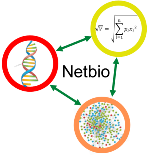
Networks: what? what for? how?



https://mia.toulouse.inra.fr/NETBIO

Julien Chiquet, Étienne Delannoy, Marie-Laure Martin-Magniette, Françoise Monéger,
Guillem Rigaill & Nathalie Villa-Vialaneix

Ecole chercheur SPS - November 30th 2017

Outline

1. What are networks/graphs?

2. What are networks useful for in biology?
   - Visualization
   - Simple analyses based on network topology
   - More advanced analyses based on network topology
   - Biological interaction models

3. How to build networks?

Outline

1. What are networks/graphs?

2. What are networks useful for in biology?
   - Visualization
   - Simple analyses based on network topology
   - More advanced analyses based on network topology
   - Biological interaction models

3. How to build networks?

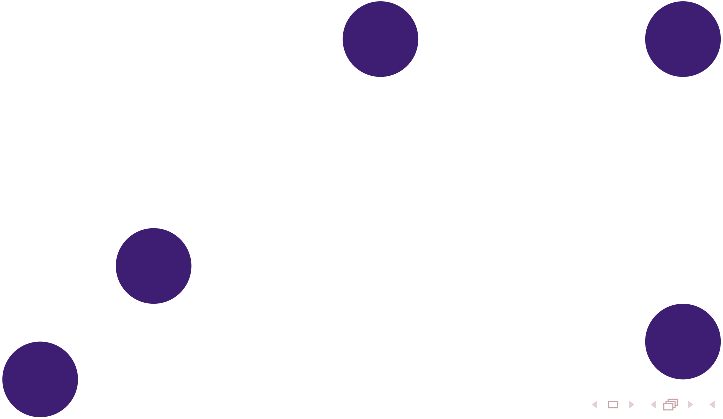
What is a graph? *graphe*

Mathematical object used to model **relational data between entities**.

What is a graph? *graphe*

Mathematical object used to model **relational data between entities**.

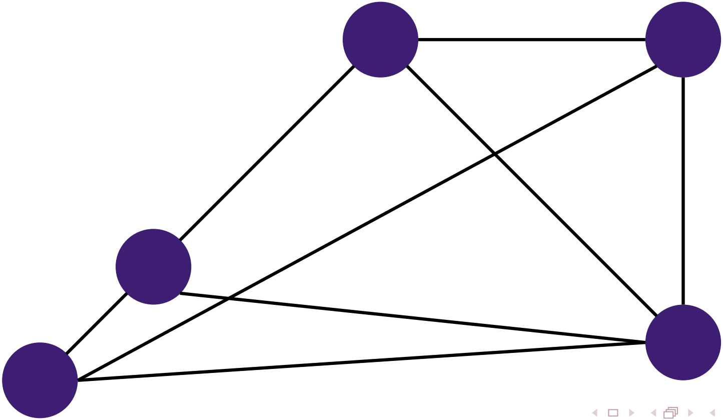
The entities are called **nodes** or **vertices**
nœuds/sommets



What is a graph? *graphe*

Mathematical object used to model **relational data between entities**.

A relation between two entities is modeled by an **edge**
arête



Graphs are a way to represent biological knowledge

Nodes can be...

genes, mRNAs, proteins, small RNAs, hormones, metabolites, species, populations, individuals, ...

Graphs are a way to represent biological knowledge

Nodes can be...

genes, mRNAs, proteins, small RNAs, hormones, metabolites, species, populations, individuals, ... Additional information can be attached to these nodes (GO term, protein family, functional motifs, cis-regulatory motifs, ...)

Graphs are a way to represent biological knowledge

Nodes can be...

genes, mRNAs, proteins, small RNAs, hormones, metabolites, species, populations, individuals, ... Additional information can be attached to these nodes (GO term, protein family, functional motifs, cis-regulatory motifs, ...)

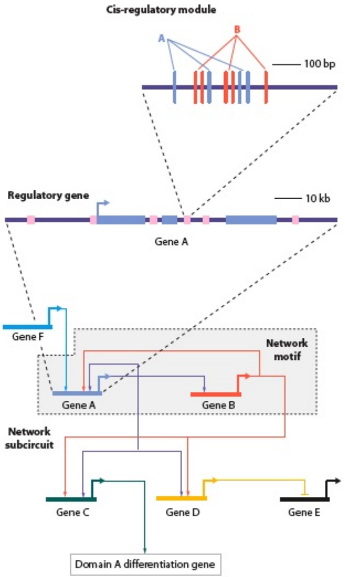
Relations can be...

- molecular regulation (transcriptional regulation, phosphorylation, acetylation, ...)
- molecular interaction (protein-protein, protein-siRNA, ...)
- enzymatic reactions
- genetic interactions (when gene A is mutated, gene B expression is up-regulated)
- co-localisation (genomic, sub-cellular, cellular, ...)
- co-occurence (when two entities are systematically found together)

Example of a molecular network with molecular regulation

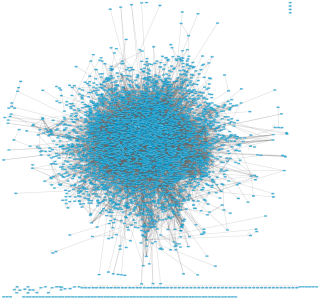
Nodes are **genes**
Relations are **transcriptional regulations**

[de Leon and Davidson, 2006]



Example of a molecular network with physical interactions

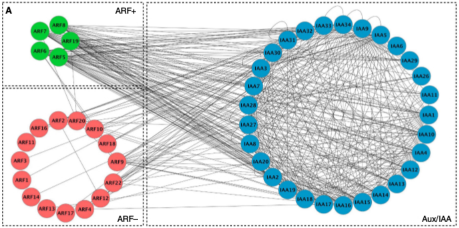
Nodes are **proteins**
Relations are **physical interactions (Y2H)**



made from data in

[*Arabidopsis* Interactome Mapping Consortium, 2011]

[Vernoux et al., 2011]



Example of a metabolic network

Nodes are **metabolites**
Relations are **enzymatic reactions**

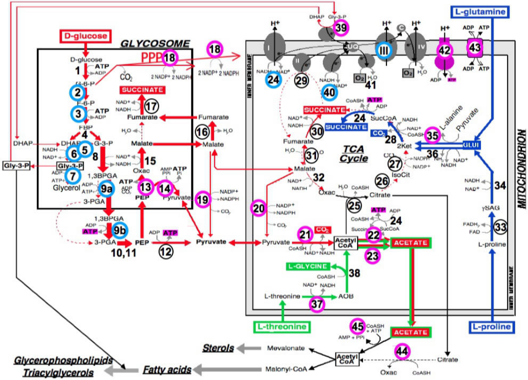
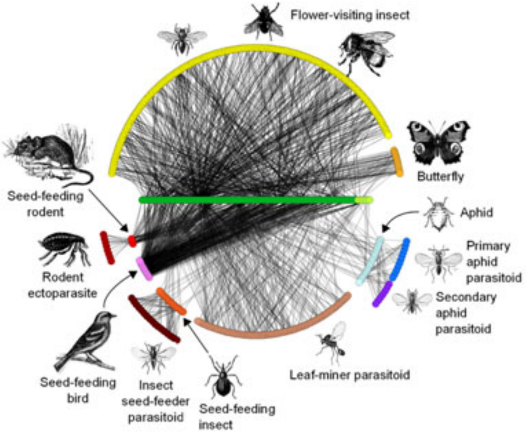


Image taken from Project "Trypanosome" (F. Bringaud - iMET team, RMSB, Bordeaux)

Example of an ecologic network

Nodes are **species**
Relations are **trophic links**

[The QUINTESSENCE Consortium, 2016]



Example of a molecular network with heterogeneous information

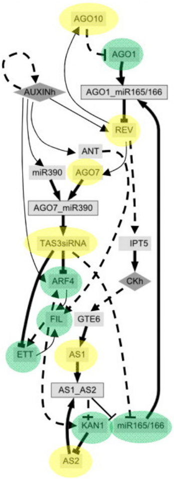
Nodes

- shapes represent the nature of the entities
- colors indicate tissue localisation

Edges are direct molecular relations of different types

- reliability: bold, dashed, normal lines
- inhibition or activation: T-line or arrow

[La Rota et al., 2011]

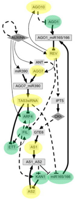


What is a model?

Model: simplified representation of reality

What is a model?

Model: simplified representation of reality

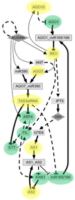


Biological model

simplified representation of a biological process

What is a model?

Model: simplified representation of reality

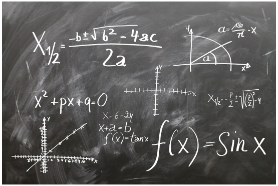


Biological model

simplified representation of a biological process

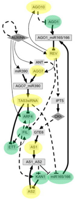
Mathematical model

- simplified description of a system using mathematical concepts
- in particular, **statistical models** represent the data-generating process



What is a model?

