

Studiu

Modele matematice și metode computaționale pentru analiza rețelelor

efectuat în cadrul proiectului *Abordarea bioeconomică a
agenților antimicrobieni – utilizare și rezistență*

(cod - PN-III-P1-1.2-PCCDI-2017-0361).

Colectiv de redacție:

Mihai Chiș: sumarizare concepte de bază în analiza spectrală (sect. 3)

Radu Moleriu: sumarizare rezultate teoretice în analiza spectrală (sect. 3)

Alina Cărunta: instrumente software pentru analiza rețelelor (sect. 4)

Daniela Zaharie: sumarizare algoritmi pentru detecția comunităților (sect. 1, sect. 2)

Coordonator: Daniela Zaharie

Membri: Mihai Chiș, Radu Moleriu, Alina Cărunta

Data finalizării: 10.12.2018

Versiunea: 1.1

Acknowledgements

Activities under this work were carried out in the *Research Laboratory Complex "Horia Cernescu"* - financed by project *"A bio-economical approach of the antimicrobial agents - use and resistance"*, in the frame of contract PCCDI 7/19.03.2018, code: PN-III P1-1.2-FPRD-2017.

1. Introducere

Scopul acestui studiu este de a prezenta concepte, modele și metode specifice de analiză a rețelelor biologice cu accent pe tehnicile și instrumentele software care pot fi utilizate în analiza rezistenței antimicrobiene. Rețelele de interacțiuni, modelate matematic prin grafuri, au devenit, în ultimii ani, instrumente utile de modelare în biologie (Charitou et al., 2016). În rețelele biologice nodurile pot reprezenta proteine, gene, metaboliți, fenotipuri sau chiar boli, iar conexiunile descriu interacțiuni sau relații între aceste entități.

Analiza rețelelor complexe s-a dezvoltat semnificativ în ultimii ani propunându-se o serie de algoritmi de extragere de informații din rețele și implementându-se o serie de instrumente software (Barabasi, 2016). Scopul analizei rețelelor este de a extrage informații privind structura și proprietățile unei rețele pornind de la descrierea primară a acesteia, specificată prin setul de noduri și relațiile dintre acestea. Din punct de vedere matematic o rețea este un graf iar scopul analizei este de a identifica structuri sau pattern-uri prezente în graf. În cazul în care rețeaua este construită pornind de la un set de date, asociate nodurilor, iar relațiile dintre noduri sunt stabilite pe baza similarității dintre date este de interes să se identifice direcțiile de-a lungul cărora datele variază semnificativ cu scopul de a determina grupuri de date suficient de similare între ele, denumite comunități. Explorarea acestei variabilități se poate realiza utilizând analiza spectrală care se bazează pe determinarea vectorilor proprii (care indică direcțiile principale de variabilitate) și a valorilor proprii (care oferă informații privind măsura variabilității). Teoria spectrală reprezintă un instrument algebric frecvent utilizat în obținerea de informații privind structura unei rețele.

2. Algoritmi pentru determinarea comunităților

Într-o rețea de interacțiuni o comunitate este un subset de noduri puternic intercorelate. În varianta cea mai simplă, intercorelarea se exprimă prin numărul de vecini ai unui nod dat (numărul de noduri cu care acesta este conectat). În cazul general intercorelarea poate fi interpretată ca o măsură de similaritate, iar rețeaua este descrisă prin intermediul unei matrici de similaritate.

Analiza comunităților în rețelele biologice permite identificarea unor grupuri funcționale în rețelele metabolice, analiza interacțiunilor între medicamente, analiza împrăștierei epidemiilor etc. (Charitou, 2016).

