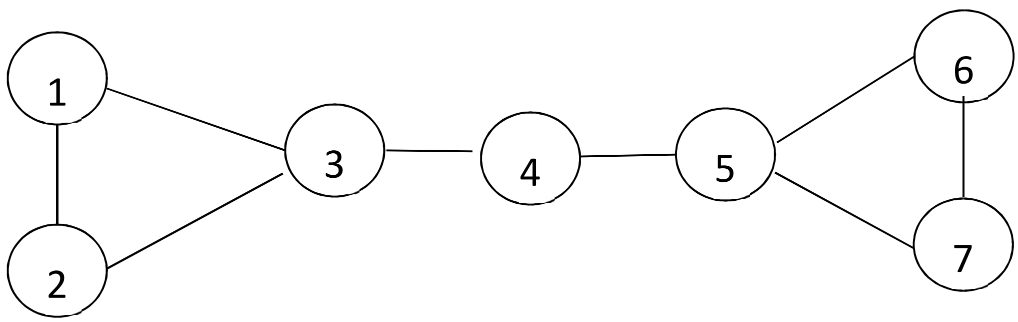
CONECTIVITATE ŞI INTERDEPENDENŢĂ ÎN SITEMELE ADAPTIVE COMPLEXE. ANALIZA REŢELELOR COMPLEXE

Studiu de caz

Reţeaua de drumuri dintr-o regiune este data în reprezentarea de mai jos, sub forma unui graf neorientat, cu următoarele şapte noduri şi opt muchii în componenţă:



Cerinţe:

a) Calculaţi gradul fiecărui nod al grafului (*"node degree"*)
b) Calculaţi şi interepretaţi numărul mediu de muchii incidente unui nod ("*average degree*")
c) Calculaţi densitatea grafului şi interpretaţi (*"graph density"*)
d) Calculaţi şi interepretaţi drumul mediu de lungime minimă (*"average path length"*)
e) Precizaţi care este diametrul *D* al reţelei (*"network diameter"*)
f) Calculaţi coeficienţii de clusterizare în jurul nodurilor *1 , 3* şi 4 (*"node clustering coefficient"*)
g) Calculaţi coeficientul mediu de clusterizare în reţea (*"average clustering coefficient"*)
h) Centralitatea unui nod - raportul dintre gradul nodului și numărul maxim posibil de conexiuni
i) Calculaţi centralitatea nodurilor, măsurată prin gradul nodurilor (*"degree centrality"*)
j) Calculaţi centralitatea nodurilor, măsurată prin apropierea nodului de alte noduri (*"closeness centrality"*)
k) Calculaţi şi interpretaţi indicatorul de modularitate al reţelei ("modularity")

Rezolvări:

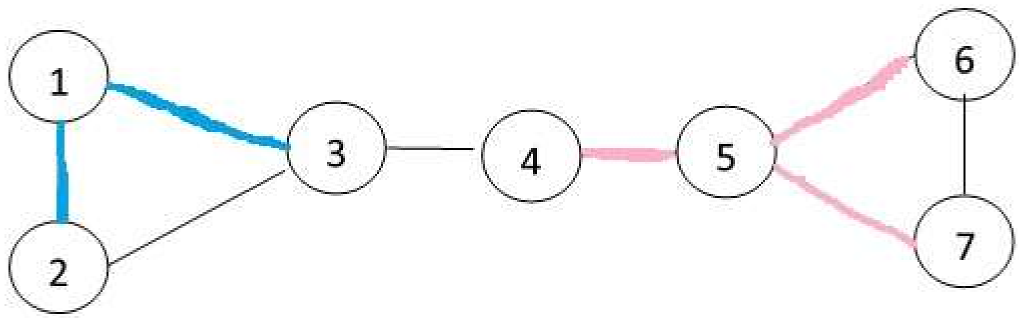
a) Gradul unui nod (*"node degree"*) este dat de numărul de muchii incidente respectivului nod.

