Evolution of complex networks – how does the Internet grow

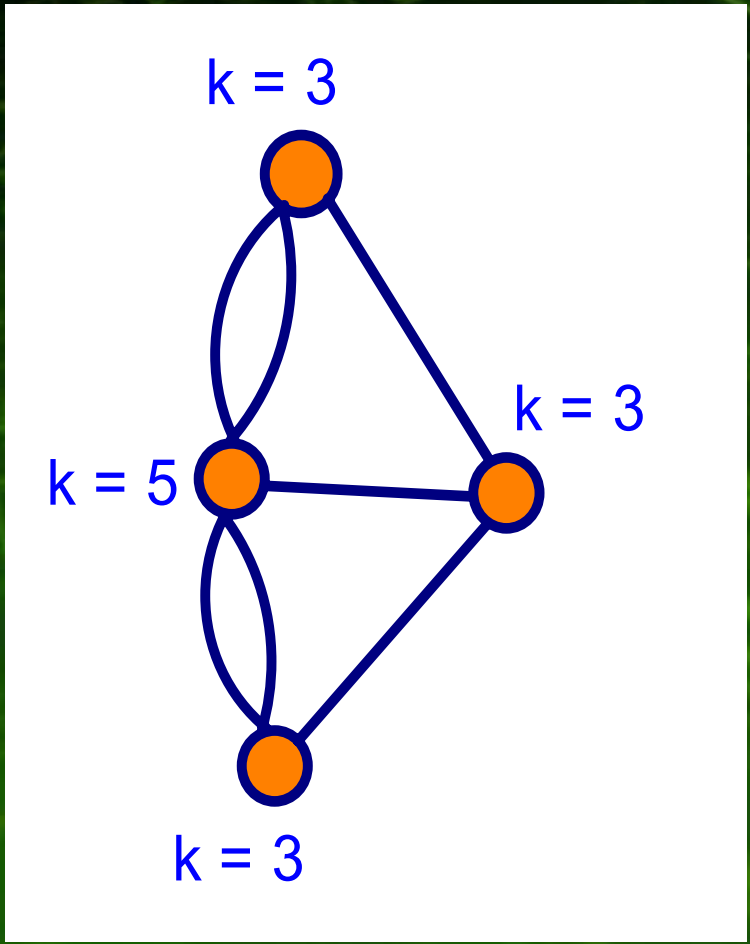
Final presentation

Students: Nino Antulov Fantulin
Branko Đurđević
Fabian Bremer

Project leader: Hrvoje Štefančić

Complex networks

- between completely regular and completely random
- can describe very diverse systems
- also serve as a new way of representing complex systems
- main elements:
  - nodes
  - links



Description of complex networks

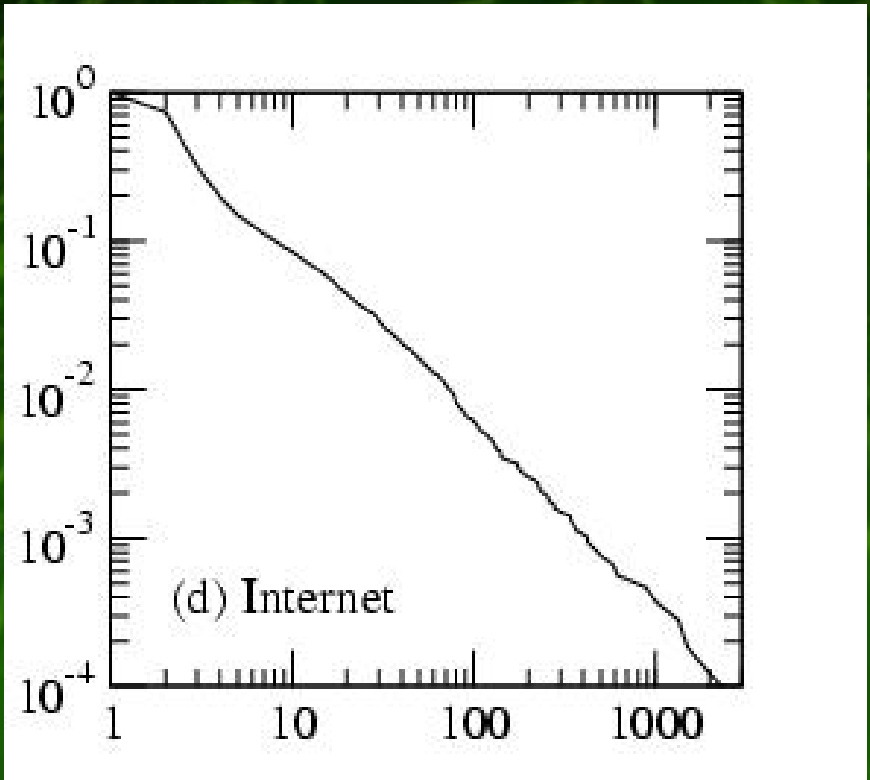
- node degree distribution
  - node degree k is the number of links attached to that node
  - node degree distribution P(k) is the probability that a randomly chosen node in a network has the degree k
- clustering coefficient
  - the ratio of the number of realized links between node's first neighbouring nodes and the maximal possible number of links between these nodes

