

Ψηφιακές Τεχνολογίες στην Εκπαίδευση

Εκτίμηση αμοιβαιότητας στις προβολές του
FollowFriday μεταξύ χρηστών του Twitter

Δαούσης Δημήτριος, Α.Μ.: 403
Λαζαρόπουλος Σπυρίδων, Α.Μ.: 414

Μεταπτυχιακοί φοιτητές Α έτους του Μαθηματικού Πατρών
Κατεύθυνση: Σύγχρονες Εφαρμογές - Τ.Π.Ε. στην Εκπαίδευση

17 Φεβρουαρίου 2010

Περιεχόμενα

1	Εισαγωγή	2
2	Ορολογία του Twitter	3
2.1	Κατηγορίες μηνυμάτων Twitter	4
2.1.1	Update tweets	4
2.1.2	Replies και mentions tweets (@reply, @mention)	4
2.1.3	Direct messages	5
2.1.4	ReTweets (RT)	5
2.2	#FollowFriday ή #FF	5
3	Συλλογή δεδομένων FF του Twitter	6
3.1	Χρήση του εργαλείου Archivist	6
4	Κατασκευή δικτύων FF	8
4.1	Κανόνες σχηματισμού τόξων	8
4.2	Χρήση του λογισμικού iGraphitter	9
4.3	Αναπαράσταση δικτύου FF με το NodeXL	12
5	Στατιστική ανάλυση δικτύων FF του Twitter	13
5.1	Βαθμοί εισόδου και εξόδου κόμβων δικτύου	13
5.2	Αμοιβαιότητα σχέσης	14
5.3	Αποτελέσματα για τις 6 πρώτες Παρασκευές	14
6	Μελλοντική εργασία - Προοπτικές	17

1 Εισαγωγή

Τα κοινωνικά μέσα (social media) αποτελούν πλέον αναπόσπαστο κομμάτι των εφαρμογών των ψηφιακών τεχνολογιών με αποτέλεσμα να εμφανίζονται όλο και περισσότερες υλοποιήσεις και υπηρεσίες κοινωνικής δικτύωσης. Μερικές από αυτές είναι το MySpace, το Friendster, το FaceBook, το LinkedIn και φυσικά το Twitter με το οποίο ασχολούμαστε στα πλαίσια αυτής της εργασίας. Τα παραπάνω χρησιμοποιούνται ευρέως και είναι πολύ δημοφιλή από τους χρήστες του Διαδικτύου, τόσο σε ατομικό, όσο και σε εταιρικό αλλά και ακαδημαϊκό επίπεδο. Στην παρούσα εργασία ασχολούμαστε με το Twitter και πιο συγκεκριμένα γίνεται μια προσπάθεια να μελετήσουμε κάποια χαρακτηριστικά του. Ειδικότερα, προσπαθούμε να εκτιμήσουμε το βαθμό της αμοιβαιότητας στις μεταξύ τους προβολές που κάνουν οι χρήστες του Twitter κάθε Παρασκευή μέσω της ετικέτας #FollowFriday (#FF).

Ως γνωστό, ένα κοινωνικό δίκτυο αναπαρίσταται από ένα γράφο, ο οποίος αποτελείται από ένα ή περισσότερα (πεπερασμένα) σύνολα δρώντων (οι κόμβοι ή κορυφές του γράφου) και από μία ή περισσότερες σχέσεις (οι συνδέσεις του γράφου), που συνδέουν τους δρώντες μεταξύ τους. Έτσι και στην περίπτωση που μελετάμε, στα μηνύματα FollowFriday κάποιοι χρήστες προβάλλουν (δηλαδή, προτείνουν ή συστήνουν ή διαφημίζουν) κάποιους άλλους χρήστες για να τους ακολουθήσουν οι άλλοι χρήστες που το επιθυμούν. Αυτή η σχέση προβολής (ή διαφήμισης) είναι μια κατευθυνόμενη σχέση, οπότε ο αντίστοιχος γράφος με δρώντες τους χρήστες του Twitter είναι ένας κατευθυνόμενος γράφος (ή δίγραφος).

Ας θεωρήσουμε λοιπόν, ότι έχουμε έναν κατευθυνόμενο γράφο, όπου κόμβοι είναι οι χρήστες του Twitter, οι οποίοι προβάλλουν άλλους χρήστες μέσω της ετικέτας #FollowFriday, ενώ τα τόξα που συνδέουν ζεύγη κόμβων αντιστοιχούν στις σχέσεις προβολής μεταξύ των χρηστών. Ο κατευθυνόμενος αυτός γράφος λέμε ότι είναι το δίκτυο FF του Twitter.

Στην εργασία μας, η οποία είναι μέρος ενός ευρύτερου ερευνητικού έργου υπό την καθοδήγηση του κ Μ. Μπουντουρίδη, ακολουθούμε την εξής διαδικασία, αρχής γενομένης από την Παρασκευή 20 Νοεμβρίου 2009 και για κάθε Παρασκευή στο επόμενο διάστημα:

1. Συγκέντρωση των μηνυμάτων του Twitter με ετικέτα # (hashtag) το #FollowFriday ή #FF.
2. Κατασκευή του δικτύου FF (υπό μορφή αρχείου Pajek).

3. Οπτικοποίηση (visualization) του δικτύου FF.
4. Υπολογισμοί της αμοιβαιότητας και των βαθμών εισόδου και εξόδου (in- και out-degree) όπως και των κατανομών τους.
5. Στατιστική εκτίμηση της αμοιβαιότητας στις συνδέσεις τους με χρήση του πακέτου ERGM του R.

