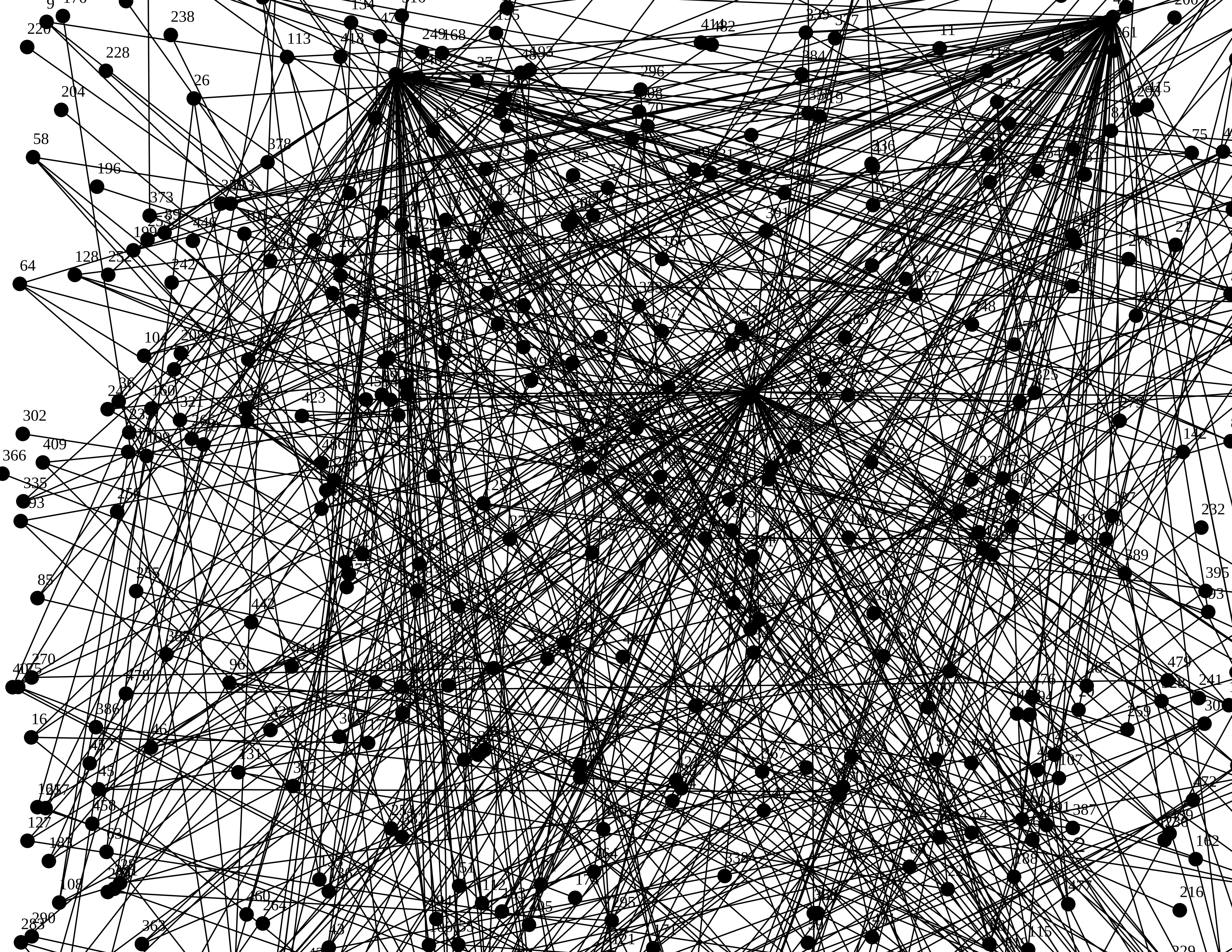


Generic Chemical Reaction Network Properties

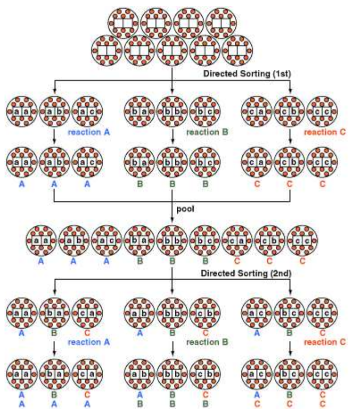
Gil Benkö

Institute for Theoretical Chemistry and Structural Biology

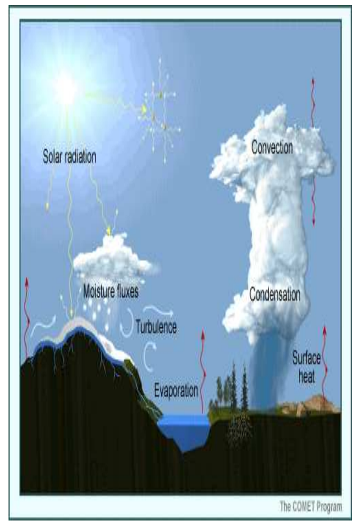
gil@tbi.univie.ac.at

Bled, 2003

Why ?

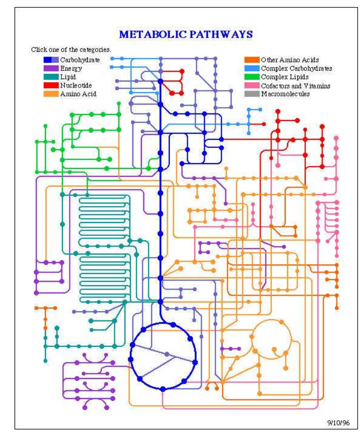


Combinatorial chemistry



Atmospheric processes

Combustion



Metabolic networks

Chemical reaction networks

What ?