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

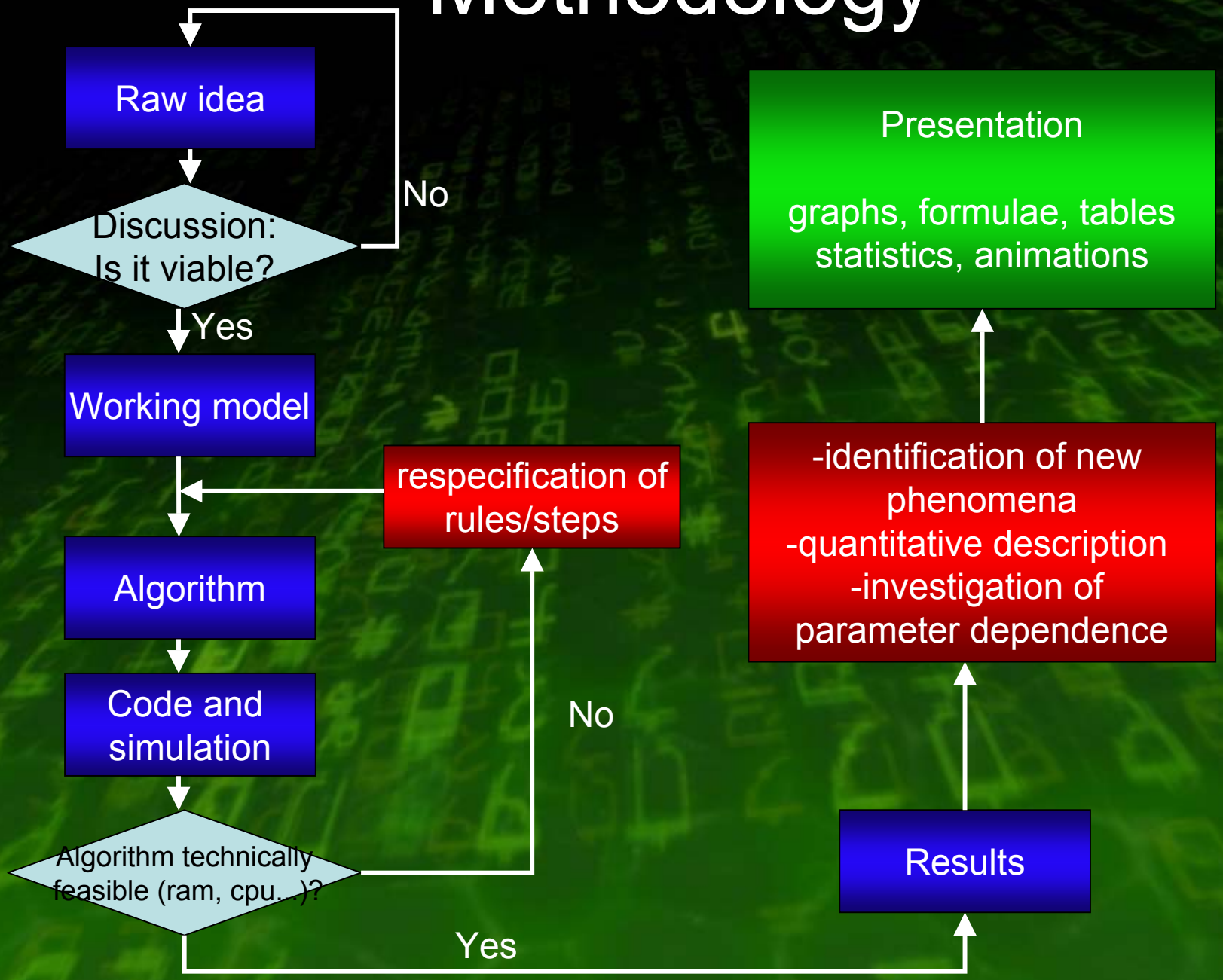
- average shortest path length

Complex networks in nature and society

- large majority caracterized by the straight line in the log k – log P(k) plot
- these are scale-free networks



Methodology



Goals of the complex networks study

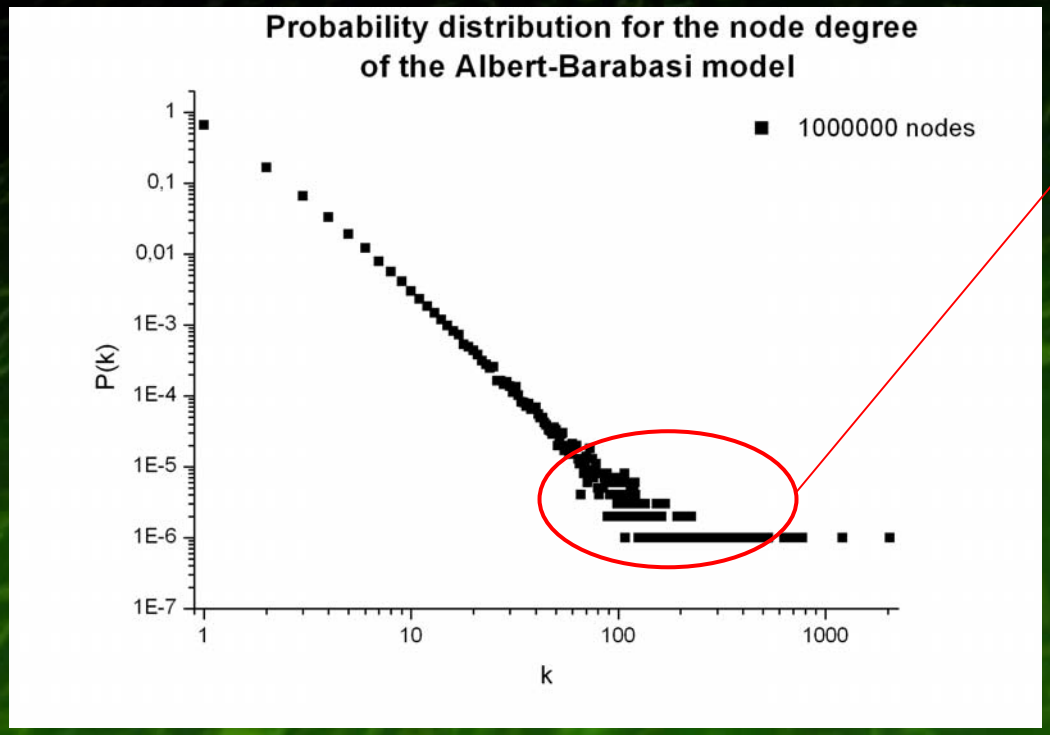
- getting familiar with the statistical description of complex networks
- studying the structure of complex networks by modeling their growth
- developing the software for the analysis of complex networks
- application to technological and/or sociological and/or biological problems

Albert - Barabasi model

- growing network – at each time step one node is added and attached to one of the existing nodes
- the new node is attached preferentially – it is more likely to attach itself to a node with more connections
- probability of attachment is proportional to the degree of the old node

Evolution of the network

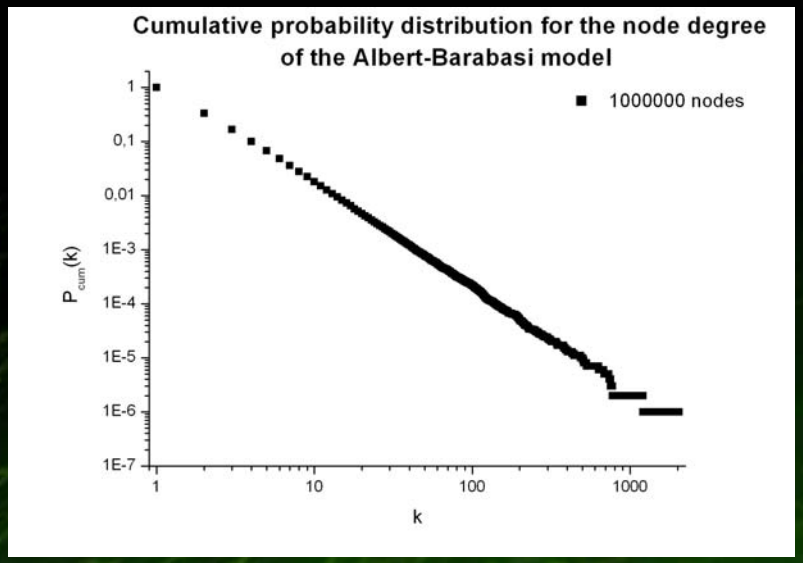
AB model results (1)

