



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Εκτίμηση της επιρροής στα κοινωνικά δίκτυα (SNs): μια
συγκριτική μελέτη αλγορίθμων και εμπορικών εργαλείων**

Χρήστος Κ. Αρμύρος

Επιβλέπουσες: **Αφροδίτη Τσαλαγατίδου**, Αναπληρώτρια Καθηγήτρια
Δρ. Ελένη Κουτρούλη, Μεταδιδακτορική Ερευνήτρια

ΑΘΗΝΑ

Νοέμβριος 2018

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Εκτίμηση της επιρροής στα κοινωνικά δίκτυα (SNs): μια συγκριτική μελέτη αλγορίθμων και εμπορικών εργαλείων

Χρήστος Κ. Αρμύρος

A.M.: 1115200500027

Επιβλέπουσες: **Αφροδίτη Τσαλαγατίδου**, Αναπληρώτρια Καθηγήτρια
Δρ. Ελένη Κουτρούλη, Μεταδιδακτορική Ερευνήτρια

ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή εργασία επικεντρώνεται στη μελέτη της επιρροής στα κοινωνικά δίκτυα και στους μηχανισμούς υπολογισμού της. Ο υπολογισμός της δημοτικότητας και της επιρροής στα κοινωνικά δίκτυα είναι μια ανάγκη που έχει παρουσιαστεί τα τελευταία χρόνια με τη ραγδαία αύξηση της χρήσης αυτών από εκατομμύρια χρήστες. Αρχικά εξετάζονται οι μελέτες που έχουν γίνει γύρω από το αντικείμενο και τον τρόπο λειτουργίας της κάθε μιας ξεχωριστά και στη συνέχεια αναλύονται τα ήδη υπάρχοντα συστήματα υπολογισμού επιρροής που έχουν υλοποιηθεί και χρησιμοποιούνται. Τέλος πραγματοποιείται μια ταξινόμηση των μηχανισμών που έχουν αναλυθεί, σύμφωνα με κοινά κριτήρια που ορίζονται και επεξηγούνται ώστε να είναι ευκολότερη η κατηγοριοποίηση και ο διαχωρισμός τους. Σκοπός της εργασίας είναι η εύρεση κοινών σημείων αναφοράς αλλά και ειδοποιών στοιχείων που κάνουν την κάθε τεχνική να διαφέρει. Τέλος, συγκρίνουμε την αποδοτικότητα των αποτελεσμάτων τους.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογισμός επιρροής στα κοινωνικά δίκτυα

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: επιρροή, κοινωνικά δίκτυα, μηχανισμοί υπολογισμού επιρροής

ABSTRACT

This thesis focuses on the study of influence on social networks and its calculation mechanisms. Estimating popularity and influence on social networks is a need that has been raised in the past few years since the usage of social networks has been increased tremendously by millions of users. Initially, we analyze studies that are examining the subject and how each one operates separately; then, we analyze the influence calculation systems that have been implemented and are being used. Finally, we apply a classification on the reputation systems that we have analyzed before, according to certain taxonomy criteria; thus, it is getting easier to categorize them. Our main target on this research is to find common points of reference as well as points of differentiation that make each technique unique and also compare the efficiency of their results.