Model: simplified representation of reality

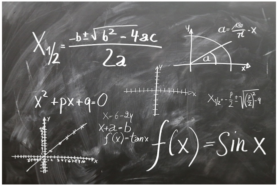


Biological model

simplified representation of a biological process

Mathematical model

- simplified description of a system using mathematical concepts
- in particular, **statistical models** represent the data-generating process



biological interaction model = biological network + mathematical model

Outline

1. What are networks/graphs?

2. What are networks useful for in biology?
    - Visualization
    - Simple analyses based on network topology
    - More advanced analyses based on network topology
    - Biological interaction models

3. How to build networks?

Outline

1. What are networks/graphs?

2. What are networks useful for in biology?
   - Visualization
   - Simple analyses based on network topology
   - More advanced analyses based on network topology
   - Biological interaction models

3. How to build networks?

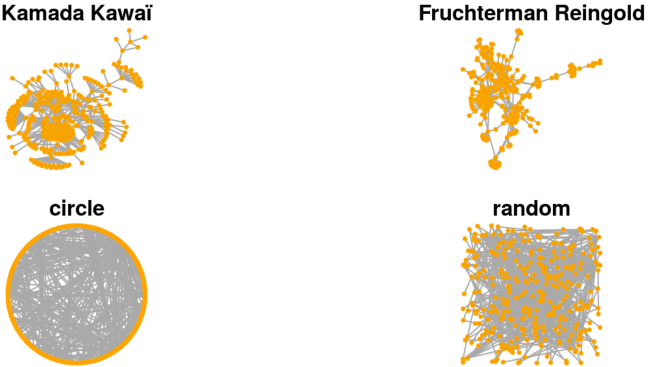
Advantages and drawbacks of network visualization

Visualization helps understand the network macro-structure and provides an **intuitive understanding** of the network.

Advantages and drawbacks of network visualization

Visualization helps understand the network macro-structure and provides an **intuitive understanding** of the network.

But all network visualizations are subjective and can mislead the person looking at it if not careful. **[Shen-Orr et al., 2002]** *Escherichia coli* transcriptional regulation network



How to represent networks?

Many different algorithms that often produce solutions that are not unique (integrate some randomness)

Most popular: **force directed placement algorithms**

- Fruchterman & Reingold [Fruchterman and Reingold, 1991]
- Kamada & Kawaï [Kamada and Kawai, 1989]

Such algorithms are computationally extensive and hard to use with large networks (more than a few thousands nodes)

Another useful layout

- attribute circle layout (quick but can be hard to read)

Network visualization software

(not only for biological networks)

- **NetworkX** (python library, not really interactive but produces javascript) https://networkx.github.io
- **igraph** (python and R libraries, not really interactive) http://igraph.org
- **Tulip** (interactive) http://tulip.labri.fr
- **Cytoscape** (interactive) http://cytoscape.org
- **Gephi** (interactive) gephi.org
- ...

Outline

1. What are networks/graphs?

2. What are networks useful for in biology?
   - Visualization
   - Simple analyses based on network topology
   - More advanced analyses based on network topology
   - Biological interaction models

3. How to build networks?

What is network topology?

Network topology

- study of the **network global and local structure**
- produces **numerical summaries** ⇒ biological interpretation

Credits: S.M.H. Oloomi, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=35247515 (network)

and AJC1, CC BY-NC-SA 2.0, https://www.flickr.com/photos/ajc1/4830932578 (biology)

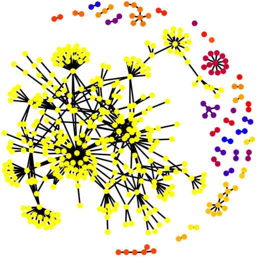
What is network topology?

Network topology

- study of the **network global and local structure**
- produces **numerical summaries** ⇒ biological interpretation

connected components are the connected subgraphs, *i.e.*, parts of the graph in which any node can be reached from any other node by a path

composantes connexes



34 connected components

[Shen-Orr et al., 2002] *Escherichia coli* transcriptional regulation network

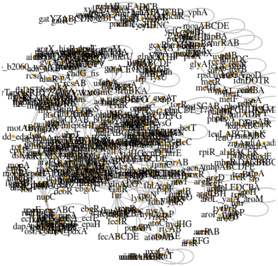
Global characteristics

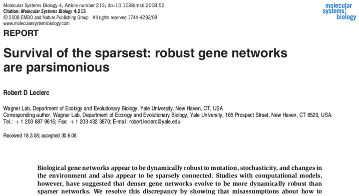
(mainly used for comparisons between networks or with random graphs having common characteristics with the real network)

Density *densité*

Number of edges divided by the number of pairs of nodes.

[Shen-Orr et al., 2002] *Escherichia coli* transcriptional regulation network: 423 nodes, 578 edges. Density: $\sim 0.64\%$




Molecular Systems Biology 4; Article number 213; doi:10.1038/msb.2008.52
Citation: ***Molecular Systems Biology*** **4:213**
© 2008 EMBO and Nature Publishing Group All rights reserved 1744-4292/08
www.molecularsystemsbiology.com

molecular systems biology

REPORT

Survival of the sparsest: robust gene networks are parsimonious

Robert D Leclerc

Wagner Lab, Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA
Corresponding author. Wagner Lab, Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 6520, USA. Tel.: +1 203 687 9615; Fax: +1 203 432 3870; E-mail: robert.leclerc@yale.edu

Received 18.3.08; accepted 30.6.08

Biological gene networks appear to be dynamically robust to mutation, stochasticity, and changes in the environment and also appear to be sparsely connected. Studies with computational models, however, have suggested that denser gene networks evolve to be more dynamically robust than sparser networks. We resolve this discrepancy by showing that misassumptions about how to

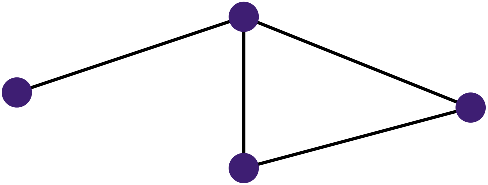
[Leclerc, 2008]: biological networks are generally sparsely connected (*S. cerevisiae*, *E. coli*, *D. melanogaster* transcriptional regulatory network densities < 0.1): evolutionary advantage for preserving robustness?

Global characteristics

(mainly used for comparisons between networks or with random graphs having common characteristics with the real network)

Transitivity *transitivité*

Number of triangles divided by the number of triplets connected by at least two edges.

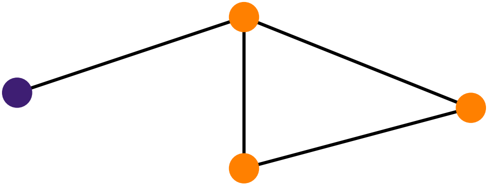


Global characteristics

(mainly used for comparisons between networks or with random graphs having common characteristics with the real network)

Transitivity *transitivité*

Number of triangles divided by the number of triplets connected by at least two edges.

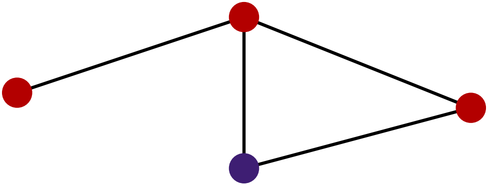


Global characteristics

(mainly used for comparisons between networks or with random graphs having common characteristics with the real network)

Transitivity *transitivité*

Number of triangles divided by the number of triplets connected by at least two edges.

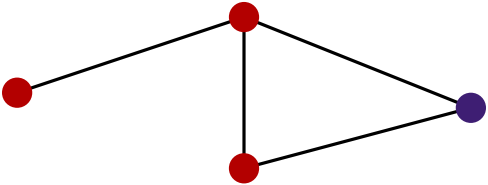


Global characteristics

(mainly used for comparisons between networks or with random graphs having common characteristics with the real network)

Transitivity *transitivité*

Number of triangles divided by the number of triplets connected by at least two edges.



Transitivity is equal to $1/3$. Density is equal to $\frac{4}{4\times3/2} = 2/3$

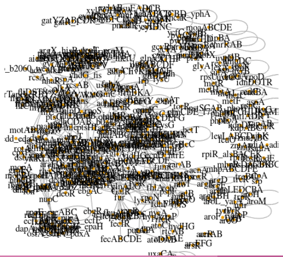
Global characteristics

(mainly used for comparisons between networks or with random graphs having common characteristics with the real network)

Transitivity *transitivité*

Number of triangles divided by the number of triplets connected by at least two edges.

[Shen-Orr et al., 2002] *Escherichia coli* transcriptional regulation network. Transitivity: $\sim 2.38\%$ $\gg$ density



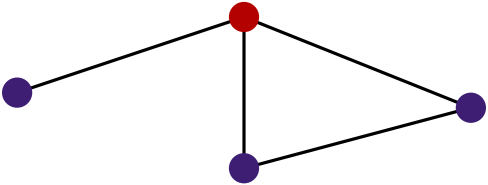
Comparaison with random graphs *(same number of nodes and edges, edges distributed at random between pairs of nodes)*: average transitivity is $\sim 0.63\%$.