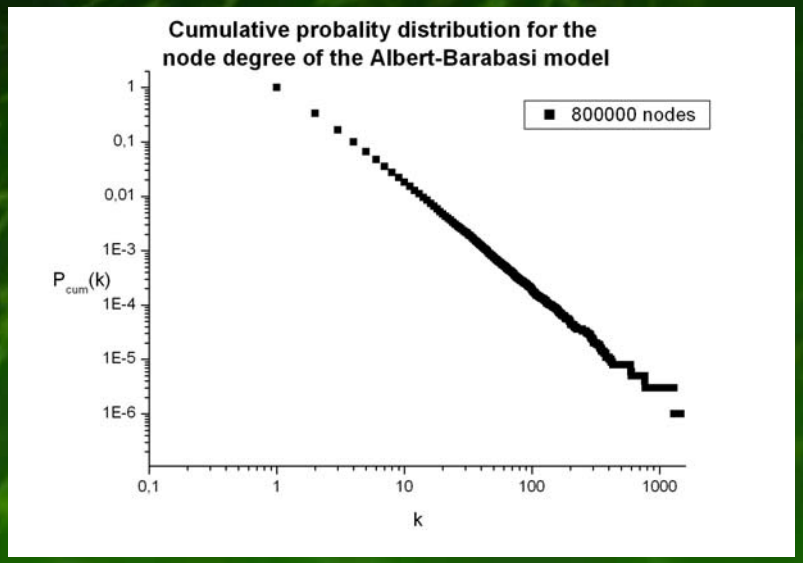


noise in the tail of the distribution

-straight line in the log-log plot
-slope is -3
- implies dependence $P(k) = Ck^{-3}$

AB model results (2)



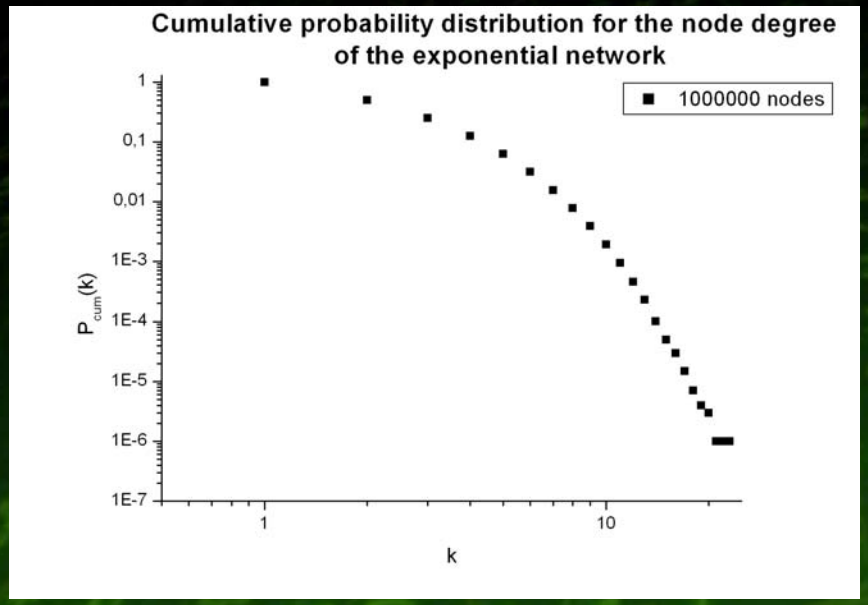


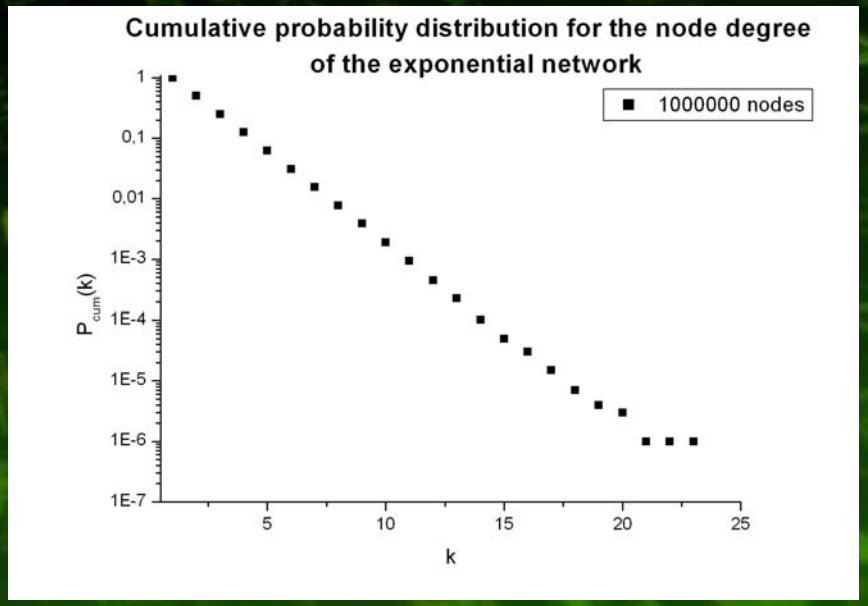
- **straight lines with slope -2**
- **noise significantly reduced**
- **dependence of the power law form**

$$P_{cum}(k)=Ak^{-2}$$

Agreement with the results of Barabasi and Albert

Testing of alternatives (random attachment)