SUBJECT AREA: Influence estimation in Social Networks

KEYWORDS: Influence, Social Networks, Reputation estimation systems

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα εκφράσω τις ευχαριστίες μου στην επιβλέπουσα καθηγήτριά μου, κα. Αφροδίτη Τσαλγατίδου και στη βοηθό της Δρ. Ελένη Κουτρούλη, για τη συνεργασία, τη βοήθεια, τις χρήσιμες ιδέες, την υπομονή τους, και την πολύτιμη συμβολή τους στη βελτίωση και ολοκλήρωση της παρούσας Πτυχιακής Εργασίας. Επίσης θα ήθελα να ευχαριστήσω θερμά τον κ. Βασίλειο Ζησιμόπουλο, ο οποίος με το ρόλο του συμβούλου φοίτησης με βοήθησε να προσπεράσω σημαντικά εμπόδια που αφορούσαν την εκπλήρωση των σπουδών μου. Τέλος, θα ήθελα να ευχαριστήσω της οικογένειά μου για τη συνεχή τους στήριξη όλα αυτά τα χρόνια, καθώς χωρίς αυτή θα ήταν ακόμα πιο δύσκολη η περίοδος φοίτησής μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ.....	10
1. ΕΙΣΑΓΩΓΗ.....	11
2. ΣΥΣΤΗΜΑΤΑ ΥΠΟΛΟΓΙΣΜΟΥ ΕΠΙΡΡΟΗΣ ΣΕ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ	13
2.1 Φήμη, επιρροή και δημοτικότητα σε κοινωνικά δίκτυα	13
2.1.1 Ταξινόμηση συστημάτων φήμης σε κοινωνικά δίκτυα	14
2.2 Συστήματα Επιρροής	16
3. ΑΝΑΛΥΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΣΥΣΤΗΜΑΤΩΝ ΕΠΙΡΡΟΗΣ ΓΙΑ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ....	18
3.1 Εύρεση Επιδραστικών Κόμβων σε ένα Δίκτυο Εμπιστοσύνης [1]	18
3.2. Εύρεση Επιδραστικών Μπλόγгер (Bloggers) σε ένα “Κοινωνικό Ιστολόγιο” [2]	19
3.3 Μηχανισμοί Φήμης σε Κοινωνικά Δίκτυα – Η περίπτωση ενός Συστήματος Υπολογισμού Επιρροής στο Twitter [3]	20
3.4 Εφαρμογή Μετρικών Κοινωνικών Δικτύων σε ένα Συνεργατικό Μοντέλο Εμπιστοσύνης για τη Δημιουργία Προσωποποιημένων Συστάσεων Χρηστών [4] ..	21
3.5 Εύρεση Επιδραστικών Χρηστών σε Κοινωνικά Δίκτυα [5]	23
3.6 Εκμάθηση Πιθανοτήτων Επιρροής στα Κοινωνικά Δίκτυα [6]	24
3.7 Στοιχεία καθορισμού μιας εικόνας ως δημοφιλούς [7].....	26
3.8 Μέτρηση Επιρροής στο Twitter [8]	27
4. ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΥΠΟ ΕΞΕΤΑΣΗ ΣΥΣΤΗΜΑΤΩΝ ΕΠΙΡΡΟΗΣ	28
5. ΔΙΑΘΕΣΙΜΑ ΕΡΓΑΛΕΙΑ ΥΠΟΛΟΓΙΣΜΟΥ ΕΠΙΡΡΟΗΣ ΣΕ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ	30
5.1 Klout – [10], [11], [12].....	30
5.2 Alexa [14], [15]	32
5.3 TweetAnalyzer [16]	33
5.4 TweetReach [18], [19]	34

5.5 UnionMetrics [20], [21]	34
5.6 Ταξινόμηση των εργαλείων υπολογισμού επιρροής	37
6. ΣΥΜΠΕΡΑΣΜΑΤΑ	39
ΑΝΑΦΟΡΕΣ	41

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Figure 1. Γραφική απεικόνιση πληροφοριών για τους χρήστες που αλληλοεπιδρούν στο περιεχόμενο των δημοσιεύσεων	35
Figure 2. Γραφική απεικόνιση σε μορφή δένδρου σχετικά με τον αριθμό και το βαθμό αναδημοσιεύσεων μιας δημοσίευσης	36
Figure 3. Γραφική απεικόνιση σχετικά με τη δραστηριότητα των δημοσιεύσεων ενός χρήστη σε βάθος χρόνου	36

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1. Ταξινόμηση μηχανισμών υπολογισμού επιρροής	29
Πίνακας 2 Ταξινόμηση των συστημάτων υπολογισμού επιρροής Klout και alexa.....	37
Πίνακας 3. Ταξινόμηση των συστημάτων υπολογισμού επιρροής tweetanalyzer, tweetreach και unionmetrics	38

ΠΡΟΛΟΓΟΣ

Η παρούσα Πτυχιακή εργασία με τίτλο «Εκτίμηση της επιρροής στα κοινωνικά δίκτυα (SNs): μια συγκριτική μελέτη αλγορίθμων και εμπορικών εργαλείων» εκπονήθηκε στα πλαίσια της ολοκλήρωσης των προϋποθέσεων, για τη λήψη του πτυχίου μου από το τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Σκοπός μου κατά τη διάρκεια της συγγραφής, δεν ήταν μόνο η ορθή και όσο το δυνατόν πληρέστερη ανάλυση του θέματος, αλλά έγινε προσπάθεια ώστε το περιεχόμενο της εργασίας να είναι κατανοητό και σαφές.

