



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**  
**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**  
**ΚΑΙ ΔΙΚΤΥΩΝ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Αξιολόγηση της σημαντικότητας των μπλόγκερ**

**Βασίλειος Σ. Λεπενιώτης**

**Επιβλέποντες: Παναγιώτης Μποζάνης, Επίκουρος Καθηγητής Π.Θ.**  
**Δημήτριος Κατσαρός, Λέκτορας Π.Θ.**

**ΒΟΛΟΣ, ΣΕΠΤΕΜΒΡΙΟΣ 2010**



# ΠΕΡΙΕΧΟΜΕΝΑ

ΛΙΣΤΑ ΕΙΚΟΝΩΝ .....	3
ΛΙΣΤΑ ΠΙΝΑΚΩΝ .....	3
ΠΕΡΙΛΗΨΗ .....	4
EXECUTIVE SUMMARY .....	6
1 <sup>ο</sup> ΚΕΦΑΛΑΙΟ - Εισαγωγή .....	7
1.1 Blog, Bloggers, Μπλογκόσφαιρα και Εικονική κοινωνία .....	7
1.2 Influential Bloggers και ανάγκη αναγνώρισης αυτών .....	12
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ .....	18
2 <sup>ο</sup> ΚΕΦΑΛΑΙΟ – Υπολογισμός του Σκορ Επιρροής ενός Blogger .....	20
2.1 Προσεγγίσεις Εύρεσης των Influential Bloggers .....	20
2.2 Μοντέλο «Θεματικού Τομέα Ενδιαφέροντος» .....	25
2.2.1 Πειραματική Αξιολόγηση του Μοντέλου «Θεματικού Τομέα Ενδιαφέροντος» .....	30
2.3 Κατάταξη Blogs εναντίων Κατάταξης Ιστοσελίδων .....	32
2.4 Αξιολόγηση του Μοντέλου I-index .....	33
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ .....	34
3 <sup>ο</sup> ΚΕΦΑΛΑΙΟ – Μετρική ΜΕΙΒΙ .....	36
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ .....	38
4 <sup>ο</sup> ΚΕΦΑΛΑΙΟ: Μετρική ΜΕΙΒΙΧ .....	39
4.1 Υπολογισμός της ΜΕΙΒΙΧ .....	39
4.2 Σύγκριση Μετρικών ΜΕΙΒΙ, ΜΕΙΒΙΧ με εκείνες που Βασίζονται σε Ιδιοδιανύσματα .....	39
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ .....	40
5 <sup>ο</sup> ΚΕΦΑΛΑΙΟ – Πειραματική Αξιολόγηση Μετρικών .....	41
5.1 Κοινωνία TUAW .....	41
5.2 Προσδιορισμός του Πειράματος .....	44
5.3 Πειραματικά Αποτελέσματα .....	45
5.3.1 Πρώτο Μέρος Πειραματικών Αποτελεσμάτων .....	45
5.3.2 Δεύτερο Μέρος Πειραματικών Αποτελεσμάτων .....	49
5.3.3 Τρίτο Μέρος Πειραματικών Αποτελεσμάτων .....	51
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ .....	53
6 <sup>ο</sup> ΚΕΦΑΛΑΙΟ – Πειραματική Αξιολόγηση του Μοντέλου I-index .....	54
6.1 Πειραματική Μελέτη Τριών Ειδών bloggers .....	54
6.2 Κατανομή των «Κορυφαίων» της Digg .....	57
6.2.1 Πειραματικά Αποτελέσματα .....	58
6.3 «Βάρη» του I-index .....	60
6.4 Χρονικά «Σχέδια» των influential bloggers .....	61
6.5 Επιπρόσθετα Πειράματα .....	64
ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ .....	68

## ΛΙΣΤΑ ΕΙΚΟΝΩΝ

Εικόνα 1: Γράφος Επιρροής .....	16
Εικόνα 2: Ιστοσελίδα της TUAW .....	42
Εικόνα 3: Ταξινομήσεις Κορυφαίων Bloggers κατά MEIBI .....	51
Εικόνα 4: Ταξινομήσεις Κορυφαίων Bloggers κατά MEIBIX .....	52
Εικόνα 5: Χρονικά Σχέδια Influential Bloggers.....	63
Εικόνα 6: Κατά ζεύγη συσχετίσεις στατιστικών.....	67
Εικόνα 7: Κατανομή του ποσοστού έκφρασης σχολίων «με διακυμάνσεις» .....	67
Εικόνα 8: Επίπεδη κατανομή του ποσοστού έκφρασης σχολίων .....	68

## ΛΙΣΤΑ ΠΙΝΑΚΩΝ

Πίνακας 1: Αποτελέσματα μέσης τιμής βαθμολογίας.....	31
Πίνακας 2: Μεταβολές Κατάταξης των Κορυφαίων Bloggers σε σχέση με τη μεταβολή του $\beta$ .....	32
Πίνακας 3: Συμβολισμοί που χρησιμοποιούνται στη MEIBI .....	36
Πίνακας 4: Δημοσιεύσεις και Εισερχόμενοι Σύνδεσμοι ανά Έτος.....	43
Πίνακας 5: Χρονική Κατανομή Εισερχόμενων Συνδέσμων .....	44
Πίνακας 6: Κατάταξη Bloggers σύμφωνα με τον activity index .....	45
Πίνακας 7: Κατάταξη Bloggers σύμφωνα με τον h-index.....	45
Πίνακας 8: Κατάταξη Bloggers σύμφωνα με την MEIBI.....	46
Πίνακας 9: Κατάταξη Bloggers σύμφωνα με την MEIBIX.....	48
Πίνακας 10: Αποτελέσματα Μετρικής Spearman's rho.....	49
Πίνακας 11: 10 Κορυφαίοι Ενεργοί Bloggers.....	50
Πίνακας 12: 5 Κορυφαίοι Bloggers .....	55
Πίνακας 13: Αποτελέσματα των Στατιστικών για τα 3 Είδη Bloggers .....	55
Πίνακας 14: Ποσότητες των $S_i$ που υπάρχουν στις 100 κορυφαίες της “Digg” .....	59
Πίνακας 15: Κατανομές του συνόλου των 100 κορυφαίων της “Digg” .....	59
Πίνακας 16: Κατανομές του συνόλου των 535 εγγραφών της TUAW.....	59
Πίνακας 17: Αποτελέσματα των Στατιστικών για τα Blog Posts που δεν εξασκούν επηρεασμό και για εκείνα που εξασκούν .....	60

## ΠΕΡΙΛΗΨΗ

Το παρόν κείμενο αποτελεί την διπλωματική εργασία του φοιτητή Λεπενιώτη Βασιλείου. Η εργασία αυτή εκπονήθηκε στα πλαίσια του Προπτυχιακού Προγράμματος Σπουδών του Τμήματος Μηχανικών Η/Υ Τηλεπικοινωνιών και Δικτύων του Πανεπιστημίου Θεσσαλίας υπό την επίβλεψη των καθηγητών του τμήματος κ. Μποζάνη Παναγιώτη και κ. Κατσαρού Δημητρίου. Παρακάτω, ακολουθεί μία συνοπτική και περιεκτική αναφορά της δομής της ύλης που ακολουθείται.

Στο **1<sup>ο</sup> κεφάλαιο** ορίζονται και αναλύονται οι έννοιες *web 2.0 τεχνολογίες*, “*bloggers*”, “*influential bloggers*”, “*blogs*”, «*μπλογκόσφαιρα*» και *εικονικές «κοινωνίες*». Επίσης, γίνεται μία αναφορά στις μεθόδους αντιμετώπισης του προβλήματος εύρεσης των ατόμων τύπου “*influential*”, δηλαδή στις μεθόδους για τη διεξαγωγή της διαδικασίας κρίσης και αξιολόγησης διάφορων *bloggers*.

Στο **2<sup>ο</sup> κεφάλαιο**, εκφράζεται σε μορφή «ποσότητας» η *επιρροή* που εξασκεί ένα *post* και ένας  *blogger* μέσω της ανάλυσης και του ενδεδειγμένου σχολιασμού των μοντέλων *influence-flow*, *i-index* και *interest vector*. Επιπλέον, διεξάγεται μία πειραματική αξιολόγηση του μοντέλου *interest vector* και γίνεται μνεία στους αλγορίθμους αξιολόγησης διαδικτυακών τόπων.

Στο **3<sup>ο</sup> κεφάλαιο** εισάγεται και ορίζεται μία μετρική αναγνώρισης των “*influential bloggers*”, γνωστή ως *MEIBI*, ενώ στο **4<sup>ο</sup> κεφάλαιο** αναλύεται μία επέκταση αυτής της μετρικής, η επονομαζόμενη *MEIBIX* και παράλληλα γίνεται μία σύγκριση αυτών των δύο μετρικών.

Στο **5<sup>ο</sup> κεφάλαιο** γίνεται μία εκτενής παρουσίαση της *πειραματικής αξιολόγησης και σύγκρισης* των μετρικών *MEIBI* και *MEIBIX* με τη μέθοδο “*influence flow*” και τις μετρικές “*activity index*” και “*h-index*”.

Τέλος, στο **6ο κεφάλαιο** αναφέρονται τα αποτελέσματα πειραμάτων που έχουν βασιστεί σε δεδομένα πραγματικού κόσμου με χρήση του δείκτη *i – index* και γίνεται μία *ταξινόμηση* των *bloggers* σε διάφορες κατηγορίες με βάση τα γνωρίσματα

«ενεργός» και «χαρακτηριζόμενος από εξάσκηση επιρροής». Επιπροσθέτως, οι bloggers κατηγοριοποιούνται με βάση το χρονικό διάστημα στο οποίο ασκούν επιρροή.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** *Ιστολόγια του διαδικτύου (blogs), μέθοδοι και μετρικές αξιολόγησης της επίδρασης που ασκούν οι bloggers και τα posts τους, κατηγορίες των bloggers.*

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** *δημοσιευμένη εγγραφή (post), Blog, Blogger που εξασκεί επιρροή (influential blogger), μπλογκόσφαιρα, μοντέλο, μετρική, μέθοδος, h-index, i-index, influence-flow, activity index, MEIBI, MEIBIX, σχόλια, ενεργοί (active), εισερχόμενοι/εξερχόμενοι σύνδεσμοι (inlinks/outlinks), TUAW, DIGG, long-term (μακροπρόθεσμος), average-term (μεσοπρόθεσμος), transient-term (προσωρινός), burgeoning (ανερχόμενοι), θεματικό πεδίο, παράθυρο χρονικής διάρκειας, πείραμα.*

## **EXECUTIVE SUMMARY**

The present text is the thesis of the student Lepeniotis Vasileios. This work was conducted in the frames of the Undergraduate Program of the Department of Computer and Telecommunication Engineering of University of Thessaly, under the supervision of the professors Mr. Bozanis Panagiotis and Mr. Katsaros Dimitrios. Below is a brief and concise description of the structure of the content that is followed.

**In the first chapter** of this thesis the concepts of *web 2.0 technologies*, “*bloggers*”, “*influential bloggers*”, “*blogs*”, “*blogosphere*” and virtual “*communities*” are defined and analyzed. There is also a reference to the methods of addressing the problem of identifying the bloggers with the characteristic “*influential*”, i.e. the methods that conduct the process of bloggers’ crisis and evaluation.

**In the second chapter**, is quantified the *influence* exerted by a post and a blogger through analysis and detailed comments towards the models *influence-flow*, *i-index* and *interest vector*. In addition, an experimental evaluation of the interest vector model is conducted and a reference to the algorithms of websites’ evaluation is made.

**In the third chapter**, a metric for the identification of “*influential bloggers*” is introduced and defined, known as *MEIBI*, whereas **the fourth chapter** discusses an extension of this metric, called *MEIBIX* and also makes a comparison between these two metrics.

**In the fifth chapter** takes place an extensive presentation of *experimental evaluation and comparison* of the metrics *MEIBI* and *MEIBIX* towards the method “*influence flow*” and the metrics “*activity index*” and “*h-index*”.

Finally, **the sixth chapter** presents the results of experiments based on real world data using the *i-index* and makes a *classification* of bloggers in various categories based on the characteristics of “*active*” and “*influential*”. Additionally, bloggers are categorized based on the period in which they are “*influential*”.

## 1<sup>ο</sup> ΚΕΦΑΛΑΙΟ - Εισαγωγή

Τα ιστολόγια (**blogs**) ανήκουν στη **κατηγορία web 2.0 τεχνολογιών**. Αυτή η κατηγορία επίσης περιλαμβάνει: όλες τις «**εικονικές κοινωνίες**» του Διαδικτύου και υπηρεσίες όπως φόρουμ, “wikis”, κοινόχρηστοι σελιδοδείκτες και διανομή υλικού media. Η web 2.0 παρέχει μια δεύτερη γενιά διαδικτυακών «κοινωνιών» και κινείται προς τη κατεύθυνση διευκόλυνσης της επικοινωνίας και συνεργασίας των χρηστών κάποιου blog. Μέσω της κατηγορίας των web 2.0 τεχνολογιών, τα “online” μέσα μετέτρεψαν τους επισκέπτες ιστολογίου από μαζικούς καταναλωτές σε μαζικούς παραγωγούς πληροφορίας. Οι εικονικές «κοινωνίες» «κτίζονται» από κάποιο σύνολο ιστολογίων και αποτελούν ένα μέσο για τους χρήστες για να εκφράζουν και να ανταλλάσσουν: ιδέες, γνώσεις, απόψεις, εμπειρίες, πληροφορίες προερχόμενες από τα Μ.Μ.Ε.

### 1.1 Blog, Bloggers, Μπλογκόσφαιρα και Εικονική κοινωνία

Ως **Bloggers** ορίζονται τα άτομα με ρόλο τον έναν ή και τους δύο από τους παρακάτω:

- Δημιουργού δημοσιευμένης εγγραφής/εγγραφών κάποιου ιστολογίου.
- Σχολιαστή δημοσιευμένης εγγραφής/εγγραφών κάποιου ιστολογίου.

Επίσης, «**ενεργοί**» χαρακτηρίζονται όσοι έχουν ρόλο σχολιαστή/δημιουργού δημοσιοποιημένης καταχώρισης/καταχωρήσεων σε blog οι οποίοι δημιουργούν ένα καίριο πλήθος δημοσιοποιημένων εγγραφών για συγκεκριμένο χρονικό διάστημα και γενικά εκτελούν σημαντικό πλήθος δραστηριοτήτων για το διάστημα αυτό. Στα ιστολόγια όπου υπάρχουν τουλάχιστον δύο «**ενεργά**» άτομα με ρόλο σχολιαστή/δημιουργού δημοσιευμένης εγγραφής/εγγραφών γίνεται ταξινόμηση αυτών των ατόμων με κριτήριο το βαθμό έως τον οποίον τα διακρίνει το επίθετο «**ενεργά**» για συγκεκριμένο χρονικό διάστημα.



Υπάρχει μια έννοια εικονικού σύμπαντος που περιλαμβάνει όλα τα ιστολόγια, αυτήν είναι η «Μπλογκόσφαιρα» (“Blogosphere”). Η «Μπλογκόσφαιρα» είναι ένας εικονικός κόσμος όπου οι χρήστες (που επισκέπτονται τα blogs αυτού του χώρου) μπορούν να βλέπουν τις εκάστοτε τοποθετημένες γνώμες, συστάσεις και εμπειρίες άλλων χρηστών ώστε να διευκολύνονται στις αγορές, στα ταξίδια τους και στις αποφάσεις τους. Επιπροσθέτως, η “Blogosphere” εμπνέει το **“viral” μάρκετινγκ**. Αυτό το είδος μάρκετινγκ βελτιώνει τον βαθμό έως τον οποίο αναγνωρίζονται τα επιχειρησιακά προϊόντα και οι υπηρεσίες καθώς αυτό το είδος βασίζεται σε “online” τεχνικές και αξιοποιεί τους ίδιους τους χρήστες. Σχετικά με το “viral” μάρκετινγκ [9], για τη μεγιστοποίηση της ολικής επίδρασης στο δίκτυο (των ιστολογίων) επιλέγεται σταθερό πλήθος κόμβων του δικτύου. Μία αποδοτική προσέγγιση για το **“viral” μάρκετινγκ** είναι το να επιλέγεται σε κάθε επανάληψη εκείνος ο κόμβος του γράφου που σχετίζεται με τη μεγαλύτερη δυνατή εξάσκηση επηρεασμού αφού πρώτα μετακινηθούν οι επιλεγόμενοι κόμβοι. Η προσέγγιση που διατυπώθηκε πριν, υπερτερεί έναντι των αλγόριθμων **Pagerank, Hits**<sup>1</sup>. Καίριο χαρακτηριστικό αυτής της προσέγγισης είναι το σημαντικό πλήθος παραπομπών που εμπεριέχει και η σταθερότητά του στα φιλτραρισμένα spam blogs.

Η «μπλογκόσφαιρα» είναι ένα από τα ταχύτερα αναπτυσσόμενα κοινωνικά μέσα το οποίο αναδιαμορφώνει το επιχειρησιακό μοντέλο, παρέχει ανάλυση της επικρατούσας τάσης και πρόβλεψη του μεγέθους των πωλήσεων. Επιπλέον, η “Blogosphere” συνεισφέρει στην αντιτρομοκρατική ενίσχυση και παρέχει πηγές πληροφοριών «βασικού» επιπέδου. Οι εικονικές «κοινωνίες» της «Μπλογκόσφαιρας» δεν περιορίζονται από τη φυσική εγγύτητα και επιτρέπουν την ανάπτυξη μιας νέας μορφής αποδοτικών επικοινωνιών μαζί με τη δημιουργία ειδικών ομάδων ενδιαφέροντος. Η ραγδαία εξάπλωση της «Μπλογκόσφαιρας» δημιούργησε την ανάγκη ανάπτυξης καινοτόμων εργαλείων για την διευκόλυνση των χρηστών της ως προς το να επικοινωνούν, να συμμετέχουν και να εξερευνούν. Σε αυτού του είδους τις «κοινωνίες», η εξάσκηση επιρροής, είναι ένα καίριο θέμα η παρουσία του οποίου «κυβερνά» τη διαδικασία που ακολουθείται για τη διαμόρφωση της σύστασης μιας «κοινωνίας». Για τη καλύτερη κατανόηση του τρόπου «λειτουργίας» της «Μπλογκόσφαιρας» είναι αναγκαίες δύο προϋποθέσεις:

---

<sup>1</sup> Αλγόριθμοι αξιολόγησης της αυθεντίας μίας ιστοσελίδας (WebRanking Algorithms)

- Η παρακολούθηση της ανάπτυξης αυτής στο πέρασμα του χρόνου.
- Η υπολογιστική διαδικασία επισήμανσης εκείνων που εξασκούν επηρεασμό και έχουν ρόλο σχολιαστή/δημιουργού δημοσιευμένης εγγραφής/εγγραφών σε ιστολόγιο / ιστολόγια της “Blogosphere”.

Η ανάπτυξη των εφαρμογών Web 2.0 είναι ραγδαία. Παράλληλα με αυτή την ανάπτυξη, όλο και περισσότεροι άνθρωποι συμμετέχουν στη μία ή την άλλη μορφή των δραστηριοτήτων οι οποίες περιλαμβάνονται σε **εικονικές «κοινωνίες»**. Κατ’ επέκταση με αυτή τη συμμετοχή, η εύρεση των bloggers που προκαλούν επιρροή όχι μόνο επιτρέπει τη καλύτερη κατανόηση των ενδιαφερόντων δραστηριοτήτων που συμβαίνουν σε έναν εικονικό κόσμο, αλλά και παρέχει μοναδικές «ευκαιρίες» για τη βιομηχανία, τις πωλήσεις, και τις διαφημίσεις. Βλέποντας τον φυσικό και τον εικονικό ως δύο κόσμους μεταξύ των οποίων βρίσκονται παράλληλες «γραμμές», για τα μέλη της “Blogosphere” είναι πολύ σημαντικά τα εξής: η διαδικασία αναγνώρισης των Bloggers που εξασκούν επηρεασμό, το ποιοι είναι αυτοί συγκεκριμένα, οι τρόποι μέσω των οποίων ευνοείται ο εντοπισμός τους. Στις μέρες μας, η κυρίαρχη παρουσία και η ευκολία που υπάρχει στη χρήση του ιστού (web) διευκολύνουν την ανάπτυξη ενός εικονικού κόσμου όπου ευνοούνται:

- **Η βελτίωση των διαπροσωπικών σχέσεων.**
- **Η δυνατότητα δημοσιοποίησης διάφορων γεγονότων.**
- **Η διεύρυνση των δραστηριοτήτων αγοραπωλησίας.**

Οι τύποι ιστολογίων που η «Μπλογκόσφαιρα» περιέχει είναι:

➤ *Ατομικά ιστολόγια τα οποία συντηρούνται και διατηρούνται ενημερωμένα από τον/την Blogger που συσχετίζεται με αυτά. Στα ατομικά ιστολόγια αναρτώνται δημοσιοποιημένα σχόλια κάθε είδους θεματολογίας. Οι Bloggers ενός blog αυτού του είδους δεν έχει νόημα να διατάσσονται σε σειρά με βάση την επιρροή που ασκούν. Παραδείγματα αυτής της κατηγορίας των blog είναι:*

- ✓ Το blog με θέμα “David Sifry’s musings” του Blogger «Sifry Alerts» (που συσχετίζεται με: Founder & CEO, Technorati) [4].
- ✓ Το Blog θέματος: «Mitch’s Open Notebook» [5] του Blogger «Ratcliffe».

Για κάθε ιστολόγιο της παρούσας κατηγορίας δεν υπάρχει λόγος εκτέλεσης κάποιας από τις τεχνικές ανίχνευσης ατόμου που εξασκεί επιρροή και έχει ρόλο σχολιαστή/δημιουργού δημοσιευμένης εγγραφής.

- *Κοινωνικά* ιστολόγια στα οποία πολλοί Bloggers μπορούν να ξεκινήσουν συζητήσεις για ένα γεγονός ή προϊόν και να αρχίσουν κάποιο νέο θέμα συζήτησης. Παραδείγματα αυτής της κατηγορίας είναι τα ιστολόγια: “Google’s Official Blog”, “The Unofficial Apple Weblog” [7,8]. Οι σχολιαστές/δημιουργοί δημοσιοποιημένης εγγραφής/εγγραφών στο παρόν είδος ιστολογίων είναι ισότιμοι μεταξύ τους και έτσι σε κάθε blog εφαρμόζεται αλγόριθμος κατάταξης αυτών ως προς τον επηρεασμό που εξασκούν καθώς αντιμετωπίζονται ως μέλη εικονικής «κοινωνίας».

Μέσα από τη “blogosphere” οι ερευνητές αντιλαμβάνονται πως ένα ιστολόγιο που συσχετίζεται με πολλούς σχολιαστές/δημιουργούς δημοσιευμένης καταχώρισης/καταχωρήσεων είναι βέβαιο πως εξασκεί επηρεασμό. Παράλληλα, αυτοί οι ερευνητές εξερευνούν τη “Blogosphere” και όλες τις ομάδες που είναι ορισμένες εσωτερικά αυτής. Σε μια ιστοσελίδα τύπου Blog είναι βέβαιη η αλληλεπίδραση μεταξύ των Bloggers αυτής που εξασκούν επηρεασμό και των υπόλοιπων. Η ανάπτυξη της έρευνας, η οποία γίνεται στα πλαίσια των κοινωνικών δικτύων και έχει ως επίκεντρο τη «μπλογκόσφαιρα» συνεχώς αυξάνεται.