- nodurile 1, 2, 4, 6 şi 7 au gradul 2 (număr de muchii incidente = 2);

$$\boldsymbol{d_1 = d_2 = d_4 = d_6 = d_7 = 2}$$

- nodurile 3 şi 5 au gradul 3 (număr de muchii incidente = 3);

$$\boldsymbol{d_3 = d_5 = 3}$$



b) Numărul mediu de muchii incidente unui nod ("*average degree*") este un indicator care măsoară câte muchii incidente are, în medie, un nod al reţelei. Se calculează ca o medie a gradelor nodurilor (sau se poate exprima prin dublul numărului muchiilor impărţit la numărul de noduri)

$$\boldsymbol{\bar{d} = \frac{2*5+3*2}{7} = 2,28}$$

- Fie N – numărul de noduri şi fie nE – numărul de muchii; atunci:

$$\boldsymbol{\bar{d} = \frac{2*nE}{N} = \frac{2*8}{7} = 2,28}$$

Interpretare: în medie, un nod al reţelei are 2,28 muchii incidente.

c) Densitatea grafului (*"graph density"*) măsoară numărul de muchii al grafului în raport cu numărul maxim de muchii posibile în graf.

Formula de calcul:

$$\boldsymbol{Den} = \frac{\mathbf{2} * \boldsymbol{nE}}{\boldsymbol{N}(\boldsymbol{N}-\mathbf{1})}$$

$$\boldsymbol{Den} = \frac{\mathbf{2} * \mathbf{8}}{\mathbf{7}(\mathbf{7}-\mathbf{1})} = \mathbf{0,381}$$

Interpretare: indicatorul de densitate în reţea are valoara 0,381, valoarea maximă a indicatorului fiind 1 (se atinge când graful este complet, fiecare nod fiind conectat cu fiecare alt nod). Deci există o densitate mică a conexiunilor în reţea.

d) Drumul mediu de lungime minimă (*"average path length"*) este numărul mediu de muchii ce trebuiesc străbătute pe drumul cel mai scurt dintre oricare două noduri ale reţelei.

Formula de calcul:

$$\boldsymbol{l} = \frac{\mathbf{2}}{\boldsymbol{N}(\boldsymbol{N}-\mathbf{1})} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{l}_{\boldsymbol{min}}(\boldsymbol{i},\boldsymbol{j}) \quad , \qquad \boldsymbol{i} \neq \boldsymbol{j}$$

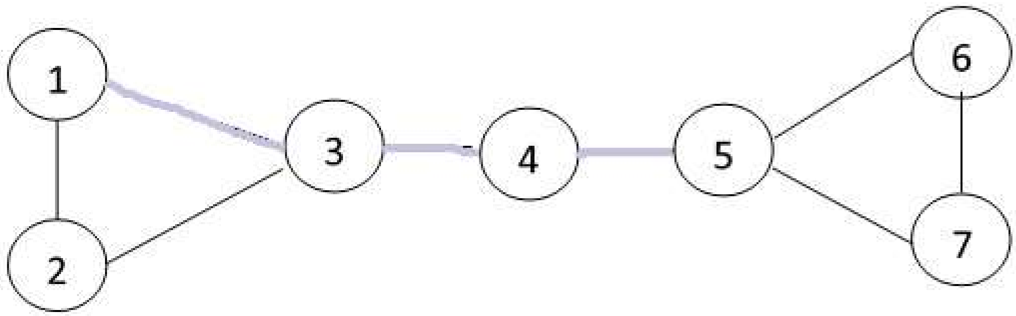
unde:

- $\boldsymbol{l}_{\boldsymbol{min}}(\boldsymbol{i},\boldsymbol{j})$ este distanţa cea mai mică între nodurile *i* şi *j* , adică numărul minim de muchii ce trebuiesc străbătute pe cel mai scurt drum dntre nodurile *i* şi *j* ;
- N reprezintă numărul de noduri al grafului (N=7)