$\Rightarrow$ strong local density in *Escherichia coli* transcriptional regulation network ("modularity" structure).

Key measures for other numerical characteristics

Node degree *degré*

number of edges adjacent to a given node or number of neighbors of the node



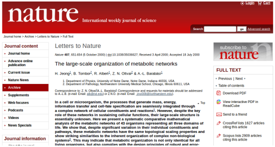
The degree of the red node is equal to 3.

Key measures for other numerical characteristics

Node degree *degré*

number of edges adjacent to a given node or number of neighbors of the node

[Jeong et al., 2000] shows that degree distribution in metabolomic networks is "scale-free"



nature International weekly journal of science

Login Cart

Search go Advanced search

Journal home > Archive > Letters to Nature > Full Text

Journal content

- Journal home
- Advance online publication
- Current issue
- Nature News
- Archive
- Supplements
- Web focuses
- Podcasts
- Videos
- News Specials

Journal information

Letters to Nature

Nature **407**, 651-654 (5 October 2000) | doi:10.1038/35036627; Received 3 April 2000; Accepted 18 July 2000

The large-scale organization of metabolic networks

H. Jeong[1], B. Tombor[2], R. Albert[1], Z. N. Oltvai[2] & A.-L. Barabási[1]

1. Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA
2. Department of Pathology, Northwestern University Medical School, Chicago, Illinois 60611, USA

Correspondence to: Z. N. Oltvai[2]A.-L. Barabási[1] Correspondence and requests for materials should be addressed to A.-L.B. (e-mail: Email: alb@nd.edu) or Z.N.O. (e-mail: Email: zno008@northwestern.edu).

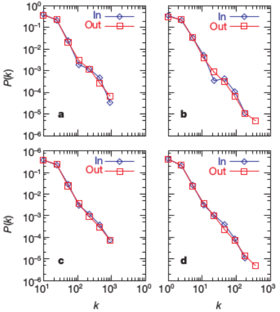
In a cell or microorganism, the processes that generate mass, energy, information transfer and cell-fate specification are seamlessly integrated through a complex network of cellular constituents and reactions[1]. However, despite the key role of these networks in sustaining cellular functions, their large-scale structure is essentially unknown. Here we present a systematic comparative mathematical analysis of the metabolic networks of 43 organisms representing all three domains of life. We show that, despite significant variation in their individual constituents and pathways, these metabolic networks have the same topological scaling properties and show striking similarities to the inherent organization of complex non-biological systems[2]. This may indicate that metabolic organization is not only identical for all living organisms, but also complies with the design principles of robust and error-

Top

subscribe to nature

FULL TEXT

- Previous | Next
- Table of contents
- Download PDF
- View interactive PDF in ReadCube
- Send to a friend
- CrossRef lists 1627 articles citing this article
- Scopus lists 2909 articles citing this article



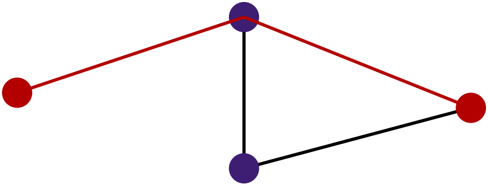
frequency of nodes having a degree of k $\sim k^{-\gamma}$ (highly skewed distributions)

Archaeoglobus fulgidus, *E. coli*, *Caenorhabditis elegans* and average over 43 organisms

Key measures for other numerical characteristics

Shortest path length (between two nodes)

minimal number of edges needed to reach a node from the other node through a path along the edges of the network

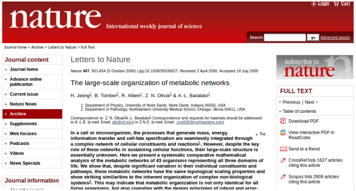


The shortest path length between red nodes is equal to 2.

Key measures for other numerical characteristics

Shortest path length (between two nodes)

minimal number of edges needed to reach a node from the other node through a path along the edges of the network



nature International weekly journal of science

Login Cart

Search go Advanced search

Journal home > Archive > Letters to Nature > Full Text

Journal content

- Journal home
- Advance online publication
- Current issue
- Nature News
- Archive
- Supplements
- Web focuses
- Podcasts
- Videos
- News Specials

Journal information

Letters to Nature

Nature **407**, 651-654 (5 October 2000) | doi:10.1038/35036627; Received 3 April 2000; Accepted 18 July 2000

The large-scale organization of metabolic networks

H. Jeong[1], B. Tombor[2], R. Albert[1], Z. N. Oltvai[2] & A.-L. Barabási[1]

1. Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA
2. Department of Pathology, Northwestern University Medical School, Chicago, Illinois 60611, USA

Correspondence to: Z. N. Oltvai[2]A.-L. Barabási[1] Correspondence and requests for materials should be addressed to A.-L.B. (e-mail: Email: alb@nd.edu) or Z.N.O. (e-mail: Email: zno008@northwestern.edu).

In a cell or microorganism, the processes that generate mass, energy, information transfer and cell-fate specification are seamlessly integrated through a complex network of cellular constituents and reactions[1]. However, despite the key role of these networks in sustaining cellular functions, their large-scale structure is essentially unknown. Here we present a systematic comparative mathematical analysis of the metabolic networks of 43 organisms representing all three domains of life. We show that, despite significant variation in their individual constituents and pathways, these metabolic networks have the same topological scaling properties and show striking similarities to the inherent organization of complex non-biological systems[2]. This may indicate that metabolic organization is not only identical for all living organisms, but also complies with the design principles of robust and error-

Top

subscribe to nature

FULL TEXT

- Previous | Next
- Table of contents

Download PDF

View interactive PDF in ReadCube

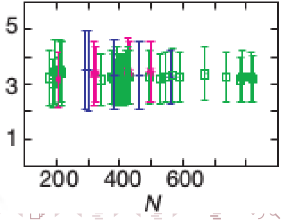
Send to a friend

CrossRef lists 1627 articles citing this article

Scopus lists 2909 articles citing this article

[Jeong et al., 2000] shows that shortest path length distribution is similar accross 43 species in metabolomic networks

observed average shortest path lengths is smaller than in random graph with uniform distribution of edges



Outline

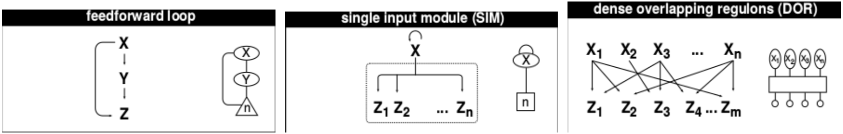
1. What are networks/graphs?

2. What are networks useful for in biology?
   - Visualization
   - Simple analyses based on network topology
   - More advanced analyses based on network topology
   - Biological interaction models

3. How to build networks?

Network motifs

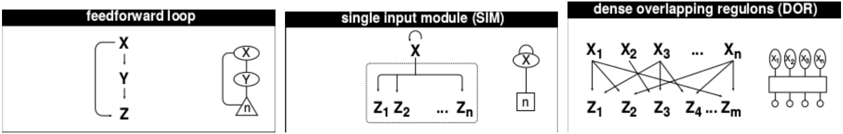
[Shen-Orr et al., 2002] showed that some specific motifs



are found significantly more often in *Escherichia coli* transcription network than in random networks with the same degree distribution.

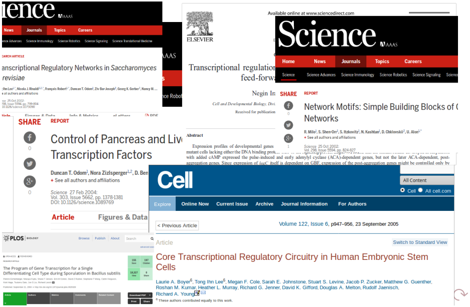
Network motifs

[Shen-Orr et al., 2002] showed that some **specific motifs**


feedforward loop
X
Y
Z
X
Y
n
single input module (SIM)
X
Z_1 Z_2 ... Z_n
X
n
dense overlapping regulons (DOR)
X_1 X_2 X_3 ... X_n
Z_1 Z_2 Z_3 Z_4 ... Z_m
X_1 X_2 X_3 X_n

are found significantly more often in *Escherichia coli* transcription network than in random networks with the same degree distribution.

