

Πανεπιστήμιο Θεσσαλίας  
Πολυτεχνική σχολή  
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών  
Υπολογιστών

## Διπλωματική εργασία

Τσινοπούλου Δέσποινα

"Κοινότητες βασισμένες στη ροή πληροφορίας σε  
σύνθετα δίκτυα"

"Cascade-based community detection in complex  
networks"

### Επιβλέποντες Καθηγητές :

Κατσαρός Δημήτριος, Επίκουρος Καθηγητής Π.Θ  
Μποζάνης Παναγιώτης, Αναπληρωτής Καθηγητής

Βόλος, Οκτώβριος 2014



Διπλωματική Εργασία για την απόκτηση του Διπλώματος του Ηλεκτρολόγου Μηχανικού και Μηχανικού Ηλεκτρονικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας, στα πλαίσια του Προγράμματος Προπτυχιακών Σπουδών του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας.

Copyright © Tsinopoulou Despoina, 2014  
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό λόγο. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται στο συγγραφέα.

Στην οικογένεια μου

## Ευχαριστίες

Με την περάτωση της παρούσας εργασίας, θα ήθελα να ευχαριστήσω θερμά τους επιβλέποντες της διπλωματικής εργασίας κ. Κατσαρό Δημήτριο, Επίκουρο Καθηγητή Π.Θ και κ. Μποζάνη Παναγιώτη, Αναπληρωτή Καθηγητή για την εμπιστοσύνη που μου έδειξαν, την άριστη συνεργασία, την συνεχή καθοδήγηση αλλά και για το χρόνο που μου αφιέρωσαν κατά τη διάρκεια εκπόνησης της διπλωματικής. Τέλος ένα μεγάλο ευχαριστώ στην οικογένεια μου για την ολόψυχη αγάπη και υποστήριξή τους όλα αυτά τα χρόνια και ιδιαίτερα κατά τη διάρκεια των τελευταίων μηνών της προσπάθειάς μου, καθώς και τους φίλους μου για τη στήριξη τους καθ' όλη τη διάρκεια των σπουδών μου.

# Περιεχόμενα

Ευχαριστίες.....	5
1.Εισαγωγή.....	8
1.1 Μια Γενική Περιγραφή.....	8
1.2 Σκοπός Της Εργασίας.....	9
2. Το μοντέλο Community – Cascade Network ( CCN ).....	10
2.1 Σύντομη περιγραφή του μοντέλου.....	10
2.2 Η θεωρία του μοντέλου CCN .....	10
3. Ο αλγόριθμος CIBC.....	13
3.1 Σύντομη Περιγραφή Του Προβλήματος.....	13
3.2 Η θεωρία του αλγορίθμου CIBC.....	13
4. Το μοντέλο SIR.....	15
5. Ο αλγόριθμος epidemic-CIBC.....	17
5.1 Περιγραφή του αλγορίθμου που υλοποιήσαμε.....	17
5.2 Τρεξίματα του αλγορίθμου σε μικρά παραδείγματα.....	17
Βιβλιογραφία.....	24
Βιβλιογραφία εικόνων.....	25

## Λίστα Εικόνων

Εικόνα 1.1: Κατευθυνόμενο γράφημα [1].....	8
Εικόνα 1.2: Μη - κατευθυνόμενου γράφημα[2].....	8
Εικόνα 2.1: Εξισώσεις υπολογισμού του μοντέλου CCN[3].....	11
Εικόνα 2.2: Εξισώσεις υπολογισμού του μοντέλου CCN[4].....	11
Εικόνα 2.3: Εξισώσεις υπολογισμού του μοντέλου CCN[5].....	11
Εικόνα 2.4: Εξισώσεις υπολογισμού του μοντέλου CCN[6].....	11
Εικόνα 2.5: Εξισώσεις υπολογισμού του μοντέλου CCN[7].....	12
Εικόνα 2.6: Εξισώσεις υπολογισμού του μοντέλου CCN[8].....	12
Εικόνα 2.7: Εξισώσεις υπολογισμού του μοντέλου CCN[9].....	12
Εικόνα 2.8: Ψευδοκώδικας υπολογισμού του μοντέλου CCN[10].....	12
Εικόνα 3: Ψευδοκώδικας υπολογισμού του αλγορίθμου CIBC[11].....	14
Εικόνα 4: Μοντέλο SIR[12].....	16
Εικόνα 5.1: Τρέξιμο του αλγορίθμου σε παράδειγμα[13].....	18
Εικόνα 5.2: Τρέξιμο του αλγορίθμου σε παράδειγμα[14].....	19
Εικόνα 5.3: Τρέξιμο του αλγορίθμου σε παράδειγμα[15].....	19
Εικόνα 5.4: Τρέξιμο του αλγορίθμου σε παράδειγμα[16].....	20
Εικόνα 5.5: Τρέξιμο του αλγορίθμου σε παράδειγμα[17].....	23

# Κεφάλαιο 1

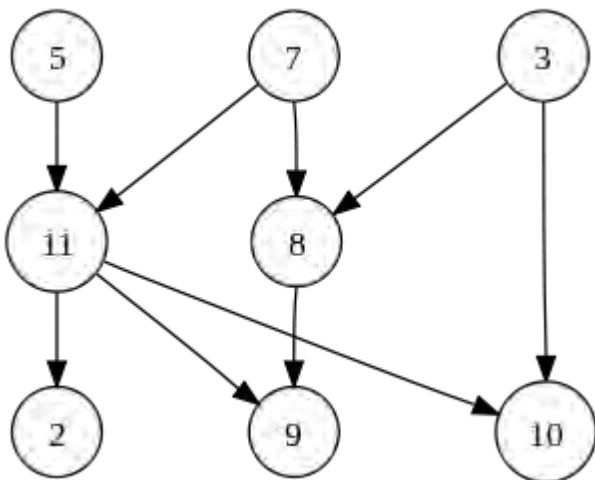
## Εισαγωγή

### 1.1 Μια γενική περιγραφή

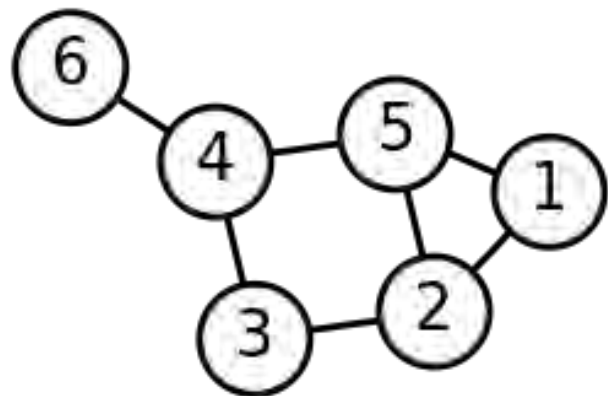
Η κατανόηση της διάδοσης νέων πρακτικών, ιδεών, τεχνολογιών και προϊόντων, αποτελεί ένα από τα βασικά ζητήματα μελέτης της κοινωνικής επιστήμης. Με δεδομένη τη διάρθρωση των κοινωνικών δικτύων, δηλαδή των δικτύων εκείνων που αναπαριστούν την αλληλεπίδραση των ατόμων στην κοινωνία, μπορούμε να κατανοήσουμε και να μελετήσουμε περαιτέρω φαινόμενα γνωστά ως “κοινωνική μετάδοση” ή “διαδοχή πληροφοριών”. Ο άνθρωπος μαθαίνει μέσα από την παρατήρηση και την αντιγραφή συμπεριφορών άλλων ανθρώπων, αν και πολλές φορές, κάτι τέτοιο γίνεται υποσυνείδητα. Ο μιμητισμός αυτός, δεν περιορίζεται μονό σε συμπεριφορές. Το πως αισθανόμαστε, τι πιστεύουμε, εξαρτάται σε μεγάλο βαθμό από τους γύρω μας και επηρεάζει τους ανθρώπους με τους οποίους ερχόμαστε σε επαφή. Αυτά το “δίκτυα επιρροής”, όπως είναι λογικό, συμβαίνει πρώτα τοπικά, σε μικρές κοινότητες, και στη συνέχεια διαδίδεται ευρέως. Για το λόγο αυτό, η μελέτη της κοινωνικής διάδοσης συνδέεται άμεσα με την κατανόηση της δομής των δικτύων.

