Product Graphs and Large Networks

Wilfried Imrich
emer.o.Univ.Prof.Dr.phil.
Montanuniversität Leoben, Austria

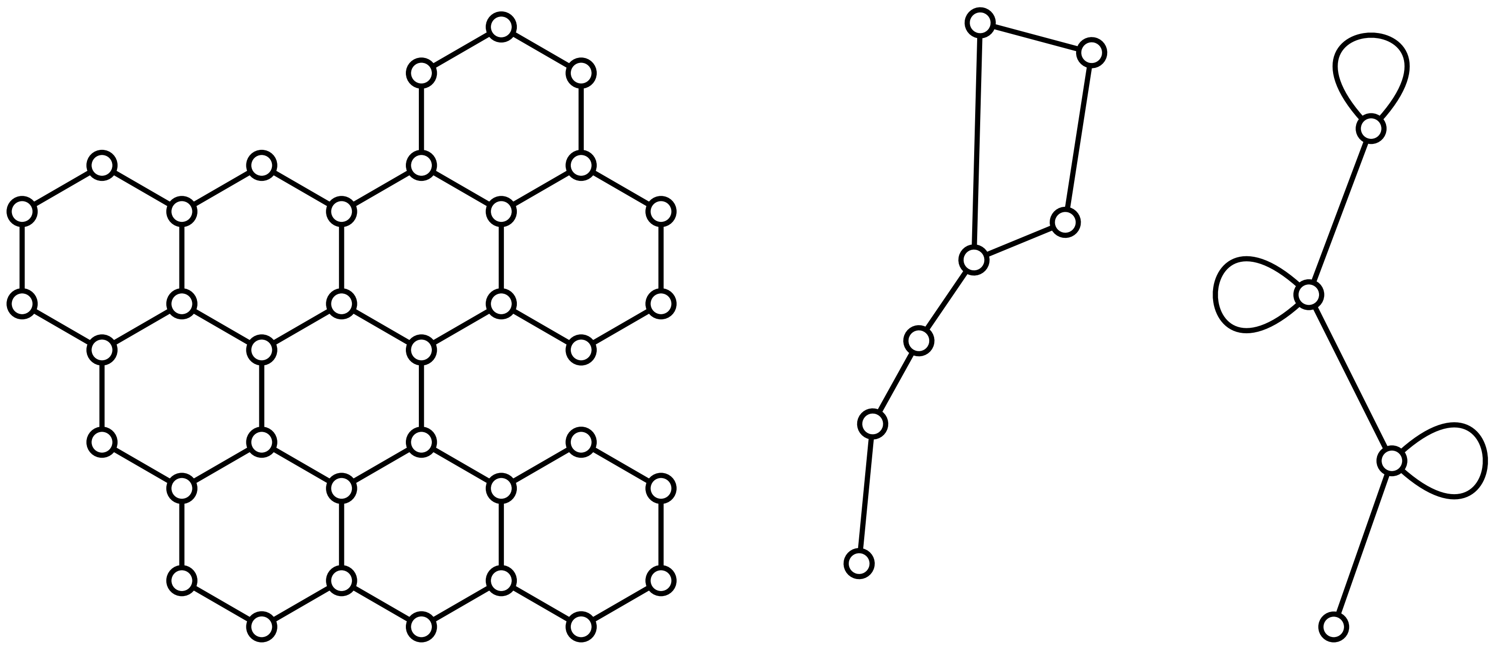
16th TBI Winter Seminar
5th Annual Meeting of the Bompfünewerer Consortium
Bled, Slovenia

February 15, 2011

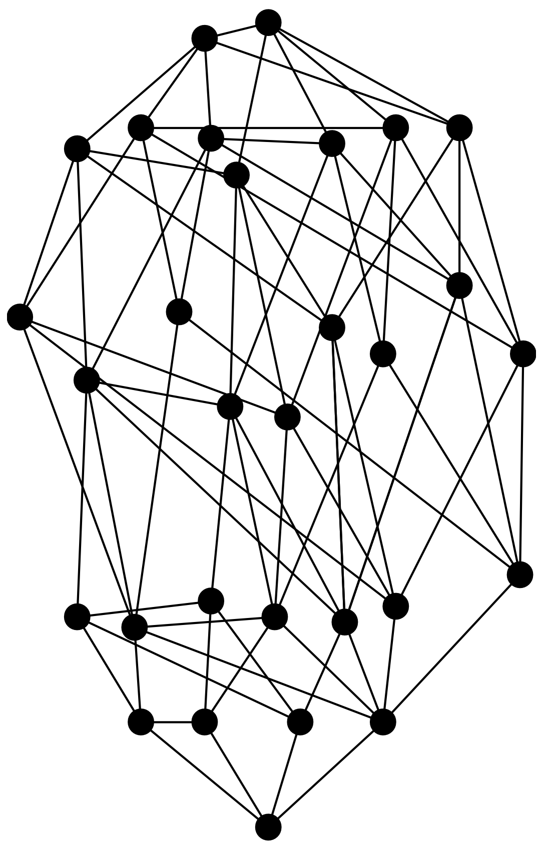
Contents

Graphs

Graphs G, vertex set $V(G)$, edge set $E(G)$



Simple graphs (left, center) and a graph with loops (right)



Degree k_i of vertex i : number of edges incident to vertex i.

Distance $d_{i,j}$ between two vertices: length of shortest path between them.