[Milo et al., 2002,
Lee et al., 2002,
Eichenberger et al., 2004,
Odom et al., 2004,
Boyer et al., 2005,
Iranfar et al., 2006] show similar conclusion in various species (bacteria, yeast, higher organisms)


ience AAAS
News Journals Topics Careers
ence Advances Science Immunology Science Robotics Science Signaling Science Translational Medicine

EARCH ARTICLE

anscriptional Regulatory Networks in *Saccharomyces revisiae*

Ihn Lee, Nicola J. Rinaldi, François Robert, Duncan T. Odom, Ziv Bar-Joseph, Georg K. Gerber, Nancy M. ...

e all authors and affiliations

nce 25 Oct 2002:
98, Issue 5594, pp. 799-804
10.1126/science.1075090

Available online at www.sciencedirect.com
ELSEVIER

Transcriptional regulation
feed-forwa

Negin Ir

Cell and Developmental Biology, Divi

Received for publication

Abstract

Expression profiles of developmental genes
mutant cells lacking either the DNA binding prot...
with added cAMP expressed the pulse-induced and early adenylyl cyclase (ACA)-dependent genes, but not the later ACA-dependent, post-aggregation genes. Since expression of *lagC* itself is dependent on GBF, expression of the post-aggregation genes might be controlled only by

Science AAAS
Home News Journals Topics Careers
Science Science Advances Science Immunology Science Robotics Science Signaling Scienc
SHARE

REPORT

Network Motifs: Simple Building Blocks of C
Networks

R. Milo[1], S. Shen-Orr[1], S. Itzkovitz[1], N. Kashtan[1], D. Chklovskii[2], U. Alon[1,*]

+ See all authors and affiliations

Science 25 Oct 2002:
Vol. 298, Issue 5594, pp. 824-827

REPORT

Control of Pancreas and Liv
Transcription Factors

Duncan T. Odom[1], Nora Zizlsperger[1,2], D. Ben

+ See all authors and affiliations

Science 27 Feb 2004:
Vol. 303, Issue 5662, pp. 1378-1381
DOI: 10.1126/science.1089769

Volume 122, Issue 6, p947–956, 23 September 2005

Article

Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells

Laurie A. Boyer[6], Tong Ihn Lee[6], Megan F. Cole, Sarah E. Johnstone, Stuart S. Levine, Jacob P. Zucker, Matthew G. Guenther, Roshan M. Kumar, Heather L. Murray, Richard G. Jenner, David K. Gifford, Douglas A. Melton, Rudolf Jaenisch, Richard A. Young

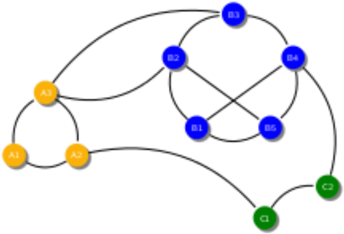
[6] These authors contributed equally to this work.

Switch to Standard View
PLOS BIOLOGY
Browse Publish About Search
155 Save 197 Citation 18,327 View 0 Share
Download PDF
Print Share

The Program of Gene Transcription for a Single Differentiating Cell Type during Sporulation in Bacillus subtilis

Node clustering *classification*

Cluster nodes into groups that are **densely connected** and share **few links** (comparatively) **with the other groups**. Clusters are often called **communities** *communautés* (social sciences) or **modules** *modules* (biology). [Fortunato, 2010]



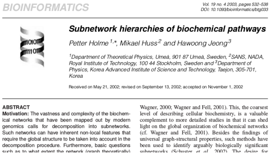
Node clustering *classification*

Cluster nodes into groups that are **densely connected** and share **few links** (comparatively) **with the other groups**. Clusters are often called **communities** *communautés* (social sciences) or **modules** *modules* (biology). **[Fortunato, 2010]**

Node clustering *classification*

Cluster nodes into groups that are **densely connected** and share **few links** (comparatively) **with the other groups**. Clusters are often called **communities** *communautés* (social sciences) or **modules** *modules* (biology). [Fortunato, 2010]

Simplification of a large complex network



BIOINFORMATICS

Vol. 19 no. 4 2003, pages 532–538
DOI: 10.1093/bioinformatics/btg033

Subnetwork hierarchies of biochemical pathways

Petter Holme [1,*], Mikael Huss [2] and Hawoong Jeong [3]

[1] Department of Theoretical Physics, Umeå, 901 87 Umeå, Sweden, [2] SANS, NADA, Royal Institute of Technology, 100 44 Stockholm, Sweden and [3] Department of Physics, Korea Advanced Institute of Science and Technology, Taejon, 305-701, Korea

Received on May 21, 2002; revised on September 13, 2002; accepted on November 1, 2002

ABSTRACT
Motivation: The vastness and complexity of the biochemical networks that have been mapped out by modern genomics calls for decomposition into subnetworks. Such networks can have inherent non-local features that require the global structure to be taken into account in the decomposition procedure. Furthermore, basic questions such as to what extent the network (graph theoretically)

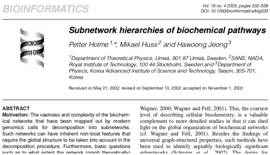
Wagner, 2000; Wagner and Fell, 2001). This, the coarsest level of describing cellular biochemistry, is a valuable complement to more detailed studies in that it can shed light on the global organization of biochemical networks (cf. Wagner and Fell, 2001). Besides the findings of universal graph-structural properties, such methods have been used to identify arguably biologically significant subnetworks (Schuster *et al.*, 2002). The desire for

[Holme et al., 2003] use clustering of metabolic networks to provide a simplified overview of the whole network and meaningful clusters

Node clustering *classification*

Cluster nodes into groups that are **densely connected** and share **few links** (comparatively) **with the other groups**. Clusters are often called **communities** *communautés* (social sciences) or **modules** *modules* (biology). [Fortunato, 2010]

Simplification of a large complex network



BIOINFORMATICS

Vol. 19 no. 4 2003, pages 532–538
DOI: 10.1093/bioinformatics/btg033

Subnetwork hierarchies of biochemical pathways

Petter Holme [1,], Mikael Huss [2] and Hawoong Jeong [3]*

[1] Department of Theoretical Physics, Umeå, 901 87 Umeå, Sweden, [2] SANS, NADA, Royal Institute of Technology, 100 44 Stockholm, Sweden and [3] Department of Physics, Korea Advanced Institute of Science and Technology, Taejon, 305-701, Korea

Received on May 21, 2002; revised on September 13, 2002; accepted on November 1, 2002

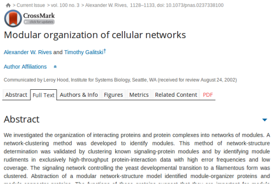
ABSTRACT

Motivation: The vastness and complexity of the biochemical networks that have been mapped out by modern genomics calls for decomposition into subnetworks. Such networks can have inherent non-local features that require the global structure to be taken into account in the decomposition procedure. Furthermore, basic questions such as to what extent the network (graph theoretically)

Wagner, 2000; Wagner and Fell, 2001). This, the coarsest level of describing cellular biochemistry, is a valuable complement to more detailed studies in that it can shed light on the global organization of biochemical networks (cf. Wagner and Fell, 2001). Besides the findings of universal graph-structural properties, such methods have been used to identify arguably biologically significant subnetworks (Schuster *et al.*, 2002). The desire for

[Holme et al., 2003] use clustering of metabolic networks to provide a simplified overview of the whole network and meaningful clusters

Identify key groups or key genes



> Current Issue > vol. 100 no. 3 > Alexander W. Rives, 1128–1133, doi: 10.1073/pnas.0237338100

CrossMark
click for updates

Modular organization of cellular networks

Alexander W. Rives and Timothy Galitski[†]

Author Affiliations

Communicated by Leroy Hood, Institute for Systems Biology, Seattle, WA (received for review August 24, 2002)

Abstract | Full Text | Authors & Info | Figures | Metrics | Related Content | PDF

Abstract

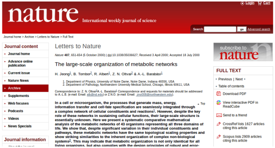
We investigated the organization of interacting proteins and protein complexes into networks of modules. A network-clustering method was developed to identify modules. This method of network-structure determination was validated by clustering known signaling-protein modules and by identifying module rudiments in exclusively high-throughput protein-interaction data with high error frequencies and low coverage. The signaling network controlling the yeast developmental transition to a filamentous form was clustered. Abstraction of a modular network-structure model identified module-organizer proteins and

[Rives and Galitski, 2003] use clustering in PPI network of yeast and found that proteins mostly interacting with members of their own cluster are often essential proteins.

Extracting important nodes

Hubs

Nodes with a high degree are called **hubs**: measure of the node popularity.



Letters to Nature

Nature **407**, 651-654 (5 October 2000) | doi:10.1038/35036627; Received 3 April 2000; Accepted 18 July 2000

The large-scale organization of metabolic networks

H. Jeong[1], B. Tombor[2], R. Albert[1], Z. N. Oltvai[2] & A.-L. Barabási[1]

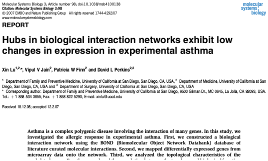
1. Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA
2. Department of Pathology, Northwestern University Medical School, Chicago, Illinois 60611, USA

Correspondence to: Z. N. Oltvai[2]A.-L. Barabási[1] Correspondence and requests for materials should be addressed to A.-L.B. (e-mail: Email: alb@nd.edu) or Z.N.O. (e-mail: Email: zno008@northwestern.edu).