Με τον όρο δίκτυο εννοούμε τη σύνδεση όμοιων αντικειμένων μεταξύ τους. Δίκτυο υπολογιστών είναι ένα σύνολο υπολογιστών και συσκευών που συνδέονται μεταξύ τους με σκοπό την ανταλλαγή πληροφοριών. Η αναπαράσταση του δικτύου γίνεται με γραφήματα.

Με τον όρο γράφημα εννοούμε μια αναπαράσταση ενός συνόλου στοιχείων όπου μερικά ζευγάρια στοιχείων συνδέονται μεταξύ τους. Τα στοιχεία ονομάζονται κορυφές ενώ οι δεσμοί που συνδέουν τις κορυφές ονομάζονται ακμές. Τα γραφήματα μπορεί να είναι κατευθυνόμενα ή μη-κατευθυνόμενα. Μη-κατευθυνόμενο γράφημα είναι εκείνο στο οποίο δεν υπάρχουν διατεταγμένα ζεύγη (οι ακμές δεν έχουν προσανατολισμό). Η ακμή  $(\alpha, \beta)$  είναι ταυτόσημη με την άκρη  $(\beta, \alpha)$ . Κατευθυνόμενο γράφημα είναι ένα διατεταγμένο ζεύγος  $G = (V, A)$  όπου  $V$ , είναι ένα σύνολο των κορυφών και  $A$  είναι μια σειρά από διατεταγμένα ζεύγη κορυφών, τα οποία ονομάζονται τόξα ή διατεταγμένες ακμές. Ένα τόξο  $a = (x, y)$  θεωρείται ότι κατευθύνεται από το  $x$  στο  $y$  [1]



Εικόνα 1.1



Εικόνα 1.2



## Κεφάλαιο 4

### Το μοντέλο SIR (Susceptible Infectious Recovered)

Πολλές μελέτες έχουν γίνει για την εμφάνιση και για την εξάπλωση μολυσματικών νόσων. Η ικανότητα να κάνει κανείς προβλέψεις για ασθένειες, θα έδινε τη δυνατότητα στους επιστήμονες να αξιολογήσουν τον εμβολιασμό του πληθυσμού ή ακόμη και την απομόνωση των μολυσμένων ατόμων, γεγονός που θα μείωνε σημαντικά το βαθμό θνησιμότητας μιας συγκεκριμένης επιδημίας.

Η μοντελοποίηση των μολυσματικών ασθενειών είναι ένα εργαλείο που έχει χρησιμοποιηθεί για τη μελέτη των μηχανισμών με τους οποίους οι ασθένειες μεταδίδονται, για να προβλέψει τη μελλοντική πορεία μιας νόσου και για να αξιολογήσει στρατηγικές για τον έλεγχο μιας επιδημίας (Daley & Gani, 2005).

Για την μοντελοποίηση της πρόοδο μιας επιδημίας σε ένα πληθυσμό, η ποικιλομορφία του πληθυσμού θα πρέπει να διαχωριστεί με βάση ορισμένα χαρακτηριστικά που σχετίζονται με τη μόλυνση όπως για παράδειγμα, για τις πιο κοινές ασθένειες της παιδικής ηλικίας που προσδίδουν μακροχρόνια ανοσία. Είναι λογικό να κατατάσσουμε τον πληθυσμό σε διάφορα διαμερίσματα ή υποομάδες, που η καθεμία αντιπροσωπεύει ένα συγκεκριμένο στάδιο της επιδημίας. Πιο συγκεκριμένα, έχουμε εκείνους που χαρακτηρίζονται ευπαθείς στην ασθένεια, εκείνους που έχουν μολυνθεί και εκείνους που έχουν αναρρώσει και έχουν ανοσία.

Το μοντέλο SIR, WO Kermack και AG McKendrick[4], είναι ένα απλό μοντέλο που περιγράφει απλοποιημένα τη μετάδοση μολυσματικών ασθενειών μέσα από τα άτομα. Έχει πάρει το όνομα του από τα αρχικά των λέξεων:

**S** = number susceptible (αριθμός ευπαθών ατόμων)

**I** = number infectious (αριθμός μολυσμένων ατόμων)

**R** = number recovered (immune) (αριθμός ανακτηθέντων/αναρρωμένων ατόμων)

Οι αριθμοί αυτοί, μπορούν να μεταβάλλονται με την πάροδο του χρόνου, ακόμη και αν το συνολικό μέγεθος του πληθυσμού  $N$ , παραμένει σταθερό:  $N = S(t) + I(t) + R(t)$ .

Για μια συγκεκριμένη ασθένεια σε ένα πληθυσμό, οι συναρτήσεις  $S(t)$ ,  $I(t)$  και  $R(t)$  μπορούν να προβλέψουν πιθανές εστίες μόλυνσης και να τις θέσουν υπό έλεγχο. Η υποομάδα  $S(t)$  αντιπροσωπεύει τον αριθμό των ατόμων που δεν έχουν προσβληθεί ακόμη από την ασθένεια στο χρόνο  $t$ , ή εκείνους που προσβάλλονται από την ασθένεια. Η υποομάδα  $I(t)$  υποδηλώνει τον αριθμό των ατόμων που έχουν προσβληθεί από την ασθένεια και μπορούν να εξαπλώσουν την ασθένεια στα ευπαθή άτομα. Τέλος, η υποομάδα  $R(t)$  χρησιμοποιείται για εκείνα τα άτομα που έχουν μολυνθεί και στη συνέχεια έχουν “αναρρώσει” από την ασθένεια, είτε λόγω ανοσοποίησης, είτε λόγω θανάτου. Όσοι ανήκουν σε αυτή την κατηγορία δεν είναι σε θέση να μολυνθούν ξανά ή να μεταδώσουν τη νόσο σε άλλους[5].

Αν  $N$  είναι ο συνολικός πληθυσμός έχουμε:

$s(t) = S(t) / N$ , κλάσμα του ευπαθούς πληθυσμού,

$i(t) = I(t) / N$ , κλάσμα του μολυσμένου πληθυσμού, και

$r(t) = R(t) / N$ , κλάσμα του ανακτημένου/αναρρωμένου πληθυσμού.

Το μοντέλο SIR είναι δυναμικό ως προς το χρόνο  $t$ , όπως υποδηλώνουν και οι συναρτήσεις. Η σημασία αυτής της δυναμικής πτυχής είναι πιο προφανής σε ενδημικές ασθένειες, με σύντομο χρονικό διάστημα μόλυνσης. Τέτοιες ασθένειες τείνουν να συμβαίνουν σε κύκλους των εστιών μόλυνσης που οφείλονται στη διακύμανση του

Κοινότητες δικτύου ονομάζονται σύνολα κόμβων, μέσα στα οποία οι κόμβοι συνδέονται μεταξύ τους με μεγαλύτερο αριθμό ακμών απ' ότι συνδέονται τα σύνολα μεταξύ τους.

Σε ένα κοινωνικό κατευθυνόμενο γράφημα και με δεδομένο ένα σύνολο διαδοχικών πληροφοριών του παρελθόντος για το συγκεκριμένο γράφημα, μελετάμε το πρόβλημα εύρεσης κοινοτήτων με τη βοήθεια του μοντέλου CCN (Community-Cascade Network model) και του CiBC αλγορίθμου.

## 1.2 Σκοπός Της Εργασίας

Σκοπός αυτής της εργασίας είναι η μελέτη, η κατανόηση και ο συνδυασμός του μοντέλου CCN (Community-Cascade Network model) και του αλγορίθμου CiBC. Γίνετε αναλυτική περιγραφή καθενός αλγορίθμου, ποιες παραδοχές κάνει ο κάθε ένας, σε ποια είδη γράφων μπορούν να χρησιμοποιηθούν και ποια τμήματα του κάθε αλγορίθμου χρησιμοποιήθηκαν για την ολοκλήρωση της εργασίας. Τέλος, παρατίθενται παραδείγματα που έχουν εκτελέσει τον αλγόριθμο, καθώς και μελέτη των αποτελεσμάτων τους. Όλες οι υλοποιήσεις έχουν πραγματοποιηθεί σε γλώσσα C.

