

# ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Σχολή Ηλ. Μηχ. & Μηχ. Η/Υ - 9<sup>ο</sup> Εξάμηνο

## 3<sup>η</sup> Εργαστηριακή Άσκηση

### Γενετικοί Αλγόριθμοι & Επιδημιολογικά Μοντέλα

Στην παρούσα άσκηση θα εξοικειωθείτε με τη χρήση των γενετικών αλγορίθμων στο MATLAB και την επιλογή των παραμέτρων τους. Επιπλέον, θα αναπτύξετε ένα γενετικό αλγόριθμο για εντοπισμό κοινοτήτων σε δίκτυα και θα συγκρίνετε το αποτέλεσμα του με συμβατικούς αλγόριθμους εντοπισμού κοινοτήτων που εφαρμόστηκαν στη 2<sup>η</sup> εργαστηριακή άσκηση. Τέλος θα εξοικειωθείτε με τα βασικά επιδημιολογικά μοντέλα SIR, SIS.

#### ΜΕΡΟΣ Α

##### Άσκηση 1<sup>ο</sup>: Εξοικείωση με τους Γενετικούς Αλγορίθμους

###### Πρόβλημα ONEMAX

Βρείτε τη **δυναδική** ακολουθία  $(x_1, x_2, \dots, x_n)$  που μεγιστοποιεί το άθροισμα  $x_1 + x_2 + \dots + x_n$ ,  $n=20$ , χρησιμοποιώντας τις συναρτήσεις του MATLAB, για γενετικούς αλγορίθμους,

<http://www.mathworks.com/help/gads/performing-a-genetic-algorithm-optimization.html>

Αναλύστε την επίδραση του μεγέθους του πληθυσμού, της πιθανότητας διασταύρωσης (crossover) και της πιθανότητας μετάλλαξης (mutation) **στην ποιότητα της λύσης (πόσο κοντά είναι στην προφανή βέλτιστη λύση)** που προκύπτει από το γενετικό αλγόριθμο. Συγκεκριμένα, μεταβάλλετε τον πληθυσμό από 10 έως 100 χρωμοσώματα με βήμα 10, την πιθανότητα διασταύρωσης από 0.3 έως 0.9 με βήμα 0.1 και την πιθανότητα μετάλλαξης από 0.01 ως 0.2 με βήμα 0.01. Εξετάστε την ποιότητα της λύσης του γενετικού αλγορίθμου **για κάθε συνδυασμό** μεγέθους πληθυσμού, πιθανότητας μετάλλαξης και πιθανότητας διασταύρωσης που προκύπτει.

##### Άσκηση 2: Εντοπισμός Κοινοτήτων σε Γράφους Κοινωνικών Δικτύων με Χρήση Γενετικών Αλγορίθμων

Στην παρούσα άσκηση θα αναπτύξετε κώδικα MATLAB για εντοπισμό κοινοτήτων σε γράφους δικτύων με χρήση γενετικών αλγορίθμων. Ακολουθώντας θα εφαρμόσετε το γενετικό αλγόριθμο για τον εντοπισμό κοινοτήτων σε πραγματικά δίκτυα και θα συγκρίνετε το αποτέλεσμα του με εκείνο των αλγορίθμων της δεύτερης εργαστηριακής άσκησης χρησιμοποιώντας τη μετρική modularity.

Ο αλγόριθμος για το γενετικό αλγόριθμο που θα πρέπει να μετατραπεί σε κώδικα φαίνεται στην Εικόνα 1 (οι συμβολισμοί επεξηγούνται στις διαφάνειες του αντίστοιχου μαθήματος):

# ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Σχολή Ηλ. Μηχ. & Μηχ. Η/Υ - 9<sup>ο</sup> Εξάμηνο

```
 $t := 0;$   
Create initial population  $B_0 = (b_{1,0}, \dots, b_{m,0});$   
  
WHILE stopping condition not fulfilled DO  
  BEGIN  
  
    (* proportional selection *)  
  
    FOR  $i := 1$  TO  $m$  DO  
      BEGIN  
         $x := \text{Random}[0, 1];$   
  
         $k := 1;$   
        WHILE  $k < m \ \& \ x < \sum_{j=1}^k f(b_{j,t}) / \sum_{j=1}^m f(b_{j,t})$  DO  
           $k := k + 1;$   
  
         $b_{i,t+1} := b_{k,t}$   
      END  
  
      (* one-point crossover *)  
  
      FOR  $i := 1$  TO  $m - 1$  STEP 2 DO  
        BEGIN  
          IF  $\text{Random}[0, 1] \leq p_C$  THEN  
            BEGIN  
               $pos := \text{Random}\{1, \dots, n - 1\};$   
  
              FOR  $k := pos + 1$  TO  $n$  DO  
                BEGIN  
                   $aux := b_{i,t+1}[k];$   
                   $b_{i,t+1}[k] := b_{i+1,t+1}[k];$   
                   $b_{i+1,t+1}[k] := aux$   
                END  
              END  
            END  
          END  
        END  
  
      (* mutation *)  
  
      FOR  $i := 1$  TO  $m$  DO  
        FOR  $k := 1$  TO  $n$  DO  
          IF  $\text{Random}[0, 1] < p_M$  THEN  
            invert  $b_{i,t+1}[k];$   
          END  
        END  
  
         $t := t + 1$   
      END
```

Εικόνα 1: Γενική μορφή γενετικού αλγορίθμου.

# ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Σχολή Ηλ. Μηχ. & Μηχ. Η/Υ - 9<sup>ο</sup> Εξάμηνο

## Αναπαράσταση υποψήφιων λύσεων (χρωμοσωμάτων)

Κάθε χρωμόσωμα  $b_{i,t}$  του πληθυσμού τη χρονική στιγμή  $t$  είναι ένα διάνυσμα με μήκος ίσο με το πλήθος των κόμβων του δικτύου,  $n$ . Αν  $b_{i,t}(j) = d$  τότε οι κόμβοι  $j, d$  ανήκουν στην ίδια κοινότητα. Επιπλέον για να ισχύει  $b_{i,t}(j) = d$ , οι κόμβοι  $j, d$  θα πρέπει να συνδέονται στην αρχική τοπολογία. Τότε, βρίσκοντας τις συνδεδεμένες συνιστώσες (FindComponents.m) του χρωμοσώματος  $b_{i,t}$ , προκύπτουν οι κοινότητες του δικτύου.

## Αρχικοποίηση

Η αρχική γενιά του πληθυσμού θα πρέπει να αποτελείται από χρωμοσώματα τα οποία θα είναι τυχαία επιλεγμένα και διορθωμένα ώστε ο αριθμός που υπάρχει σε κάθε θέση να είναι γείτονας του αντίστοιχου κόμβου. (Αν δεν ισχύει θα πρέπει να αντικατασταθεί από έναν γείτονα του αντίστοιχου κόμβου.) Θεωρείστε πληθυσμό 300 χρωμοσωμάτων.

## Συνάρτηση Fitness

Η συνάρτηση fitness υπολογίζεται όπως στις διαφάνειες 28 και 29 της αντίστοιχης διάλεξης του μαθήματος. Προσοχή για κάθε συνιστώσα (component) του γράφου θα πρέπει να βρίσκετε τον αντίστοιχο υπογράφο χρησιμοποιώντας τη συνάρτηση subgraph.m.

## Επιλογή (Selection)

Η επιλογή χρωμοσωμάτων από μία γενιά του πληθυσμού για την κατασκευή της επόμενης γενιάς γίνεται με τη μέθοδο της ρουλέτας (διαφάνειες μαθήματος και αλγόριθμος). Τροποποιήστε τον παραπάνω αλγόριθμο ώστε να εφαρμόσετε και ελιτισμό, δηλαδή τα  $x$  πρώτα χρωμοσώματα της νέας γενιάς θα είναι εκείνα τα χρωμοσώματα της προηγούμενης γενιάς που έχουν επιτύχει τη μέγιστη τιμή της συνάρτησης fitness.

## Διασταύρωση (Crossover)

Η διασταύρωση πραγματοποιείται όπως περιγράφεται στον αλγόριθμο, μεταξύ διαδοχικών χρωμοσωμάτων της εκάστοτε γενιάς. Η πιθανότητα διασταύρωσης ανά ζεύγος είναι ίση με  $p_c$ .

## Μετάλλαξη (Mutation)

Η μετάλλαξη πραγματοποιείται σε κάθε θέση του χρωμοσώματος με πιθανότητα  $p_m$ , όπως στον αλγόριθμο. Θα πρέπει ωστόσο ο αλγόριθμος να τροποποιηθεί ώστε να δίνει ένα τυχαίο γείτονα του κόμβου που αντιστοιχεί σε κάθε θέση ως αποτέλεσμα της μετάλλαξης.

## Κριτήριο Τερματισμού (Stopping Criterion)

Ως κριτήριο τερματισμού ορίστε ένα μέγιστο πλήθος επαναλήψεων εφαρμογής του αλγορίθμου (γενεών πληθυσμού), ίσο με 30, ή ότι η βέλτιστη τιμή της συνάρτησης fitness δεν έχει μεταβληθεί για τις τελευταίες 5 επαναλήψεις.

Εφαρμόστε τον παραπάνω γενετικό αλγόριθμο στις πραγματικές τοπολογίες της δεύτερης εργαστηριακής άσκησης που δίνονται και στον Πίνακα 1. Οι παράμετροι που

# ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Σχολή Ηλ. Μηχ. & Μηχ. Η/Υ - 9<sup>ο</sup> Εξάμηνο

θα πρέπει να εξεταστούν δίνονται στον παρακάτω πίνακα. Προσοχή θα πρέπει να εξεταστούν όλοι οι συνδυασμοί.