1. ΕΙΣΑΓΩΓΗ

Τα κοινωνικά δίκτυα έχουν παρουσιάσει τεράστια ανάπτυξη και άνθηση τα τελευταία χρόνια στις σύγχρονες κοινωνίες. Οι χρήστες τους τα χρησιμοποιούν πολύ συχνά είτε για διασκέδαση είτε για κοινωνικοποίηση είτε για την πραγματοποίηση καθημερινών τους αναγκών. Παρόλο που τα κοινωνικά δίκτυα διαχωρίζονται σύμφωνα με τον τρόπο επικοινωνίας σε δίκτυα διαμοιρασμού περιεχομένου, έκφρασης απόψεων ή κριτικών ή απλής κοινωνικής δικτύωσης, τα κοινά δομικά τους στοιχεία είναι ο χρήστης και το περιεχόμενο. Λόγω της μεγάλης διάδοσής τους, έχει παρουσιαστεί το φαινόμενο της αλληλεπίδρασης και της επιρροής των χρηστών αναμεταξύ τους, το οποίο χρήζει ανάλυσης, και μπορεί να αποφέρει σημαντικά και στοχευμένα αποτελέσματα που μπορούν να φανούν πολύ χρήσιμα σε διαφημιστικές δραστηριότητες. Για παράδειγμα, για να αποκτήσει μια διαφημιστική καμπάνια τη μέγιστη δυνατή απόδοση με τη μικρότερη δυνατή χρηματική δαπάνη, αρκεί να απευθυνθεί στα μέλη του κοινωνικού δικτύου που ασκούν τη μεγαλύτερη επιρροή στους υπόλοιπους. Στόχος λοιπόν των ερευνητών είναι η εύρεση αυτών των χρηστών και ο υπολογισμός της επιρροής που ασκούν στους γύρω τους.

Το φαινόμενο της επιρροής στα κοινωνικά δίκτυα ορίζεται ως η διαδικασία κατά την οποία ένας χρήστης A πραγματοποιεί μία ενέργεια τη οποία νωρίτερα έχει πραγματοποιήσει ο χρήστης B. Τότε λέμε ότι ο B ασκεί επιρροή στον A ή αντίστροφα ότι ο A επηρεάζεται από τον B. Αναλόγως το κοινωνικό δίκτυο και την οπτική γωνία που θέλουμε να εξετάσουμε, διαχωρίζουμε την επιρροή ως προς το χρήστη ή ως προς το περιεχόμενο. Η επιρροή ως προς το χρήστη περιγράφει το κατά πόσο οι ενέργειες του θα ακολουθηθούν ή θα επηρεάσουν τη μελλοντική συμπεριφορά άλλων χρηστών ενώ η επιρροή ως προς το περιεχόμενο επικεντρώνεται στη δραστηριότητα που μπορεί να εκκινήσει μια δημοσίευση, ανεξαρτήτως του χρήστη που την παρήγαγε.

Η εύρεση των πιο δημοφιλών χρηστών που ασκούν την μεγαλύτερη επιρροή, βασίζεται σε στοιχεία που αντλούνται από το ίδιο το κοινωνικό δίκτυο. Η πιο συνήθης μέθοδος είναι η απεικόνιση του δικτύου σε μορφή γράφου όπου κάθε κόμβος απεικονίζει ένα χρήστη και οι ακμές συμβολίζουν τις σχέσεις των χρηστών αναμεταξύ τους. Πολλές φορές ο γράφος είναι κατευθυνόμενος (δηλαδή οι ακμές του έχουν συγκεκριμένη φορά) και με αυτό τον τρόπο φαίνεται ποιος χρήστης έκανε πρώτος μια πράξη και επηρέασε το γειτονικό του. Εφαρμόζοντας υπολογιστικές πράξεις γράφων στον γράφο του κοινωνικού δικτύου, εξάγονται αποτελέσματα που υποδείχνουν τους χρήστες που ασκούν τη μεγαλύτερη επιρροή. Ένας άλλος τρόπος εύρεσης δημοφιλών χρηστών βασίζεται στους πίνακες καταγραφής συμβάντων. Ένας τέτοιος πίνακας περιέχει στοιχεία για τη χρονική στιγμή του συμβάντος, για το είδος της πράξης και για το χρήστη που την πραγματοποίησε. Ανατρέχοντας στις γραμμές του πίνακα και εφαρμόζοντας συγκεκριμένους αλγόριθμους εξόρυξης δεδομένων σε αυτόν, παράγονται αποτελέσματα που υποδεικνύουν τους δημοφιλέστερους χρήστες του κοινωνικού δικτύου. Σπάνια συνδυάζονται αυτές οι δύο τεχνικές.