Formând toate perechile de noduri ***(i,j)*** cu $\boldsymbol{i} \neq \boldsymbol{j}$, obţinem:

$$l = \frac{2}{N(N-1)}(l_{1\to2} + l_{1\to3} + l_{1\to4} + l_{1\to5} + l_{1\to6} + l_{1\to7} +$$
$$+l_{2\to3} + l_{2\to4} + l_{2\to5} + l_{2\to6} + l_{2\to7} +$$
$$+l_{3\to4} + l_{3\to5} + l_{3\to6} + l_{3\to7} +$$
$$+l_{4\to5} + l_{4\to6} + l_{4\to7} + l_{5\to6} + l_{5\to7} + l_{6\to7})$$

Spre exemplu, distanţa cea mai mică între nodurile (1 , 5) este 3, fiind necesară străbaterea muchiilor 1 – 3, 3 – 4, 4 – 5.



Aşadar obţinem:

$$\Rightarrow l = \frac{2}{7(7-1)}(1+1+2+3+4+4+1+2+3+4+4+1+2+3+3+1+2+2+1+1+1)$$

$$\boldsymbol{l = \frac{2}{42} * 46 = 2,19}$$

Interpretare: în medie, trebuiesc străbătute 2,19 muchii pe drumul cel mai scurt dintre oricare două noduri ale grafului.

d) Diametrul reţelei (*"network diameter"*) reprezintă cel mai lung drum minim din reţea, drumurile minime din reţea fiind, după cum am văzut mai sus, drumurile cele mai scurte dintre două noduri.

Formula de calcul:

$$\boldsymbol{D = max\ (l_{min}(i, j))}$$

$$\boldsymbol{D = 4}$$

Diametrul *D* al reţelei este 4 şi este dat de cel mai lung drum minim din reţea (de exemplu de la nodul *1* la nodul *7* se ajunge trecănd prin minim 4 muchii).

f) Coeficientul de clusterizare în jurul unui nod (*"node clustering coefficient"*) este definit ca raportul dintre numărul de legături existente şi numărul de legături potenţiale între vecinii nodului analizat.

Formula de calcul:

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

unde:

- $\boldsymbol{k_i}$ reprezintă numărul de vecini ai nodului *i* ;
- $\boldsymbol{E_i}$ reprezintă numărul de legături existente între cei $\boldsymbol{k_i}$ vecini ai nodului *i* ;
- $\frac{\boldsymbol{k_i(k_i-1)}}{\mathbf{2}}$ reprezintă numărul de legături potenţiale între cei $\boldsymbol{k_i}$ vecini ai nodului *i*

Pentru nodul *1,* coeficientul de clusterizare în jurul nodului este *1*:

$$\boldsymbol{C_1} = \frac{\mathbf{2 * 1}}{\mathbf{2(2-1)}} = \mathbf{1}$$

Pentru nodul *3,* coeficientul de clusterizare în jurul nodului este *0,33*:

$$\boldsymbol{C_3} = \frac{\mathbf{2 * 1}}{\mathbf{3(3-1)}} = \mathbf{0,33}$$

Pentru nodul *4,* coeficientul de clusterizare în jurul nodului este *0*:

$$\boldsymbol{C_4} = \frac{\mathbf{2 * 0}}{\mathbf{2(2-1)}} = \mathbf{0}$$

g) Coeficientul mediu de clusterizare în reţea (*"average clustering coefficient"*) semnifică cât de densă este, în medie, vecinătatea nodurilor reţelei. Se calculează ca o medie a coeficienţilor de clusterizare în jurul tuturor nodurilor reţelei.