In a cell or microorganism, the processes that generate mass, energy, information transfer and cell-fate specification are seamlessly integrated through a complex network of cellular constituents and reactions[1]. However, despite the key role of these networks in sustaining cellular functions, their large-scale structure is essentially unknown. Here we present a systematic comparative mathematical analysis of the metabolic networks of 43 organisms representing all three domains of life. We show that, despite significant variation in their individual constituents and pathways, these metabolic networks have the same topological scaling properties and show striking similarities to the inherent organization of complex non-biological systems[2]. This may indicate that metabolic organization is not only identical for all living organisms, but also complies with the design principles of robust and error-

[Jeong et al., 2000] show that the hubs are practically identical in metabolic networks among many species

[Lu et al., 2007] show that hubs have low changes in expression and have significantly different functions than peripherical nodes



Molecular Systems Biology 3; Article number 98; doi:10.1038/msb4100138
Citation: *Molecular Systems Biology* 3:98
© 2007 EMBO and Nature Publishing Group All rights reserved 1744-4292/07
www.molecularsystemsbiology.com

REPORT

Hubs in biological interaction networks exhibit low changes in expression in experimental asthma

Xin Lu[1,2,*], Vipul V Jain[2], Patricia W Finn[2] and David L Perkins[2,3]

[1] Department of Family and Preventive Medicine, University of California at San Diego, San Diego, CA, USA, [2] Department of Medicine, University of California at San Diego, San Diego, CA, USA and [3] Department of Surgery, University of California at San Diego, San Diego, CA, USA
[*] Corresponding author. Department of Family and Preventive Medicine, University of California at San Diego, 9500 Gilman Dr., MC 0645, La Jolla, CA 92093, USA. Tel.: +1 858 534 3855; Fax: +1 858 822 5290; E-mail: xinlu@ucsd.edu

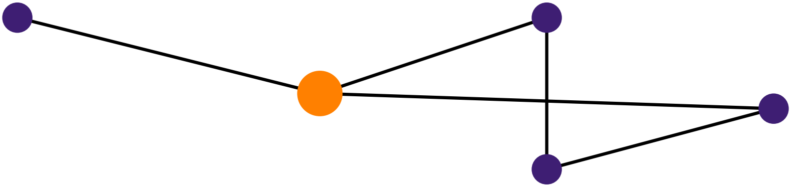
Received 18.12.06; accepted 12.2.07

Asthma is a complex polygenic disease involving the interaction of many genes. In this study, we investigated the allergic response in experimental asthma. First, we constructed a biological interaction network using the BOND (Biomolecular Object Network Databank) database of literature curated molecular interactions. Second, we mapped differentially expressed genes from microarray data onto the network. Third, we analyzed the topological characteristics of the

Extracting important nodes

Betweenness (of a node) *centralité*

number of shortest paths between all pairs of nodes that pass through the node. Betweenness is a centrality measure (nodes that are likely to disconnect the network if removed).

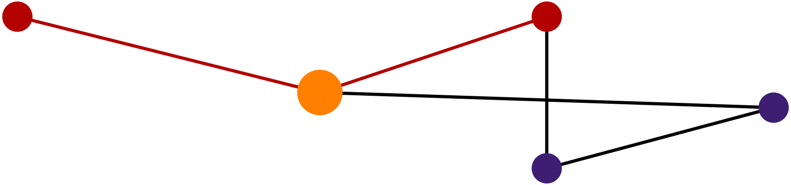


The orange node's degree is equal to 3, its betweenness to 4.

Extracting important nodes

Betweenness (of a node) *centralité*

number of shortest paths between all pairs of nodes that pass through the node. Betweenness is a centrality measure (nodes that are likely to disconnect the network if removed).

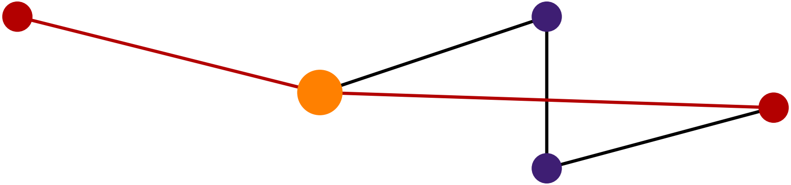


The orange node's degree is equal to 3, its betweenness to 4.

Extracting important nodes

Betweenness (of a node) *centralité*

number of shortest paths between all pairs of nodes that pass through the node. Betweenness is a centrality measure (nodes that are likely to disconnect the network if removed).

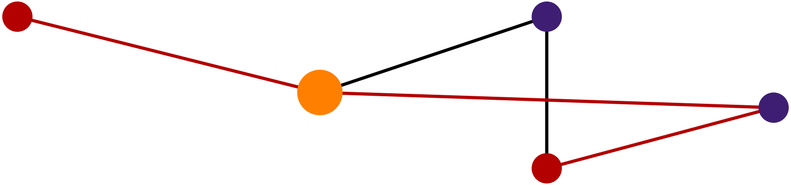


The orange node's degree is equal to 3, its betweenness to 4.

Extracting important nodes

Betweenness (of a node) *centralité*

number of shortest paths between all pairs of nodes that pass through the node. Betweenness is a centrality measure (nodes that are likely to disconnect the network if removed).

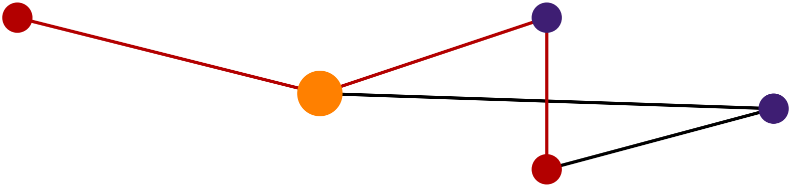


The orange node's degree is equal to 3, its betweenness to 4.

Extracting important nodes

Betweenness (of a node) *centralité*

number of shortest paths between all pairs of nodes that pass through the node. Betweenness is a centrality measure (nodes that are likely to disconnect the network if removed).

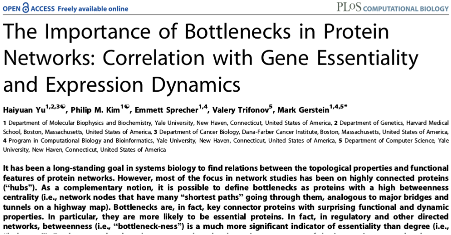


The orange node's degree is equal to 3, its betweenness to 4.

Extracting important nodes

Betweenness (of a node) *centralité*

number of shortest paths between all pairs of nodes that pass through the node. Betweenness is a centrality measure (nodes that are likely to disconnect the network if removed).



OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics

Haiyuan Yu[1,2,3], Philip M. Kim[1], Emmett Sprecher[1,4], Valery Trifonov[5], Mark Gerstein[1,4,5]*

1 Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, United States of America, **2** Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America, **3** Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America, **4** Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, United States of America, **5** Department of Computer Science, Yale University, New Haven, Connecticut, United States of America

It has been a long-standing goal in systems biology to find relations between the topological properties and functional features of protein networks. However, most of the focus in network studies has been on highly connected proteins ("hubs"). As a complementary notion, it is possible to define bottlenecks as proteins with a high betweenness centrality (i.e., network nodes that have many "shortest paths" going through them, analogous to major bridges and tunnels on a highway map). Bottlenecks are, in fact, key connector proteins with surprising functional and dynamic properties. In particular, they are more likely to be essential proteins. In fact, in regulatory and other directed networks, betweenness (i.e., "bottleneck-ness") is a much more significant indicator of essentiality than degree (i.e.,

[Yu et al., 2007] show that nodes with high betweenness in PPI networks are key connector proteins and are more likely to be essential proteins.

Outline

1 What are networks/graphs?

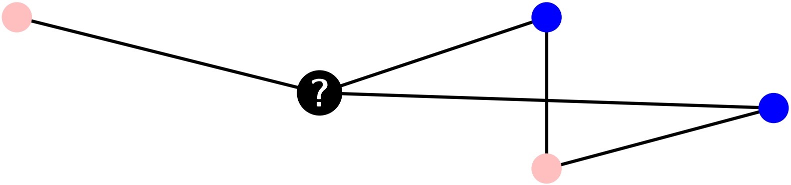
2 What are networks useful for in biology?

