Network of networks: modeling modularity of real-world

Marija Mitrović

Scientific Computing Laboratory
Institute of Physics, Belgrade, Serbia

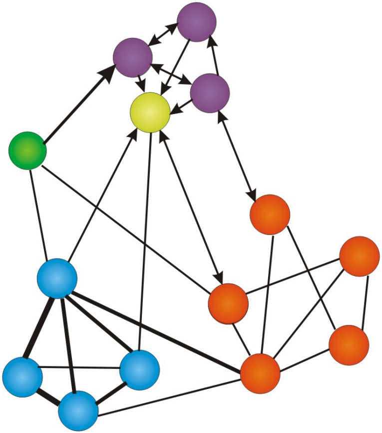
Seminar: Institute Jožef Stefan

Outline

Why are networks interesting?

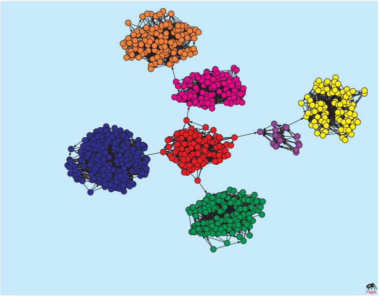
- Complex networks - natural models for a variety of systems
- Exploration of new phenomena on networks

What is a network?



- Network-set of vertices and edges
- Networks:
  - undirected
  - directed
  - unweighted
  - weighted
- Network in physics - graph in mathematics

Multiscale structure



- New ways for classifying and modeling networks
- Communities or modules on networks
- Models of networks with multiscale structure

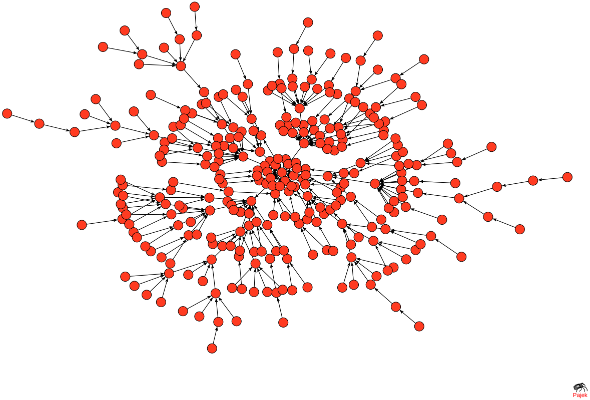
Search for subgraphs in networks

- Community detection methods:
  - Maximum-cut-minima flow
  - Maximization of modularity
- Maximum likelihood method:
  - Fitting mixture model to observed data using expectation-maximization algorithm
  - Result split of network in substructures (modules)
  - Generalization of method for finding weighted subgraphs

Network of networks

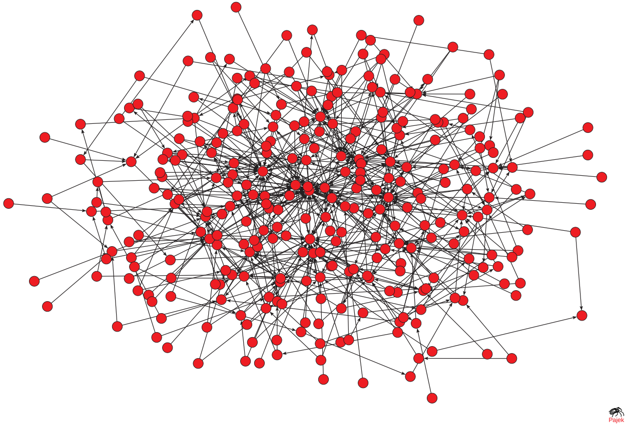
- Models of scale free networks-power law
- Model of World Wide Web-power law, clustered (B.Tadić, Physica A 293,(2001))
- Model of multiscale networks - power law, clustered, multisclae structure
- Growing rules:
  - At every time step new node i and M new links are added
  - With probability P_o new group is started
  - With probability α new node is attached to node k, which is chosen with probability p_{in} among nodes with sam group index as i
  - With probability $1-\alpha$ new link attaches node k with node n chosen among all existing nodes with probability p_{out}