Για το σκοπό αυτό χρησιμοποιούμε τα εξής εργαλεία – λογισμικά:

1. Archivist
2. iGraphitter (δικό μας λογισμικό, που δημιουργήσαμε για τις ανάγκες της εργασίας)
3. NodeXL
4. The R-Project / ERGM package

Αρχικά, στην ενότητα 2, θα αναφερθούμε και θα εξηγήσουμε την ορολογία που χρησιμοποιείται στο Twitter και τη δομή των μηνυμάτων που διακινούνται μεταξύ των χρηστών του. Στην ενότητα 3 θα περιγράψουμε αναλυτικά την συλλογή των δεδομένων για τα δίκτυα FF του Twitter, τα οποία θα τα συζητήσουμε στην ενότητα 4 με βάση ένα απλό παράδειγμα. Τέλος, στην ενότητα 5 θα κάνουμε τη στατιστική ανάλυση για την εκτίμηση της αμοιβαιότητας των συνδέσεων των δικτύων αυτών.

2 Ορολογία του Twitter

Το Twitter είναι ένα σύστημα **κοινωνικής δικτύωσης (social networking)**, στο οποίο ένας χρήστης του Twitter (μερικές φορές, ο χρήστης του Twitter λέγεται **tweeple**) μπορεί να στέλνει και να λαμβάνει μηνύματα του Twitter (τα οποία ονομάζονται **tweets**). Πιο συγκεκριμένα, κάθε μήνυμα του Twitter (δηλαδή, κάθε tweet) είναι μια ανάρτηση (post) ενός μικρού κειμένου, που αποτελείται, το πολύ, από 140 χαρακτήρες. Ένα tweet μοιάζει πολύ με ένα μήνυμα κινητού τηλεφώνου (SMS).

Το μήνυμα που στέλνει ένας χρήστης, το παίρνουν όλοι οι χρήστες που έχουν δηλώσει ότι είναι **φίλοι-με (followers ή friends)** τον αποστολέα, κι έτσι, τον ακολουθούν. Ισοδύναμα, ένας χρήστης παίρνει όλα τα μηνύματα, που τα στέλνουν οι χρήστες που τους ακολουθεί, τους οποίους θα ονομάζουμε **φίλους-προς (following)** αυτόν. Με άλλα λόγια, ο A είναι φίλος-με τον B, δηλαδή, ο B ακολουθεί τον A, αν και μόνον αν ο B είναι φίλος-προς τον A, δηλαδή, ο

A ακολουθείται από τον B.

Βέβαια, οι δυο αυτές σχέσεις φιλίας δεν είναι υποχρεωτικά **αμοιβαίες** (λέμε, δεν ανταποδίδονται υποχρεωτικά). Δηλαδή, μπορεί ο A να είναι φίλος-με τον B (που σημαίνει ότι ο B ακολουθεί τον A), αλλά ο B να μην είναι φίλος-με τον A (ο A να μην ακολουθεί τον B). Και παρόμοια, μπορεί ο A να είναι φίλος-προς τον B (που σημαίνει ότι ο A ακολουθεί τον B), αλλά ο B να μην είναι φίλος-προς τον A (ο B να μην ακολουθεί τον A).

Τα μηνύματα του Twitter (tweets) μπορούν να έχουν (αν το επιθυμούν οι αποστολείς τους) κάποια θεματική ετικέτα, η οποία συμβολίζεται τοποθετώντας σαν πρώτο στοιχείο το σύμβολο “#” (hashtag). Οι ετικέτες αυτές βοηθούν στον ορισμό θεμάτων για αναζήτηση. Το μήκος μιας θεματικής ετικέτας δεν μπορεί να ξεπερνά τους 16 χαρακτήρες.

2.1 Κατηγορίες μηνυμάτων Twitter

Τα tweets διακρίνονται σε 5 κύριες κατηγορίες: α) update tweets, β) replies, γ) mentions, δ) direct messages και ε) retweets. Παρακάτω αναφέρουμε λίγα πράγματα για κάθε μία από τις κατηγορίες.

2.1.1 Update tweets

Κάθε φορά που ένας χρήστης του Twitter στέλνει ένα μήνυμα, αυτό ονομάζεται update και το λαμβάνουν όλοι οι χρήστες που είναι followers του (φίλοι-με αυτόν). Με αυτόν τον τρόπο μπορούν να ενημερώνονται οι χρήστες που ανήκουν στη λίστα των followers για την τρέχουσα κατάσταση του αποστολέα των tweets αυτών, τα οποία μοιάζουν με προσωπικές ανακοινώσεις.

2.1.2 Replies και mentions tweets (@reply, @mention)

Ένα μήνυμα Twitter, που το στέλνει ένας χρήστης σε απάντηση ή με αναφορά σε ή απευθυνόμενος προς κάποιο συγκεκριμένο άλλο χρήστη (το σύμβολο της αναφοράς είναι το “@”) ονομάζεται **αναφερόμενο** μήνυμα (**referral** ή **mention**). Προφανώς, το αναφερόμενο μήνυμα από τον χρήστη A προς τον χρήστη B είναι δημόσιο και το παίρνουν ο B καθώς και όλοι οι χρήστες που είναι φίλοι-με τον A, αλλά δε τον παίρνουν οι φίλοι-με τον B (εκτός από αυτούς που είναι κοινόι φίλοι των A,B).

2.1.3 Direct messages

Ένα άμεσο μήνυμα (**direct message**) είναι ένα ιδιωτικό μήνυμα (ορατό μόνο από τον αποστολέα και τον παραλήπτη του), που αποστέλλεται από έναν χρήστη του Twitter σε κάποιον άλλον, μόνο εφόσον αυτός είναι φίλος-προς τον πρώτο, δηλαδή ανήκει στη λίστα “following” του πρώτου.