Η παρακάτω εργασία αποτελεί μια συγκριτική μελέτη για τους διαφορετικούς τρόπους έκφρασης και υπολογισμού της έννοιας της επιρροής στα κοινωνικά δίκτυα. Αναλύουμε σε βάθος τους διάφορους αλγόριθμους υπολογισμού επιρροής, όσων αφορά τον τρόπο λειτουργεία τους, τα δεδομένα που λαμβάνουν ως είσοδο και τους ξεχωρίζουμε ανάλογα με

το αν είναι εξειδικευμένοι σε κάποιο συγκεκριμένο κοινωνικό δίκτυο ή μπορούν να λειτουργήσουν το ίδιο καλά σε διαφορετικού τύπου εφαρμογές. Επίσης, αξιοποιώντας τα κριτήρια του ταξινομητή [3], διαχωρίζουμε τους μηχανισμούς υπολογισμού επιρροής που αναφέρονται στη μελέτη μας και καταγράφουμε τα χαρακτηριστικά τους σύμφωνα με τα κριτήρια του. Τέλος παρουσιάζουμε τα πιο δημοφιλή εμπορικά εργαλεία υπολογισμού επιρροής που κυκλοφορούν στο διαδίκτυο. Περιγράφουμε τον τρόπο υπολογισμού της επιρροής καθώς και τον τρόπο απεικόνισης του αποτελέσματος στον τελικό χρήστη. Στο 2^ο κεφάλαιο αναφερόμαστε στις διαφορετικές έννοιες της επιρροής και της φήμης όπως εμφανίζονται στα κοινωνικά δίκτυα και στα κριτήρια που χρησιμοποιεί ο ταξινομητής μας. Στο 3^ο μέρος της εργασίας παραθέτουμε λεπτομερή ανάλυση του τρόπου λειτουργίας του κάθε συστήματος ξεχωριστά και στο 4^ο πραγματοποιείται η ταξινόμηση αυτών. Στο 5^ο μέρος καταγράφουμε και περιγράφουμε τους πιο δημοφιλείς και αποδοτικούς μηχανισμούς υπολογισμού δημοτικότητας και τέλος στο 6^ο μέρος εξάγουμε τα συμπεράσματά μας.

2. ΣΥΣΤΗΜΑΤΑ ΥΠΟΛΟΓΙΣΜΟΥ ΕΠΙΡΡΟΗΣ ΣΕ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ

Στο παρακάτω κεφάλαιο θα παρουσιάσουμε και θα αναλύσουμε τις διαφορετικές οπτικές της επιρροής και τον τρόπο που αντιμετωπίζεται από τους εκάστοτε μηχανισμούς υπολογισμού της. Επίσης, στα πλαίσια της ταξινόμησης των μηχανισμών αυτών θα παρουσιάσουμε τα κριτήρια με βάση τα οποία γίνεται ο διαχωρισμός τους.

2.1 Φήμη, επιρροή και δημοτικότητα σε κοινωνικά δίκτυα

Οι έννοιες φήμη, επιρροή και δημοτικότητα στα κοινωνικά δίκτυα αποτελούν παρόμοιες περιγραφές του πιο σημαντικού στοιχείου που ερευνάτε τα τελευταία χρόνια στο επιστημονικό πεδίο των ψηφιακών κοινωνικών δικτύων και αυτό είναι η ικανότητα που έχουν κάποιοι από τους χρήστες του να καθοδηγούν την κοινή γνώμη και τις ενέργειες των υπολοίπων χρηστών. Ανεξαρτήτως της φύσεως του κοινωνικού δικτύου, οι χρήστες με τη μεγαλύτερη επιρροή είναι πάντα μία πολύ χρήσιμη ομάδα στο δίκτυο αυτό καθώς μπορούν να επηρεάσουν την ευρεία μάζα του όσων αφορά διαφημιστικές καμπάνιες, προώθηση προϊόντων ή υπηρεσιών, κτλ.