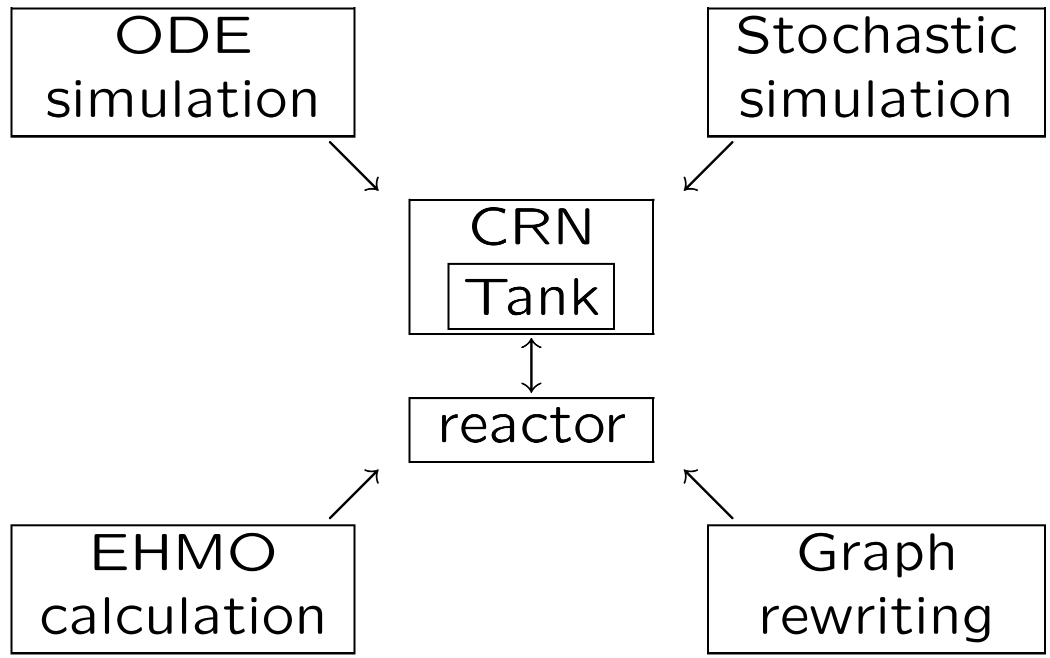
What are the generic properties of a CRN :

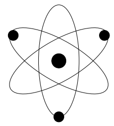
Are average minimum path lengths short ?
(small-world property)

What is the scaling behavior of the graph?

How robust are these properties?

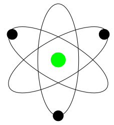
Organization of the Toy Model

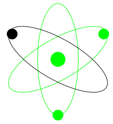


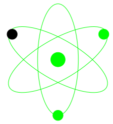
Electronic Energy Calculation 

Schrödinger equation
$$\hat{H}\Psi = E\Psi$$

$\Downarrow$

Born-Oppenheimer 

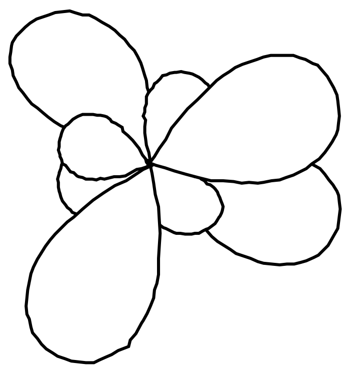
LCAO and Extended Hückel 

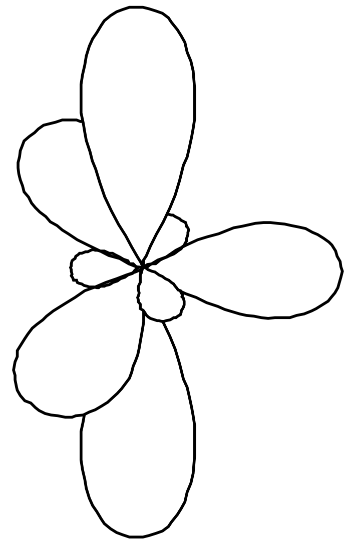
VSEPR and Tight Binding 

Generalized Eigenvalue Problem

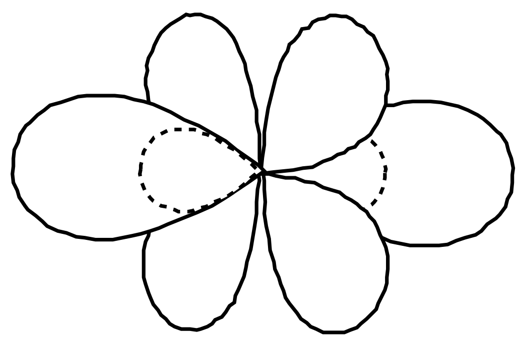
$$\begin{pmatrix} \ddots & & \\ & H_{AO_i - AO_j} & \\ & & \ddots \end{pmatrix} \mathbf{C} = \begin{pmatrix} \ddots & & \\ & S_{AO_i - AO_j} & \\ & & \ddots \end{pmatrix} \mathbf{CE}$$