Formula de calcul:

$$\overline{\boldsymbol{C}} = \frac{\sum_{i=1}^{N} \boldsymbol{C_i}}{\boldsymbol{N}}$$

unde:

- $\boldsymbol{C_i}$ **-** coeficientul de clusterizare în jurul nodului *i;*
- $\boldsymbol{N}$ **-** numărul de noduri al grafului.

$$\boldsymbol{C_1} = \mathbf{1};\ \boldsymbol{C_2} = \mathbf{1}; \boldsymbol{C_3} = \mathbf{0,33};\ \boldsymbol{C_4} = \mathbf{0};\ \boldsymbol{C_5} = \mathbf{0,33};\ \boldsymbol{C_6} = \mathbf{1};\ \boldsymbol{C_7} = \mathbf{1}$$

$$\overline{\boldsymbol{C}} = \frac{\mathbf{1 * 4 + 0,33 * 2 + 0}}{\mathbf{7}} = \mathbf{0,66}$$

Interpretare: coeficientul mediu de clusterizare în reţea are valoarea 0,66. Dacă fiecare nod ar fi fost conectat cu fiecare nod al grafului, valoarea indicatorului ar fi fost 1, aşadar interpretăm că în reţeaua exemplificată valoarea coeficientului mediu de clusterizare nu este nici ridicată, nici scăzută, densitatea vecinătăţii nodurilor fiind medie spre riddicată.

h) Centralitatea nodurilor, măsurată prin gradul nodurilor (*"degree centrality"*) indică importanţa (centralitatea) unui nod în reţea din perspectiva numărului de muchii incidente nodului. Cu cât gradul nodului este mai mare, cu atât nodul este considerat mai important în reţea. Se calculează prin raportul dintre gradul nodului și numărul maxim posibil de conexiuni al nodului respectiv.

Formula de calcul:

$$D_c(i) = \frac{d(i)}{N-1}$$

unde:

- $d(i)$ - gradul nodului *i;*
- $N-1$ - numărul maxim posibil de conexiuni al nodului *i*.

$D_c(1) = \frac{2}{7-1} = \mathbf{0,33.}$ $D_c(2) = \frac{2}{7-1} = \mathbf{0,33.}$ $D_c(3) = \frac{3}{7-1} = \mathbf{0,5.}$

$D_c(4) = \mathbf{0,33}.$ $D_c(5) = \mathbf{0,5.}$ $D_c(6) = \mathbf{0,33}.$ $D_c(7) = \mathbf{0,33}.$

De exemplu, nodul *5* are trei muchii incidente şi şase muchii potenţial incidente, deci centralitatea nodului *5* măsurată prin gradul nodului este *0,5*.

Interpretare: nodurile 3 şi 5 sunt cele mai centrale noduri ale grafului, masurând centralitatea prin gradul nodurilor, acestea înregistrând cele mai ridicate valori ale indicatorului.

j) **Centralitatea nodurilor, măsurată prin apropierea nodului de alte noduri** (*"closeness centrality"*) indică importanţa (centralitatea) unui nod în reţea din perspectiva apropierii nodului faţă de toate celelalte noduri din reţea. Se calculează ca fiind inversul lungimii medii a celor mai scurte drumuri de la nodul respectiv la fiecare din celelalte noduri (raportul dintre numărul maxim posibil de conexiuni cu nodul respectiv şi suma drumurilor de lungime minimă de la nod la celelalte noduri ale reţelei).

Formula de calcul:

$$C_c(i) = \frac{N-1}{\sum_{j=1, j\neq i}^{n} l_{min}(i,j)}$$

unde:

- $l_{min}(i,j)$ este drumul de lungime minimă de la nodul i la nodul j;
- $N-1$ - numărul maxim posibil de conexiuni al nodului *i*.

Pentru nodurile *1, 2, 6, 7*:

$$C_c(1) = \frac{7-1}{1+1+2+3+4+4} = \mathbf{0,4}$$

$$C_c(2) = C_c(6) = C_c(7) = \mathbf{0,4}$$

Pentru nodurile *3, 5*:

$$C_c(3) = \frac{7-1}{1+1+1+2+3+3} = \mathbf{0,54}$$

$$C_c(5) = \mathbf{0,54}$$

Pentru nodul *4*:

$$C_c(4) = \frac{7-1}{2+2+1+1+2+2} = \mathbf{0,6}$$

Interpretare: nodul *4* este cel mai important nod al reţelei (central) din perspectiva apropierii nodului de celelalte noduri ale reţelei, urmat de nodurile 3 şi 5. De la nodul *4* se poate ajunge la orice alt nod al reţelei traversând doar 1 sau 2 muchii.