Σύμφωνα με την πολύ εμπειριστατωμένη μελέτη [4] η ανάλυση των κοινωνικών δικτύων υπήρξε ένας σημαντικός τομέας έρευνας των κοινωνιολόγων για πολλά χρόνια και προσφάτως έχει προσελκύσει το ενδιαφέρον της ερευνητικής κοινότητας η ηλεκτρονική διάσταση που έχουν πάρει. Η εξάπλωση του παγκόσμιου διαδικτυακού ιστού (WWW) και η τεράστια αύξηση της χρήσης κοινωνικών δικτύων όλων των τύπων (ιστότοποι για κοινωνικοποίηση, ιστότοποι αξιολόγησης υπηρεσιών και προϊόντων, τα ηλεκτρονικά ημερολόγια κ.λπ..) έχουν μετατρέψει τους χρήστες του διαδικτύου από καταναλωτές πληροφοριών σε παραγωγούς, δημιουργώντας έτσι για τους ερευνητές ένα τεράστιο όγκο πληροφοριών και περιεχομένου δημιουργημένο από τους εκάστοτε χρήστες. Ένα τέτοιο περιεχόμενο παρουσιάζει χαρακτηριστικά, τα οποία είναι μοναδικά στο διαδίκτυο, από την άποψη της κοινής συγγραφικής ταυτότητας, πλήθος ετικετών-χαρακτηρισμών που παρέχονται από τον χρήστη και μέσω αυτού του περιεχομένου μπορεί να παραχθεί σημαντική πληροφορία σχετικά με τη συνδεσιμότητα μεταξύ των χρηστών. Όλα αυτά τα χαρακτηριστικά παρέχουν μια πλατφόρμα από την οποία μπορούν να εξαχθούν ενδιαφέρουσες πληροφορίες σχετικά με τις αλληλεπιδράσεις των χρηστών μεταξύ του δικτύου.

Ένας κοινός τύπος ανάλυσης της επιρροής είναι αρχικά η ανίχνευση ομάδων χρηστών με παρόμοια ενδιαφέροντα, και στη συνέχεια, μέσα σε αυτές τις ομάδες, γίνεται αναζήτηση για τους χρήστες που ασκούν τη μεγαλύτερη επιρροή. Ως χρήστες μεγάλης επιρροής μπορούμε να ορίσουμε τα άτομα με πολλές συνδέσεις μέσα στην κοινότητα, αλλά γενικά είναι δυνατοί και άλλοι ορισμοί ανάλογα με τον τύπο του κοινωνικού δικτύου. Οι χρήστες μεγάλης επιρροής λειτουργού ως «πηγές» ή «διαμοιραστές» πληροφορίας μέσα στην κοινότητά τους και επομένως διαδραματίζουν βασικό ρόλο στη διάδοση πληροφοριών και μπορούν να παίξουν πολύ σημαντικό ρόλο στην προώθηση και την υποστήριξη νέων προϊόντων ή ιδεών. Για εμπορικά και διαφημιστικά τμήματα των εταιριών είναι πολύ σημαντική η εύρεση και χρήση των χρηστών μεγάλης επιρροής, καθώς οι απόψεις που εκφράζουν μπορούν να διαδοθούν γρήγορα και σε μεγάλο εύρος κοινού, με μικρή προσπάθεια.