2.1.4 ReTweets (RT)

Ένα μήνυμα Twitter, το οποίο λαμβάνει ένας χρήστης από κάποιον φίλο-προς αυτόν και στη συνέχεια το προωθεί, δηλαδή το ξαναστέλνει σ’ όλους τους φίλους-με αυτόν, ονομάζεται **επανατροφοδοτούμενο** μήνυμα (**retweet**). Προφανώς, το επανατροφοδοτούμενο μήνυμα από τον χρήστη A, το οποίο είχε αρχικά σταλεί από τον χρήστη B (ο A θα πρέπει υποχρεωτικά να είναι φίλος-με τον B, για να λάβει το μήνυμα του B), το παίρνουν κι όλοι οι φίλοι-με τον A, δηλαδή, ενδεχομένως να μην το πάρει ο B, αν αυτός δεν είναι φίλος-με τον A. Αν κάποιος επανατροφοδοτήσει ένα tweet που αρχικά είχε σταλεί από τον χρήστη @Name, τότε στην αρχή του κειμένου του tweet αυτού θα προστεθεί το “RT @Name”.

Αξίζει να αναφερθεί ότι τα επαναπροωθούμενα tweets μπορούν με τη σειρά τους να επαναπροωθηθούν, οπότε είναι δυνατό σε ένα tweet να δούμε πολλαπλές διαδοχικές επανατροφοδοτήσεις RTs. Όταν συνήθως ζητά κάποιος βοήθεια για κάτι ή ενημερώνει για κάτι κρίσιμο τους “followers” του, τότε συνηθίζεται η πολλαπλή επανατροφοδότηση, οπότε στο σώμα του κειμένου είναι ορατά και πολλαπλά “RT @username”.

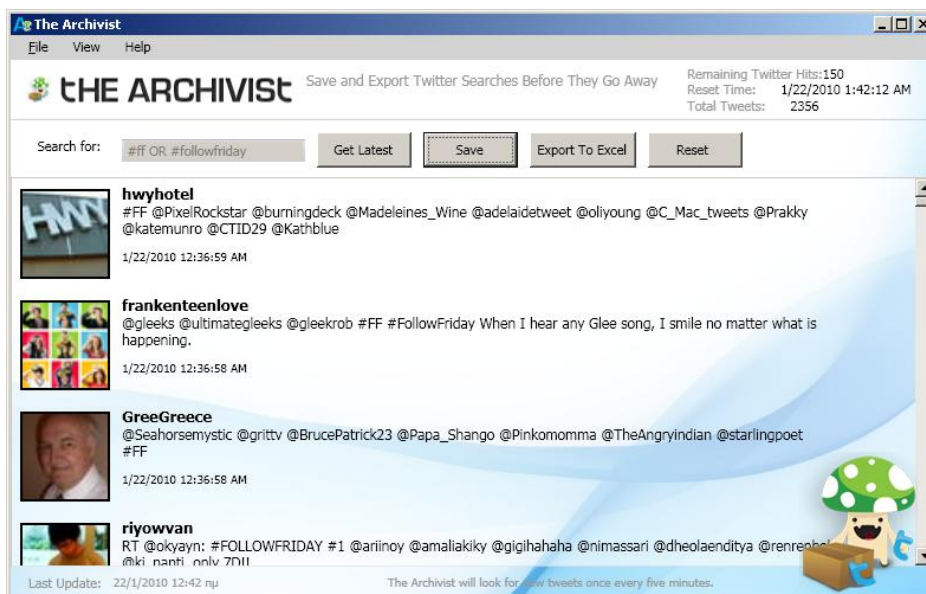
2.2 #FollowFriday ή #FF

Στο Twitter έχει καθιερωθεί κάθε Παρασκευή οι χρήστες που επιθυμούν να προβάλλουν ή να διαφημίσουν ή να συστήσουν κάποιους άλλους χρήστες να αποστέλλουν tweets με θεματική ετικέτα το “#FollowFriday” ή “#FF” στο κείμενο του μηνύματός τους. Έτσι, προτείνουν εμμέσως να προσθέσουν οι φίλοι-με αυτούς τους προβαλλόμενους χρήστες στη λίστα αυτών που ακολουθούν.

3 Συλλογή δεδομένων FF του Twitter

3.1 Χρήση του εργαλείου Archivist

Προκειμένου να συγκεντρώσουμε τα μηνύματα κάτω από τη θεματική ετικέτα #FollowFriday (#FF), που στέλνονται κάθε Παρασκευή, χρησιμοποιούμε την εφαρμογή Archivist (<http://flotzam.com/archivist/>), για αναζητήσεις των “#FF” και “#FollowFriday” (Σχήμα 1). Για να καλύψουμε τις ωρολογιακές ζώνες των χρηστών που κατοικούν σε ολόκληρο τον πλανήτη, οι αναζητήσεις αυτές πραγματοποιούνται από τις 14.00μμ κάθε Πέμπτης ως τις 14.00μμ του Σαββάτου κάθε φορά.



Σχήμα 1: Η εφαρμογή Archivist (<http://flotzam.com/archivist/>).

Τα tweets που συγκεντρώνουμε αποθηκεύονται σε αρχεία της μορφής XML, που τα ονομάζουμε ως “FF_ddMonYY.xml”, όπου “dd” = η ημερομηνία, “Mon” = ο μήνας και “YY” = το έτος. Για παράδειγμα, αποθηκεύσαμε τα δεδομένα της Παρασκευής 20 Νοεμβρίου 2009 στο αρχείο “FF_20Nov09.xml”.

Η εφαρμογή Archivist παρέχει μια απλή διεπαφή, μέσω της οποίας μπορούμε να εκτελέσουμε αναζητήσεις στο Twitter. Η εφαρμογή αξιοποιεί τις δυνατότητες που παρέχει το API του Twitter στους χρήστες του, μία από τις οποίες είναι και η αναζήτηση. Ωστόσο, οι παρεχόμενες δυνατότητες του API του Twitter συνοδεύονται από περιορισμούς, που κάποιοι υφίστανται για λόγους ασφαλείας, ενώ άλλοι για οικονομικούς λόγους. Το Twitter αλληλεπιδρά με τους χρήστες

του μέσω του API που τους παρέχει, κάτι που αποτελεί τον τρόπο με τον οποίο το Twitter διαθέτει τα δεδομένα στους χρήστες.