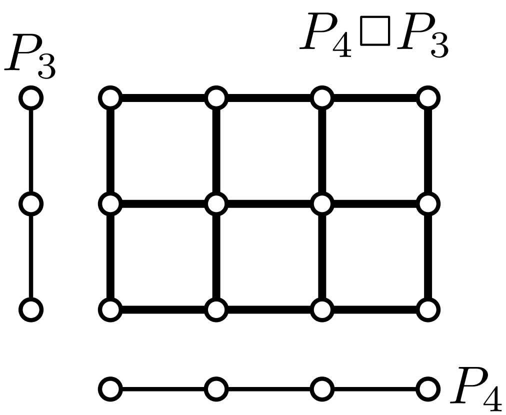
Diameter $diam(G)$ of G: maximal distance between two vertices.

Betweenness b_i of a vertex: total number of shortest paths among pairs of vertices that pass through i.

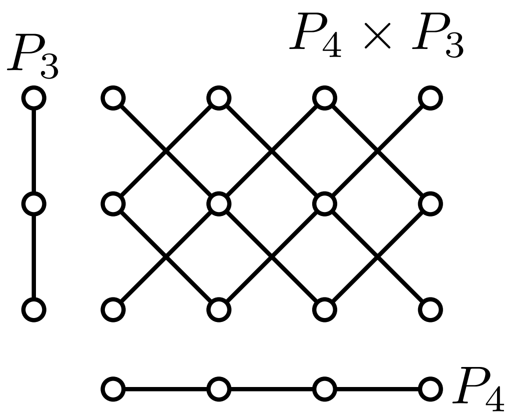
Clustering coefficient c_i: number e_i of edges between nearest neighbors of i divided by $k_i(k_i - 1)/2$.

Cartesian, direct and strong product

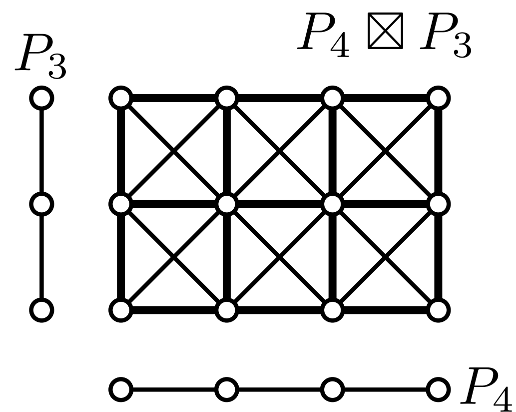
$$
\begin{aligned}
V(G \,\square\, H) &= V(G \times H) = \{(g,h) \mid g \in V(G) \text{ and } h \in V(H)\}, \\
E(G \,\square\, H) &= \{(g,h)(g',h') \mid g = g',\, hh' \in E(H), \text{ or } \; gg' \in E(G),\, h = h'\} \\
E(G \times H) &= \{(g,h)(g',h') \mid gg' \in E(G) \text{ and } hh' \in E(H)\}.
\end{aligned}
$$