## Κεφάλαιο 2

### Το μοντέλο Community – Cascade Network ( CCN )

#### 2.1 Σύντομη περιγραφή του μοντέλου

Το μοντέλο CCN (Community-Cascade Network model), έχει εφαρμογή τόσο στο κοινωνικό γράφημα που εξετάζουμε, όσο και στο σύνολο των πληροφοριών που έχουμε σχετικά με το παρελθόν, και παράγει επικαλυπτόμενες κοινότητες κόμβων, καθώς και το βαθμό ενεργητικής και παθητικής συμμετοχής του κάθε κόμβου στην κοινότητα την οποία ανήκει [2]. Παρατηρώντας τη διάδοση της πληροφορίας αλλά και τη διάρθρωση του κοινωνικού δικτύου, μπορούν να εξαχθούν συμπεράσματα σχετικά με τη συμπεριφορά των χρηστών μέσα στο δίκτυο.

#### 2.1 Η θεωρία του μοντέλου CCN

Στο μοντέλο CCN έχουμε ως είσοδο ένα κατευθυνόμενο κοινωνικό γράφημα  $G=(N,A)$ , αφού πρόκειται για διάδοση πληροφορίας, και ένα σύνολο πληροφοριών σχετικά με το παρελθόν του δικτύου,  $\log L$ . Πιο συγκεκριμένα, μπορούμε να επωφεληθούμε από το σύνολο των διαδοχικών πληροφοριών σχετικά με το παρελθόν, που έχουμε ήδη στη διάθεση μας, έτσι ώστε να ανιχνεύσουμε τις κοινότητες των κόμβων του γραφήματος

Η ακμή του γραφήματος  $(u,v)$  αντιπροσωπεύει το γεγονός ότι ο χρήστης  $v$  “ακολουθεί” τον χρήστη  $u$ . Με άλλα λόγια, ο χρήστης  $v$  ενημερώνεται για τις δραστηριότητες του χρήστη  $u$ .

Με δεδομένο ένα χρήστη  $u$  και μια κοινότητα  $c$ , βρίσκουμε το βαθμό συμμετοχής του κόμβου στην κοινότητα. Ο βαθμός συμμετοχής χωρίζεται σε δυο παραμέτρους  $\pi < u^{(c,s)}$  και  $\pi < u^{(c,d)}$ . Το  $\pi < u^{(c,s)}$  αντιπροσωπεύει το βαθμό ενεργής συμμετοχής στην κοινότητα, ενώ το  $\pi < u^{(c,d)}$  το βαθμό της παθητικής συμμετοχής στην κοινότητα. Ο όρος ενεργή συμμετοχή του χρήστη στο δίκτυο αναφέρεται στην πιθανότητα αυτός ο χρήστης να αποκτήσει ακολούθους (ακμές που φεύγουν από τον κόμβο του χρήστη και κατευθύνονται προς κάποιον άλλο κόμβο) ή την πιθανότητα ο χρήστης αυτός να δημοσιεύσει κάτι στο διαδίκτυο. Ο όρος παθητική συμμετοχή αναφέρεται στην πιθανότητα ο χρήστης αυτός να έχει εισερχόμενες ακμές η πιθανότητα να επηρεαστεί από άλλους χρήστες μέσα στην κοινότητα.

Στο μοντέλο αυτό, με δεδομένο ένα κατευθυνόμενο κοινωνικό γράφημα  $G = (N, A)$  και ένα σύνολο διαδοχικών πληροφοριών, ξεκάνει η διάδοση της πληροφορίας μέσω των ακμών μιας κοινότητας και διαδίδεται σε άλλες κοινότητες μέσω των χρηστών που είναι επικαλυπτόμενοι, ανήκουν δηλαδή ταυτόχρονα σε περισσότερες από μια κοινότητες.

Το  $(i,u,t) \in L$  αναπαριστά το γεγονός ότι ο χρήστης  $u$  υιοθέτησε το  $i$  τη χρονική στιγμή  $t$ . Με  $D(t)$  αναπαριστούμε το σύνολο των χρηστών που υιοθετούν το  $i$  τη χρονική στιγμή  $t$ .

Θεωρούμε ότι:

$$\exp \{ \pi < u^{(k,s)} \} / \exp \{ \pi < u^{(k,d)} \}$$

είναι ο βαθμός εισόδου ενός κόμβου / βαθμός εξόδου ενός κόμβου

$$\vartheta_u^k = \frac{\exp \{ \pi_u^{k,s} \}}{\sum_{\bar{u} \in N} \exp \{ \pi_{\bar{u}}^{k,s} \}} \quad (1)$$

$$\varphi_u^k = \frac{\exp \{ \pi_u^{k,d} \}}{\sum_{\bar{u} \in N} \exp \{ \pi_{\bar{u}}^{k,d} \}} \quad (2)$$

Εικόνα 2.1

Στο μοντέλο αυτό, παίρνουμε ένα ζευγάρι (αντικείμενο  $i$ , χρονοσφραγίδα  $t$ ) και καταλήγουμε σε μια αλληλουχία ενεργοποιήσεων κόμβων. Συσχετίζουμε σε κάθε ενεργοποίηση  $a$  μια χρονοσφραγίδα  $t < a >$  και μια δράση  $i < a >$

$$\theta_u^{k,a} = \frac{\exp \{ \pi_u^{k,s} \}}{\sum_{u' \in \mathcal{F}_{i_a}(t_a)} \exp \{ \pi_{u'}^{k,s} \}} \quad (3)$$

$$\phi_{u,v}^{k,a} = \frac{\exp \{ \pi_v^{k,d} \}}{\sum_{v': (u,v') \in A, v' \notin C_{i_a}(t_a-1)} \exp \{ \pi_{v'}^{k,d} \}} \quad (4)$$

Εικόνα 2.2

Ο κόμβος εκείνος που θα ασκήσει επιρροή,  $u$ , διαλέγεται από το σύνολο των κόμβων που μπορούν να διαδώσουν το αντικείμενο  $i$ . Αντίστοιχα, ένας κόμβος  $v$  μπορεί να επηρεαστεί ώστε να υιοθετήσει  $i < a >$  μόνο αν συνδέεται με τον κόμβο ο οποίος ασκεί την επιρροή και δεν έχει ακόμη ενεργοποιηθεί.

Η εκτιμώμενη πιθανότητα δίνεται από τον τύπο:

$$Q(\Theta, \Theta') = \sum_{(u,v) \in A} \sum_k \gamma_{u,v,k}(\Theta') \left[ \log \pi_k + \log \vartheta_u^k + \log \varphi_v^k \right] \\ + \sum_{a \in \mathbb{D}} \sum_k \sum_{u \in \mathcal{F}_{i_a, v_a}} \eta_{u,a,k}(\Theta') \left( \log \pi_k + \log \theta_u^{k,a} + \log \phi_{u,v_a}^{k,a} \right)$$

Εικόνα 2.3

$$\gamma_{u,v,k}(\Theta) = P(z_\ell^k | \ell \equiv (u, v) \in A, \Theta) \\ = \frac{\vartheta_u^k \varphi_v^k \pi_k}{\sum_{k'} \vartheta_u^{k'} \varphi_v^{k'} \pi_{k'}} \quad (9)$$

Εικόνα 2.4

$$\begin{aligned}
\eta_{u,a,k}(\Theta) &= P(z_a^k, w_a^u | a \in \mathbb{D}, \Theta) \\
&= \frac{P(a \in \mathbb{D} | w_a^u, z_a^k, \Theta) P(w_a^u | z_a^k, \Theta) P(z_a^k | \Theta)}{P(a \in \mathbb{D} | \Theta)} \\
&= \frac{\phi_{u,v_a}^{k,a} \theta_u^{k,a} \pi_k}{\sum_{k'} \sum_{u' \in \mathcal{F}_{i_a, v_a}} \phi_{u',v_a}^{k',a} \theta_{u'}^{k',a} \pi_{k'}}
\end{aligned} \tag{10}$$