k) **Indicatorul de modularitate a reţelei** (*"modularity"*) măsoară potenţialul de divizare a reţelei în comunităţi (grupe , module, clustere) pe baza densităţii conexiunilor dintre noduri în cadrul fiecărei comunităţi. Formula de calcul se bazează pe diferenţa dintre numărul de muchii existente în cadrul unei comunităţi şi numărul maxim potential de muchii în cadrul respectivei comunităţi

Formula de calcul:

$$Q = \frac{1}{2*nE}\sum_{i,j}(A_{i,j} - \frac{d(i)d(j)}{2*nE})\delta(C^{(i)}, C^{(j)})$$

sau

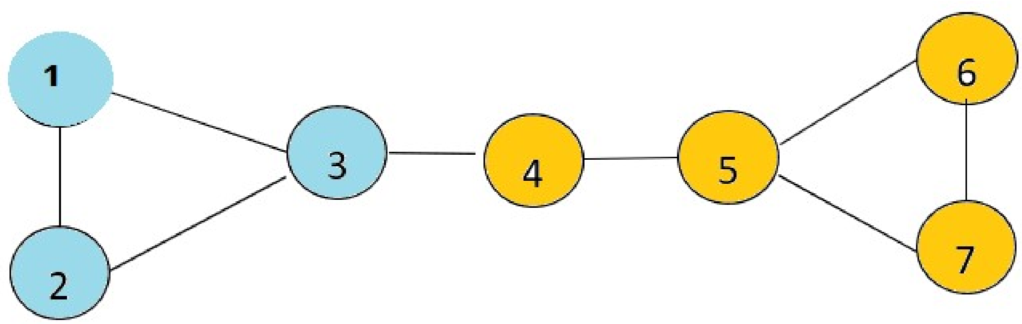
$$Q = \frac{1}{2*nE}\sum_{C}\sum_{i\in C}\sum_{j\in C}(A_{i,j} - \frac{d(i)d(j)}{2*nE})$$

unde:

- $\boldsymbol{nE}$ - numărul de muchii al reţelei;
- $A_{i,j}$ – elementele matricii de adiacenţă, având valorile *1* dacă între nodurile *i şi j* există conexiune, sau *0* dacă nu există conexiune între respectivele noduri;
- $d(i)$ **-** gradul nodului *i;* $d(j)$ **-** gradul nodului *j;*
- $C^{(i)}$ – comunitatea (tipul) nodului *i*; $C^{(j)}$ – comunitatea (tipul) nodului *j*;
- $\delta\left(C^{(i)}, C^{(j)}\right)$ - funcţia *Kronecker delta* , $\delta\left(C^{(i)}, C^{(j)}\right) = \begin{cases} 1, dacă\ C^{(i)} = C^{(j)} \\ 0, altfel \end{cases}$

Împărţim reţeaua în 2 comunităţi:

- $C^{(1)}$ – formată din nodurile [1, 2, 3]
- $C^{(2)}$ – formată din nodurile [4, 5, 6, 7]



În comunitatea $C^{(1)} = [1, 2, 3]$, perechile de noduri cu $A_{i,j} = 1$ sunt:

- (i,j) = (1, 2); $A_{i,j} - \frac{d(i)d(j)}{2*nE} = 1 - \frac{2*2}{2*8} = \mathbf{0.75}$;
- (i,j) = (1, 3); $A_{i,j} - \frac{d(i)d(j)}{2*nE} = 1 - \frac{2*3}{2*8} = \mathbf{0.625}$ **;**
- (i,j) = (2, 3); $A_{i,j} - \frac{d(i)d(j)}{2*nE} = 1 - \frac{2*3}{2*8} = \mathbf{0.625}$

În comunitatea $C^{(1)} = [1, 2, 3]$, nu avem perechi de noduri cu $A_{i,j} = 0$.