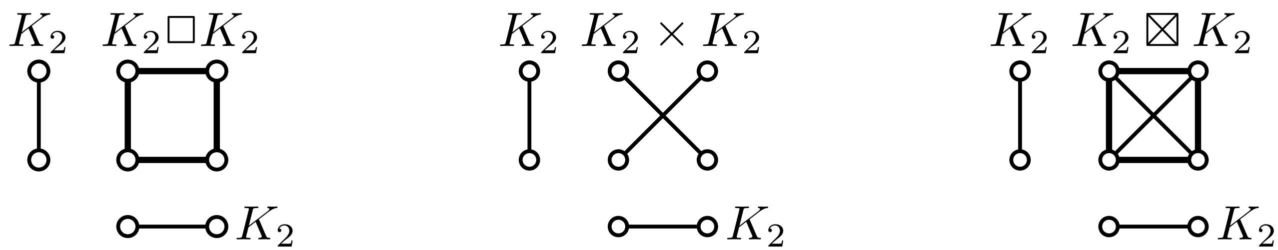
Cartesian product



Direct product

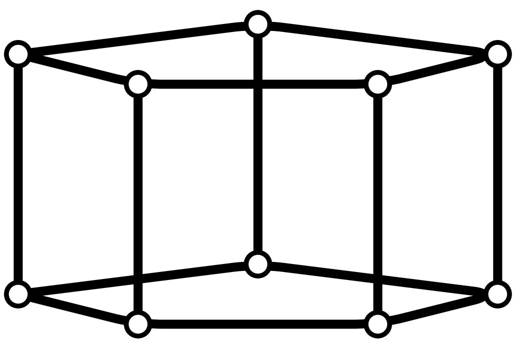


Strong product



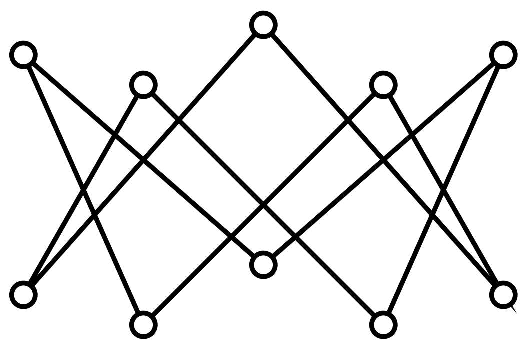
Origin of the notation

$C_5 \square K_2$



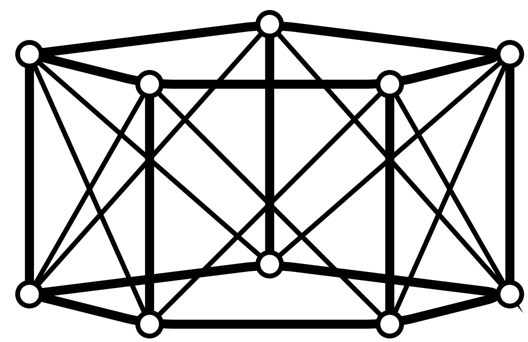
Cartesian product

$C_5 \times K_2$

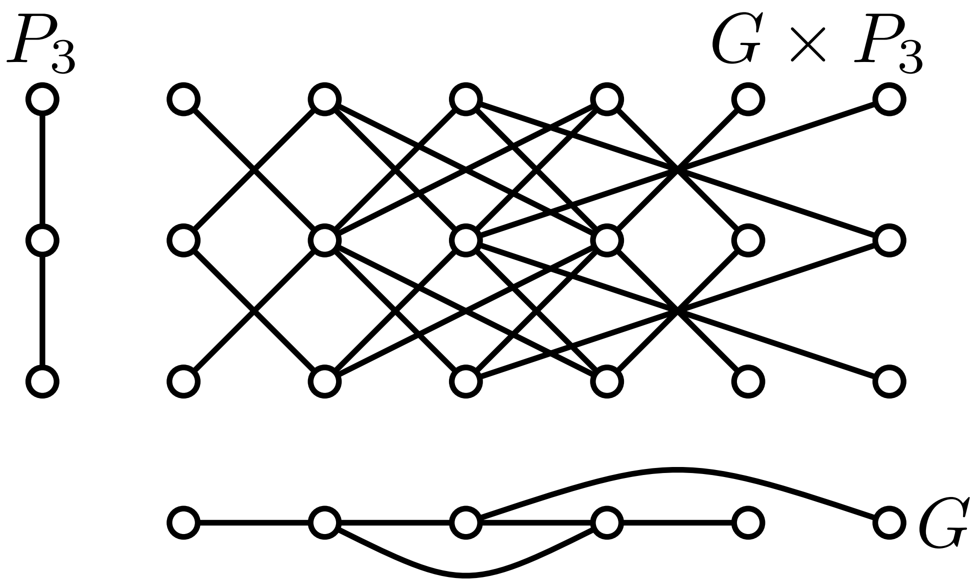


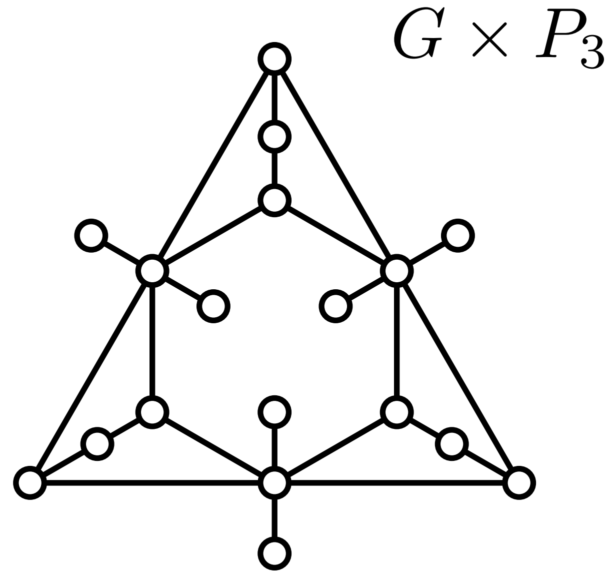
Direct Product

$C_5 \boxtimes K_2$

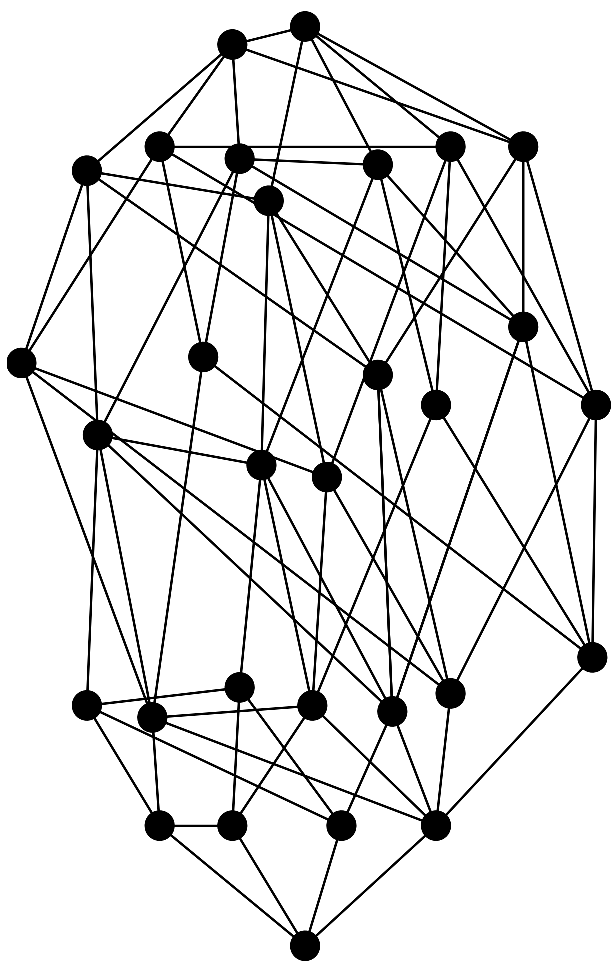


Strong product





A direct product (left) and a more symmetric representation (right)