Atom Orbitals



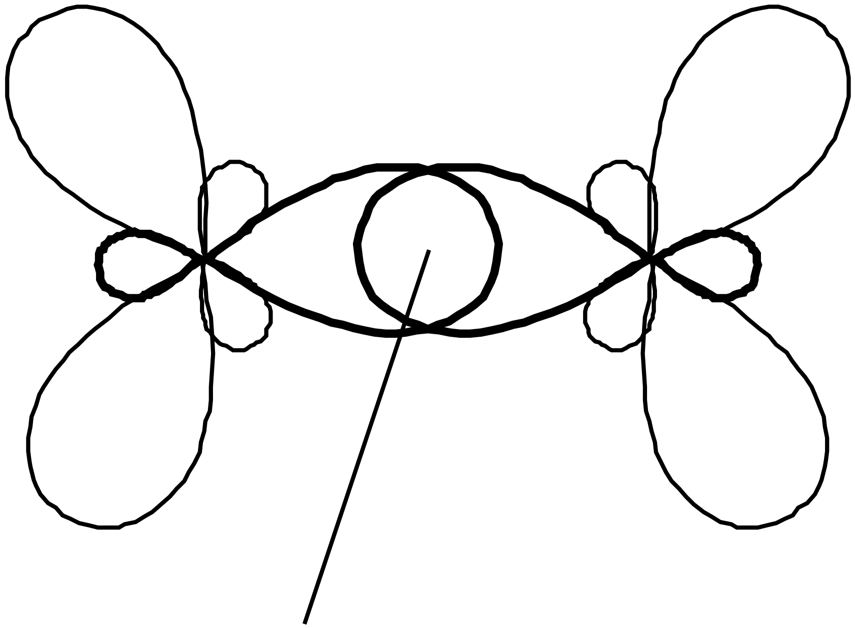


sp^3 hybridized sp^2 hybridized + one p

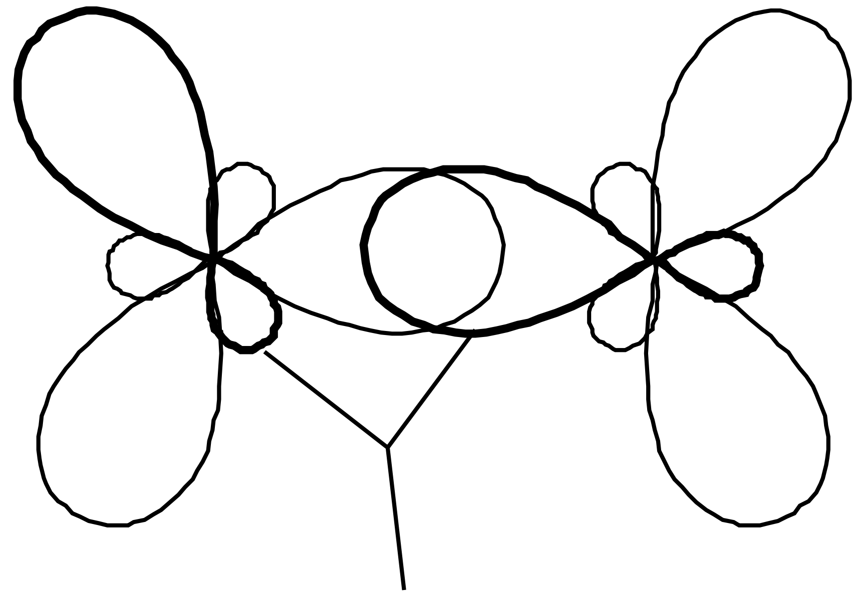


sp hybridized + two p

Implemented Overlaps

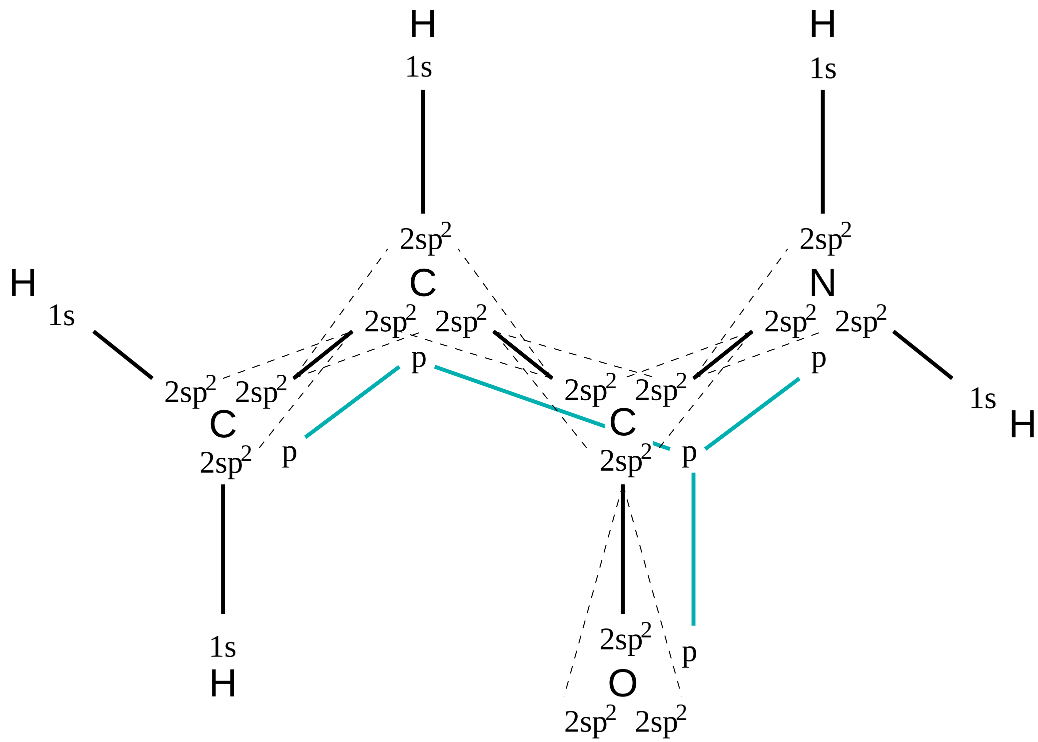


big σ-overlap
between sp^2 orbitals



lesser overlap
between sp^2 orbitals

Orbital graph



Overlap matrix S of propenamide

$$C^1(H^2)(H^3) = C^4(H^5)C^6(= O^7)N^8(H^9)H^{10}$$

C^1 sp^2	H^2 s	C^1 sp^2	H^3 s	C^1 sp^2	C^4 sp^2	C^4 sp^2	H^5 s	C^4 sp^2	C^6 sp^2	C^6 sp^2	O^7 sp^2	C^6 sp^2	N^8 sp^2	N^8 sp^2	H^9 s	N^8 sp^2	H^{10} s	O^7 sp^2	O^7 sp^2	C^1 p	C^4 p	C^6 p	O^7 p	N^8 p
1	.65	0	0	0	.077	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.65	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	.65	0	.077	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	.65	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	.77	.077	0	.077	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.077	0	.077	0	.77	1	0	0	0	.077	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	.077	0	1	.65	0	.077	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	.65	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	.077	0	0	0	1	.77	.077	0	.077	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	.077	.077	0	.77	1	0	.068	0	.073	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	.077	0	1	.68	0	.073	0	0	0	0	.068	.068	0	0	0	0	0
0	0	0	0	0	0	0	0	0	.068	.68	1	.068	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	.077	0	0	.068	1	.73	.073	0	.073	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	.073	.073	0	.73	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	.073	0	1	.63	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	.63	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	.073	0	0	0	1	.63	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.63	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	.068	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	.068	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	.38	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.38	1	.38	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.38	1	.26	.31
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.26	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.31	0	1

