



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής  
Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Πληροφορική»

**Μεταπτυχιακή Διατριβή**

Τίτλος Διατριβής	<b>(Ελληνικά)</b> Αλγοριθμική άντληση δομικής πληροφορίας από ψηφιακά κοινωνικά δίκτυα και υπολογισμός μέτρων κεντρικότητας. <b>(Αγγλικά)</b> Algorithmic extraction of structural information from social networks and computation of associated centrality measures.
Όνοματεπώνυμο Φοιτητή	<b>ΚΩΝΣΤΑΝΤΙΝΟΣ ΖΟΡΖΟΜΙΚΟΣ</b>
Πατρώνυμο	<b>ΔΗΜΗΤΡΙΟΣ</b>
Αριθμός Μητρώου	<b>ΜΠΠΛ13020</b>
Επιβλέπων	<b>ΤΣΙΧΡΙΝΤΖΗΣ ΓΕΩΡΓΙΟΣ, ΚΑΘΗΓΗΤΗΣ</b>

**Τριμελής Εξεταστική Επιτροπή**

(υπογραφή)

(υπογραφή)

(υπογραφή)

Όνομα Επώνυμο  
Βαθμίδα

**Γ.Τσιχριντζής**

**Καθηγητής**

Όνομα Επώνυμο  
Βαθμίδα

**Χ. Κωνσταντόπουλος**

**Επίκουρος**

Όνομα Επώνυμο  
Βαθμίδα

**Α.Πικράκης**

**Επίκουρος**



## ΠΕΡΙΛΗΨΗ

Η Μεταπτυχιακή μας διατριβή έχει σαν τίτλο: «Αλγοριθμική άντληση δομικής πληροφορίας από ψηφιακά κοινωνικά δίκτυα και υπολογισμός μέτρων κεντρικότητας.» Με πιο απλά λόγια, στα πλαίσια της εργασίας αυτής, καταφέραμε να πάρουμε δεδομένα από ένα κοινωνικό δίκτυο (συγκεκριμένα από το Twitter), και μετά από κάποια επεξεργασία των δεδομένων αυτών, βγάλαμε έναν γράφο. Ο Γράφος αυτός αναπαριστά ένα μικρό κομμάτι του δικτύου. Για να είμαστε ακριβείς αναπαριστά το δίκτυο που δημιουργείτε ξεκινώντας από εμάς, μέχρι και τους Followers των Followers μας. Δηλαδή μέχρι και τους συγγενείς 2<sup>ου</sup> βαθμού στο δίκτυο. Στα πλαίσια της μεταπτυχιακής διατριβής, η εξόρυξη γνώσης, από το δίκτυο, έγινε γράφοντας κώδικα σε Python. Και η αναπαράσταση του γράφου έγινε με την βοήθεια του εργαλείου Gephi. Δώθηκε μεγάλη έμφαση κατά την εξόρυξη, ώστε να μην ξαναπάρουμε κόμβους, που τους έχουμε ξαναπάρει, γιατί ο σκοπός μας ήταν να δημιουργήσουμε έναν γράφο, όπου ο κάθε κόμβος να αναπαριστά ένα μοναδικό χρήστη του Twitter. Η μεταπτυχιακή διατριβή αυτή, θα μπορούσαμε να πούμε πως χωρίζεται σε 3 μέρη. Πρώτο, η εξόρυξη γνώσεις από ένα κοινωνικό δίκτυο, και η εισαγωγή των αποτελεσμάτων σε ένα αρχείο .txt. Δεύτερο μέρος, η ειατροπή του αρχείου .txt σε αρχείο μορφής .gml, ώστε να διαβάζεται στο Gephi. Και τρίτο μέρος, η εισαγωγή του αρχείου .gml στο Gephi, και η μελέτη του γράφου που βγαίνει, με σκοπό να μελετήσουμε, κυρίως τα μέτρα κεντρικότητας.

Τα μέτρα κεντρικότητας, δηλαδή το πόσο 'κεντρικός ή σημαντικός είναι ένας κόμβος σε ένα γράφο, είναι κάτι που απασχολεί πολλούς ανθρώπους, ανα τον κόσμο, διαφορετικών κλάδων. Για παράδειγμα οι συγκοινωνιολόγοι, μελετάνε τα μέτρα κεντρικότητας, για να βγάλουν σε ένα οδικό δίκτυο μιας πόλης, τα πιο κεντρικά σημεία, σημεία που οι πιθανότητες να περάσουν η πλειοψηφία του πληθυσμού είναι αρκετά υψηλή. Οι αεροπορικές εταιρίες μελετάνε επίσης τα μέτρα κεντρικότητας των αεροπορικών γραμμών, βλέποντας την μετακίνηση των επιβατών, αλλά και τις αλλαγές στρατηγικής των άλλων εταιριών. Με αυτό τον τρόπο (με τα μέτρα κεντρικότητας δηλαδή) οι αεροπορικές εταιρίες καταλαβαίνουν τους πιο 'εμπορικούς' αερολιμένες για αυτούς, πληροφορία πολύ σημαντική για το πως θα δράσει η εταιρία στο άμεσο μέλλον. Με βάση τα μέτρα κεντρικότητας, και όχι μόνο, πολλές φορές διαλέγουν να συγχωνευθούν πολλές αεροπορικές εταιρίες. Ελπίζοντας πως με λίγους, αλλά κεντρικούς αερολιμένες, θα μπορούν να καλύψουν, σχεδόν όλο το δίκτυο.

Θα ήταν πάρα πολύ ενδιαφέρον να δούμε το πως θα εξελισσόταν, το δίκτυο μας, συναρτήση του χρόνου, αλλά αυτό ξεφεύγει από τα στενά όρια της μεταπτυχιακής διατριβής, και μπαίνει στα όρια της έρευνας. Σε ένα μετέπειτα βήμα, αν βλέπαμε πως θα άλλαζε το δίκτυο μας, στην πάροδο του χρόνου, ίσως να είχαμε την δυνατότητα να κάνουμε και προβλέψεις. Και σε περίπτωση που φτάναμε σε σημείο να προβλέπουμε την ανάπτυξη του δικτύου, θα είχαμε ανακαλύψει μια νέα γνώση, η οποία θα μπορούσε να χρησιμοποιηθεί και σε άλλους κλάδους εκτός της πληροφορικής. Για παράδειγμα στην ιατρική, να προβλέπει επιδημίες, σε ποιά σημεία και προς τα που τείνει να αναπτυχθεί. Αυτό θα μπορούσε να γίνει, ίσως, σε συνέχεια της μεταπτυχιακής αυτής διατριβής, σε διδακτορικό επίπεδο. Γιατί η ανάπτυξη κώδικα για ένα αρκετά μεγάλο κομμάτι του δικτύου του Twitter, αλλά και η μελέτη του αντίστοιχου γράφου, η παρατήρηση του γράφου, συναρτήση του χρόνου, και η σύγκριση των αποτελεσμάτων ανα χρονικές περιόδους, είναι μια αρκετά σύνθετη δουλειά, και πάρα πολύ χρονοβόρα.

Αυτό που θέλαμε όμως να πετύχουμε, και το πετύχαμε σε αυτήν την εργασία, είναι πως η ιδέα μας, έστω σε αυτό το μικρό δίκτυο, δουλεύει. Μπορέσαμε και 'τραβήξαμε' δεδομένα από το κοινωνικό δίκτυο, και με βάση αυτά φτιάξαμε ένα γράφο. Από τον πολυδιάστατο αυτόν γράφο βγάλαμε σημαντικά συμπεράσματα, όπως τα κριτήρια με τα οποία, κάποιος αποφασίζει να 'ακολουθήσει' κάποιον, στο κοινωνικό δίκτυο του Twitter. Επίσης, βγάλαμε και στατιστικά στοιχεία, όπως τον μέσο όρο σύνδεσης ανά κόμβο, τις πιθανότητες που έχουμε κάποιος συγγενείς 2<sup>ου</sup> βαθμού να μας κάνει Follow κ.τ.λ

## ABSTRACT

Our Master thesis has the title: «Algorithmic extraction of structural information from social networks and computation of associated centrality measures». In simpler words, in the context of this work, we managed to get data from one social network (specifically from Twitter), and after some processing of such data, we took a graph. The graph represents a small part of the network. To be precise, represents the network that is creating, by starting from our Twitter profile, to our Followers of our Followers. That is up to the 2nd degree relatives in the network. As part of postgraduate thesis, we were writing code in Python. In this way, we have the ability to get some data from the network. The representation of the graph was made with the help of Gephi tool. It was given great emphasis in data mining. We do not take nodes that have them retake, because our goal was to create a graph, where each node represents a unique user of Twitter. This master thesis, we could say that is divided into 3 parts. First, the extraction of knowledge from a social network, and the introduction of the results in a file .txt. Second part, the conversion from .txt format file to .gml file, Because Gephi can read only .gml files. And the third part, the introduction of file .gml in Gephi, and the study of the graph coming out, especially the centrality measures.

The centrality measures, is how 'central' or important is a node in a graph, it is something that concerns many people, around the world, many times at different industries. For example, transportation, studying the centrality measures, to make a road of a city, the most central points, points that are likely to spend the majority of the population is quite high. Airlines are also studying the centrality measures of air lines, watching the movement of the passengers, but also the strategic changes of other companies. In this way (by using measures centrality) the airlines understand the more 'commercial' airports for them, very important information about how the company will act in the near future. By using the centrality measures, of course and others parameters, often choose to merge many airlines. Hoping that with few, but hubs, nodes you can cover almost the entire network. It would be very interesting if we could see the changing of our network, step by step, function of time, but this is outside the confines of the master thesis, and enters the boundaries of research. In a subsequent step, if we could see how our network is changing, function of time, we might have the ability to do and forecasts. And if we could predict the development of the network, we would have discovered a new knowledge, which could be used in other industries outside of IT. For example in medicine, provide epidemics. Whereas, epidemics grow up in the similar way as a social network. This could be, perhaps, following this master thesis at the doctoral level. Because, develop code for a sizable chunk of the network of Twitter, and the study of the corresponding graph, the observation of the graph, function of time and the comparison of results particular moments, is a quite complex job, and too time consuming.

What we wanted to achieve, and we succeeded in this master thesis, is that our idea, even in this small network, it works. We were able and 'pulled' data from the social network, and based on these data, we built a graph. From the multidimensional this graph we took important lessons, such as the criteria by which someone decided to 'follow' someone, in a social network, for example on Twitter network. Also, we took and statistics results, such as average connection per node, the chances that we have, one second-degree relative Follow us on Twitter etc.

**ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ**

<b>ΠΕΡΙΛΗΨΗ</b> .....	3
<b>Abstract</b> .....	4
<b>ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ</b> .....	5
<b>ΕΙΣΑΓΩΓΗ</b> .....	6
<b>Κεφάλαιο 1.</b> .....	8
<b>Τι είναι τα κοινωνικά δίκτυα. Πως λειτουργούν. Σχετικές επιστήμες, σχετικές έρευνες. Μεγέθη.</b> .....	8
<b>Κεφάλαιο 2.</b> .....	10
<b>Η ιδέα μας. Τι θέλουμε να πετύχουμε, πως και γιατί.</b> .....	10
<b>Κεφάλαιο 3.</b> .....	14
<b>Κοινωνική επιλογή. Αλγόριθμοι αναζήτησης πληροφοριών σε κοινωνικά δίκτυα. Επιδημίες. Παραδείγματα εξεγέρσεων. Social Media Analytics.</b> .....	14
<b>Κοινωνική επιλογή</b> .....	14
<b>Αλγόριθμοι αναζήτησης πληροφοριών σε κοινωνικά δίκτυα</b> .....	14
<b>ΕΠΙΔΗΜΙΕΣ</b> .....	15
<b>Social Media Analytics</b> .....	15
<b>Κεφάλαιο 4.</b> .....	17
<b>Twitter API – Python – Gephi</b> .....	17
<b>Εισαγωγικά</b> .....	17
<b>Twitter API</b> .....	17
<b>GEPI</b> .....	22
<b>ΚΕΦΑΛΑΙΟ 5</b> .....	24
<b>GEPI – ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ – ΜΕΤΡΑ ΚΕΝΤΡΙΚΟΤΗΤΑΣ</b> .....	24
<b>GEPI</b> .....	24
<b>ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ</b> .....	25
<b>Κεφάλαιο 6.</b> .....	30
<b>Μελλοντική ανάπτυξη. Παρατήρηση δικτύου με βάση το χρόνο.</b>	
<b>Προβλέψεις</b> .....	30
<b>Κεφάλαιο 7</b> .....	31
<b>Συμπεράσματα</b> .....	31
<b>Κεφάλαιο 8.</b> .....	32
<b>Παράρτημα-Κώδικας Python</b> .....	32
<b>Κεφάλαιο 9.</b> .....	36
<b>Ευχαριστίες</b> .....	36
<b>Ευχαριστίες</b> .....	36
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ-ΠΗΓΕΣ</b> .....	38

## ΕΙΣΑΓΩΓΗ

Στην μεταπτυχιακή διατριβή αυτή θα μιλήσουμε για εξόρυξη γνώσης από ένα κοινωνικό δίκτυο, πιο συγκεκριμένα για το Twitter. Θα οπτικοποιήσουμε την γνώση αυτή γραφικά, φτιάχνοντας ένα γράφο, χρησιμοποιώντας κάποια εργαλεία και θα βγάλουμε τέλος κάποια συμπεράσματα με βάση τον γράφο. Τα συμπεράσματα αυτά μπορεί να έχουν σχέση με διάφορες επιστήμες όπως η στατιστική επιστήμη, η επιστήμη των υπολογιστών, η ψυχολογία, η κοινωνιολογία κ.ο.κ

Στο τέλος θα μιλήσουμε για ποιόν λόγο θεωρούμε την ιδέα μας τόσο καινοτόμα, για την τυχόν μελλοντική εξέλιξη της ιδέας μας, αλλά και για τους λόγους που δεν συνεχίσαμε (προς το παρόν) την ιδέα μας σε πιο μεγάλο δείγμα ανθρώπων.

Πιο συγκεκριμένα,

Στο Κεφάλαιο 1, κάνουμε μια μικρή εισαγωγή, του τι είναι ένα κοινωνικό δίκτυο, πως λειτουργούν, ποιά κοινωνικά δίκτυα υπάρχουν σήμερα, τι αποσκοπούν, ποιές ανάγκες καλύπτουν, αλλά κυρίως ποιές επιστήμες σχετίζονται άμεσα με αυτά και γιατί.

Στο Κεφάλαιο 2, θα μιλήσουμε για την ιδέα μας, τι ακριβώς θέλουμε να πετύχουμε, πως και γιατί. Στο κεφάλαιο αυτό, θα μιλήσουμε πρώτη φορά για τα μέτρα κεντρικότητας, και θα δώσουμε τους ορισμούς του βαθμού, της ενδιαμεσότητας αλλά και της εγγύτητας. Έννοιες, άμεσα συνδεδεμένες με την κεντρικότητα. Θα μιλήσουμε επίσης για την ομοφιλία, αλλά και για τα παραδείγματα 'κλειστών' κοινοτήτων μέσα στα κοινωνικά δίκτυα.

Στο Κεφάλαιο 3, θα μιλήσουμε για την κοινωνική επιλογή, τους αλγόριθμους αναζήτησης πληροφοριών μέσα σε ένα δίκτυο, για παράδειγμα ο αλγόριθμος PageRank της Google. Θα μιλήσουμε επίσης, για επιδημίες, για διαχείριση πόρων, θα πούμε για παραδείγματα εξεγέρσεων(Αθήνα 2008, Κων/πολη 2013) και τον ρόλο που έπαιξαν τα κοινωνικά δίκτυα, και τέλος θα μιλήσουμε για τα Social media analytics.

Στο Κεφάλαιο 4, θα μιλήσουμε για το Twitter API, τι ακριβώς είναι, γιατί το χρησιμοποιήσαμε και τι καταφέραμε με αυτό. Θα πούμε για την γλώσσα προγραμματισμού Python, που χρησιμοποιήσαμε, και θα μιλήσουμε και για το Gephi. Μιλώντας για το Gephi, θα κάνουμε μια μικρή εισαγωγή και στα θεωρητικά των γράφων. Στο κεφάλαιο αυτό δηλαδή θα πούμε κάποια λίγα εισαγωγικά. Τον ορισμό του γράφου, για το τι είναι ένας κόμβος, τι είναι μία ακμή, ένας γείτονας κ.ο.κ. Θα δούμε τα δύο είδη γράφων, τον κατευθυνόμενο και τον μη-κατευθυνόμενο, και θα εξηγήσουμε γιατί ο δικός ανήκει στην κατηγορία των κατευθυνόμενων γράφων.

Στο Κεφάλαιο 5, σαν συνέχεια του κεφαλαίου 4, θα πούμε για τα χαρακτηριστικά των γράφων, θα μιλήσουμε για το μήκος ένας γράφου, αλλά και για την διάμετρο ενός γράφου. Θα πούμε για τους δυο διαφορετικούς αλγόριθμους που χρησιμοποιούνται για την αναζήτηση στοιχείου σε δεντροειδή μορφή. Θα πούμε για την συνάρτηση set() που χρησιμοποιήσαμε στη Python, ώστε να μην ξαναπαίρνουμε χρήστες που ήδη έχουμε μια φορά, ώστε ο γράφος μας μετά να αποτελείται ΜΟΝΟ από μοναδικούς χρήστες. Θα δούμε το γράφημα που βγήκε με βάση τα δικά μας στοιχεία. Θα μιλήσουμε για την πυκνότητα ενός γράφου, και την σχέση που έχει αυτό με τα μέτρα κεντρικότητας. Θα δούμε επίσης στατιστικά που βγάλαμε, μέσω του Gephi, για τον γράφο και γιατί αυτά είναι τόσο σημαντικά για εμάς.

Στο Κεφάλαιο 6, θα πούμε για την μελλοντική ανάπτυξη. Για την παρατήρηση του γράφου συναρτήση του χρόνου, και θα μιλήσουμε για πιθανές προβλέψεις που θα μπορούσαν να γίνουν. Θα δούμε επίσης πως θα μπορούσε αυτή η ιδέα να γίνει η αρχή για μία διδακτορική διατριβή, πάνω στην παρατήρηση και στην πρόβλεψη της ανάπτυξης των κοινωνικών δικτύων.

Στο Κεφάλαιο 7, θα πούμε συνοπτικά τα συμπεράσματα από την μεταπτυχιακή διατριβή.

Στο Κεφάλαιο 8, είναι το παράρτημα - ο κώδικας που γράψαμε στα πλαίσια της πτυχιακής σε Python.

Στο Κεφάλαιο 9, είναι οι ευχαριστίες.

Και στο τέλος οι βιβλιογραφίες και οι πηγές που χρησιμοποιήσαμε.

Και τα Copyrights.

---

## Κεφάλαιο 1.

### Τι είναι τα κοινωνικά δίκτυα. Πως λειτουργούν. Σχετικές επιστήμες, σχετικές έρευνες. Μεγέθη.

---

Τα **κοινωνικά δίκτυα** είναι ένα σύνολο αλληλεπιδράσεων και διαπροσωπικών σχέσεων. Ο όρος σήμερα χρησιμοποιείται επίσης για να περιγράψει ιστοσελίδες οι οποίες επιτρέπουν την διεπαφή ανάμεσα στους χρήστες, πχ. με σχόλια, φωτογραφίες κ.ο.κ. Οι πιο γνωστές από αυτές τις ιστοσελίδες είναι το [Facebook](#), [Twitter](#), [Instagram](#) και [LinkedIn](#). Οι ιστότοποι αυτοί αποτελούν εικονικές κοινότητες όπου οι χρήστες μπορούν να επικοινωνούν και να αναπτύξουν επαφές μέσα από αυτές. Ένα κοινωνικό δίκτυο είναι μια κοινωνική δομή που αποτελείται από ένα σύνολο παραγόντων, όπως άτομα ή οργανισμούς. Στο διαδίκτυο, τα κοινωνικά δίκτυα είναι μία πλατφόρμα που συντηρείται για την δημιουργία κοινωνικών σχέσεων μεταξύ των ανθρώπων, που συνήθως αποτελούν ενεργά μέλη του κοινωνικού δικτύου, με κοινά ενδιαφέροντα ή δραστηριότητες.

Οι ιστότοποι κοινωνικής δικτύωσης είναι οργανωμένες ιστοσελίδες στο διαδίκτυο με περισσότερο ομαδοκεντρικό χαρακτήρα που παρέχουν, στην συντριπτική τους πλειοψηφία, μία σειρά από βασικές και δωρεάν υπηρεσίες όπως τη δημιουργία προφίλ, το ανέβασμα εικόνων και βίντεο, τον σχολιασμό σε ενέργειες που γίνονται από άλλα μέλη του δικτύου ή μίας ομάδας, την άμεση ανταλλαγή μηνυμάτων και πολλά άλλα. <sup>[1]</sup>

Στη μεταπτυχιακή διατριβή αυτή, θα ασχοληθούμε με το Twitter. Το Twitter είναι ένα από τα πιο νέα κοινωνικά δίκτυα, ξεκίνησε το Ιούλιο του 2006, από τον Τζακ Ντόρσει, έχει σαν έδρα το Σαν Φρανσίσκο, Καλιφόρνια, ΗΠΑ. Η υπηρεσία έγινε γρήγορα δημοφιλής και σύμφωνα με την εταιρία είχε 250 εκατομμύρια ενεργούς χρήστες (2014). Είναι το πιο γρήγορα αναπτυσσόμενο κοινωνικό δίκτυο αυτή την στιγμή. (Τα επίσημα στοιχεία, δημοσιεύονται το 1<sup>ο</sup> τρίμηνο κάθε έτους, για το προηγούμενο έτος). Είναι ένας από τους δέκα πιο δημοφιλείς ιστότοπους του διαδικτύου. <sup>[2]</sup>

Αξίζει, επίσης, να σημειωθεί ότι το 63% των Ελλήνων χρηστών σύμφωνα με δημοσίευμα της εφημερίδας “Ελευθεροτυπία” έχουν δημιουργήσει προφίλ σε κάποια online υπηρεσία κοινωνικής δικτύωσης, ενώ ο αντίστοιχος ευρωπαϊκός μέσος όρος είναι 45%. Το υψηλό αυτό ποσοστό φέρνει την Ελλάδα στην τρίτη θέση πανευρωπαϊκά, μετά τη Δανία και τη Νορβηγία.

Οι πιο δημοφιλείς ιστότοποι κοινωνικής δικτύωσης σήμερα είναι οι εξής:

- MySpace
- Facebook
- Twitter
- Hi 5
- LinkedIn
- YouTube
- Windows Live Messenger

Τα κοινωνικά δίκτυα συνήθως αποτελούνται από τα μέλη της οικογένειας, τους φίλους και τους γνωστούς και περιλαμβάνουν τρεις κρίσιμες έννοιες:

1. το μέγεθος ή το εύρος, το οποίο αναφέρεται στον αριθμό των ατόμων που συμμετέχουν στο δίκτυο



2. τη σύνθεση, δηλαδή το ποσοστό συμμετοχής μελών της ευρύτερης οικογένειας ή φίλων στο δίκτυο, και
3. η συχνότητα, που δηλώνει το πόσο συχνά τα μέλη ενός κοινωνικού δικτύου αλληλεπιδρούν μεταξύ τους

Όταν πρόκειται για online κοινωνική δικτύωση, χρησιμοποιούνται ιστοσελίδες γνωστές ως ιστότοποι κοινωνικής δικτύωσης. Οι ιστότοποι κοινωνικής δικτύωσης λειτουργούν ως online κοινότητες των χρηστών του Διαδικτύου. Ανάλογα με τα χαρακτηριστικά κάθε ιστότοπου, τα μέλη της online κοινότητας έχουν κοινά ενδιαφέροντα όπως κάποιο χόμπι, τη θρησκεία ή την πολιτική. Η εγγραφή και πρόσβαση σε έναν ιστότοπο κοινωνικής δικτύωσης παρέχει στα μέλη του άμεση και συνεχή κοινωνικοποίηση. Αυτή η κοινωνικοποίηση μπορεί να περιλαμβάνει την ανάγνωση του προφίλ των άλλων μελών και συχνά και την επικοινωνία μαζί του. <sup>[3]</sup>

Στα πλαίσια αυτά, υπάρχουν φορείς, οργανισμοί, επιχειρήσεις ακόμα και πανεπιστήμια που ερευνούν το 'φαινόμενο' των κοινωνικών δικτύων, ο καθένας με την δικιά του οπτική, προσπαθώντας να καταλάβουν την δομή, ενός τέτοιου δικτύου, τις αλληλεπιδράσεις που έχει με την καθημερινότητα των ανθρώπων που είναι χρήστες, αλλά και τις μελλοντικές συνέπειες αυτού, του νέου τρόπου επικοινωνίας μεταξύ των ανθρώπων.

Οι επιστήμες που κυρίως μελετούν τα κοινωνικά δίκτυα είναι οι εξής:

- Πληροφορική
- Κοινωνιολογία
- Ψυχολογία
- Οικονομικές επιστήμες
- Στατιστική

Πιο συγκεκριμένα, ο κλάδος της πληροφορικής μελετάει κυρίως το τεχνικό κομμάτι, των κοινωνικών δικτύων. Για παράδειγμα, τα δίκτυα, τους Servers, και την συντήρησή τους, τις βάσεις δεδομένων, data mining κ.τ.λ.

*Κάθε χρόνο το Τμήμα Πληροφορικής του Οικονομικού Πανεπιστημίου Αθηνών, για παράδειγμα, κάνει μια στατιστικού τύπου έρευνα, πάνω στα κοινωνικά δίκτυα. Η έρευνα τελειώνει φέτος τον Δεκέμβριο. Σχετικές πληροφορίες θα βρείτε στο Link :*

*<http://www.eltrun.gr/meletes/erevna-gia-social-media/>*

Ο κλάδος της Κοινωνιολογίας μελετά, την ομοιογένεια, ή κατα πόσο υπάρχει αυτή, μεταξύ των χρηστών. Δηλαδή μελετάει να βρεί με βάση ποιό χαρακτηριστικό διαλέγουν να γίνουν 2 χρήστες, φίλοι μεταξύ τους. Τα πιο συνηθισμένα κριτήρια, είναι η γλώσσα, ο τόπος καταγωγής ή η μόνιμα κατοικία, τα πολιτικά «πιστεύω», ο τόπος εργασίας, κ.τ.λ. Η Ψυχολογία μελετά την ανθρώπινη δραστηριότητα ενός κόμβου στο δίκτυο. Δηλαδή μελετά το πότε, το ποιός και το που κάποιος έκανε κάτι στο Κοινωνικό δίκτυο, προσπαθώντας να απαντήσει στο ερώτημα, γιατί? Οι οικονομικές επιστήμες, μελετούν την οικονομική δραστηριότητα σε ένα κοινωνικό δίκτυο. Στο Facebook, για παράδειγμα, οι διαφημίσεις μέσω της σελίδας θα αποφέρουν στην εταιρία μόνο για το 2015 πάνω από 4 δις Αμερικάνικα δολάρια. Εταιρία που μόνο στην Καλιφόρνια των ΗΠΑ, εργάζονται πάνω από 6.000 άτομα ετησίως. Είναι εύκολα λοιπόν αντιληπτό πως τα οικονομικά κέρδη, μέσα από τα κοινωνικά δίκτυα είναι τεράστια. Τέλος, η στατιστική επιστήμη, μελετάει την τυχαιότητα 2 κόμβων, τα στατιστικά, είτε ενός μεμονωμένου χρήστη, είτε μιας ομάδας χρηστών, προσπαθώντας να βγάλει κάποια στατιστικά νούμερα που εκφράζουν το δίκτυο.

---

## Κεφάλαιο 2.

### Η ιδέα μας. Τι θέλουμε να πετύχουμε, πως και γιατί.

---

Στην συγκεκριμένη μεταπτυχιακή διατριβή, χρησιμοποιήσαμε την γλώσσα προγραμματισμού Python, για να κάνουμε εξόρυξη γνώσεων. Συγκεκριμένα βγάζουμε τους Followers, και τους Followers των Followers. Με βάση τον δικό μας λογαριασμό στο Twitter. Κάνοντας μια συγκεκριμένη διαδικασία, την οποία θα αναλύσουμε παρακάτω, καταφέρνουμε να πάρουμε, μέχρι και από τους Followers των Followers μας, τα αλφαριθμητικά ID (είναι αυτά με τα οποία αποθηκεύετε ο κάθε χρήστης στους Servers του Twitter, γιατί δεν αποθηκεύετε ο καθένας με το όνομά του), και τα TwitterNames τους.

Για παράδειγμα, το TwitterName μου είναι 'zorzomikoskosta'. Ενώ στη βάση τους με έχουν αποθηκευμένο ως : 'bO6OiScMfeZkWyFsQeyEjHbl9bYNPFJr1LOdWKFzu7RtU'. Ο κύριος λόγος που δεν αποθηκεύουν τον κάθε χρήστη με το όνομα που έχει κανονικά, είναι γιατί φοβούνται κάποια κενά ασφαλείας. Οι διαχειριστές στη βάση ποτέ δεν θα δούνε το TwitterName κάποιου, παρόλο που θα ήταν πιο βολικό σαν χρήση. Γιατί φοβούνται πως υπάρχει μεγάλη πιθανότητα, κάποιος χρήστης της βάσης, να ξέρει κάποιον χρήστη του Twitter, και να τον καταλάβει από το TwitterName του, και να μπει στα δεδομένα του.

Όσο και αν φαίνεται απίστευτο, μια τελευταία μελέτη, έδειξε πως 2 τυχαίοι χρήστες στο Facebook απέχουν ΚΑΤΑ ΜΕΣΟ ΟΡΟ, σχεδόν 6 κόμβους. Δηλαδή αν πάρουμε δυο τυχαίους χρήστες στο Facebook, θα φτάσουμε κατά μέσο όρο από τον έναν στον άλλον, περνώντας μόνο από 6 άτομα. [\[5\]](#)

Ας υποθέσουμε πως έχουμε μια δομή δέντρου. Στην ρίζα του είμαστε εμείς, τα παιδιά του δέντρου είναι οι Followers μας, τα εγγόνια μας είναι οι Followers των Followers κ.ο.κ. Αυτό που θέλουμε να πετύχουμε είναι να δούμε με ποιά κριτήρια μας κάνουν Follow οι άλλοι, αλλά και με ποιά κριτήρια κάνουμε και εμείς Follow σε κάποιον. Επίσης ενδιαφέρον έχει ο λόγος του ενός, προς του άλλου. Γιατί από τον λόγο αυτό φαίνεται η δημοτικότητα κάποιου, στο δίκτυο.

Βρίσκοντας λοιπόν, τους Followers, και τους Followers των Followers (δηλαδή πάμε μέχρι 3<sup>ου</sup> βαθμού, αν πούμε πως εμείς είμαστε βαθμός 1), έχουμε μια μικρή εικόνα για το το δίκτυο, μέρος του οποίου είμαστε εμείς. Τα δεδομένα αυτά, όταν τρέξει ο κώδικας, υπάρχουν σε ένα αρχείο newfile.txt, στο path που βρίσκεται και η Python. Και το αρχείο φτιάχτηκε με αυτό τον τρόπο, ώστε να είναι ξεκάθαρο ποιός είναι παιδί μου, και ποιός εγγονός μου. Το επόμενο βήμα, είναι η μετατροπή του αρχείου από .txt σε .gml. Τα αρχεία .gml διαβάζονται από το Gephi, το οποίο μας αναπαριστά το δίκτυο, γραφικά. Δηλαδή μας δίνει τον γράφο, για εμάς, τα παιδιά μας, και τα εγγόνια μας.

Ο σκοπός όλο αυτού είναι να δούμε γραφικά, το μέρος του δικτύου, στο οποίο ανήκουμε στο Twitter. Βλέποντάς το γραφικά, και χρησιμοποιώντας τα εργαλεία του Gephi, βγάζουμε κάποια συμπεράσματα, όπως το μέσο όρο σύνδεσης του κάθε κόμβου (χρήστη δηλαδή), τα μέτρα κεντρικότητας κτλ.

Εμείς κυρίως θα ασχοληθούμε με τα μέτρα κεντρικότητας, στο κοιμωνικό μας δίκτυο, και όχι τόσα με τα στατιστικά νούμερα που βρήκαμε μελετώντας το. Στη θεωρία της ανάλυσης των δικτύων, υπάρχουν τρία χαρακτηριστικά που σχετίζονται με τα μέτρα της κεντρικότητας ενός κόμβου μέσα σε ένα γράφημα που χρησιμοποιούνται ευρέως στην ανάλυση του δικτύου: η κεντρικότητα ως προς το βαθμό, ως προς την ενδιαμεσότητα, και ως προς την εγγύτητα. Αυτά καθορίζουν τη σημαντικότητα της κορυφής στο

διάγραμμα (για παράδειγμα, πόσο σημαντικό είναι ένα άτομο μέσα σε ένα κοινωνικό δίκτυο, ή, πόσο σημαντικό είναι ένα δωμάτιο μέσα σε ένα κτίριο ή το πόσο χρήσιμος είναι ένας δρόμος μέσα σε ένα αστικό δίκτυο).

Τα μέτρα κεντρικότητας, γενικότερα χρησιμοποιούνται από πολλές επιστήμες. Οι αεροπορικές εταιρίες για παράδειγμα, μελετάνε τα μέτρα κεντρικότητας των αεροπορικών τους γραμμών, και βλέπουν, ανα τακτά χρονικά διαστήματα, το πόση κίνηση υπάρχει σε κάθε αεροδρόμιο, ποιοί ενδιάμεσοι σταθμοί χρησιμοποιούνται πιο πολύ, και ποιοί κόμβοι είναι πιο σημαντική για αυτούς, με βάση της πτήσεις που κάνουν. Επίσης οι συγκοινωνιολόγοι, χρησιμοποιούν μέτρα κεντρικότητας, βάζοντας σαν κόμβους τα πιο σημαντικά μέρη εστιάσεις του κόσμου, στάσεις μέσω μαζικής μεταφοράς κτλ, και σαν συνδέσεις/γραμμές τις οδικές αρτηρίες. Με αυτό τον τρόπο προσπαθούν να βρουν με βάση τα μέτρα κεντρικότητας, ποιοί κόμβοι είναι οι πιο σημαντικοί, άρα στις συνδέσεις αυτού του κόμβου, τις ώρες αιχμής είναι πολύ πιθανό να έχουν κυκλοφοριακή συμφόρηση. Τα μέτρα κεντρικότητας χρησιμοποιεί και η αστυνομία παγκοσμίως, σε περιπτώσεις διαδηλώσεων, εξεγέρσεων κτλ, ώστε να εξημερευτούν τα μέσα, χωρίς να περνάνε από κεντρικούς κόμβους, και με όσο λιγότερες στάσεις, το μέγιστο αριθμό ανθρώπων. Πολύ πρόσφατο παράδειγμα, το Νοέμβριο του 2015 στο Παρίσι, μετά το τρομοκρατικό χτύπημα, ο δήμος, και η αστυνομία, αποφάσισαν να μην χρησιμοποιούν τους πιο «κεντρικούς» σταθμούς του μετρό, γιατί σε περίπτωση κάποιας έκρηξης, θα είχε τα περισσότερα θύματα. Συνοψίζοντας λοιπόν, τα μέτρα κεντρικότητας, χρησιμοποιούνται καθημερινά γύρω μας, είτε το ξέρουμε, είτε όχι, και είναι κάτι το οποίο απασχολεί ανθρώπους από πολλά επιστημονικά πεδία.

Ο πρώτος και απλούστερος τρόπος να ορίσουμε την κεντρικότητα ενός κόμβου είναι να τη μετρήσουμε μέσω του αριθμού των απευθείας συνδέσεων που αυτός ο κόμβος έχει, με άλλους κόμβους στο δίκτυο. Ο αριθμός αυτός, στη θεωρία των γράφων, ονομάζεται βαθμός του κόμβου. Πέρα από τον βαθμό ενός κόμβου, υπάρχουν και η εγγύτητα και η ενδιαμεσότητα.

Η εγγύτητα (closeness) για έναν κόμβο A ορίζεται ως ο λόγος του πλήθους των υπόλοιπων κόμβων στο δίκτυο προς το άθροισμα του μήκους των μονοπατιών ελαχίστου μήκους που συνδέουν τον A με καθέναν από τους υπόλοιπους κόμβους. Η εγγύτητα ενός κόμβου αυξάνεται όσο μειώνεται το άθροισμα στον παρανομάστη και επομένως μια μεγάλη τιμή εγγύτητάς υποδηλώνει ότι ο A είναι “κοντά” σε όλους τους υπόλοιπους κόμβους.

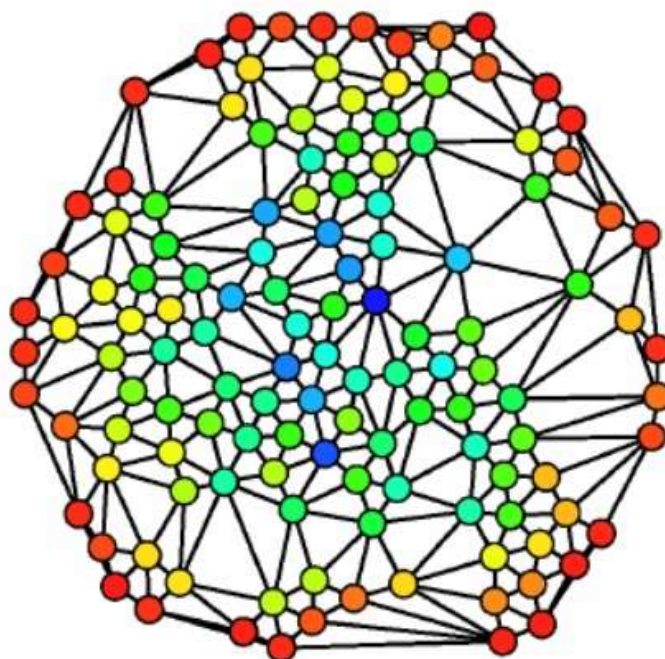
Η ενδιαμεσότητα (betweenness) ενός κόμβου A εκτιμά πόσο σημαντικός είναι ο A για την επικοινωνία μεταξύ των άλλων κόμβων. Η μεσολάβηση ορίζεται ως το πλήθος των μονοπατιών ελαχίστου μήκους που συνδέουν δυο οποιουσδήποτε κόμβους και στα οποία ανήκει ο A προς το πλήθος των μονοπατιών ελαχίστου μήκους που συνδέουν δυο οποιουσδήποτε κόμβους. Στην κανονικοποιημένη μορφή της η μεσολάβηση διαιρείται δια του όρου  $(n-1)*(n-2)/2$  όπου n το πλήθος των κόμβων στο δίκτυο. Ο όρος αναφέρεται στον αριθμό των δυνατών συνδυασμών n-1 κόμβων ανα 2. Πιο απλά η ενδιαμεσότητα, μας δείχνει πόσοι ενδιάμεσοι κόμβοι υπάρχουν μεταξύ εμάς, και συνήθως του πιο ‘ισχυρού’ κόμβου, επίσης μας δείχνει και τις διάφορες εναλλακτικές διαδρομές που μπορεί να ακολουθήσουμε, για να φτασούμε εκεί. <sup>[6]</sup>

Υπάρχουν παράδοξα, στα μέτρα κεντρικότητας ενός κοινωνικού δικτύου. Ενώ ο βαθμός, είναι το πρώτο και ίσως το πιο σημαντικό κριτήριο στα μέτρα κεντρικότητας, στα κοινωνικά δίκτυα δεν είναι. Και αυτό γιατί ΚΥΡΙΩΣ στα κοινωνικά δίκτυα μας ενδιαφέρει και ενδιαμεσότητα, και η εγγύτητα. (προφανώς και η φερεγγυότητα). Δεν έχει νόημα για εμάς να έχει κάποιος πέντε χιλιάδες Followers, και το προφίλ του να είναι ψεύτικο. Άρα, πόσες απ’ευθείας συνδέσεις, έχει ένας χρήστης, παίζει ρόλο, δηλαδή ο βαθμός του κόμβου, αλλά πάντα συνδυαστικά με τους άλλους δυο παράγοντες.

Για παράδειγμα, είναι άλλο να έχει ένας κόμβος 5000 συνδέσεις, και ο κάθε ένας από αυτούς να έχει επίσης 5000 συνδέσεις, από το να έχει ένας κόμβος 5000 συνδέσεις, αλλά οι πλειοψηφία αυτών να έχουν μικρό βαθμό. Συνοπτικά δηλαδή, το ιδανικό σενάριο, είναι να έχει ένας κόμβος μεγάλο βαθμό, οι συγγενείς του κόμβου αυτού (παιδιά-εγγόνια κτλ) να έχουν επίσης μεγάλο βαθμό, και τα προφίλ όλων αυτών να είναι αληθινά, διαφορετικά χάνεται η εγγύτητα, και η φερεγγυότητα.

Για να καταλάβουμε καλύτερα θα σας πούμε ακόμα ένα παράδειγμα. Αν έχω ένα γνωστό μου, και μέσω αυτού έγινα «φίλος» με το OBAMA (στο Facebook), τότε η σχέση μου, μαζί του, σχετίζεται μόνο από αυτόν το ένα φίλο μου. Δηλαδή, αν μελλοντικά δεν είμαι πλέον φίλος του, πιθανότατα, να μην είμαι φίλος και με τον OBAMA. Τότε, χάνω και την σύνδεση με τον «κέντρο» κεντρικότητας (αν υποθέσουμε πως είναι ο OBAMA), αλλά χάνω και την σύνδεση με το γνωστό μου, που η ενδιαμεσότητα του με τον OBAMA ήταν 1<sup>ου</sup> βαθμού. <sup>[4]</sup> Συμπερασματικά λοιπόν, ο συνδυασμός και των 3 αυτών μεγεθών, μας δίνει την πληροφορία που θέλουμε. Να έχουμε μεγάλο βαθμό, να έχουμε όσο γίνεται πιο πολλές συνδέσεις με τα μεγάλα κέντρα/κόμβους του κοινωνικού δικτύου και οι συνδέσεις αυτές να είναι όσο γίνεται 1<sup>ου</sup> βαθμού, βέβαια όλα αυτά γνωρίζοντας ότι τα δεδομένα είναι πραγματικά.

Η κεντρικότητα ενός ατόμου μας δείχνει πόσο καλά είναι τοποθετημένο στο δίκτυο, ώστε να του παρέχεται η δυνατότητα λόγω θέσης, να λαμβάνει και να μοιράζει πληροφορίες σε όλα τα άλλα μέλη του δικτύου (Freeman, 1979· Wasserman & Faust, 1994). Αυτό φαίνεται στην παρακάτω εικόνα.



**Η απόχρωση (από το κόκκινο = 0 έως μπλε = μέγιστο) μας δείχνει την ενδιαμεσότητα των κόμβων.**

**Εικ.1**

Ένα ενδιαφέρον παράδειγμα το οποίο μελετάτε στα κοινωνικά δίκτυα, είναι το φαινόμενο της ομοφιλίας. Μια κοινότητα ανθρώπων που είναι όλοι φίλοι μεταξύ τους, και ο καθένας να έχει παρόμοιο βαθμό (παρόμοιο αριθμό φίλων δηλαδή), και σχετικά ίδια ενδιαμεσότητα και εγγύτητα. Συνήθως παρατηρείται, σε κόλπους πολιτικούς, καλλιτεχνικούς κτλ. Όπου υπάρχει μια «κλειστή» ομάδα ανθρώπων, που είναι όλοι φίλοι μεταξύ τους, και δύσκολα κάποιος κάνει φίλο του ή Follow, αναλόγως, κάποιον που δεν ανήκει σε αυτή την κλειστή ομάδα. Οι ερευνητές μεγάλων πανεπιστημίων, που ερευνούν το φαινόμενο, το ονομάζουν, μάλλον ειρωνικά, σαν 'Ghetto' των κοινωνικών δικτύων.

Θεωρούμε λοιπόν πως η ιδέα μας, να μελετήσουμε ένα κοινωνικό δίκτυο, βγάζοντας συμπεράσματα για τα μέτρα κεντρικότητας, και όχι μόνο, είναι κάτι χρήσιμο, κυρίως σαν επίπεδο πληροφοριών. Είναι

μια σύνθετη πληροφορία, που μας δείχνει και το πως είναι στην πράξη το δίκτυο, στο οποίο και εμείς ανήκουμε, και ο τρόπος σύνδεσης μεταξύ των χρηστών, τα τοπογραφικά τους χαρακτηριστικά κ.ο.κ.

Βέβαια, δεν είναι κάτι εύκολο, αφού απαιτεί γνώση προγραμματισμού, εφόσον χρειάστηκε να γραφτεί αρκετός κώδικας, για να πάρουμε τα δεδομένα, αλλά απαιτεί και αρκετή γνώση πάνω σε γράφους. Χωρίς του γράφους δεν θα είχαμε την οπτική απεικόνιση του δικτύου, και δεν θα είχαμε σαφή εικόνα κυρίως για την εγγύτητα και την ενδιάμεσότητα.

---

## Κεφάλαιο 3.

### Κοινωνική επιλογή. Αλγόριθμοι αναζήτησης πληροφοριών σε κοινωνικά δίκτυα. Επιδημίες. Παραδείγματα εξεγέρσεων. Social Media Analytics.

---

#### Κοινωνική επιλογή

Η Θεωρία της Κοινωνικής Επιλογής ή η Κοινωνική Επιλογή είναι ένα θεωρητικό πλαίσιο ανάλυσης ή συνδυασμού ατομικών προτιμήσεων, ενδιαφερόντων ή προθέσεων για να προσεγγίσεις μία συλλογική απόφαση ή κοινωνικής πρόθεσης, κατά κάποιο τρόπο. Ένα μη-θεωρητικό παράδειγμα συλλογικής απόφασης είναι η θέσπιση ενός νόμου ή το σύνολο των νόμων στο πλαίσιο ενός συντάγματος. Η θεωρία της κοινωνικής επιλογής χρονολογείται από τη διαμόρφωση του Condorcet για το παράδοξο της ψηφοφορίας. Η Κοινωνική Επιλογή συνδυάζει στοιχεία της οικονομικής ευημερίας και της θεωρίας της ψηφοφορίας. Είναι μεθοδολογικά ατομικιστική, υπό την έννοια ότι αθροίζει τις προτιμήσεις και τις συμπεριφορές των μεμονωμένων μελών της κοινωνίας. Χρησιμοποιώντας τα στοιχεία της τυπικής λογικής γενικά, η ανάλυση προχωρά από μια σειρά φαινομενικά λογικών αξιωμάτων της κοινωνικής επιλογής στο σχηματισμό μιας λειτουργίας κοινωνικής πρόθεσης (ή συντάγματος). <sup>[7]</sup>

Όπως λοιπόν και στην πραγματικότητα, έτσι και στα κοινωνικά δίκτυα, από τις επιλογές των ανθρώπων που έχουμε διαλέξει σαν φίλους, (ή να τους κάνουμε Follow), βγάζουμε συμπεράσματα, για την οικονομική ευμάρεια των χρηστών, για τα πολιτικά πιστεύω, για τα θρησκευτικά πιστεύω, τις τυχόν ομάδες που υποστηρίζουν, την ηλικία τους, τα χόμπι τους, τον τόπο κατοικίας κ.ο.κ. Οι περισσότεροι χρήστες είναι φίλοι με κάποιον που έστω σε κάτι έχουν μια σχέση, ή σχετίζονται με κάποιο τρόπο. Οι σχέσεις που μπορεί να έχουν, ποικίλει. Μπορεί να είναι σχέσεις εργοδότη-εργαζομένου, μπορεί να είναι συγγενικές σχέσεις ή οτιδήποτε άλλο.

#### Αλγόριθμοι αναζήτησης πληροφοριών σε κοινωνικά δίκτυα

Αυτό το κομμάτι ενδιαφέρει κυρίως τις οικονομικές επιστήμες, προσπαθώντας να βρουν τρόπο να κάνουν στοχευμένη διαφήμιση μέσω κοινωνικών δικτύων. Για να το πετύχουν αυτό, πολλές εταιρίες ανά τον κόσμο, έχουν φτιάξει αλγόριθμους αναζήτησης πληροφοριών σε κοινωνικά δίκτυα. <sup>[8]</sup> Ο πιο διαδεδομένος αλγόριθμος που χρησιμοποιείται για την αναζήτηση πληροφοριών σε κοινωνικά δίκτυα, είναι ο Αλγόριθμος PageRank. Ο αλγόριθμος αυτός είναι γενικά γνωστός, γιατί τον χρησιμοποιούν κυρίως μηχανές αναζήτησης. Είναι ο αλγόριθμος που χρησιμοποιεί και η Google, ώστε με βάση μια λέξη-κλειδί, να μας βγάλει όσο γίνεται παραπάνω όγκο δεδομένων, και όσο πιο σχετικό, με αυτό που ζητήσαμε.

ΣΗΜΕΙΩΣΗ : Σίγουρα υπάρχουν ατέλειες στο PageRank, ιδίως τώρα που οι άνθρωποι γνωρίζουν τον αλγόριθμο αυτόν και μπορούν να δημιουργήσουν «ψεύτικους» συνδέσμους με στόχο να αποκτήσει υψηλότερο PageRank κάποια σελίδα που επιθυμούν (link farming). Πλέον η Google έχει προσαρμόσει τους υπολογισμούς του αλγόριθμου της, ώστε να φιλτράρουν τις σελίδες εκείνες από τις σελίδες που πιθανόν κάνουν χρήση «φάρμα συνδέσμων». <sup>[9]</sup>

Το Google, κάνοντας αναζήτηση, εμφανίζει πάνω-πάνω, τις αντίστοιχες σελίδες που εμπεριέχουν την λέξη-κλειδί, και έχουν μεγάλη επισκεψιμότητα, πιστεύοντας πως έτσι είναι πιο πιθανό να βρεί ο χρήστης εύκολα αυτό που ψάχνει. Κάποιοι, αποσπώντας μεγάλα οικονομικά ωφέλη, είχαν βρεί



τρόπο να κάνουν 'ψεύτικα κλίκ' στις ιστοσελίδες του, για να τους βγάξει πάνω-πάνω η αναζήτηση το δικό τους Site. Ο αλγόριθμος PageRank όμως έχει αλλάξει το τελευταίο διάστημα, και αναγνωρίζει τα 'εικονικά κλίκ', ώστε να βγάξει σωστά τα αποτελέσματα, με βάση την επισκεψιμότητα.

## ΕΠΙΔΗΜΙΕΣ

Μια ιδιαίτερα σημαντική περιοχή έρευνας στα κοινωνικά δίκτυα σχετίζεται με τη διερεύνηση του τρόπου με τον οποίο διάφορα φαινόμενα εξαπλώνονται σε ένα δίκτυο. Χαρακτηριστικά παραδείγματα τέτοιων φαινομένων, στα οποία αναφερόμαστε με τον όρο επιδημίες, αποτελούν η εξάπλωση υπολογιστικών ιών σε δίκτυα επικοινωνιών, η εμφάνιση διαφόρων μορφών βιολογικών επιδημιών (π.χ γρίπη, AIDS) σε ένα πληθυσμό ή η εξάπλωση διαφόρων ειδών μόδας που βασίζονται στη μίμηση (π.χ τατουάζ, viral videos κλπ). Οι τρόποι με τους οποίους οι επιδημίες εξαπλώνονται σε ένα πληθυσμό εξαρτάται όχι μόνο από την ικανότητα μετάδοσης που διαθέτουν τα παθογόνα στοιχεία τους, τη διάρκεια και την κρισιμότητα της μόλυνσης που προκαλούν αλλά και από τη δομή του δικτύου με το οποίο συνδέονται τα μέλη του πληθυσμού. Γενικότερα, οι ευκαιρίες που έχει μια επιδημία να εξαπλωθεί προέρχονται από ένα δίκτυο επαφών (contact network) στο οποίο υπάρχει ένας κόμβος για κάθε μέλος του πληθυσμού και μια ακμή που υποδηλώνει ότι δύο άτομα έρχονται σε επαφή μεταξύ τους με τρόπο που επιτρέπει στατιστικά να μεταδοθεί ο παθογόνος παράγοντας από το ένα στο άλλο. Επομένως η δομή του συγκεκριμένου δικτύου καθορίζει αποφασιστικά το μέγεθος της εξάπλωσης μιας επιδημίας. Διαισθητικά όσο πιο πυκνό είναι το δίκτυο επαφών για ένα συγκεκριμένο παθογόνο παράγοντα τόσο μεγαλύτερη είναι η διάδοση της επιδημίας στον υποκείμενο πληθυσμό. Η κύρια διαφορά ανάμεσα στις επιδημίες και σε άλλα φαινόμενα διάχυσης σε δίκτυα (π.χ διάδοση ιδεών ή αντιλήψεων) είναι ότι στις επιδημίες η διαδικασία μετάδοσης μπορεί να μοντελοποιηθεί σε ένα αφηρημένο επίπεδο παρατήρησης ως μια τυχαία διαδικασία η οποία επιτρέπει στον παθογόνο παράγοντα να εξαπλωθεί με μια συγκεκριμένη πιθανότητα από ένα μολυσμένο άτομο σε ένα υγιές κάθε φορά που τα δύο αυτά άτομα έρχονται σε επαφή. Κάτι τέτοιο δεν συμβαίνει στην περίπτωση ιδεών ή αντιλήψεων η διάδοση των οποίων προϋποθέτει τη λήψη σχετικών αποφάσεων από τη μεριά του ατόμου που τις υιοθετεί οι οποίες δεν είναι τυχαίες. [\[11\]](#)

Ένα παράδειγμα, για να καταλάβουμε πως η πληροφορία αναπαράγεται εκθετικά στο δίκτυο, δηλαδή πως μια είδηση γίνεται ευρέως γνωστή σε ελάχιστο χρόνο, είναι η εξέγερση τον Δεκέμβριο του 2008, μετά την δολοφονία του Αλέξανδρου Γρηγορόπουλου, όπου μέσω των κοινωνικών δικτύων, κυρίως οι νέοι της εποχής εκείνης, οργανώθηκαν, σε χρονικό διάστημα περίπου 20 λεπτών. Ακριβώς το ίδιο έγινε και στην γειτονική Τουρκία το 2013 στο πάρκο Γκεζί, στην πλατεία Ταξίμ της Κωνσταντινούπολης, δείγμα του πόσο σημαντικό μπορεί να είναι ένα κοινωνικό δίκτυο, τόσο σαν ένα μέσω ενημέρωσης, όσο και σαν μέσω χειραγώγησης την κοινής γνώμης. [\[12\]](#)

## Social Media Analytics

Ένα από τα μεγαλύτερα πλεονεκτήματα της χρήσης των social media για την επικοινωνία των εταιρειών είναι το ότι μπορούν να μετρήσουν την αποτελεσματικότητα της χρήσης τους αλλά και κάποιας συγκεκριμένης καμπάνιας τους. Ο όρος Social Media Analytics αναφέρεται στη διαδικασία μέτρησης, ανάλυσης και ερμηνείας των διαδράσεων και των συσχετισμών μεταξύ ανθρώπων, θεμάτων και ιδεών. Μια επιχείρηση μπορεί να εξάγει πολύτιμα συμπεράσματα για τα δημογραφικά στοιχεία των πελατών (ή δυνητικών πελατών) της, τις προτιμήσεις τους, τις αντιδράσεις τους αναφορικά με προϊόντα και υπηρεσίες και πολλά άλλα. Με προσεκτική ανάλυση, μια εταιρεία μπορεί επίσης να αναγνωρίσει τάσεις της αγοράς και με βάση αυτές να προσαρμόσει καλύτερα την στρατηγική της. Οι εταιρείες μπορούν μετρήσεις αφορούν κυρίως στην αποτελεσματικότητα των διαφημιστικών ενεργειών, ενώ άλλοι λόγοι ήταν η ανάλυση του brand, ο εντοπισμός ανταγωνιστικών πλεονεκτημάτων, η εξυπηρέτηση πελατών, η διερεύνηση πρόθεσης αγοράς, κ.ά. Πολλά από τα εργαλεία ανάλυσης των μέσων κοινωνικής δικτύωσης παρέχονται δωρεάν, αλλά υπάρχουν και αρκετά που κοστίζουν, ειδικά αυτά που απευθύνονται σε μεγάλες επιχειρήσεις.

Παρακάτω παρατίθεται μια λίστα από δωρεάν εργαλεία:

- Google Analytics, παρέχει στατιστικά για την κίνηση σε κάποιο site ή κοινωνικό δίκτυο (επισκεψιμότητα, δημογραφικά επισκεπτών)
- Facebook Insights, στατιστικά που παρέχει το Facebook για το κοινό κάθε σελίδας
- Social Mention, αναζήτηση περιεχομένου σε διάφορα κοινωνικά δίκτυα και άλλες πηγές (π.χ. για ένα προϊόν, εταιρεία, κ.τ.λ.)
- Social Mention, αναζήτηση περιεχομένου σε διάφορα κοινωνικά δίκτυα και άλλες πηγές (π.χ. για ένα προϊόν, εταιρεία, κ.τ.λ.)
- Trending, αναλύει καθημερινά το ελληνικό twitter (πιο δημοφιλείς χρήστες, θέματα, κ.τ.λ.)
- Sentiment Viz, παρουσιάζει το συναίσθημα των tweets και σχετικούς όρους με βάση ένα keyword/hashtag
- Sentiment140, εύρεση των tweets που περιέχουν τον όρο αναζήτησης και κατανομή των συναισθημάτων σε αυτά
- Topsy, παρέχει συγκρίσεις και τους top influencers για συγκεκριμένα θέματα
- Boardreader, παρέχει δεδομένα και στατιστικά για την εξέλιξη της συζήτησης σχετικά με έναν όρο αναζήτησης
- Klout, Υπολογίζει τη διαδικτυακή φήμη ενός χρήστη χρησιμοποιώντας δεδομένα από τα κοινωνικά δίκτυα που συμμετέχει [\[10\]](#)

Σημείωση: Σε αρκετά πανεπιστήμια, κυρίως του εξωτερικού, υπάρχουν ομάδες ερευνητών που σε συνεργασία με άλλους φορείς ή επιχειρήσεις ασχολούνται με την εξόρυξη δεδομένων (data mining). Κυρίως στα κοινωνικά δίκτυα, η εξόρυξη γνώσεως, είναι πολύ σημαντική, γιατί και ο όγκος της πληροφορίας είναι πολύ μεγάλος, αλλά και η ίδια η πληροφορία, είναι πολυσύνθετη, πράγμα που δυσκολεύει τον διαχωρισμό της πληροφορίας σε τάξεις. Κυρίως στην Αμερική τα πανεπιστήμια ξοδεύουν πολύ μεγάλα ποσά σε τέτοιου είδους έρευνες, φτιάχνοντας πολύ περίεργους αλγόριθμους ομαδοποίησης ή κατηγοριοποίησης της πληροφορίας.



---

## Κεφάλαιο 4

### Twitter API – Python – Gephi

---

#### Εισαγωγικά

Το πρακτικό κομμάτι της μεταπτυχιακής αυτής, όπως αναφέραμε και πιο πάνω, χωρίζετε σε 3 επιμέρους μέρη. Το 1<sup>ο</sup> είναι η ανάπτυξη κώδικα, ο οποίος θα παίρνει τα δεδομένα που θέλουμε, και θα τα βάζει σε ένα .txt αρχείο, με τέτοιο τρόπο ώστε να είναι ξεκάθαρο ποιός είναι η ρίζα (το δικό μας προφίλ), ποιά τα παιδιά, και ποιά τα εγγόνια. Ο κώδικας υλοποιήθηκε στο eclipse, σε γλώσσα Python.

Στο παράρτημα, στο τέλος της μεταπτυχιακής διατριβής παρουσιάζεται όλος ο κώδικας που χρησιμοποιήθηκε, για την εργασία αυτή, όπως και τα αρχεία .txt και αρχεία .gml.

Στο 2<sup>ο</sup> στάδιο έχουμε την μετατροπή του αρχείου .txt σε αρχείο .gml. Αυτό γίνεται γιατί τα αρχείο .txt δεν διαβάζεται από το Gephi. Για να φτιάξεις έναν γράφο στο Gephi, ή θα φτιάξεις μέσα από το Gephi, έναν-έναν τους κόμβους και τις ακμές, δημιουργώντας ένα αρχείο .gerhi, ή θα φτιάξεις ένα αρχείο .gml το οποίο θα το κάνεις εισαγωγή από το Gephi. Ο δεύτερος τρόπος είναι πιο σωστός, γιατί για οποιαδήποτε αλλαγή χρειαστεί να γίνει, αλλάζεις μόνο το αρχείο .gml και απλά το ξανακάνεις import στο Gephi.

Το 3<sup>ο</sup> και τελικό στάδιο είναι η εισαγωγή του κατάλληλου .gml αρχείου, ώστε να αποικονίζονται όλοι οι κόμβοι, και η εξαγωγή δεδομένων/πληροφοριών με βάση το γράφο, με τα εργαλεία του Gephi.

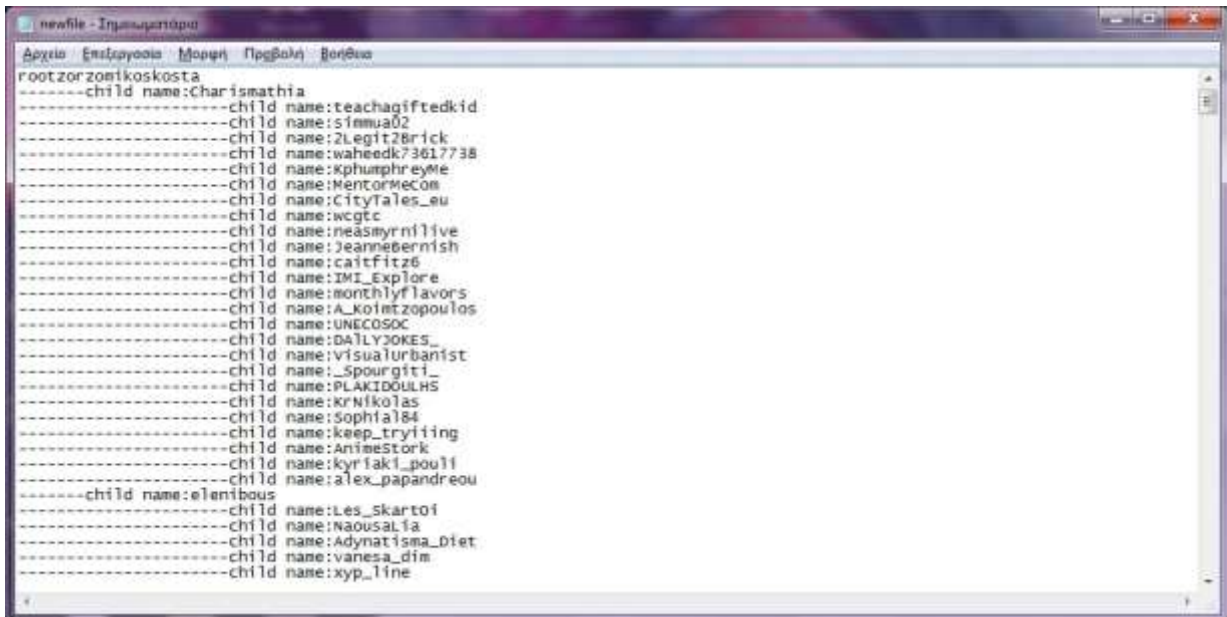
#### Twitter API

Το Twitter μας δίνει την δυνατότητα να πάρουμε κάποια δεδομένα, από τους χρήστες, κατεβάζοντας το Twitter API. Το επόμενο βήμα είναι να κάνουμε ένα developer account στο Twitter, εξηγώντας παράλληλα για ποιό λόγο το χρειαζόμαστε. (Εμείς βάλαμε για ερευνητικούς σκοπούς). Η διαδικασία συνεχίζεται βάζοντας το κινητό σου, full number, μαζί με το +30 δηλαδή, και μέσα σε μια μέρα σου στέλνουν στο κινητό σου ένα κωδικό. Όταν η διαδικασία ολοκληρωθεί επιτυχώς, στο τέλος σου στέλνουν στο e-mail σου 4 κωδικούς. Αυτοί είναι οι : CONSUMER\_KEY, CONSUMER\_SECRET, ACCESS\_TOKEN, ACCESS\_TOKEN\_SECRET. Από εκεί και πέρα, απλά κάνεις εισαγωγή το Tweepy στην Python, και βάζεις τους κωδικούς που σου έχουσε δώσει.

Αν όλα γίνουν σωστά, τότε θα πρέπει οι πρώτες γραμμές του κώδικα, να μοιάζουν όπως στη παρακάτω εικόνα.

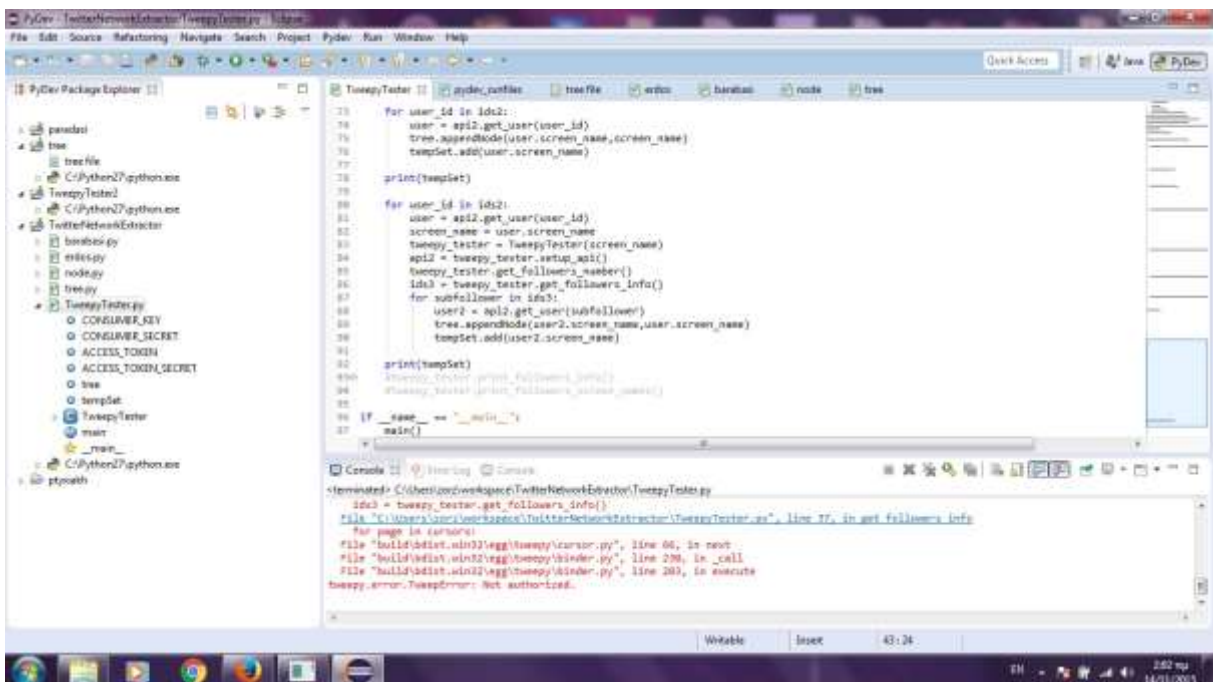


Στο τέλος, στο ίδιο path που είναι και η Python, θα δημιουργηθεί ένα αρχείο newfile.txt. Στο αρχείο αυτό θα εμφανίζονται τα ονόματα όλων, με τέτοιο τρόπο ώστε να φαίνεται τη σχέση έχει ο κάθε ένας με εμάς (είτε παιδί, είτε εγγόνι). Το αρχείο αυτό μοιάζει όπως παρακάτω στην Εικ.4



Εικ.4

Σε αυτό το σημείο θα πρέπει να πούμε πως υπάρχει μόνο ΜΙΑ περίπτωση να μην τρέξει ο κώδικας. Αν έχουμε συγγενή 1<sup>ου</sup> βαθμού, δηλαδή παιδί μας, ο οποίος έχει secure account, δηλαδή έχει κλειδωμένο λογαριασμό, λόγω ασφαλείας, τότε ο κώδικας ΔΕΝ θα τρέξει, και θα βγάλει το παρακάτω exception:



Εικ. 5

Εκεί χτυπάει το ίδιο το API, και βγάξει μήνυμα: 'Not authorized'. Το ενδεχόμενο αυτό όμως, είναι αρκετά μικρό.

Οι χρήστες που έχουν τέτοιο είδος προφίλ, έχουν προστατευμένα tweets, όπως λέγονται. Ουσιαστικά ακόμα και αν τους 'ακολουθείς' δεν έχεις ποτέ πρόσβαση στα δεδομένα τους, εικόνες, tweets κ.τ.λ. και τα προφίλ τους μοιάζουν όπως στην παρακάτω εικόνα 5β.

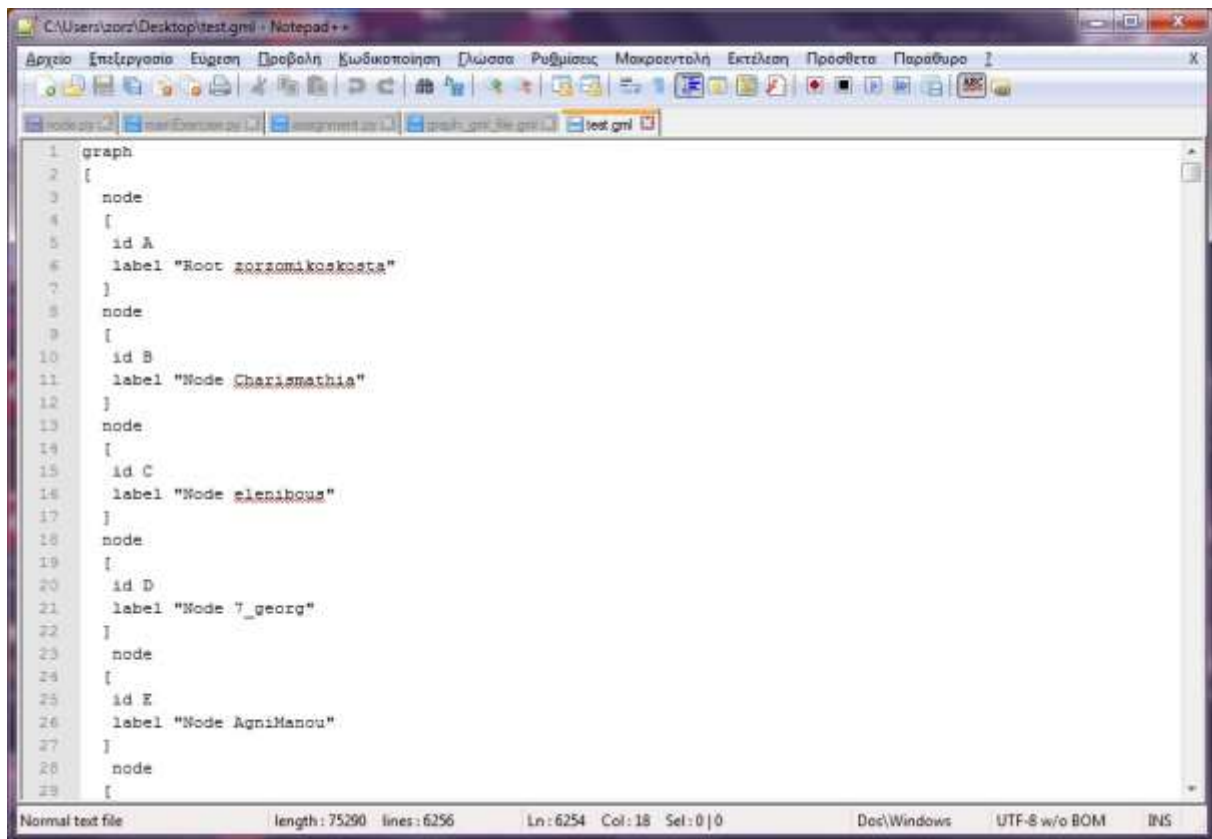


Εικ. 5β

Σημαντική παρατήρηση: Τα προφίλ των χρηστών αυτών ξεχωρίζουν από τα υπόλοιπα προφίλ, γιατί δίπλα από το όνομα υπάρχει ένα λουκέτο. Αν γίνει κάτι τέτοιο, δυστυχώς δεν τρέχει ΚΑΘΟΛΟΥ ο κώδικας, αρά δεν δημιουργείτε και το αρχείο .txt, που αναφέραμε παραπάνω. Υπενθυμίζω όμως πως οι πιθανότητες για κάτι τέτοιο είναι ελάχιστες.

Μετατροπή αρχείων .txt σε αρχεία μορφής .gml

Όταν τρέξει λοιπόν ο κώδικας, θα μας δημιουργηθεί ένα αρχείο .txt σαν και αυτό στην Εικ.4. Θα πρέπει να φτιάξουμε με βάση αυτό, ένα αρχείο .gml για να το εισάγουμε στο Gheri ώστε να βγει ο γραφός. Η διαδικασία είναι αρκετά απλή, αλλά χρονοβόρα. Χρειάζονται δυο πράγματα για να είναι σωστό ένα gml αρχείο. Χρειάζεται να φτιάξεις ένα κόμβο, και μία ακμή. Δηλαδή, ένα κόμβο, και μια ακμή, που θα δείχνει με ποιόν ή ποιούς άλλους ενώνεται αυτός ο κόμβος. Γιατί να έχεις ένα κόμβο, που να μην ενώνεται με κάτι, ή μια ένωση, χωρίς κόμβους, δεν έχει νόημα. Έτσι λοιπόν οι κόμβοι μοιάζουν στο αρχείο .gml σαν την παρακάτω εικόνα:



```
1 graph
2 {
3   node
4   [
5     id A
6     label "Root ΖορζομίκοςΚοστα"
7   ]
8   node
9   [
10    id B
11    label "Node Χαρισμάθια"
12  ]
13  node
14  [
15    id C
16    label "Node ελεσιθους"
17  ]
18  node
19  [
20    id D
21    label "Node Γ_georg"
22  ]
23  node
24  [
25    id E
26    label "Node ΑgniΜανου"
27  ]
28  node
29  [
```

Εικ.6

Ενώ οι ακμές μοιάζουν σαν την Εικ.7. Στην ουσία μας δείχνει, με άλλα λόγια, την σύνδεση, άρα την σχέση των 2 κόμβων.



```

2848 edge
2849 {
2850   source B
2851   target A
2852   label "Edge B to A"
2853 }
2854 edge
2855 {
2856   source C
2857   target A
2858   label "Edge C to A"
2859 }
2860 edge
2861 {
2862   source D
2863   target A
2864   label "Edge D to A"
2865 }
2866 edge
2867 {
2868   source E
2869   target A
2870   label "Edge E to A"
2871 }
2872 edge
2873 {
2874   source F
2875   target A
2876   label "Edge F to A"

```

Εικ.7

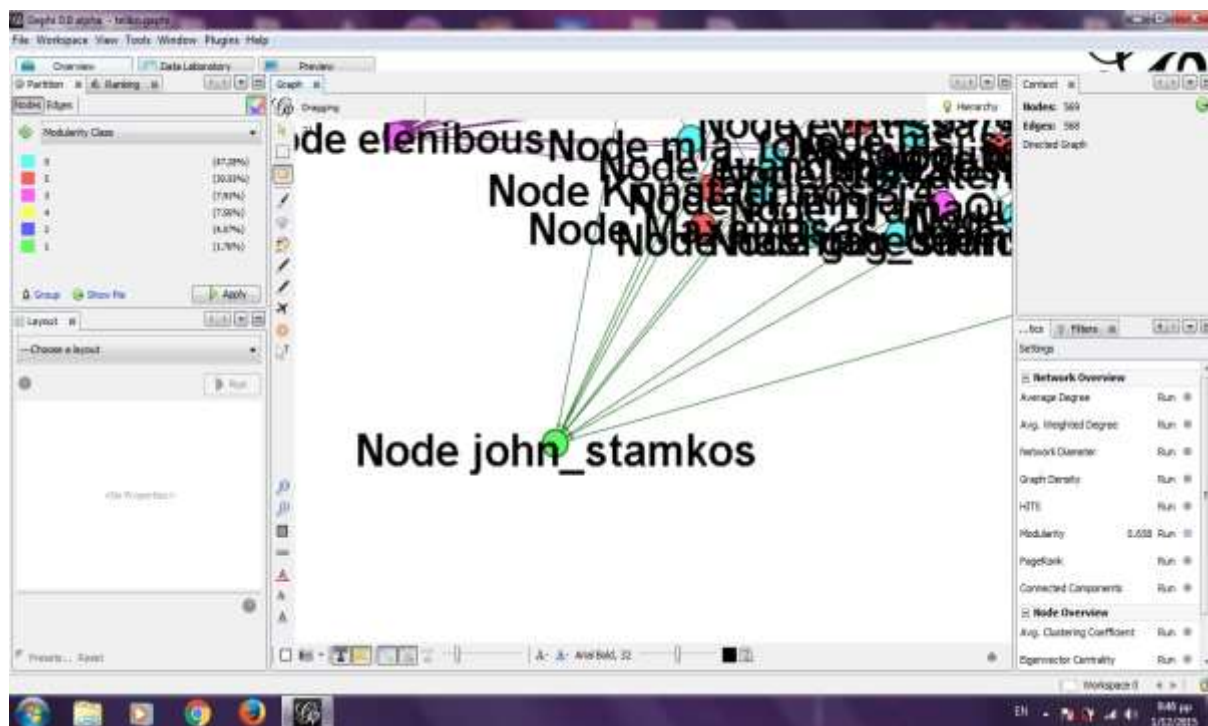
Τώρα που έχουμε φτιάξει το αρχείο μας σωστά, μένει μόνο η εισαγωγή του στο Gephi.

## GEPI

Το Gephi είναι ένα open source πρόγραμμα που κυρίως χρησιμοποιείτε για network analysis και visualization. Είναι γραμμένο σε γλώσσα προγραμματισμού, Java, σε NetBeans πλατφόρμα. Αναπαριστά συνήθως γραφήματα ή γράφους. Και έχει μια σειρά από εργαλεία, που μας βοηθούν είτε ταξινομώντας τα δεδομένα, με βάση κάποιο κριτήριο, είτε βγάζοντας στατιστικά για το γράφο, για παράδειγμα, τον μέσο όρο σύνδεσης ανά κόμβο κ.τ.λ. Πριν σας παρουσιάσουμε όμως τον δικό μας γράφο, θα πούμε δύο πράγματα γενικά για τους γράφους, τα είδη τους κ.ο.κ.

- Ένας **γράφος (graph)** αποτελείται από ένα σύνολο αντικείμενων που ονομάζονται **κόμβοι** (nodes) και από ένα σύνολο ζευγών από αντικείμενα τα οποία αποτελούν τις **ακμές** (edges) του γράφου.
- Δύο κόμβοι που συνδέονται με μια ακμή χαρακτηρίζονται ως **γείτονες** (neighbors).
- Όταν κάθε ακμή ενός γράφου έχει μια συγκεκριμένη κατεύθυνση, αναπαριστά δηλαδή μια συμμετρική σχέση μεταξύ δύο κόμβων, τότε ο γράφος ονομάζεται **κατευθυνόμενος** (directed). Στην αντίθετη περίπτωση ο γράφος ονομάζεται **μη κατευθυνόμενος** (non-directed) [\[13\]](#)

Στη δικιά μας περίπτωση έχουμε κατευθυνόμενο γράφο, γιατί αναπαριστά μια συμμετρική σχέση μεταξύ δυο κόμβων. Αυτό γίνεται εύκολα αντιληπτό και από το παρακάτω σχήμα.



Εικ. 8

Παρατηρώντας τον Node john\_stamkos βλέπετε τα βελάκια, που δείχνουν προς τον κόμβο, όποτε καταλαβαίνουμε την σχέση εκείνων των κόμβων, σε σχέση με αυτών.

Περισσότερα για την λειτουργία του προγράμματος Gephi, υπάρχουν στο επόμενο κεφάλαιο.

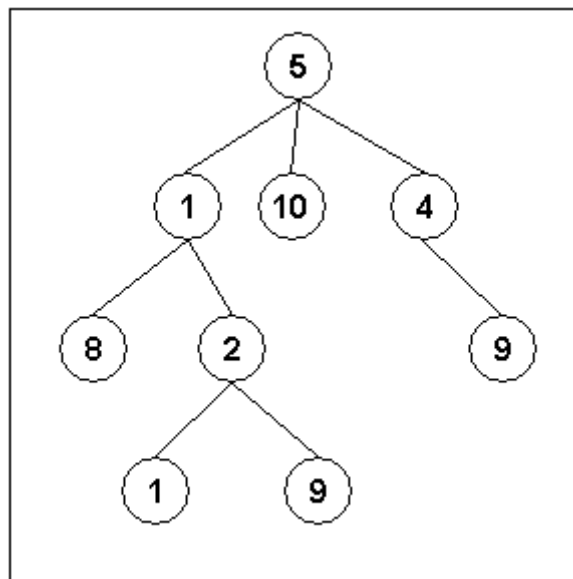
## ΚΕΦΑΛΑΙΟ 5

### ΓΕΡΗΙ – ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ – ΜΕΤΡΑ ΚΕΝΤΡΙΚΟΤΗΤΑΣ

#### ΓΕΡΗΙ

Για να αναπαραστήσουμε σωστά το δικό μας γράφο, πρέπει να προσέξουμε κάποια ενδεχόμενα.

Για παράδειγμα, σκεφτείτε, ότι εμείς είμαστε η ρίζα σε μια δομή τύπου δέντρου, όπου οι Followers μας είναι τα παιδιά μας, και οι Followers των Followers είναι τα εγγόνια. Υπάρχει μια περίπτωση, ας πούμε το εγγόνι μας, να έχει κάνει Follow εμάς ή το παιδί μας... Σε αυτή την περίπτωση αν έχουμε ήδη αυτή την σύνδεση, δεν πρέπει να την ξαναπάρουμε. Γιατί αν την ξαναπάρουμε, στο επόμενο βήμα θα μας ξαναβγάλει πάλι το εγγόνι μας, κ.ο.κ. με αποτέλεσμα, να έχουμε ένα ατέρμονα βρόχο. Για να το καταλάβουμε γραφικά, παρακαλώ παρατηρήστε την παρακάτω εικόνα.



Εικ. 9

Στο παράδειγμα της παραπάνω εικόνας, οι κόμβοι 1 και 9 πρέπει να τους πάρουμε μόνο μία φορά. (Προφανώς κρατάμε την πρώτη φορά, που είναι και το πιο σωστό). Σκεφτείτε να τρέξουμε τον κώδικα για το κόμβο 2, θα μας βγάλει τους κόμβου/χρήστες 1 και 9, που υπάρχουν ήδη. Άρα θα τρέξει μετά τον κώδικα για τον κόμβο 1, που έχει μέσα τον κόμβο 2, που με την σειρά του έχει ξανά τον κόμβο 1 κ.ο.κ.

Σαν απόρροια λοιπόν όλου αυτού, προκύπτουν δύο ζητήματα.

1. Με ποιόν τρόπο θα γίνεται η αναζήτηση των κόμβων στο δίκτυο; Κάθετα, με βάση την ρίζα ή οριζόντια, με βάση των βαθμό συγγένειας δηλαδή.
2. Με ποιόν τρόπο θα συγκρίνουμε το κάθε νέο κόμβο που βρίσκουμε, για να δούμε αν τυχόν, προ-υπάρχει, ώστε να μην το λαμβάνουμε υπ'όψιν μας.



Για αρχή, πρέπει να ξεκαθαρίσουμε πως υπάρχουν δύο διαφορετικοί αλγόριθμοι αναζήτησης σε δομές δέντρου. Οι αλγόριθμοί αυτοί είναι ο DFS (Depth-first search) και ο BFS (Breadth-first search). Ο πρώτος 'σαρώνει' ένα δέντρο κάθετα, από την ρίζα δηλαδή μέχρι τα φύλλα, ενώ ο δεύτερος, σαρώνει ένα δέντρο οριζόντια. Στην δικιά μας μεταπτυχιακή διατριβή χρησιμοποιήσαμε τον αλγόριθμο DFS, δηλαδή σαρώσαμε την δένδροειδή μας δομή, από πάνω προς τα κάτω, κάθετα. Μέχρι να πάρουμε και το τελευταίο φύλλο της δομής μας. Όσο για το άλλο θέμα, που είπαμε παραπάνω, για να μην παίρνουμε κόμβους που ήδη έχουμε, έχουμε βάλει μέσα στον κώδικα της αναζήτησης την συνάρτηση Set(). Η συνάρτηση αυτή παίρνει ορίσματα που είναι unique. Αν δηλαδή έχει ξαναπάρει την ίδια τιμή, στο παρελθόν, το αγνοεί και συνεχίζει, σαν να μην υπάρχει. Με αυτόν το τρόπο κρατάμε, σε περίπτωση που υπάρχει παραπάνω από ένα κόμβος, ΜΟΝΟ αυτό που βρήκαμε την πρώτη φορά. [\[14\]](#)  
[\[15\]](#)

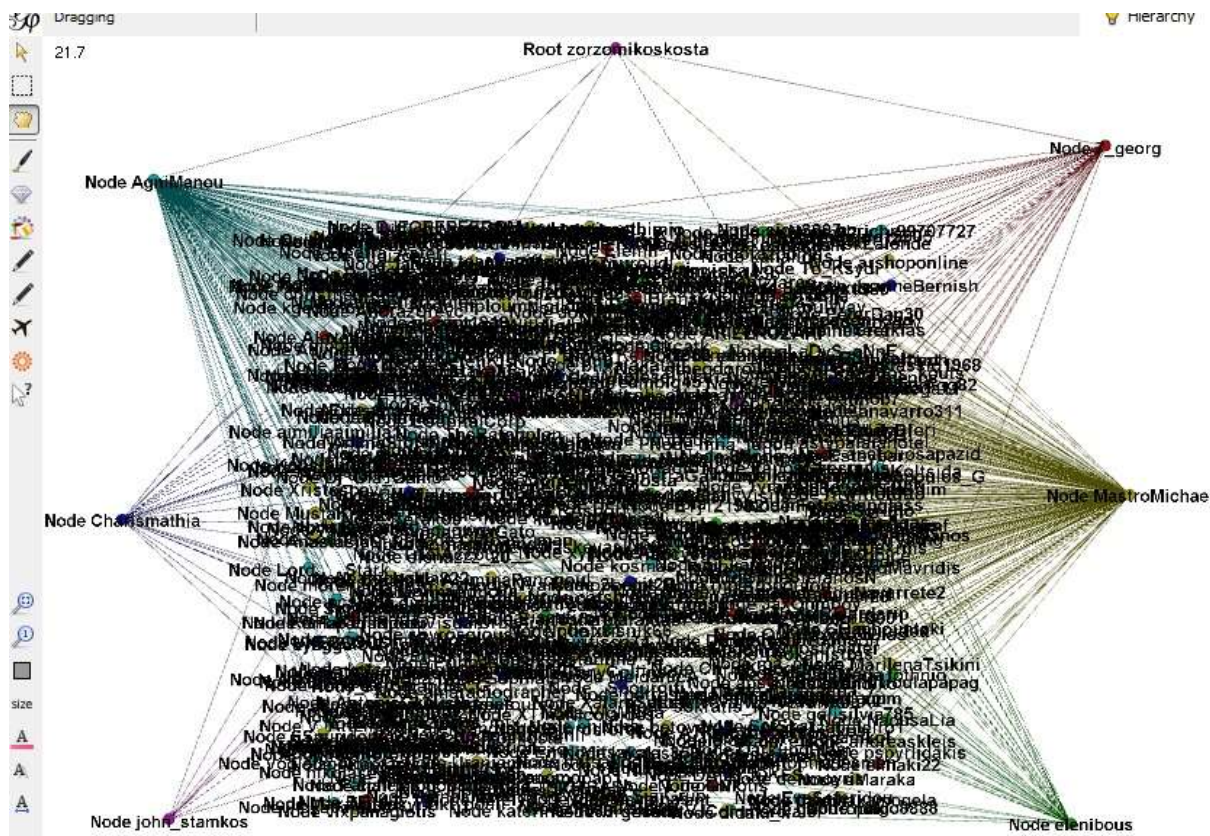
## ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Τώρα που ξέρουμε πως σαρώνουμε σωστά την δένδροειδή μας δομή, και πως δεν θα πάρουμε παραπάνω από μια φορά κάποιον χρήστη του Twitter, πάμε να δούμε κάποια χαρακτηριστικά των γράφων, ώστε να καταλάβουμε τι ακριβώς απεικονίζει ένας γράφος, με απώτερο σκοπό να μελετήσουμε τα μέτρα κεντρικότητας.

- Ορίζουμε ως **μήκος (length)** ενός μονοπατιού τον αριθμό των ακμών που περιέχει και ως **απόσταση (distance)** μεταξύ δυο κόμβων το μήκος του μονοπατιού με το ελάχιστο μήκος που συνδέει τους δυο αυτούς κόμβους.
- Ορίζουμε ως **διάμετρο (diameter)** του γράφου το μήκος του μονοπατιού με το μεγαλύτερο μήκος μεταξύ όλων των μονοπατιών ελάχιστου μήκους που υπάρχουν στον γράφο.

Το τι είναι μια ακμή, και ένας κόμβος, τα έχουμε εξηγήσει λεπτομερέστερα, παραπάνω στο Κεφάλαιο 4.

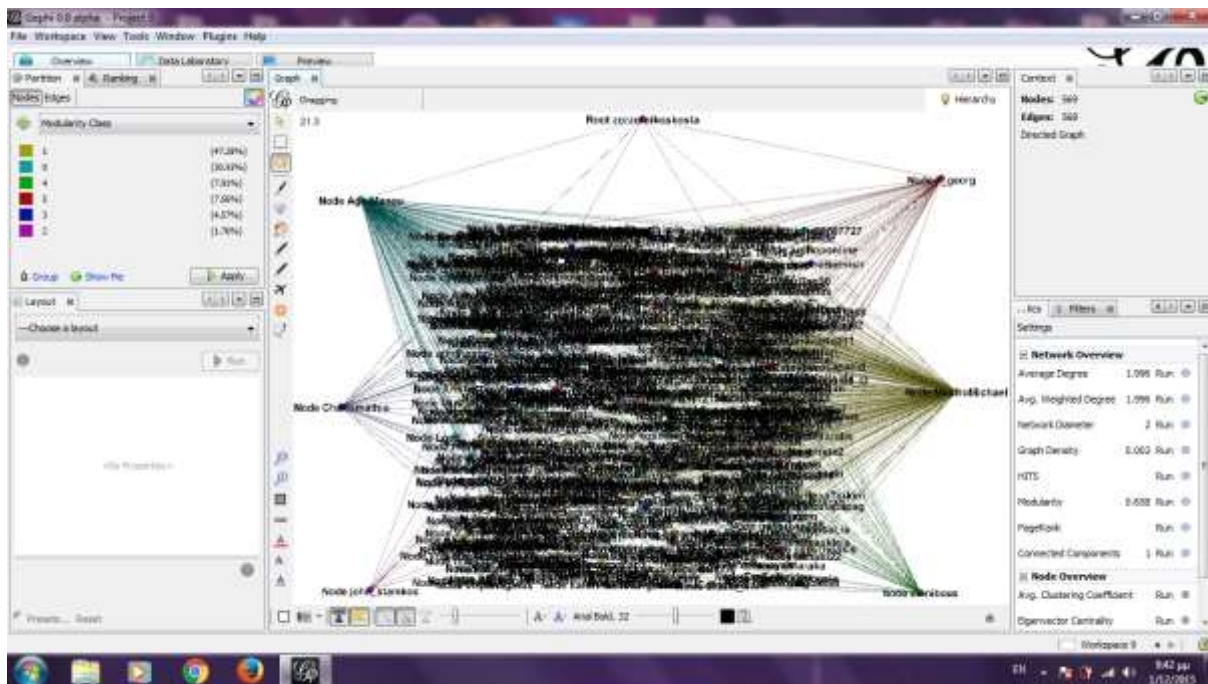
Έχοντας πλέον γνώση του τι είναι ακμή, κόμβος, μήκος μονοπατιού και διάμετρο, θα πάμε να αναλύσουμε το γράφο μας. Ο γράφος μας, όπως βγαίνει από το Gephi, είναι αυτός που φαίνεται στην παρακάτω εικόνα.



Εικ. 10

Πάνω-πάνω είμαστε εμείς, δεξιά-αριστερά, είναι τα παιδιά μας, και στο κέντρο, είναι τα παιδιά των παιδιών μας, τα εγγόνια μας. Έχουμε βάλει το κάθε μας παιδί και ένα διαφορετικό χρώμα, για να ξεχωρίζουνε, όσο είναι δυνατόν, μεταξύ τους. Το Gheri έχει και κάποια εργαλεία, που μας βοηθούν είτε κατανοώντας καλύτερα τον γράφο, είτε δίνοντας σε εμάς κάποια στατιστικά του γράφου, τα οποία θα ήταν πολύ δύσκολα να τα εξάγουμε εμείς διαφορετικά.

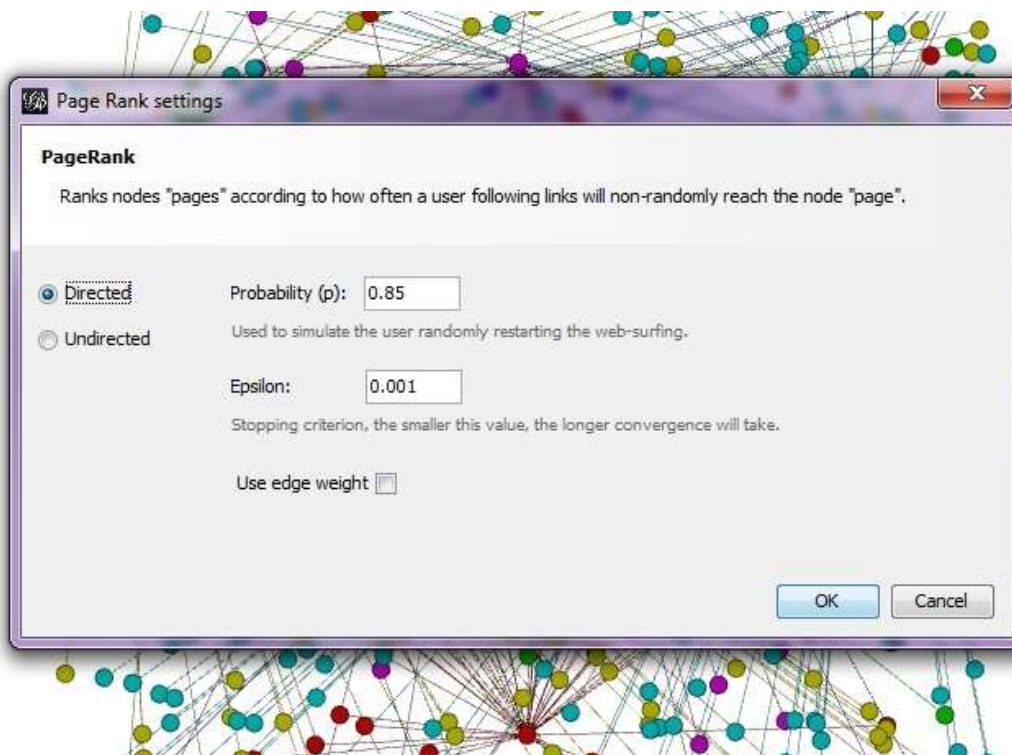
Για παράδειγμα, στην Εικ.11 πάνω δεξιά, μας δείχνει πόσες ακμές, και πόσους κόμβους έχουμε. Συνολικά έχουμε 569 κόμβους και 568 ακμές. Επίσης, κάτω δεξιά μας δείχνει κάποια στατιστικά, που χωρίς το Gheri θα ήταν αδύνατον να τα βγάλουμε. Ενδεικτικά να αναφέρουμε πως : η διάμετρος του δικτύου είναι 2. Η πυκνότητα είναι 0,002. Το Modularity(επεκτασιμότητα) είναι 0,658 και ο αριθμός 'κοινοτήτων' είναι 6. (δείτε Εικ.13)



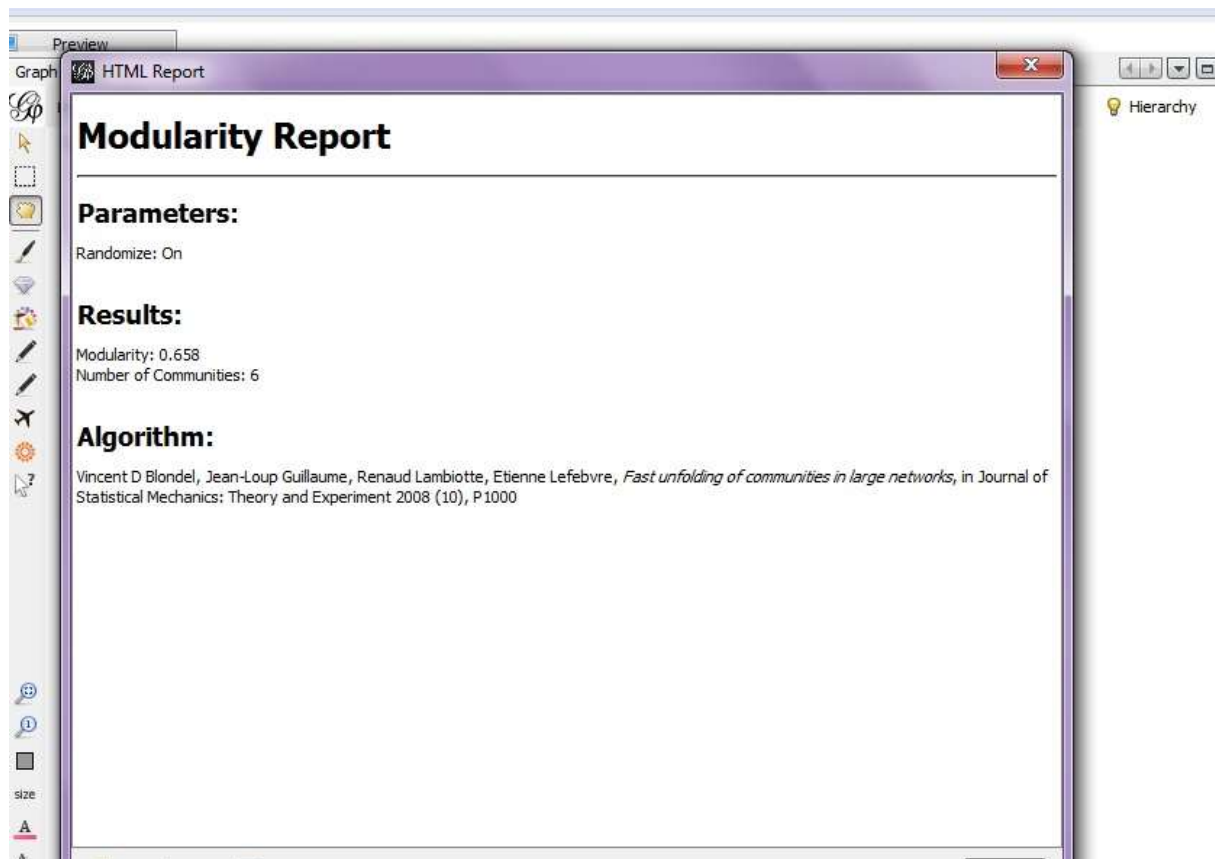
Εικ.11

Μια πολύ χρήσιμη δυνατότητα του Gephi είναι το ότι παίρνει το γράφο ανάποδα, δηλαδή από τα φύλλα προς την ρίζα, και με βάση κάποια δεδομένα μας βγάζει τι πιθανότητες υπάρχουν να κάνει π.χ ο εγγόνος μας Follow εμάς. Στο δικό μας γράφο η μέση πιθανότητα να μας κάνει follow ένα εγγόνι μας είναι 85%. Όπως φαίνεται και στην Εικ.12

ΣΗΜΕΙΩΣΗ: Η δικιά μας έκδοση του Gephi είναι η Gephi 0.8 version alpha.



Εικ.12

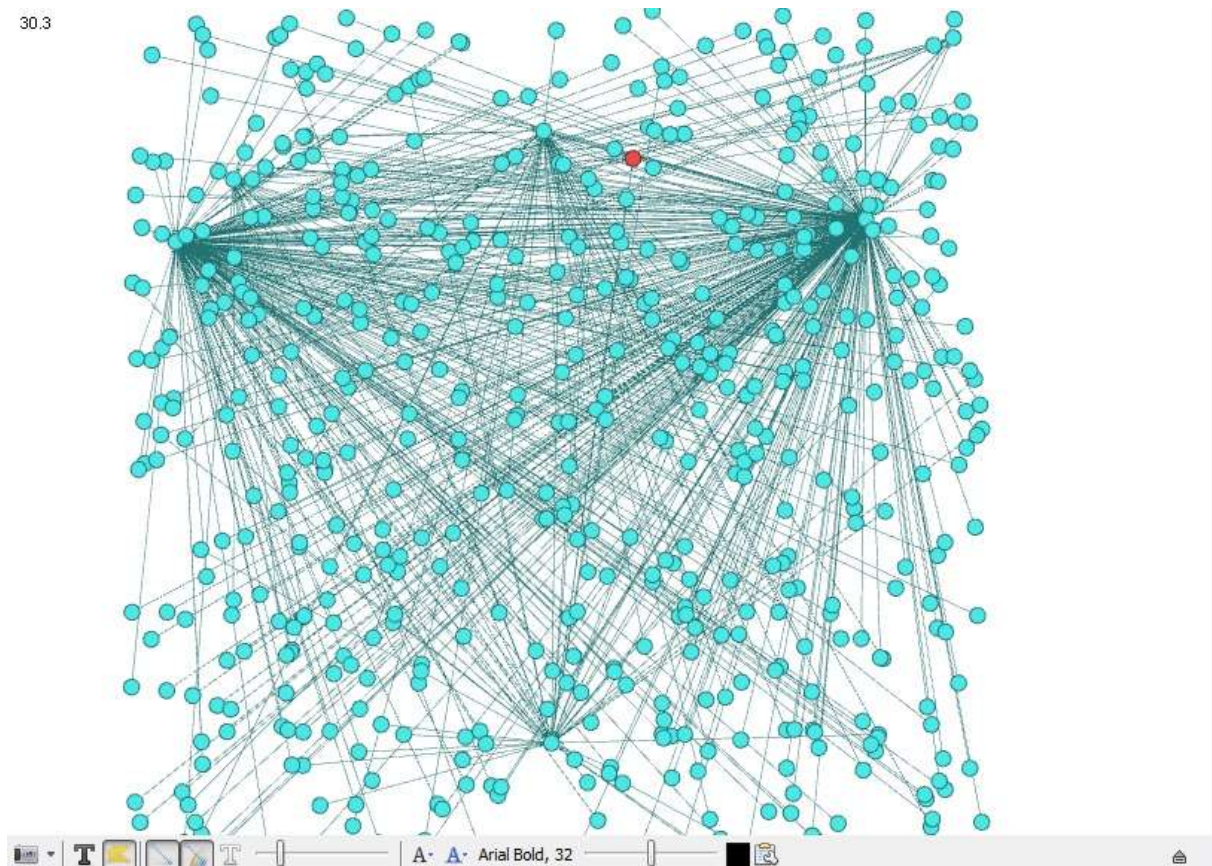


Εικ. 13

Στο τέλος, θα σας δείξουμε, γραφικά το πως είναι ο δικό μας κόμβος, στο πολυδιάστατο δίκτυο του Twitter. Όπως αυτό φαίνεται στην παρακάτω εικόνα.



30.3

**Εικ.14**

Με κόκκινο είναι ο δικό μας κόμβος, και με γαλάζιο οι κόμβοι του δικού μας δικτύου. Πραφανώς ο κάθε κόμβος, όπως και ο δικό μας, είναι και ένας ξεχωριστός/μοναδικός χρήστης του Twitter.

---

## Κεφάλαιο 6

### Μελλοντική ανάπτυξη. Παρατήρηση δικτύου με βάση το χρόνο. Προβλέψεις

---

Όπως φαίνεται και πιο πάνω στην Εικ.14, εμείς είμαστε απλά ένας κόμβος στο πολυδιάστατο δίκτυό μας, το οποίο είναι εντελώς ασήμαντο μπροστά στο τεράστιο δίκτυο του Twitter. Είναι εύκολα κατανοητό, πως στο δικό μας μικρό κομμάτι του δικτύου που μελετήσαμε, όσο μάλλον σε όλο το Twitter τα πράγματα δεν είναι στατικά, και συνέχεια, κάθε στιγμή αλλάζουν. Νέοι χρήστες μπαίνουν καθημερινά στο Twitter, ενώ άλλοι απενεργοποιούν τους λογαριασμούς τους για διάφορους λόγους. Σαν αποτέλεσμα όλου αυτού είναι να άλλοι κόμβοι να δημιουργούνται και άλλοι να απενεργοποιούνται. Αυτό σημαίνει πως τα πάντα αλλάζουν στο δίκτυο, και κυρίως τα μέτρα κεντρικότητας που μας ενδιαφέρουν. Τα μέτρα κεντρικότητας, όπως έχουμε πει πολλάκις στα πιο πάνω κεφάλαια, σχετίζονται κυρίως, γραφικά, με την πυκνότητα του γράφου. Στην ουσία, είναι τα πιο κεντρικά 'σημεία' ενός δικτύου, πράγμα που σημαίνει πως αν θέλουμε να φτάσουμε από ένα τυχαίο χρήστη σε έναν άλλο, πιθανότητα, σε κάποιο σημείο, σε κάποιο βήμα θα περάσουμε από τον κόμβο ή τους κόμβους αυτούς. Η χρησιμότητα τα όμως τα μέτρων κεντρικότητας, είπαμε πως ξεφεύγει από τα στενά όρια της επιστήμης της πληροφορικής, και σχετίζεται με πολλές άλλες επιστήμες. Στατιστικά, έχει αποδειχθεί πως οι κόμβοι αυτοί, είναι συνήθως άνθρωποι γνωστοί ήδη, που συχνά είναι πολιτικοί, καλλιτέχνες, αθλητές κ.ο.κ. Άρα, πέρα από το καθαρά 'τεχνικό' κομμάτι της μελέτης των μέτρων κεντρικότητας, έχει και μια κοινωνικοπολιτική πλευρά το όλο ζήτημα, γιατί οι άνθρωποι αυτοί έχουν μεγάλη σφαίρα επιρροής σε μεγάλο ποσοστό του πλυθησμού. Αυτό σημαίνει πως επηρεάζουν την κοινή γνώμη και διαμορφώνουν απόψεις πολλών ανθρώπων, και δη νέων, που είναι και η πλειοψηφία των χρηστών, των κοινωνικών δικτύων.

Στο σημείο αυτό, θα πρέπει αν πούμε πως το πιο 'έξυπνο' και σωστό που θα μπορούσαμε να κάνουμε, θα ήταν να παίρναμε σαν ρίζα, όχι το δικό μας λογαριασμό στο twitter, αλλά κάποιου που ξέρουμε, με βάση τα μέτρα κεντρικότητας, πως είναι πολύ σημαντικός κόμβος σε ΟΛΟ το δίκτυο του Twitter και να βλέπαμε πως θα διαμορφωθεί ο γράφος του μετά από Χ χρονικό διάστημα, και το πως θα μπορούσαμε να μελετήσουμε το ίδιο γράφο/δίκτυο, στο μέλλον, χωρίς να χάσουμε την ιστορικότητά του. Έτσι θα μπορούσαμε να δούμε τις αλλαγές ανά κόμβο, ακμή κτλ... και ίσως θα μπορούσαμε να προβλέψουμε μια πιθανή αλλαγή στο κόμβο, ξέροντας ποιοί κόμβοι είναι πιο ενεργοί από κάποιους άλλους. Οι μελέτη του δικτύου, και η πρόβλεψη, του πως θα μετατραπεί σε συνάρτηση του χρόνου, είναι κάτι αρκετά πιο δύσκολο, και πιο χρονοβόρο, και στην ουσία μιλάμε για έρευνα που ξεφεύγει από τα πλαίσια της πτυχιακής που καλούμαστε να κάνουμε. Θα μπορούσε όμως να είναι όμως μια έρευνα, σε επίπεδο διδακτορικού. Αυτό που θέλαμε να κάνουμε εμείς στα πλαίσια αυτής της μεταπτυχιακής διατριβής, είναι να δούμε πως η ιδέα μας μπορεί να υλοποιηθεί, έστω σε ένα μικρό κομμάτι του δικτύου, και πως αυτό που θέλουμε να βγάλουμε έχει όντως μια χρησιμότητα. Όλα αυτά με την λογική πως αν δουλέψει σε ένα μικρό κομμάτι του δικτύου, τότε θα δουλέψει και για την ολότητα του δικτύου. Εμείς βγάλαμε πως με βάση εμάς, και μελετώντας τους κόμβους μέχρι 2<sup>ου</sup> βαθμού από εμάς, στο δίκτυο, έχουμε 569 κόμβους, όλοι τους ενεργοί, και πως ο πιο σημαντικός από όλους είναι ο κόμβος με όνομα: MastroMichael ο οποίος έχει 270 Followers, δηλαδή μόνο αυτός απασχολεί σχεδόν το μισό δίκτυο. Με βάση τα μέτρα κεντρικότητας, είναι λοιπόν, μακράν ο πιο σημαντικός μας κόμβος, έχοντας τις πιο πολλές συνδέσεις από οποιοδήποτε άλλον.

---

## Κεφάλαιο 7

### Συμπεράσματα

---

Η ιδέα που είχαμε ήταν με βάση το δικό μας προφίλ στο Twitter, να βρούμε τα προφίλ των ανθρώπων που μας κάνουν Follow, δηλαδή των Followers μας, και να βρούμε και τα προφίλ των Followers των Followers. Βρίσκοντας τα προφίλ αυτά, (μέσω κώδικα, γραμμένα σε Python) φτιάξαμε ένα αρχείο .gml. Τα αρχεία αυτά ουσιαστικά αναπαραστούν κόμβους και την σύνδεση των κόμβων μεταξύ τους. Τέλος κάναμε εισαγωγή το αρχείου .gml στο Gephi. Το πρόγραμμα αυτό μας βγάζει τον αντίστοιχο γράφο, με βάση το αρχείο .gml. Αυτό που θέλαμε να πετύχουμε, και το πετύχαμε όντως είναι αφενός μεν η εξόρυξη γνώσεως από ένα κοινωνικό δίκτυο, και το πετύχαμε με τον κώδικα, παίρνοντας τα ονόματα όλων των χρηστών μέχρι 2<sup>ου</sup> βαθμού από εμάς. Αφετέρου δε η μελέτη του δικτύου, που δημιουργείται, δίνοντας κυρίως βάση στα μέτρα κεντρικότητας.

Ο σκοπός μας ήταν να παρατηρήσουμε το δίκτυο, μέρος του οποίου είμαστε και εμείς, και να βγάλουμε κάποια χρήσιμα συμπεράσματα. Πιστεύοντας πως αν ισχύουν αυτά που βγάλαμε στο μικρό αυτό κομμάτι του δικτύου του Twitter, θα ισχύουν γενικότερα.

Παρότι μικρό το δίκτυό μας, μόλις 569 κόμβους (μην ξεχνάμε ότι μελετήσαμε μόνο μέχρι και τα εγγόνια μας), βγάλαμε χρήσιμα συμπεράσματα. Μελετήσαμε την πυκνότητα του γράφου, είδαμε την δομή του, βγάλαμε στατιστικά στοιχεία, όπως τις πιθανότητες που υπάρχουν να μας κάνει στο μέλλον ένας εγγονός μας, εμάς Follow κ.ο.κ.

Δυστυχώς αυτό που δεν μπορέσαμε να μελετήσουμε, είναι η εξέλιξη του μικρού αυτού δικτύου, συναρτήση του χρόνου. Βέβαια αυτό ξεφεύγει από τα στενά όρια της εργασίας μας, γιατί αυτό είναι έρευνα, αλλά θα ήταν πολύ ωραία να βλέπαμε το πως αλλάζει μορφή το δίκτυο με την πάροδο του χρόνου. Αν το κάναμε αυτό ίσως να μπορούσαμε σε δεύτερο χρόνο να 'μαντέψουμε' το πως, και προς τα που θα εξαπλωθεί το δίκτυο. Γνώση η οποία θα ήταν πολύ χρήσιμη, όχι μόνο στον τομέα της πληροφορικής, αλλά γενικότερα. Θα μπορούσε για παράδειγμα αυτή η γνώση να βοηθούσε την ιατρική επιστήμη, να προβλέψει επιδημίες, ώστε να μην εξαπλωθούν περισσότερα.

---

## Κεφάλαιο 8

### Παράρτημα-Κώδικας Python

---

Στην Python έχουμε τον 'Κύριο κώδικα' ο οποίος τρέχει, και δυο άλλους κώδικες, σαφώς πιο μικρούς σε μέγεθος, που καλούνται όταν τρέχουμε τον Κύριο κώδικα.

Ο Κύριο κώδικας είναι ο «TweepyTester» και είναι ο εξής :

```
import tweepy
import time
import pprint
from tree import Tree
from warnings import catch_warnings
CONSUMER_KEY = 'SbxgAkBHPFq9HofpTsVx0ihXx'
CONSUMER_SECRET = 'KkCQ3Ipn3wKRCTPFLKTCfooAjkjxmKPeJM6XjGBPLbQsgu5Sbi'
ACCESS_TOKEN = '3317250454-05TkoAajS2vvL9E0IzkfTem6Bq9jwDK8uFGn3La'
ACCESS_TOKEN_SECRET = 'b060iScMfeZkWyFsQeyEjHbI9bYNPFJr1LOdWKFzu7RtU'
tree = Tree()
tempSet = set()

class TweepyTester(object):

    def initialize_parameters(self):
        self.consumer_key = CONSUMER_KEY
        self.consumer_secret = CONSUMER_SECRET
        self.access_token = ACCESS_TOKEN
        self.access_token_secret = ACCESS_TOKEN_SECRET
        print 'Access tokens were successfully initialized!'

    def __init__(self,screen_name):
        self.initialize_parameters()
        self.screen_name = screen_name
        print 'Class instance was successfully initialized!'

    def setup_api(self):
        auth = tweepy.OAuthHandler(self.consumer_key, self.consumer_secret)
        auth.set_access_token(self.access_token, self.access_token_secret)
        self.api=tweepy.API(auth)
        print 'Api was successfully initialized!'
        return self.api

    def get_followers_info(self):
        num = 1
        self.ids = []
        cursors = tweepy.Cursor(self.api.followers_ids, id =
self.screen_name).pages()
        for page in cursors:
            self.ids.extend(page)
```



```

        print "IDS of size %d were extended %d times on %s!"
%(len(self.ids),num,time.strftime("%a, %d %b %Y %H:%M:%S +0000", time.gmtime()))
        time.sleep(780)
        num = num + 1
    print len(self.ids)
    return self.ids

def get_followers_number(self):
    #for i in range(0,1000):
    #print i+1
    user = self.api.get_user(self.screen_name)
    print user.screen_name
    print user.followers_count
    #time.sleep(60)

def print_followers_screen_names(self):
    for user_id in self.ids:
        user = self.api.get_user(user_id)
        print user.screen_name
        #time.sleep(60)

def print_followers_info(self):
    pprint.pprint(self.ids)
    print "Number of followers ids extracted %d" %len(self.ids)

def main():
    try:
        file = open("newfile.txt", "w")
        screen_name = "zorjomikoskosta"
        great_name = "zorjomikoskosta"
        tempSet.add(screen_name)
        file.write("root"+screen_name+"\n")
        tree.appendNode(screen_name)

        tweepy_tester = TweepyTester(screen_name)
        api2 = tweepy_tester.setup_api()
        tweepy_tester.get_followers_number()
        ids2 = tweepy_tester.get_followers_info()

        for user_id in ids2:

            user = api2.get_user(user_id)
            tweepy_tester = TweepyTester(user.screen_name)
            api2 = tweepy_tester.setup_api()

            file.write("-----child name:"+user.screen_name+"\n")
            tree.appendNode(user.screen_name,great_name)
            tempSet.add(user.screen_name)

            tweepy_tester.get_followers_number()
            ids3 = tweepy_tester.get_followers_info()
            for subfollower in ids3:
                user2 = api2.get_user(subfollower)
                file.write("-----child
name:"+user2.screen_name+"\n")
                tree.appendNode(user2.screen_name,user.screen_name)
                tempSet.add(user2.screen_name)
            #edw kanei sleep gia n ananewnetai to rate limit
            time.sleep(780)

```

```

except tweepy.TweepError as e:
    if e.message[0]['code'] == 88:
        print("Tweepy Rate Limit Error.")
finally:
    file.close()
    print(tempSet)

#tweepy_tester.print_followers_info()
#tweepy_tester.print_followers_screen_names()

if __name__ == "__main__":
    main()

```

Ενώ οι δύο άλλοι κώδικες που τρέχουν μέσα από το Κύριο κώδικα είναι οι : tree και node που είναι οι εξής :

```

from node import Node

class Tree:

    def __init__(self):
        self.distinctIds = set()
        self.nodes = {}

    def getNodes(self):
        return self.nodes

    def appendNode(self, nodeId, parentId=None):
        newNode = Node(nodeId)
        self.nodes[nodeId] = newNode

        if(parentId is not None):
            self.nodes[parentId].appendChild(nodeId)

    def display(self, nodeId, depth=0):
        children = self.nodes[nodeId].children
        if depth == 0:
            print(nodeId)
        else:
            print("-----"*depth, nodeId)

        depth += 1
        for child in children:
            self.display(child, depth)

    def setAllDistinctNodeIds(self, nodeId):
        children = self.nodes[nodeId].children
        self.distinctIds.add(nodeId)
        for child in children:
            self.setAllDistinctNodeIds(child)

    def displayDistinctNodeIds(self):
        print (self.distinctIds)

```

και ο κώδικας Node :

```
class Node:

    def __init__(self, identifier):
        self.identifier = identifier
        self.children = []

    def getIdentifier(self):
        return self.identifier

    def getChildren(self):
        return self.children

    def appendChild(self, identifier):

        self.children.append(identifier)
```

---

## Κεφάλαιο 9

### Ευχαριστίες

---

#### Ευχαριστίες

Πρώτα απ' όλα, θέλω να ευχαριστήσω τον επιβλέποντα της μεταπτυχιακής εργασίας μου, Καθηγητή κ. Γεώργιο Α. Τσιχριντζή, για την πολύτιμη βοήθεια και καθοδήγησή του κατά τη διάρκεια της δουλειάς μου. Επίσης, είμαι ευγνώμων στον κύριο Σωτηρόπουλο Διονύσιο για την απέραντη βοήθεια όλους αυτούς του μήνες, χωρίς αυτήν δεν θα είχε πραγματοποιηθεί η εργασία αυτή. Πάνω απ' όλα, είμαι ευγνώμων στην μητέρα μου, Μαρία Ζορζομίκου για την ολόψυχη αγάπη και υποστήριξή της όλα αυτά τα χρόνια. Αφιερώνω αυτήν την μεταπτυχιακή εργασία στην μητέρα μου, ως ελάχιστο φόρο τιμής, στις θυσίες που έκανε καθημερινά.

Ζορζομίκος Κωνσταντίνος



## **ΒΙΒΛΙΟΓΡΑΦΙΑ-ΠΗΓΕΣ**

[1] [https://el.wikipedia.org/wiki/Κοινωνικό\\_δίκτυο](https://el.wikipedia.org/wiki/Κοινωνικό_δίκτυο)

[2] <https://el.wikipedia.org/wiki/Twitter>

[3] [www.itbiz.gr/κοινωνικα-δικτυα-social-networks-σε-μια-επιχείρηση/](http://www.itbiz.gr/κοινωνικα-δικτυα-social-networks-σε-μια-επιχείρηση/)

[4] University Linz, Institut for Pervasive Computing, Univ. Prof. Dr. Alois Ferscha, Social Computing- Pervasive Computing Systems and Environments WT 2014

[5] Ν. Μ. Σγούρος – Κοινωνικά Δίκτυα – Τμ. Ψηφιακών Συστημάτων, Παν. Πειραιώς. / ΣΤΟΙΧΕΙΑ ΘΕΩΡΙΑΣ ΓΡΑΦΩΝ

[6] Ν. Μ. Σγούρος – Κοινωνικά Δίκτυα – Τμ. Ψηφιακών Συστημάτων, Παν. Πειραιώς. / Χαρακτηριστικά & Μοντέλα Γράφων

[7] <https://marykeynotes.wordpress.com/2013/08/24/social-choice-theory/>

[8] <http://phdtheses.ekt.gr/eadd/handle/10442/29429> (Εθνικό Αρχείο Διδακτορικών Διατριβών)

[9] [https://el.wikipedia.org/wiki/Αλγόριθμος\\_PageRank](https://el.wikipedia.org/wiki/Αλγόριθμος_PageRank)

[10] [https://el.wikipedia.org/wiki/Μέσα\\_κοινωνικής\\_δικτύωσης](https://el.wikipedia.org/wiki/Μέσα_κοινωνικής_δικτύωσης)

[11] Ν. Μ. Σγούρος – Κοινωνικά Δίκτυα – Τμ. Ψηφιακών Συστημάτων, Παν. Πειραιώς. / ΕΠΙΔΗΜΙΕΣ

[12] [http://eurydice.lib.aegean.gr/\\_upload\\_/id1135/file0.pdf](http://eurydice.lib.aegean.gr/_upload_/id1135/file0.pdf)

[13] Ν. Μ. Σγούρος – Κοινωνικά Δίκτυα – Τμ. Ψηφιακών Συστημάτων, Παν. Πειραιώς. / **Στοιχεία Γράφων**

[14] [https://en.wikipedia.org/wiki/Depth-first\\_search](https://en.wikipedia.org/wiki/Depth-first_search)

[15] [https://en.wikipedia.org/wiki/Breadth-first\\_search](https://en.wikipedia.org/wiki/Breadth-first_search)

Ο,ΤΙ ΕΙΝΑΙ ΥΠΟΓΡΑΜΜΙΣΜΕΝΟ ΕΙΝΑΙ ΙΣΤΙΟΓΡΑΦΙΑ, ΕΝΩ Ο,ΤΙ ΔΕΝ ΕΙΝΑΙ ΥΠΟΓΡΑΜΜΙΣΜΕΝΟ ΕΧΕΙ ΠΑΡΘΕΙ ΕΙΤΕ ΑΠΟ ΣΗΜΕΙΩΣΕΙΣ, ΕΙΤΕ ΑΠΟ ΑΛΛΕΣ ΠΤΥΧΙΑΚΕΣ(ΜΕΤΑΠΤΥΧΙΑΚΩΝ Η ΔΙΔΑΚΤΟΡΙΚΩΝ ΦΟΙΤΗΤΩΝ) ΕΙΤΕ ΑΠΟ ΡΑΡΕΡΣ ΔΗΜΟΣΙΕΥΜΕΝΑ. ΟΛΑ ΑΥΤΑ ΥΠΑΡΧΟΥΝ ΣΤΟ ΦΑΚΕΛΟ ΠΟΥ ΘΑ ΔΟΘΕΙ ΜΑΖΙ ΜΕ ΤΗΝ ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ, ΣΕ ΜΟΡΦΗ .PDF

Copyright © 2016 – Ζορζομίκος Κωνσταντίνος

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. 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.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν αποκλειστικά τον συγγραφέα και δεν αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Πειραιώς.