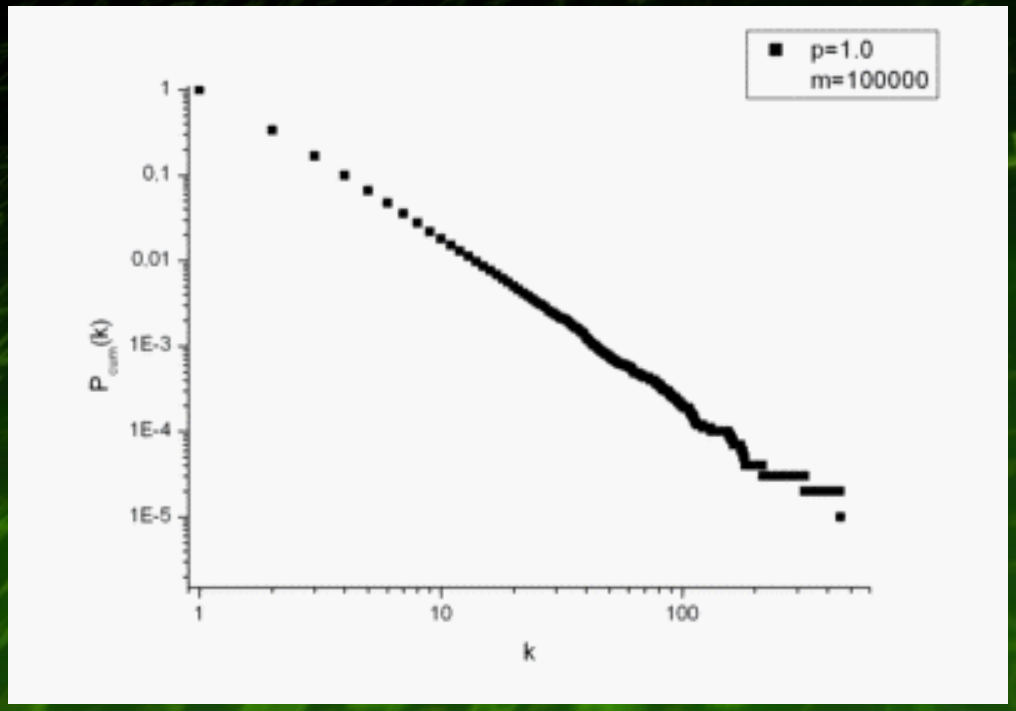
- no straight line in log-log plot

- straight line in lin-log plot

Exponential distribution → the network is not scale-free

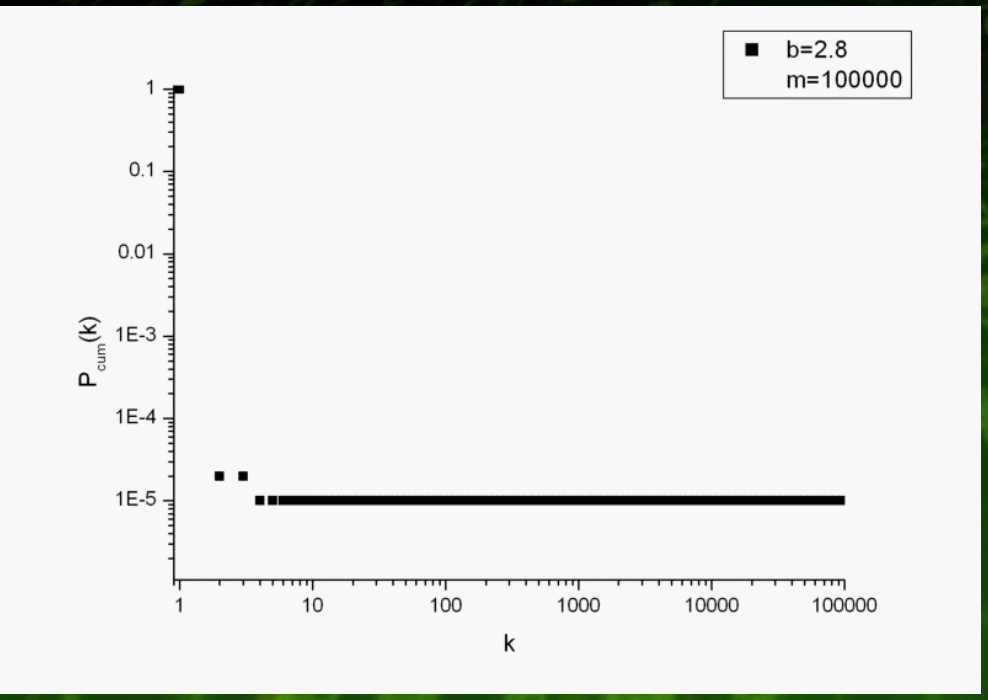
Conclusion: Preferential attachment is needed to obtain scale-free networks.

Hybrid network building



- combining exponential and power law networks
- randomly choosing between Albert-Barabasi model (with probability p) and random model (with probability 1-p) for the connection of the new node
- p=0 -> random model
- p=1 -> Albert-Barabasi model

Generalized preferential linking



- new parametar b introduced
- the new node is attached to the old node with probability ~ k^b (k is the degree of the old node)
- b=1 -> Albert-Barabasi model
- b=0 -> Random model
- b>1 -> condensation of links to a single node
- statistics of condensation:

  out of 30 simulations with 10000 nodes all contained a node with more than 9500 links