Identificarea subseturilor de noduri care formează o comunitate poate fi interpretată ca o problema de grupare (clustering) care permite partiționarea nodurilor în grupuri care au proprietatea că nodurile din același grup sunt similare între ele iar nodurile din grupuri diferite sunt disimilare. Algoritmii de identificare a comunităților sunt similari algoritmilor ierarhici de clustering (Barabasi, 2016) doi dintre cei mai cunoscuți algoritmi fiind:

- *Algoritmul Ravasz* – este un algoritm aglomerativ (Ravasz, 2002) care pornește de la o configurație inițială în care fiecare nod face parte din propria comunitate după

care comunitățile suficient de “apropiate” între ele sunt reunite iterativ; apropierea dintre comunități se determină folosind una dintre strategiile clasice din clustering: *single-link* (se folosește similaritatea dintre cele mai similare două noduri ce aparține celor două comunități), *complete-link* (se folosește similaritatea dintre cele mai disimilare două noduri ce aparține celor două comunități), *average-link* (se folosește similaritatea medie dintre nodurile aparținând celor două comunități). Măsura utilizată pentru a exprima similaritatea dintre două noduri i și j este:

$S(i,j)=J(i,j)/(\min(k_i,k_j)+1-A(i,j))$ unde $J(i,j)$ reprezintă numărul de vecini comuni ai nodurilor i și j , k_i este gradul nodului i (numărul de noduri cu care este conectat direct) iar $A(i,j)$ este element al matricii de adiacență (este 1 dacă există conexiune directă între nodul i și nodul j și 0 în caz contrar)

- *Algoritmul Girvan-Newman* – este un algoritm diviziv (se consideră că inițial toate nodurile fac parte din aceeași comunitate care este succesiv divizată în subcomunități prin eliminarea unor conexiuni); elementul cheie al algoritmului îl reprezintă criteriul de selecție a conexiunii (i,j) care trebuie eliminată la fiecare etapă. În cazul algoritmului Girvan-Newman ([Girvan and Newman, 2002](#)) se folosește măsura de centralitate definită astfel: $\text{centralitate}(i,j) = \text{numărul celor mai scurte căi între oricare două noduri din rețea care trec prin muchia } (i,j)$. Algoritmul constă în repetarea următorilor pași:
 - *Pas 1:* se calculează măsura de centralitate pentru fiecare pereche de noduri (i,j) între care există conexiune
 - *Pas 2:* se elimină conexiunea cu cea mai mare valoare a măsurii de centralitate.

Configurația rețelei de la fiecare etapă furnizează o variantă de separare a nodurilor în comunități, la fel cum secționarea la diferite nivele a unei dendrograme construite folosind un algoritm ierarhic de grupare conduce la diferite partiții.

O direcție recentă de analiză a comunităților este reprezentată de formularea problemei ca una de optimizare în care se folosește o măsură de modularitate drept criteriu de maximizare. Pentru un subset C de noduri, modularitatea se definește prin

$$Mod(C) = \frac{1}{2L} \sum_{(i,j) \in C} (A_{ij} - P_{ij})$$

unde L este numărul de conexiuni din rețea, A_{ij} este elementul corespunzător din matricea de adiacență, iar $P_{ij} = k_i k_j / L$ (reprezintă numărul așteptat de conexiuni între i și j), k_i fiind gradul nodului i . Dacă $Mod(C)$ este pozitiv atunci se consideră că C este o comunitate potențială, iar dacă este negativ atunci cel mai probabil, C nu reprezintă o comunitate. Suma modularității asociată fiecărui grup dintr-o partiție reflectă calitatea partiției (a structurării pe comunități) și poate fi utilizată pentru a decide care dintre partițiile generate printr-o abordare ierarhică (aglomerativă sau divizivă) reflectă cel mai bine structura de comunități.

Rețelele reale, inclusiv cele biologice pot conține comunități care se suprapun. În cazul acesta există algoritmi specifici de tip *clique percolation* ([Palla, 2005](#)) (comunitățile sunt

construite pornind de la k -clici, adică subrețele cu k noduri total interconectate) respectiv *link clustering* (Shi, 2013) (ideea de bază este de a grupa muchii în loc de noduri). Tehnica bazată pe clici este implementată în CFinder (<http://www.cfinder.org/>).