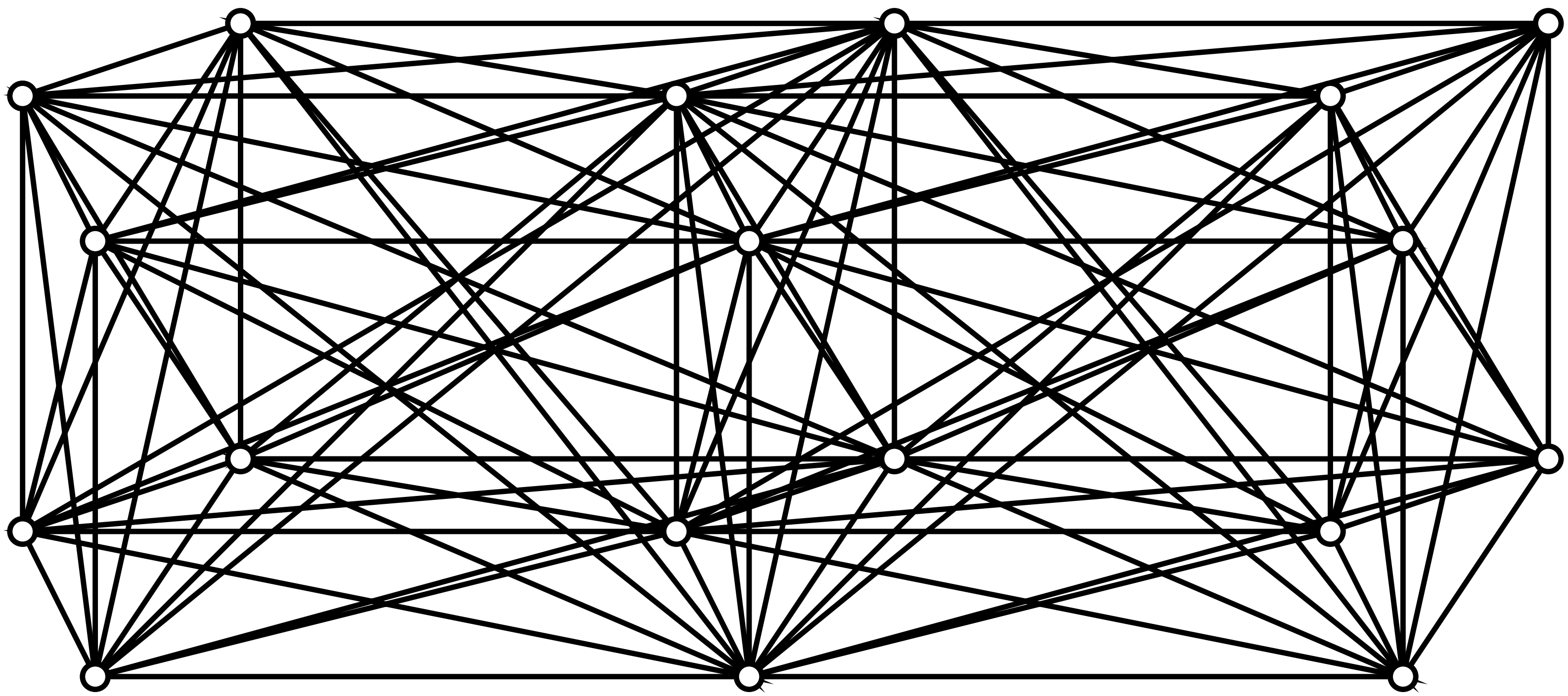
Recognition algorithms

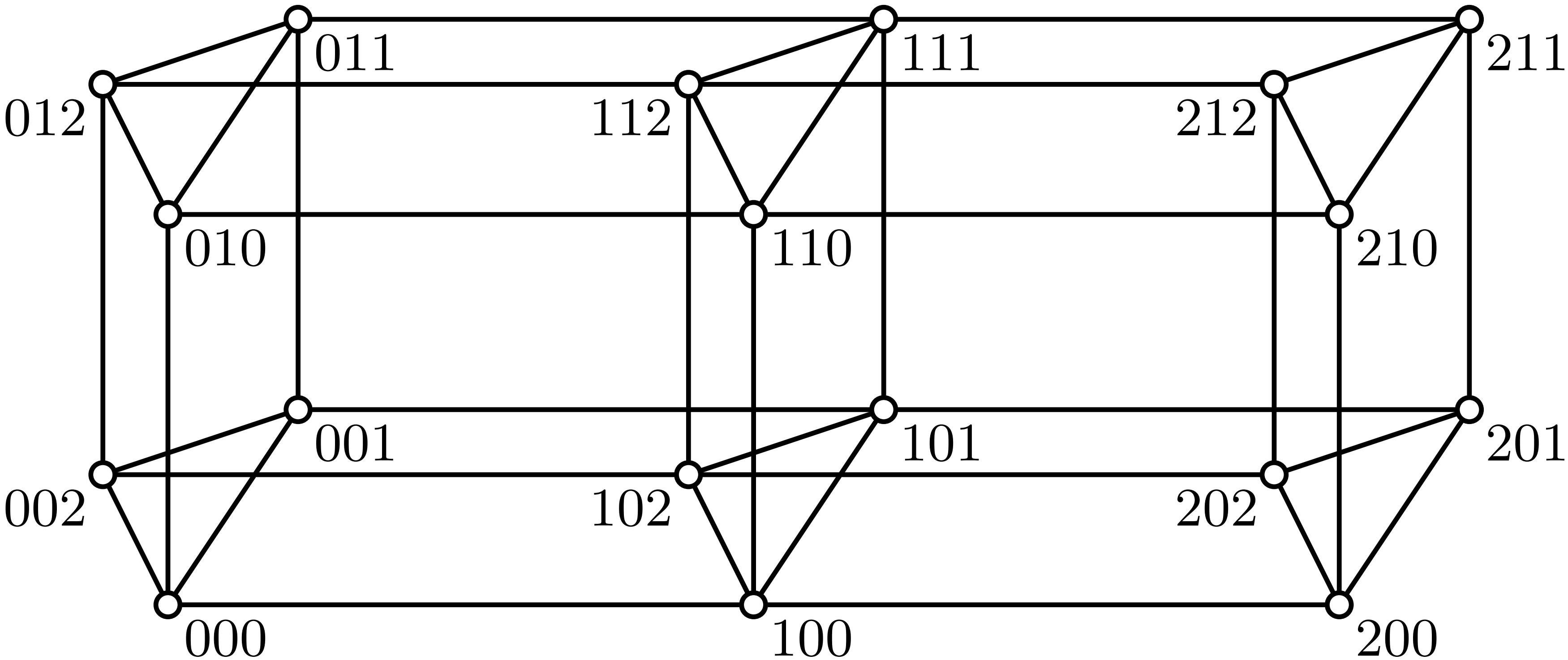
How does one recognize a graph as a product?