Adjacency matrix

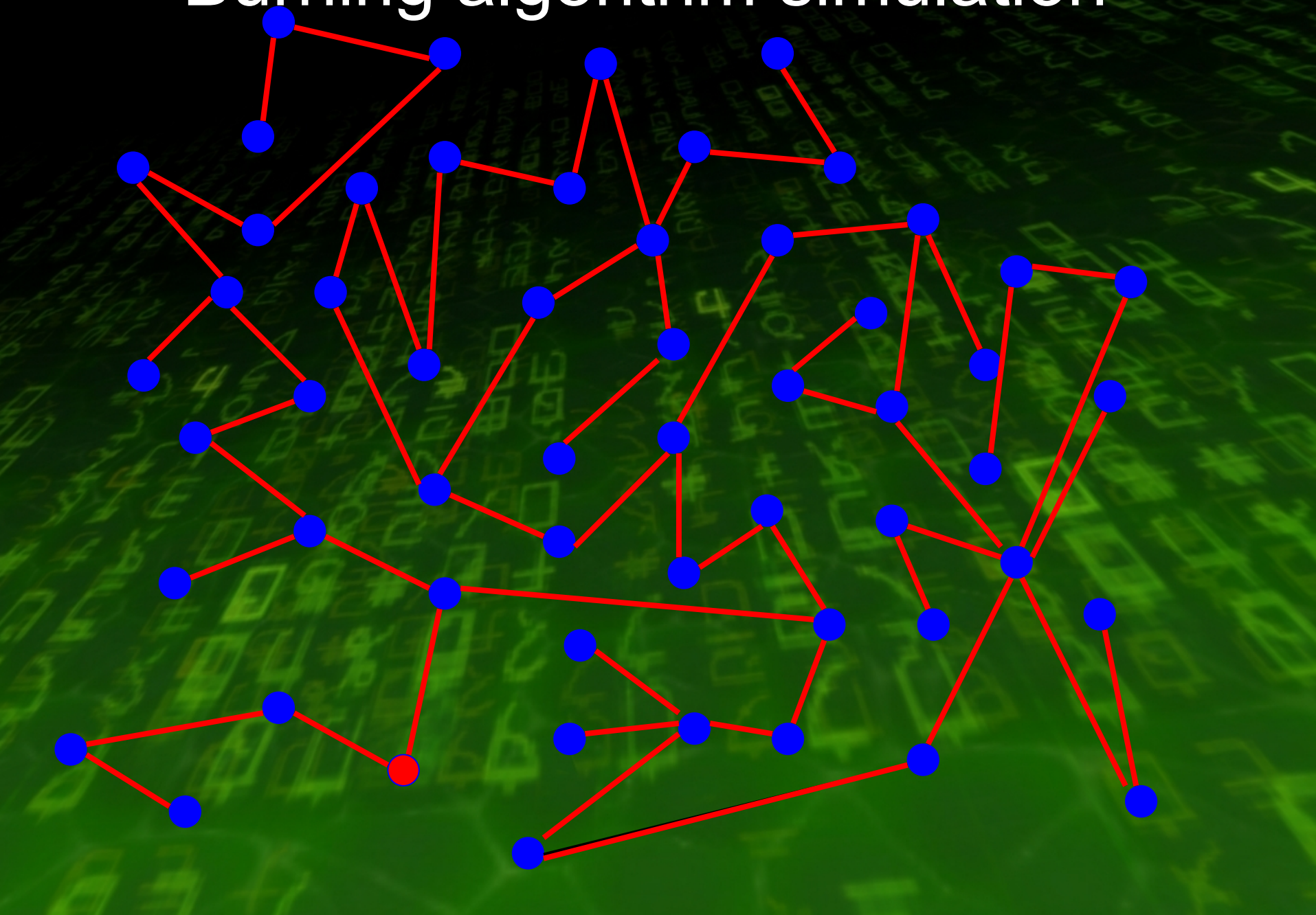
$$C = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

- Complete representation of a network
- $C_{ij} = 1$ if nodes i and j are connected
- $C_{ij} = 0$ if there is no connection between nodes i and j
- $C_{ij} = C_{ji}$
- $C_{ii} = 0$

Burning algorithm

- can be used to determine individual components of a complex network and to calculate the average shortest path
- functions like spreading of fire through the network
- each node that is on fire spreads the fire to its first neighbours if they are not already on fire

Burning algorithm simulation



Optimization of a network

- in order to connect componets of a network in the most efficient way (to increase the efficiency of a network) we first need to identify the components
- calculate its average shortest path
- by knowing it we can optimize a network with minimal cost (minimum number of new links)
- by burning algorithm calculate the best way to connect the components in order to decrease the average shortest path to minimum possible
- the power of the method becomes especially clear when connecting large components of a network

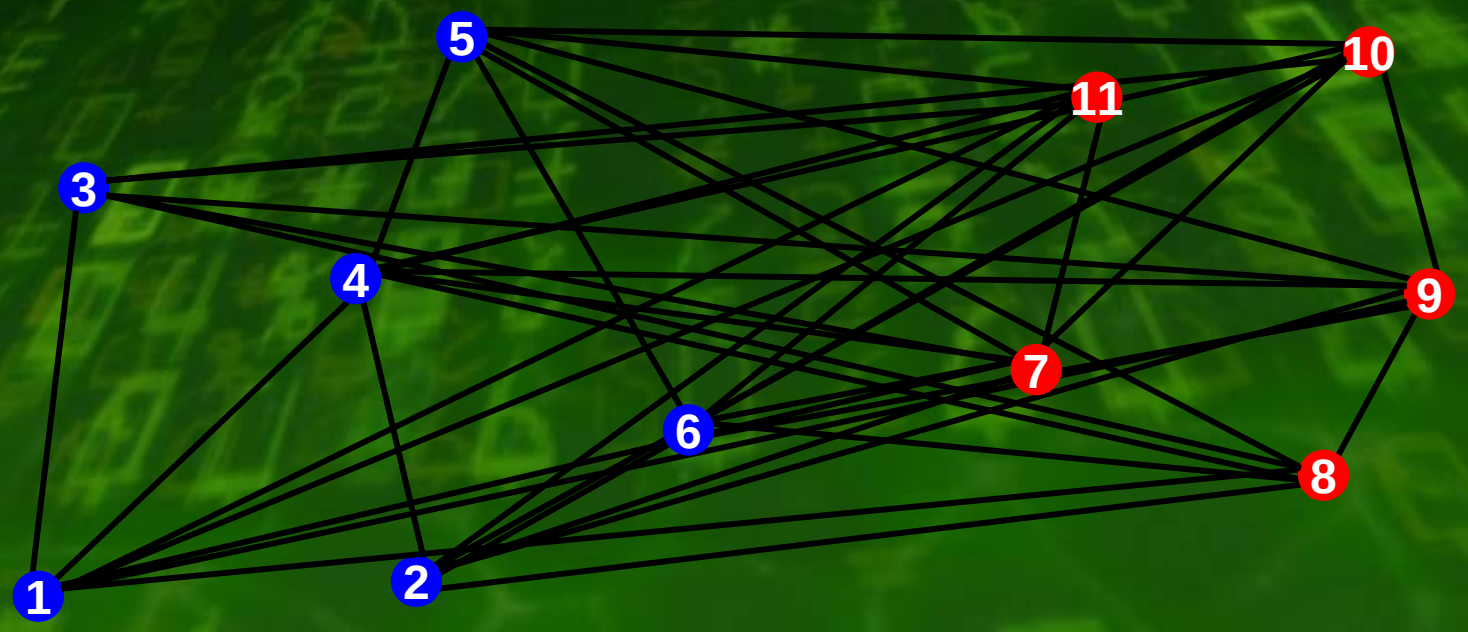
Two component optimization

- calculate the average shortest path before optimization
- a node from the first component connected to a node from the second component and the average shortest path of the so formed network calculated
- procedure repeated for all such pairs of nodes
- the combination of a node from the first and the second component with the minimum average shortest path is used as best optimization of a network using the addition of one link