Εικόνα 2.5

$$\pi_k = \frac{\sum_{(u,v) \in A} \gamma_{u,v,k} + \sum_{a \in \mathbb{D}} \sum_{u \in \mathcal{F}_{i_a, v_a}} \eta_{u,a,k}}{m + \sum_{a \in \mathbb{D}} \sum_k \sum_{u \in \mathcal{F}_{i_a, v_a}} \eta_{u,a,k}} \tag{11}$$

Εικόνα 2.6

$$\delta_{\bar{u}}^k = \log \left\{ \frac{\sum_{v: (\bar{u}, v) \in A} \gamma_{\bar{u}, v, k} + \sum_{a \in \mathbb{D}: \bar{u} \in \mathcal{F}_{i_a, v_a}} \eta_{\bar{u}, a, k}}{\vartheta_{\bar{u}}^k \sum_{(u,v) \in A} \gamma_{u,v,k} + \sum_{a \in \mathbb{D}: \bar{u} \in \mathcal{F}_{i_a}(t_a)} \theta_{\bar{u}}^{k,a} \sum_{u \in \mathcal{F}_{i_a, v_a}} \eta_{u,a,k}} \right\} \tag{5}$$

$$\lambda_{\bar{v}}^k = \log \left\{ \frac{\sum_{u: (u, \bar{v}) \in A} \gamma_{u, \bar{v}, k} + \sum_{a: v_a = \bar{v}} \sum_{u \in \mathcal{F}_{i_a, \bar{v}}} \eta_{u,a,k}}{\varphi_{\bar{v}}^k \sum_{(u,v) \in A} \gamma_{u,v,k} + \sum_{a: \bar{v} \notin C_{i_a}(t_a-1)} \sum_{u \in \mathcal{F}_{i_a, v_a}: (u, \bar{v}) \in A} \eta_{u,a,k} \phi_{u, \bar{v}}^{k,a}} \right\} \tag{6}$$

Εικόνα 2.7

Ακολουθεί ο ψευδοκώδικας υπολογισμού των παραμέτρων του μοντέλου CCN

---

**Algorithm 1:** CCN model parameters estimation

---

**Input** : Graph  $G = (N, A)$ , propagation log  $\mathbb{D}$ , and  $K \in \mathbb{N}^+$ .  
**Output**: The set of all parameters,  $\Theta = \{\Pi, \Pi^s, \Pi^d\}$ .

```

init( $\Pi, \Pi^s, \Pi^d$ ); //Initialization of parameters
repeat
  forall the  $k \in \{1, \dots, K\}$  do
    forall the  $\ell \equiv (u, v) \in A$  do
      | compute  $\gamma_{u,v,k}$  according to eq. 9
    end
    E-step
    forall the  $a \equiv (i_a, v_a, t_a) \in \mathbb{D}$  do
      forall the  $u \in N$  do
        | compute  $\eta_{u,a,k}$  according to eq. 10
      end
    end
  end
  forall the  $k \in \{1, \dots, K\}$  do
    compute  $\pi_k$  according to equation 11
    M-step
    forall the  $u \in N$  do
      compute  $\delta_u^k$  and  $\lambda_u^k$  according to eqs. 5-6
       $\pi_u^{k,s(new)} \leftarrow \pi_u^{k,s} + \delta_u^k$ 
       $\pi_u^{k,d(new)} \leftarrow \pi_u^{k,d} + \lambda_u^k$ 
    end
  end
until convergence;

```

---

Εικόνα 2.8

## Κεφάλαιο 3

### Ο αλγόριθμος CIBC

#### 3.1 Σύντομη Περιγραφή Του Προβλήματος

Ο αλγόριθμος CIBC[3] είναι μια πολιτική ανάθεσης δικτύων διανομής περιεχομένου (CDN) που βρίσκει κοινότητες βασισμένες σε ιδιότητες τοπικών και παγκόσμιων γραφών. Με τον όρο δίκτυα διανομής δεδομένων (CDN) αναφερόμαστε σε δίκτυα που έχουν εξισορροπημένο κόστος και ποιότητα παρεχόμενων υπηρεσιών όσον αφορά την παροχή περιεχομένου.

Με τον αλγόριθμο CIBC εντοπίζουμε γενικευμένες κοινότητες ιστοσελίδων από το περιεχόμενο ενός διακομιστή δικτύου. Τα δεδομένα του διακομιστή αναπαριστώνται από ένα γράφημα  $G = (V, E)$ . Οι κόμβοι είναι ουσιαστικά οι ιστοσελίδες, και οι ακμές του γραφήματος είναι οι υπερσυνδέσεις που υπάρχουν ανάμεσα στις ιστοσελίδες.

#### 3.2 Η θεωρία του αλγορίθμου CIBC

Ο αλγόριθμος CIBC χωρίζεται σε δυο φάσεις. Στην πρώτη φάση γίνεται ο υπολογισμός του BC των κόμβων του γραφήματος. Στη δεύτερη φάση, οι κόμβοι συσσωρεύονται γύρω από τους κόμβους με το μικρότερο BC (κόμβοι – πόλοι) και παράγεται ένα σύνολο κοινοτήτων ιστοσελίδων.

Πιο αναλυτικά, στην πρώτη φάση πραγματοποιείται υπολογισμός του BC, παρέχοντας πληροφορίες σχετικά με το πόσο κεντρικός είναι ο συγκεκριμένος κόμβος μέσα στο γράφημα. Το BC αναφέρεται στο ποσοστό πλοήγησης της συγκεκριμένης σελίδας σε σχέση με τις υπόλοιπες σελίδες του διακομιστή. Στη δεύτερη φάση, γίνεται η συσσώρευση των κόμβων του γραφήματος γύρω από τους κόμβους-πόλους. Είναι μια επαναληπτική διαδικασία, όπου σε κάθε βήμα γίνεται ο υπολογισμός των κοινοτήτων από τους κόμβους με χαμηλό σχετικά BC. Οι κόμβοι με μεγάλο BC έχουν μεγάλο αριθμό ακμών που συνδέονται με τον συγκεκριμένο κόμβο (fan-out). Αυτή η διαδικασία τελειώνει όταν όλοι οι κόμβοι θα ανατεθούν σε κάποιο σύνολο. Επόμενο βήμα, το οποίο χρησιμοποιούμε κι εμείς στην εργασία, είναι η ελαχιστοποίηση του αριθμού των συνόλων. Δημιουργείται ένας τετραγωνικός πίνακας μεγέθους  $n \times n$  και ο αριθμός των συνόλων. Κάθε στοιχείο του πίνακα  $B[i, j]$  είναι ο αριθμός των ακμών από το σύνολο  $C_i$  στο σύνολο  $C_j$ . Η διαγώνιος,  $B[i, i]$ , περιέχει το σύνολο των εσωτερικών ακμών του συνόλου  $C_i$ . Σε κάθε επανάληψη, επιλέγεται προς συνένωση εκείνο το ζευγάρι συνόλων για το οποίο ισχύει  $B[i, j] \geq B[i, i]$ . Τελικά, δημιουργούνται κοινότητες με εσωτερικό αριθμό ακμών μέσα στην κοινότητα μεγαλύτερο από τον αριθμό ακμών μεταξύ των κοινοτήτων. Ο αλγόριθμος τερματίζει όταν δεν υπάρχει αλλαγή στον αριθμό των κοινοτήτων.