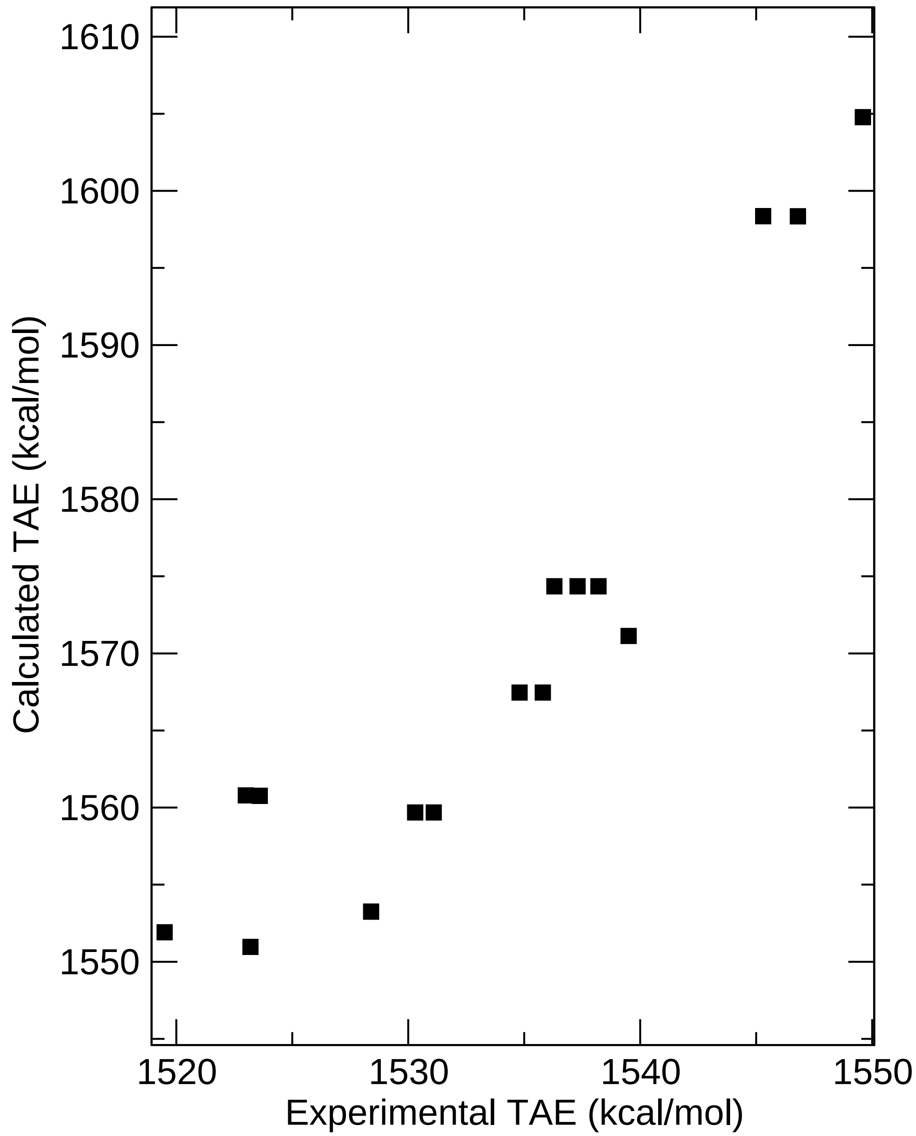
Wave function analysis

By definition, any molecular property can be calculated from the wave function:

- Energy
- Charge distribution
- Reactivities

Calculated vs Experimental TAE of C6H10 isomers

(Total Atomization Energy)



Structure representation I : SMILES;)

CH_4

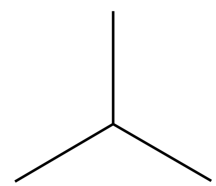
C

or C(H)(H)(H)H

H_3C-CH_3

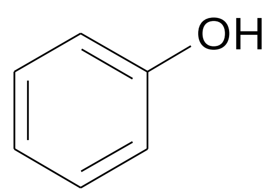
CC

or C(H)(H)(H)C(H)(H)(H)



CC(C)C or

C(H)(H)(H)C(H)(C(H)(H)(H))C(H)(H)(H)



OC(C = C1) = CC = C1 or

HOC(C(H) = C1H) = C(H)C(H) = C1H

Structure representation II : GML

```
# Isobutane
graph [
  node [ id 1 label "C" ]
  node [ id 2 label "C" ]
  node [ id 3 label "C" ]
  node [ id 4 label "C" ]

  edge [ source 1 target 2 label "-" ]
  edge [ source 1 target 3 label "-" ]
  edge [ source 1 target 4 label "-" ]
]
```

What is Graph Rewrite ?

A grammar which operates on graphs instead of strings.

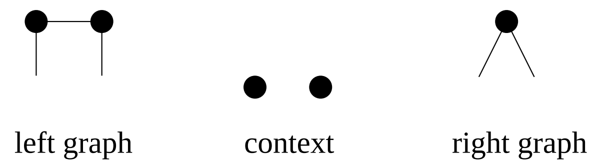
A grammar is a finite set of rules describing a formal language.

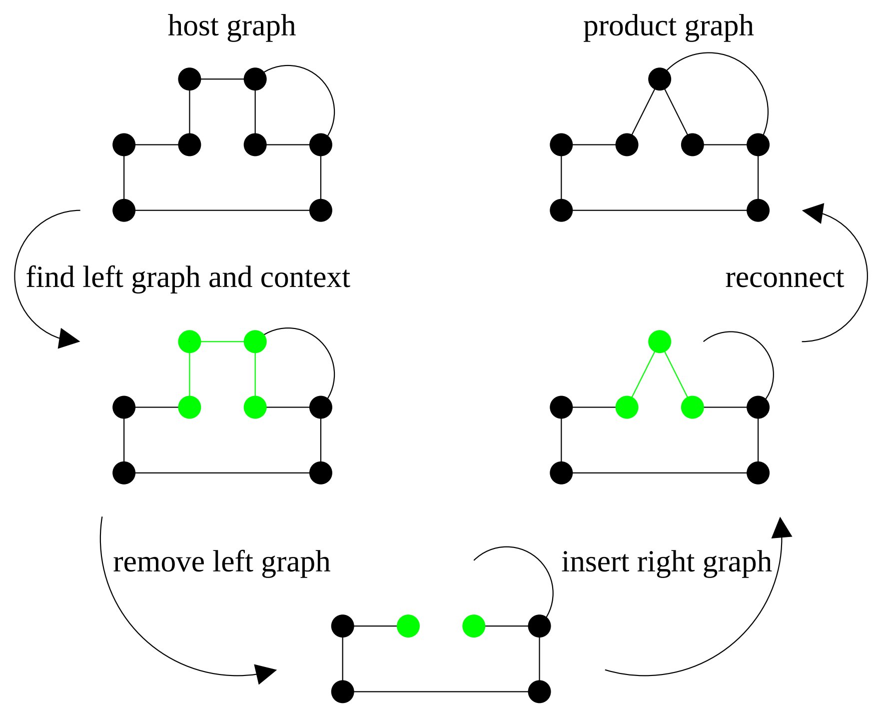
A formal language is a set of strings over some fixed alphabet.

Graph Rewriting Step

Step 1: **find** isomorphic subgraph (match left graph).
Step 2: **remove** subgraph (don't touch context; keep dangling ends!!).
Step 3: **insert** new subgraph (right graph).
Step 4: **rewire** new subgraph (with respect to the dangling ends).

Graph rewriting steps





Network Generation

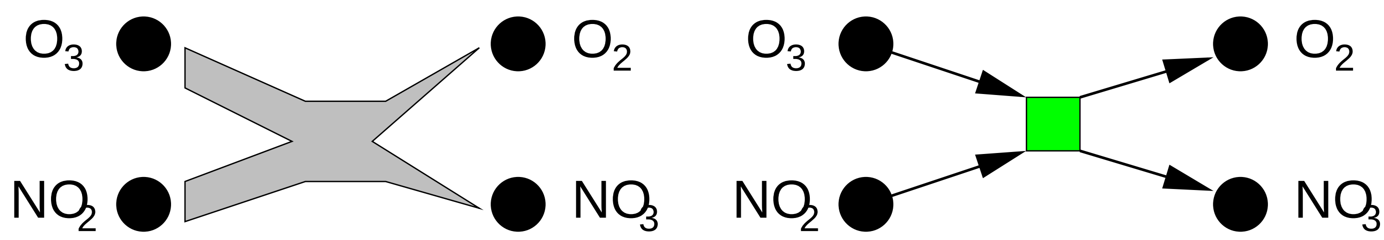
Start perform all unimolecular and bimolecular reactions on the molecules in $\mathfrak{L}_0$ and put the products in a new set $\mathfrak{L}'_1$, eliminating all duplicates. This is summarized by the notation $\mathfrak{L}'_1 = \mathfrak{L}_0 \otimes \mathfrak{L}_0$. Calculate $\mathfrak{L}_1 = \mathfrak{L}'_1 \setminus \mathfrak{L}_0$.