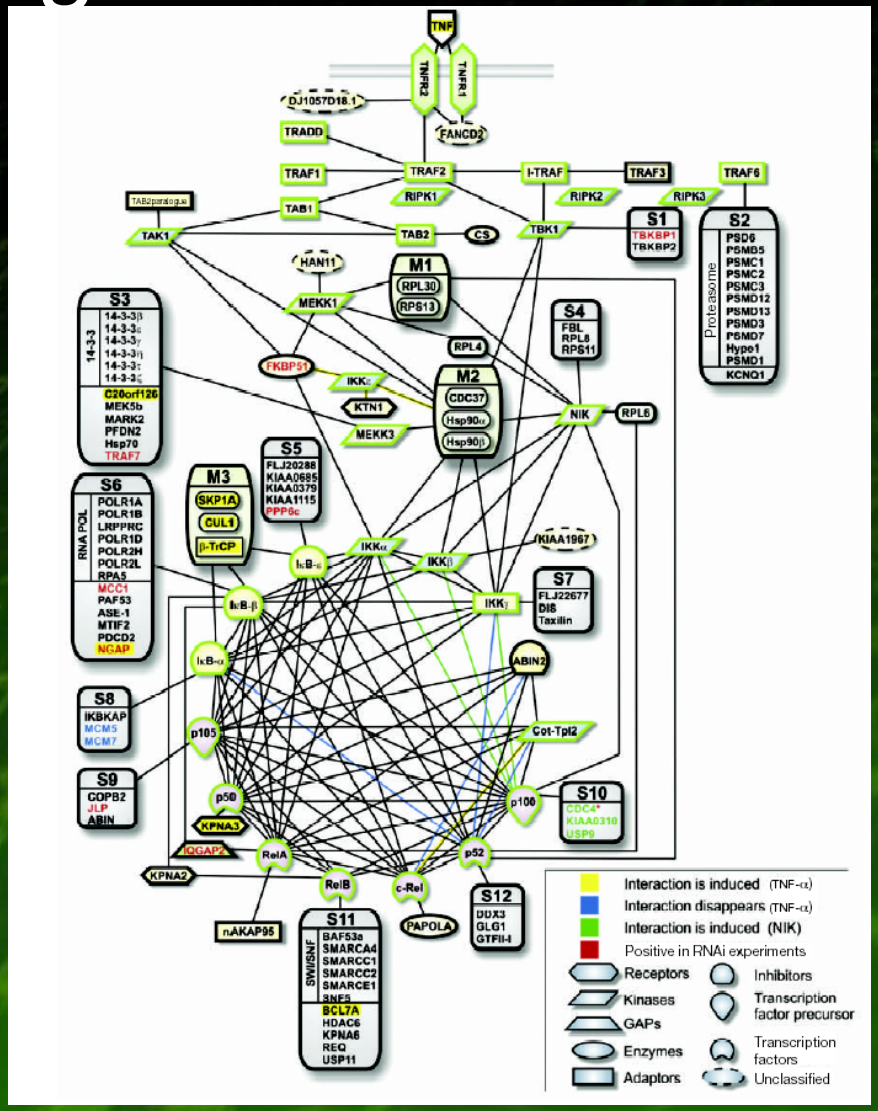
Before optimization: 3, 40

After optimization: 2, 091

Current Average Shortest Path 2, 330

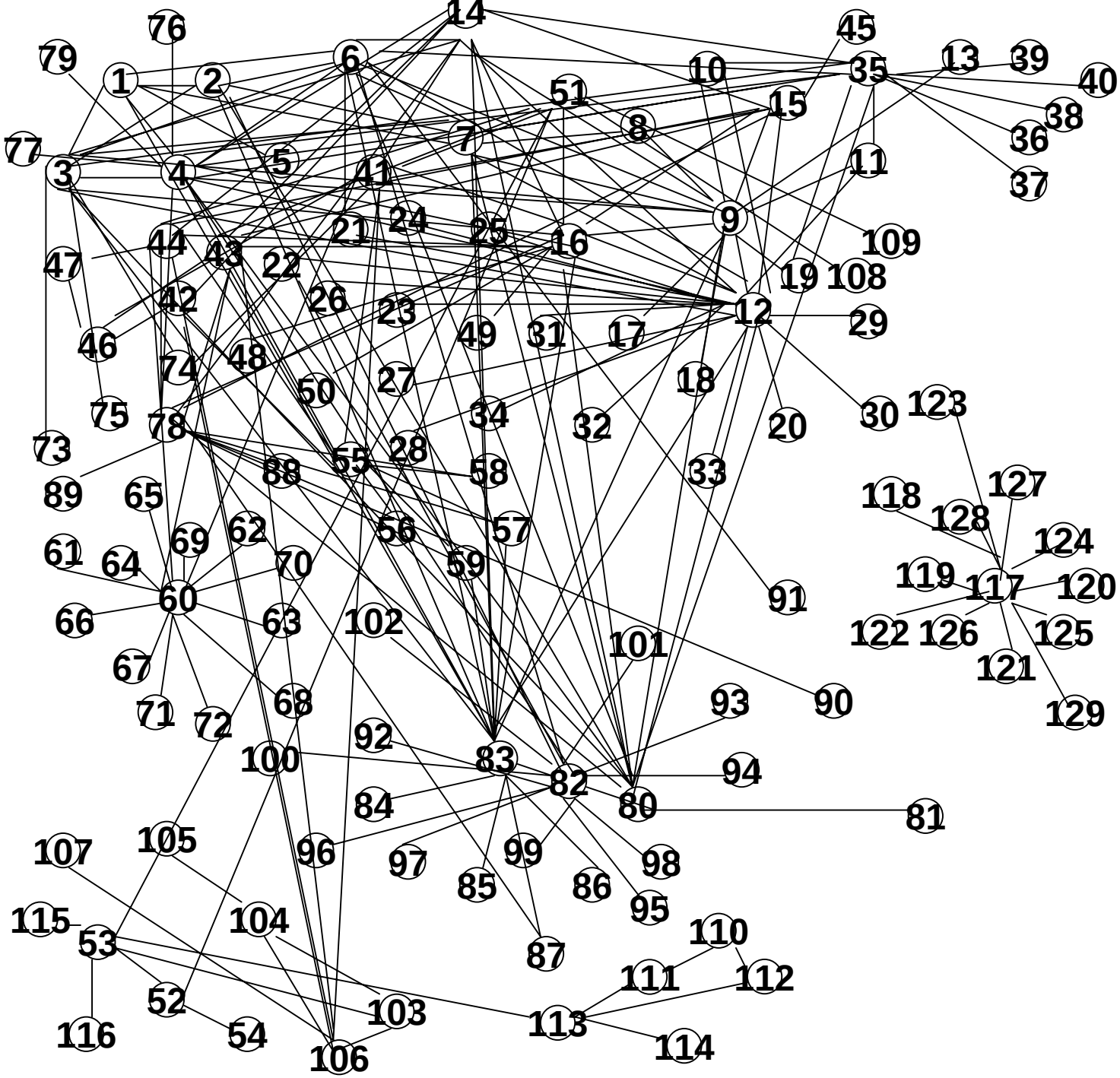


Analysis of the human TNF-α/NF-κB signal transduction pathway network (1)