Ακολουθεί ο ψευδοκωδικας του αλγορίθμου

Εικόνα 3

---

**Algorithm 1** The CiBC algorithm

---

**Input:**  $G(V, E)$ : Web site graph;  
**Output:**  $C = \{C_1, C_2, \dots\}$ : A set of Web page communities

**Phase I: Computation of Betweenness Centrality**

- 1: **for all**  $v \in G$  **do**
- 2:   compute Betweenness Centrality (BC);
- 3: **end for**
- 4: sort  $V$  nodes of  $G$  by ascending BC;

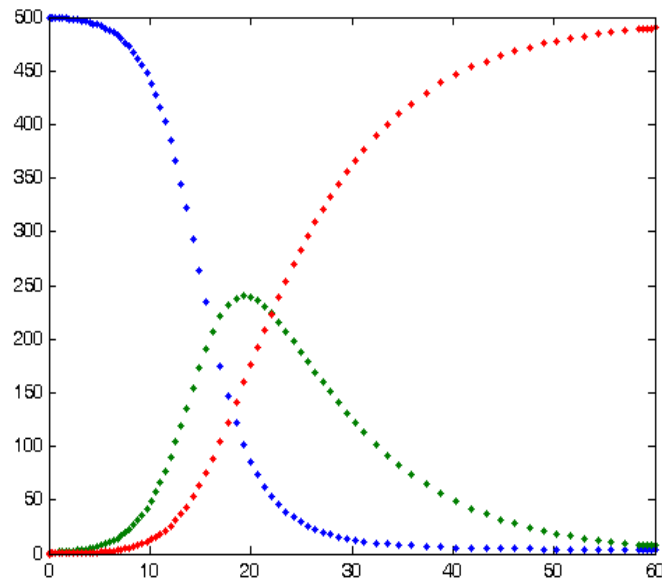
**Phase II: Accumulation around Pole-Nodes**  
 /\* $l$ : the number of groups\*/

- 5:  $l=1$ ;
- 6: **repeat**
- 7:   **for**  $i = 1$  to  $V$  **do**
- 8:     **if**  $V(i) \in C$  **then**
- 9:       continue;
- 10:    **end if**
- 11:     $C_l = C_l \cup \{V(i)\}$ ;
- 12:     $C_l = C_l \cup \{\text{nodes directly connected with } V(i)\}$ ;
- 13:     $C_l = C_l \cup \text{Bounded-BFS}(\text{graph: } G, \text{ depth: } \sqrt{V}, \text{ starting node: } V(i))$ ;
- 14:     $l = l + 1$ ;
- 15:   **end for**
- 16:   **for**  $i = 1$  to  $l$  **do**
- 17:     **for**  $j = 1$  to  $l$  **do**
- 18:        $B[i, j] = \text{number of edges from } C_i \text{ to } C_j$ ;
- 19:        $B[i, i] = \text{number of internal edges of } C_i$ ;
- 20:     **end for**
- 21:    **end for**
- 22:    Find  $(C_x, C_y) = \text{maximum}(\frac{B[x, y]}{B[x, x]})$ ;
- 23:    **if** (size of  $C_x \cup C_y \leq \text{cache size of surrogate}$ ) and  $(B[x, y] \geq B[x, x])$  **then**
- 24:      $C_x \leftarrow C_x \cup C_y$ ;
- 25:     delete  $C_y$ ;
- 26:      $l = l - 1$ ;
- 27:    **end if**
- 28:    **for**  $i = 1$  to  $l$  **do**
- 29:      $V_{C_i} = \text{nodes in } C_i$ ;
- 30:     /\*Degenerate cases of communities are removed\*/
- 31:     **if** (number of  $V_{C_i} \leq 3$ ) or (size of  $V_{C_i} > \text{cache size of surrogate}$ ) **then**
- 32:        $V_{C_i} \rightarrow \text{unassigned nodes}$ ;
- 33:       delete  $C_i$ ;
- 34:        $l = l - 1$ ;
- 35:     **end if**
- 36:    **end for**
- 37:     $C_{l+1} = \{V \notin C\}$ ;
- 38: **until** size of  $C_{l+1} > \text{cache size of surrogate}$

---

αριθμού των ευπαθών ατόμων (susceptible) κατά τη διάρκεια του χρόνου. Κατά τη διάρκεια μιας επιδημίας, ο αριθμός των ευπαθών ατόμων πέφτει ραγδαία καθώς περισσότεροι από αυτούς έχουν μολυνθεί. Η νόσος δεν μπορεί να ξεσπάσει ξανά έως ότου ο αριθμός των ευπαθών ατόμων (susceptible) να αυξηθεί ξανά. Ο μόνος τρόπος για ένα άτομο να μην ανήκει στην ευπαθή ομάδα, είναι να μολυνθεί.

Στο μοντέλο SIR, ένα άτομο του πληθυσμού θεωρείται ότι έχει ίση πιθανότητα με οποιοδήποτε άλλο άτομο να προσβληθεί από την ασθένεια, με ποσοστό  $\beta$ , το οποίο είναι το ποσοστό επαφών ανά μονάδα χρόνου που επαρκούν για να εξαπλωθεί η ασθένεια. Ως εκ τούτου, ένα μολυσμένο άτομο έρχεται σε επαφή και είναι σε θέση να μεταδώσει τη νόσο σε  $\beta N$  άλλους ανά μονάδα χρόνου. Δεν είναι όμως όλες αυτές οι επαφές με ευπαθή άτομα. Αν υποθέσουμε μια ομοιογενή ανάμιξη του πληθυσμού, το ποσοστό αυτών των επαφών που είναι με ευπαθή άτομα (susceptibles) είναι  $s(t) = S(t) / N$ . Ο αριθμός των νέων μολύνσεων ανά μονάδα χρόνου ανά μολυσμένο άτομο είναι  $\beta N (S / N)$ , με ρυθμό νέων μολύνσεων (ή ατόμων που “εγκαταλείπουν” την ευπαθή κατηγορία)  $\beta N (S / N) I = \beta SI$ .



Εικόνα 4

Μπλε = ευπαθείς, πράσινο = μολυσμένοι, κόκκινο = αναρρωμένοι



## Κεφάλαιο 5

### Ο αλγόριθμος epidemic-CIBC

#### 5.1 Περιγραφή του αλγορίθμου που υλοποιήσαμε

Στη διπλωματική αυτή εργασία, στόχος είναι η εύρεση κοινοτήτων. Έχοντας ως έναυσμα το μοντέλο CCN που περιγράψαμε προηγουμένως, και με χρήση της δεύτερης φάσης του αλγορίθμου CIBC, ο αλγόριθμος που υλοποιήσαμε, δέχεται ως είσοδο ένα γράφημα  $G=(N,A)$ , που αντιπροσωπεύει έναν πληθυσμό, και ένα σύνολο πληροφοριών σχετικά με το παρελθόν  $\log L$ . Το σύνολο  $\log L$  είναι ουσιαστικά κάποιες αρχικές, μικρές κοινότητες. Θεωρούμε ότι οι αρχικές αυτές κοινότητες αντιστοιχούν σε σύνολο ατόμων του πληθυσμού που έχουν μολυνθεί από κάποια νόσο. Στην παρούσα διπλωματική, έχουμε ακολουθήσει το μοντέλο διάχυσης κακόβουλου περιεχομένου SIR, με μικρή πιθανότητα μόλυνσης των γειτονικών ευπαθών κόμβων  $\lambda$ , έτσι ώστε η μόλυνση να μην επεκτείνεται σε μεγάλο κομμάτι του δικτύου, αλλά να μένει στην γειτονιά του κόμβου απο όπου ξεκίνησε. Θέλουμε να επωφεληθούμε από το σύνολο των διαδοχικών πληροφοριών σχετικά με το παρελθόν, που έχουμε ήδη στη διάθεση μας, έτσι ώστε να ανιχνεύσουμε τις τελικές κοινότητες των κόμβων του γραφήματος.