Οι βασικοί περιορισμοί του API του Twitter είναι οι παρακάτω:

1. Η επικοινωνία με το API γίνεται μέσω του HTTP πρωτοκόλλου και μέσω web clients ή εφαρμογών που το υποστηρίζουν.
2. Όλες οι κλήσεις πρέπει να είναι συμβατές με τα πρότυπα κωδικοποίησης URL.
3. Μια αναζήτηση επιστρέφει το πολύ 1500 αποτελέσματα ανά κλήση.
4. Υπάρχει χρονικός περιορισμός όσον αφορά το πόσο πίσω στο χρόνο θέλουμε να ανατρέξουμε κάνοντας μια αναζήτηση. Μέχρι αυτή τη στιγμή, επιτρέπεται η αναδρομή μέχρι το πολύ ένα μήνα πριν (κάτι τέτοιο δεν είναι πάντα εφικτό, δεδομένου ότι όσο αυξάνει το πλήθος των διακινούμενων tweets, τόσο περιορίζεται το χρονικό όριο αναδρομής).
5. Κάθε χρήστης δεν μπορεί να ξεπεράσει τις 150 κλήσεις ανά ώρα.

Στο Twitter, κλήση θεωρείται κάθε εντολή GET του HTTP, προκειμένου ο χρήστης να εκτελέσει κάποια εργασία. Έτσι, για κάθε χρήστη υπολογίζεται ανά ώρα το πλήθος των εντολών GET που εκτελεί, πλην των εντολών που αφορούν την πρόσβαση του χρήστη στο λογαριασμό του και την ενημέρωσή του σχετικά με τα διαθέσιμα περιθώρια κλήσεων. Η καταγραφή των κλήσεων του χρήστη γίνεται με βάση την IP του χρήστη, οπότε κατά το login του χρήστη, γίνεται καταγραφή και του IP του.

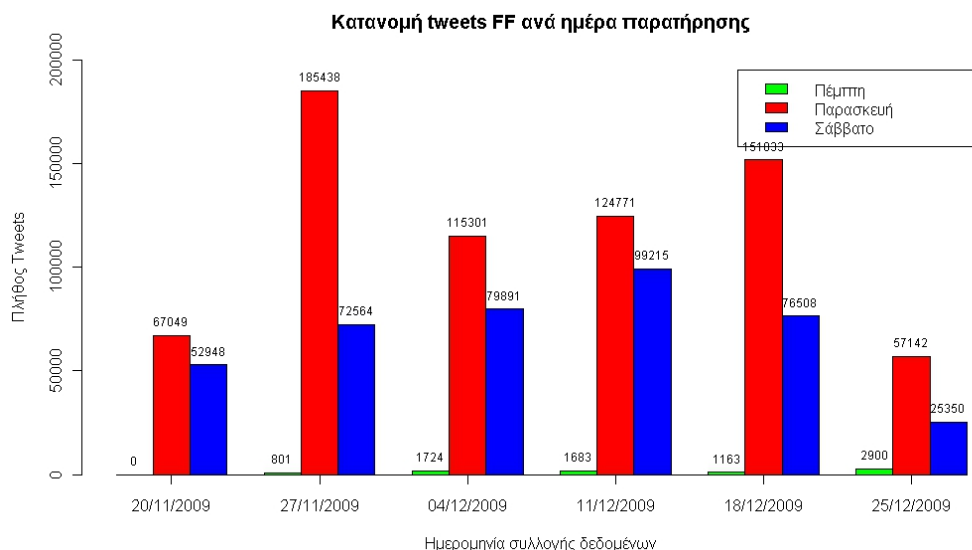
Η εφαρμογή Archivist είναι κατασκευασμένη με τη γλώσσα Visual Basic .NET και αξιοποιεί στο έπακρο τις δυνατότητες που παρέχει το API του Twitter, προκειμένου να συγκεντρώνει τα tweets που ικανοποιούν τα κριτήρια της αναζήτησής μας. Στο πλαίσιο του κειμένου, όπου εισάγουμε τα κριτήρια αναζήτησης, μπορούμε να εισάγουμε και σύνθετες εκφράσεις (για παράδειγμα, το λογικό τελεστή της διάζευξης OR). Στην σελίδα <http://search.twitter.com/advanced> υπάρχουν τα κριτήρια που μπορούμε να δώσουμε σαν παραμέτρους σε μια αναζήτηση για να εκτελέσουμε μια προχωρημένη αναζήτηση.

Η συλλογή των tweets με ετικέτες #FF κάθε Παρασκευής, από τις 20.11.2009 ως τις 25.12.2009, οδήγησε στη δημιουργία 6 μεγάλων αρχείων xml. Δεδομένων των περιορισμών του API του Twitter, ένας μικρός αριθμός από tweets με #FF χάνεται, αφού σε κάθε κλήση αναζήτησης, μπορούμε να εξάγουμε το πολύ έως 1500 tweets. Αυτό δεν επηρεάζει τη μελέτη μας, αφού το Archivist

αναζητεί κάθε 5 λεπτά νέα tweets. Παρακάτω δίνουμε κάποια συγκεντρωτικά στοιχεία, για τα συγκεντρωθέντα tweets με #FF των 6 πρώτων εβδομάδων.