Τα ιστολόγια αποτελούν για τους δημιουργούς ή σχολιαστές δημοσιοποιημένης εγγραφής/εγγραφών ένα μέσο για:

- Να εκφράζουν τις απόψεις τους, τα “θέλω” τους και όσα τους είναι ανεπιθύμητα.
- Να σχολιάζουν.
- Να δίνουν απαντήσεις βασιζόμενοι στις γνώσεις τους σχετικά με τις δημοσιευμένες απόψεις των χρηστών (επισκεπτών) των συγκεκριμένων blogs.
- Να κάνουν συστάσεις όπου παρατηρούν προβληματισμούς ή λανθασμένες πληροφορίες.
- Να δημοσιεύουν ειδήσεις (οι οποίες δεν έχουν τον ίδιο βαθμό εγκυρότητας μεταξύ τους).

- Να διαμορφώνουν ομάδες στη “μπλογκόσφαιρα”, γνωστές ως «ειδικές ομάδες ενδιαφέροντος» και έτσι να επικοινωνούν, να συνεργάζονται, να αντιπαρατίθενται, να διαμοιράζονται ιδέες και προβληματισμούς, να παρέχουν συμβουλές με διαδικασία αλληλεπιδραστική.

Τα ιστολόγια αποτελούν ένα ειδικής μορφής κοινωνικό δίκτυο που περιέχουν “entries” των χρηστών τους με αντίστροφη χρονολογική σειρά. Αυτές οι “entries” μπορεί να είναι δημοσιοποιημένες εγγραφές (posts) είτε σχόλια (comments) τα οποία είναι δημοσιοποιημένες καταχωρήσεις που αναρτώνται μετά από κάποια συγκεκριμένη δημοσιευμένη καταχώρηση και συνδέονται προς αυτή. Ακόμη, στη τυπική τους μορφή συνδυάζουν κείμενο, εικόνες, συνδέσμους προς ιστοσελίδες και άλλα blogs.

Στα blogs που είναι *κοινωνικού τύπου*, επιπροσθέτως, κάθε δημοσιοποιημένη εγγραφή συσχετίζεται με μεταδεδομένα τα οποία είναι: η ημερομηνία και η ώρα ανάρτησής, το όνομα του ατόμου Blogger που τη καταχώρησε, το πλήθος των σχολίων μαζί με τον γενικό «μετασχολιασμό» (post annotation) που δέχεται η δημοσιοποιημένη εγγραφή.

Μέσα σε κάθε ιστολόγιο, κάτω από κάθε είδους εγγραφή που περιέχει, τοποθετούνται τα πιθανά σχόλια από τους χρήστες που τα επισκέπτονται. Ακόμη, ερευνάται το ενδεχόμενο σύνδεσης με ακμή των ιστοσελίδων τύπου Blog που έχουν το ίδιο θέμα ως προς τα περιεχόμενά τους. Επιπλέον σε μία ιστοσελίδα τύπου *Blog κοινωνικού τύπου*, είναι σημαντικά τα ακόλουθα **στατιστικά**:

- το μέσο μήκος των σχολίων που καταχωρούνται.
- ο ρυθμός με τον οποίο αναρτώνται δημοσιευμένες εγγραφές.
- Οι εισερχόμενοι/εξερχόμενοι σύνδεσμοι. Οι **εισερχόμενοι σύνδεσμοι (inlinks)** αφορούν τα υπόλοιπα posts που αναφέρονται στο post ενός blogger. Το σύνολο αυτών των συνδέσμων συλλέγεται μέσω του API [22] της μηχανής αναζήτησης ιστολόγιων Technorati. Δηλαδή, αυτό το API παρέχει πληροφορίες για τους εισερχόμενους συνδέσμους. Οι **εξερχόμενοι σύνδεσμοι (outlinks)** είναι posts ή άρθρα στα οποία ένας blogger έχει αναφερθεί. *Το πλήθος αυτού του είδους των συνδέσμων προς μια δημοσιευμένη εγγραφή ιστολογίου αποτελεί ένδειξη της καινοτομίας μιας δημοσιευμένης καταχώρισης*

*και είναι αντιστρόφως ανάλογη προς τον βαθμό της καινοτομίας και άσκησης επίδρασης προερχόμενης από την εγγραφή αυτή. Το πλήθος των εισερχόμενων συνδέσμων προς δημοσιοποιημένη καταχώριση ιστολογίου μεταβάλλεται ανάλογα προς το βαθμό έως τον οποίο εξασκείται επηρεασμός από την εγγραφή αυτή.*

Η έννοια του **Blogging** αφορά στον τρόπο μέσω του οποίου κάποιος χρήστης του διαδικτύου μπορεί να κάνει τις παραπάνω ενέργειες σε ιστολόγιο. Έτσι, συντηρείται η διαδικασία ανατροφοδότησης της ανταλλαγής απόψεων και γνώσεων μεταξύ των επισκεπτών και των ατόμων που δημοσίευσαν/σχολίασαν καταχώριση στον συγκεκριμένο ιστοχώρο. Οι “bloggers” έχουν τη δυνατότητα μέσα στα blogs να τοποθετούν(αναρτούν) εικόνες είτε αρχεία ήχου είτε ενσωματωμένα βίντεο είτε κείμενα είτε συνδέσμους προς ιστότοπους ή δημοσιευμένες εγγραφές κάποιων άλλων ιστολογίων.

Οι προκλήσεις που έχουν παρατηρηθεί στη «Μπλογκόσφαιρα» είναι οι κάτωθι:

- Το να αντιστοιχεί σχόλιο/σύνολο σχολίων σε καθεμία δημοσίευση κάποιου ιστολογίου.
- Η πιθανή συμπερίληψη κύκλου στον τελικό γράφο.
- Ο συγγραφέας να είναι ένα άτομο.
- Η ύπαρξη λεπτομερούς χρονικής διασποράς.

## **1.2 Influential Bloggers και ανάγκη αναγνώρισης αυτών**

Υπάρχουν διάφορες μέθοδοι για τη διεξαγωγή της διαδικασίας κρίσης και αξιολόγησης των Bloggers. Αυτές οι μέθοδοι αφορούν σε μια κατηγορία σχολιαστών/δημιουργών δημοσιευμένης εγγραφής/εγγραφών σε ιστολόγιο/ιστολόγια που εξασκούν αρκετή επιρροή, τους “**Influential**”. Επιπλέον, αυτές οι μέθοδοι έχουν «λειτουργία» που είναι «κάθετη» στη διαδικασία ελέγχου του εάν οι Bloggers που «εξετάζονται» ως προς τον επηρεασμό που εξασκούν ανήκουν σε ιστοσελίδα τύπου Blog που ασκεί επηρεασμό.

Στην εποχή μας είναι πολλοί οι Bloggers (έστω τύπου “a”) που πριν προβούν σε κάποια αγορά ή πριν αποφασίσουν για κάποιο ζήτημα επηρεάζονται από άλλους Bloggers (έστω τύπου “b”) οι οποίοι είναι Bloggers «συνάδελφοι» εκείνων του τύπου “a”. Οι Bloggers του τύπου “b” χαρακτηρίζονται από τον σημαντικό βαθμό στον οποίο ασκούν επίδραση. Πάντοτε στον επιχειρησιακό και τον κοινωνικό τομέα η ασκούμενη επίδραση έχει καίριο ρόλο.

Η κατηγορία των **“Influential”** [12] περιλαμβάνει άτομα που :

- *Προτείνουν επίκαιρα θέματα και λύσεις στο ιστολόγιο / ιστολόγια τους και κατ’ επέκταση έχουν τη δυνατότητα άσκησης επίδρασης.*
- *Δημοσιεύουν ποιοτικές εγγραφές (που αντανακλούν το βαθμό εξειδίκευσής τους και τη διασημότητα τους) οι οποίες έχουν αρκετούς σχολιαστές.*
- *Η αυθεντία τους έχει υψηλό επίπεδο. Η αυθεντία, αποτελεί την αντανάκλαση:*
  - ✓ *της θέσης τους στο συνολικό διαδίκτυο.*
  - ✓ *των συνδέσμων τους με άλλα άτομα που δημιούργησαν κάποιο Blog.*
- *Οδηγούν στην εκτέλεση δραστηριότητας με περιεχόμενο ανάλογο προς τον θεματικό τομέα όπου ασκούν επίδραση.*

Όσοι προαναφέρθηκαν ως **“Influential”** αληθεύει ότι επηρεάζουν το αγοραστικό κοινό. Οι φορείς που εξαρτώνται από όσους εξασκούν επιρροή επιτυγχάνουν να αποφεύγουν μελλοντικές παγίδες και να προβλέπουν πιθανά κέρδη με το να στηρίζονται απλά σε εκείνους που ασκούν επίδραση

Οι μέθοδοι εύρεσης των bloggers που εξασκούν επηρεασμό αναπτύχθηκαν επειδή η διαδικασία εύρεσης εκείνων που εξασκούν επιρροή και έχουν ρόλο σχολιαστή/δημιουργού δημοσιευμένης καταχώρισης σε ιστολόγιο, είναι κύρια για τα επαγγελματικά συμφέροντα εμπορικών εταιρειών και επιχειρήσεων. Τα άτομα με ρόλο σχολιαστή/δημιουργού δημοσιευμένης καταχώρισης σε ιστολόγιο αξιοποιούνται από τις εταιρείες και τις επιχειρήσεις. Αυτήν η αξιοποίηση σημαίνει πως οι εταιρείες τους τροφοδοτούν με επιπρόσθετες πληροφορίες και συμβουλές για να τους μετατρέψουν στους ανεπίσημους «εκπροσώπους» τους. Μία σημαντική τέτοιου είδους εταιρεία είναι η Macromedia η οποία συγκεντρώνει, κατηγοριοποιεί και ερευνά μόνο τις δημοσιοποιημένες καταχωρήσεις ατόμων που εξασκούν επιρροή.

Επιπλέον, η συστηματική σειρά ενεργειών για την αναγνώριση των Bloggers που εξασκούν επιρροή βοηθά στο να «ανακαλύπτονται» όσες ενδιαφέρουσες δραστηριότητες αναπτύσσονται στη «Μπλογκόσφαιρα» και διευκολύνει τη λειτουργία των βιομηχανιών, τις διαφημίσεις και τον σκοπό τους.

Μιας και οι Bloggers μπορούν να συνδεθούν σε εικονική «κοινωνία» οποτεδήποτε και οπουδήποτε, η διαδικασία που διατυπώθηκε μόλις πριν μπορεί να ωφελήσει:

- Στο να σφυρηλατούνται πολιτικές ατζέντες.
- Στη συζήτηση κοινωνικών θεμάτων.
- Στη δυνατότητα οδήγησης σε πολλές ενδιαφέρουσες εφαρμογές.

Τα Blogs, έχουν προσελκύσει το ενδιαφέρον ενός μεγάλου πλήθους εταιρειών οι οποίες τα χρησιμοποιούν στα πλαίσια της στρατηγικής marketing των προϊόντων τους. Οι “**Influential**” έχουν πολλαπλό ρόλο που οφείλεται στα όσα τοποθετούν στα Blogs τους. Είναι πιθανό επομένως να επηρεάσουν:

- Τις αποφάσεις των επισκεπτών των ιστολογίων.
- Τους δημιουργούς ή σχολιαστές δημοσιευμένης εγγραφής/εγγραφών κάποιου ιστολογίου που είναι ομότιμοι με αυτούς. Γι’ αυτό το σκοπό, όσοι εξασκούν επιρροή και είναι δημιουργοί ή σχολιαστές post κάποιου ιστολογίου χρησιμοποιούνται ως μοχλός από τις εταιρείες που προαναφέρθηκαν. Οι λόγοι που οδηγούν σ’ αυτό, είναι οι ακόλουθοι:

- i. Οι εταιρείες, στη προσπάθεια που κάνουν για προώθηση των προϊόντων τους, αποσκοπούν στη μετάδοση διαφημιστικών μηνυμάτων “από στόμα σε στόμα” και στο να κάνουν αποτελεσματικότερη τη διαδικασία διεξαγωγής του εμπορίου. Για να επιτευχθεί ο παραπάνω στόχος, αξιοποιούν από την εμπορική σκοπιά, τους Bloggers (του συγκεκριμένου τύπου που προαναφέρθηκε) και χρησιμοποιούν τους ίδιους τους πελάτες τους ως μοχλό για να γίνουν οι ανεπίσημοι «εκπρόσωποί» τους. Λόγω της αυθεντίας τους, τα άτομα της κατηγορίας «Influential», θεωρούνται πως επιλύουν προβλήματα των πελατών τους και είναι αξιόπιστα. Αυτήν η αξιοποίηση δίνει σημαντικό νόημα στη διαδικασία αναγνώρισης των Bloggers

που εξασκούν επίδραση. Οι περισσότερες διαφημιστικές εταιρείες, μέσω των ιστολογίων «διαισθάνονται» πως η διεξαγωγή διαφήμισης μέσω κάποιου Blog είναι αρκετά αποδοτική.

- ii. Οι bloggers έχουν κύριο ρόλο στην ανάπτυξη καινοτόμων επιχειρησιακών ευκαιριών (όπως για εμπορικές δοσοληψίες και ταξίδια).
- iii. Οι bloggers κάποιου ιστολογίου έχουν κύριο ρόλο στο να επηρεάσουν τη συμπεριφορά εκείνων που βρίσκονται στη διαδικασία κάποιας ψηφοφορίας.

- Τις αντιδράσεις κυβερνητικών πολιτικών, καθώς οι Bloggers της συγκεκριμένης κατηγορίας έχουν απόψεις οι οποίες εξασκούν επίδραση σε πολιτικές καμπάνιες και εκλογές.

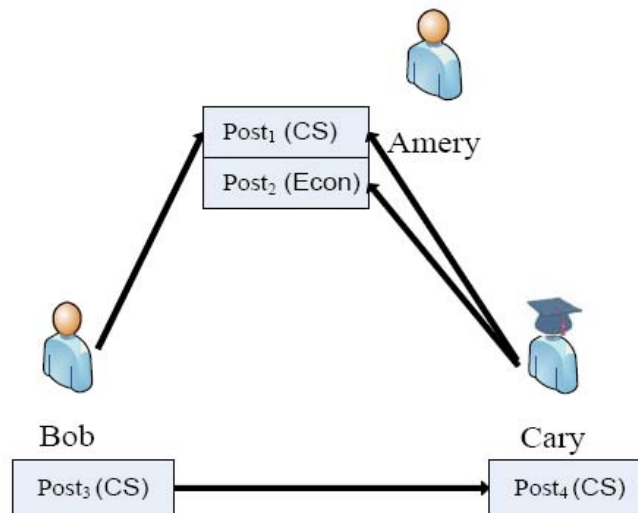
Έτσι, πρόσφατα ο εντοπισμός των κορυφαίων «k» (στο πλήθος), ως προς την επίδραση που ασκούν, είναι ένα συγκεκριμένο πεδίο που αρχίζει να προσελκύει όλο και περισσότερο τα ερευνητικά συμφέροντα σύμφωνα με έρευνες που διεξήχθησαν [1, 2, 3]. Δηλαδή, είναι αναγκαίο να γίνει η διερεύνηση ενός τέτοιου (τουλάχιστον ενός) μηχανισμού εντοπισμού. Επιπλέον, είναι επιθυμητό να δημιουργούνται λίστες (έστω τύπου “L1”) που συνήθως θα περιλαμβάνουν τους κορυφαίους, ως προς το πλήθος των δημοσιευμένων εγγραφών τους, σχολιαστές/δημιουργούς δημοσιοποιημένης εγγραφής/εγγραφών σε ιστολόγιο / ιστολόγια και λίστες(έστω τύπου “L2”) που συνήθως θα περιέχουν τα κορυφαία ιστολόγια ως προς το πλήθος των σχολίων που λαμβάνουν οι δημοσιευμένες εγγραφές τους. Οι λίστες που ορίστηκαν πριν ως επί το πλείστον δε λαμβάνουν υπόψη άτομα και ιστολόγια μη «ενεργά».

Συνυπολογίζοντας τόσο τη συσχέτιση «δημοσιοποιημένης εγγραφής» – «απάντησης προς αυτή» όσο και τις πληροφορίες γενικού συνδέσμου κάποιου blogger, είναι πιθανό να γίνει αξιολόγηση της συνολικής επίδρασης ενός blogger. Οι προαναφερθείσες έρευνες, μετρούν την επιρροή μεταξύ των bloggers με βάση τις σχέσεις «δημοσιεύματος – απάντησης (σχολιασμού)», που μοντελοποιούνται σε ένα γράφο επιρροής (**Εικόνα 1**) . Επίσης, οι γενικές πληροφορίες του συνδέσμου



κάποιου Blogger κάποιου ιστολογίου συνεισφέρουν στον υπολογισμό του βαθμού στον οποίο ασκείται επιρροή. Στον γράφο, η Amery δημοσιεύει την εγγραφή “post1” η οποία σχολιάζεται από τους Bob και Cary. Επίσης, η Amery δημοσιεύει την εγγραφή “post2” η οποία σχολιάζεται από τον Cary. Υποθέτουμε ότι η “post1” εξετάζει ορισμένες ικανότητες προγραμματισμού στην επιστήμη των υπολογιστών, ενώ η “post2” ερευνά την πρόσφατη οικονομική ύφεση και τις πιθανές τάσεις στο επόμενο δίμηνο. Οι τρεις ακόλουθες πληροφορίες είναι σημαντικές όσον αφορά τη συσχέτιση «δημοσιευμένης εγγραφής» – «απάντησης προς αυτή» :

- Οι δύο δημοσιευμένες εγγραφές της Amery είναι αναγκαίο να κριθούν ξεχωριστά. Ένα καλό μοντέλο θα έπρεπε να αναπαριστά το συγκεκριμένο πεδίο της επιρροής προερχόμενης από έναν Blogger.
- Δεύτερον, η επιρροή του κάθε δημοσιοποιημένου σχόλιου μπορεί να είναι διαφορετική, ανάλογα με το ποιος κοινοποιεί τα σχόλια. Τα σχόλια των Bob και Cary σχετικά με τη 1η δημοσίευση της Amery πρέπει να αντιμετωπίζονται διαφορετικά, και είναι προφανές ότι το σχόλιο του Cary θα ενίσχυε περισσότερο την επιρροή της 1ης δημοσίευσης (post1).
- Τρίτον, τα σχόλια από τους υπόλοιπους bloggers μπορεί να είναι είτε θετικά είτε αρνητικά είτε ουδέτερα. Αυτοί οι “συναισθηματικοί” παράγοντες επίσης επηρεάζουν τον αντίκτυπο του δημοσιεύματος προς τους σχολιαστές.



Εικόνα 1:Γράφος Επιρροής

Μια άλλη λογική βάση της οποίας γίνεται αξιολόγηση του ατόμου blogger στηρίζεται στην εξής θεώρηση: «Οι δημιουργοί ή σχολιαστές δημοσιοποιημένης εγγραφής κάποιου ιστολογίου που διακρίνονται για την επίδραση που προκαλούν δεν εντοπίζονται με βάση το αν ασκείται επιρροή ή όχι από τις αντίστοιχες ιστοσελίδες (τύπου ιστολογίου) που σχετίζονται με αυτούς».

Το πρόβλημα του εντοπισμού σε μια ιστοσελίδα τύπου Blog των Bloggers που προκαλούν αντίκτυπο είναι καινοτόμο. Προς τη κατεύθυνση επίλυσης του προβλήματος αυτού είναι σημαντικό *καταρχήν* να βρεθεί ένα μοντέλο το οποίο θα «εκφράζει σε μορφή ποσότητας» την έννοια του επηρεασμού που ασκείται από ορισμένους Bloggers και *κατά δεύτερον* ένα μοντέλο το οποίο να εντοπίζει τους Bloggers που ανήκουν σε διάφορες κατηγορίες. Αξίζει να σημειωθεί ακόμη ότι τα μοντέλα αυτά απαιτούν την ενσωματωμένη χρήση των συγκεκριμένων πληροφοριών που είναι δημοσιευμένες σε ιστοσελίδα τύπου Blog. Οι Bloggers που ασκούν επιρροή την ασκούν με φυσικό τρόπο σε άλλα μέλη της κοινότητας, εκφράζοντας με κυρίαρχο τρόπο την επικρατούσα μόδα και μπορούν να επηρεάσουν τα συμφέροντα της ομάδας σε μια κοινότητα. Αυτοί αποτελούν τους «αγωγούς μεταφοράς» των πληροφοριών στις κοινότητές τους. Οι Bloggers που προκύπτουν ως αποτέλεσμα της λειτουργίας των δύο παραπάνω μοντέλων δε συνεπάγεται απαραίτητα πως είναι και «ενεργοί» .

Οι μέθοδοι αντιμετώπισης του προβλήματος εύρεσης των ατόμων τύπου “influential”, «απεικονίζονται» μέσω:

- ✓ Της διεξαγωγής πειραμάτων τα οποία επεξεργάζονται δεδομένα «πραγματικού κόσμου». Τα δεδομένα αυτά προέρχονται από ιστοσελίδα τύπου «ιστολόγιο» που ανήκει στη «κοινωνία» Microsoft MSN space.
- ✓ Της εκτίμησης πολλαπλών πτυχών του προβλήματος
- ✓ Των συζητήσεων για υλοποίηση μεθόδων προκλητικά καινοτόμων.

Συνεπώς, τα blogs γίνονται μια από τις πιο ευρέως χρησιμοποιηθείσες διαδικτυακές υπηρεσίες.

Όσον αφορά το κοινωνικό είδος ιστολογίων, η μέθοδος εύρεσης των **Influential Bloggers** θεωρεί τα δύο ακόλουθα :

- Υπάρχουν τέσσερα είδη σχολιαστών/δημιουργών δημοσιευμένης καταχώρισης/καταχωρήσεων σε blog. Αυτά είναι: 1<sup>ο</sup>) οι Bloggers που εξασκούν επίδραση και είναι «ενεργοί», 2<sup>ο</sup>) οι Bloggers που δεν εξασκούν επιρροή και είναι «ενεργοί», 3<sup>ο</sup>) οι σχολιαστές/δημιουργοί δημοσιοποιημένης καταχώρισης/καταχωρήσεων σε ιστολόγιο που δεν είναι «ενεργοί» και δεν εξασκούν επιρροή , 4<sup>ο</sup>) οι Bloggers που δεν είναι «ενεργοί» και ασκούν επίδραση.
- Η συστηματική σειρά ενεργειών που είναι απαραίτητη για τον εντοπισμό των Bloggers που εξασκούν επιρροή διευκολύνεται όταν λαμβάνει «υπόψη της» έναν συνδυασμό των στατιστικών που διατυπώθηκαν σε προηγούμενη παράγραφο.

## ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “Identifying the influential bloggers in a community”, in *WSDM*, 2008.

[2] D. Kempe, J. Kleinberg, and Éva Tardos, “Maximizing the spread of influence through a social network”, in *KDD '03*. ACM, 2003.

[3] J. Scrpss, P.-N. Tan, and Abdol-Hossein, “Node roles and community structure in networks”, *WEBKDD*, 2007

[4] <http://www.sifry.com/alerts/>

[5] <http://www.ratcliffeblog.com>

[6] <http://googleblog.blogspot.com/>

[7] <http://www.tuaw.com/>

[8] Tim O’Reilly, “What is Web 2.0 , design patterns and business models for the next generation of software”

available at <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, September 2005

[9] Matthew Richardson and Pedro Domingos, “Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge Discovery and Data mining*”, pages 61–70, New York, NY, USA, 2002, ACM Press.