Recursion (1) $\mathfrak{L}'_{k+1} = \left(\bigcup_{j=0}^{k-1} \mathfrak{L}_j\right) \otimes \mathfrak{L}_k \cup (\mathfrak{L}_k \otimes \mathfrak{L}_k)$

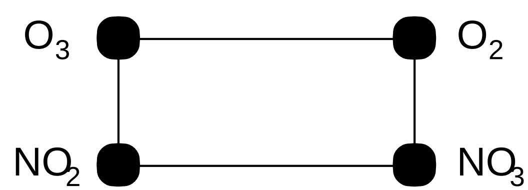
(2) and $\mathfrak{L}_{k+1} = \mathfrak{L}'_{k+1} \setminus \bigcup \mathfrak{L}_k$.

Network Representation

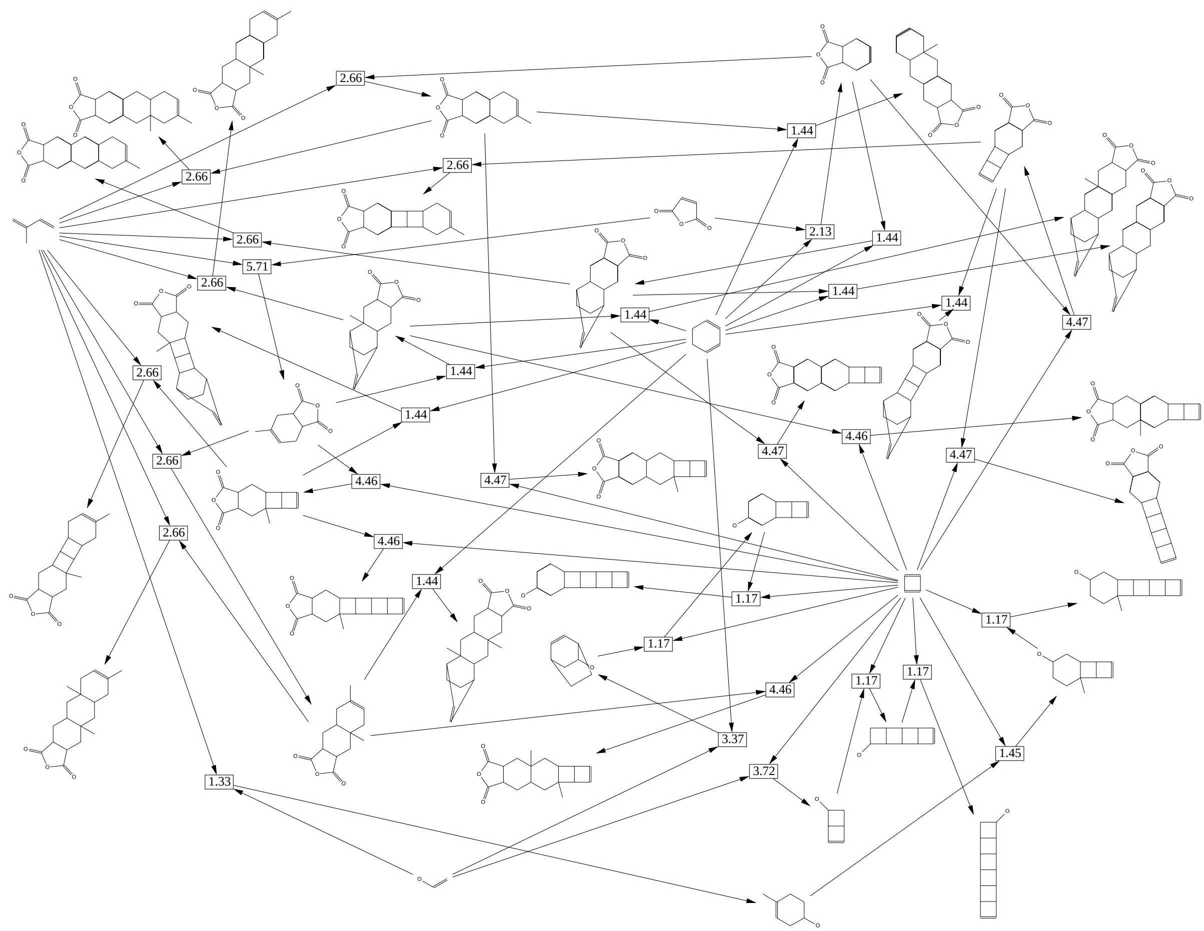
Hypergraph and bipartite graph
for $O_3 + NO_2 \longrightarrow O_2 + NO_3$



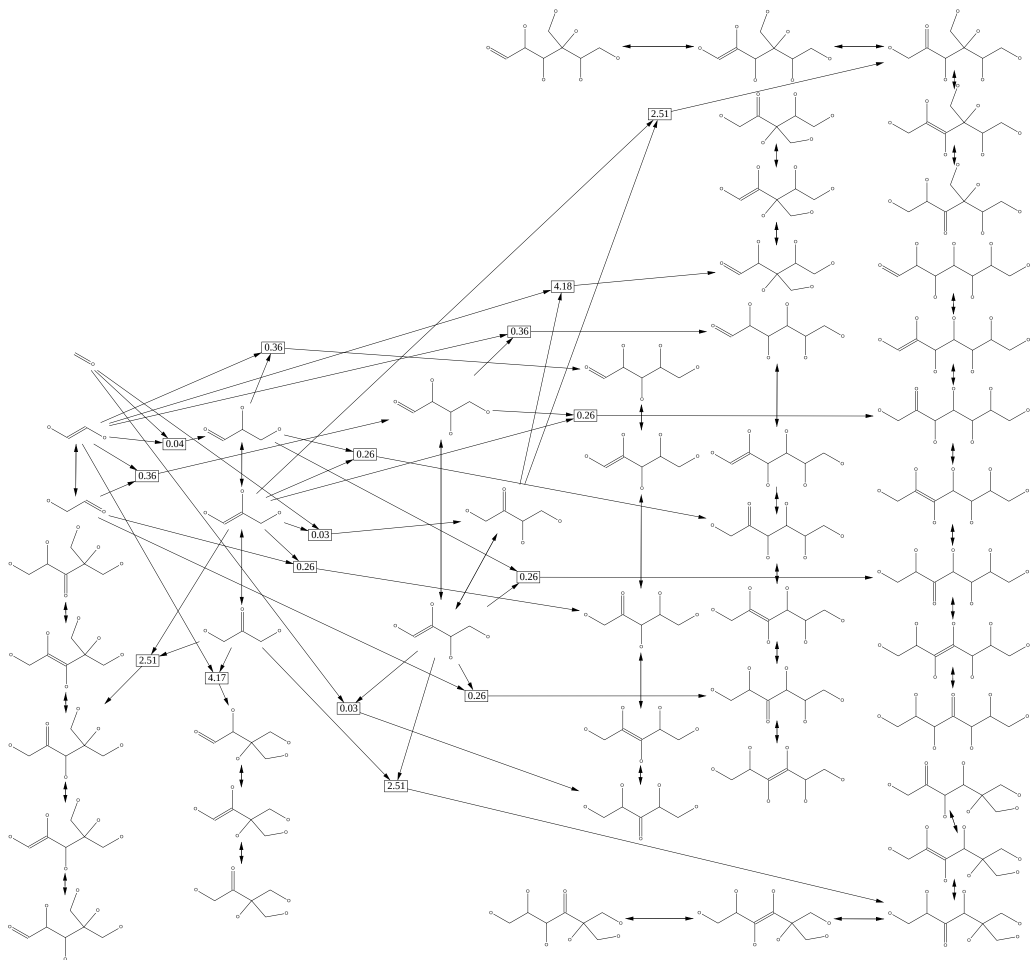
one-mode projection $\longrightarrow$ Substrate graph