Scale free network

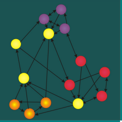


Parameters: $\alpha = 1$, $M = 1$, $P_o = 0$, $N = 300$

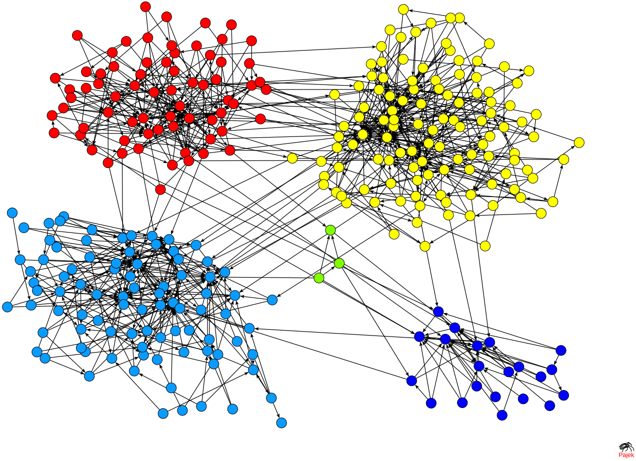
Model of World Wide Web



Parameters: $\alpha = 0.85$, $M = 2$, $P_o = 0$, $N = 300$

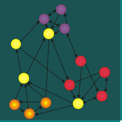


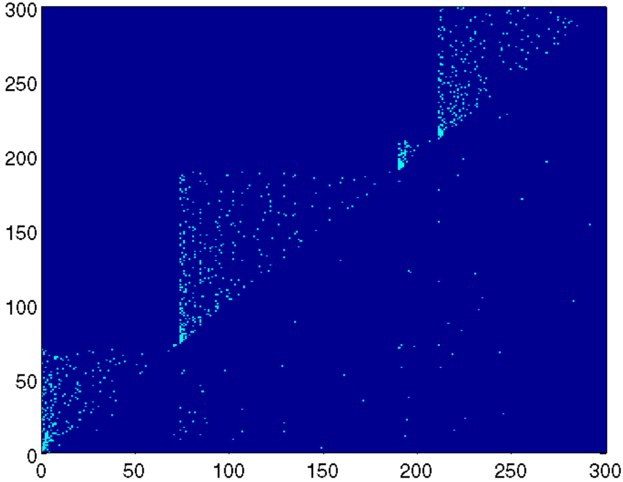
Model of multiscale network



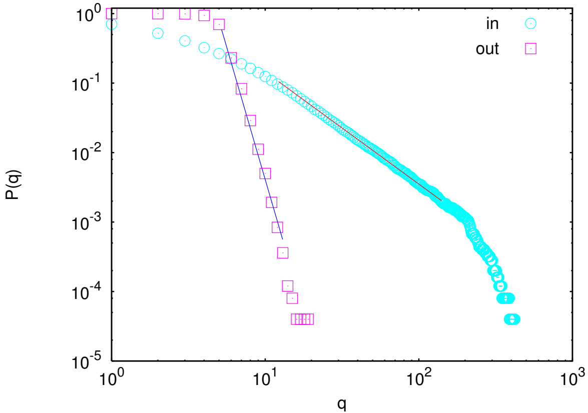
Parameters: $\alpha = 0.9$, $M = 3$, $P_o = 0.015$, $N = 300$

Adjacency matrix of multiscale network





Degree distribution



Parameters: $\alpha = 0.9$, $M = 3$, $G = 6$, $N = 25000$, $\tau_{in} = 2.61$, $\tau_{out} = 8.6$

Mixture models and likelihood

- Mixture models technique is well known in statistical data analysis
- "Probability" allows us to predict unknown outcomes based on known parameters - "likelihood" allows us to estimate unknown parameters based on known outcomes.
- Maximum likelihood estimation is statistical method used to calculate the best way of fitting a mathematical model to some data
- Algorithms for maximization of likelihood: k-means, expectation-maximization algorithm