- Visualization
- Simple analyses based on network topology
- More advanced analyses based on network topology
- Biological interaction models

3 How to build networks?

Principle of status prediction based on a biological network

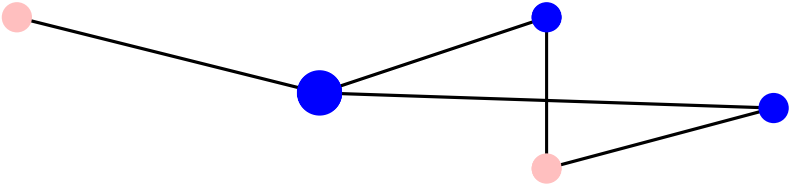
Available data: a network in which nodes are labeled by (incomplete) information (*e.g.*, GO term, disease status...)
Question: complete the information of nodes with unknown status



Principle of status prediction based on a biological network

Available data: a network in which nodes are labeled by (incomplete) information (*e.g.*, GO term, disease status...)
Question: complete the information of nodes with unknown status

Solution: Rule based on a majority vote among the neighbours. If the score is greater than a given threshold, then status is selected.
[Zaag, 2016]



Prediction model using a graph

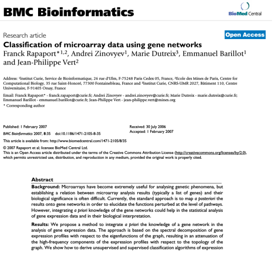
Available data: a set of gene expression profiles and a gene network (on the same genes)
Question: predict the status of a sample (*e.g.*, healthy / not healthy)

Prediction model using a graph

Available data: a set of gene expression profiles and a gene network (on the same genes)
Question: predict the status of a sample (*e.g.*, healthy / not healthy)

[Rapaport et al., 2007] using the network knowledge improves the results by producing solutions that have **similar contributions for genes connected by the network**

regression model with **network based penalization**



BMC Bioinformatics

BioMed Central

Research article

Open Access

Classification of microarray data using gene networks

Franck Rapaport*[1,2], Andrei Zinovyev[1], Marie Dutreix[3], Emmanuel Barillot[1] and Jean-Philippe Vert[2]

Address: [1]Institut Curie, Service de Bioinformatique, 26 rue d'Ulm, F-75248 Paris Cedex 05, France, [2]Ecole des Mines de Paris, Centre for Computational Biology, 35 rue Saint-Honoré, 77300 Fontainebleau, France and [3]Institut Curie, CNRS-UMR 2027, Bâtiment 110, Centre Universitaire, F-91405 Orsay, France

Email: Franck Rapaport* - franck.rapaport@curie.fr; Andrei Zinovyev - andrei.zinovyev@curie.fr; Marie Dutreix - marie.dutreix@curie.fr; Emmanuel Barillot - emmanuel.barillot@curie.fr; Jean-Philippe Vert - jean-philippe.vert@mines.org

* Corresponding author

Published: 1 February 2007
BMC Bioinformatics 2007, **8**:35 doi:10.1186/1471-2105-8-35
Received: 30 July 2006
Accepted: 1 February 2007
This article is available from: http://www.biomedcentral.com/1471-2105/8/35

Abstract

Background: Microarrays have become extremely useful for analysing genetic phenomena, but establishing a relation between microarray analysis results (typically a list of genes) and their biological significance is often difficult. Currently, the standard approach is to map *a posteriori* the results onto gene networks in order to elucidate the functions perturbed at the level of pathways. However, integrating *a priori* knowledge of the gene networks could help in the statistical analysis of gene expression data and in their biological interpretation.

Results: We propose a method to integrate *a priori* the knowledge of a gene network in the analysis of gene expression data. The approach is based on the spectral decomposition of gene expression profiles with respect to the eigenfunctions of the graph, resulting in an attenuation of the high-frequency components of the expression profiles with respect to the topology of the graph. We show how to derive unsupervised and supervised classification algorithms of expression

Differential expression using a graph

Available data: a set of gene expression obtained in two conditions and a gene network (on the same genes)
Question: find genes that are differentially expressed between the two conditions

Differential expression using a graph

Available data: a set of gene expression obtained in two conditions and a gene network (on the same genes)
Question: find genes that are differentially expressed between the two conditions

standard approach
independant tests and multiple test corrections

But: multiple test corrections are made for independant tests and genes are strongly correlated

Differential expression using a graph

Available data: a set of gene expression obtained in two conditions and a gene network (on the same genes)
Question: find genes that are differentially expressed between the two conditions

standard approach
independant tests and multiple test corrections

But: multiple test corrections are made for independant tests and genes are strongly correlated

using the network (T. Ha's Thesis "A multivariate learning penalized method for a joined inference of gene expression levels and gene regulatory networks")

a regression model for incorporating the information on gene dependency structure provided by the network into the differential analysis

Outline

1. What are networks/graphs?
2. What are networks useful for in biology?
   - Visualization
   - Simple analyses based on network topology
   - More advanced analyses based on network topology
   - Biological interaction models
3. How to build networks?

Standard methods for network inference

- bibliographic (expert based) inference (automatic language processing, ontology, text mining, ...) **[Huang and Lu, 2016]**
  Advantages: uses large expertise knowledge from biological databases

Standard methods for network inference

- bibliographic (expert based) inference (automatic language processing, ontology, text mining, ...) [Huang and Lu, 2016]
  Advantages: uses large expertise knowledge from biological databases

- statistical methods: from transcriptomic measures, infer network with
  - nodes: genes;
  - edges: dependency structure obtained from a statistical model (different meanings)

  Advantages: can handle interactions with **yet unknown genes** and deal with data collected in **specific conditions**

Standard methods for network inference

- bibliographic (expert based) inference (automatic language processing, ontology, text mining, ...) [Huang and Lu, 2016]
  Advantages: uses large expertise knowledge from biological databases

- statistical methods: from transcriptomic measures, infer network with
  - nodes: genes;
  - edges: dependency structure obtained from a statistical model (different meanings)

  Advantages: can handle interactions with **yet unknown genes** and deal with data collected in **specific conditions**
  Most widely used methods: relevance network, **Gaussian graphical models** (GGM), Bayesian models [Pearl, 1998, Pearl and Russel, 2002, Scutari, 2010] (R package bnlearn)

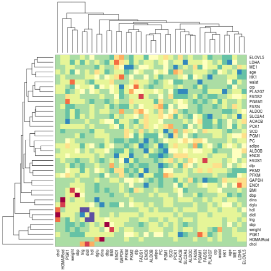
Correlation networks and GGM

Data: gene expression data

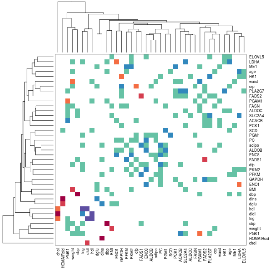
$$\begin{array}{c}\text{individuals}\\ n \simeq 30/50\end{array} \qquad \underbrace{\left\{ X = \begin{pmatrix} . & . & . & . & . & . \\ . & . & X_i^j & . & . & . \\ . & . & . & . & . & . \end{pmatrix} \right.}_{\text{variables (selected gene expressions), } p}$$

Using *correlations*: relevance network [Butte and Kohane, 1999, Butte and Kohane, 2000]

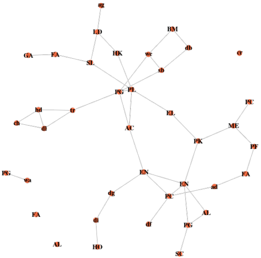
First (naive) approach: calculate correlations between expressions for all pairs of genes, threshold the smallest ones and build the network.



"Correlations"



Thresholding



Graph

Correlation, n and p

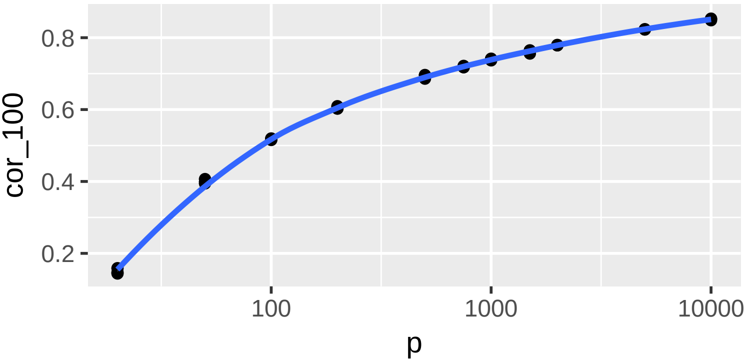
22 novembre 2017

100th largest correlation (Code)

- *n* experiences with an increasing number of genes *p*.
- all genes are simulated independant.
- we compute all correlations between genes
- we report the 100th largest absolute correlation
- we repeat a 100 times

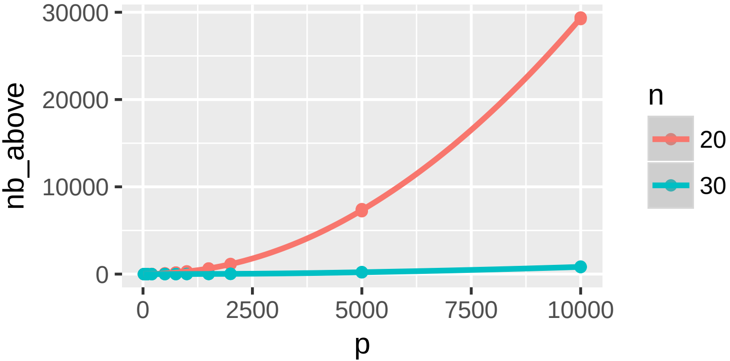
```r
onesimu <- function(p, n=20, rk=100){
  data <- matrix(rnorm(n*p), ncol=p)
  mat_cor <- abs(cor(data))
  return( c( p, n,
             sort(mat_cor[upper.tri(mat_cor)],
                  decreasing = TRUE)[rk]
           )
        )
}
```

100th largest correlation



- Statistically, this is not really surprising, but . . .
- To keep in mind when analysing large correlation matrices.
- “what does it means to have a correlation larger than . . . ?”