În comunitatea $C^{(2)} = [4, 5, 6, 7]$, perechile de noduri cu $A_{i,j} = 1$ sunt:

- (i,j) = (4, 5); $A_{i,j} - \frac{d(i)d(j)}{2*nE} = 1 - \frac{2*3}{2*8} = \mathbf{0.625}$;
- (i,j) = (5, 6); $A_{i,j} - \frac{d(i)d(j)}{2*nE} = 1 - \frac{2*3}{2*8} = \mathbf{0.625}$ **;**

- (i,j) = (5, 7); $A_{i,j} - \frac{d(i)d(j)}{2*nE}$ = $1 - \frac{2*3}{2*8} = \mathbf{0.625}$ **;**
- (i,j) = (6, 7); $A_{i,j} - \frac{d(i)d(j)}{2*nE}$ = $1 - \frac{2*2}{2*8} = \mathbf{0.75}$

În comunitatea $C^{(2)} = [4, 5, 6, 7]$, perechile de noduri cu $A_{i,j} = 0$ sunt:

- (i,j) = (4, 6); $A_{i,j} - \frac{d(i)d(j)}{2*nE}$ = $0 - \frac{2*2}{2*8} = \mathbf{-0.25}$;
- (i,j) = (4, 7); $A_{i,j} - \frac{d(i)d(j)}{2*nE}$ = $0 - \frac{2*2}{2*8} = \mathbf{-0.25}$

Atunci:

$$Q = \frac{1}{2 * nE} \sum_{C} \sum_{i \in C} \sum_{j \in C} (A_{i,j} - \frac{d(i)d(j)}{2 * nE}) =$$

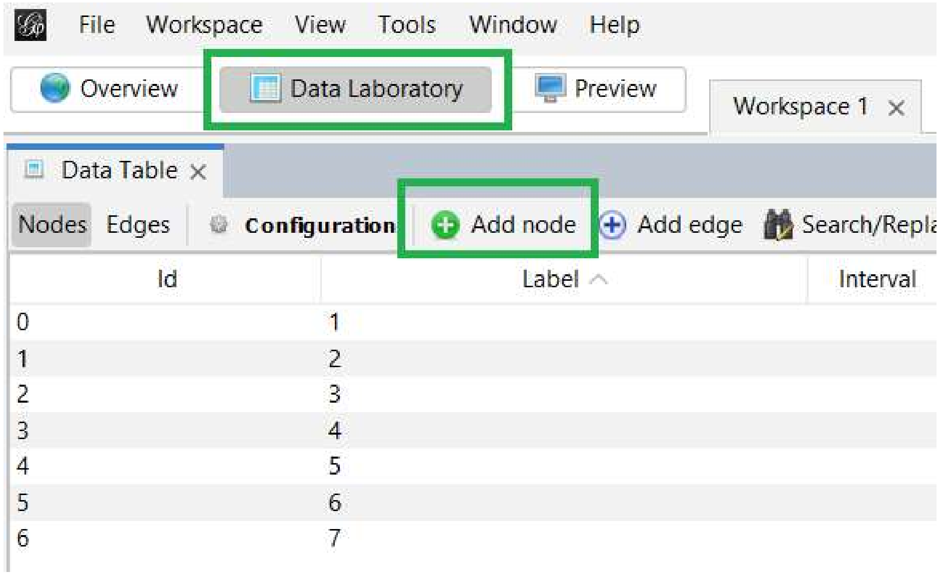
$$= \frac{1}{2 * 8} (\mathbf{0.75 + 0.625 + 0.625 + 0.625 + 0.625 + 0.625 + 0.75 - 0.25 - 0.25})$$