Για να κατανοήσουμε τη σημαντικότητα των χρηστών μεγάλης επιρροής αρκεί να αναφέρουμε τα παρακάτω αποτελέσματα ερευνών σχετικά με τη λήψη αποφάσεων και αναζήτησης συμβουλών ανθρώπων προτού προχωρήσουν σε κάποια ενέργεια. Σύμφωνα με το [2] στο φυσικό κόσμο, το 83% των ανθρώπων προτιμούν να πάρουν κάποια συμβουλή από την οικογένεια, φίλους ή έναν ειδικό, παρά να εμπιστευτούν τα παραδοσιακά μέσα διαφήμισης πριν δοκιμάσουν ένα νέο εστιατόριο. Επίσης το 71% των ανθρώπων κάνει το ίδιο πριν αγοράσει ένα φάρμακο ή επισκεφθεί κάποιο μέρος και το 61% των ανθρώπων ζητούν τη γνώμη ενός οικογενειακού μέλους ή φίλου πριν παρακολουθήσει μια ταινία. Εν ολίγοις, οι άνθρωποι πριν πάρουν αποφάσεις, μιλάνε, και ακούν την εμπειρία, τις γνώμες και τις προτάσεις κάποιου άλλου ανθρώπου που εμπιστεύονται. Αυτοί οι άνθρωποι που ασκούν μεγάλη επιρροή ονομάζονται «έμπιστοι» και όσο πιο έμπιστος θεωρείται κάποιος, τόσο πιο μεγάλη επιρροή ασκεί στους υπόλοιπους.

2.1.1 Ταξινόμηση συστημάτων φήμης σε κοινωνικά δίκτυα

Υπάρχουν πολλοί διαφορετικοί τύποι μηχανισμών υπολογισμού επιρροής στα κοινωνικά δίκτυα και λόγω των διαφορετικών χαρακτηριστικών του κάθε κοινωνικού δικτύου, υπάρχουν αντίστοιχες διαφοροποιήσεις στο σχεδιασμό των μηχανισμών υπολογισμού επιρροής. Ανάλογα με το είδος του κοινωνικού δικτύου μπορεί να υπάρχουν διάφοροι αλληλένδετοι στόχοι. Για παράδειγμα, στα κοινωνικά δίκτυα με αντικείμενο την πώληση ή ανταλλαγή προϊόντων, ο αντίστοιχος μηχανισμός υπολογισμού φήμης θα πρέπει να λαμβάνει σοβαρά υπόψη την έντιμη συμπεριφορά των χρηστών. Στα κοινωνικά δίκτυα διαμοιρασμού περιεχομένου ο στόχος του μηχανισμού είναι η εύρεση του χρήστη με το καλύτερο παραγόμενο περιεχόμενο. Στα ηλεκτρονικά δίκτυα κοινωνικής δικτύωσης ο στόχος είναι να διευκολύνεται η αντιστοίχιση των μελών με κοινά ενδιαφέροντα, δηλαδή να βοηθούνται τα μέλη να βρίσκουν άλλα άτομα τα οποία μπορούν να εμπιστευτούν για να μοιράζονται πληροφορίες μαζί τους, ενώ στους ιστότοπους αξιολόγησης προϊόντων και υπηρεσιών, στόχος είναι να βοηθήσει τους χρήστες να βρουν αξιόπιστες κριτικές για τα αντικείμενα που τους ενδιαφέρουν. Ο γενικός στόχος θα μπορούσε να είναι να βρεθούν οι πιο σημαντικοί χρήστες, δηλαδή οι χρήστες των οποίων οι ενέργειες μπορούν να ενεργοποιήσουν παρόμοιες ενέργειες σε άλλους χρήστες. Είναι σημαντικό για έναν σχεδιαστή μηχανισμού υπολογισμού φήμης να καθορίσει τους βασικούς στόχους του και επίσης να κατανοήσει τα βασικά χαρακτηριστικά του κοινωνικού δικτύου ώστε τα αποτελέσματα του μηχανισμού να έχουν ακρίβεια. Σύμφωνα με την εργασία [3] έχουν εντοπιστεί τα ακόλουθα βασικά χαρακτηριστικά:

- **Ο στόχος της επιρροής.** Η επιρροή θα μπορούσε να αναφέρεται σε έναν χρήστη ενός κοινωνικού δικτύου, π.χ. σε έναν χρήστη κοινωνικού δικτύου διαμοιρασμού περιεχομένου σχετικά με τη συνολική αξία του περιεχομένου που ανεβάζει, καθώς και θα μπορούσε επίσης να αναφέρεται στο περιεχόμενο που παράγει ο χρήστης σε αυτό, π.χ. μια φωτογραφία σε ένα κοινωνικό δίκτυο διαμοιρασμού φωτογραφιών μπορεί να αξιολογηθεί επίσης ανάλογα με τα ποιοτικά χαρακτηριστικά της αλλά και την κινητικότητα που προκαλεί. Σε ορισμένες περιπτώσεις, η επιρροή μπορεί να εκτιμηθεί τόσο για χρήστες όσο και για περιεχόμενο, όπως στον ιστότοπο αξιολόγησης Yelp [9], ο οποίος δημοσιεύει υπολογισμούς επιρροής τόσο για τα μέλη του που παράγουν τις αξιολογήσεις όσο και για τις κριτικές.

- **Το είδος των πληροφοριών που χρησιμοποιούνται για τον υπολογισμό της επιρροής.** Οι πληροφορίες αυτές μπορούν να αναφέρονται είτε στο ίδιο το κοινωνικό δίκτυο, δηλαδή στους χρήστες του και στα είδη διασύνδεσης μεταξύ τους (π.χ. φιλία) είτε σε μορφές εμπιστοσύνης που μπορούν να ανατεθούν από τα μέλη μεταξύ τους είτε πληροφορίες προφίλ του χρήστη, οι οποίες περιλαμβάνουν προσωπικά δεδομένα, βαθμολογίες και σχόλια του χρήστη, στατιστικά στοιχεία σχετικά με τη δραστηριότητα του, κ.λπ.
- **Μέθοδος αξιοποίησης πληροφοριών για τον υπολογισμό της επιρροής.** Συνήθως αναφέρεται στον αλγόριθμος που χρησιμοποιείται για τον υπολογισμό της επιρροής.
- **Μέθοδος εμφάνισης επιρροής:** Οι μέθοδοι παρουσίασης της επιρροής ανήκουν στις παρακάτω ακόλουθες κατηγορίες:
 - Στατιστικά στοιχεία κοινωνικών δικτύων. Τα πιο συνήθη στατιστικά στοιχεία που χρησιμοποιούνται είναι ο αριθμός των αξιολογήσεων ή ο αριθμός των επιβραβεύσεων - “Likes”. Το πλεονέκτημα της χρήσης στατιστικών στοιχείων απευθείας από τα κοινωνικά δίκτυα είναι ότι ο μηχανισμός υπολογισμού επιρροής μπορεί να τα ερμηνεύσει σύμφωνα με τη δικιά του σκοπιά και να παράγει τα δικά του συμπεράσματα. Από την άλλη πλευρά, ο σχεδιαστής του μηχανισμού θα πρέπει να είναι αρκετά εξοικειωμένος με το κοινωνικό δίκτυο που μελετάει ώστε να χρησιμοποιήσει σωστά τα κατάλληλα στατιστικά στοιχεία.
 - Βαθμολογίες και τίτλοι. Ανάλογα με το κοινωνικό δίκτυο, η απεικόνιση της κατάταξης των χρηστών σύμφωνα με την επιρροή τους μπορεί να γίνει με βαθμολογία με αριθμό αστεριών, αριθμητικές βαθμολογίες, αριθμημένα επίπεδα ή ονομαστικές βαθμίδες μελών ή ανάθεση συγκεκριμένων επιτευγμάτων. Η χρήση κάποιας μορφής ρητής βαθμολογίας, όπως οι αξιολογήσεις με αστέρια, γνωστοποιεί στους χρήστες αν η απόδοση κάποιου σε κάποιο πεδίο είναι “καλή” ή “κακή”. Αυτή η μέθοδος προβολής βοηθά τους χρήστες να κατανοούν πιο εύκολα το επίπεδο ενός χρήστη.
 - Κατάταξη χρηστών ή περιεχομένου σε λίστα. Η ταξινόμηση των χρηστών σε λίστες προσφέρει την άμεση σύγκριση τους και έτσι αποκτούν κίνητρο ώστε να αυξήσουν την επιρροή τους. Από την άλλη πλευρά, η άμεση αυτή σύγκριση οδηγεί σε έξαρση ανταγωνισμού που μπορεί να δημιουργήσει προβλήματα στη συμπεριφορά των χρηστών. Ορισμένοι χρήστες ενδέχεται να προκαλέσουν χειραγωγική συμπεριφορά για να αυξήσουν την κατάταξή τους. Άλλοι χρήστες που δεν φτάνουν στις κορυφαίες θέσεις, ενδέχεται να αισθάνονται ότι δεν έχουν κίνητρο να συνεχίσουν να συμμετέχουν στην κοινότητα και να σταματήσουν να χρησιμοποιούν την εφαρμογή. Επομένως, η κατάταξη σε λίστες και βαθμολογίες, ενδέχεται να επηρεάσει τη συμπεριφορά ενός χρήστη.
- **Εύρος υπολογισμού της επιρροής.** Το εύρος υπολογισμού της επιρροής θεωρείται είτε γενικό, όταν ένας ή περισσότεροι δείκτες επιρροής υπολογίζονται και παρουσιάζονται στους χρήστες χωρίς να εξαρτώνται από τις προτιμήσεις του χρήστη είτε εξατομικευμένο, όπου οι δείκτες επιρροής σχετικά με ένα χρήστη εξαρτώνται από τις προτιμήσεις του.