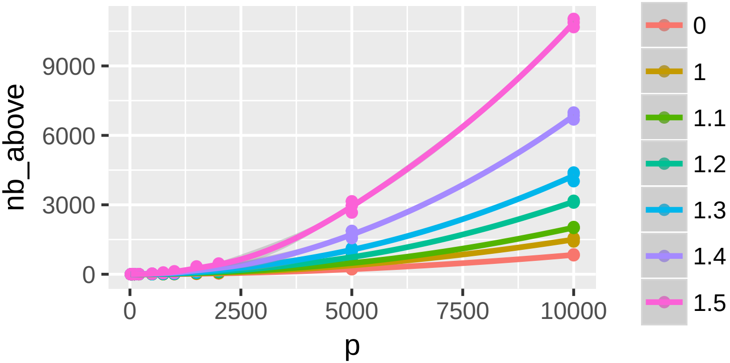
Number of correlations above a 0.7 threshold



- Statistically, this is not really surprising, but...
- To keep in mind when analysing large correlation matrices.
- “what does it means to have so many pairs of genes with a correlation larger than ... ?”

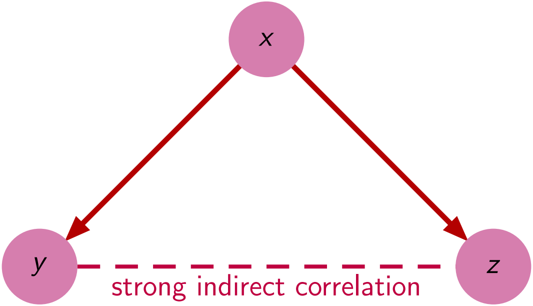
Number of correlations above a 0.7 threshold (n=30)

- 2 biological conditions 15 vs. 15
- 20% of the genes with a mean difference of 1...



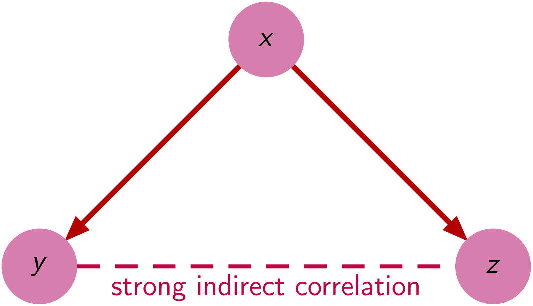
- With a t.test and Bonferonni correction (n=10000) a gene with a 1.5 difference is called DE only 10% of the time
- “what does it means to have a correlation higher than 0.7 ?”

Correlation is not partial correlation...



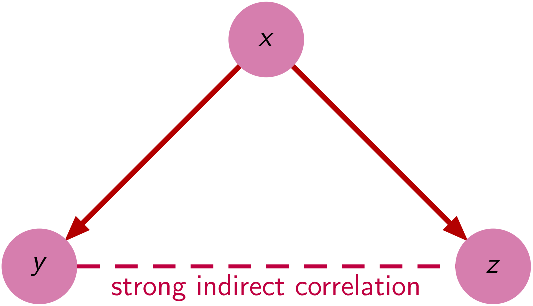
```
set.seed(2807); x <- runif(100)
y <- 2*x+1+rnorm(100,0,0.1); cor(x,y); [1] 0.9988261
z <- 2*x+1+rnorm(100,0,0.1); cor(x,z); [1] 0.998751
cor(y,z); [1] 0.9971105
```

Correlation is not partial correlation...



```
set.seed(2807); x <- runif(100)
y <- 2*x+1+rnorm(100,0,0.1); cor(x,y); [1] 0.9988261
z <- 2*x+1+rnorm(100,0,0.1); cor(x,z); [1] 0.998751
cor(y,z); [1] 0.9971105
♯ Partial correlation
cor(lm(y∼x)$residuals,lm(z∼x)$residuals) [1] -0.1933699
```

Correlation is not partial correlation...



Networks are built using **partial correlations**, i.e., correlations between gene expressions **knowing the expression of all the other genes** (residual correlations).

GGM

Assumptions: $(X_i)_{i=1,\dots,n}$ are i.i.d. Gaussian random variables $\mathcal{N}(0,\Sigma)$ (gene expression)

GGM

Assumptions: $(X_i)_{i=1,\ldots,n}$ are i.i.d. Gaussian random variables $\mathcal{N}(0,\Sigma)$ (gene expression)

GGM definition

- **Partial correlation formulation**

$$j \longleftrightarrow j' \text{(genes } j \text{ and } j' \text{ are linked)} \Leftrightarrow \mathbb{C}\mathrm{or}\left(X^j, X^{j'} | (X^k)_{k \neq j,j'}\right) \neq 0$$

GGM

Assumptions: $(X_i)_{i=1,\ldots,n}$ are i.i.d. Gaussian random variables $\mathcal{N}(0,\Sigma)$ (gene expression)

GGM definition

- **Partial correlation formulation**

$$j \longleftrightarrow j' (\text{genes } j \text{ and } j' \text{ are linked}) \Leftrightarrow \mathbb{C}\text{or}\left(X^j, X^{j'} | (X^k)_{k\neq j,j'}\right) \neq 0$$

- **Regression formulation**

$$X^j = \sum_{j'\neq j} \beta_{jj'} X^{j'} + \epsilon \qquad \beta_{jj'} \neq 0 \Leftrightarrow j \longleftrightarrow j' (\text{genes } j \text{ and } j' \text{ are linked})$$

Mathematical background

Theoretically: If $X \sim \mathcal{N}(0, \Sigma)$ then for $S = \Sigma^{-1}$

- partial correlation formulation

$$\mathbb{C}\mathrm{or}\left(X^j, X^{j'} | (X^k)_{k \neq j, j'}\right) = -\frac{S_{jj'}}{\sqrt{S_{jj} S_{j'j'}}}$$

- regression formulation

$$\beta_{jj'} = -\frac{S_{jj'}}{S_{jj}}$$

Mathematical background

Theoretically: If $X \sim \mathcal{N}(0, \Sigma)$ then for $S = \Sigma^{-1}$

- partial correlation formulation

$$\mathbb{C}\mathrm{or}\left(X^j, X^{j'} | (X^k)_{k \neq j, j'}\right) = -\frac{S_{jj'}}{\sqrt{S_{jj} S_{j'j'}}}$$

- regression formulation

$$\beta_{jj'} = -\frac{S_{jj'}}{S_{jj}}$$

In practice:

- Since p (number of genes) is often large compared to n (number of samples), **S is hard to estimate**.
- After the estimation, entries of S are not null ⇒ How to **select the "largest" entries in S**?

Some solutions

1. Seminal work [Schäfer and Strimmer, 2005a, Schäfer and Strimmer, 2005b] (implemented in the R package GeneNet)
   - Estimation of S: regularization for inversion of Σ
   - Edge selection: Bayesian approach

Some solutions

1. Seminal work
[Schäfer and Strimmer, 2005a, Schäfer and Strimmer, 2005b]
(implemented in the R package GeneNet)
   - Estimation of S: regularization for inversion of Σ
   - Edge selection: Bayesian approach

2. Sparse approach
[Friedman et al., 2008, Meinshausen and Bühlmann, 2006]
(implemented in the R package huge)
   - estimation and selection performed together
   - uses the regression framework in which a "sparse" penalty is added (LASSO or Graphical LASSO)

Important notices

- **ultra-high dimensionality**: if p is the number of genes, n the number of samples and k the (true) number of edges of a network, ultra-high dimensionality means that $k\left[1+\log\left(\frac{p(p-1)/2}{k}\right)\right]$ is "large" compared to n

Important notices

- **ultra-high dimensionality**: if p is the number of genes, n the number of samples and k the (true) number of edges of a network, ultra-high dimensionality means that $k\left[1+\log\left(\frac{p(p-1)/2}{k}\right)\right]$ is "large" compared to n

  In this case, there is **no hope to estimate the network** **[Verzelen, 2012]**.

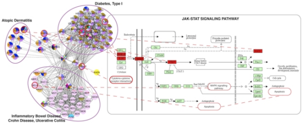
Important notices

- **ultra-high dimensionality**: if p is the number of genes, n the number of samples and k the (true) number of edges of a network, ultra-high dimensionality means that $k\left[1+\log\left(\frac{p(p-1)/2}{k}\right)\right]$ is "large" compared to n
  In this case, there is **no hope to estimate the network** [Verzelen, 2012].