	Παρασκευή1		Παρασκευή2		Παρασκευή3		Παρασκευή4		Παρασκευή5		Παρασκευή6	
Ημερομηνία	20/11/2009		27/11/2009		4/12/2009		11/12/2009		18/12/2009		25/12/2009	
Πλήθος Tweets	119.998		258.805		197.333		225.670		229.505		85.393	
Πέμπτη	0	0%	801	0%	1.724	1%	1.683	1%	1.163	1%	2.900	3%
Παρασκευή	67.049	56%	185.438	72%	115.301	58%	124.771	55%	151.833	66%	57.142	67%
Σάββατο	52.948	44%	72.564	28%	79.891	40%	99.215	44%	76.508	33%	25.350	30%
Σύνολο	119.998	100%	258.805	100%	197.333	100%	225.670	100%	229.505	100%	85.393	100%

Πίνακας 1: Συγκεντρωτικά στοιχεία tweets με #FF για τις 6 πρώτες Παρασκευές, από 20.11.2009 ως 25.12.2009 και κατανομές τους στις μέρες Πέμπτη ως Σάββατο.



Σχήμα 2: Κατανομή πλήθους tweets με #FF ανά ημέρα παρατήρησης, για τις 6 πρώτες εβδομάδες.

4 Κατασκευή δικτύων FF

4.1 Κανόνες σχηματισμού τόξων

Για την κατασκευή των δικτύων FF κάθε Παρασκευής, χρησιμοποιούμε τους εξής κύριους κανόνες για το σχηματισμό των τόξων στον αντίστοιχο κατευθυνόμενο γράφο μεταξύ των χρηστών του Twitter που εμφανίζονται:

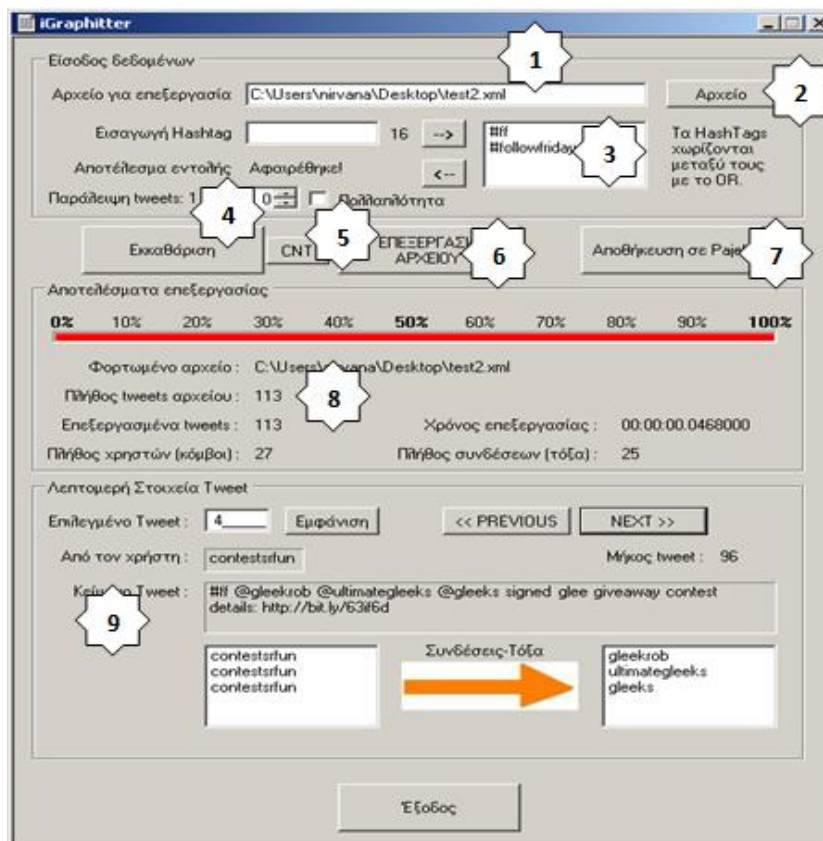
1. Τα tweets FF με μορφή αναφοράς “@”:
 “@A : @B @C @D ... #FF @L @M @N ...”, σχηματίζουν τα τόξα:
 $A \rightarrow B, A \rightarrow C, A \rightarrow D, A \rightarrow \dots$ και
 $A \rightarrow L, A \rightarrow M, A \rightarrow N, A \rightarrow \dots$
2. Τα tweets FF με μορφή επανατροφοδότησης “RT”:
 “@ A : RT @B RT @C RT @D RT ... RT @K #FF @L @M @N ...”,
 σχηματίζουν τα τόξα:
 $A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow \dots, \dots \rightarrow K$ και
 $K \rightarrow L, K \rightarrow M, K \rightarrow N, K \rightarrow \dots$

Τα A, B, C, D, L, M, N ... είναι οι tweeples, δηλαδή οι Twitter χρήστες, ενώ το #FF είναι η θεματική ετικέτα του tweet.

Για tweets που είναι μεικτής μορφής, η εξαγωγή των τόξων γίνεται με αναγωγή τους στους παραπάνω κανόνες.

4.2 Χρήση του λογισμικού iGraphitter

Για τους σκοπούς της εργασίας κατασκευάσαμε ειδικό λογισμικό, που επεξεργάζεται τα αρχεία μορφής XML, που συγκεντρώνουμε με το Archivist, έτσι ώστε να εξάγουμε τους κόμβους και τα τόξα του δικτύου FF. Το λογισμικό είναι κατασκευασμένο με τη γλώσσα προγραμματισμού Visual Basic .NET. Δέχεται ως είσοδο ένα αρχείο μορφής XML και ως έξοδο παράγει ένα αρχείο μορφής Rajek, που αντιστοιχεί στο σχετικό δίκτυο FF.



Σχήμα 3: Το λογισμικό iGraphitter.

Αρχικά πρέπει να επιλέξουμε και να εισάγουμε το αρχείο XML. Αυτό μπορεί να γίνει είτε εισάγοντας τη διαδρομή και το όνομα του αρχείου απευθείας στο πλαίσιο κειμένου “Αρχείο προς επεξεργασία” (επισήμανση 1) είτε με το πλήκτρο “Αρχείο” (επισήμανση 2) και επιλέγοντας το αρχείο με χρήση του κλασικού πλαισίου διαλόγου των Microsoft Windows.