3. Metode spectrale pentru analiza rețelelor

În contextul metodelor de grupare, metodele spectrale se particularizează prin faptul că realizează în mod implicit o selecție a componentelor care au rol important în estimarea similarității. Analiza spectrală studiază proprietățile unei rețele (interpretată ca fiind un graf în sens matematic) pe baza spectrului și a vectorilor proprii asociați matricii de similaritate și a matricii Laplace a unui graf. Ambele tipuri de matrice sunt frecvent studiate folosind metode din domeniul algebrei liniare.

3.1. Concepte de bază

Din punct de vedere matematic un graf G este o pereche (V, E) unde V reprezintă mulțimea nodurilor iar E mulțimea conexiunilor (denumite muchii în cazul grafurilor neorientate, respectiv arce în cazul grafurilor orientate). Unui graf cu n noduri $\{x_1, x_2, \dots, x_n\}$, $n \in \mathbb{N}^*$ i se asociază următoarele matrice:

- i. *matricea de adiacență* $A \in M(n; \mathbb{R})$, $A = (a_{ij})$, unde

$$a_{ij} = \begin{cases} 1, & \text{dacă nodurile } x_i, x_j \text{ sunt conectate} \\ 0, & \text{în rest} \end{cases};$$
- ii. *matricea diagonală* care conține pe diagonală principală gradele nodurilor grafului, notată $D \in M(n; \mathbb{R})$, $D = \text{diag}\{d_1, d_2, \dots, d_n\}$ cu

$$d_i = \sum_j a_{ij};$$
- iii. *matricea Laplace*, $L \in M(n; \mathbb{R})$, $L = D - A$;
- iv. *matricea normală sau matricea probabilităților de conexiune* $P \in M(n; \mathbb{R})$, $P = D^{-1}A$, unde D^{-1} este inversa matricii D în condițiile în care se consideră că:

$$d_i = 0 \Rightarrow (D^{-1})_{i,i} = 0.$$

3.2. Rezultate teoretice

În cazul unei rețele modelate printr-un graf neorientat matricea de adiacență este simetrică și prin urmare are valori proprii reale. Pe baza teoremei de descompunere spectrală matricea de adiacență poate fi aproximată printr-o sumă de forma:

$$A_k = \sum_{i=1}^k \alpha_i v_i v_i^T,$$

unde $k \leq n$ și v_i este vectorul propriu normal asociat valorii proprii α_i . Rezultă că matricea A_k poate surprinde variabilitatea datelor din sistem. De asemenea, utilizând valoarea proprie cea mai mare a matricii A , se poate estima o margine superioară pentru volumul grafului.

Dacă matricea de adiacență este simetrică atunci și matricea Laplace este simetrică. În plus, este pozitiv semidefinită și valorile proprii satisfac $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$.

În lucrarea (Luxburg, 2007) (în propozițiile 1 și 2) se afirmă că: numărul componentelor conexe ale unui graf este egal cu multiplicitatea algebrică a valorii proprii 0 (zero) și fiecare componentă conexă corespunde unui vector propriu asociat acesteia.

În lucrarea (Chung) se prezintă legătura dintre spectrul matricei Laplace L și numărul nodurilor unui graf:

$$\sum_i \lambda_i \leq n ,$$

cu egalitate doar dacă graful G este complet. Dacă G nu este complet, atunci $\lambda_1 < 1$ și valorile proprii sunt mărginite de gradele nodurilor grafului.

În literatură, există două forme normalizate ale matricei Laplace:

$$L_{sym} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$$

și

$$L_{rw} = I - D^{-1}A .$$

Spectrele celor două matrice coincid și dacă v este un vector propriu asociat matricei L_{rw} atunci $D^{\frac{1}{2}}v$ este un vector propriu asociat matricei L_{sym} .