Cartesian product $O(m)$ time.

Strong product $O(\min(m^2, m\, a(G)\Delta))$ time.

Direct Product $O(\min(mn^2, m\Delta^3)$ time.




011
111
211
012
112
212
010
110
210
001
101
201
002
102
202
000
100
200

Approximate graph products

Phenotype and Genotype
Characters (traits)

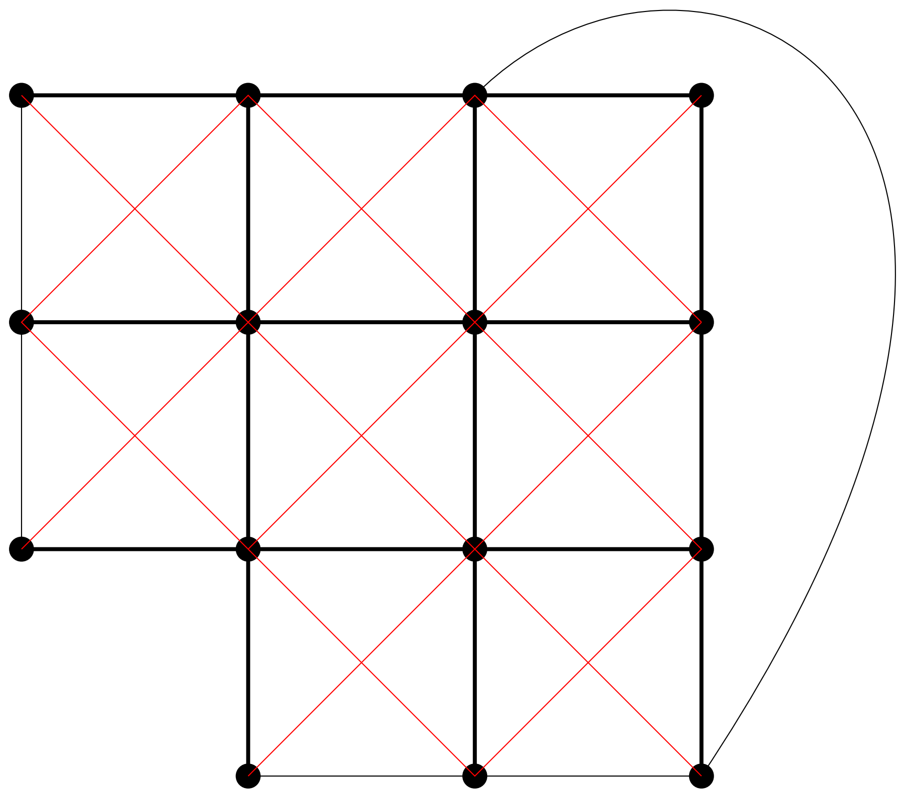
Graphs are an appropriate topology for the description of the relationship between genotypes and phenotypes. And "characters" correspond to "factors" of the graphs (Stadler, B. and P., Wagner, G., and Fontana, W., 2001)