Στη λίστα ακριβώς από κάτω (επισήμανση 3), φαίνονται οι θεματικές ετικέτες που έχουμε επιλέξει να χρησιμοποιήσουμε ώστε να ξεχωρίσουμε τα tweets που επιθυμούμε να επεξεργαστούμε. Προφανώς, έχουμε επιλέξει να επεξεργαστούμε τα tweets που με θεματικές ετικέτες το #FF και το #FollowFriday.

Μπορούμε να μειώσουμε τον συνολικό αριθμό των tweets που θα επεξεργαστούμε, εκτελώντας μια δειγματοληψία. Αυτό γίνεται αν χρησιμοποιήσουμε την επιλογή “Παράλειψη tweets: 1 στα” (επισήμανση 4). Έτσι μπορούμε να επιλέξουμε ανά πόσα tweets που θα διαβάζουμε από το XML αρχείο, θα παραλείψουμε ένα tweet. Ο λόγος που θα θέλαμε να κάνουμε κάτι τέτοιο, είναι ότι συχνά

προκύπτουν δίκτυα με πάρα πολλούς κόμβους, οπότε τα εργαλεία R και ERGM είναι αδύνατο στη συνέχεια να τα επεξεργαστούν με επιτυχία. Μάλιστα έχουμε παρατηρήσει ότι ειδικά το πακέτο ERGM μπορεί να επεξεργαστεί δίκτυα με το πολύ 200.000 κόμβους (+/- 500 κόμβους). Αν επιλέξουμε να αφήσουμε την επιλογή 0 (μηδέν), τότε δε θα γίνει δειγματοληψία.

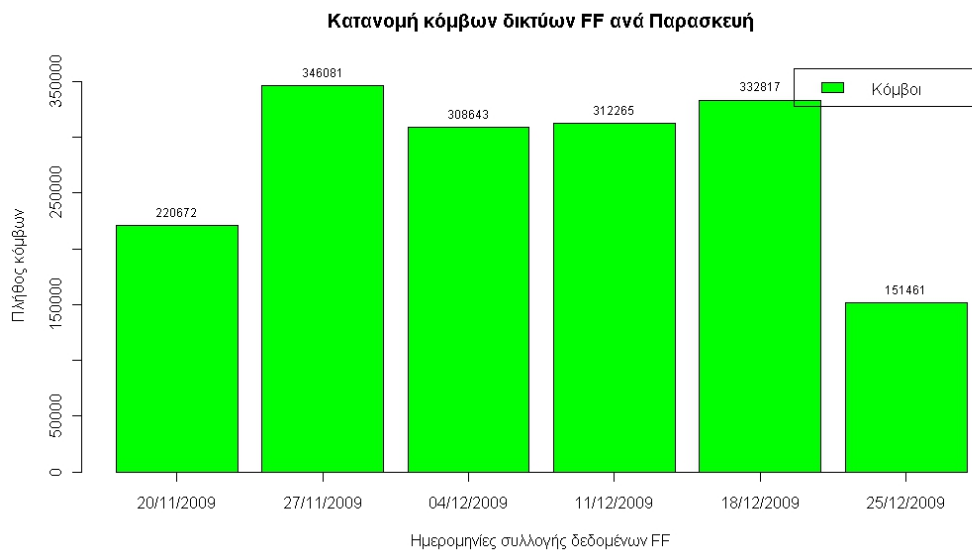
Αν επιθυμούμε το παραγόμενο δίκτυο να θέτει βάρη στα τόξα (τιμή μεγαλύτερη του 1), τότε μπορούμε να τσεκάρουμε την επιλογή “Πολλαπλότητα” (επισήμανση 5). Στην εργασία μας, λαμβάνουμε υπόψη μας μόνο μια φορά το κάθε τόξο, ενώ αν επαναλαμβάνεται τότε απλώς το αγνοούμε.

Με το πλήκτρο “Επεξεργασία” (επισήμανση 6) εκτελούμε την επεξεργασία του επιλεγμένου αρχείου XML και παράγουμε το σχετικό δίκτυο FF. Ο χρόνος επεξεργασίας του XML αρχείου εξαρτάται από το πλήθος των tweets που περιέχονται σε αυτό. Κατά τη διάρκεια της επεξεργασίας μια μπάρα προόδου μας ενημερώνει για το ποσοστό ολοκλήρωσης της εργασίας. Όταν τελειώσει η επεξεργασία, τότε εμφανίζεται σχετικό μήνυμα, ενώ ενεργοποιούνται τα πλαίσια 8 και 9 (επισημάνσεις 8 και 9 αντίστοιχα) που μας παρέχουν χρήσιμα στοιχεία όσον αφορά τα αποτελέσματα της επεξεργασίας, καθώς και επιπλέον λεπτομέρειες. Μόλις ολοκληρωθεί η επεξεργασία ενεργοποιείται το πλήκτρο “Αποθήκευση σε RAJEK”, δίνοντάς μας τη δυνατότητα να αποθηκεύσουμε το δίκτυο που παράγαμε σε αρχείο μορφής Rajek.

Παρακάτω φαίνονται συγκεντρωτικά στοιχεία σχετικά με το πλήθος των tweets και των κόμβων κάθε δικτύου FF που συγκεντρώσαμε από τις 20.11.2009 ως τις 25.12.2009.