- [10] Daniel Gruhl, R. Guha, Ravi Kumar, Jasmine Novak, and Andrew Tomkins, “*The predictive power of online chatter*”, In *KDD '05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 78–87, New York, NY, USA, 2005. ACM Press.
- [11] Akshay Java, Pranam Kolari, Tim Finin, and Tim Oates, “Modeling the spread of influence on the blogosphere”, In *Proceedings of the 15th International World Wide Web Conference*, 2006.
- [12] Ed Keller and Jon Berry, “*One American in ten tells the other nine how to vote, where to eat and, what to buy. They are The Influentials.*” The Free Press, 2003.
- [13] T. Elkin, “Just an online minute... online forecast.”, available at <http://publications.mediapost.com/index.cfm?fuseaction=Articles.showArticleartaid=29803>.
- [14] Daniel Drezner and Henry Farrell, “*The power and politics of blogs.*”, in American Political Science Association Annual Conference, 2004.
- [15] G. W. Flake, S. Lawrence, and C. L. Giles, “*Efficient identification of web community*”, in 6th International Conference on Knowledge Discovery and Data Mining, 2000.
- [16] Yu-Ru Lin, Hari Sundaram, Yun Chi, Jun Tatemura, and Belle Tseng, “*Discovery of blog communities based on mutual awareness*”, In Proceedings of the 3rd annual workshop on weblogging ecosystem: aggregation, analysis and dynamics, 2006.
- [17] Ying Zhou and Joseph Davis, “*Community discovery and analysis in blogspace*”, in Proceedings of the 15<sup>th</sup> international conference on World Wide Web, pages 1017–1018, New York, NY, USA, 2006. ACM Press.
- [18] T. Coffman and S. Marcus, “*Dynamic classification of groups through social network analysis and HMMs*”, in *Proceedings of IEEE Aerospace Conference*, 2004.
- [19] Kathy E. Gill, “*How can we measure the influence of the blogosphere?*”, in Proceedings of the WWW'04: workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, 2004.
- [20] David Kempe, Jon Kleinberg, and Eva Tardos, “*Maximizing the spread of influence through a social network*”, In Proceedings of the KDD, pages 137–146, New York, NY, USA, 2003. ACM Press.
- [21] Akshay Java, Pranam Kolari, Tim Finin, and Tim Oates, “*Modeling the spread of influence on the blogosphere*”, In Proceedings of the 15th International World Wide Web Conference, 2006.
- [22] <http://technorati.com/developers/api/cosmos.html>
- [23] Identifying the Influential Bloggers in a Community, Nitin Agarwal, Huan Liu, Lei Tang - Arizona State University Tempe, AZ 85287, USA
- [24] ” Identifying Influential Bloggers: Time Does Matter “ Leonidas Akritidis, Dimitrios Katsaros, Panayiotis Bozanis - Department of Computer & Communication Engineering, University of Thessaly Volos, Greece
- [25] “Mining Influential Bloggers: from General to Domain Specific”  
Yichuan Cai, Yi Chen - Arizona State University P.O. Box 878809, Tempe, AZ 85287 - 8809, USA

## 2<sup>ο</sup> ΚΕΦΑΛΑΙΟ – Υπολογισμός του Σκορ Επιρροής ενός Blogger

Στο παρόν κεφάλαιο εξετάζεται το πρόβλημα αναγνώρισης των Bloggers που ασκούν επιρροή και αναλύεται ο τρόπος υπολογισμού της επιρροής ενός post ενός blogger καθώς επίσης της συνολικής επιρροής του Blogger.

### 2.1 Προσεγγίσεις Εύρεσης των Influential Bloggers

Το 1<sup>ο</sup> μοντέλο που προσπάθησε να επιλύσει το πρόβλημα αναγνώρισης των bloggers που ασκούν επιρροή είναι γνωστό ως «ροή της επιρροής» (**Influence – flow method**). Αυτό το μοντέλο έκανε διάκριση μεταξύ των ισχυρών για την επιρροή τους Bloggers και των «ενεργών» (παραγωγικών). Παράλληλα διατύπωσε συγκεκριμένες θεωρήσεις για τη «Μπλογκόσφαιρα» όπως το μέγεθος της δημοσιευμένης εγγραφής κάποιου blog, τον αριθμό σχολίων και τόσο το σύνολο των εισερχόμενων συνδέσμων όσο και το σύνολο των εξερχόμενων συνδέσμων που περιλαμβάνονται σε κάποιο ιστολόγιο. Καθένα από αυτά τα σύνολα αφορά σε ένα ή περισσότερα διαφορετικά ιστολόγια. Μειονέκτημα του συγκεκριμένου μοντέλου είναι πως δεν ενσωματώνει δύο σημαντικά στοιχεία:

- a) τις χρονικές παραμέτρους
- b) τη παραγωγικότητα

Κρίνεται επομένως αναγκαία η ταυτόχρονη θεώρηση των χρονικών παραμέτρων και της παραγωγικότητας της συμπεριφοράς της blogging διαδικασίας σε παραλληλία με τη διασύνδεση (μέσω παραπομπών) μεταξύ των δημοσιεύσεων των Blogs.

Το πρόβλημα της αναγνώρισης, σε κάποιον διαδικτυακό τόπο τύπου Blog, των χρηστών “Bloggers” που ασκούν επιρροή είναι πολύ σημαντικό. Η διαδικασία εύρεσης των ιστότοπων τύπου Blog που έχουν «χαρακτήρα» άσκησης επίδρασης [6] και η σχετική μελέτη για το εύρος της επίδρασης μεταξύ ιστοσελίδων τύπου blog [7],[8],[9],[10], είναι «**ορθογώνια**» στο προαναφερθέν πρόβλημα και δεν είναι ακριβώς εκείνη που ενδιαφέρει.

Για να κριθούν οι δημοσιοποιημένες εγγραφές των Blogs, προτάθηκε ένα διαισθητικό μοντέλο (2<sup>ο</sup> μοντέλο), το οποίο είναι βασισμένο στις παραμέτρους (ιδιότητες):

- 1) **Αναγνώριση**, η οποία έχει μέγεθος η τιμή του οποίου καθορίζεται από το πλήθος των δημοσιοποιημένων εγγραφών (οι οποίες ίσως προέρχονται από ένα ή περισσότερα διαφορετικά ιστολόγια) που αναφέρονται σε μια δημοσιευμένη καταχώριση (έστω « $p_i$ ») και από το πλήθος των εισερχόμενων συνδέσμων της  $p_i$ . Ακόμη, το σκορ “ $I(p_i)$ ” **μεγαλώνει ανάλογα με** το σκορ επηρεασμού των δημοσιευμένων καταχωρήσεων που αναφέρονται στη « $p_i$ ».
- 2) **Γέννηση της δραστηριότητας**, που ορίζεται ως το κατά πόσο η συγκεκριμένη δημοσιοποιημένη εγγραφή είναι δυνατό να: i) δέχεται ένα σημαντικό πλήθος (έστω « $\gamma$ ») από σχόλια ως αποτέλεσμα της επιρροής που ασκεί η  $p_i$  ii) να ξεκινά κάποιες νέες συζητήσεις, δηλαδή να προκαλεί την έκφραση αρκετού σχολιασμού και να δέχεται πληθώρα εισερχόμενων συνδέσμων.
- 3) **Καινοτομία**, που έχει μέγεθος η τιμή του οποίου είναι:
  - i) Αντιστρόφως ανάλογη με το πλήθος (έστω « $\theta$ ») των εξερχόμενων συνδέσμων οι οποίοι αφορούν σε ένα συγκεκριμένο ιστολόγιο ή σε δημοσιοποιημένη εγγραφή/εγγραφές κάποιας ιστοσελίδας τύπου Blog ή σε άρθρα.
  - ii) Ανάλογη με το μέγεθος της επίδρασης που ασκείται από δημοσιοποιημένη εγγραφή.
- 4) **Ευγλωττία**, η τιμή του οποίου είναι αντιστρόφως ανάλογη με την ασκούμενη επίδραση από τη δημοσιοποιημένη εγγραφή κάποιου Blog. Κάποια δημοσιοποιημένη εγγραφή που έχει ένα σημαντικό μήκος (έστω « $\lambda$ ») είναι βέβαιο πως θα χαρακτηρίζεται από τη σημαντική επιρροή που ασκεί. Μπορεί να αναπτυχθεί «ευρετική» (heuristic) μέθοδος που έχει ως **βάση της το « $\lambda$ » και υπολογίζει εάν θα ισχύει ή όχι το συμπέρασμα που διατυπώθηκε πριν λίγο.**

Οι παραπάνω παράμετροι χρησιμοποιούνται για τη δημιουργία γράφου επηρεασμού, στον οποίο η επιρροή ρέει μεταξύ των κόμβων. Ο κάθε κόμβος αναπαριστά μία απλή δημοσιοποιημένη εγγραφή κάποιου ιστολογίου, η οποία χαρακτηρίζεται από τις

## 2<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Υπολογισμός του Σκορ Επιρροής ενός Blogger

τέσσερις παραμέτρους που προαναφέρθηκαν. Όλες οι ιδιότητες που διατυπώθηκαν προηγουμένως αν συνδυαστούν είναι δυνατό να υποδείξουν τις δημοσιοποιημένες καταχωρήσεις που εξασκούν επιρροή. Ακόμη, οι παραπάνω ιδιότητες αποτελούν τη βάση επάνω στην οποία θα στηριχθεί η δημιουργία μοντέλου που εξετάζει και επίσης αναλύει, τροποποιεί και επεκτείνει αυτό το 2<sup>ο</sup> μοντέλο.

Επιπροσθέτως, σχετικά με τις δημοσιευμένες εγγραφές κάποιου ιστολογίου χρησιμοποιείται **γράφος ροής επηρεασμού** μεταξύ των κόμβων (δημοσιοποιημένες καταχωρήσεις). Καθένας από αυτούς τους κόμβους χαρακτηρίζεται από ορισμένες ιδιότητες που είναι οι «ι», «θ», «γ», «λ». Έστω πως “I” είναι το σκορ επιρροής της δημοσιευμένης καταχώρισης (post) ενός blog. Τότε η μεταβλητή “**InfluenceFlow**”(ροή της επιρροής) ενός ιστολογίου υπολογίζεται ως εξής :

$$\mathbf{InfluenceFlow}(p) = w_{in} \sum_{m=1}^{|I|} I(p_m) - w_{out} \sum_{n=1}^{|\theta|} I(p_n) \quad \text{εξίσ.(α)}$$

Τα «βάρη»  $w_{in}$ ,  $w_{out}$  χρησιμοποιούνται για τη προσαρμογή της συνεισφοράς της εισερχόμενης και εξερχόμενης επίδρασης αντίστοιχα. Η “ $p_m$ ” είναι μεταβλητή που «απεικονίζει» όλες τις δημοσιευμένες εγγραφές που συνδέονται προς (αναφέρουν δηλαδή) μία δημοσιοποιημένη καταχώριση “ $p$ ” όπου  $1 \leq m \leq |I|$  ενώ η “ $p_n$ ” «απεικονίζει» τις δημοσιευμένες εγγραφές που αναφέρονται από τη “ $p$ ” και ισχύει εδώ:  $1 \leq n \leq |\theta|$ . Οι  $|I|$ ,  $|\theta|$  είναι αντίστοιχα τα ολικά πλήθη των εισερχόμενων και εξερχόμενων συνδέσμων της “ $p$ ”. Η “**InfluenceFlow**” μετράει την διαφορά μεταξύ της ολικής εισερχόμενης επιρροής των εισερχόμενων συνδέσμων από την συνολική εξερχόμενη επίδραση των εξερχόμενων συνδέσμων που αναφέρονται στη “ $p$ ”. Η “**InfluenceFlow**” έχει οριστεί για «λογαριασμό» του επηρεασμού προερχόμενου από μια δημοσιευμένη εγγραφή και ο οποίος εξαρτάται από τους εισερχόμενους/εξερχόμενους συνδέσμους της εγγραφής.

Από την εξίσ.(α) προκύπτει ότι το πλήθος των εισερχόμενων συνδέσμων μιας δημοσιευμένης καταχώρισης είναι ανάλογο προς το βαθμό έως τον οποίο αναγνωρίζεται μια “ $p$ ” και προς το βαθμό έως τον οποίο εξασκείται επηρεασμός από την “ $p$ ”. Επιπλέον, το “I” είναι ανάλογο προς το πλήθος των σχολίων, δηλαδή το « $\gamma_p$ », το σύνολο των οποίων καταχωρείται προς μια δημοσιευμένη εγγραφή (τη “ $p$ ”).

Η συνολική επιρροή που αφορά σε δημοσιοποιημένη καταχώριση (έστω τη “p”) είναι η **εξίσωση** που ακολουθεί:

$$I(p) \propto w_{com} \gamma_p + InfluenceFlow(p) \quad \text{εξίσ.(\beta)}$$

Στον παραπάνω τύπο, η  $w_{com}$  αναπαριστά το βάρος που διαχειρίζεται τη συνεισφορά του  $\gamma_p$  προς τον επηρεασμό που εξασκεί η “p”. Η συνάρτηση “w “ εξαρτάται από το **μήκος « λ » δημοσιευμένης καταχώρισης** και «κρίνει» (επικροτεί/κατακρίνει) το  $I(p)$  με βάση το «λ» της “p”. Με αφετηρία τις εξίσ.(α), εξίσ.(β) **προκύπτει** ο παρακάτω τύπος που καθορίζει την επιρροή που εξασκείται από μια δημοσιευμένη καταχώριση (τη “p”):

$$I(p) = w(\lambda) \times (w_{com} \gamma_p + InfluenceFlow(p)) \quad \text{εξίσ.(\gamma)}$$

Από την εξίσ.(γ) είναι εφικτός ο υπολογισμός του επηρεασμού που εξασκείται από κάθε δημοσιευμένη καταχώριση ενός ιστολογίου και είναι δυνατό να ρυθμιστεί τόσο το  $w_{com}$  όσο και τα βάρη που εμπεριέχει η  $w(\lambda)$ . Το  $w(\lambda)$  «δηλώνει» το κατά πόσο ισχύει η ιδιότητα “goodness” για μια δημοσιευμένη καταχώριση καθώς το «λ» είναι ευρετικό μέτρο της ιδιότητας αυτής. Το βάρος  $w(\lambda)$  απλώς διαβαθμίζει το σκορ επιρροής ( $I(p)$ ) μιας δημοσιευμένης καταχώρισης σε blog, οπότε η αλλαγή της τιμής του  $w(\lambda)$  δεν αναμένεται να επηρεάσει την διάταξη των bloggers που εξασκούν επηρεασμό αλλά να περιορίσει κλιμακωτά ή να αυξήσει κλιμακωτά τα σκορ επιρροής. Βασιζόμενοι στις εξίσ.(α) και εξίσ.(γ) προκύπτει η μετρική **i-index (αυτή η μετρική ονομάζεται και προκαταρκτικό μοντέλο)**

$$I(p) = w(\lambda) (w_{com} \gamma_p + w_{in} \sum_{m=1}^{|\lambda|} I_p(m) - w_{out} \sum_{n=1}^{|\theta|} I_p(n)) \quad (\text{εξίσωση της i-index})$$

Χρησιμοποιώντας το post με το μέγιστο σκορ επιρροής ως το αντιπροσωπευτικό ενός blogger, είναι δυνατόν να καθοριστεί το i-index (blogger influence index) ενός Blogger (“b”) μέσα σε κάποιο ιστολόγιο.

$$iIndex(b) = \max(I(p_i)) \quad \text{εξίσ.(\delta)}$$

όπου  $1 \leq i \leq N$  και  $N$  το πλήθος των posts του blogger. Δοθέντος ενός συνόλου  $U$  από  $M$  bloggers  $\{b_1, b_2, \dots, b_M\}$ , το πρόβλημα της εύρεσης του influential blogger έγκειται



## 2<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Υπολογισμός του Σκορ Επιρροής ενός Blogger

στον προσδιορισμό ενός υποσυνόλου  $V$  από  $K$  bloggers  $\{b_{j_1}, b_{j_2}, \dots, b_{j_K}\}$  οι οποίοι ταξινομούνται με βάση το σκορ  $i$ -index δηλαδή  $i\text{Index}(b_{j_1}) \geq i\text{Index}(b_{j_2}) \geq \dots \geq i\text{Index}(b_{j_K})$ . Το σύνολο  $V$  περιέχει τους “ $K$ ” bloggers που εξασκούν τον μεγαλύτερο δυνατό επηρεασμό. Συνεπώς, για όλα τα posts  $\{p_1, p_2, \dots, p_L\}$  όλων των  $M$  bloggers, ως post που ασκεί επίδραση θεωρείται εκείνο για το οποίο ισχύει  $I(p_l) \geq i\text{-index}(b_{j_K})$ , όπου  $1 \leq l \leq L$ . Οπότε κάποιος blogger που δημοσιεύει μία εγγραφή για την οποία ισχύει το  $I(p_l) \geq i\text{-index}(b_{j_K})$  θεωρείται προφανώς ένας influential blogger.

Ένα μοντέλο που είναι αποτελεσματικό για την αναγνώριση εκείνων που ασκούν επίδραση και έχουν ρόλο σχολιαστή/δημιουργού δημοσιοποιημένης καταχώρισης/καταχωρήσεων σε ιστολόγιο βασίζεται σε όλες υποχρεωτικά τις παρακάτω παραμέτρους :

- 1) Τη συσχέτιση «δημοσιευμένης εγγραφής» – «απάντησης προς αυτή» χωρίς όμως να δίνει από μόνη της αξιόπιστα αποτελέσματα κυρίως λόγω του τρόπου με τον οποίον σχολιάζουν ορισμένοι με ρόλο σχολιαστή/δημιουργού δημοσιευμένης καταχώρισης/καταχωρήσεων σε blog.
- 2) Τη θεώρηση του πλήθους των **εξωτερικών συνδέσμων** που αφορούν προς κάποιο blog στο οποίο τουλάχιστον ένα άτομο έχει ρόλο σχολιαστή/δημιουργού δημοσιοποιημένης καταχώρισης. Δηλαδή, όταν ένα μεγάλο πλήθος εξωτερικών συνδέσμων αναφέρονται προς τον σύνδεσμο του blog το οποίο είναι ιδιοκτησία ενός/μιας Blogger, συνεπάγεται ότι αυτό το άτομο Blogger ασκεί επίδραση. Οι **εξωτερικοί σύνδεσμοι** ενδέχεται να αναφέρονται και προς μία ιστοσελίδα (εκτός των ιστολογίων).
- 3) Την αξιολόγηση του κατά πόσο χαρακτηρίζεται από αυθεντία ή όχι μια ιστοσελίδα (ή ένας σύνδεσμος). Οι αλγόριθμοι που κάνουν αυτή την αξιολόγηση είναι οι **Pagerank και Hits**. Αυτήν η αυθεντία είναι καίριο γνώρισμα για τη διαδικασία κριτικής του επηρεασμού που εξασκεί ένας blogger.

## 2.2 Μοντέλο «Θεματικού Τομέα Ενδιαφέροντος»

Η διαδικασία ανάλυσης ενός συγκεκριμένου θεματικού τομέα ενδιαφέροντός ενός Blogger αναπαρίσταται από το **μοντέλο-διάνυσμα «θεματικού τομέα ενδιαφέροντος»**. Το διάνυσμα «θεματικού τομέα ενδιαφέροντος» αναπαριστά τη πιθανότητα του να ανήκει μια δημοσιοποιημένη καταχώριση σε συγκεκριμένο θεματικό τομέα. Ένας χώρος blog του Blogger αποτελείται από πολλαπλές δημοσιευμένες καταχωρίσεις και κάθε μία από αυτές μπορεί να ανήκει σε έναν ή πολλαπλούς θεματικούς τομείς ενδιαφέροντος. Για παράδειγμα, κάποιες δημοσιευμένες εγγραφές αναφορικά με ειδήσεις από το χώρο του NBA μπορεί να προέρχονται από το θεματικό πεδίο των “σπορ”, η μετάδοση της ομιλίας του κ. Ομπάμα στις ΗΠΑ για την έναρξη των διαπραγματεύσεων του ASU ανήκει στον θεματικό τομέα της “πολιτικής” καθώς και στον θεματικό τομέα της “εκπαίδευσης”. Το ενδιαφέρον που σχετίζεται με το post με αύξων αριθμό  $d_k$  ενός blogger, ο οποίος έχει όνομα που δηλώνεται από τη  $b_i$ , εκφράζεται σε μορφή «ποσότητας» στον συνολικό χώρο ενδιαφερόντων από το παρακάτω διάνυσμα (πίνακα (interest vector) ):

$$IV(b_i, d_k) = \{ iv_1, iv_2, \dots, iv_N \}$$

στο οποίο, η τιμή του  $iv_t$  ανήκει στο διάστημα  $[0,1]$  (π.χ το  $iv_1$  μπορεί να αναφέρεται στο ενδιαφέρον στα αθλητικά και το  $iv_2$  στο ενδιαφέρον στην εκπαίδευση) και αυτό δείχνει το πόσο ενδιαφέρον υπάρχει σε ένα συγκεκριμένο τομέα. Ακόμη, στον πίνακα αυτό ισχύει  $1 \leq t \leq N$  δηλαδή το  $t$  παίρνει σα μέγιστη τιμή το  $N$  που δείχνει το πλήθος των τομέων που μπορεί να προκαθοριστεί σύμφωνα με κάποιες στάνταρντ κατηγορίες (όπως το Open Directory Project). Η  $t$  είναι ο αύξων αριθμός της διάστασης στο χώρο “ενδιαφέροντος”. Έτσι, καθορίζεται ένα πρότυπο που βασίζεται σε ένα πίνακα ενδιαφέροντος, ο οποίος αναπαριστά την πιθανότητα ενός post να ανήκει σε ένα συγκεκριμένο τομέα. Πιο επίσημα, ο “interest vector” ενός post υπολογίζεται ως εξής [3,4,5]:

$$iv(b_i, d_k, C_t) = (P(b_i, d_k | C_t)P(C_t)) / \sum_{n=1}^N (P(b_i, d_k | C_n)P(C_n))$$

**(εξίσωση του interest vector)**

όπου  $P(C_t|b_i, d_k)$  είναι η πιθανότητα ένα post  $d_k$  ενός blogger  $b_i$  να ανήκει στη κατηγορία ενδιαφερόντων  $C_t$  και  $iv(b_i, d_k, C_t)$  είναι η πιθανότητα το post  $d_k$  του

## 2<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Υπολογισμός του Σκορ Επιρροής ενός Blogger

blogger  $b_i$  να ανήκει στην κατηγορία  $C_i$  κανονικοποιημένη από το άθροισμα της πιθανότητας του να ανήκει το post  $d_k$  του  $b_i$  σε όλες τις προκαθορισμένες κατηγορίες.

*Η επιρροή ενός Blogger σε ένα συγκεκριμένο πεδίο ενδιαφέροντος μπορεί να θεωρηθεί ως μία συνιστώσα της συνολικής επίδρασης που ασκεί ένας blogger. Ως εκ τούτου, πριν την εξέταση της επιρροής ενός blogger σε μία συγκεκριμένη θεματική περιοχή, είναι αναγκαίο να εκφράζουμε τη συνολική του επιρροή σε μορφή «ποσότητας».*

Το σκορ της συνολικής επιρροής που ασκείται αποτελείται από 2 τμήματα. Το 1ο τμήμα είναι το σκορ επιρροής που αφορά στο άθροισμα των επιδράσεων προερχόμενων από τις δημοσιευμένες καταχωρίσεις αυτού και είναι γνωστό ως Accumulated Post (AP), ενώ το 2ο τμήμα είναι το σκορ επιρροής που αφορά στην «εξουσία» (authority) αυτού του Blogger στο διαδίκτυο και είναι γνωστό ως General Links (GL) για έναν blogger. Η βαθμολογία του GL για καθέναν/καθεμία blogger μπορεί να υπολογιστεί από τον “αυθεντικό” αλγόριθμο Pagerank [1] ή σαν άμεση προσέγγιση με ορισμένες μεθόδους παροχής της τιμής του PageRank ή να προσεγγιστεί από τον πάροχο της αξίας του PageRank δηλαδή το cubestat (<http://www.cubestat.com>). Συνεπώς, ο βαθμός της συνολικής προσωπικής επίδρασης που ασκεί ένας/μία blogger έχει τιμή που υπολογίζεται από τον τύπο:

$$\text{Inf}(b_i) = \alpha * \sum_{k=1}^{|P_i|} \text{Inf}(b_i, d_k) + (1-\alpha) * \text{GL}(b_i) \quad \text{εξίς.}(2)$$

Στον παραπάνω τύπο,  $\text{Inf}(b_i, d_k)$  είναι η τιμή του σκορ AP που προαναφέρθηκε ενώ  $|P_i|$  είναι το ολικό πλήθος δημοσιευμένων εγγραφών του Blogger  $b_i$ . Το  $\text{Inf}(b_i, d_k)$  είναι το σκορ επίδρασης της δημοσιευμένης εγγραφής  $d_k$  του Blogger  $b_i$ . Το  $\text{GL}(b_i)$  είναι το σκορ επίδρασης του Blogger  $b_i$ . Η  $\alpha$  είναι παράμετρος με ρόλο τον συντονισμό της σχετικής σημασίας του σκορ AP και του GL.