Application I : Repetitive Diels-Alder



Application II : Formose reaction



Graph properties

- $< k >= 2\,\frac{\text{edges}}{\text{nodes}}$ is the average node degree.

- $< L >$ is the average length of the shortest path between to nodes.

- $C_i = 2\,\frac{\text{edges between i-neighbours}}{\text{i-neighbours(i-neighbours}-1)}$ is the clustering coefficient.

Computational results

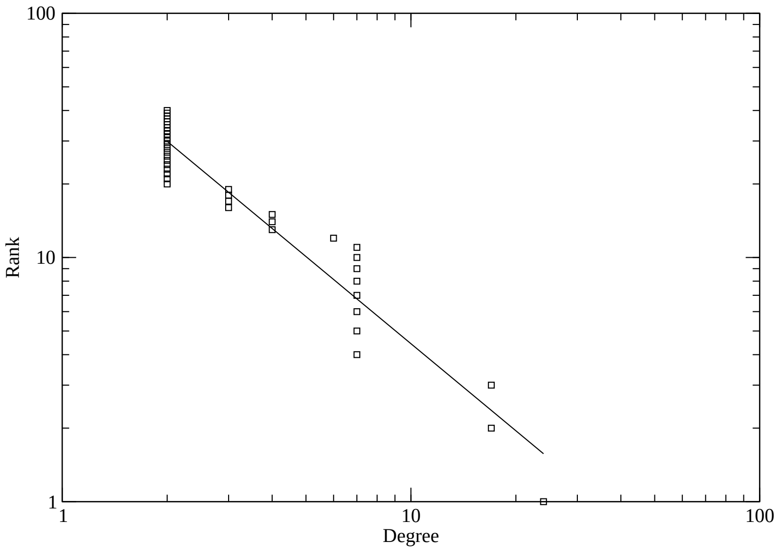
For comparing: results for random Erdös-Renyi graphs

$$< L_{rand} > \approx \frac{\ln n}{\ln < k >}$$
$$< C_{rand} > = \frac{< k >}{(n-1)}$$

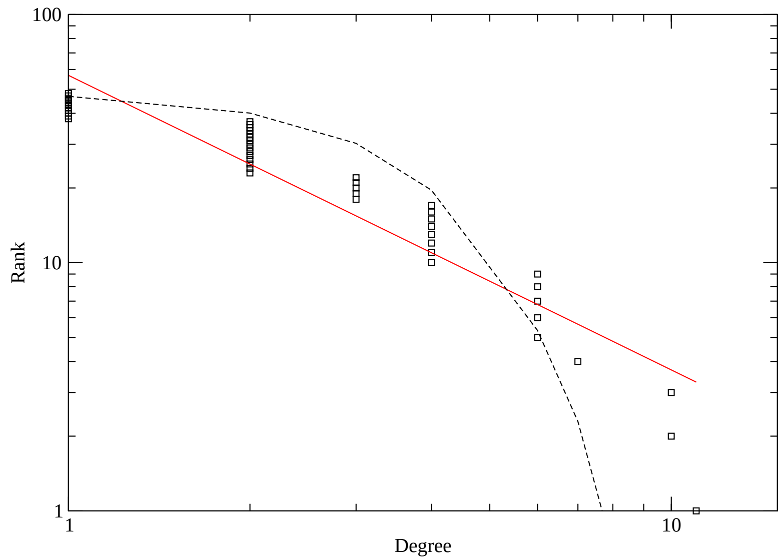
	nodes	$< k >$	$< L >$	$< L_{rand} >$	$< C >$	$< C_{rand} >$
Diels-Alder	40	4.65	2.15	2.40	0.72	0.11
Formose	48	3.25	3.55	3.28	0.15	0.068
E. coli	282	7.35	2.9	3.04	0.32	0.026