Παράμετρος	Τιμές
Διασταύρωση $p_c$	0.7:0.9 με βήμα 0.1
Μετάλλαξη $p_M$	0.1, 0.2
Ελιτισμός $x$	1:3 με βήμα 1

Συγκρίνετε το αποτέλεσμα του γενετικού αλγορίθμου (το καλύτερο για τις διάφορες επιλογές παραμέτρων) με εκείνο των αλγορίθμων εντοπισμού κοινοτήτων της εργαστηριακής άσκησης 2 (δίνονται και στον Πίνακα 2), υπολογίζοντας το modularity χρησιμοποιώντας τη συνάρτηση QFModul.m. Θα πρέπει να έχετε πρωτίστως μετατρέψει το αποτέλεσμα του γενετικού αλγορίθμου σε κατάλληλη μορφή για να θεωρηθεί ως είσοδος στην QFModul.m, με χρήση της FindComponents.m πάνω στο χρωμόσωμα που επιτυγχάνει τη βέλτιστη τιμή της συνάρτησης fitness.

Πίνακας 1: Γράφοι δικτύων από πραγματικά δεδομένα.

	Τοπολογία	Αρχείο
ΔΙΚΤΥΑ ΠΡΑΓΜΑΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ	<a href="#">American College football</a>	football.gml
	<a href="#">Les Miserables</a>	lesmis.gml
	<a href="#">Dolphin social network</a>	dolphins.gml

Πίνακας 2: Αλγόριθμοι εντοπισμού κοινοτήτων.

Μέθοδος Εντοπισμού Κοινοτήτων	Αρχείο ComDetTBv090\Algorithms
Spectral Clustering	 GCSpectralClust2.m
Newman-Girvan	 newmangirvan.m
Modularity Maximization	 GCModulMax1.m

Στην παρούσα άσκηση θα πρέπει να παραδοθεί και ο κώδικας.

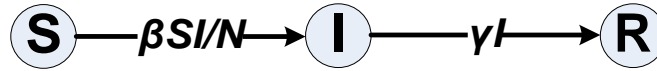
# ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Σχολή Ηλ. Μηχ. & Μηχ. Η/Υ - 9<sup>ο</sup> Εξάμηνο

## ΜΕΡΟΣ Β

Το παρόν μέρος θα επικεντρώσει στην αριθμητική επίλυση των βασικών επιδημιολογικών μοντέλων SIR και SIS, με τη βοήθεια του MATLAB.

A) Το SIR μοντέλο δίνεται από το ακόλουθο διάγραμμα μεταβάσεων καταστάσεων:



και περιγράφεται από το σύνολο διαφορικών εξισώσεων:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}\quad (1)$$

Χρησιμοποιώντας το αρχείο SIR.m από το πακέτο έτοιμου κώδικα, και για τις παραμέτρους που φαίνονται στον ακόλουθο Πίνακα 3, να δείξετε σε κοινό διάγραμμα για κάθε συνδυασμό παραμέτρων, τα  $S(t), I(t), R(t)$ . Τι παρατηρείτε? Σχολιάστε τα αποτελέσματα σε σχέση με το βασικό ρυθμό αναπαραγωγής  $R_0$ .

Πίνακας 3 – Παράμετροι μελέτης μοντέλου SIR

Παράμετροι	Τιμές					
$\beta$	$10^{-3}$	$10^{-3}$	$10^{-3}$	$10^{-3}$	$10^{-3}$	$10^{-3}$
$\gamma$	$10^{-6}$	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-2}$	$10^{-1}$
$S(0)$	5x0					
$I(0)$	1x					
$R(0)$	0					

B) Για το μοντέλο SIS, υιοθετείται η περίπτωση περιοδικού ρυθμού επαφής κόμβων. Κάτι τέτοιο μπορεί να αντιστοιχεί σε μια περιοδικότητα στη λήψη επιδημιών, π.χ. σε ετήσια βάση όπως με τον συνιθισμένο ιό της γρίπης. Το σύνολο των διαφορικών εξισώσεων το οποίο περιγράφει αυτό το μοντέλο είναι:

$$\begin{aligned}\frac{dI}{dt} &= (\beta(t)N - a)I - \beta(t)I^2 \\ \beta(t) &= 2 - 1.8\cos(5t)\end{aligned}\quad (2)$$

Με τη βοήθεια του έτοιμου κώδικα SIS.m και για τους συνδυασμούς παραμέτρων που φαίνονται στον ακόλουθο Πίνακα 4, να δείξετε σε κοινό διάγραμμα την εξέλιξη του φαινομένου για κάθε συνδυασμό τιμών ( $I(t), \beta(t)$ ).

Πίνακας 4 - Παράμετροι μελέτης μοντέλου SIS

Παράμετροι	Τιμές					
$a$	6	5	4	3	2	1
$N$	1					