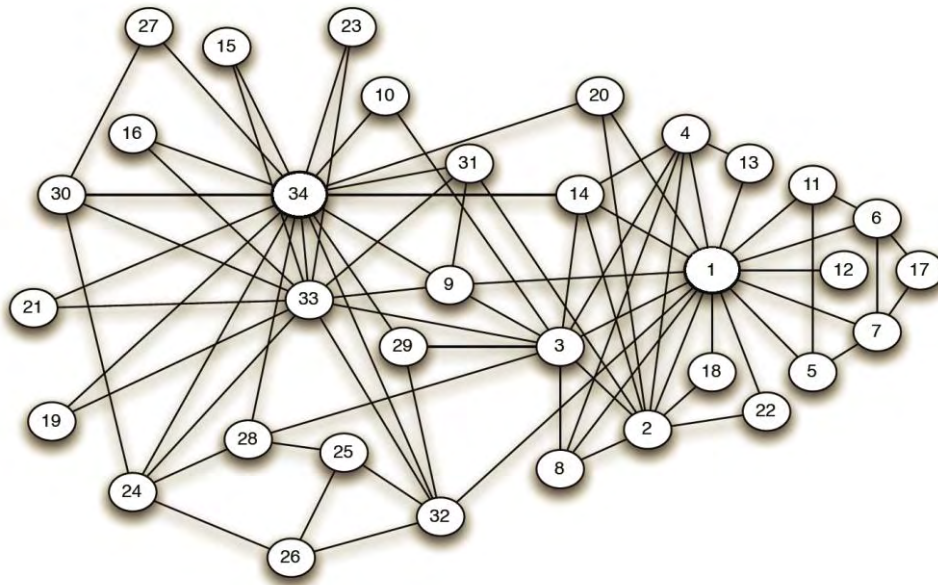
Πιο συγκεκριμένα, ο αλγόριθμος μας δέχεται ως είσοδο ένα γράφημα και ένα σύνολο αρχικών κοινοτήτων. Δημιουργεί μια λίστα με τους κόμβους όλου του γραφήματος, για τον κάθε κόμβο διατηρεί τον αριθμό συνδέσεων του και τις συνδέσεις του με άλλους κόμβους. Στη συνέχεια, δημιουργεί έναν τετραγωνικό πίνακα  $B[i, j]$  μεγέθους όσος και ο αριθμός των αρχικών κοινοτήτων. Κάθε στοιχείο του πίνακα  $B[i, j]$  είναι ο αριθμός των ακμών από το σύνολο  $C_i$  στο σύνολο  $C_j$ . Η διαγώνιος,  $B[i, i]$ , περιέχει το σύνολο των εσωτερικών ακμών του συνόλου  $C_i$ . Σε κάθε επανάληψη, επιλέγεται προς συνένωση εκείνο το ζευγάρι συνόλων για το οποίο ισχύει  $B[i, j] \geq B[i, i]$  και τελικά δημιουργούνται κοινότητες με εσωτερικό αριθμό ακμών μεγαλύτερο από τον αριθμό ακμών προς τις υπόλοιπες κοινότητες. Ο αλγόριθμος τερματίζει όταν δεν υπάρχει αλλαγή στον αριθμό των κοινοτήτων. Επιπλέον, όταν ολοκληρώνεται η διαδικασία της ομαδοποίησης των κοινοτήτων, ελέγχονται εκ νέου οι κοινότητες οι οποίες έχουν δημιουργηθεί και διαγράφονται εκείνες που περιλαμβάνουν πολύ μικρό αριθμό κόμβων (μέχρι 3).

#### 5.2 Τρεξίματα του αλγορίθμου σε μικρά παραδείγματα

##### Παράδειγμα 1 (Zachary's karate club)

Οι αρχικές κοινότητες είναι:

```
|1|---> 17,6,7,
|2|---> 5,11,12,1,13,22,
|3|---> 2,4,18,20,14,
|4|---> 3,8,31,9,10,
|5|---> 32,29,25,26,
|6|---> 34,23,15,27,16,30,
|7|---> 33,21,19,24,28,
```



Εικόνα 5.1

Ο πίνακας B στην τελευταία εκτέλεση είναι:

```
38 | 10 |
-----
10 | 29 |
-----
```

Μετά την εκτέλεση του αλγορίθμου, οι κοινότητες που προκύπτουν είναι:  
 3,8,31,9,10,2,4,18,20,14,17,6,7,5,11,12,1,13,22,  
 32,29,25,26,34,23,15,27,16,30,33,21,19,24,28,

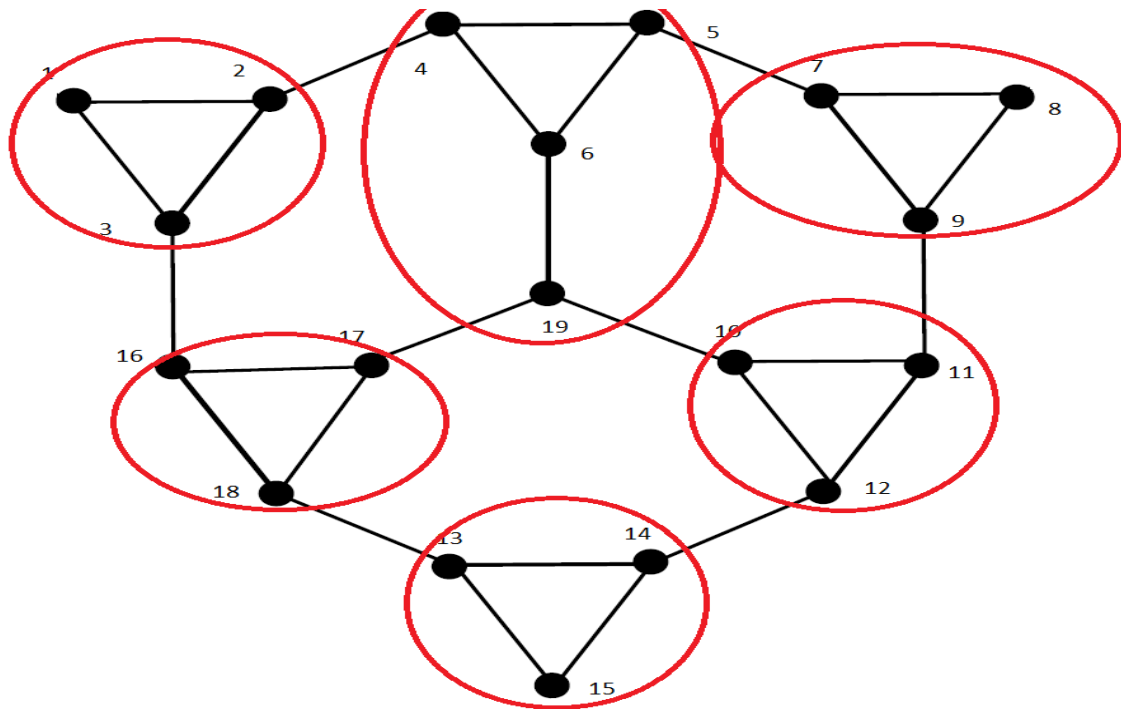
χρόνοι  
 real 0m0.006s  
 user 0m0.000s  
 sys 0m0.004s

παράδειγμα 2 (cubic planar graph)

Ο πίνακας B στην τελευταία εκτέλεση είναι:

```
3 | 1 | 0 | 0 | 0 | 1 |
-----
1 | 4 | 1 | 1 | 0 | 1 |
-----
0 | 1 | 3 | 1 | 0 | 0 |
-----
0 | 1 | 1 | 3 | 1 | 0 |
-----
0 | 0 | 0 | 1 | 3 | 1 |
-----
1 | 1 | 0 | 0 | 1 | 3 |
-----
```

Οι αρχικές κοινότητες είναι ζωγραφισμένες στο γράφημα με κόκκινο χρώμα.