Pe baza matricelor normalizate se introduc algoritmi spectrali de grupare: *algoritmul spectral introdus de Shi și Malik (2000)* (folosește matricea L_{rw}) și *algoritmul spectral introdus de Ng Jordan și Weiss (2002)* (folosește matricea L_{sym}).

Aceste tehnici sunt similare cu algoritmul spectral nenormalizat introdus în lucrarea (Luxburg, 2007) și face referire la reprezentarea abstractă a datelor. Gruparea datelor se realizează în raport cu similaritatea dintre acestea.

Tehnicile utilizate sunt *RatioCut* și *Ncut* și se bazează pe volumul claselor și pe numărul de conexiuni din interiorul acestora.

Gruparea datelor utilizează primii k - vectori proprii a matricei Laplace iar viteza de convergență a algoritmului depinde de decalajul dintre valorile proprii:

$$\gamma_k = |\lambda_k - \lambda_{k-1}| .$$

Alegerea valorii k se face astfel: valorile proprii $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{k-1}$ trebuie să fie mici, iar valoarea proprie λ_k trebuie să fie mare.

În lucrările (Mahoney et al., 2012) și (Seary&Richards, 2003) se studiază a doua valoare proprie a matricei Laplace, λ_1 și a vectorilor proprii asociați. Aceștia determină parametrii grafului, prezența conexiunilor rare și comportamentul aleator al conexiunilor.

Matricea normală sau *matricea probabilităților de tranziție* ((Chung, 1997), (Luxburg, 2007)) se notează cu $P \in M(n; \mathbb{R})$ și $P(x_i, x_j) = p_{ij} = \mathbb{P}\{x_{i+1}|x_i\}$. Matricea P este stohastică, cu spectrul mărginit în intervalul $(-1,1]$, cu 1 valoare proprie. P se poate interpreta ca matrice de tranziție a unui lanț Markov cu spațiul stărilor format din nodurile grafului, iar matricea A ca un tabel de contingență.

În ipoteza unui graf ireductibil și aperiodic este asigurată existența unei distribuții staționare. Problema asociată constă în determinarea unui moment de timp t astfel încât pornind de la o distribuție inițială a grafului, matricea P^t să furnizeze o aproximare eficientă a distribuției staționare. O alegere uzuală a probabilităților de tranziție este

$$P(x_i, x_j) = \frac{1}{d_i} ,$$

Modele matematice și modele computaționale pentru analiza rețelelor

dacă x_i, x_j sunt conectate, iar dacă nodurile nu sunt conectate se ia valoarea 0 (zero). Teorema 1.18 din lucrarea (Chung, 1997) impune o margine timpului, $t > 0$, folosind valorile proprii ale matricii Laplace.

Legătura dintre matricea probabilităților de tranziție și matricea Laplace a grafului este :

$$P = D^{-\frac{1}{2}}(I - 1)P^{\frac{1}{2}},$$

iar distribuția staționară în acest caz este

$$\frac{d_i}{\text{vol}(G)},$$

unde d_i este gradul nodului și $\text{vol}(G)$ este volumul grafului.

Lucrarea (Wang et al, 2011) consideră matricea de adiacență $A = (a_{ij})$, dată prin:

$$a_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right),$$

unde σ este un parametru de scalare și prezintă etapele unui algoritm de grupare a datelor.

În lucrarea (Thorne, 2012) se prezintă o metodă de codificare a proteinelor și a interacțiunii dintre acestea pe baza distanțelor definite de spectrul matricii Laplace asociate grafului. Se face o comparație între diferite grafuri de evoluție și se estimează parametrii acestora.

Lucrarea (Luxburg, 2007) prezintă diferite metode de construcție a unui graf pornind de la o similaritățile sau disimilaritățile dintre datele unui set. Metodele de construcție sunt :

- vecinătatea de ordin ε , $\varepsilon > 0$;
- metoda celor mai aproape „ k ” vecini ;
- conectivitate totală bazată pe funcții nucleu de tip gaussian.