Το σκορ κάθε δημοσιευμένης εγγραφής, απεικονίζεται τόσο από τη ποιότητά της όσο και από το πλήθος των σχολίων που λαμβάνει. Η επίδραση που ασκείται από κάποια δημοσιοποιημένη εγγραφή έχει καίριο ρόλο όσον αφορά στη «ποιότητα» της συγκεκριμένης εγγραφής.

## 2<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Υπολογισμός του Σκορ Επιρροής ενός Blogger

Για να οριστεί το σκορ με τύπο  $Inf(b_i, d_k)$  πρέπει να ληφθεί υπόψη τόσο η ποιότητα της δημοσιευμένης καταχώρισης όσο και η επιρροή που ασκείται από τους σχολιαστές αυτής. Οπότε:

$$Inf(b_i, d_k) = \beta * QualityScore(b_i, d_k) + (1-\beta) * CommentScore(b_i, d_k) \quad \text{εξίσ.(3)}$$

Στα δύο παραπάνω μέλη αυτού του τύπου, η  $\beta$  είναι μια παράμετρος «βάρους».

Για την αξιολόγηση του  $QualityScore(b_i, d_k)$  πρέπει να λαμβάνονται υπόψη οι παρακάτω δύο παράγοντες:

- Το μήκος της δημοσιευμένης εγγραφής με οποιονδήποτε τρόπο και αν σκέφτηκε ο blogger που εξασκεί επιρροή για να τη δημιουργήσει. Αυτό το μήκος είναι ένδειξη για το πόσο εξασκεί επηρεασμό η συγκεκριμένη καταχώριση.
- Ο βαθμός του γνωρίσματος καινοτομία που χαρακτηρίζει τη δημοσιευμένη εγγραφή. Αυτός κυμαίνεται μεταξύ του 0 και του 1.

Η βαθμολογία  $CommentScore(b_i, d_k)$ , αντιπροσωπεύει την επίδραση που ασκείται από τους σχολιαστές κάποιας δημοσιευμένης εγγραφής. Η βαθμολογία κάθε σχολίου είναι ανάλογη προς τη συνολική επιρροή  $Inf(b_j)$  που ασκεί ο σχολιαστής blogger και προς τη στάση του σχολιαστή προς τη δημοσιευμένη καταχώριση. Αυτή η στάση «μεταφράζεται» σε ποσότητα με τον όρο  $SF(b_i, d_k, b_j)$ . Στον όρο αυτό,  $b_j$  είναι το άτομο σχολιαστής ενώ  $b_i$  είναι το άτομο δημιουργός της εγγραφής.

Για τη περίπτωση που κάποιος σχολιαστής σχολιάζει περισσότερο από έναν χώρους των Bloggers, η επίδρασή του διαμοιράζεται μεταξύ των ομότιμών του. Οπότε υπάρχει η δυνατότητα κανονικοποίησης του σκορ (βαθμολογίας) από την τιμή του  $TC(b_j)$  που είναι το ολικό πλήθος των σχολίων που δημοσιοποιήθηκαν από τον  $b_j$  σχολιαστή. Το σκορ (βαθμολογία) “comment score” που είχε αναφερθεί στην εξίσωση(3) ορίζεται ως ακολούθως.

$$CommentScore(b_i, d_k) = \sum_{j=1}^{|b_i, d_k|} (Inf(b_j) * SF(b_i, d_k, b_j)) / TC(b_j) \quad \text{εξίσ.(4)}$$

Ο όρος  $|b_i, d_k|$  είναι το ολικό πλήθος των σχολίων που έχουν εκφραστεί για τη δημοσιευμένη εγγραφή  $d_k$  του  $b_i$  ατόμου δημιουργού ή σχολιαστή δημοσιευμένης εγγραφής/εγγραφών κάποιου ιστολογίου. Ο όρος  $SF(b_i, d_k, b_j)$  (παράγοντας

ευαισθησίας) αντικατοπτρίζει την τοποθέτηση (άποψη) του σχολιαστή και η τιμή του μπορεί να ανήκει σε μια από τις κατηγορίες: “αρνητικό”, “θετικό”, “ουδέτερο”. Χρησιμοποιούνται οι ακόλουθες ευρετικές (heuristics) τεχνικές για να προβλεφθεί η τιμή του παράγοντα SF:

- i. Όσο ένα σχόλιο περιέχει επιβεβαιωμένα θετικές λέξεις (όπως «συμφωνούν», «υποστήριξη») θα αντιμετωπιστεί ως ένα θετικό σχόλιο.
- ii. Αν ένα σχόλιο περιέχει αρνητικές λέξεις (όπως «διαφωνώ», «μίσος», «μη αξιόπιστο»), θα αντιμετωπιστεί ως αρνητικό σχόλιο. Διαφορετικά, θα αντιμετωπιστεί ως ένα ουδέτερο σχόλιο.

Από τις εξίσ.(4) και εξίσ.(3), προκύπτει η ακόλουθη εξίσωση:

$$\text{Inf}(b_i, d_k) = \beta * \text{QualityScore}(b_i, d_k) + (1 - \beta) * \sum_{j=0}^{|b_i, d_k|-1} (\text{Inf}(b_j) * \text{SF}(b_i, d_k, b_j)) / \text{TC}(b_j)$$

**Εξίσ.(5)**

Όλες οι δημοσιοποιημένες καταχωρήσεις ενός blogger έχουν μια εξίσωση στη μορφή της (5). Οι εξισώσεις μιας δημοσιευμένης καταχώρισης οποιουδήποτε ατόμου δημιουργού ή σχολιαστή δημοσιευμένης εγγραφής σε ιστολόγιο, μπορούν να επιλυθούν αποτελεσματικά με χρήση της επαναληπτικής μεθόδου [2] και η λύση σ’ αυτό το σετ εξισώσεων δίνει το συνολικό σκορ επίδρασης του post ενός blogger. Από το σκορ αυτό μαζί με την εξίσ.(2) προκύπτει η συνολική βαθμολογία επίδρασης ασκούμενης από κάποιο άτομο με ρόλο σχολιαστή/δημιουργού δημοσιοποιημένης εγγραφής σε ιστολόγιο. Από το σκορ που μόλις προαναφέρθηκε μπορούμε να πάρουμε τη βαθμολογία για τη συνολική επιρροή του Blogger με τη χρήση και της εξίσ.(2) .

Η διαδικασία αξιολόγησης του σκορ επίδρασης ενός blogger κάποιου ιστολογίου μπορεί να αφορά και σε συγκεκριμένο θεματικό πεδίο (τομέα) ενδιαφέροντος όπως είναι τα πολύ δημοφιλή: ταξίδι, τέχνη, σπορ που ανήκουν στη “μπλογκόσφαιρα”. Για τη βελτίωση του τρόπου με τον οποίο γίνεται η εύρεση των Bloggers που εξασκούν επίδραση, γίνεται ανάλυση των θεματικών τομέων ενδιαφέροντός τους. **Η επιρροή που ασκεί ένας blogger σε ένα συγκεκριμένο πεδίο ενδιαφέροντος** είναι ανάλογη της συνολικής του επιρροής και του όρου  $\text{iv}(b_i, d_k, C_t)$  (σκορ ενδιαφέροντος του post  $d_k$  του blogger  $b_i$  στον τομέα  $C_t$ ) που αναφέρθηκε προηγουμένως.

$$\text{inf}(b_i, C_t) = \sum_{k=1}^{|P_t|} \text{inf}(b_i, d_k) \text{iv}(b_i, d_k, C_t)$$

εξίσ. (6)

Θεωρώντας ως παράδειγμα το διάγραμμα επιρροής της **Εικόνα 1** θα εξεταστεί λεπτομερώς ο υπολογισμός του σκορ επίδρασης για κάποιο συγκεκριμένο θεματικό τομέα ενδιαφέροντος του Blogger. Για την ακρίβεια, έστω ας χρησιμοποιηθεί ο πρώτος χαρακτήρας του ονόματος του blogger για να γίνει υπόδειξη για το σε ποιο τέτοιου είδους άτομο αναφερόμαστε. Θα έχουμε τις ακόλουθες εξισώσεις για το σετ εκείνων που έχουν ρόλο σχολιαστή/δημιουργού δημοσιευμένης καταχώρισης/καταχωρήσεων κάποιου Blog σύμφωνα με την (εξίσ.5):

$$\text{Inf}(A,P1) = \alpha * \text{QualityScore}(A,P1) + (1/2) * \beta * ((\alpha * \text{Inf}(B,P3)) + (1-\alpha) * (\text{GL}(B)) * \text{SF}(A,P1,B))) + (1/2) * \beta * ((\alpha * \text{Inf}(C,P4) + (1-\alpha) * \text{GL}(B)) * \text{SF}(A,P1,C))$$

$$\text{Inf}(A,P2) = \alpha * \text{QualityScore}(A,P2) + (1/2) * \beta * ((\alpha * \text{Inf}(C,P4) + (1-\alpha) * \text{GL}(C)) * \text{SF}(A,P1,C))$$

$$\text{Inf}(B,P3) = \alpha * \text{QualityScore}(B,P3)$$

$$\text{Inf}(C,P4) = \alpha * \text{QualityScore}(C,P4) + (1/2) * \beta * ((\alpha * \text{Inf}(B,P3)) + (1-\alpha) * (\text{GL}(B)) * \text{SF}(C,P4,B))$$

Μετά από την επίλυση των παραπάνω εξισώσεων, οι  $\text{inf}(A,P1)$ ,  $\text{inf}(A,P2)$ ,  $\text{inf}(B,P3)$ ,  $\text{inf}(C,P4)$  είναι μεταβλητές, είναι εύκολο να πάρουμε τις τιμές των  $\text{inf}(A)$ ,  $\text{inf}(B)$ ,  $\text{inf}(C)$  που είναι τα συνολικά σκορ (βαθμολογίες) των αντίστοιχων Bloggers. Βασιζόμενοι στα παραπάνω μπορούμε εύκολα να προσεγγίσουμε περαιτέρω άμεσα τα σκορ που αντιπροσωπεύουν την επίδραση προερχόμενη από το εκάστοτε θεματικό πεδίο ενδιαφέροντος του blogger. Τα σκορ που μόλις προαναφέρθηκαν είναι τα  $\text{Inf}(A, \text{Econ})$ ,  $\text{Inf}(B, \text{CS})$  και  $\text{Inf}(C, \text{CS})$  και αντιστοιχούν απευθείας στα πεδία  $\{\text{Econ}, \text{CS}\}$ .

### 2.2.1 Πειραματική Αξιολόγηση του Μοντέλου «Θεματικού Τομέα Ενδιαφέροντος»

Για την εκτίμηση της αποδοτικότητας των μεθόδων «εξόρυξης» των θεματικών πεδίων ενδιαφέροντος των Bloggers, γίνεται διεξαγωγή εμπειριστατωμένης ανάλυσης σε σετ «πραγματικών δεδομένων» έτσι όπως περιγράφεται ακολούθως.

Σαν σύνολο τέτοιου είδους δεδομένων χρησιμοποιείται το *Microsoft MSN space* που είναι από τους γνωστότερους παρόχους υπηρεσιών για ιστολόγιο και έτσι αυξάνεται η αποτελεσματικότητα της προσέγγισης που γίνεται μέσω της ανάλυσης. Ένας μελλοντικός στόχος είναι η απεικόνιση των Bloggers που εξασκούν επίδραση και η συνεργασία με επιχειρησιακού τύπου απαιτήσεις πραγματικής εφαρμογής. Στο *Microsoft MSN space*, το κάθε άτομο Blogger μπορεί όχι μόνο να δημοσιεύει εγγραφές στο ιστολόγιό του, αλλά επίσης να σχολιάζει καταχωρίσεις δημοσιευμένων εγγραφών οι οποίες προέρχονται από άλλους που δημιούργησαν δημοσιευμένες καταχωρίσεις του ιστολογίου. Για τη διεξαγωγή των πειραμάτων έχουν «συλλεχθεί» περίπου 1000 χώροι (Spaces) του MSN με τα προφίλ των χρηστών. Υπάρχουν 10 τομείς που καλύπτουν το μεγαλύτερο μέρος των θεματικών πεδίων ενδιαφέροντός τους και αυτοί προκαθορίζονται ως εξής: «Ταξίδια», «χρήση υπολογιστών», «επικοινωνία», «μόρφωση», «οικονομικά», «στρατιωτικά», «σπορ», «ιατρική», «τέχνη», «πολιτική». Όταν υπολογίζεται το General Link (GL) σκορ επιρροής, διαπιστώνεται ότι οι τιμές του PageRank αλγόριθμου για τους περισσότερους bloggers είναι πολύ μικρές. Το προηγούμενο συμβαίνει επειδή στο σύνολο δεδομένων, για το 90% από το σύνολο των blogs οι αξίες του αλγόριθμου PageRank είναι μικρότερες του 1, ενώ για το 99% από το σύνολο των blogs οι αξίες του αλγόριθμου PageRank είναι μικρότερες του 3. Αντί για τον απευθείας υπολογισμό της αξίας του PageRank είναι δυνατό να γίνεται προσέγγιση της αξίας όταν χρησιμοποιηθεί ο αλγόριθμος "Microsoft Live Indexed Pages" και η αξία λαμβάνεται από την ιστοσελίδα "Cubestat" (<http://www.cubestat.com/>).

Με βάση κάποιο ειδικό (ή μη) θεματικό πεδίο ενδιαφέροντος του Blogger, ισχύουν τα παρακάτω όσον αφορά στα πειράματα:

- Για να εκτιμηθεί η αποτελεσματικότητα του μοντέλου-διάνυσμα «**θεματικού τομέα ενδιαφέροντος**» γίνεται κλήση 10 χρηστών να κάνουν μια μελέτη και να συγκρίνουν τις επιδόσεις για τους 3 κορυφαίους σε άσκηση επιρροής bloggers οι

## 2<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Υπολογισμός του Σκορ Επιρροής ενός Blogger

οποίοι προέρχονται από **γενικά** (General) και **συγκεκριμένα** (Domain Specific) θεματικά πεδία ενδιαφέροντος (για παράδειγμα «Ταξίδια», «Τέχνη», «Αθλητισμός»).

- Για την εύρεση των κορυφαίων τριών bloggers τόσο στις λίστες του γενικού όσο και σε εκείνες του ειδικού θεματικού τομέα ενδιαφέροντος, θα πρέπει να σταλεί η διεύθυνση URL του κάθε blogger σε τελικούς χρήστες. Επιπλέον, προς αυτή τη κατεύθυνση, όσοι δημοσίευσαν/σχολίασαν καταχώρηση σε blog ζητούν από τους χρήστες να τους βαθμολογήσουν στη κλίμακα από 1 μέχρι 5, σύμφωνα με την κατανόησή τους για έναν συγκεκριμένο σενάριο χρήσης (όπως διαφημιστική καμπάνια βιομηχανίας).

Τα αποτελέσματα της μέσης τιμής βαθμολογίας της μελέτης που εφαρμόζεται εμφανίζονται στον Πίνακα 1. Όπως μπορούμε να δούμε, το **μοντέλο-διάνυσμα «θεματικού πεδίου ενδιαφέροντος» (interest vector model)** έχει καλύτερα αποτελέσματα αξιολόγησης από το αντίστοιχο «γενικό» σύστημα που **υποδεικνύει** τους bloggers που εξασκούν επιρροή σε γενικά θεματικά πεδία ενδιαφέροντος. Από αυτόν τον πίνακα παρατηρούμε ότι το πεδίο «σπορ» έχει πολύ υψηλότερη βαθμολογία αξιολόγησης στο συγκεκριμένο θεματικό τομέα ενδιαφέροντος (τιμή 4.6) σε σχέση με το «Γενικό» θεματικό πεδίο ενδιαφέροντος (τιμή 3.2).

Average Applicable Scores	Traveling	Art	Sports
General	3.2	3.4	3.2
Domain Specific	4.4	4.0	4.6

**Πίνακας 1:** Αποτελέσματα μέσης τιμής βαθμολογίας

Όσον αφορά στον αντίκτυπο των παραμέτρων στάθμισης (weighting parameters) στη διαδικασία εκτέλεσης πειραμάτων αξίζει να αναφερθούν τα παρακάτω:

- «α» είναι η παράμετρος για τη ρύθμιση της σχετικής σημασίας του συσσωρευμένου αντίκτυπου και της επιρροής ενός γενικού συνδέσμου.
- «β» είναι η παράμετρος για την προσαρμογή της σχετικής σημασίας του σκορ ποιότητας κάθε δημοσιευμένης καταχώρισης και του σκορ κάποιου σχόλιου.

Για να γίνουν φανερές οι επιπτώσεις από αυτές τις παραμέτρους ρύθμισης, θα επιλεγεί τυχαία ως παράδειγμα το πεδίο «Τέχνη». Ρυθμίζουμε τις παραμέτρους α και



β, σταθεροποιώντας της τιμής της α και αλλάζοντας την τιμή της άλλης και στη συνέχεια παρατηρούμε τη διακύμανση της κατάταξης των αποτελεσμάτων. Στον Πίνακα 2 δίνονται απλώς οι μεταβολές στη κατάταξη των 5 κορυφαίων bloggers ως προς το βαθμό άσκησης επιρροής, για μεταβολές της παραμέτρου β όταν η τιμή της α είναι ίση με 0.5.

$\beta = 0.9$	$\beta = 0.5$	$\beta = 0.1$	$\beta = 0.01$	$\beta = 0$
youyou	youyou	youyou	youyou	kelly
sky	sky	sky	kelly	winson
newwishes	newwishes	sabrina	sky	best
sabrina	sabrina	newwishes	sabrina	whenlove
Frank	Frank	kelly	newwishes	youyou

Πίνακας 2: Μεταβολές Κατάταξης των Κορυφαίων Bloggers σε σχέση με τη μεταβολή του β

Όπως προκύπτει από τον παραπάνω πίνακα το όνομα του 1ου κορυφαίου blogger που ασκεί επιρροή αλλάζει από "youyou" σε "Kelly". Ρίχνοντας μια προσεκτικότερη ματιά στις δημοσιευμένες καταχωρίσεις των δύο ατόμων που δημοσίευσαν/σχολίασαν καταχώρηση σε ιστολόγιο, είναι προφανές πως ο Blogger "youyou" έχει πολλές υψηλά ποιοτικές δημοσιευμένες καταχωρίσεις μαζί με πολλά θετικά σχόλια από σχολιαστές των οποίων ο δείκτης ένδειξης της επίδρασης που ασκούν είναι υψηλός. Η "Kelly" παρόλο που έχει δημοσιευμένες εγγραφές οι οποίες αναπαράγονται από άλλες πηγές που δεν είναι πλούσιες σε περιεχόμενο, έχει σχολιαστές που προκαλούν αρκετό επηρεασμό και αυτό πιθανώς οφείλεται στο ότι μερικοί από αυτούς προέρχονται από τους φίλους της από τη "μπλογκόσφαιρα". Ανεξάρτητα από τη καλή ή μη ποιότητα κάποιων/κάποιων δημοσιευμένων καταχωρίσεων του blog της Kelly, αυτή επιτυγχάνει πάντα να έχει υψηλό σκορ σχολιασμού. Το παράδειγμα αυτό δείχνει τη συσχέτιση μεταξύ των δύο δομικών στοιχείων μιας δημοσιευμένης εγγραφής, και αυτό το ζεύγος στοιχείων έχει επιπτώσεις στην αξιολόγηση του σκορ του αντίκτυπου ενός post.

## 2.3 Κατάταξη Blogs εναντίων Κατάταξης Ιστοσελίδων

Το πρόβλημα της κατάταξης των ιστολογίων ή των bloggers διαφέρει από αυτό της διερεύνησης των έγκυρων ιστοσελίδων (authoritative). Τα ιστολόγια στην μπλογκόσφαιρα είναι πολύ αραιά συνδεδεμένα και για το λόγο αυτό δεν είναι σωστό να διενεργείται μία κατάταξη των blogs χρησιμοποιώντας αλγορίθμους κατάταξης ιστοσελίδων

(π.χ. PageRank, Hits). Η διάσταση του χρόνου είναι πιο σημαντική για ένα ιστολόγιο. Ενώ μια ιστοσελίδα είναι πιθανό να γίνεται όλο και πιο έγκυρη με το πέρασμα του χρόνου (ο πίνακας γειτνίασής της γίνεται πιο πυκνός), η επιρροή ενός post ή ενός blogger φθίνει με το χρόνο. Αυτό οφείλεται στο γεγονός ότι ο πίνακας γειτνίασης των blogs γίνεται όλο και πιο αραιός καθώς χιλιάδες από νέα αραιά συνδεδεμένα posts εμφανίζονται κάθε μέρα.

### 2.4 Αξιολόγηση του Μοντέλου I-index

Ένα αρνητικό του 2<sup>ου</sup> μοντέλου (i-index) που περιγράφηκε στην παράγραφο 2.1 είναι η αγνόηση της **παραγωγικότητας (activity)** κάποιου Blogger καθώς ένας blogger μπορεί να έχει κοινοποιήσει μόνο ένα πολύ μικρό πλήθος δημοσιεύσεων που ασκούν επιρροή ενώ κάποιοι άλλοι bloggers ίσως να είχαν δημιουργήσει ορισμένες δεκάδες από posts με «χαρακτήρα» επιρροής των οποίων το σκορ όμως είναι χαμηλότερο από το post που εξασκούσε τη μεγαλύτερη δυνατή επίδραση και ήταν του προηγούμενου blogger.

Ένα επιπλέον μειονέκτημα του 2<sup>ου</sup> μοντέλου είναι η αποκλειστική εξάρτησή του από τις τιμές των **μεταβλητών στάθμισης** που καθορίζονται από τη πλευρά του χρήστη (είτε αυτός έχει την ιδιότητα σχολιαστή/δημιουργού δημοσιευμένης εγγραφής κάποιου blog είτε εκείνη του επισκέπτη του blog). Η διαφορετικότητα στις τιμές των παραμέτρων που προαναφέρθηκαν μπορεί να οδηγεί σε διαφορετικά αποτελέσματα διάταξης των bloggers. Συνεπώς, το προαναφερθέν μοντέλο δεν είναι αντικειμενικό καθώς μεταβάλλοντας τις κατάλληλες τιμές των μεταβλητών στάθμισης, προσδιορίζει τους Bloggers με διαφορετικά χαρακτηριστικά. Με άλλα λόγια, δεν είναι δυνατόν να παραχθεί μία ικανοποιητική απάντηση στην ερώτηση «Ποιος blogger εξασκεί το μέγιστο δυνατό επηρεασμό», αλλά μπορεί να δοθεί απάντηση σε ερωτήσεις του τύπου «Ποιος είναι ο blogger που εξασκεί τη μέγιστη δυνατή επιρροή ανάλογα με τον αριθμό των σχολίων που τα posts του λαμβάνουν».