Ημερομηνία	20/11/2009	27/11/2009	04/12/2009	11/12/2009	18/12/2009	25/12/2009
Tweets	119.998	258.805	197.333	225.670	229.505	85.393
Κόμβοι	220.672	346.081	308.643	312.265	332.817	151.461

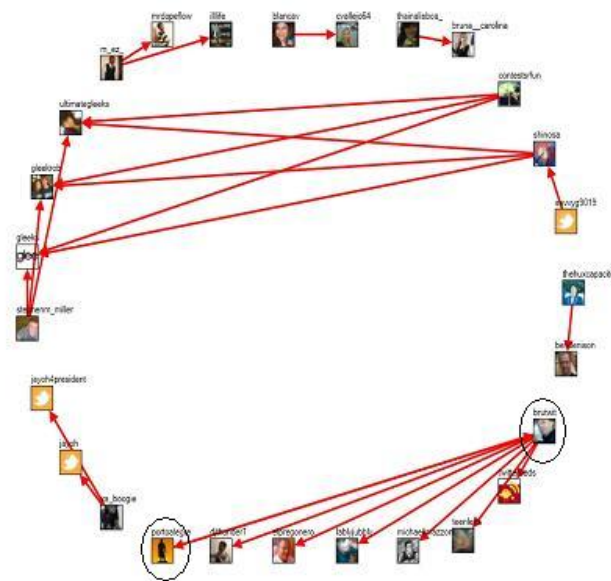
Πίνακας 2: Κατανομή κόμβων δικτύων FF ανά Παρασκευή.



Σχήμα 4: Κατανομή κόμβων δικτύων FF ανά Παρασκευή.

4.3 Αναπαράσταση δικτύου FF με το NodeXL

Το NodeXL είναι ένα πρόσθετο του Microsoft Office, που μας παρέχει τη δυνατότητα γραφικής αναπαράστασης ενός δικτύου, χρησιμοποιώντας ακόμα και τις φωτογραφίες των tweeples. Επιπλέον, παρέχει τη δυνατότητα εισαγωγής των δεδομένων ενός δικτύου από αρχείο μορφής Pajek. Το NodeXL είναι διαθέσιμο μέσω της ιστοσελίδας <http://nodexl.codeplex.com/>. Παρακάτω φαίνεται η αναπαράσταση ενός μικρού δικτύου με το NodeXL.



Σχήμα 5: Κυκλική γραφική αναπαράσταση ενός δικτύου με 27 κόμβους με το NodeXL. Στο δίκτυο υπάρχει μία αμοιβαία σχέση μεταξύ 2 κόμβων.

5 Στατιστική ανάλυση δικτύων FF του Twitter

Στην εργασία μας χρησιμοποιούμε το λογισμικό R και το ERGM πακέτο του, για να διεξάγουμε στατιστική ανάλυση των δικτύων FF που έχουμε συγκεντρώσει για κάθε Παρασκευή. Για κάθε δίκτυο FF υπολογίζουμε τις κατανομές των βαθμών εισόδου και εξόδου των κόμβων του, καθώς και το δείκτη αμοιβαιότητας στην προβολή με το FF.

5.1 Βαθμοί εισόδου και εξόδου κόμβων δικτύου

Ο βαθμός εισόδου (in-degree) ενός κόμβου, είναι ο αριθμός των κόμβων, που είναι γειτονικοί προς αυτόν, δηλαδή είναι ο αριθμός των τόξων που κατευθύνονται και καταλήγουν σε αυτόν, ξεκινώντας από κόμβους, με τους οποίους είναι συνδεδεμένος.

Ο βαθμός εξόδου (out-degree) ενός κόμβου, είναι ο αριθμός των κόμβων, που είναι γειτονικοί από αυτόν τον κόμβο, δηλαδή είναι ο αριθμός των τόξων που ξεκινούν από αυτόν και κατευθύνονται και καταλήγουν προς όλους τους κόμ-

βους, με τους οποίους αυτός ο κόμβος είναι συνδεδεμένος.

Οι βαθμοί εξόδου συνιστούν μέτρα επεκτασιμότητας και οι βαθμοί εισόδου μέτρα της δεκτικότητας ή της δημοτικότητας. Για παράδειγμα, ως προς την σχέση της φιλίας, ένας κόμβος με υψηλό βαθμό εξόδου είναι κάποιο άτομο, που υποδεικνύει ότι έχει πολλούς φίλους. Ένας κόμβος με χαμηλό βαθμό εξόδου υποδεικνύει λιγότερους φίλους. Ένας κόμβος με υψηλό βαθμό εισόδου είναι κάποιο άτομο, που πολλοί άλλοι τον υποδεικνύουν ως φίλο, και ένας κόμβος με χαμηλό βαθμό εισόδου επιλέγεται σαν φίλος από λίγα άτομα.

5.2 Αμοιβαιότητα σχέσης

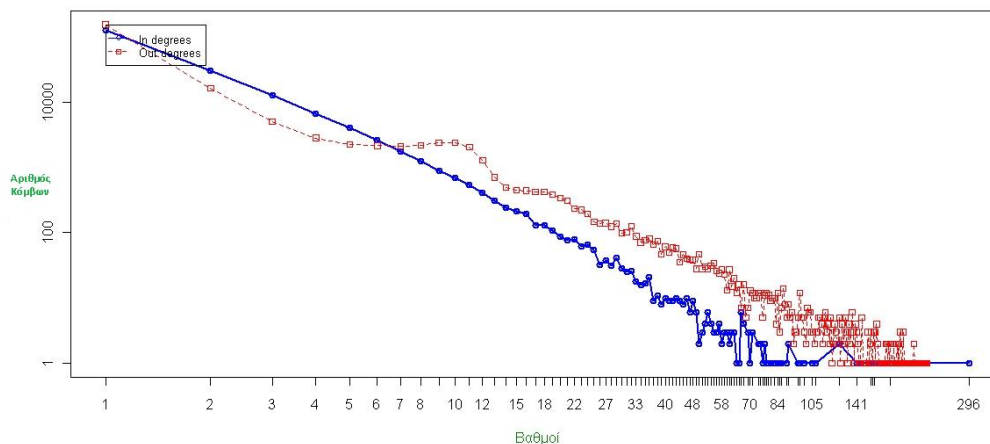
Ως αμοιβαίες αναφέρονται οι σχέσεις που ανταποδίδονται μεταξύ κάποιων κόμβων. Αν για παράδειγμα, ένας κόμβος συνδέεται μέσω μιας σχέσης με κάποιον άλλον κόμβο, τότε για να θεωρηθεί ως αμοιβαία η σχέση θα πρέπει ο δεύτερος κόμβος να ανταποδίδει, δηλαδή να συνδέεται μέσω ίδιας σχέσης με τον πρώτο κόμβο.