Degree distribution

Repetitive Diels-Alder



Formose reaction

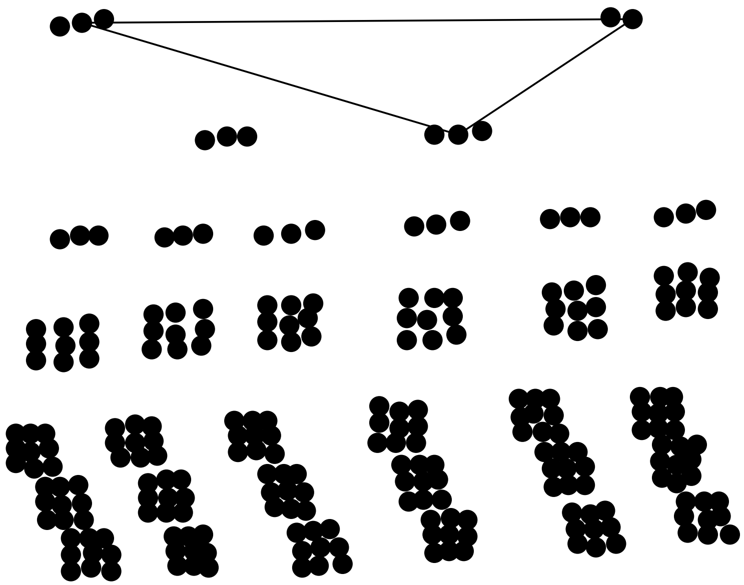


□ : datapoints

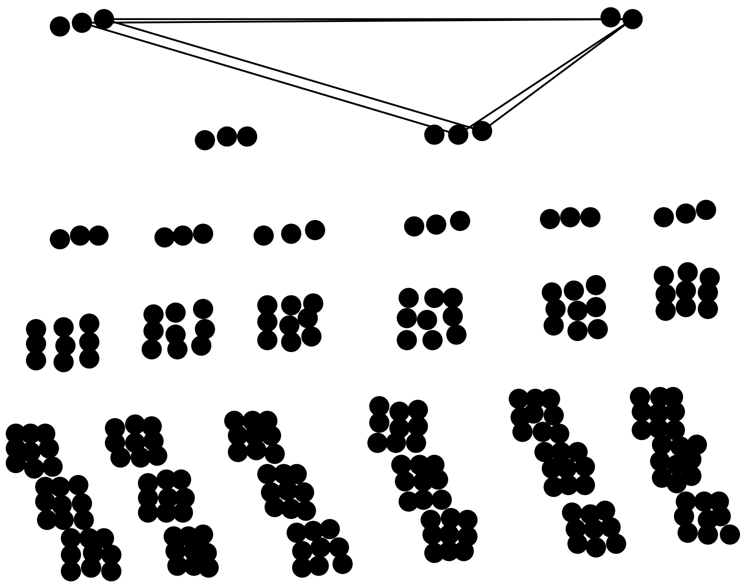
solid : power-law regression

dashed : Poisson distribution

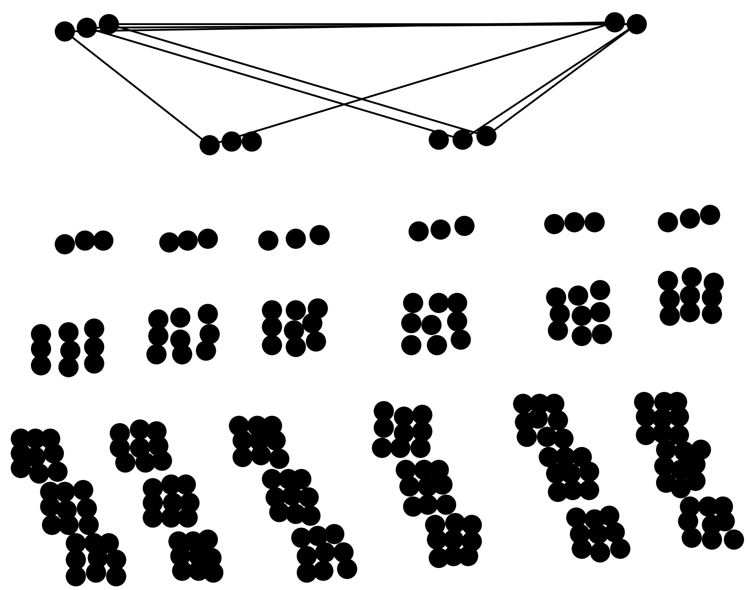
Threshold of 112



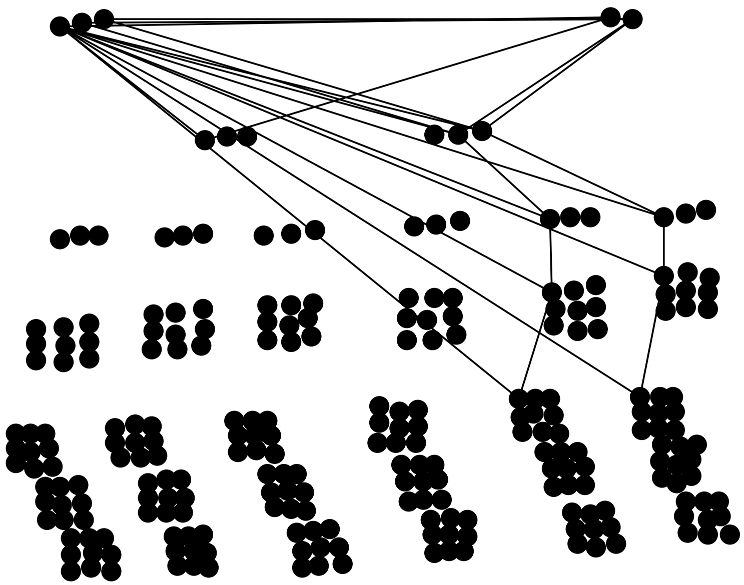
Threshold of 108



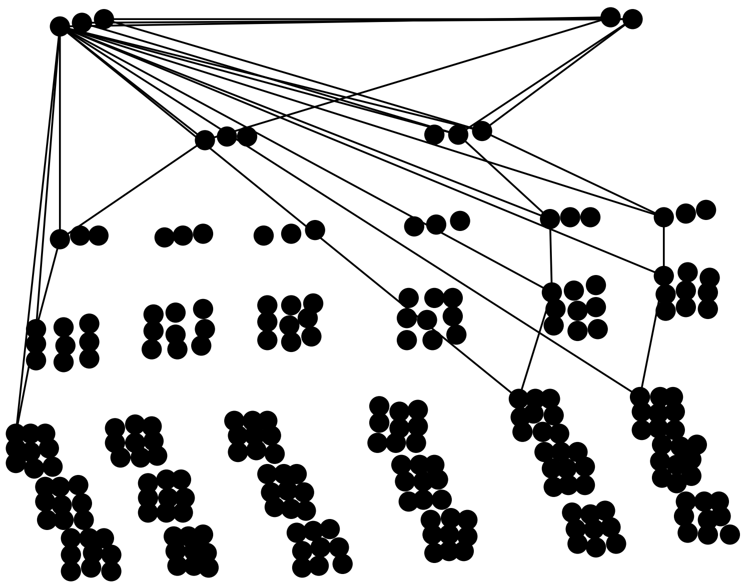
Threshold of 107



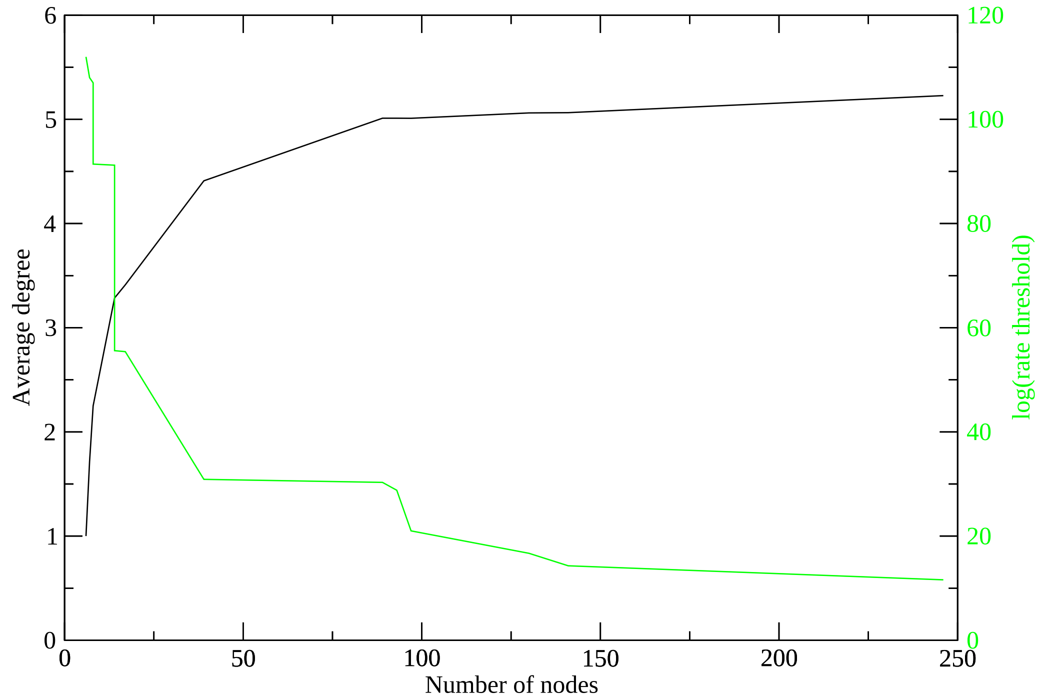