Το κυριότερο όμως μειονέκτημα του i-index, είναι η αγνόηση του **παράγοντα χρόνος** της «Μπλογκόσφαιρας». Ο χρόνος σχετίζεται με την ηλικία ενός blog post, αλλά και με την ηλικία των εισερχόμενων συνδέσεων (inlinks) σε αυτό το post. Ένας blogger που ασκεί επιρροή είναι ένας blogger που έχει γράψει

posts που εξασκούν επηρεασμό πρόσφατα ή τα posts του έχουν ένα πρόσφατο «αντίκτυπο». Συνεπώς, ένα αποτελεσματικό πρότυπο που προσδιορίζει εκείνους που εξασκούν επιρροή και είναι σχολιαστές/δημιουργοί δημοσιευμένης καταχώρισης/καταχωρήσεων σε Blog/Blogs, πρέπει να λαμβάνει υπόψη την ημερομηνία που μια δημοσιευμένη καταχώριση υποβλήθηκε και τις ημερομηνίες που τα posts που έχουν ρόλο «παραπεμπτικό» δημοσιεύθηκαν, προκειμένου να υπάρχει η δυνατότητα προσδιορισμού των χρηστών των blogs οι οποίοι δεν έχουν τη δυνατότητα άσκησης επίδρασης. Επιπλέον, με τέτοιες απαιτήσεις είναι υποχρεωτικό να υπάρχουν γρήγορες μέθοδοι για την ανακάλυψη των bloggers που εξασκούν επιρροή. Αυτές οι μέθοδοι αποκλείουν τη χρήση απαιτητικών (demanding) και ασταθών (unstable) αναδρομικών ορισμών όπως εκείνων που χρησιμοποιούνται από το μοντέλο influence-flow που προτείνεται στο [18].

## ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ

- [1] L. Page, S. Brin, R. Motwani, and T. Winograd, “*The pagerank citation ranking: Bringing order to the web*”, Technical report, Stanford Digital Library Technologies Project, 1998.
- [2] <http://en.wikipedia.org/wiki/gauss-seidelmethod>.
- [3] Y. Liu, W. Liu, and C. Jiang, “*User interest detection on web pages for building personalized information agent*”, in Innovations in Information Technology, 2006.
- [4] W. Paik, S. Yilmazel, E. Brown, M. Poulin, S. Dubon, and C. Amice, “*Applying natural language processing (nlp) based metadata extraction to automatically acquire user preferences*”, in Proceedings of the 1st international conference on Knowledge capture, 2001.
- [5] <http://en.wikipedia.org/wiki/naivebayesclassifier>
- [6] K. E. Gill, “*How can we measure the influence of the Blogosphere?*” in Proceedings of WWE Workshop, 2004.
- [7] D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins, “*The predictive power of online chatter*”, in Proceedings of ACM KDD Conf., pages 78–87, 2005.
- [8] D. Gruhl, D. Liben-Nowell, R. Guha, and A. Tomkins, “*Information diffusion through Blogosphere*”, ACM SIGKDD Explorations, 6(2):43–52, 2004.
- [9] A. Java, P. Kolari, T. Finin, and T. Oates, “*Modeling the spread of influence on the Blogosphere*”, in Proceedings of ACM WWW Conf., 2006.

- [10] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. van- Briesen, and N. Glance, “*Cost-effective outbreak detection in networks*”, in Proceedings of ACM KDD Conf., 2007.
- [11] Kathy E. Gill, “*How can we measure the influence of the blogosphere?*”, in Proceedings of the WWW’04: workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, 2004.
- [12] Apostolos Kritikopoulos, Martha Sideri, and Iraklis Varlamis. Blogrank, “*ranking weblogs based on connectivity and similarity features*”, in AAA-IDEA ’06: Proceedings of the 2nd international workshop on Advanced architectures and algorithms for internet delivery and applications, page 8, 2006.
- [13] Ed Keller and Jon Berry, “*One American in ten tells the other nine how to vote, where to eat and, what to buy. They are The Influentials*”, The Free Press, 2003.
- [14] P. Kolari, T. Finin, and A. Joshi, “*SVMs for the blogosphere: Blog identification and splog detection*”, in AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs, 2006.
- [15] Yu-Ru Lin, Hari Sundaram, Yun Chi, Junichi Tatemura, and Belle L. Tseng, “*Splog detection using self-similarity analysis on blog temporal dynamics*”, in Proceedings of the 3rd international workshop on Adversarial information retrieval on the web (AIRWeb), pages 1–8, New York, NY, USA, 2007, ACM Press.
- [16] Gerald D. Fensterer, “*Planning and Assessing Stability Operations: A Proposed Value Focus Thinking Approach*”, PhD thesis, Air Force Institute of Technology, 2007.
- [17] R. L. Keeney and H. Raiffa, “*Decisions with Multiple Objectives: Preferences and Value Tradeoffs*”, Cambridge University Press, 1993.
- [18] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “*Identifying the influential bloggers in a community*”, in Proceedings of ACM WSDM Conf., pages 207–218, 2008.
- [19] K. E. Gill, “*How can we measure the influence of the Blogosphere?*”, in Proceedings of WWE Workshop, 2004.
- [20] D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins, “*The predictive power of online chatter*”, in Proceedings of ACM KDD Conf., pages 78–87, 2005.
- [21] D. Gruhl, D. Liben-Nowell, R. Guha, and A. Tomkins, “*Information diffusion through Blogosphere*”. ACM SIGKDD Explorations, 6(2):43–52, 2004.
- [22] A. Java, P. Kolari, T. Finin, and T. Oates, “*Modeling the spread of influence on the Blogosphere*”, in Proceedings of ACM WWW Conf., 2006.
- [23] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. van- Briesen, and N. Glance, “*Cost-effective outbreak detection in networks*”, in Proceedings of ACM KDD Conf., 2007.

### 3<sup>ο</sup> ΚΕΦΑΛΑΙΟ – Μετρική ΜΕΙΒΙ

Στο πλαίσιο αυτού του κεφαλαίου θα αναπτύξουμε μία νέα μέθοδο υπολογισμού της επιρροής ενός blog post, ορμώμενοι από τις παρατηρήσεις που κάναμε στην παράγραφο 2.4. Στο παρακάτω πίνακα περικλείουμε τους τύπους που θα χρησιμοποιήσουμε στον ορισμό της νέας μετρικής.

Σύμβολο	Έννοια
$BP(j)$	Το σετ δημοσιεύσεων του $j$ (στη σειρά) blogger
$Bp_j(i)$	Αφορά στην $i$ (στη σειρά) δημοσιευμένη καταχώριση σε κάποιο blog, από τον $j$ (στη σειρά) blogger
$C_j(i)$	Αφορά στο σετ των σχολίων που τοποθετήθηκαν για μια $i$ (στη σειρά) καταχώριση δημοσιευμένης εγγραφής του Blogger με αύξοντα αριθμό $j$
$R_j(i)$	Σχετίζεται με το σύνολο των δημοσιεύσεων που αναφέρονται στη δημοσιευμένη καταχώριση με αύξοντα αριθμό $i$ του blogger με αύξοντα αριθμό $j$
$\Delta TP_j(i)$	Σχετίζεται με το χρονικό διάστημα (σε μέρες) παρεμβολής μεταξύ του παρόντος χρόνου και της χρονικής στιγμής όπου αναρτήθηκε η $i$ δημοσιευμένη εγγραφή του blogger με αύξοντα αριθμό $j$
$\Delta TP(x)$	Σχετίζεται με το χρονικό διάστημα (σε μέρες) παρεμβολής μεταξύ του παρόντος χρόνου και της χρονικής στιγμής όπου αναρτήθηκε η « $x$ » (στη σειρά) δημοσιευμένη εγγραφή

**Πίνακας 3:** Συμβολισμοί που χρησιμοποιούνται στη ΜΕΙΒΙ

Αναμφίβολα, για κάποιον Blogger που ως προς το παρόν χαρακτηρίζεται από υψηλό βαθμό εξάσκησης επηρεασμού, δεν είναι σίγουρο ότι στο μέλλον θα συμβαίνει το ίδιο. Υπάρχει μία εξίσωση υπολογισμού σκορ η οποία μας δίνει την έννοια του κατά πόσο είναι επίκαιρο ή όχι ένα θέμα που έχει δημοσιευτεί σε ιστολόγιο μέσω της «γλώσσας» των αριθμών. Η εξίσωση αυτή υπολογίζει το σκορ (που συμβολίζεται ως  $S_j^m(i)$ ) ως εξής:

$$S_j^M(i) = \gamma(|C(i)|+1) (\Delta TP_j(i) + 1)^{-\delta} |R_j(i)|$$

Στον παραπάνω τύπο, οι παράμετροι  $\gamma$  και  $\delta$  δεν είναι αναγκαίες ούτε σημαντικές για τη ποσοτικοποίηση της έννοιας που προαναφέρθηκε. Η τιμή της βαθμολογίας που δίνει η  $S_j^m(i)$ , παίρνει τιμές οι οποίες είναι μεγαλύτερες ή ίσες του ένα.

Χρησιμοποιώντας τον ορισμό του σκορ  $S_j^m(i)$  θα εισαγάγουμε μία νέα μετρική αναγνώρισης των “influential bloggers”, γνωστή ως **MEIBI**. Το κυριότερο κίνητρο εισαγωγής των μεθόδων MEIBI και MEIBIX (επέκταση της MEIBI – θα εξετασθεί στο επόμενο κεφάλαιο) είναι ότι οι σχετικά «κλειστές» ανταγωνιστικές μέθοδοι, δεν έχουνε λάβει υπόψη τους τις χρονικές παραμέτρους του προβλήματος. Αυτές οι παράμετροι είναι οι πιο σημαντικές όταν απασχολούμαστε με πεδία όπως η «Μπλογκόσφαιρα» η οποία είναι ευμετάβλητη και διπλασιάζεται σε μέγεθος κάθε έξι μήνες. Σε αυτού του είδους τα πεδία, η διαδικασία διάχυσης της πληροφορίας παρομοιάζεται με τον τρόπο εξάπλωσης μιας μολυσματικής ασθένειας καθώς στη “Blogosphere” η συστηματική σειρά ενεργειών αντιγραφής και ανάγνωσης μιας πληροφορίας δημοσιευμένης εγγραφής εξαπλώνεται γρήγορα σύμφωνα με τους Gruhl et al [1]. Εκείνες οι ανταγωνιστικές μέθοδοι έχουν τον ίδιο σκοπό με την MEIBIX και τη MEIBI .

Σύμφωνα με τη MEIBI μετρική, για τη διαδικασία κρίσης λαμβάνονται υπόψη:

- ✓ *το πλήθος των εισερχόμενων συνδέσμων.*
- ✓ *Το πλήθος των σχολίων τα οποία αφορούν στις δημοσιεύσεις κάποιου ιστολογίου.*
- ✓ *Την ημερομηνία στην οποία τοποθετήθηκε μια καταχώριση σε κάποιο blog.*

Η μετρική MEIBI κρίνεται με τέτοιον τρόπο που είναι ενάντια στη μέθοδο εύρεσης των Bloggers που εξασκούν επηρεασμό με τη μέθοδο αυτή να διατυπώνεται στη [2]. Αυτός ο τρόπος αφορά στη χρήση δεδομένων προερχόμενων από μία «κοινωνία» «πραγματικού» κόσμου όπως είναι οι διαδικτυακοί τόποι τύπου blog. Τα αποτελέσματα που προκύπτουν δείχνουν σαφέστερα πως η MEIBI είναι σε θέση να προσδιορίσει καλύτερα τα σημαντικά χρονικά σχέδια (πλαίσια) στη «συμπεριφορά» όλων των ιστολογίων. Επιπλέον, αναθέτει διαφορετικό βαθμό (προερχόμενο από ένα μεγάλο εύρος) άσκησης επίδρασης στους Bloggers.

Η **MEIBI ορίζεται** ως ακολούθως (συνυπολογίζοντας τη τιμή του σκορ που προαναφέρθηκε): «Ένας Blogger με αύξοντα αριθμό  $j$  (στη σειρά) έχει δείκτη της MEIBI ίσο με  $m$ , εάν  $m$  (στο πλήθος) από τις  $BP(j)$  (στο πλήθος) συνολικά δημοσιεύσεις έχουνε συγκεντρώσει σκορ  $S_j^m(i) \geq m$  η καθεμία και οι υπόλοιπες  $BP(j)-m$  (στο πλήθος) έχουνε λάβει σκορ  $S_j^m(i) \leq m$ ».

## ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ

- [1] D. Gruhl, David Liben-Nowell, R. Guha, and A. Tomkins, “*Information diffusion through blogspace*”, SIGKDD Exploration Newsletter, 6(2):43–52, 2004.
- [2] Alvin Chin and Mark Chignell, “*A social hypertext model for finding community in blogs*”, in HYPERTEXT ’06: Proceedings of the seventeenth conference on Hypertext and hypermedia, pages 11–22, New York, NY, USA, 2006, ACM Press.
- [3] E. Keller and J. Berry, “*One American in ten tells the other nine how to vote, where to eat and, what to buy. They are The Influentials*”, The Free Press, 2003.
- [4] ” Identifying Influential Bloggers: Time Does Matter “ Leonidas Akritidis, Dimitrios Katsaros, Panayiotis Bozanis - Department of Computer & Communication Engineering, University of Thessaly Volos, Greece

## 4<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Μετρική ΜΕΙΒΙΧ

Πώς μπορούμε να συμπεράνουμε ότι ένα παλιό χρονολογικά post έχει ακόμη την ιδιότητα του ατόμου που εξασκεί επηρεασμό; Μόνο εάν εξετάσουμε την ηλικία των εισερχομένων συνδέσμων σε αυτό το post. Εάν ένα post δεν αναφέρεται πλέον, είναι μία ένδειξη ότι διαπραγματεύεται θέματα παρωχημένα ή προτείνει ξεπερασμένες λύσεις. Από την άλλη πλευρά, αν ένα παλιό post εξακολουθεί να συνδέεται με το παρόν, τότε αυτό είναι μια ένδειξη ότι περιέχει περιεχόμενο που ασκεί «επίδραση».

### 4.1 Υπολογισμός της ΜΕΙΒΙΧ

Με βάση τις ιδέες που αναπτύχθηκαν για τη μετρική ΜΕΙΒΙ, χρησιμοποιούμε έναν ανάλογο τρόπο προσέγγισης της μετρικής ΜΕΙΒΙΧ. Αντί να αντιστοιχούμε στα παλιά posts ενός blogger μικρότερα σκορ ανάλογα με την ηλικία τους, μπορούμε να αντιστοιχίσουμε σε κάθε εισερχόμενη σύνδεση ενός blogger post ένα μικρότερο βάρος που εξαρτάται από την ηλικία του συνδέσμου. Το παραπάνω εκφράζεται σε μορφή «ποσότητας» με την εξίσωση:

$$S_j^x(i) = \gamma(|C(i)| + 1) \sum_{\forall x \in R_j(i)} (\Delta T(P(x) + 1)^{-\delta}$$

Με τη βοήθεια αυτής της εξίσωσης ο **ορισμός της ΜΕΙΒΙΧ** διαμορφώνεται ως εξής: «Ο blogger j έχει τον δείκτη ΜΕΙΒΙΧ ίσο με «x», εάν «x» από τις συνολικά BP(j) (σε πλήθος) δημοσιευμένες του καταχωρήσεις λαμβάνουν σκορ  $S_j^x(i) > x$  έκαστη, ενώ οι υπόλοιπες BP(j) – x συγκεντρώνουν το σκορ  $S_j^x(i) \leq x$ » .

### 4.2 Σύγκριση Μετρικών ΜΕΙΒΙ, ΜΕΙΒΙΧ με εκείνες που Βασίζονται σε Ιδιοδιανύσματα

Η σύγκριση των μετρικών ΜΕΙΒΙ, ΜΕΙΒΙΧ με εκείνες που υπολογίζουν σκορ επιρροής στηριζόμενες στα ιδιοδιανύσματα (δηλαδή οι στηριζόμενες στον «δείκτη



τύπου  $h$ » και οι διαφοροποιήσεις τους [1]) οδηγεί στο συμπέρασμα ότι οι τελευταίες, περιλαμβάνουν σύνθετους υπολογισμούς και παρουσιάζουν πρόβλημα σταθερότητας. Επιπλέον, οι μετρικές MEIBI και MEIBIX δημιουργούν απλή πολιτική για την μέτρηση της επίδρασης των δημοσιευμένων αναρτήσεων και των blogs. Ακόμη, οι υπολογισμοί που πραγματοποιούνται σε αυτές τις μεθόδους είναι σύντομοι καθώς δεν λαμβάνονται υπόψη σύνθετοι αναδρομικοί ορισμοί του επηρεασμού και δε χρησιμοποιούνται παράμετροι που να συντονίζονται (tunable parameters) με κάποιον τρόπο. Τέτοιου είδους παράμετροι είναι δύσκολο να τεθούν. Τα αποτελέσματα που προκύπτουν δείχνουν σαφέστερα πως η MEIBIX, όπως και η MEIBI, είναι σε θέση να προσδιορίσουν καλύτερα τα σημαντικά χρονικά σχέδια (πλαίσια) στη «συμπεριφορά» όλων των ιστολογίων. Τόσο η MEIBI όσο και η MEIBIX που προτάθηκαν από τους Keller και Berry, μπορούν να χρησιμοποιηθούν σύμφωνα και με τις τρέχουσες εξελίξεις για την αναγνώριση των bloggers που ασκούν επηρεασμό στο «τώρα».

## ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ

- [1] Wikipedia, *The Hirsch h-index*, Jan. 2009, Available at <http://en.wikipedia.org/wiki/H-index>
- [2] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “*Identifying the influential bloggers in a community*”, in Proceedings of ACM WSDM Conf., pages 207–218, 2008.
- [3] E. Keller and J. Berry, “*One American in ten tells the other nine how to vote, where to eat and, what to buy. They are The Influentials*”, The Free Press, 2003.
- [4] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “*Identifying the influential bloggers in a community*”, in Proceedings of ACM WSDM Conf., pages 207–218, 2008.
- [5] Alvin Chin and Mark Chignell, “*A social hypertext model for finding community in blogs*”, In HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia, pages 11–22, New York, NY, USA, 2006. ACM Press.

## 5<sup>ο</sup> ΚΕΦΑΛΑΙΟ – Πειραματική Αξιολόγηση Μετρικών

Στο παρόν κεφάλαιο παρουσιάζεται μία εκτενής περιγραφή των πειραμάτων που διεξήχθησαν με σκοπό αφενός να ασκηθεί κριτική και αφετέρου να αξιολογηθούν και παράλληλα να συγκριθούν οι μετρικές και οι μέθοδοι που σχετίζονται με την αναγνώριση των influential bloggers.

### 5.1 Κοινωνία TUAW

Υφίστανται εκατομμύρια ιστολόγια στο διαδίκτυο. Σύμφωνα με δεδομένα της μηχανής αναζήτησης Technorati τα blogs ξεπερνούν τα 115 εκατομμύρια. Φυσικά επειδή είναι αδύνατο να γίνει αναζήτηση σε όλη τη μπλογκόσφαιρα για να ανακτήσουμε ένα επαρκές σύνολο δεδομένων, κρίνεται αναγκαία η εύρεση μιας ενεργής blog κοινωνίας η οποία θα παρέχει

- Αναγνώριση των bloggers (identification)
- Πληροφορίες σχετικές με την ημερομηνία και την ώρα κάποιας εγγραφής, το πλήθος των σχολίων και τους εξερχόμενους συνδέσμους προς ιστολόγιο/ιστολόγια.

Η TUAW (The Unofficial Apple Weblog) αποτελεί μία κοινωνία που ανταποκρίνεται σε όλες τις παραπάνω απαιτήσεις.

Σύμφωνα με μετρήσεις που υλοποιήθηκαν στη Technorati, κάθε δευτερόλεπτο περίπου 18.6 νέες δημοσιοποιημένες καταχωρήσεις αναρτώνται στη «μπλογκόσφαιρα», δηλαδή αυτός ο ρυθμός ανάρτησης είναι αρκετά μεγάλος. Το μέγεθος της “Blogosphere” αυξάνεται περίπου κατά 100% κάθε 6 μήνες και συνυπολογίζοντας αυτό νέες τεχνικές πρέπει να αναπτυχθούν στη “Blogosphere” ώστε να διατηρείται ο έλεγχος για οτιδήποτε συμβαίνει σε αυτή. Οι μετρικές και μέθοδοι που θα κριθούν στις επόμενες ενότητες, αφορούν σε «κοινωνίες» των blogs που μοιάζουν αρκετά με τα χαρακτηριστικά των blogs που περιλαμβάνονται στη «κοινωνία» της TUAW. Παρακάτω απεικονίζεται η ιστοσελίδα της TUAW.



Εικόνα 2: Ιστοσελίδα της TUAW

Από τη «κοινωνία» ιστολογίων της TUAW προέκυψαν 16000 περίπου κυριότεροι και εγκυρότεροι διαδικτυακοί τόποι τύπου Blog που ξεχώριζαν για 17831 δημοσιευμένες εγγραφές τους οι οποίες ήταν δημιουργημένες από 51 σημαντικούς bloggers. Το 90% του συγκεκριμένου συνόλου δημοσιευμένων καταχωρήσεων δέχθηκε σχολιασμό από χρήστες (οι οποίοι είναι επισκέπτες, σε αυτούς δε συμπεριλαμβάνονται οι Bloggers) και το 10% όχι. Κάνοντας χρήση του API<sup>2</sup> της Technorati, επιτυγχάνεται η συλλογή των εισερχόμενων συνδέσμων προς κάθε δημοσιευμένη εγγραφή, μαζί με την ημερομηνία δημιουργίας των συνδέσμων αυτών και τα ονόματα των συγγραφέων τους [6]. Για το λόγο αυτό, το API έχει μεγάλη σημασία για τις μεθόδους MEIBI και MEIBIX. Από τις 17831 δημοσιεύσεις, οι 4586 ήτανεικονογραφημένες από εισερχόμενους συνδέσμους.

Η κατανομή χρόνου για τις δημοσιευμένες καταχωρήσεις ιστολογίων και τους εισερχόμενους προς αυτές συνδέσμους, φαίνεται στον ακόλουθο πίνακα.

<sup>2</sup> <http://technorati.com/developers/api/cosmos.html>

Έτος	Δημοσιεύσεις	Δημοσιεύσεις που δέχθηκαν εισερχόμενους Συνδέσμους	Εισερχόμενοι Σύνδεσμοι
2008	3676	3653	53204
2007	4497	662	259
2006	4354	186	18
2005	4307	77	1
2004	997	8	0
Σύνολο	17831	4386	53575

**Πίνακας 4:** Δημοσιεύσεις και Εισερχόμενοι Σύνδεσμοι ανά Έτος

Προκειμένου να γίνει η κατηγοριοποίηση των bloggers που εξασκούν επιρροή, η έννοια χρόνος είναι σημαντικό γνώρισμα στις αντίστοιχες χρησιμοποιούμενες μετρικές και αυτό προκύπτει από τον **Πίνακα 5** και τα όσα αναφέρονται πριν από αυτόν. Η χρονική κατανομή των εισερχόμενων συνδέσμων μιας δημοσιευμένης εγγραφής μετριέται μέσω του υπολογισμού ενδιάμεσου χρονικού διαστήματος (ηλικίας) από τη χρονική στιγμή που τοποθετήθηκε μια δημοσιευμένη εγγραφή σε Blog έως και τη στιγμή όπου δημιουργήθηκε τουλάχιστον ένας εισερχόμενος σύνδεσμος προς αυτή. Τα αποτελέσματα απεικονίζονται παρατεταγμένα στον **Πίνακα 5**. Το βασικότερο που προκύπτει από τη μελέτη του πίνακα αυτού είναι ότι περίπου το μισό από το συνολικό πλήθος των εισερχόμενων συνδέσμων, δημιουργήθηκε την ίδια ημέρα με τη μέρα τοποθέτησης της δημοσιευμένης καταχώρισης σε Blog. Έτσι λοιπόν, όσο μικρότερο είναι το χρονικό διάστημα που μεσολαβεί μεταξύ της χρονικής στιγμής που τοποθετήθηκε μια δημοσιευμένη εγγραφή σε Blog έως και τη στιγμή όπου δημιουργήθηκε τουλάχιστον ένας εισερχόμενος σύνδεσμος προς αυτή, τόσο αυξάνεται και ο βαθμός ένδειξης της επίδρασης που εξασκεί εκείνος/εκείνη που έχει το ρόλο του σχολιαστή/δημιουργού δημοσιευμένης καταχώρισης.