- Biological data are rarely complete
- Biological data are always noisy

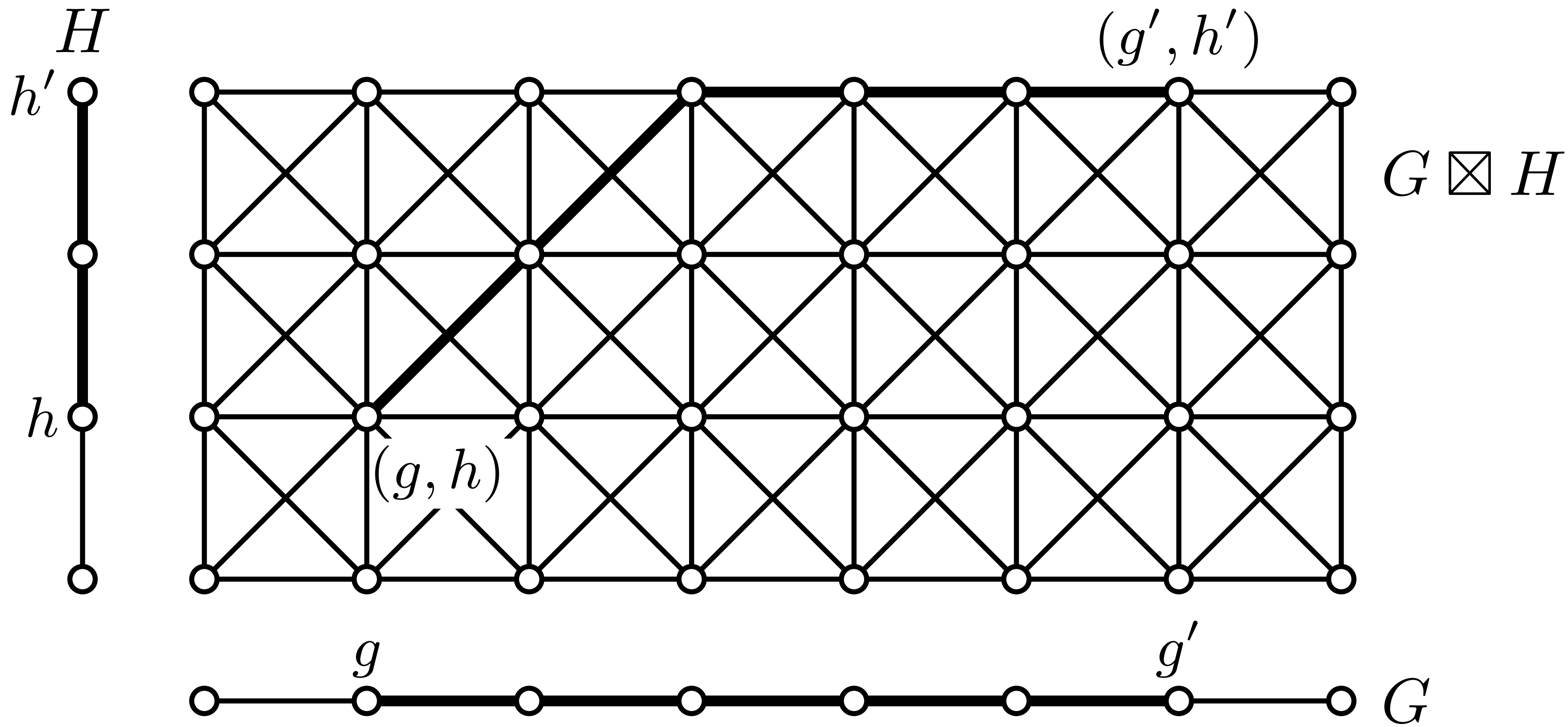
Thus any meaningful approach in bioinformatics must be able to deal with inaccuracies in the input data.

This leads to the investigation of approximate graph products - and quasilinear algorithms to recognize them (Hellmuth, Imrich, Stadler).

Consider the following approximate product:

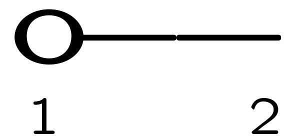


Complexity $O(m\Delta^6)$. (Marc Hellmuth)


H
h'
h
(g', h')
$G \boxtimes H$
(g, h)
g
g'
G

Powers with respect to the direct product

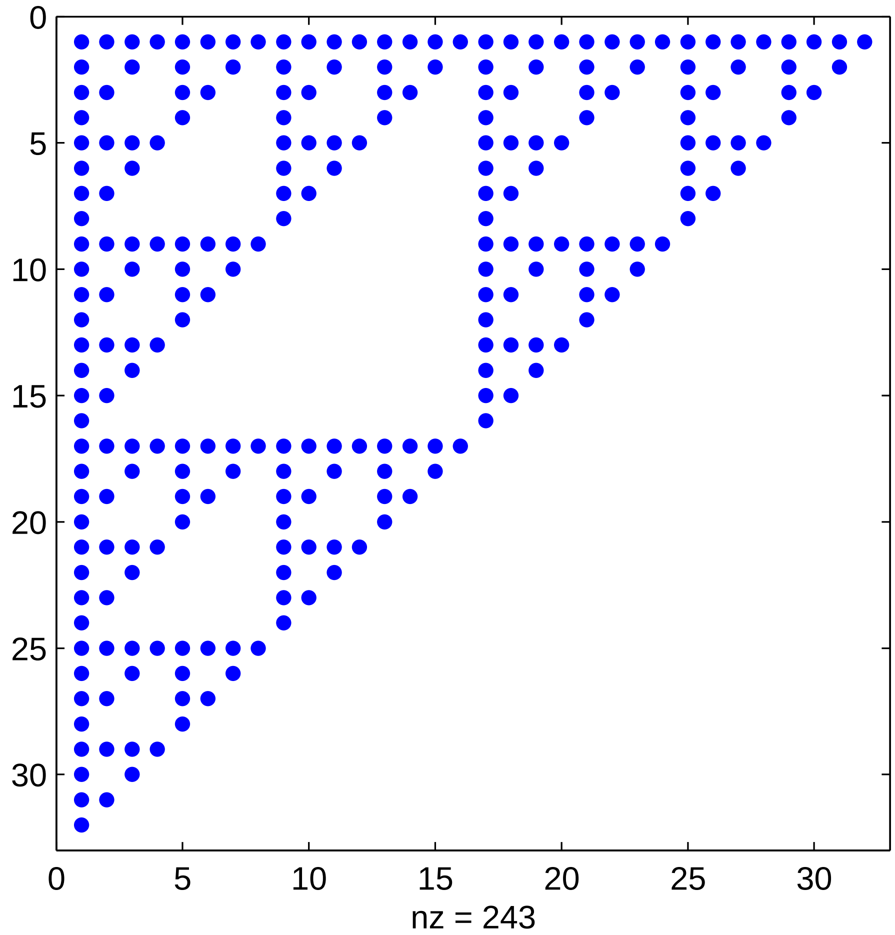
Graph G:



The following matrices are the adjacency matrices of G and of its second and third power with respect to the direct product.