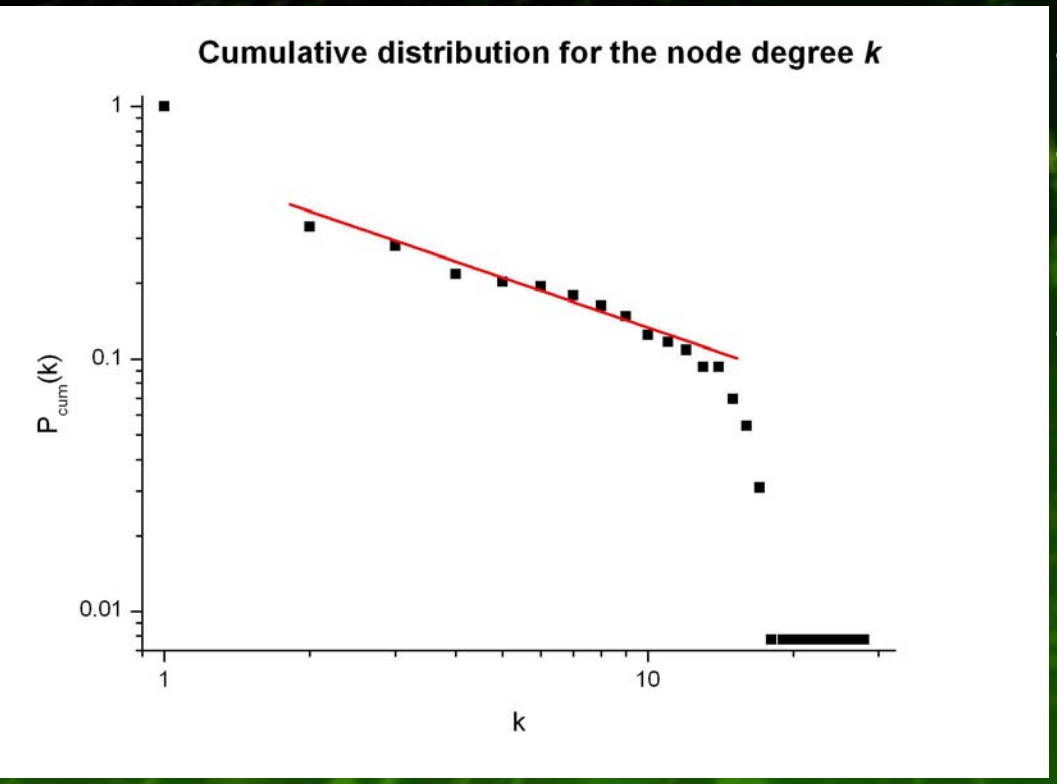


- nodes – proteins
- links – protein interactions
- methods:
  - degree distribution
  - knocking out single nodes and looking at how that affects the number of components and the average shortest path
- T. Bouwmeester et al., Nature Cell Biol. **6,** 97-105 (2004).

In collaboration with the molecular biology group (Danijel)

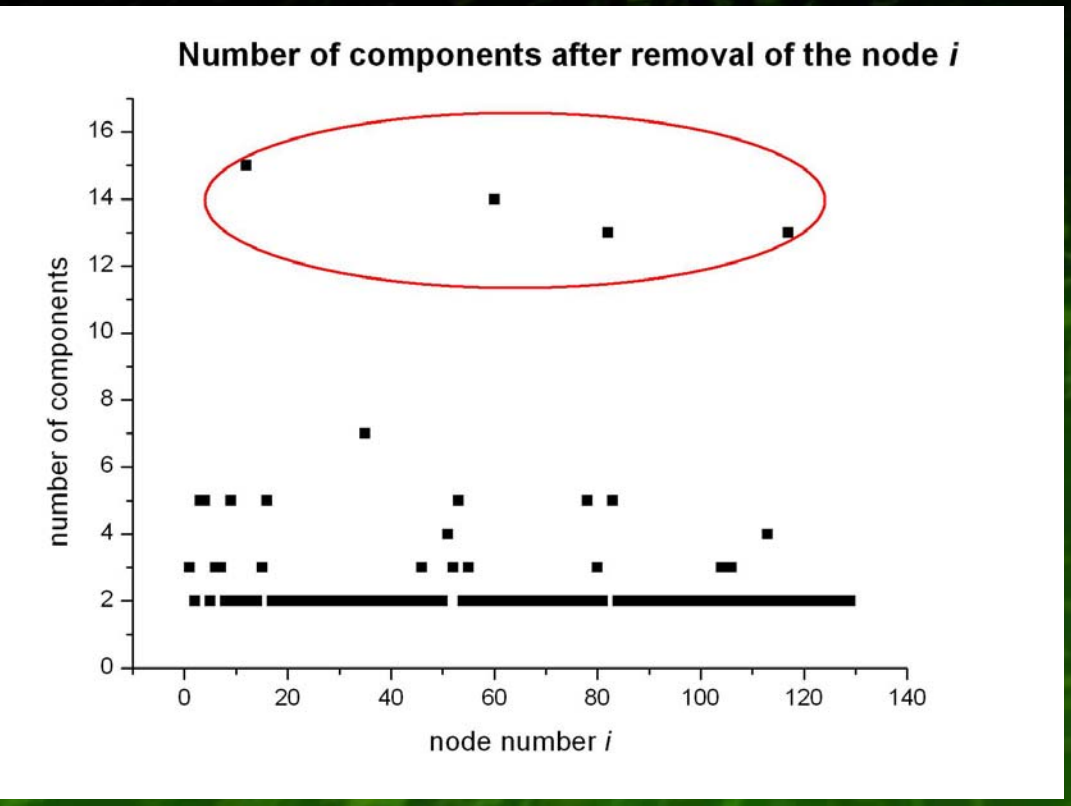

14
76
45
79
6
10
35
13
39
40
1
2
51
15
38
8
36
77
7
5
11
3
4
41
37
9
24
21
25
16
109
44
43
22
47
19
108
42
26
23
12
29
46
49
31
17
74
48
27
18
50
123
75
78
34
32
20
30
73
55
28
58
33
89
88
65
118
127
128
62
56
57
124
69
61
64
70
59
119
117
120
60
91
66
63
102
122
126
125
101
121
67
68
93
90
129
71
72
92
100
83
94
82
80
84
81
105
107
96
97
99
98
85
86
95
115
104
110
53
87
111
112
52
103
113
116
54
114
106