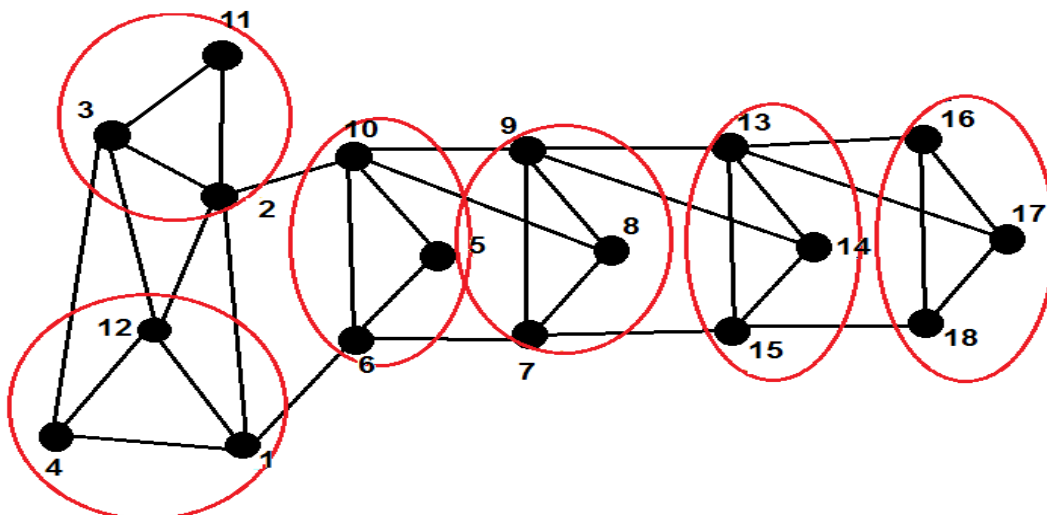
Εικόνα 5.2

Μετά την εκτέλεση του αλγορίθμου, οι κοινότητες που προκύπτουν είναι: 4,5,6,19,

χρόνοι  
 real 0m0.003s  
 user 0m0.000s  
 sys 0m0.000s

παράδειγμα 3

Οι αρχικές κοινότητες είναι ζωγραφισμένες στο γράφημα με κόκκινο χρώμα.



Εικόνα 5.3

Ο πίνακας B στην τελευταία εκτέλεση είναι:

10 | 2 |

-----

2 | 21 |

-----

Μετά την εκτέλεση του αλγορίθμου, οι κοινότητες που προκύπτουν είναι:

1,12,4,2,3,11,

16,17,18,13,14,15,5,6,10,7,8,9,

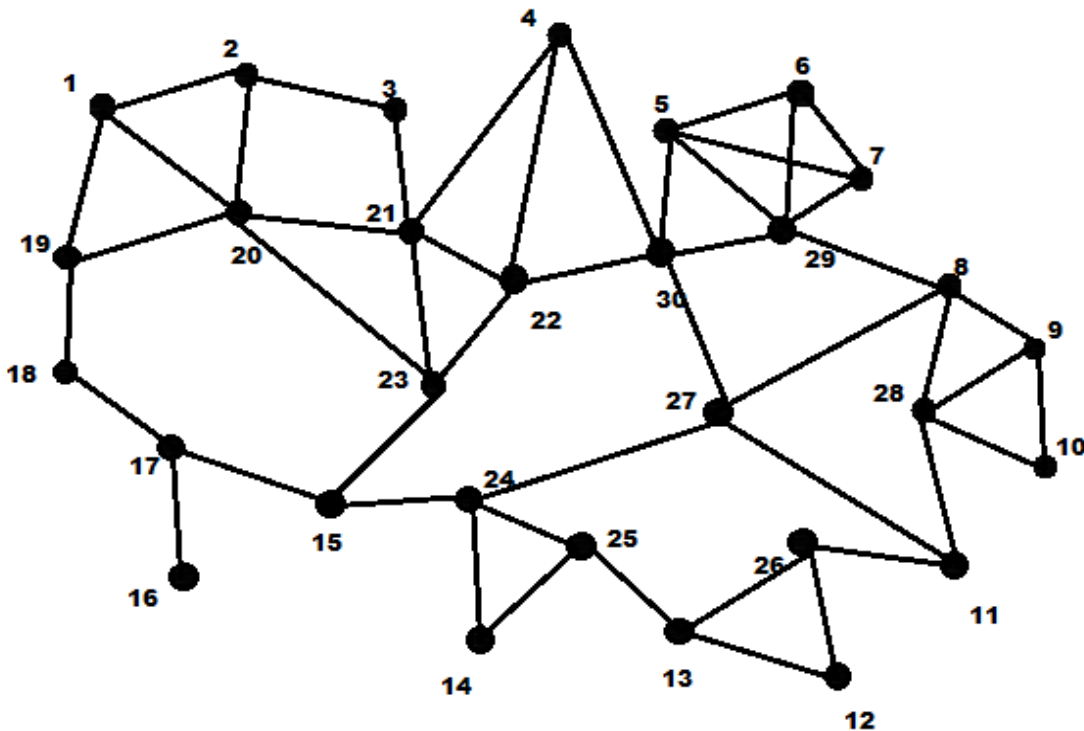
χρόνοι

real 0m0.003s

user 0m0.000s

sys 0m0.000s

παράδειγμα 4



Εικόνα 5.4

Οι αρχικές κοινότητες είναι:

|1|---> 1,2,20,19,  
 |2|---> 2,1,3,21,  
 |3|---> 3,2,21,  
 |4|---> 4,21,22,30,  
 |5|---> 5,6,7,29,30,  
 |6|---> 6,7,29,30,  
 |7|---> 7,5,6,29,  
 |8|---> 8,29,9,28,  
 |9|---> 9,8,10,28,11,  
 |10|---> 10,28,9,  
 |11|---> 11,28,27,26,  
 |12|---> 12,13,26,11,25,  
 |13|---> 13,12,26,25,  
 |14|---> 14,25,24,13,  
 |15|---> 15,24,23,17,  
 |16|---> 16,17,18,  
 |17|---> 17,16,15,18,  
 |18|---> 18,17,19,20,1,  
 |19|---> 19,18,20,1,  
 |20|---> 20,19,2,21,  
 |21|---> 21,20,22,23,3,  
 |22|---> 22,21,23,30,4,  
 |23|---> 23,20,21,22,15,  
 |24|---> 24,27,15,25,14,  
 |25|---> 25,24,14,13,  
 |26|---> 26,13,12,11,  
 |27|---> 27,30,8,11,24,  
 |28|---> 28,8,9,10,11,  
 |29|---> 29,5,6,7,8,30,  
 |30|---> 30,4,5,29,27,22,

Μετά την εκτέλεση του αλγορίθμου, οι κοινότητες που προκύπτουν είναι:

16,17,18,3,2,21,1,20,19,  
 15,24,23,17,2,1,3,21,18,19,20,  
 4,21,22,30,17,16,15,18,20,23,3,  
 5,6,7,29,30,  
 6,7,29,30,5,  
 8,29,9,28,5,6,7,30,  
 9,8,10,28,11,  
 10,28,9,27,30,8,11,24,  
 11,28,27,26,8,9,10,  
 12,13,26,11,25,24,27,15,14,  
 13,12,26,25,  
 14,25,24,13,  
 19,18,20,1,2,21,  
 22,21,23,30,4,  
 23,20,21,22,15,  
 25,24,14,13,  
 26,13,12,11,  
 30,4,5,29,27,22,

και ο πίνακας B στην τελευταία εκτέλεση είναι:

```

12|4|6|0|0|0|0|0|0|1|0|0|0|5|5|0|0|2|
-----
4|17|5|1|1|1|0|2|1|3|1|2|0|4|2|2|0|5|
-----
6|5|16|2|2|2|0|3|1|3|0|2|5|0|0|2|0|3|
-----
0|1|2|8|0|1|1|3|2|2|0|1|0|2|1|1|0|3|
-----
0|1|2|0|8|1|1|3|2|2|0|1|0|2|1|1|0|3|
-----
0|1|2|1|1|12|3|5|4|3|0|1|0|2|1|1|1|3|
-----
0|0|0|1|1|3|6|1|2|2|1|0|0|0|0|0|1|2|
-----
0|2|3|3|3|5|1|11|2|5|3|2|0|2|2|2|2|5|
-----
0|1|1|2|2|4|2|2|10|3|2|2|0|1|0|2|2|2|
-----
1|3|3|2|2|3|2|5|3|13|0|0|0|3|1|0|0|2|
-----
0|1|0|0|0|0|1|3|2|0|4|3|0|0|0|3|1|1|
-----
0|2|2|1|1|1|0|2|2|0|3|5|0|1|1|0|2|2|
-----
0|0|5|0|0|0|0|0|0|0|0|0|8|4|3|0|0|2|
-----
5|4|0|2|2|2|0|2|1|3|0|1|4|7|3|1|0|3|
-----
5|2|0|1|1|1|0|2|0|1|0|1|3|3|6|1|0|3|
-----
0|2|2|1|1|1|0|2|2|0|3|0|0|1|1|5|2|2|
-----
0|0|0|0|0|1|1|2|2|0|1|2|0|0|0|2|4|2|
-----
2|5|3|3|3|3|2|5|2|2|1|2|2|3|3|2|2|7|
-----