Ηλικία	Εισερχόμενοι Σύνδεσμοι	Ποσοστό
0 ημέρες	26346	49,2%
1 ημέρα	13470	25,1%
Μεταξύ μίας και επτά ημερών	6653	12,4%
Μεταξύ επτά και τριάντα ημερών	2406	4,5%
Μεταξύ τριάντα και εξήντα ημερών	928	1,7%

Μεταξύ εξήντα και τριακοσίων εξήντα πέντε ημερών	2523	4,7%
Άνω των τριακοσίων εξήντα πέντε ημερών	1249	2,3%
Συνολικά	53575	99,9%

Πίνακας 5: Χρονική Κατανομή Εισερχόμενων Συνδέσμων

## 5.2 Προσδιορισμός του Πειράματος

Εκτός από τις μεθόδους αναγνώρισης των influential bloggers οι οποίες αναφέρθηκαν στα κεφάλαια 2,3,4 (Influence-flow [1], i-index, MEIBI, MEIBIX, interest-vector) θα εξετάσουμε επιπλέον δύο νέες «απλές» (plain) μεθόδους. Η πρώτη, γνωστή ως δείκτης δραστηριότητας “activity index” κατατάσσει τους bloggers χρησιμοποιώντας μόνο τη δραστηριότητα τους, δηλαδή τον αριθμό των δημοσιευμένων posts. Η δεύτερη αποτελεί μία απευθείας υιοθέτηση της μεθόδου h-index [2].

Χωρίζουμε το πείραμα που πραγματοποιείται σε 3 μέρη.

- Στο **πρώτο μέρος** συγκρίνουμε τους influential bloggers που υποδεικνύονται από τις μεθόδους MEIBI και MEIBIX («νέες μέθοδοι») με τους bloggers που βρίσκονται μέσω των «απλών» μεθόδων. Χρησιμοποιούμε το πλήρες σύνολο δεδομένων σαν μια βάση του πειράματος για την εξέταση του εάν οι χρονικές παράμετροι αξίζουν να ληφθούν υπόψη.
- Στο **δεύτερο μέρος** συγκρίνουμε τους influential bloggers που υποδεικνύονται από τις μεθόδους MEIBI και MEIBIX με αυτούς που προέκυψαν από τη μέθοδο Influence-flow, κάνοντας χρήση των blogs που δημοσιεύτηκαν το Νοέμβριο του 2008, για να αποδείξουμε ότι ακόμη και για μικρά χρονικά μεσοδιαστήματα οι κατατάξεις είναι διαφορετικές.
- Στο **τρίτο μέρος** εξετάζουμε τη χρονική εξέλιξη των influential bloggers, η οποία προσδιορίζεται από τις μετρικές MEIBI και MEIBIX που υλοποιήθηκαν κατά τη διάρκεια του 2008, ώστε να γίνει εξέταση του εάν οι influential

bloggers χάνουν την κυριαρχία τους στην εξάσκηση επιρροής και εάν ενισχύεται ακόμη περισσότερο η ανάγκη του να ληφθούν υπόψη οι χρονικές παράμετροι.

## 5.3 Πειραματικά Αποτελέσματα

Στην παρούσα ενότητα παρουσιάζονται διεξοδικά τα αποτελέσματα και τα συμπεράσματα των τριών φάσεων της πειραματικής μας μελέτης.

### 5.3.1 Πρώτο Μέρος Πειραματικών Αποτελεσμάτων

Ξεκινώντας την παρουσίαση των αποτελεσμάτων του πρώτου μέρους της πειραματικής μας μελέτης, στους πίνακες 6 και 7 φαίνεται η κατάταξη των Bloggers με βάση τις «απλές» μεθόδους.

	Δημιουργοί ιστολόγιου/ιστολόγιων	N	Πρώτη	Τελευταία
1	S. McNulty	3037	6/1/2005	31/7/2008
2	D. Caolo	2242	7/6/2005	4/12/2008
3	D. Chartier	1835	26/8/2005	30/8/2007
4	E. Sadun	1560	9/11/2006	26/9/2008
5	C.K. Sample III	1057	1/3/2005	5/6/2006
6	M. Lu	1043	13/12/2006	4/12/2008
7	L. Duncan	954	19/9/2004	23/1/2007
8	C. Bohon	793	24/2/2004	4/12/2008
9	M. Rose	793	29/11/2006	5/12/2008
10	M. Schramm	648	7/6/2007	4/12/2008

**Πίνακας 6:** Κατάταξη Bloggers σύμφωνα με τον activity index

	Bloggers	h	Posts	Cited	Inlinks
1	E. Sadun	31	1560	489	5759
2	C. Bohon	29	793	676	9439
3	M. Schramm	25	648	339	4322
4	R. Palmer	25	354	354	4809
5	M. Rose	24	793	364	4222
6	D. Caolo	23	2242	459	4907
7	M. Lu	23	1043	397	4282
8	S. McNulty	23	3037	334	3212
9	B. Terpstra	22	226	223	3013
10	C. Warren	22	133	112	1605

**Πίνακας 7:** Κατάταξη Bloggers σύμφωνα με τον h-index

Στον Πίνακα 6, εκείνοι που είναι σχολιαστές/δημιουργοί δημοσιευμένης καταχώρισης/καταχωρήσεων **διατάσσονται σύμφωνα με το δείκτη δραστηριότητάς τους**, δηλαδή σε μια φθίνουσα σειρά με βάση το πλήθος των δημοσιοποιημένων καταχωρήσεων που συνέγραψαν και είχαν αναρτήσει οι bloggers σε ιστολόγιο/ιστολόγια της «κοινωνίας» TUAW. Αυτοί είναι «ενεργοί» Bloggers. Οι ημερομηνίες της 1<sup>ης</sup> και τελευταίας δημοσιευμένης εγγραφής φαίνονται στις δύο τελευταίες στήλες του πίνακα αντίστοιχα.

Στον Πίνακα 7, οι Bloggers **διατάσσονται σύμφωνα τη μετρική του «δείκτη τύπου h»**[2]. Η κατάταξή τους γίνεται με φθίνουσα σειρά (με βάση τη τιμή του δείκτη «h» τύπου). Στις τρεις τελευταίες στήλες του πίνακα, φαίνονται αντίστοιχα: *το πλήθος όλων των posts του κάθε blogger που αναρτήθηκε σε Blog/Blogs της «κοινωνίας» TUAW, το πλήθος δημοσιευμένων εγγραφών (που αποτελεί υποσύνολο όλων των τοποθετημένων posts στη «κοινωνία» TUAW του συγκεκριμένου δημιουργού) οι οποίες έχουνε παραπεμφθεί από άλλες δημοσιευμένες εγγραφές και το πλήθος των εισερχόμενων προς κάθε δημοσιευμένη καταχώριση συνδέσμων.*

	<b>Bloggers</b>	$m$	$C_j$
1	C. Bohon	49	14745
2	R. Palmer	46	9916
3	S. Sande	36	7246
4	E. Sadun	34	32432
5	M. Rose	30	13499
6	M. Schramm	30	12838
7	C. Warren	28	4857
8	D. Caolo	27	27985
9	M. Lu	25	17966
10	B. Terpstra	17	3770

**Πίνακας 8:** Κατάταξη Bloggers σύμφωνα με την MEIBI

Ο Πίνακας 8, περιλαμβάνει την **διάταξη** των bloggers **σύμφωνα με τη μετρική MEIBI**. Σε αυτόν, στη 2<sup>η</sup> στήλη βρίσκονται τιμές που δείχνουν το πλήθος δημοσιευμένων καταχωρήσεων (το οποίο είναι υποσύνολο του ολικού πλήθους δημοσιευμένων εγγραφών των Bloggers) και στη 3<sup>η</sup> στήλη τα πλήθη των σχολίων που έχουν συγγραφεί από τους χρήστες (οι οποίοι είναι επισκέπτες, σε αυτούς δε συμπεριλαμβάνονται εκείνοι οι οποίοι είναι δημιουργοί δημοσιοποιημένης

εγγραφής/εγγραφών σε blog/blogs) για τις αντίστοιχες «m» (στο πλήθος) δημοσιεύσεις του συγκεκριμένου blogger.

Κάνοντας τη σύγκριση μεταξύ των πινάκων 6 και 7, προκύπτουν ορισμένες διαφορές. Αυτές οι διαφορές δικαιολογούν πως οι έννοιες παραγωγικότητα και επίδραση δε συμπίπτουν. Για παράδειγμα, η S.McNulty δημιουργός θα είναι 8<sup>η</sup> στη τελική διάταξη, όταν η διαδικασία ταξινόμησης γίνεται με φθίνουσα σειρά με χρήση της τεχνικής «δείκτης τύπου h». Σύμφωνα με τη μετρική που χρησιμοποιεί τον δείκτη του τύπου που επισημάνθηκε πριν, τη μεγαλύτερη επιρροή εξασκεί η blogger “E. Sadun” η οποία έχει 31 άρθρα σε καθένα από τα οποία υπάρχουν τουλάχιστον 31 εισερχόμενες συνδέσεις. **Η E. Sadun είναι η τέταρτη πιο «ενεργή» σχολιαστής/δημιουργός δημοσιευμένης καταχώρισης/καταχωρήσεων σε blogs/blog στη TUAW, αν και δε δημοσιεύτηκε κάτι από αυτή στους τελευταίους δύομισι μήνες. Αν και είναι «ανεργή» πρόσφατα, αυτή εξακολουθεί να εξασκεί τον μεγαλύτερο επηρεασμό σύμφωνα με τη μετρική του «δείκτη τύπου h». Άρα από το παράδειγμα αυτό προκύπτει ότι *μειονέκτημα* του «δείκτη τύπου h», είναι ότι *δεν καθιστά εφικτό να βρίσκει τους Bloggers οι οποίοι τόσο εξασκούν επηρεασμό όσο και είναι «ενεργοί».***

Από τον πίνακα 8 είναι φανερό πως στην αναπαριστώμενη κατάταξη, προηγούνται εκείνοι οι bloggers των οποίων οι δημοσιεύσεις άσκησαν επίδραση σε χρονική στιγμή όσο το δυνατόν πιο πρόσφατη.

Υπάρχει μια παρατήρηση που φανερώνει τη διαφορά που υπάρχει ανάμεσα στις μετρικές που περιλαμβάνουν αντίστοιχα:

- ✓ τη χρήση «δείκτη τύπου h»
- ✓ τη MEIBI

Σύμφωνα με τη παρατήρηση αυτή, στον πίνακα 8 προκύπτει το ότι η μετρική MEIBI μας φανερώνει όχι μόνο τους Bloggers που χαρακτηρίζονται από το μεγαλύτερο βαθμό ως προς την επίδραση που ασκούν αλλά και τους πιο «ενεργούς» από τους σχολιαστές/δημιουργούς δημοσιευμένης καταχώρισης/καταχωρήσεων. Είναι μία μετρική από τις καταλληλότερες για τη περίπτωση της «Μπλογκόσφαιρας», καθώς η τελευταία χαρακτηρίζεται από ραγδαίες αλλαγές και η MEIBI καταφέρνει να



συμπορεύεται με τις αλλαγές κυρίως εξαιτίας της ικανότητάς της να διαχειρίζεται σωστά την ημερομηνία τοποθέτησης των δημοσιευμένων καταχωρήσεων στα blogs και τα σχόλια που αναλογούν σε κάθε δημοσιοποιημένη εγγραφή.

Ο Πίνακας 9 μας δίνει μια κατάταξη των bloggers με κριτήριο τα αποτελέσματα της MEIBIX μετρικής. Η στήλη του πίνακα με όνομα «x», περιλαμβάνει τις τιμές του δείκτη «x» (της MEIBIX). Προφανώς υπάρχουν διαφορές μεταξύ της τελικής διάταξης των bloggers κάποιου Blog που προκύπτει από τον πίνακα 6, τον 7 και τον 9 και αυτό είναι προφανές από τη σειρά που καταλαμβάνει ο blogger *S. McNulty* στη διάταξη του πίνακα 6 (πρώτος), στην αντίστοιχη του 7 πίνακα (όγδοος) και στην αντίστοιχη του πίνακα 9 (δεν υπάρχει στη δεκάδα). Επίσης, υπάρχουν και διαφορές στη θέση στη τελική διάταξη για τουλάχιστον έναν/μία Blogger μεταξύ των πινάκων 7 και 9.

	<b>Bloggers</b>	<i>x</i>
1	C. Bohon	48
2	R. Palmer	47
3	S. Sande	37
4	E. Sadun	33
5	C. Warren	30
6	M. Rose	29
7	M. Schramm	27
8	M. Lu	26
9	D. Caolo	25
10	B. Terpstra	15

**Πίνακας 9:** Κατάταξη Bloggers σύμφωνα με την MEIBIX

Με τη χρήση της μετρικής “Spearman’s rho”, υπολογίστηκε ο βαθμός συσχέτισης μεταξύ των μετρικών του «δείκτη τύπου h», της MEIBIX και της MEIBI. Προκύπτει πως τα αποτελέσματα της τελικής κατάταξης που δημιουργούνται από τη MEIBIX και τη MEIBI μοιάζουν μεταξύ τους, όμως αυτά αποκλίνουν από τα αποτελέσματα της τελικής διάταξης που μας δίνει η μετρική του «δείκτη τύπου h».

Στον Πίνακα 10 υπάρχουν τα αποτελέσματα που προκύπτουν από τη χρήση της μετρικής “Spearman’s rho”. Στη δεύτερη στήλη βρίσκονται οι τιμές του δείκτη της συγκεκριμένης μετρικής.

Methods	$\rho$
h-index – MEIBI	0.478788
h-index – MEIBIX	0.321212
MEIBI – MEIBIX	0.951515

Πίνακας 10: Αποτελέσματα Μετρικής Spearman's rho

### 5.3.2 Δεύτερο Μέρος Πειραματικών Αποτελεσμάτων

Από τη πρώτη φάση της πειραματικής μελέτης έγινε φανερό ότι η μη ύπαρξη δραστηριότητας έχει ένα σημαντικό αντίκτυπο στη τελική κατάταξη των bloggers. Το πραγματικό ζήτημα που σχετίζεται με την χρησιμότητα των μεθόδων MEIBI και MEIBIX είναι εάν σε ένα μικρό χρονικό διάστημα, για παράδειγμα σε ένα μήνα, αυτές οι μέθοδοι θα μπορούσαν να δώσουν διαφορετική κατάταξη σε σχέση με την κατάταξη της influence-flow. Έτσι, επιλέχθηκαν μόνο εκείνα τα posts που είχαν δημοσιευτεί το Νοέμβριο του 2008.

Στον ομαδοποιημένο Πίνακα 11 παριστάνονται οι κορυφαίοι 10 «ενεργοί» bloggers κάποιου ιστολογίου δίνοντας βάση αντίστοιχα:

- a) στον δείκτη δραστηριότητάς τους κατά τη διάρκεια του Νοεμβρίου του 2008, καθώς αυτή η ταξινόμηση παρέχεται από τον ιστότοπο της «κοινωνίας» TUAW.
- b) στη μετρική «ροής – επιρροής» (influence-flow).
- c) στη MEIBI.
- d) στη MEIBIX.

Καμία από τις μετρικές MEIBI, MEIBIX δε παράγει προκύπτουσα τελική διάταξη που να συμφωνεί με το αποτέλεσμα της ταξινόμησης η οποία πραγματοποιείται από τη «κοινωνία» TUAW για τους Bloggers.

## 5<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Πειραματική Αξιολόγηση Μετρικών

	Bloggers	$N$	Inlinks	$C_j$
1	C. Bohon	47	508	556
2	R. Palmer	42	339	491
3	S. Sande	34	354	177
4	M. Schramm	29	203	166
5	D. Caolo	20	163	178
6	M. Rose	19	138	154
7	B. Terpstra	15	103	87
8	C. Warren	8	80	331
9	M. Lu	8	71	248
10	V. Agreda	5	30	42

	Blogger
1	C. Bohon
2	R. Palmer
3	M. Lu
4	C. Warren
5	D. Caolo
6	C. Ullrich
7	S. Sande
8	M. Rose
9	V. Agreda
10	Jason Clarke

	Blogger	$m$
1	C. Bohon	26
2	R. Palmer	20
3	S. Sande	20
4	D. Caolo	17
5	M. Schramm	16
6	M. Rose	13
7	M. Lu	8
8	B. Terpstra	7
9	C. Warren	7
10	V. Agreda	4

	Blogger	$x$
1	C. Bohon	27
2	S. Sande	20
3	R. Palmer	19
4	D. Caolo	18
5	M. Schramm	16
6	M. Rose	13
7	M. Lu	8
8	B. Terpstra	7
9	C. Warren	7
10	V. Agreda	4

**Πίνακας 11:** 10 Κορυφαίοι Ενεργοί Bloggers

Αναλύοντας τον πίνακα 11, πρέπει να σημειωθεί πως: Η κατάταξη των bloggers στο πρώτο από αριστερά πινακάκι, προέρχεται από τη «κοινωνία» TUAW. Υπενθυμίζεται πως  $N$  είναι το ολικό πλήθος των δημοσιοποιημένων εγγραφών στη TUAW και  $C_j$  είναι το πλήθος των σχολίων που έλαβε ο blogger της συγκεκριμένης( $j^{th}$ ) σειράς του πίνακα. Η διάταξη των Bloggers στον δεύτερο υποπίνακα οφείλεται στο μοντέλο αξιολόγησης «ροής-επιρροής». Η ταξινόμηση των bloggers στον επόμενο πίνακα οφείλεται στη μέθοδο MEIBI. Ο  $m$  είναι δείκτης ο ρόλος του οποίου είχε αναφερθεί όταν ορίστηκε και παρουσιάστηκε η MEIBI. Η κατάταξη των Bloggers στον τελευταίο υποπίνακα, προέρχεται από τη MEIBIX μέθοδο. Ο « $x$ » είναι δείκτης ο ρόλος του οποίου είχε αναφερθεί όταν ορίστηκε και παρουσιάστηκε η MEIBIX.

Από τον πίνακα 11 συμπεραίνεται ότι η *σειρά κατάταξης* συγκεκριμένων ονομάτων bloggers στη τελική διάταξη, διαφέρει ανάλογα με το εάν θεωρούμε τη TUAW ή τις μεθόδους influence-flow, MEIBI, MEIBIX. Έτσι, παρόλο που τα posts κάποιου blogger μπορεί να προσελκύσουν λιγότερα σχόλια και λιγότερες inlinks σε σχέση με κάποιον άλλον blogger, αυτός σύμφωνα με τη μέθοδο influence-flow κατατάσσεται σε πιο υψηλή θέση (παράδειγμα Caolo, Sande). Ακόμη, τουλάχιστον ένας συγκεκριμένος blogger (παράδειγμα Lu) παρότι ενδέχεται να έχει δημοσιοποιήσει σχετικά μικρό πλήθος καταχωρίσεων στη «κοινωνία» TUAW, έχει μεγάλο βαθμό άσκησης επίδρασης όταν έχει λάβει ένα πλήθος από σχόλια που είναι από τα μεγαλύτερα. Τέλος, προκύπτει ότι η μετρική MEIBIX είναι πιο «ευαίσθητη» στο πλήθος των εισερχόμενων συνδέσμων προς δημοσιευμένες εγγραφές σε σύγκριση με τη μετρική MEIBI (παράδειγμα κατάταξης Sande).

### 5.3.3 Τρίτο Μέρος Πειραματικών Αποτελεσμάτων

Έχει ενδιαφέρον να εξεταστεί το πώς οι ταξινομήσεις που πραγματοποιούνται από τις μεθόδους MEIBI και MEIBIX ποικίλουν σε ένα εύρος χρόνου. Στις εικόνες 3 και 4 φαίνονται οι ταξινομήσεις ως προς την επιρροή για τους δέκα κορυφαίους ως προς τον επηρεασμό που εξασκούν bloggers (για τους αναγραφόμενους μήνες στις στήλες του πίνακα) όταν εφαρμόστηκαν οι μέθοδοι MEIBI, MEIBIX αντίστοιχα. Αν συμβολιστούν με (i,j) τα κελιά των εικόνων 3 και 4 τότε ένα τέτοιο κελί αποθηκεύει τη σειρά του i-blogger στη θέση j του χρονικού παραθύρου. Όπου υπάρχει το σύμβολο -, εννοείται πως το συγκεκριμένο άτομο Blogger δεν ήτανε μεταξύ των 10 κορυφαίων για τη συγκεκριμένη χρονική περίοδο.

	Jan 2008	Feb 2008	Mar 2008	Apr 2008	May 2008	Jun 2008	Jul 2008	Aug 2008	Sep 2008	Oct 2008	Nov 2008
Erica Sadun	1	2	1	2	1	4	3	4	2	-	-
Scott McNulty	2	10	8	6	6	3	4	-	-	-	-
Cory Bohon	3	1	2	1	2	1	2	1	3	2	1
Dave Carlo	4	8	5	3	5	5	6	5	6	7	4
Mike Schramm	5	4	4	9	9	8	7	6	5	5	5
Brett Terpstra	6	5	7	7	8	-	-	7	8	9	8
Christina Warren	7	6	-	8	-	7	9	-	7	8	9
Mat Lu	8	3	6	4	3	6	8	8	9	6	7
Michael Rose	9	7	3	5	-	-	-	9	10	3	6
Nik Fletcher	10	9	9	10	-	-	-	-	-	-	-
Chris Ulrich	-	-	10	-	-	-	-	-	-	-	-
Robert Palmer	-	-	-	-	4	2	1	2	1	1	2
Steven Sande	-	-	-	-	7	9	5	3	4	4	3
Joshua Ellis	-	-	-	-	10	10	-	-	-	-	-
Gilles Turnbull	-	-	-	-	-	-	10	10	-	-	-
Victor Agreda, Jr.	-	-	-	-	-	-	-	-	-	10	10

Εικόνα 3: Ταξινομήσεις Κορυφαίων Bloggers κατά MEIBI

	Jan 2008	Feb 2008	Mar 2008	Apr 2008	May 2008	Jun 2008	Jul 2008	Aug 2008	Sep 2008	Oct 2008	Nov 2008
Erica Sadun	1	2	1	3	1	4	3	4	2	-	-
Scott McNulty	2	-	8	6	5	3	4	-	-	-	-
Cory Bohon	3	1	2	1	2	1	2	1	3	2	1
Dave Caolo	4	7	4	2	7	5	6	5	7	7	4
Brett Terpstra	5	5	7	7	6	-	-	7	9	9	7
Christina	6	6	-	8	-	6	9	-	6	8	9
Mat Lu	7	4	5	5	3	7	7	8	10	6	8
Michael Rose	8	8	3	4	-	-	-	9	8	3	6
Mike Schramm	9	3	6	9	8	8	8	6	4	5	5
Nik Fletcher	10	9	9	10	-	-	-	-	-	-	-
Chris Ulrich	-	10	10	-	-	-	-	-	-	-	-
Robert Palmer	-	-	-	-	4	2	1	2	1	1	2
Steven Sande	-	-	-	-	9	9	5	3	5	4	3
Joshua Ellis	-	-	-	-	10	10	-	-	-	-	-
Giles Turnbull	-	-	-	-	-	-	10	10	-	-	-
Victor Agreda	-	-	-	-	-	-	-	-	-	10	10

Εικόνα 4: Ταξινομήσεις Κορυφαίων Bloggers κατά MEIBIX

Οι μέθοδοι MEIBI, MEIBIX «παράγουν» παρόμοιες τελικές κατατάξεις, με τη MEIBIX να είναι περισσότερο επηρεασμένη από το πλήθος των εισερχόμενων συνδέσμων, ενώ η MEIBI αναθέτει βελτιωμένες τιμές στα σκορ των δημοσιεύσεων που συγκέντρωσαν περισσότερα σχόλια. Για να γίνει διάκριση μεταξύ των bloggers που ασκούσαν επίδραση για ένα πολύ μεγάλο χρονικό διάστημα (bloggers τύπου 1) ή ένα πολύ μικρό χρονικό διάστημα (bloggers τύπου 2), ένα αξιόπιστο χρησιμοποιηθέν εργαλείο είναι η μελέτη (ως προς το χρόνο) της διακύμανσης στις παραγόμενες ταξινομήσεις των bloggers. Οι Bloggers του τύπου 1, θεωρούνται σημαντικότεροι ως προς το βαθμό της επίδρασης που ασκούν σε σύγκριση με εκείνους του τύπου 2, που θεωρούνται πιο αξιόπιστοι.