$$= \frac{1}{16} * \mathbf{4.125 = 0,258}$$

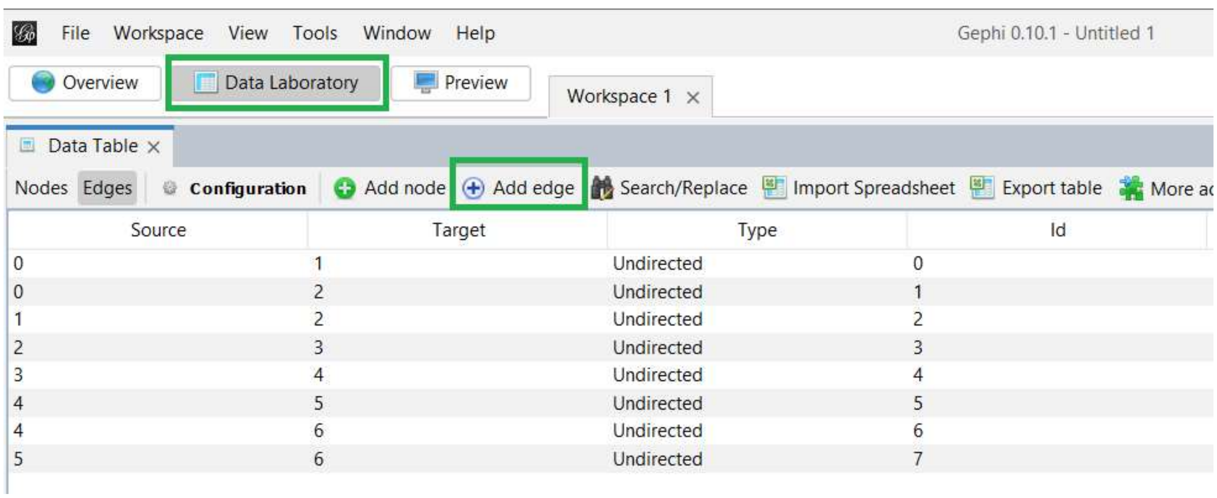
Interpretare: indicatorul de modularitate *Q* poate lua valori în intervalul [-1, 1]. O valoare pozitivă a modularităţii *Q* indică faptul că numărul de muchii din cadrul comunităţilor reţelei este mai mare decât numărul de muchii din cadrul comunităţilor în situaţia în care structura de noduri şi muchii a reţelei ar fi rescrisă într-un mod aleatoriu. O valoare a lui Q de peste 0.3 indică o structură a reţelei împărţită pe comunităţi semnificativă. În exemplul dat, Q = 0,258 aşadar nu este semnificativă împărţirea reţelei în 2 comunităţi.

Implementarea studiului de caz în programul *Gephi*

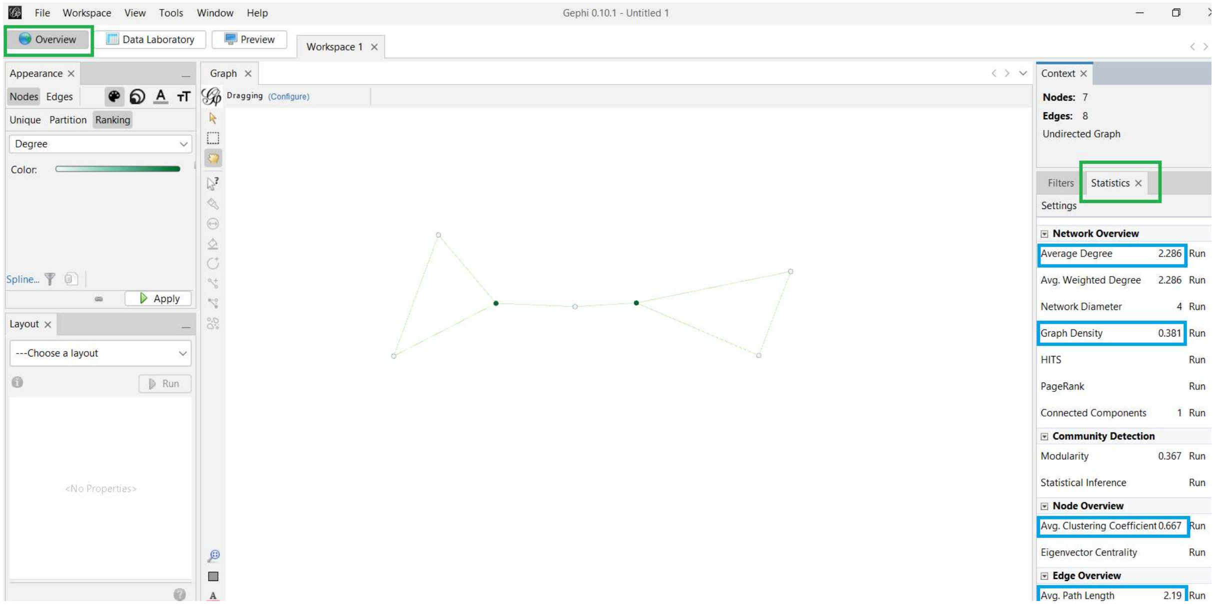
- După accesarea programului Gephi, se alege opţiunea 'New project';
- Definirea nodurilor reţelei se face cu ajutorul opţiunilor "Data Laboratory", "Add node"