4. Instrumente software pentru analiza rețelelor biologice

Complexitatea și volumul mare de date care trebuie procesate într-un anumit timp a condus la crearea unor instrumente necesare analizei a acestora, dat fiind faptul că există posibilitatea de a găsi anumite legături între acestea. Datorită volumului mare de date, identificarea acestor legături poate fi dificilă, iar o mai bună vizualizare a acestora permite identificarea legăturilor precum și a nodurilor aferente. Procesarea acestor date sub formă de rețele facilitează analiza datelor, identificarea legăturilor dintre noduri, care conțin informația, precum și vizualizarea datelor.

Inițial, analiza datelor cu ajutorul rețelelor a fost folosită în câteva domenii sociale precum și cele biologice (moleculare, celulare). Ulterior, utilizarea acestor rețele s-a extins și în alte domenii pentru rezolvarea unor probleme complexe. Implementarea acestor rețele a fost posibilă datorită instrumentelor dezvoltate folosind tehnologiile existente.

De asemenea, diversitatea tipurilor de rețele variază în funcție de domeniu, acestea fiind folosite pentru a descrie diferite sisteme, începând cu Internetul până la rețele sociale, inclusiv rețele genetice pentru determinarea sistemelor biologice. Albert Laszlo Barabasi

descrie câteva exemple din viața reală în cartea sa intitulată „Network science” (Barabasi, 2016), aplicabilitatea acestor rețele fiind posibilă în domenii precum fizica, informatica, inginerie, economie, științe sociale și medicină.

Câteva dintre instrumentele destinate analizei datelor folosind rețele sunt descrise mai jos.

4.1. Cytoscape

Cytoscape este o platformă software de tip „open source” pentru vizualizarea rețelelor de interacțiuni moleculare și legături biologice și integrarea acestor rețele cu adnotări, profilurile de expresii de gene și alte date de stare. Deși Cytoscape (Shannon, 2003) a fost inițial destinat cercetării biologice, acum este o platformă generală pentru analiza și vizualizarea rețelor complexe. Nucleul platformei Cytoscape oferă un set de caracteristici de bază pentru integrarea, analiza și vizualizarea datelor. Pe lângă acestea sunt oferite caracteristici adiționale sub denumirea de „Apps” (cunoscute inițial ca „Plugins”). Câteva exemple de aplicații realizate folosind Cytoscape sunt: *MORO AutoAnnotate* (Kucera, 2016), „plugins” precum *clusterMaker* (Morris, 2011) (rețea biologică), *GIANT* (Cumbo, 2014) (rețea metabolică), *NeMo*. MORO analizează legătura dintre modularitatea și robustețea în rețele de semnalizare biologică de dimensiuni mari, folosind grafuri directe, rețele booleene și calculul paralel cu biblioteca OpenCL.

AutoAnnotate este un instrument care facilitează vizualizarea sumară a unei rețele prin identificarea grupurilor (clusterelor) și adnotarea lor cu o formă și etichetă. Acest tool este util pentru analizarea oricărui tip de rețea.

4.2. R

Un alt instrument care poate fi utilizat în analiza rețelelor este R, o platformă software dedicată în principal prelucrărilor statistice dar care conține o serie de pachete specifice pentru bioinformatică. Funcțiile din R pot fi utilizate direct sau prin intermediul interfeței *RStudio*. De asemenea, R oferă diverse funcții sub forma de pachete pentru analiza rețelelor atât în domeniul științelor sociale (de exemplu, în domeniul psihologiei) cât și în alte domenii, folosind pachete precum „*qgraph*” (<http://sachaepskamp.com/qgraph/>) și „*bootnet*” (Epskamp, 2017). Acest instrument permite și crearea de noi pachete pentru tipul de rețea dorit. R oferă și posibilitatea vizualizării interactive a rețelei într-un browser prin intermediul *RCyjs* (Shannon, 2018).