Maximization likelihood method

- Maximization likelihood method (MLM) suggested by Newman (Newman, M.E.J and Leicht, E.A., PNAS 104, 9564 (2007))
- Network with N vertices is represented with adjacency matrix A
- Network can bi split into c groups, group memberships g_i are hidden data
- Model parameters:
  - θ_{ri}-probability that vertex from group r connects node i
  - π_r-probability that randomly chosen node falls in group r
  - The normalization conditions:

$$\sum_r \pi_r = 1, \quad \sum_i \theta_{ri} = 1 \ , \tag{1}$$

Maximum likelihood method

- Maximization of likelihood $Pr(A, g|\pi, \theta)$ with respect to π and θ in order to find g_i
- Factorization rule

$$Pr(A, g|\pi, \theta) = Pr(A|g, \pi, \theta)Pr(g|\pi, \theta), \tag{2}$$

- Likelihoods

$$Pr(A|g, \pi, \theta) = \prod_{ij} \theta_{g_i j}^{A_{ij}}, \quad Pr(g|\pi, \theta) = \prod_i \pi_{g_i}. \tag{3}$$

- Likelihood for a network

$$Pr(A, g|\pi, \theta) = \prod_i \pi_{g_i} \prod_j \theta_{g_i, j}^{A_{ij}}. \tag{4}$$

Maximum likelihood method

- Numbers g_i are unknown, value of log-likelihood is unknown
- Expected value of log-likelihood

$$\overline{L} = \sum_{ir} q_{ir}[ln\pi_r + \sum_{j} A_{ij} ln\theta_{g_i,j}] \tag{5}$$

- q_{ir} is probability that node i belongs to group r

$$q_{ir} = \frac{\pi_r \prod_j \theta_{rj}^{A_{ij}}}{\sum_s \pi_s \prod_j \theta_{sj}^{A_{ij}}} \tag{6}$$

Maximum likelihood method

- Result of maximization of $\overline{L}$

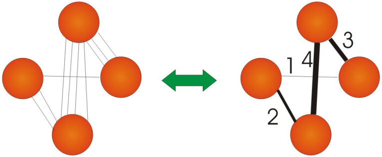
$$\pi_r = \frac{\sum_i q_{ir}}{n}, \qquad \theta_{ri} = \frac{\sum_j A_{ji} q_{jr}}{\sum_j q_{out}(j) q_{jr}} \tag{7}$$

- Expectation-maximization algorithm:
  - *expectation* step - calculating q_{ir}
  - *maximization* step - calculatingπ_i and θ_{ri}

Implementation of algorithm

- Iterative algorithm
- Initialization of parameters π and θ:
  - $\pi_i = \frac{1}{c}$ and $\theta_{ri} = \frac{1}{N}$-trivial fixed point
  - perturbed randomly a small distance from fixed point
- Calculate probabilities q_{ir}
- We stop when algorithm converges to local maxima of likelihood

MLM for multigraphs



- Weight of link between two nodes can be seen as multiple links between them
- Weighted subgraphs can be seen as set of vertices connected with the strongest links

MLM for multigraphs

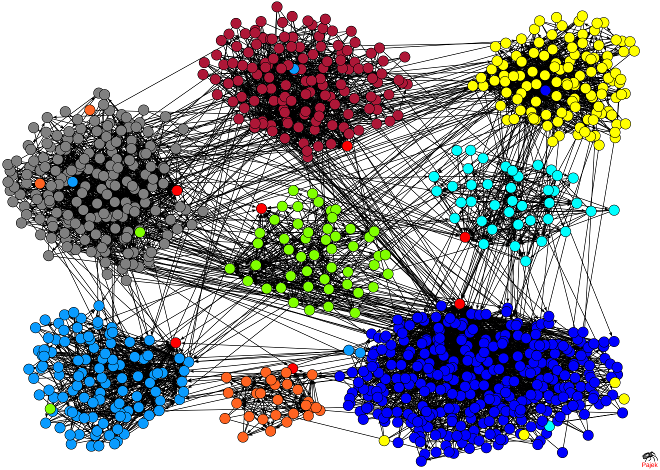
- W_{ij} is number of multiple links between node i and j
- Formulas for numbers q, π and θ

$$q_{ir} = \frac{\pi_r \prod_j \theta_{rj}^{W_{ij}}}{\sum_s \pi_s \prod_j \theta_{sj}^{W_{ij}}}, \tag{8}$$