Στην εργασία μας εξετάζουμε κατά πόσο η σχέση προβολής με το FF μεταξύ των tweeples στα δίκτυα FF είναι αμοιβαία, δηλαδή κατά πόσο όταν ένας tweeple, σε ένα δίκτυο FF προβάλλει έναν άλλον tweeple, τότε ο πρώτος προβάλλεται ταυτόχρονα και από τον δεύτερο.

5.3 Αποτελέσματα για τις 6 πρώτες Παρασκευές

Στο παρακάτω σχήμα φαίνονται οι κατανομές των βαθμών εισόδου και εξόδου των κόμβων για το δίκτυο FF της Παρασκευής 20 Νοεμβρίου 2009.

Διάγραμμα κατανομών βαθμών εισόδου και εξόδου δικτύου FF της 20.11.2009



Σχήμα 6: Κατανομές βαθμών εισόδου και εξόδου του δικτύου FF της Παρασκευής 20.11.2009.

Στον παρακάτω πίνακα φαίνεται ο αριθμός των αμοιβαίων συνδέσεων για τα δίκτυα FF που συγκεντρώσαμε για τις Παρασκευές από 20 Νοεμβρίου 2009 ως και την 25 Δεκεμβρίου 2009. Οι μετρήσεις έγιναν με χρήση του πακέτου SNA.

Ημερομηνία	20/11/2009	27/11/2009	4/12/2009	11/12/2009	18/12/2009	25/12/2009
Αμοιβαία Τόξα	34.104	19.841	20.108	20.256	18.902	24.464

Πίνακας 3: Πλήθος αμοιβαίων συνδέσεων (τόξων) ανά Παρασκευή.

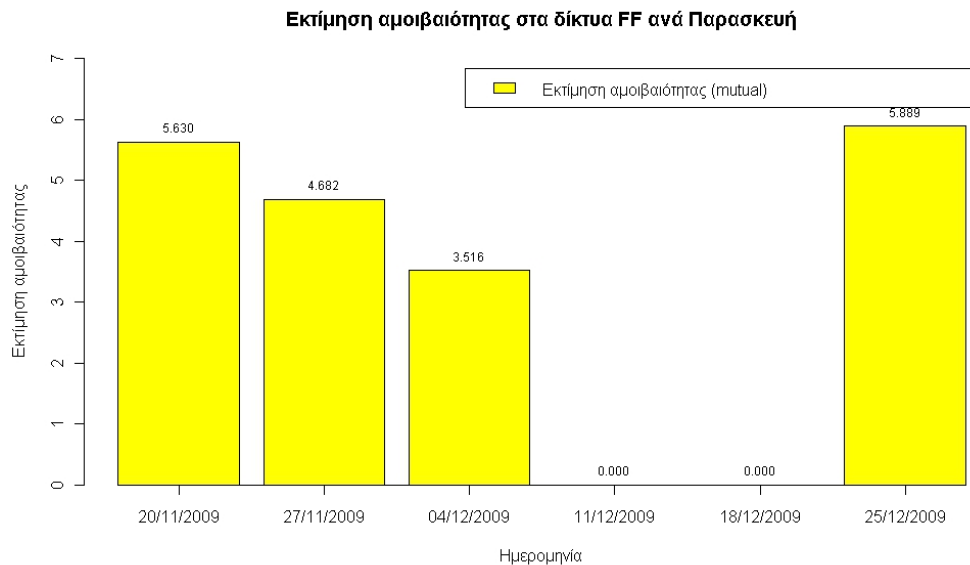


Σχήμα 7: Πλήθος αμοιβαίων συνδέσεων (τόξων) ανά Παρασκευή.

Τέλος, στον παρακάτω πίνακα, συνοψίζουμε τα αποτελέσματα της ERGM ανάλυσης, για την εκτίμηση της αμοιβαιότητας στις προβολές με το FF στα δίκτυα που συγκεντρώσαμε από 20 Νοεμβρίου 2009 ως και 25 Δεκεμβρίου 2009.

Ημερομηνία	20/11/2009	27/11/2009	4/12/2009	11/12/2009	18/12/2009	25/12/2009
Εκτίμηση αμοιβαιότητας	5,630	4,682	3,516	0,000	0,000	5,889

Πίνακας 4: Εκτίμηση αμοιβαιότητας στα δίκτυα FF ανά Παρασκευή.



Σχήμα 8: Εκτίμηση αμοιβαιότητας στα δίκτυα FF ανά Παρασκευή.

6 Μελλοντική εργασία - Προοπτικές

Βασιστήκαμε στην υπόθεση ότι η χρονική εξέλιξη της εκτίμησης της αμοιβαιότητας στην προβολή με το FF, ίσως δείξει κάποιου είδους αυτό-οργανούμενης διεργασίας μεταξύ των μελών των κοινοτήτων των tweeples, όπως την ανάπτυξη κάποιου είδους αμοιβαίας ικανοποίησης ή μιας ώριμης κοινωνικότητας μεταξύ τους. Η υπόθεση αυτή θα εξεταστεί παίρνοντας κατάλληλα δείγματα των δικτύων FF που θα συγκεντρώσουμε ως τον Ιούνιο του 2010, ώστε να διεξάγουμε μια αξιόπιστη ανάλυση μέσω του πακέτου ERGM. Τέλος, θα γίνει προσπάθεια οπτικοποίησης κάποιων αρκετά μικρότερων τμημάτων των δικτύων (κοινότητες tweeples ή components), καθώς και παρουσίασης της χρονικής τους εξέλιξης ανά Παρασκευή.