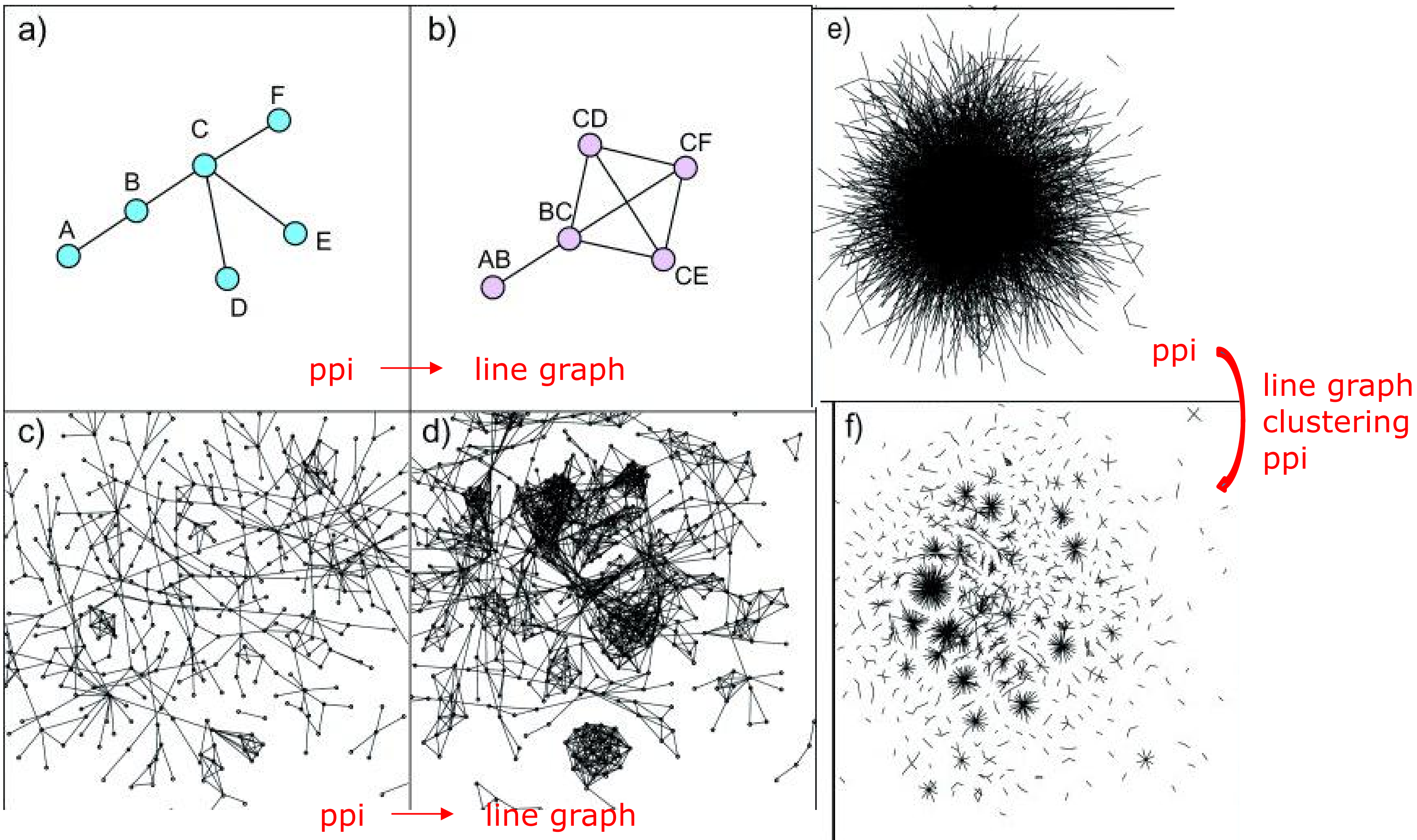
$\overline{G}$

- Le line graph d'un graphe *G* est le graphe L(*G*) dans lequel sont inversés sommets et arêtes, c'est-à-dire que **deux sommets adjacents** dans le line graph **correspondent à deux arêtes incidentes à un même sommet** dans *G*.



Detection of Functional Modules From Protein Interaction Networks

Jose B. Pereira-Leal,[1] Anton J. Enright,[2] and Christos A. Ouzounis[1*]
[1]Computational Genomics Group, The European Bioinformatics Institute, Cambridge, United Kingdom
[2]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York



Degré ou valence d'un sommet

- Pour un graphe orienté :
  - dégré entrant : nombre de prédécesseurs d'un sommet, noté $d_-(x)$
  - degré sortant : nombre de successeurs d'un sommet, noté $d_+(x)$
  - degré total : nombre d'arcs dont x est le début ou la fin (on compte donc 2 fois les boucles), $d(x) = d_-(x) + d_+(x)$

$$\sum_{x \in V} d_-(x) = \sum_{x \in V} d_+(x) = |E| \qquad \sum_{x \in V} d(x) = 2|E|$$

- Un sommet de degré entrant non nul et de degré sortant nul est un **puit**.
- Un sommet de degré entrant nul et de degré sortant non nul est appelé **source**.
- Un sommet de degré nul est un sommet **isolé**.
- Pour un graphe non orienté : $d(x)$

- Pour un graphe orienté
  - Un **chemin** C est une suite (x_0, x_1, $x_{...}$, $x_{n\text{-}1}$, x_n) de sommets de G tel que 2 sommets consécutifs quelconque x_i et x_{i+1} sont reliés par un arc de G
  - x_0 et x_n sont le début et la fin du chemin C
  - C est de longueur n (nombre d'arcs)

- Pour un graphe non orienté, on parle de **chaîne**

- Un **circuit** dans un graphe orienté est un chemin de longueur non nulle dont le début et la fin sont identiques.

- Un **chemin simple** ne passe par 2 fois par le même arc.

- Un **chemin élémentaire** ne passe pas 2 fois par le même sommet.

- Pour un graphe non orienté on parle de **cycle**.

Connexité

- Un graphe non orienté est **connexe** si pour tout couple de sommets *x*, *y*, il existe une chaîne reliant *x* à *y*.
- Une **composante connexe** d'un graphe est un sous ensemble **maximal** de sommets dont toutes les paires de sommets sont reliées par une chaîne.
- Les composantes connexes d'un graphe forment une **partition** du graphe.
- Un graphe orienté est **connexe** si son graphe non orienté associé est connexe.
- Un graphe orienté est **fortement connexe** si pour toute paire de sommets, il existe un chemin les reliant.
- Un **point d'articulation** est un sommet dont la suppression augmente le nombre de composantes connexes.
- Un **isthme** est une arête dont la suppression augmente le nombre de composantes connexes.
- Un **ensemble d'articulation** d'un graphe connexe est un ensemble de sommets dont la suppression rend le graphe non connexe.
- Un **pont** dans un graphe connexe est une arête dont la suppression déconnecte le graphe.

- Pour un graphe non orienté, une chaîne (respectivement un cycle) eulérienne passe une et une seule fois par toutes les arêtes du graphe.

- Pour un graphe orienté, un chemin (resp. un circuit) eulérien passe une et une seule fois par tous les arcs du graphe.

- Pour un graphe non orienté, une chaîne (respectivement un cycle) hamiltonienne passe une et une seule fois par tous les sommets du graphe.

- Pour un graphe orienté, un chemin (resp. un circuit) hamiltonien passe une et une seule fois par tous les sommets du graphe.