Φυσικά, είναι πολλές οι υπόλοιπες κατηγορίες των bloggers που μπορούν να δημιουργηθούν ως επακόλουθο της ανασκόπησης της δραστηριότητας αυτών των bloggers στο χρόνο και πολλές αντίστοιχες εφαρμογές μπορούν να αναπτυχθούν μέσω αυτών των κατηγοριών.

## ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ

- [1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “*Identifying the influential bloggers in a community*”, in Proceedings of ACM WSDM Conf., pages 207–218, 2008.
- [2] Wikipedia, *The Hirsch h-index*, Jan. 2009. Available at <http://en.wikipedia.org/wiki/H-index>.
- [3] N. Agarwal and H. Liu, “*Blogosphere: Research issues, tools and applications*”, ACM SIGKDD Explorations, 10(1):18–31, 2008.
- [4] “Identifying the Influential Bloggers in a Community” Nitin Agarwal, Huan Liu, Lei Tang, Arizona State University Tempe, AZ 85287, USA - Philip S. Yu, University of Illinois at Chicago, Chicago, IL 60607, USA
- [5] ” Identifying Influential Bloggers: Time Does Matter “ Leonidas Akritidis, Dimitrios Katsaros, Panayiotis Bozanis - Department of Computer & Communication Engineering, University of Thessaly Volos, Greece
- [6] Identifying the Influential Bloggers in a Community, Nitin Agarwal, Huan Liu, Lei Tang - Arizona State University Tempe, AZ 85287, USA

## 6<sup>ο</sup> ΚΕΦΑΛΑΙΟ – Πειραματική Αξιολόγηση του Μοντέλου I-index

Στα πλαίσια αυτού του κεφαλαίου διεξάγονται διάφορα πειράματα χρησιμοποιώντας τον δείκτη i-index (παράγραφος 2.1) και παράλληλα παρουσιάζονται τα συμπεράσματα που προκύπτουν από τα αποτελέσματα των πειραμάτων.

### 6.1 Πειραματική Μελέτη Τριών Ειδών bloggers

Ο δείκτης i-index χρησιμοποιείται για τον σχεδιασμό πειραμάτων και γίνεται έρευνα για κατάλληλα πειραματικά δεδομένα ώστε να βρεθούν όσοι ασκούν επίδραση και έχουν ρόλο σχολιαστή/δημιουργού δημοσιευμένης εγγραφής/εγγραφών σε ιστολόγιο. Μέσω της περιήγησης σε blogs τύπου «πραγματικού κόσμου», ελαχιστοποιούνται οι προσπάθειες για τον καθορισμό των τρόπων απόκτησης των αναγκαίων στατιστικών δεδομένων και μεγιστοποιείται η αναπαραγωγή των πειραμάτων που διενεργούνται. Όταν η περιήγηση αυτή έγινε στη **Tuaw**, από την έναρξη λειτουργίας της ως και τον Ιανουάριο του 2007 συγκεντρώθηκε ένα ορισμένο πλήθος δημοσιοποιημένων καταχωρήσεων και διαμορφώθηκε ένα *πλήρες ιστορικό* με δυνατότητα ανάκτησής του αργότερα με σχεσιακή βάση δεδομένων.

Γενικά, υπάρχουνε διαφοροποιήσεις ως προς τις ταξινομήσεις όλων των ειδών των Bloggers με βάση τα γνωρίσματα των δημοσιοποιημένων τους καταχωρήσεων ως προς τα χαρακτηριστικά: Bloggers «ενεργοί», Bloggers «που ασκούν επίδραση». Επιπλέον, στις κοινωνικές ιστοσελίδες τύπου Blog **δεν ισχύουν οι παρακάτω δύο συνεπαγωγές:**

- a) Κάποιος/κάποια που ασκεί επίδραση και έχει ρόλο σχολιαστή/δημιουργού δημοσιευμένης εγγραφής/εγγραφών σε ιστολόγιο είναι απαραίτητα και «ενεργό» άτομο.
- b) Ένας/μία Blogger που είναι «ενεργός»/«ενεργή» θα ασκεί απαραίτητα και επίδραση. Δηλαδή, το να είναι ανενεργό κάποιο άτομο δε συνεπάγεται απαραίτητα ότι δεν εξασκεί επηρεασμό. Αυτή η συνεπαγωγή και η ακριβώς αντίστροφη της αποδεικνύονται από τον **Πίνακα 13** της τρέχουσας ενότητας

## 6<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Πειραματική Αξιολόγηση του Μοντέλου I-index

και ειδικά από τα δεδομένα που περιλαμβάνει αυτός για τη κυρία “*Erica Sadun*” που είναι κορυφαία τόσο στην ταξινόμηση των κορυφαίων Bloggers ως προς το πόσο «ενεργοί» είναι όσο και στην αντίστοιχη των κορυφαίων ως προς την επίδραση που ασκούν.

Οι περιπτώσεις στις οποίες οι προαναφερθείσες συνεπαγωγές είναι ορθές προσωποποιούνται στα ονόματα των ατόμων του Πίνακα 12, τα οποία είναι με πλάγια γραφή.

Οι 5 κορυφαίοι «ενεργοί» Bloggers σύμφωνα με τη «κοινωνία» TUAW	Τα 5 κορυφαία ως προς την επιρροή που εξασκούν άτομα Bloggers
1 <sup>ο</sup> : <i>Erica Sadun</i> 2 <sup>ο</sup> : <i>Scott McNulty</i> 3 <sup>ο</sup> : <i>Mat Lu</i> 4 <sup>ο</sup> : <i>David Chartier</i> 5 <sup>ο</sup> : <i>Michael Rose</i>	1 <sup>ο</sup> : <i>Erica Sadun</i> 2 <sup>ο</sup> : <i>Dan Lurie</i> 3 <sup>ο</sup> : <i>David Chartier</i> 4 <sup>ο</sup> : <i>Scott McNulty</i> 5 <sup>ο</sup> : <i>Laurie A. Duncan</i>

Πίνακας 12: 5 Κορυφαίοι Bloggers

		Number of Comments		Number of Inlinks		Length of Blog Post		Number of Outlinks		Total Number of Blog Posts	Influential Blog Posts
		Max	Average	Max	Average	Max	Average	Max	Average		
Active + Influential	<i>Erica Sadun</i>	75	11.0197	80	10.1316	2935	830.0066	15	2.5329	152	9
	<i>David Chartier</i>	56	11.3088	32	10.25	3529	1054.912	14	4.3529	68	4
	<i>Scott McNulty</i>	112	11.5607	33	8.9252	2246	623.2991	12	2.5888	107	3
Inactive + Influential	<i>Dan Lurie</i>	96	19.6316	37	10.2632	1569	793.7368	4	2.3158	16	4
	<i>Laurie A. Duncan</i>	65	16.2895	34	10.6053	2888	993.8947	11	3.4737	26	2
Active + Non-Influential	<i>Mat Lu</i>	42	8.0294	29	10.0147	1699	771.1471	12	4.1029	73	0
	<i>Michael Rose</i>	31	8.7273	21	9.6061	1378	735.9848	15	6.1515	58	0

Πίνακας 13: Αποτελέσματα των Στατιστικών για τα 3 Είδη Bloggers

Ο Πίνακας 12 που στηρίζεται σε δεδομένα της TUAW των τελευταίων 30 ημερών αναλύεται ως εξής:

- στη 1<sup>η</sup> του στήλη περιλαμβάνεται η ταξινόμηση των ονομάτων των κορυφαίων Bloggers ως προς το πόσο «ενεργοί» είναι αυτοί.
- στη 2<sup>η</sup> στήλη του περιλαμβάνεται η ταξινόμηση των κορυφαίων ονομάτων των bloggers λαμβάνοντας υπόψη το πόσο σημαντική είναι η επίδραση που ασκούν αυτοί (δείκτης i-index). Για αυτούς, επιπλέον κοινό γνώρισμα είναι το ότι με τις δημοσιοποιημένες τους



καταχωρήσεις, έχουν συγκεντρώσει υψηλό πλήθος σχολίων. Παράλληλα, το χαρακτηριστικό της καινοτομίας και της ακρίβειας στη πληροφόρηση αυξάνουν το κατά πόσο μια δημοσιευμένη καταχώριση ασκεί επίδραση.

Στον **πίνακα 13**, εκτός από τις παρατηρήσεις που επισημάνθηκαν πριν, στις υπόλοιπες **στήλες (παραμέτρους)** περιλαμβάνονται στατιστικά τα οποία συσχετίζονται (δηλαδή ορισμένα από αυτά έχουν τιμές που μεταβάλλονται ανάλογα είτε αντιστρόφως ανάλογα) με το κατά πόσο οι bloggers που ανήκουν στα είδη: **1<sup>ο</sup>, 2<sup>ο</sup>, 4<sup>ο</sup> που επισημάνθηκαν στο 1<sup>ο</sup> κεφάλαιο (βλέπε ενότητα 1.2)** εξασκούν επίδραση. Έτσι, είναι προφανές το κατά πόσο διαφοροποιούνται οι Bloggers αυτών των ειδών. **Σχετικά με τα γνωρίσματα της 1<sup>ης</sup> και 2<sup>ης</sup> στήλης του πίνακα, όσο περισσότερα είναι αυτά, τόσο περισσότερες συζητήσεις έχουν προκαλέσει οι δημοσιευμένες εγγραφές όπου αφορούν και αυτό είναι προφανές για τους: “Erica Sadun”, “Laurie A. Duncan”.**

Η κυρία “Erica Sadun” είναι αντιπροσωπευτικό παράδειγμα του 1<sup>ου</sup> είδους (ενεργή και άτομο που εξασκεί επιρροή) των Bloggers που ασκούν επίδραση και για αυτή στο «παράθυρο» χρονικής διάρκειας 30 ημερών έχει σημασία το πλήθος των δημοσιευμένων εγγραφών της που ασκούν επίδραση. Η δημοσιοποιημένη της καταχώριση που είχε ασκήσει τη μέγιστη επιρροή αναφερόταν σχετικά με τη κεντρική ομιλία του «μεγάλου» αφεντικού της “Apple” “Steve Jobs”, για την εταιρεία. Συγκεκριμένα, η δημοσιοποιημένη της καταχώριση ήταν πρωτόπορος στα εξής ζητήματα:

- περιείχε όλες τις νέες υπηρεσίες και τα νέα προϊόντα που αρχικοποιούσε η “apple”.
- είχε ως χαρακτηριστικό τη λεπτομερή περιγραφή της.

Η ομιλία που περιλαμβανόταν σε εκείνη τη καταχώριση της Erica, προκάλεσε 63 σχόλια και δέχθηκε 80 παραπομπές. Στο ίδιο είδος bloggers όμως κάποιες θέσεις χαμηλότερα σε σχέση με τη κυρία Erica στη 2<sup>η</sup> στήλη, όπως φαίνεται και στον **Πίνακα 13**, ανήκει και το άτομο “Laurie A. Duncan” το οποίο με μια δημοσιοποιημένη του καταχώριση «τόνισε» λεπτομερώς το ζήτημα στο οποίο ο ιστότοπος του “MACZOT” παραβίασε με τον κατασκευαστή του “Xpad” τις

συμφωνίες άδειας χρήσης. Η μελέτη που αφορά στους “Erica Sadun”, “Laurie A. Duncan” καταδεικνύει την αποτελεσματικότητα του μοντέλου του δείκτη i-index.

**Όσον αφορά στο 2<sup>ο</sup> είδος (ενεργοί και όσοι δεν εξασκούν επηρεασμό) σχολιαστών/δημιουργών δημοσιευμένης καταχώρισης/καταχωρήσεων σε ιστολόγιο, αντιπροσωπευτικοί εδώ είναι οι “Mat Lu” και “Michael Rose”.** Αυτοί έχουν δημοσιοποιήσει συνολικό πλήθος εγγραφών που διακρίνεται στην εξής θέση του Πίνακα 13: αυτή υποδεικνύεται από τη σειρά (είδος όπου ανήκουν αυτοί) και τη κατάλληλη στήλη. Παράλληλα, ο καθένας από αυτούς χαρακτηρίζεται στο blog/blogs τους από υψηλό πλήθος εξερχόμενων συνδέσμων. Για αυτούς, στο «παράθυρο» χρονικής διάρκειας 30 ημερών έχει σημασία το πλήθος των δημοσιοποιημένων εγγραφών που ανάρτησαν και διαφέρει κατά πολύ από το αντίστοιχο για τους bloggers των υπόλοιπων κατηγοριών.

**Σχετικά με το 4<sup>ο</sup> είδος (μη ενεργοί και εκείνοι που εξασκούν επιρροή) των Bloggers, αντιπροσωπευτικό άτομο είναι εκείνο με την ονομασία “Dan Lurie”.** Αυτό έχει ένα ποσοστό 25% των δημοσιοποιημένων του εγγραφών που εξασκούν επηρεασμό (έχουν προκαλέσει έντονες συζητήσεις) και επίσης, στο «παράθυρο» χρονικής διάρκειας 30 ημερών είναι κρίσιμο το ότι παρά τη πολύ μικρότερη ποσότητα των δημοσιευμένων καταχωρήσεων σε σχέση με την κυρία Erica, υπερέχει στην «ευστοχία» αυτών ως προς την επίδραση που ασκούν. *Χαρακτηριστική δημοσιοποιημένη εγγραφή του είναι εκείνη με θέμα το iPhone[2]. Αυτήν η εγγραφή, είναι λεπτομερής περιγραφή της προσωπικής του εμπειρίας και όχι αποτέλεσμα που προήλθε από εξωτερικές πηγές ειδήσεων.*

## 6.2 Κατανομή των «Κορυφαίων» της Digg

Καταρχάς, υπάρχει ανάγκη ύπαρξης ενός κοινού σημείου αναφοράς με στόχο το διαχωρισμό μεταξύ των τεσσάρων ειδών (βλέπε ενότητα 1.2) των Bloggers. Προς τη κατεύθυνση αυτή είναι σημαντική η ιστοσελίδα της “digg”(www.digg.com) που ανήκει στις web 2.0 τεχνολογίες. Μέσω αυτής της ιστοσελίδας, οι χρήστες έχουν τη δυνατότητα να διαμοιράζονται, να διαχειρίζονται σωστά, να αφομοιώνουν εύκολα τα καίρια (σύμφωνα με το δηλωμένο σκορ αρεσκείας που ακολουθεί) θέματα της

επικαιρότητας. Αυτό συμβαίνει επειδή όλα τα posts (προερχόμενα από οποιονδήποτε από τους τέσσερις τύπους των Bloggers) που βρίσκονται στον ιστότοπο της “digg” έπονται της ψηφοφορίας στον υπερσύνδεσμο “[www.digg.com](http://www.digg.com)” και έτσι διατάσσονται σύμφωνα με σκορ αρεσκείας το οποίο ανατίθεται στα “posts” από τους χρήστες επισκέπτες της “digg” που ψηφίζουν. Επιπροσθέτως, για το «παράθυρο» χρονικής διάρκειας 30 ημερών που διατυπώθηκε στη προηγούμενη ενότητα, η “digg” «επιτρέπει» την κατ’ απαίτηση εξόρυξη κορυφαίων (έστω των πρώτων 100) ταξινομημένων posts έτσι ώστε να αποτελέσουν εργαλείο αξιολόγησης του μοντέλου i-index. Ο μήνας Ιανουάριος του 2007 αντιστοιχεί στο προαναφερθέν «παράθυρο» μέσα στο οποίο έγινε η συγκέντρωση:

- ✓ των κορυφαίων 100 δημοσιευμένων καταχωρήσεων της “digg”
- ✓ 535 δημοσιοποιημένων εγγραφών της TUAW

### 6.2.1 Πειραματικά Αποτελέσματα

Θεωρώντας τις τέσσερις κατηγορίες των Bloggers, γίνεται διανομή των posts (της «κοινωνίας» TUAW) που δημιουργούνται από καθεμιά από αυτές σε διαφορετικό σύνολο  $S_i$ , με  $i=1, \dots, 4$ . Η αντιστοίχιση καθενός από αυτά τα σύνολα γίνεται προς καθεμιά από τις τέσσερις κατηγορίες των Bloggers όπως αυτές διατυπώθηκαν στη παρούσα διπλωματική εργασία. **Κατόπιν, αφού γίνει επιλογή των 20 πρώτων posts από τα σύνολα  $S_1, S_2, S_4$  και είκοσι τυχαίων από το  $S_3$ , θα γίνεται σύγκριση αυτών των επιλεγμένων «ποσοτήτων» με τις 100 κορυφαίες δημοσιευμένες εγγραφές που προέκυψαν από τη “digg”.** Στον Πίνακα 14 παρουσιάζονται «ποσότητες» των posts καθενός από τα  $S_i$  οι οποίες υπάρχουν και στις εκατό κορυφαίες της “digg”. Ενδιαφέρουσες είναι οι καταχωρημένες τιμές για τα  $S_2, S_3$ . Στους Πίνακες 15 και 16 βρίσκονται αντίστοιχα: οι κατανομές του συνόλου των 100 κορυφαίων της “digg” και οι κατανομές του συνόλου των 535 εγγραφών της TUAW. Από αυτούς τους τρεις πίνακες προκύπτει πως το **σκορ αρεσκείας** όλων εκείνων που έχουν ρόλο σχολιαστή/δημιουργού δημοσιευμένης εγγραφής/εγγραφών είναι μεγαλύτερο για τις δημοσιοποιημένες καταχωρήσεις που εξασκούν μεγαλύτερο επηρεασμό. Για τον Πίνακα 15, εντύπωση προκαλούν οι 7 καταχωρήσεις του συνόλου  $S_3$ . Οι παρατηρήσεις που έγιναν όσον αφορά στη κατανομή του Πίνακα 15 έδειξαν πως η

## 6<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Πειραματική Αξιολόγηση του Μοντέλου I-index

αύξηση της «ποσότητας» των κορυφαίων Bloggers ως προς τον επηρεασμό που εξασκούν επιδρά στη κατανομή αυτή.

Τύποι των Bloggers	«Ενεργοί»	μη «ενεργοί»
Όσοι εξασκούν επηρεασμό	<b>S<sub>1</sub>:17</b>	<b>S<sub>4</sub>:7</b>
Όσοι δεν εξασκούν επηρεασμό	<b>S<sub>2</sub>:3</b>	<b>S<sub>3</sub>:0/1</b>

**Πίνακας 14:** Ποσότητες των S<sub>i</sub> που υπάρχουν στις 100 κορυφαίες της “Digg”

Τύποι των Bloggers	«Ενεργοί»	μη «ενεργοί»
Όσοι εξασκούν επηρεασμό	<b>S<sub>1</sub>:71</b>	<b>S<sub>4</sub>:14</b>
Όσοι δεν εξασκούν επηρεασμό	<b>S<sub>2</sub>:8</b>	<b>S<sub>3</sub>:7</b>

**Πίνακας 15:** Κατανομές του συνόλου των 100 κορυφαίων της “Digg”

Τύποι των Bloggers	«Ενεργοί»	μη «ενεργοί»
Όσοι εξασκούν επηρεασμό	<b>S<sub>1</sub>:327</b>	<b>S<sub>4</sub>:42</b>
Όσοι δεν εξασκούν επηρεασμό	<b>S<sub>2</sub>:131</b>	<b>S<sub>3</sub>:35</b>

**Πίνακας 16:** Κατανομές του συνόλου των 535 εγγραφών της TUAW

Αναφορικά με τις 535 δημοσιοποιημένες εγγραφές της TUAW που συλλέχθηκαν όπως αναφέρθηκε στη παρούσα ενότητα, προκειμένου να γίνει διαχωρισμός των posts που ασκούν επιρροή από εκείνα που δεν ασκούν, χρησιμοποιήθηκε ο ορισμός που αναφέρθηκε στην ανάπτυξη του δείκτη i-index (2.1 ενότητα). Οι 5 κορυφαίοι ως προς την επιρροή που εξασκούν Bloggers του Πίνακα 12 της ενότητας 6.1 αξιοποιούνται ως προς όλες τους τις δημοσιευμένες καταχωρήσεις ώστε να αναφερθούν έστω ως δύο σύνολα εκείνα τα posts της TUAW που ασκούν επιρροή από εκείνα που δεν ασκούν. Με αυτόν τον τρόπο, για τον Ιανουάριο του 2007 για τις 535

δημοσιοποιημένες εγγραφές της TUAW προέκυψαν τα πλήθη που αναγράφονται στις δύο τελευταίες στήλες του πίνακα 17 που ακολουθεί.

Μελετώντας τον πίνακα παρατηρούμε ότι τα posts που ασκούν επίδραση είναι πολύ μεγαλύτερα σε μήκος και έχουν πολλά περισσότερα σχόλια. Υπάρχουν πολλοί περισσότεροι inlinks στα posts που εξασκούν επιρροή, αλλά το πλήθος των outlinks αποτελεί μία ασθενής απόδειξη της εξασκούμενης επιρροής, παρόλου που τα posts που εξασκούν επιρροή έχουν ελαφρώς μικρότερο αριθμό outlinks από εκείνα που δεν εξασκούν επηρεασμό.

	Number of Comments		Number of Inlinks		Length of Blog Post		Number of Outlinks		Total Number of Blog Posts
	Max	Avg	Max	Avg	Max	Avg	Max	Avg	
Influential Blog Posts	112	74.18	80	38.63	3529	1999.32	15	3.36	22
Non-influential Blog Posts	69	10.84	39	8.96	1930	703.74	27	4.3	513

**Πίνακας 17:** Αποτελέσματα των Στατιστικών για τα Blog Posts που δεν εξασκούν επηρεασμό και για εκείνα που εξασκούν

### 6.3 «Βάρη» του I-index

Υπάρχουν τέσσερα βάρη στο μοντέλο i-index απαραίτητα για τη ρύθμιση της συμβολής των τεσσάρων παραμέτρων στον υπολογισμό του σκορ επιρροής (εξισώσεις α και β ενότητας 2.1). Για να τα ξαναθυμηθούμε το  $w_{in}$  είναι για την επίδραση από τους **εισερχόμενους συνδέσμους**, το  $w_{out}$  για την επιρροή από **εξωτερικούς συνδέσμους**, το  $w(\lambda)$  για την ιδιότητα “**goodness**” ενός blog post και το  $w_{comm}$  για τον **αριθμό των σχολίων**. Όλα τα βάρη παίρνουν πραγματικές τιμές στο  $[0,1]$ . Παρακάτω εξετάζουμε το πως η αλλαγή στις τιμές τους θα επηρεάσει την κατάταξη των influential.

- i. Αλλάζοντας την  $w(\lambda)$  απλώς κλιμακώνεται (scales up or down) η τιμή του σκορ ένδειξης επηρεασμού ενός post, οπότε δεν επηρεάζει την κατάταξη των influential bloggers (σε ένα blog site).
- ii. Μεταβάλλοντας τη τιμή του  $w_{comm}$  στο εύρος από 0.0 ως και το 1.0, το μοντέλο μένει σταθερό για τις τιμές που είναι τουλάχιστον ίσες με το 0.6, δηλαδή δεν επηρεάζεται η κατάταξη των influential bloggers
- iii. Όταν μεταβληθεί η τιμή του  $w_{in}$ , θα υπάρχει σταθερότητα στο προκαταρκτικό μοντέλο όταν  $w_{in} \geq 0.9$ .

- iv. Όταν μεταβληθεί η τιμή του  $w_{out}$ , θα υπάρχει σταθερότητα στο προκαταρκτικό μοντέλο όταν  $w_{out} \geq 0.2$ .

Όταν θα ισχύει ταυτόχρονα πως τα βάρη θα έχουν τις κατάλληλες τιμές επίτευξης σταθερότητας που υπάρχουν στα ii, iii και iv, τότε το προκαταρκτικό μοντέλο παραμένει αμετάβλητο. Οι ρυθμίσεις των βαρών του μοντέλου καθορίζουν τα γνωρίσματα εκείνων που θα βρεθούν και οι οποίοι θα ασκούν επίδραση και θα έχουν ρόλο σχολιαστή/δημιουργού δημοσιοποιημένης καταχώρισης/καταχωρήσεων σε Blog. Η πιθανή διαφοροποίηση/διαφοροποιήσεις στη σειρά ταξινόμησης των Bloggers που διακρίνονται για την επίδραση που ασκούν έγκειται στο ότι τα posts τους εξασκούν επηρεασμό που προέρχεται από οποιαδήποτε από τις τέσσερις παραμέτρους του **Πίνακα 13**. Για παράδειγμα τιμές  $w_{in}=w_{out}=0$  δίνουν αντίστοιχη σειρά ταξινόμησης εξαρτώμενη από τη 1<sup>η</sup> παράμετρο του **Πίνακα 13**. Επιπρόσθετα, το να ισχύει αύξηση τιμών της  $w_{out}$  είναι ένας τρόπος της μη ενθάρρυνσης των παραπομπών από άλλα posts. Συνεπώς και τα τέσσερα βάρη παρέχουν το μέσο για περαιτέρω εξέλιξη και επέκταση του μοντέλου i-index για ένα ευρύ φάσμα εφαρμογών.

## 6.4 Χρονικά «Σχέδια» των influential bloggers

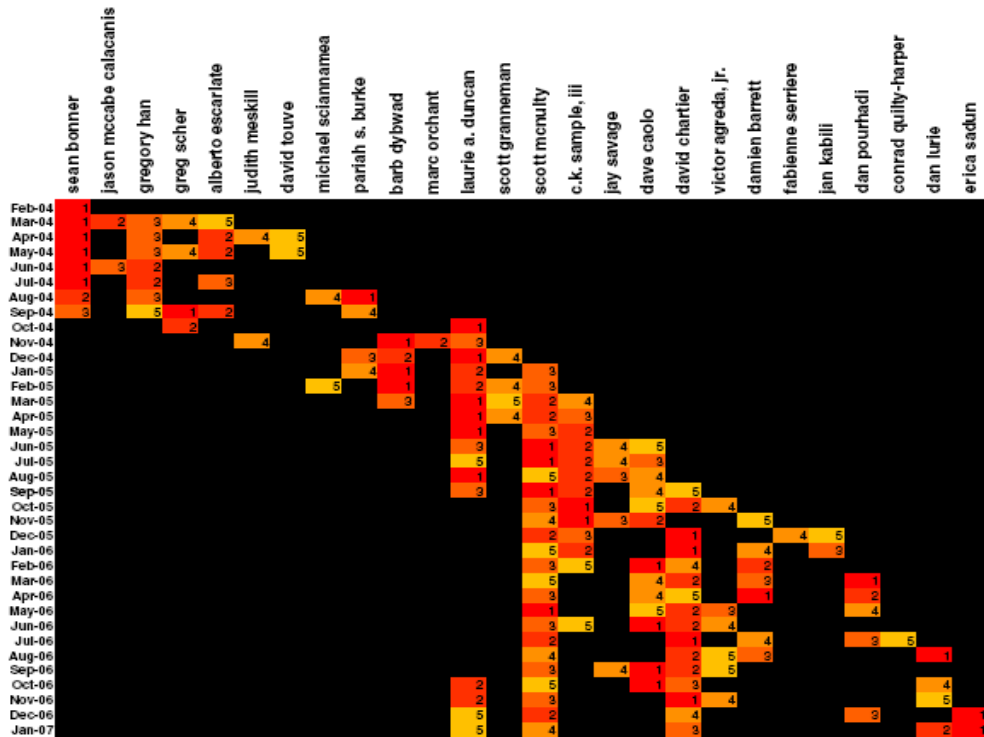
Έχει σημασία η μελέτη των χρονικών σχεδίων των σχολιαστών/δημιουργών δημοσιευμένης εγγραφής σε ιστολόγιο για τις ιστοσελίδες τύπου Blog όπου έχουν αναρτηθεί πολλές δημοσιοποιημένες εγγραφές. Όπως αναφέραμε σε προηγούμενη ενότητα, πραγματοποιείται μελέτη για την αναγνώριση των κορυφαίων πέντε ως προς την επιρροή που εξασκούν bloggers συνολικά κατά το χρονικό διάστημα από τον Φεβρουάριο του 2004 ως και τον Ιανουάριο του 2007 που αφορά στη λειτουργία του ιστολογίου της TUAW. Για να γίνει ορθά η παραπάνω μελέτη, χρησιμοποιούνται συνεχόμενα χρονικά παράθυρα 30 ημερών που δεν επικαλύπτονται μεταξύ τους και είναι τόσα ώστε να καλύπτουν το χρονικό διάστημα που διατυπώθηκε πριν. Αναλυτικά, από την **Εικόνα 5** είναι προφανή τα ακόλουθα:

- ✓ Το ολικό πλήθος των στηλών του πίνακα που διακρίνεται ισούται με το ολικό πλήθος των Bloggers που βρέθηκαν να ασκούν επίδραση κατά το χρονικό διάστημα που διατυπώθηκε πριν για τη «κοινωνία» της TUAW.

## 6<sup>ο</sup> ΚΕΦΑΛΑΙΟ: Πειραματική Αξιολόγηση του Μοντέλου I-index

Δηλαδή, η χρονική διαμόρφωση των χρονικών σχεδίων αυτών των Bloggers απεικονίζεται σε αυτή την εικόνα.

- ✓ Οι θέσεις του πίνακα που είναι χρωματισμένες αποτελούν ένδειξη για το ότι ο κάθε μήνας, χρονικό παράθυρο 30 ημερών, αποτελεί μήνα κατά τον οποίο το άτομο της στήλης που αντιστοιχεί στη κάθε θέση συγκαταλέγεται μεταξύ των κορυφαίων Bloggers που προέκυψαν κατά το εξεταζόμενο χρονικό διάστημα που προαναφέρθηκε. Πιο συγκεκριμένα, όσο πιο σκούρο είναι το χρώμα της χρωματισμένης θέσης γίνεται πιο προφανές ότι ο Blogger αυτής θα ασκεί περισσότερη επίδραση. Τα τετράγωνα που είναι μαύρα δηλώνουν ότι στους συγκεκριμένους μήνες οι bloggers που αντιστοιχούν σε αυτά δεν ασκούσαν επίδραση αρκετή έτσι ώστε να συμπεριληφθούν μεταξύ των κορυφαίων.
- ✓ Ο αριθμός που είναι μέσα σε κάθε τετραγωνάκι δηλώνει τη σειρά στην οποία βρίσκεται ταξινομημένο το άτομο που αναγράφεται σε κάθε στήλη που τέμνει το τετραγωνάκι.
- ✓ **Σημαντικό εύρημα αποτελεί ο μοναδικός Blogger “sean bonner” που κατά το χρονικό διάστημα υλοποίησης της μελέτης, κατόρθωσε να διατηρείται στη 1<sup>η</sup> θέση των κορυφαίων bloggers ως προς την επιρροή που εξασκούσαν για 6 συνεχόμενα χρονικά παράθυρά 30 ημερών, που είναι το μεγαλύτερο δυνατό.**



Εικόνα 5: Χρονικά Σχέδια Influential Bloggers

Υπάρχουν τέσσερα διαφορετικά είδη από Bloggers που εξασκούν επιρροή. Αυτά είναι τα ακόλουθα:

1. Το 1<sup>ο</sup> είδος αποτελείται από όσους εξασκούν επηρεασμό «μεσοπρόθεσμα» (**Average-Term Influential**): Αυτοί οι Bloggers, βρίσκονται στη κατάσταση “influential” για χρονικό διάστημα περίπου 4 με 5 μήνες, όπως για παράδειγμα ο “gregory han”.
2. Το 2<sup>ο</sup> είδος αποτελείται από όσους εξασκούν επηρεασμό «Μακροπρόθεσμα» (**Long Term Influential**): Αυτοί οι Bloggers, παραμένουν στην ομάδα των influential για ένα μεγάλο χρονικό διάστημα, με χαρακτηριστικό παράδειγμα τον “Scott McNulty” που διατηρήθηκε ως influential για 25 συνεχόμενα χρονικά παράθυρα 30 ημερών.
3. Το 3<sup>ο</sup> είδος (**Transient Influential**) αποτελείται από όσους εξασκούν επιρροή για πολύ μικρό χρονικό διάστημα. Εδώ, αντιπροσωπευτικό παράδειγμα είναι ο “fabienne serriere” ο οποίος διατηρήθηκε ως influential για μόνο ένα χρονικό παράθυρο 30 ημερών.



4. Το 4<sup>ο</sup> είδος (Burgeoning influential) αποτελείται από τους ανερχόμενους influential και είναι αυτοί που ξεπρόβαλλαν ως influential bloggers πρόσφατα. Σε αυτή τη κατηγορία ανήκουν οι Dan Lurie και Erica Sadun.

Διαφορετικοί bloggers μπορούν να παρουσιάσουν διαφορετικά χρονικά σχέδια. Οι μακροπρόθεσμοι influential έχουν μεγαλύτερη επιρροή από άλλους bloggers καθώς συγκριτικά είναι πιο αξιόπιστοι. Οι ανερχόμενοι influential έχουν τη δυνατότητα να μεταβούν στην κατηγορία των μακροπρόθεσμων influential, αλλά είναι δύσκολο να γίνει το ίδιο πράγμα για τους προσωρινούς influential (transient) καθώς αυτοί μπορεί να εξασκούν επηρεασμό κατά τύχη. Σίγουρα, υπάρχουν πολλά άλλα χρονικά σχέδια που εξαρτώνται από συγκεκριμένη εφαρμογή.

Πολλές εν δυνάμει εφαρμογές μπορούν να αναπτυχθούν χρησιμοποιώντας τις παραπάνω κατηγορίες των influential. Μία από αυτές σχετίζεται με τη διαδικασία «γνωριμίας» με μια νέα ιστοσελίδα τύπου Blog που ανήκει στη «κοινωνία» της TUAW. Μία προσέγγιση του τύπου αυτού, είναι ο «επισκέπτης» του blog να «κοιτάει» για όσους εξασκούν μακροπρόθεσμα επηρεασμό στη TUAW. Επίσης, τα posts ενός “Transient Influential” μπορούν να χρησιμοποιηθούν για να κατανοηθούν τα μεταβαλλόμενα “topics”.

## 6.5 Επιπρόσθετα Πειράματα

Σε αυτή την ενότητα αναφέρονται κάποια επιπλέον πειράματα αφενός α) για να εξετασθούν οι κατά ζεύγη συσχετίσεις των 4 στατιστικών (σχόλια, inlinks, μήκος post, outlinks) και αφετέρου β) για να μελετηθεί το ποσοστό έκφρασης σχολίων (rate of comments) με σκοπό την διεύρυνση του αρχικού i-index μοντέλου.

Σχετικά με το α) που αναφέρθηκε προηγουμένως

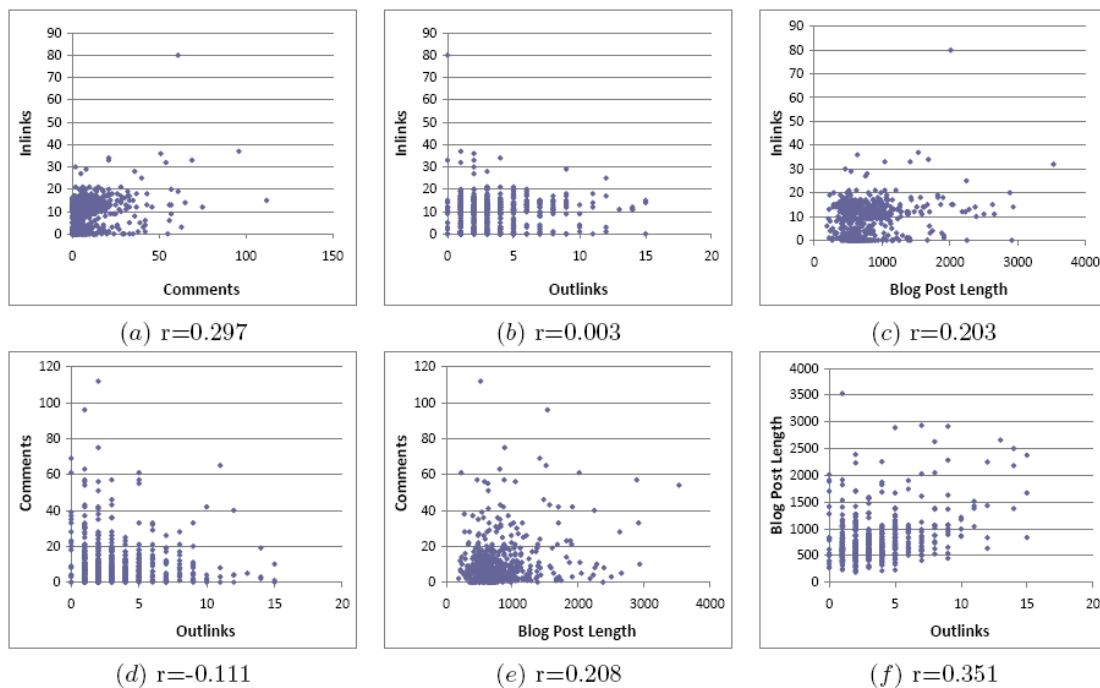
- Θα προκύψουν οι γραφικές παραστάσεις που φαίνονται στην **Εικόνα 6**
- Η μεταβλητή “r” που είναι κάτω από κάθε γραφική παράσταση λέγεται “συντελεστής συσχέτισης”. Όταν η “r” έχει αρνητική τιμή (το οποίο διακρίνεται μόνο στη “d” γραφική παράσταση) συμπεραίνεται η ακόλουθη σχέση: για μεγάλο πλήθος εξερχόμενων συνδέσμων προς μια

δημοσιευμένη εγγραφή θα είναι μικρό το πλήθος των σχολίων που “αναφέρονται” στην εγγραφή αυτή. Το συμπέρασμα αυτό, είναι αληθές και για την αντίστροφη της σχέσης αυτής.

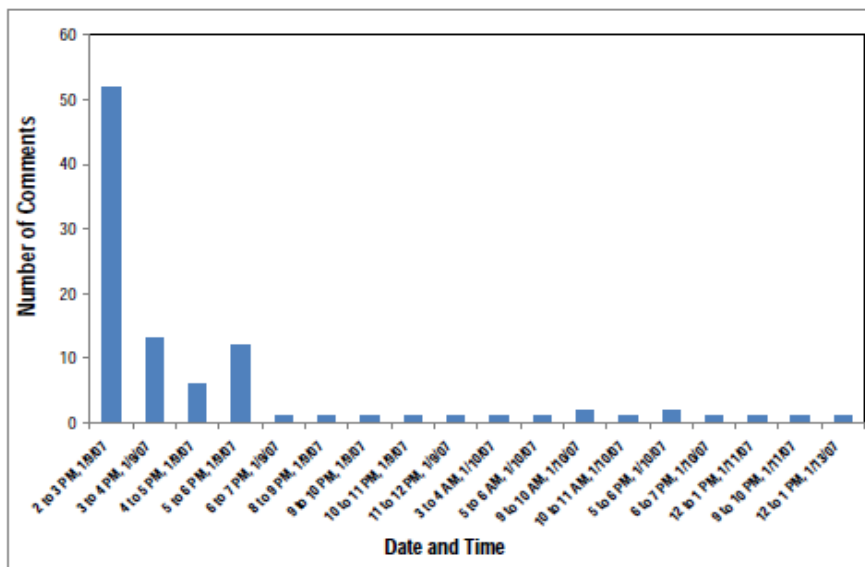
Σχετικά με το β) που αναφέρθηκε προηγουμένως

- Διενεργείται μελέτη περίπτωσης με θέμα το πόσο «με διακυμάνσεις» (spiky) ή «επίπεδη» (flat) είναι η γραφική παράσταση κατανομής του ποσοστού έκφρασης σχολίων για μια δημοσιοποιημένη καταχώρηση σε blog. Τα αποτελέσματα απεικονίζονται στις **Εικόνα 7** και **Εικόνα 8**. Οι κατανομές αυτών των δύο εικόνων «αναφέρονται» σε δύο διαφορετικές δημοσιοποιημένες καταχωρήσεις σε blog οι οποίες καταχωρήσεις ασκούν επιρροή. Η δημοσιοποιημένη εγγραφή της **Εικόνας 7** αφορά σε κάποια έκδοση ενός iphone ενώ εκείνη της **Εικόνας 8** αφορά στην εταιρεία Apple. Για την **Εικόνα 7**, παρατηρείται μία «άμεσα απότομη» ανοδική συμπεριφορά απόκρισης σχολιασμού όσων είναι σχολιαστές δημοσιευμένης εγγραφής μόλις μία ώρα μετά από την δημοσίευση της αντίστοιχης εγγραφής. Την ώρα αυτή σημειώθηκε το μέγιστο ποσοστό σχολιασμού. Η κατανομή της **Εικόνας 8** κρίνεται ως «επίπεδη» καθώς εξακολουθεί να υπάρχει διακύμανση του πλήθους σχολίων των bloggers κοντά στη τιμή 10 ακόμη και 7 με 8 ώρες ύστερα από την ανάρτηση της δημοσιευμένης καταχώρισης. Στον οριζόντιο άξονα τόσο της **Εικόνας 7** όσο και της **Εικόνας 8** παριστάνεται η ημερομηνία και η ώρα έκφρασης του εκάστοτε πλήθους σχολίων.
- Η έννοια «ποσοστό έκφρασης σχολίων» είναι κατάλληλη για το «χτίσιμο» εφαρμογών ειδικού τύπου όπως είναι οι κρίσιμες ως προς το χρόνο εφαρμογές διαχείριση και πρόληψη καταστροφής και ο χειρισμός περίπτωσης έκτακτης ανάγκης.
- Η επέκταση του μοντέλου i-index περιλαμβάνει:
  - i. Την μελέτη του τρόπου φιλτραρίσματος των σχολίων τύπου spam για την αποτροπή των τύπου spam επιθέσεων χρησιμοποιώντας τις τεχνικές που διατυπώθηκαν μέσα στις [5,6].

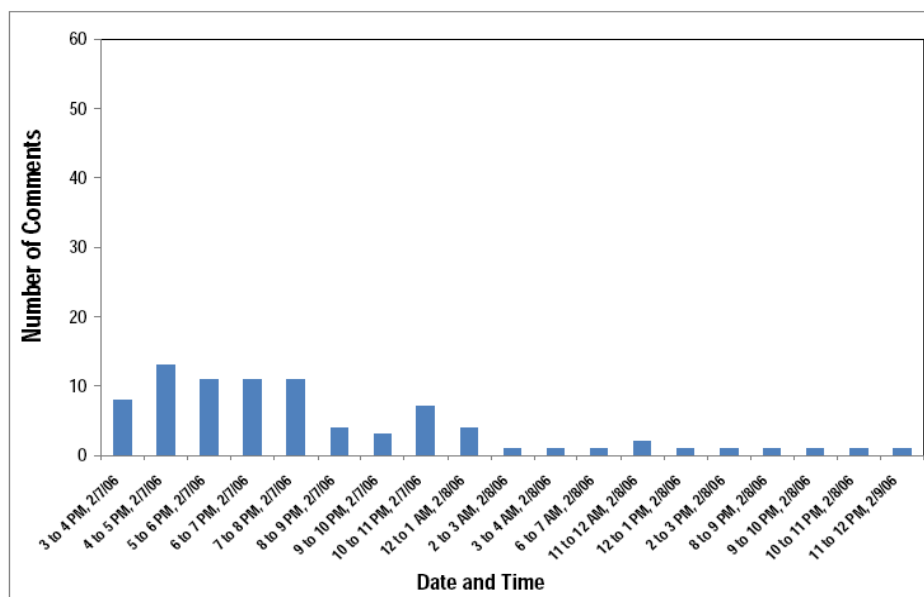
- ii. Την μελέτη περισσότερο κατάλληλων τεχνικών εκτίμησης της ποιότητας ενός blog post συμπεριλαμβανομένων της λογοτεχνικής ανάλυσης και της ανάλυσης περιεχομένου (literary and content analysis).
- iii. Τη μελέτη διαφορετικών συναρτήσεων ώστε να «επιβληθεί» με μη-γραμμικό τρόπο:
  1. Σχεδόν αμελητέα «ποινή» στην επίδραση που εξασκείται όταν είναι μικρό το πλήθος των εξερχόμενων συνδέσμων.
  2. Αρκετά σημαντική «ποινή» στην επίδραση που εξασκείται όταν είναι αρκετά υψηλό το πλήθος των εξερχόμενων συνδέσμων. Μία τέτοια συνάρτηση θα μπορούσε να είναι εκθετική που θα αντικαθιστούσε το  $w_{out} \sum_{n=1}^{|\theta|} I(p_n)$  στην εξίσ.(α) (ενότητα 2.1) με το  $\exp(w_{out} \sum_{n=1}^{|\theta|} I(p_n))$ . Για τη περίπτωση χρήσης μιας τέτοιας συνάρτησης πρέπει να διερευνηθεί διεξοδικά ο ρόλος της μεταβλητής  $w_{out}$ .



Εικόνα 6: Κατά ζεύγη συσχετίσεις στατιστικών



Εικόνα 7: Κατανομή του ποσοστού έκφρασης σχολίων «με διακυμάνσεις»



Εικόνα 8: Επίπεδη κατανομή του ποσοστού έκφρασης σχολίων

## ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΕΦΑΛΑΙΟΥ

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “Identifying the influential bloggers in a community”, in Proceedings of ACM WSDM Conf., pages 207–218, 2008.

[2]<http://www.tuaw.com/2007/01/09/iphone-will-not-allowuser-installable-applications/>

[3]<http://www.tuaw.com/2007/01/04/xpad-developer-saysmaczot-and-brian-ball-ripped-him-off/>

[4] Identifying the Influential Bloggers in a Community, Nitin Agarwal, Huan Liu, Lei Tang - Arizona State University Tempe, AZ 85287, USA, Philip S. Yu - University of Illinois at Chicago Chicago, IL 60607, USA

[5] P. Kolari, T. Finin, and A. Joshi, “SVMs for the blogosphere: Blog identification and splog detection”, in AAI Spring Symposium on Computational Approaches to Analyzing Weblogs, 2006.

[6] Yu-Ru Lin, Hari Sundaram, Yun Chi, Junichi Tatemura, and Belle L. Tseng, “Splog detection using self-similarity analysis on blog temporal dynamics”, in Proceedings of the 3rd international workshop on Adversarial information retrieval on the web (AIRWeb), pages 1–8, New York, NY, USA, 2007, ACM Press.

[7] Identifying Influential Bloggers: Time Does Matter, Leonidas Akritidis, Dimitrios Katsaros, Panayiotis Bozanis - Department of Computer & Communication Engineering University of Thessaly Volos, Greece