4.3. Alte instrumente software și modele matematice în analiza rezistenței la antibiotice

Vladimir Batagelj prezintă instrumente pentru analiza rețelor, iar câteva dintre acestea sunt menționate mai jos. Instrumente pentru analizarea rețelor complexe pentru

Modele matematice și modele computaționale pentru analiza rețelelor

interacțiuni moleculare pot fi considerate și *Pajek* (Batagelj, 1998), *Graphlet* și *daVinci*, care permit organizarea și vizualizarea datelor folosind rețele bi-dimensionale.

Pe de altă parte, în ultimii ani au fost dezvoltate o serie de modele matematice și computaționale pentru studiul transmiterii bolilor infecțioase (Lanzas, 2015), respectiv rezistența microbilor la antibiotice în (Knight, 2018), (Sun, 2017), (Miryala, 2018) și (Tej, 2017).

În articolul (Lanzas, 2015) este prezentat un model de sistem complex pentru epidemiologia veterinară în cazul transmiterii patogenilor (*E. coli* 0157) în rândul populației de vaci. Metodele de modelare cele mai folosite în epidemiologie sunt cele modulare, rețelele și modelele bazate pe agenți. Metoda folosită în acest caz este utilizarea rețelor bazate pe agenți și modelare modulară, transmiterea agenților patogeni fiind influențată de mai multe elemente interconectate..

Rezistența microbilor/virusilor la antibiotice este studiată în cazul infectării cu *E.coli* în (Knight, 2018) și (Miryala, 2018) precum și în tratamentul diferitelor tipuri de cancer în lucrările (Sun, 2017) și (Tej, 2017).

Problema rezistenței la antibiotice este prezentată de (Knight, 2018), iar scopul lucrării este de a determina dacă sursa de bacterii rezistente la antibiotice (ARB) provine din cadrul comunității sau spital. Soluția propusă este un model matematic de transmitere dinamică, inspirat din istoria bacteriilor rezistente la antibiotice, pentru estimarea numărului de bacterii rezistente la antibiotice (ARB) luate din ambele medii (comunitate și spital). De asemenea, diferite combinații de parametri sunt folosite pentru studiul rezistenței tulpinei *E. Coli* în Anglia. Elementul de noutate al acestei metode este analiza mai multor combinații de patogen/antibiotic. Rezultatul final al acestui studiu constă în elaborarea unui cadru cantitativ pentru a explora o gamă largă de parametri care cuprinde un set de scenarii posibile pentru dobândirea rezistenței la antibiotice în cadrul comunității sau a spitalului.

Rezistența la medicamente este studiată și în (Miryala, 2018) prin explorarea rezistenței la multiple medicamente în cazul tulpinei *E. Coli* 0157:H7. Soluția propusă este o rețea de interacțiuni a genelor, având o abordare bazată pe sisteme biologice, iar pentru vizualizarea și analiza rețelei s-a folosit Cytoscape. De asemenea, o abordare de tip cluster și topologic sunt folosite pentru a identifica genele care joacă un rol important în mecanismele de rezistență la medicamente. Această rețea este compusă din 29 de gene rezistente la antibiotice și 777 de interacțiuni în *E. Coli* 0157:H7. Datele au fost furnizate de baze de date de gene rezistente la antibiotice, mai exact de (<http://ardb.cbcb.umd.edu/>). Rezultatele obținute oferă o înțelegere detaliată a mecanismului de rezistență la multiple medicamente în cazul infectării cu *E. coli*.