```

χρόνοι

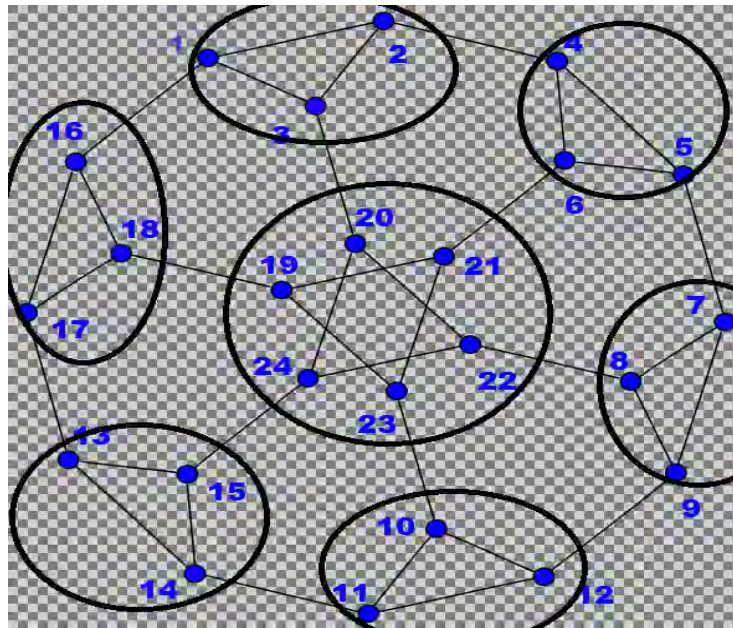
real 0m0.041s

user 0m0.032s

sys 0m0.004s

παράδειγμα 5 (Truncated cubical graph)

Οι αρχικές κοινότητες είναι ζωγραφισμένες στο γράφημα με μαύρο χρώμα.



Εικόνα 5.5

Ο πίνακας B στην τελευταία εκτέλεση είναι:

```

3 | 1 | 0 | 0 | 0 | 1 | 1 |
-----
1 | 3 | 1 | 0 | 0 | 0 | 1 |
-----
0 | 1 | 3 | 1 | 0 | 0 | 1 |
-----
0 | 0 | 1 | 3 | 1 | 0 | 1 |
-----
0 | 0 | 0 | 1 | 3 | 1 | 1 |
-----
1 | 0 | 0 | 0 | 1 | 3 | 1 |
-----
1 | 1 | 1 | 1 | 1 | 1 | 6 |

```

Μετά την εκτέλεση του αλγορίθμου, οι κοινότητες που προκύπτουν είναι:  
19,20,21,22,23,24,

χρόνοι  
real 0m0.003s  
user 0m0.000s  
sys 0m0.000s

## Βιβλιογραφία:

[1] Γράφος from Wikipedia, <http://en.wikipedia.org/wiki/Γράφος>

[2] Cascade-based Community Detection by Nicola Barbieri, Francesco Bonchi and Giuseppe Manco link: [http://www.francescobonchi.com/CCN\\_wsdm13.pdf](http://www.francescobonchi.com/CCN_wsdm13.pdf)

[3] CDNs Content Outsourcing via Generalized Communities by Dimitrios Katsaros, George Pallis, Konstantinos Stamos, Athena Vakali, Antonis Sidiropoulos, and Yannis Manolopoulos [http://inf-server.inf.uth.gr/~dkatsar/papers/IEEE\\_TKDE08kpsvsm.pdf](http://inf-server.inf.uth.gr/~dkatsar/papers/IEEE_TKDE08kpsvsm.pdf)

[4] Epidemic model from Wikipedia, [http://en.wikipedia.org/wiki/Epidemic\\_model](http://en.wikipedia.org/wiki/Epidemic_model)

[5] Compartmental models in epidemiology from Wikipedia, link: [http://en.wikipedia.org/wiki/Compartmental\\_models\\_in\\_epidemiology](http://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology)



## Βιβλιογραφία Εικόνων:

[1] Γράφος from Wikipedia, <http://en.wikipedia.org/wiki/Γράφος>

[2] Γράφος from Wikipedia, <http://en.wikipedia.org/wiki/Γράφος>

[3] Cascade-based Community Detection Nicola Barbieri, Francesco Bonchi and Giuseppe Manco link: [http://www.francescobonchi.com/CCN\\_wsdm13.pdf](http://www.francescobonchi.com/CCN_wsdm13.pdf)

[4] Cascade-based Community Detection Nicola Barbieri, Francesco Bonchi and Giuseppe Manco link: [http://www.francescobonchi.com/CCN\\_wsdm13.pdf](http://www.francescobonchi.com/CCN_wsdm13.pdf)

[5] Cascade-based Community Detection Nicola Barbieri, Francesco Bonchi and Giuseppe Manco link: [http://www.francescobonchi.com/CCN\\_wsdm13.pdf](http://www.francescobonchi.com/CCN_wsdm13.pdf)

[6] Cascade-based Community Detection Nicola Barbieri, Francesco Bonchi and Giuseppe Manco link: [http://www.francescobonchi.com/CCN\\_wsdm13.pdf](http://www.francescobonchi.com/CCN_wsdm13.pdf)

[7] Cascade-based Community Detection Nicola Barbieri, Francesco Bonchi and Giuseppe Manco link: [http://www.francescobonchi.com/CCN\\_wsdm13.pdf](http://www.francescobonchi.com/CCN_wsdm13.pdf)

[8] Cascade-based Community Detection Nicola Barbieri, Francesco Bonchi and Giuseppe Manco link: [http://www.francescobonchi.com/CCN\\_wsdm13.pdf](http://www.francescobonchi.com/CCN_wsdm13.pdf)

[9] Cascade-based Community Detection Nicola Barbieri, Francesco Bonchi and Giuseppe Manco link: [http://www.francescobonchi.com/CCN\\_wsdm13.pdf](http://www.francescobonchi.com/CCN_wsdm13.pdf)

[10] CDNs Content Outsourcing via Generalized Communities by Dimitrios Katsaros, George Pallis, Konstantinos Stamos, Athena Vakali, Antonis Sidiropoulos, and Yannis Manolopoulos [http://inf-server.inf.uth.gr/~dkatsar/papers/IEEE\\_TKDE08kpsvsm.pdf](http://inf-server.inf.uth.gr/~dkatsar/papers/IEEE_TKDE08kpsvsm.pdf)

[11] Compartmental models in epidemiology from Wikipedia, link: [http://en.wikipedia.org/wiki/Compartmental\\_models\\_in\\_epidemiology](http://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology)

[12] Haystack Blog MIT CSAIL Research (Group Website), link: <http://haystack.csail.mit.edu/blog/2010/07/12/data-by-the-people-for-the-people/>

[13] MathOverflow is a question and answer site for professional mathematicians. It's 100% free, no registration required, link: <http://mathoverflow.net/questions/123990/does-the-cubic-planar-graph-with-6-3-faces-and-6-7-faces-have-a-name>

[14] File:Truncated cubical graph.neato.svg from Wikimedia Commons link: [http://commons.wikimedia.org/wiki/File:Truncated\\_cubical\\_graph.neato.svg](http://commons.wikimedia.org/wiki/File:Truncated_cubical_graph.neato.svg)