Analysis of the human TNF-α/NF-κB signal transduction pathway network (2)



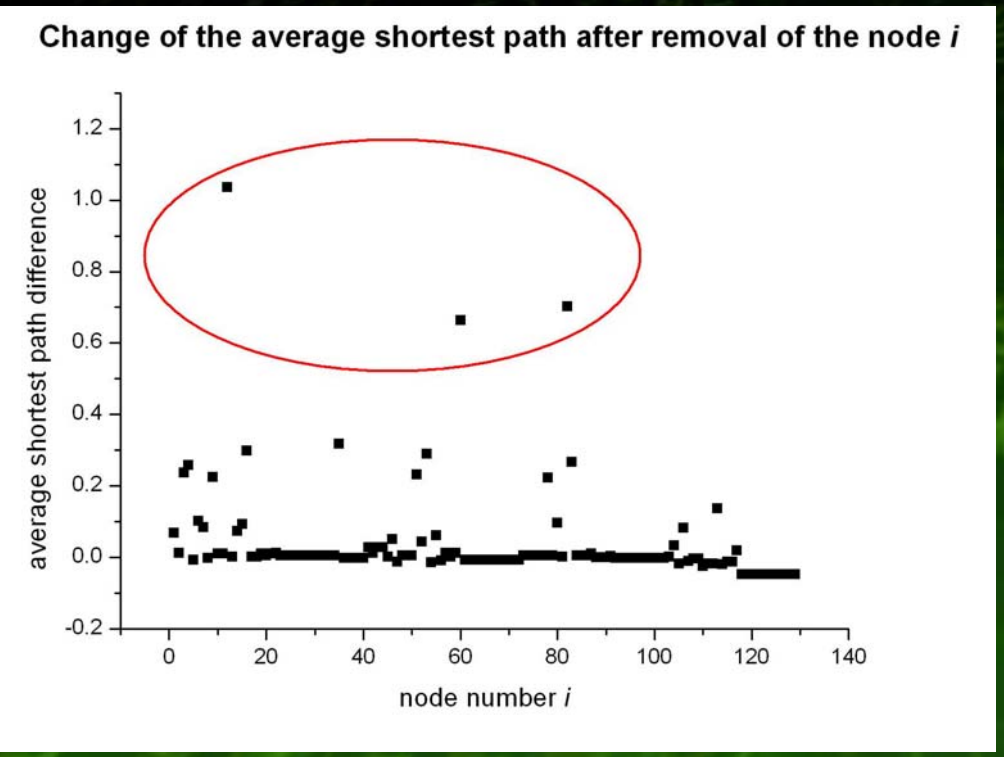
- cumulative distribution
- small network – limited statistics
- straight line in log k - log $P_{cum}(k)$ plot
  - indication that the network is scale-free
  - real network
  - sharp cutoff in the tail due to limited number of nodes

Analysis of the human TNF-α/NF-κB signal transduction pathway network (3)



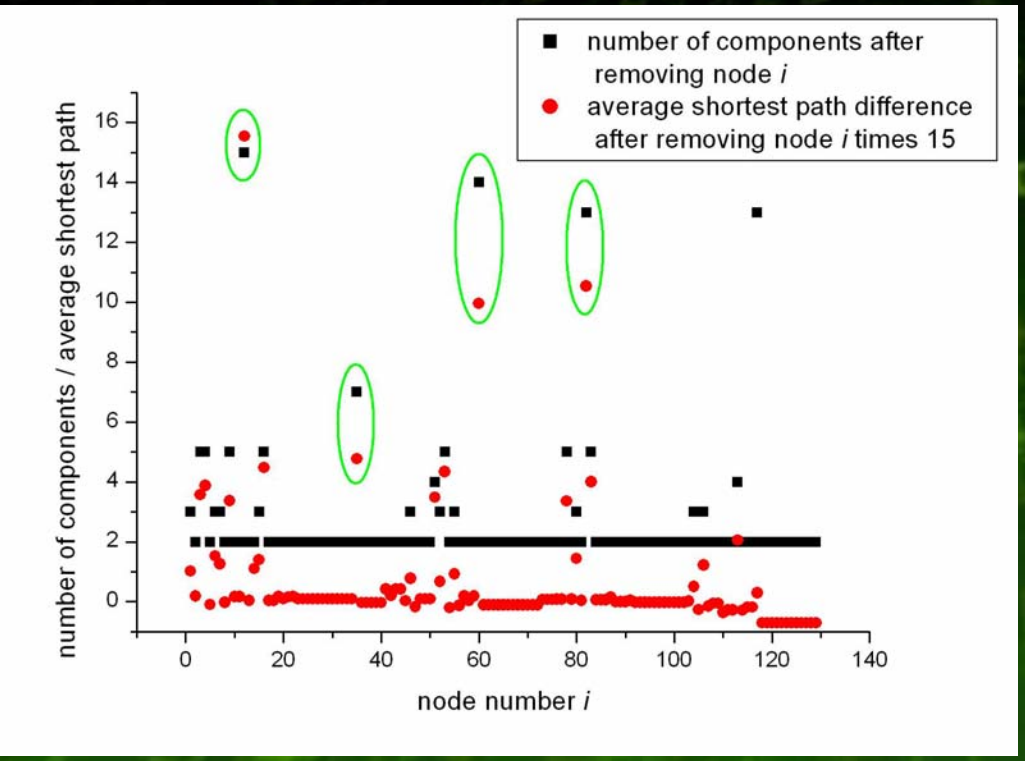
- sensitive to only certain nodes being removed – terorist attacks
- insensitive to removal of a large majority of nodes – malfunctions

Analysis of the human TNF-α/NF-κB signal transduction pathway network (4)



- the average shortest path also increases with removal of certain nodes

Analysis of the human TNF-α/NF-κB signal transduction pathway network (5)



- number of components and average shortest path after removing a node combined
- nodes that produce large difference in number of components also increase the average shortest path the most and vice versa

Our achievements

- all numerical simulations performed by the software developed by the team members
- all graphical simulations made by the team members
- real biological network completely analyzed by the software developed during the project

Hacker Team