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \quad A \otimes A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad A^3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

And here are the nonzero elements of the fifth power:



Properties of (high) powers G^n of a graph G with respect to the direct product:

- fractal structure
- same diameter as G
- power law degree distribution (Pareto distribution)

$$P(k) \simeq ak^{-\gamma}$$

Variance of the degrees very large

Makes the graph in a sense scale free

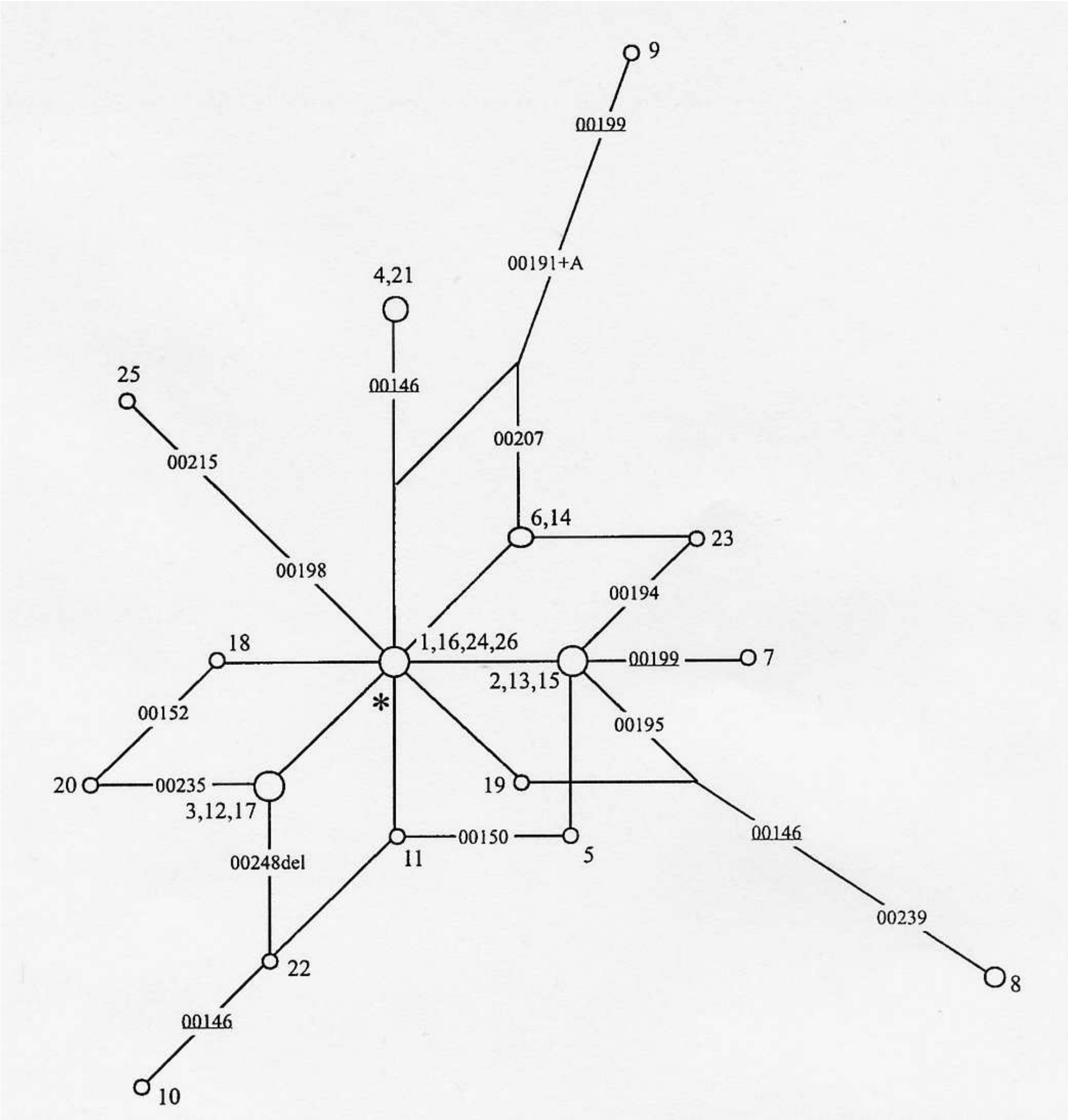
Relationship with the fractal structure.

Complex Networks

Networks arise e.g. in biology, ecology, mathematical chemistry, software technology, operations research.

Investigation became a hot research topic only in the last decade, coinciding with increased interest in the Internet network(s), social networks, citation networks, neural networks, and so on.

What does one study:reliability, reachability, distance (small world phenomenon), virus propagation, and so on.



Relationshipgraph

Mediannetwork for human mtDNA (Bandelt et al. 2000)

(Median graphs are retracts of hypercubes)

Generation of probabilistic graphs with the desired properties.

Erdös-Renyi model for random graphs, it does not satisfy all requirements, in particular not the heavy tail distribution (scale freeness)

Models by Barábasi and Albert (1999) computationally expensive.

An appealing approach was made by Leskovec, J. Chakrabarti, D. Kleinberg, J. Faloutsos, C. Ghahramani, Z. using the direct product of graphs (or, equivalently, the Kronecker product of matrices)

Given a graph G, the adjacency matrix of the direct product $G^{\times,k}$ is the kth power of the adjacency matrix of G with respect to the Kronecker product.