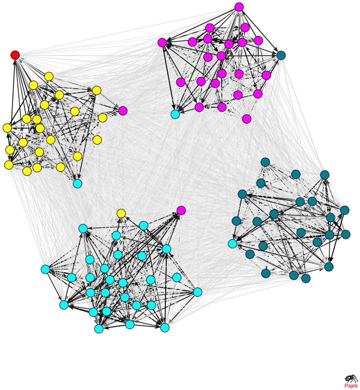
$$\pi_r = \frac{\sum_i q_{ir}}{n}, \qquad \theta_{ri} = \frac{\sum_j W_{ji} q_{jr}}{\sum_j s_j q_{jr}}, \tag{9}$$

- s_j is a strenght of node j

Results-multiscale network



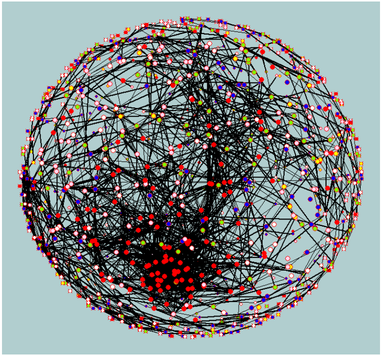
Results-weighted random graph



- Unweighted Erdos-Renyi(ER) model is homogenous
- Multiple links in ER model - multigraph

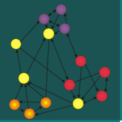
M.Mitrović and B.Tadić, LNCS, (2008)

Results-yeast gene expressions network

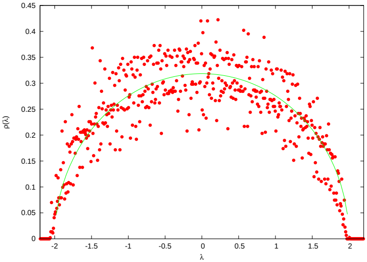


- Network of yeast gene expressions
- Weights of links appear through the correlation coefficient of the gene expressions

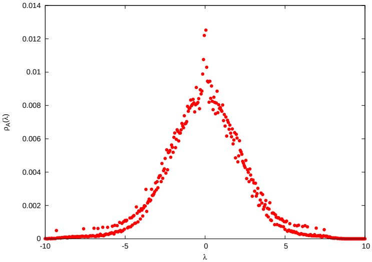
Živković, J., Tadić, B., Wick, N., Thurner, S., European Physical Journal B, 255 (2006)



Spectral density of adjacency matrix

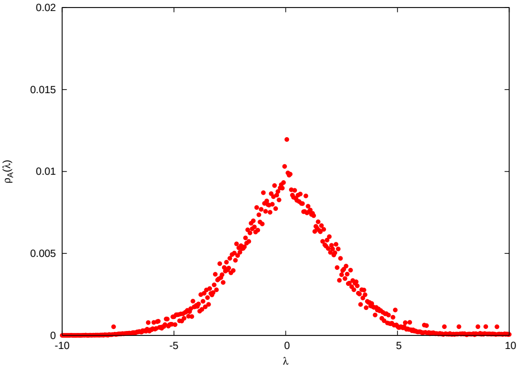


Erdos-Renyi graph $N = 500$, $p = 0.5$, $\Delta\lambda = 0.1$



Scale-free network $N = 500$, $M = 3$, $\Delta\lambda = 0.1$

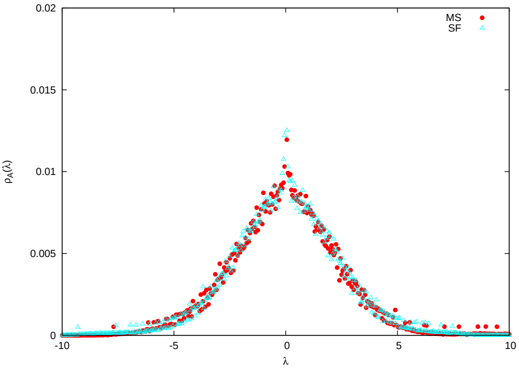
Spectral density of adjacency matrix



Multiscale network $N = 500, \alpha = 0.9, M = 3$, $G = 5$, $\Delta\lambda = 0.1$

Spectral density of adjacency matrix

Marija Mitrović



Conclusion

- Description of model of multiscale networks
- Maximum likelihood method
- Generalization of MLM for multigraphs
- Spectral density of adjacency matrix

Future work

- Random walks on networks can reveal community structure
- Synchronization processes on complex networks
- Spectra of Laplacian matrix of multiscale networks

Acknowledgments

- This work was done in collaboration with professor Bosiljka Tadić