Bibliografie:

- (Charitou, 2016) T. Charitou, K. Bryan, and D. J. Lynn, Using biological networks to integrate visualize and analyze genomics data, *Genet Sel Evol* (2016) 48:27
- (Chung, 1997) F. R. K. Chung, Lectures on Spectral Graph Theory, University of Pennsylvania, Philadelphia
- (Luxburg, 2007) U. von Luxburg, A Tutorial on Spectral Clustering, *Statistics and Computing*, 17 (4), 2007.
- (Mahoney et al., 2012) M. W. Mahoney, L. Orecchia, N. K. Vishnoi, A Local Spectral Method for Graphs: With Applications to Improving Graph Partitions and Exploring Data Graphs Locally, *Journal of Machine Learning Research* 13 (2012) 2339-2365
- (Seary&Richards, 2003) A. J. Seary , W. D. Richards, Spectral methods for analyzing and visualizing networks: an introduction, 209-228, *Dynamic Social Network Modeling and Analysis*, The National Academies Press, 2003 ;
- (Thorne, 2012) T. Thorne , M. P. H. Stumpf, Graph spectral analysis of protein interaction network evolution, *J. R. Soc. Interface* (2012), 9, 2653–2666 ;
- (Wang et al, 2011) H. Wang, j. Chen, k. Guo, A Genetic Spectral Clustering Algorithm, *Journal of Computational Information Systems* 7: 9 (2011) 3245-3252.
- (Barabasi, 2016) Barabasi, A. L. (2016). *Network Science*. Cambridge University Press.
- (Batagelj) Batagelj, V. Graph Theory and Network Analysis. Preluat pe Noiembrie 19, 2018
- (Batagelj, 1998) Batagelj, V., & Mrvar, A. (1998). Pajek—Program for large network analysis. *Connections*, 21, 47-57.
- (Cumbo et al., 2014) Cumbo, F., Paci, P., Santoni, D., Di Paola, L., & Giuliani, A. (2014). GIANT: a cytoscape plugin for modular networks. *PLoS One*, 9(10).
Cytoscape. (fără an). Preluat pe Noiembrie 19, 2018, <http://www.cytoscape.org/>
- (Epskamp, 2017) Epskamp, S., Borsboom, D., & Fried, E. I. (2017). Estimating psychological networks and their accuracy: a tutorial paper. *Behavior Research Methods*.
- (Girvan and Newman, 2002) M. Girvan, M.E.J. Newman, Community structure in social and biological networks, *PNAS* June 11, 2002 99 (12) 7821-7826
- (Kucera et al, 2016) Kucera, M., Isserlin, R., Arkhangorodsky, A., & Bader, G. D. (2016). AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Research*. doi:10.12688/f1000research.9090.1
- (Morris et al., 2011) Morris, J., Apeltsin, L., Newman, A., Baumbach, J., Wittkop, T., Su, G., . . . Ferrin, T. (2011). clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*, 12(1). doi:<https://doi.org/10.1186/1471-2105-12-436>
- (Palla, 2005) G. Palla (2005). "Uncovering the overlapping community structure of complex networks in nature and society". *Nature*. 435: 814–818
- (Ravasz, 2002) E. Ravasz, Hierarchical Organization of Modularity in Metabolic Networks, *Science* 297, 1551 (2002)

- ([Rivera et al, 2010](#)) Rivera, C., Vakil, R., & Bader, J. (2010). NeMo: Network Module identification in Cytoscape. *BMC Bioinformatics*, 11(1).
- ([Shannon et al, 2003](#)) Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11), 2498-2504. Preluat de pe <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC403769/>
- ([Shannon, 2018](#)) Shannon P (2018). *RCyjs: Display and manipulate graphs in cytoscape.js*
- ([Shi et al., 2013](#)) C. Shi, Y. Cai, D. Fu, Y. Dong, B. Wu, A link clustering based overlapping community detection algorithm, *Data&Knowledge Engineering* 87 (2013) 394-404
- ([Truong et al, 2016](#)) Truong, C. D., Tran, T. D., & Kwon, Y. K. (2016). MORO: a Cytoscape app for relationship analysis between modularity and robustness in large-scale biological networks. *Proceedings of the 27th International Conference on Genome Informatics: systems biology*. *BMC Systems Biology*. Preluat de pe <https://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-016-0363-3>