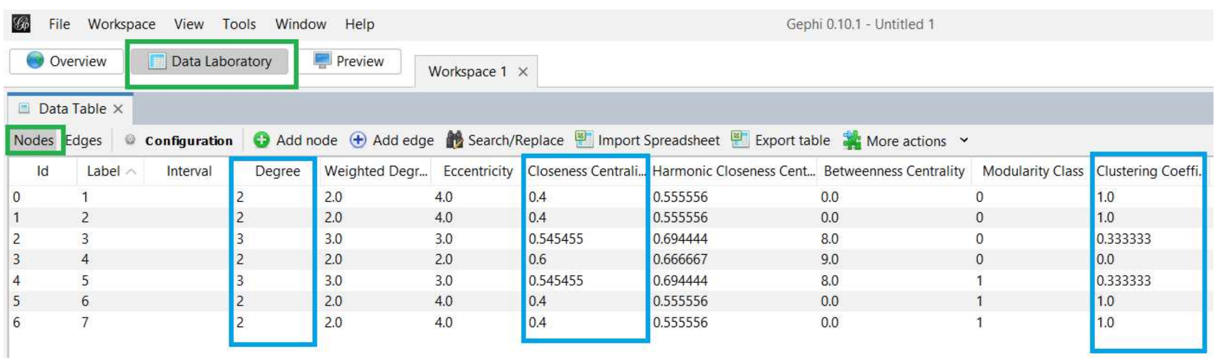
- Definirea muchiilor reţelei se face cu ajutorul opţiunilor "Data Laboratory", "Add edge"



- Calcularea indicatorilor se face din meniul "Overview", "Statistics"



- O gamă mai variată de indicatori la nvel de nod se accesează din meniul “Data Laboratory”, “Nodes”:



Id	Label	Interval	Degree	Weighted Degr...	Eccentricity	Closeness Centrali...	Harmonic Closeness Cent...	Betweenness Centrality	Modularity Class	Clustering Coeffi...
0	1		2	2.0	4.0	0.4	0.555556	0.0	0	1.0
1	2		2	2.0	4.0	0.4	0.555556	0.0	0	1.0
2	3		3	3.0	3.0	0.545455	0.694444	8.0	0	0.333333
3	4		2	2.0	2.0	0.6	0.666667	9.0	0	0.0
4	5		3	3.0	3.0	0.545455	0.694444	8.0	1	0.333333
5	6		2	2.0	4.0	0.4	0.555556	0.0	1	1.0
6	7		2	2.0	4.0	0.4	0.555556	0.0	1	1.0

PROVOCARE. Deschide în Gephi o reţea din librărie (Les Miserables, Java, Power Grid, US Airports, etc). Analizează reţeaua. Descoperă 5 noi funcţionalităţi în Gephi. Include capturi de ecrane relevante.