Threshold of 91.2



Threshold of 55.4

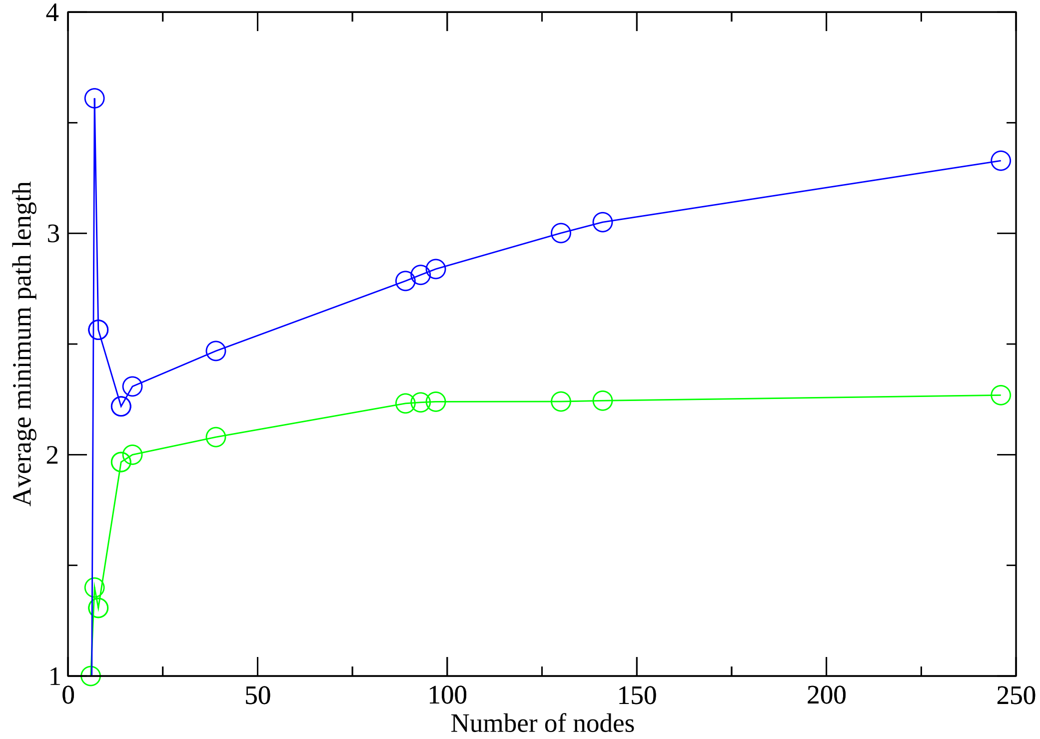


A larger CRN . . .



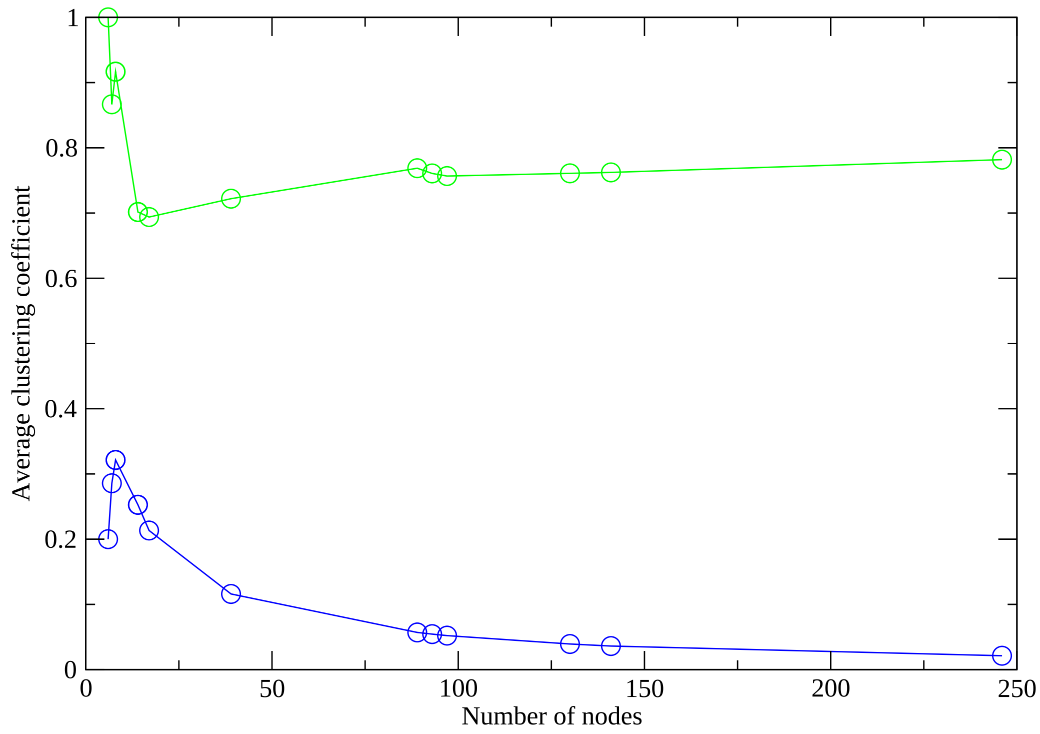
$< k >$ and reaction rate threshold vs. CRN size n

... with constant minimum path length $< L >$ ...



CRN size dependency of $< L >$ and $< L_{rand} >$

... and clustering coefficient $< C >$



CRN size dependency of $< C >$ and $< C_{rand} >$

$\longrightarrow$ **robust small-world property**