Start with a square probability matrix $\mathcal{P}_1$ whose entries i-j entry represents the probability that an edge joins vertex i to vertex j.

Compute the Kronecker kth power $\mathcal{P}_k$.

Then (an instance of) a *stochastic Kronecker graph* is obtained from $\mathcal{P}_k$ by including an edge between two vertices with probability as given in $\mathcal{P}_k$.

Stochastic Kronecker graphs can be generated in linear time with respect to their expected number of edges.

Stochastic Kronecker are also close to real world networks.

Degree distribution Real networks exhibit the so-called "heavy tail degree distribution" and this is also the case with the stochastic Kronecker graph.

Fitting real-world networks Given a real network, finding a stochastic Kronecker graph that is "similar" to the network can be done efficiently.

Multitude of technical details is omitted

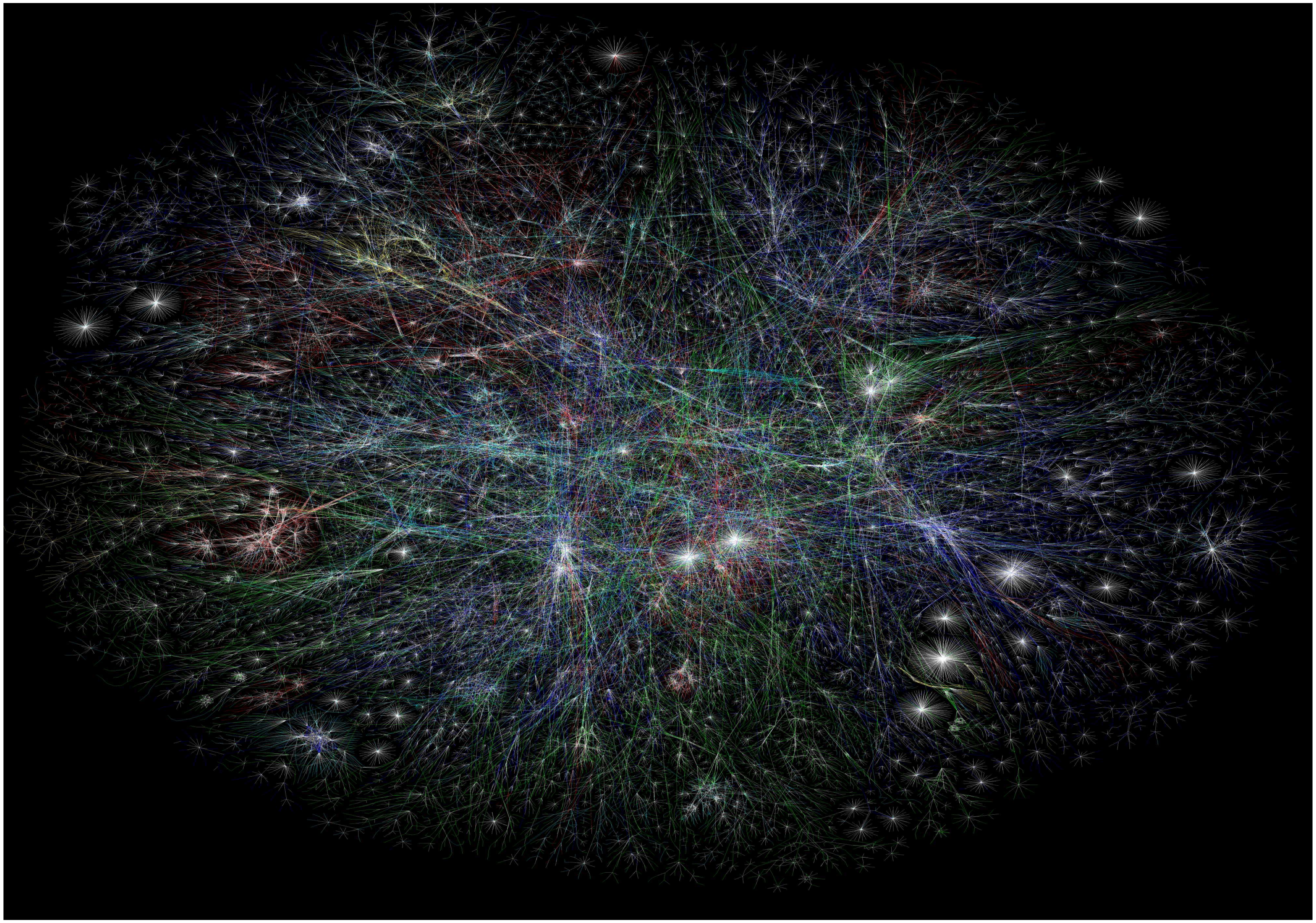
Leskovec et al. add loops to every vertex of the generating graphs. They really work with the strong product, not necessary.

Leskovec and Faloutsos (2007) showed that the simple generating matrix

$$\begin{pmatrix} .98 & .58 \\ .58 & .06 \end{pmatrix}$$

yields a Kronecker graph that fits the Internet fairly well.

Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., and Ghahramani, Z. (2009). Kronecker graphs: an approach to modeling networks. J. Mach. Learn. Res., x (x), x-x.



DISCRETE MATHEMATICS AND ITS APPLICATIONS

Series Editor KENNETH H. ROSEN

HANDBOOK OF PRODUCT GRAPHS

SECOND EDITION

Richard Hammack

Wilfried Imrich

Sandi Klavžar

CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK