

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ



ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

ΤΗΚΕVASE KSANA ΕΚFONISI GΙΑ TO TΙ ΖΗΤΑ

Your Subtitle

January 21, 2024

Περιεχόμενα

1	Εισαγωγή	2
2	Λήψη δεδομένων	2
3	Γραφική Αναπαράσταση Δικτύου	3
4	Βασικά στοιχεία Δικτύου	6
5	Component Measures	7
6	Degree measures	9

1 Εισαγωγή

Το Youtube είναι ένας ιστότοπος κοινοποίησης, αποθήκευσης, αναζήτησης και αναπαραγωγής βίντεο. Κάθε χρήστης μπορεί να δημιουργήσει λογαριασμό και να ανεβάζει τα δικά του βίντεο ή ακόμα και να αναπαράγει σε πραγματικό χρόνο. Εκτός από τους χρήστες, πρόσβαση έχει ο οποιοσδήποτε στον ιστότοπο αυτό όπου μπορεί μόνο να παρακολουθεί τα βίντεο άλλων χρηστών. Το προφίλ του χρήστη παρουσιάζεται ως κανάλι όπου άλλοι χρήστες μπορούν να εγγραφούν ώστε να παρακολουθούν και να ενημερώνονται για βίντεο ή για πραγματικού χρόνου αναπαραγωγές που τους ενδιαφέρουν. Τα βίντεο που ανεβάζει ο κάθε χρήστης είναι συνήθως αποθηκευμένα σε playlists αναλόγως με την μορφή και το θέμα που έχουν. Επίσης στο κανάλι του ο κάθε χρήστης μπορεί να έχει κανάλια άλλων χρηστών που όπως αναφέρονται στην αγγλική ορολογία "Featured channels". Τα επιλεγμένα αυτά κανάλια αποτελούν κανάλια όπου ένας χρήστης επιλέγει να τα συμπεριλάβει στο δικό του κανάλι (δεν φαίνονται στο κοινό). Ο λόγος που γίνεται αυτό είναι για να προωθούν οι χρήστες και να εμφανίζουν άλλα κανάλια που τους αρέσουν, με τα οποία μπορεί να συνεργάζονται ή να θέλουν να τα προτείνουν στους θεατές τους. Έτσι με αυτό τον τρόπο, οι χρήστες μπορούν να προσεγγίσουν πολλά είδη κοινού και να αυξήσουν έτσι τις εγγραφές και τις προβολές τους. Στην ανάλυση αυτή θα εξετάσουμε το κανάλι Samsung. Το κανάλι αυτό είναι το κανάλι του ομίλου εταιρειών Samsung που έχει ως σκοπό την ενημέρωση σχετικά με εκδηλώσεις, καινοτομες τεχνολογίες, εφαρμογές και υπηρεσίες, B2B solutions, παρουσιάσεις, και τις τελευταίες και καινοτόμες τεχνολογίες του ομίλου.

2 Λήψη δεδομένων

Τα δεδομένα για την ανάλυση μας τα πήραμε με τη χρήση του [Bernhard Reiner's Tool](#) χρησιμοποιώντας τα YouTube Data Tools. Αρχικά, χρησιμοποιώντας το link του καναλιού στο YouTube, βρήκαμε το id του καναλιού μέσω του [Channel Info Module](#). Έπειτα με τη χρήση του [Channel Network Module](#), πήραμε δεδομένα για το δίκτυο του καναλιού. Οι παραμέτροι που χρησιμοποιήθηκαν ήταν το seed (αρχικό κανάλι) με τη χρήση του id με crawl depth ίσον με 2 (το crawl depth καθορίζει πόσο βαθιά στο δίκτυο μπορούμε να φτάσουμε. Για παράδειγμα με depth=0 το εργαλείο αυτό επιστρέφει το δίκτυο με τις συσχετίσεις ανάμεσα στα seeds που δίνονται, με dept=1 επιστρέφει τα featured channels που έχει ο χρήστης στο κανάλι του και με depth=2 επιστρέφει τα featured channels που υπάρχουν στα κανάλια που βρήκαμε στο depth=1). Η επιλογή για της εγγραφές δεν λήφθηκε υπόψην διότι θέλαμε τα δεδομένα να είναι μόνο με τα featured channels. Μετά από αυτά τα βήματα το εργαλείο δημιούργησε ένα gdf αρχείο το οποίο φορτώσαμε στο πρόγραμμα Gephi για ανάλυση. Εδώ να σημειωθεί ότι μέσω του Gephi έγινε έλεγχος των δεδομένων για τυχόν σφάλματα που θα μπορούσαν να

επηρεάσουν την ανάλυση μας όπως για παράδειγμα ο έλεγχος δυπλοτύπων, όπου σε μια περίπτωση υπήρχε διπλότυπο όπου και εντιμετωπίστηκε μέσω του Gephii, ο έλεγχος για null τιμές κ.α. Σε μερικές περιπτώσεις υπήρχαν μη διαθέσιμες τιμές. Για παράδειγμα σε ορισμένους κόμβους, δεν υπήρχε στο αντίστοιχο κελί η χώρα ενώ ήταν γνωστή. Επομένως εισήχθησαν χοιροκίνητα οι τιμές όπου ήταν εφικτό. Σε άλλες περιπτώσεις, τυχόν σφάλματα αντιμετωπίζονται αναλόγως τη δεδομένη στιγμή όπου και αναφέρονται.

3 Γραφική Αναπαράσταση Δικτύου

Το δίκτυο μόνο με τα ονόματα των κόμβων(καναλιών) χωρίς κάποια παραμετροποίηση.



Βλέποντας τα δεδομένα του δικτύου μας απο το Data Laboratory του Gephi, παρατηρήσαμε πως υπάρχουν κανάλια απο διάφορες χώρες. Επομένως θεωρήσαμε ενδιαφέρον να κάνουμε μία παραμετροποίηση με τις χώρες ως εξής. Ο χρωματισμός έγινε μέσω διαφορετικών χρωματων, τωσων, όσος και ο αριθμός των διαφορετικών χωρών, μέσω του partition tab. Στο σημείο αυτο, θεώρησαμε επίσης σημαντικό και την αναφορά του seed. Αυτό έγινε μεσω του μεγέθους των κόμβων μέσω του seedrank(αντίστοιχη μεταβλητη με την isseed εαν χρησιμοποιούσαμε τον χρωματισμό). Στη συνέχεια μέσω του Plugin Circular Layout που κατεβάσαμε μέσω των Tools του Gephi, δημιουργήσαμε την πιο κάτω διάταξη θέτοντας στην ιδιότητα "Order Nodes By" την χώρα. Για άλλη μια φορά, χρησιμοποιήσαμε τον Label Adjust για διαχωρισμό των ετικετών.



Απο την πιο πάνω εικόνα μπορούμε εύκολα να παρατηρήσουμε πως ο κεντρικός και ίσως ο πιο σημαντικός κόμβος να είναι ο "Samsung" ο οποίος είναι με πράσινο χρώμα. Οι δύο δεξιές θέσεις απο αυτο το κόμβο είναι επίσης με πράσινο χρώμα αφού και αυτοι οι κόμβοι είναι κανάλια απο την ίδια χώρα, την Νότιο Κορέα.

4 Βασικά στοιχεία Δικτύου

Το δίκτυο που μελετάμε έχει τα εξής βασικά στοιχεία:

- Αριθμός κόμβων: **76** διαφορετικά **κανάλια-κόμβοι**
- Αριθμός ακμών: **149** **συνδέσμοι** μέσω των οποίων συνδέονται τα κανάλια-κόμβοι
- Ο γράφος μας είναι **κατευθυνόμενος**. Δηλαδή κάθε σύνδεσμος απο ένα κανάλι προς ένα άλλο έχει κατεύθυνση όπως φαίνεται στην πιο κάτω εικόνα:



Ο πράσινος κόμβος-κανάλι έχει ως featured channel τον κόμβο-κανάλι με ροζ χρώμα.

- Διάμετρος δικτύου: Η **διάμετρος** ενός δικτύου είναι η μακρύτερη συντομότερη διαδρομή που μπορούμε να βρούμε. Στην περίπτωση μας είναι **3**. Τιμή αναμενόμενη λόγω του depth με τιμή 2 που επιλέξαμε.
- **Average path length**: Είναι ο **μέσος όρος των συντομότερων μονοπατιών** για όλα τα ζεύγη κόμβων. Στο δίκτυο μας είναι **1.9760**.

Parameters:

Network Interpretation: directed

Results:

Diameter: 3

Radius: 0

AveragePathlength: 1.9760319573901464

5 Component Measures

Στο δίκτυο μας, όλοι οι κόμβοι είναι συνδεδεμένοι μεταξύ τους(έμμεσα είτε άμεσα). Άρα μπορούμε να πούμε πως υπάρχει **ένα giant component**. Επομένως ο αριθμός των **weakly connected components** είναι ίσος με **1**.

Αναφορικά με τον αριθμό των **strongly connected components**, αυτό που πρέπει να δούμε στην περίπτωση μας είναι αν υπάρχουν κανάλια-κόμβοι τα οποία δεν έχουν Featured Channels, δηλαδή δεν έχουν εξερχόμενους συνδέσμους. Έτσι μέσω του Connected Components tool απο το πεδίο Statistics του Gephi έχουμε την ακόλουθη αναφορά.

Connected Components Report

Parameters:

Network Interpretation: directed

Results:

Number of Weakly Connected Components: 1

Number of Strongly Connected Components: 57

Παρατηρώντας την πιο πάνω εικόνα λοιπόν, μπορούμε να επιβεβαιώσουμε τον αριθμό των weakly connected components. Όσον αφορά τον αριθμό των strongly connected components μέσω του Gephi βλέπουμε πως είναι **57**. Στο σημείο αυτό μπορούμε να εφαρμόσουμε μια διάταξη για να δούμε σχηματικά αυτούς τους κόμβους ώστε να καταλάβουμε καλύτερα τι συμβαίνει. Χρησιμοποιώντας λοιπόν τον αλγόριθμο Dual Circle Layout, με Upper Order Count ίσο με 20(Πλήθος κόμβων - strong connected components + weakly connected components) με σκοπό να πάρουμε στον εξωτερικό κύκλο τα κανάλια που δεν έχουν Featured Channels(20 κανάλια, 20 διαφορετικά χρώματα). Έτσι όπως φαίνεται και πιο κάτω, στον εξωτερικό κύκλο, τα κανάλια αυτά έχουν ακμές που φτάνουν σε αυτά και κανένα δεν έχει ακμή που να ξεκινάει από αυτά.

6 Degree measures

Degree measures, such as maximum and average node degrees, as well as degree distribution.