- **applicability**: Gaussian models are well designed for microarray datasets. However, **extension to RNA-seq data is non trivial** and still under development.

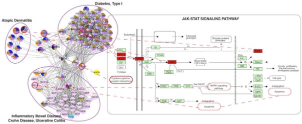
Take home message...

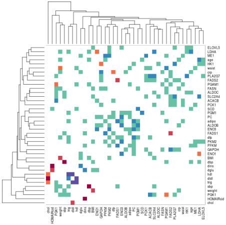
networks are useful to model complex systems


Diabetes, Type I
Atopic Dermatitis
JAK-STAT SIGNALING PATHWAY
Inflammatory Bowel Disease
Crohn Disease, Ulcerative Colitis

Take home message...

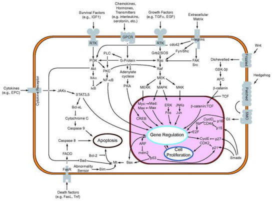
networks are useful to model complex systems


Diabetes, Type I
Atopic Dermatitis
JAK-STAT SIGNALING PATHWAY
Inflammatory Bowel Disease
Crohn Disease, Ulcerative Colitis

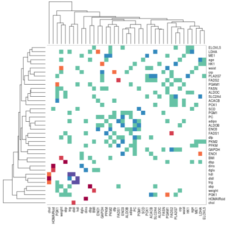


networks can be built with various approaches that define what they can be used for

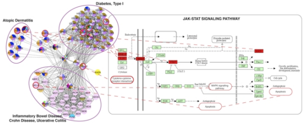
Take home message...



networks are useful information that can be integrated in biological models to improve knowledge



networks are useful to model complex systems


Diabetes, Type I
Atopic Dermatitis
JAK-STAT SIGNALING PATHWAY
Inflammatory Bowel Disease
Crohn Disease, Ulcerative Colitis
Apoptosis
MAPK signaling pathway

networks can be built with various approaches that define what they can be used for

References

Boyer, L., Lee, T., Cole, M., Johnstone, S., Levine, S., Zucker, J., Guenther, M., Kumar, R., Murray, H., Jenner, R., Gifford, D., Melton, D., Jaenisch, R., and Young, R. (2005).
Core transcriptional regulatory circuitry in human embryonic stem cells.
Cell, 122(6):947–956.

Butte, A. and Kohane, I. (1999).
Unsupervised knowledge discovery in medical databases using relevance networks.
In *Proceedings of the AMIA Symposium*, pages 711–715.

Butte, A. and Kohane, I. (2000).
Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements.
In *Proceedings of the Pacific Symposium on Biocomputing*, pages 418–429.

de Leon, S. and Davidson, E. (2006).
Deciphering the underlying mechanism of specification and differentiation: the sea urchin gene regulatory network.
Science's STKE, 361:pe47.

Eichenberger, P., Fujita, M., Jensen, S., Conlon, E., Rudner, D., Wang, S., Ferguson, C., Haga, K., Sato, T., Liu, J., and Losick, R. (2004).
The program of gene transcription for a single differentiating cell type during sporulation in bacillus subtilis.
PLoS Biology, 2(30):e328.

Fortunato, S. (2010).
Community detection in graphs.
Physics Reports, 486:75–174.

Friedman, J., Hastie, T., and Tibshirani, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics, 9(3):432–441.

Fruchterman, T. and Reingold, B. (1991).

Graph drawing by force-directed placement.
Software, Practice and Experience, 21:1129–1164.

Holme, P., Huss, M., and Jeong, H. (2003).
Subnetwork hierarchies of biochemical pathways.
Bioinformatics, 19(4):532–538.

Huang, C. and Lu, Z. (2016).
Community challenges in biomedical text mining over 10 years: success, failure and the future.
Briefings in Bioinformatics, 17(1):132–144.

Iranfar, N., Fuller, D., and Loomis, W. (2006).
Transcriptional regulation of post-aggregation genes in *dictyostelium* by a feed-forward loop involving GBF and LagC.
Developmental Biology, 290(9):460–469.

Arabidopsis Interactome Mapping Consortium (2011).
Evidence for network evolution in an *arabidopsis* interactome map.
Science, 333(6042):601–607.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z., and Barabási, A. (2000).
The large scale organization of metabolic networks.
Nature, 407:651–654.

Kamada, T. and Kawai, S. (1989).
An algorithm for drawing general undirected graphs.
Information Processing Letters, 31(1):7–15.

La Rota, C., Chopard, J., Das, P., Paindavoine, S., Rozier, F., Farcot, E., Godin, C., Traas, J., and Monéger, F. (2011).
A data-driven integrative model of sepal primordium polarity in *arabidopsis*.
The Plant Cell, 23(12):4318–4333.

Leclerc, R. (2008).

Survival of the sparsest: robust gene networks are parsimonious.
Molecular Systems Biology, 4:213.

Lee, T., Rinaldi, N., Robert, F., Odom, D., Bar-Joseph, Z., Gerber, G., Hannett, N., Harbison, C., Thompson, C., Simon, I., Zeitlinger, J., Jennings, E., Murray, H., Gordon, D., Ren, B., Wyrick, J., Tagne, J., Volkert, T., Fraenkel, E., Gifford, D., and Young, R. (2002).
Transcriptional regulatory networks in *saccharomyces cerevisiae*.
Science.

Lu, X., Jain, V., Finn, P., and Perkins, D. (2007).
Hubs in biological interaction networks exhibit low changes in expression in experimental asthma.
Molecular Systems Biology, 3:98.

Meinshausen, N. and Bühlmann, P. (2006).
High dimensional graphs and variable selection with the Lasso.
Annals of Statistic, 34(3):1436–1462.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002).
Network motifs: simple building blocks of complex networks.
Science, 298(5594):824–827.

Odom, D., Zizlsperger, N., Gordon, D., Bell, G., Rinaldi, N., Murray, H., Volkert, T., Schreiber, J., Rolfe, P., Gifford, D., Fraenkel, E., Bell, G., and Young, R. (2004).
Control of pancreas and liver gene expression by HNF transcription factors.
Science, 303(5662):1378–1381.

Pearl, J. (1998).
Probabilistic reasoning in intelligent systems: networks of plausible inference.
Morgan Kaufmann, San Francisco, California, USA.

Pearl, J. and Russel, S. (2002).
Bayesian Networks.
Bradford Books (MIT Press), Cambridge, Massachussets, USA.

Rapaport, F., Zinovyev, A., Dutreix, M., Barillot, E., and Vert, J. (2007).
Classification of microarray data using gene networks.
BMC Bioinformatics, 8:35.

Rives, A. and Galitski, T. (2003).
Modular organization of cellular networks.
Proceedings of the National Academy of Sciences, 100(3):1128–1133.

Schäfer, J. and Strimmer, K. (2005a).
An empirical bayes approach to inferring large-scale gene association networks.
Bioinformatics, 21(6):754–764.

Schäfer, J. and Strimmer, K. (2005b).
A shrinkage approach to large-scale covariance matrix estimation and implication for functional genomics.
Statistical Applications in Genetics and Molecular Biology, 4:1–32.

Scutari, M. (2010).
Learning Bayesian networks with the bnlearn R package.
Journal of Statistical Software, 35(3):1–22.

Shen-Orr, S., Milo, R., Mangan, S., and Alon, U. (2002).
Network motifs in the transcriptional regulation network of *escherichia coli*.
Nature Genetics, 31:64–68.

The QUINTESSENCE Consortium (2016).
Networking our way to better ecosystem service provision.
Trends in Ecology & Evolution, 31(2):105–115.

Vernoux, T., Brunoud, G., Farcot, EtienneE.and Morin, V., Van den Daele, H., Legrand, J., Oliva, M., Das, P., Larrieu, A., Wells, D., Guédon, Y., Armitage, L., Picard, F., Guyomarc'h, S., Cellier, C., Parry, G., Koumproglou, R., Doonan, J., Estelle, M., Godin, C., Kepinski, S., Bennett, M., De Veylder, L., and Traas, J. (2011).
The auxin signalling network translates dynamic input into robust patterning at the shoot apex.
Molecular Systems Biology, 7:508.

Verzelen, N. (2012).
Minimax risks for sparse regressions: ultra-high-dimensional phenomenons.
Electronic Journal of Statistics, 6:38–90.

Yu, H., Kim, P., Sprecher, E., Trifonov, V., and Gerstein, M. (2007).
The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics.
PLoS Computational Biology, 3(4):e59.

Zaag, R. (2016).
Enrichissement de profils transcriptomiques par intégration de données hétérogènes : annotation fonctionnelle de gènes d'Arabidopsis thaliana impliqués dans la réponse aux stress.
Thèse de doctorat, Université Paris Saclay, Saint-Aubin, France.