Χρησιμοποιήσαμε τα παραπάνω στοιχεία ως τα χαρακτηριστικά μιας προτεινόμενης ταξινόμησης των συστημάτων υπολογισμού επιρροής κοινωνικών δικτύων [3]. Η ταξινόμηση εφαρμόστηκε στους αλγόριθμους υπολογισμού επιρροής που εξετάζουμε παρακάτω καθώς

και στους ήδη υπάρχοντες μηχανισμούς που έχουν υλοποιηθεί και τα αποτελέσματά απεικονίζονται με τη μορφή πίνακα.

2.2 Συστήματα Επιρροής

Ο υπολογισμός της επιρροής είναι μια διαδικασία που ενώ μπορεί να χρειάζεται διαφορετική υλοποίηση ανάλογα με το είδος του κοινωνικού δικτύου, βασίζεται συνήθως σε δύο μεθόδους. Στην πρώτη μέθοδο μεταφράζεται το κοινωνικό δίκτυο με τους χρήστες του σε διάγραμμα τύπου γράφου όπου κάθε κόμβος, συνήθως συμβολίζει το χρήστη και οι ακμές του συμβολίζουν τις σχέσεις εμπιστοσύνης μεταξύ τους. Δηλαδή αν κάποιος χρήστης έχει επηρεαστεί από κάποιον άλλο, τότε συνδέονται με μία ακμή και ονομάζονται γειτονικοί χρήστες. Πολλές φορές οι ακμές χαρακτηρίζονται από έναν αριθμό που ονομάζεται «βάρος» της ακμής και μπορεί να συμβολίζει την πιθανότητα επιρροής που υπάρχει μεταξύ των χρηστών [1] ή τη χρονική στιγμή που δημιουργήθηκε, δηλαδή τη στιγμή που επηρεάστηκε ο χρήστης [6]. Για τη μεταφορά του κοινωνικού δικτύου σε γράφο χρησιμοποιείται ένα ημερολόγιο των κινήσεων των χρηστών [1] [6]. Το ημερολόγιο αυτό είναι ένας πίνακας, ο οποίος περιέχει τριάδες στοιχείων και περιγράφει τη χρονική στιγμή που επηρεάστηκε ένας χρήστης (εκτέλεσε μία ενέργεια) και από ποιον γειτονικό χρήστη προήλθε η επιρροή. Υπάρχουν μηχανισμοί [4] [5] που εμβαθύνουν στη μελέτη του γράφου, για τον υπολογισμό της δημοτικότητας ενός χρήστη ή τον υπολογισμό των πιο δημοφιλών χρηστών και για να το πετύχουν αυτό εξετάζουν μεγέθη όπως είναι ο βαθμός ενός κόμβου, δηλαδή ο αριθμός των ακμών που ξεκινούν ή τελειώνουν στον κόμβο αυτό, το μέσο μήκος μονοπατιού του γράφου και η ευκολία πρόσβασης ενός κόμβου. Η επιτυχία αυτής της μεθόδου έγκειται στο συνδυασμό πολλών διαφορετικών μετρήσεων και παρακολούθησης του γράφου σε βάθος χρόνου. Μάλιστα σύμφωνα με το [6] εξάγεται το συμπέρασμα ότι όσο περνάει ο χρόνος από τη στιγμή που γίνεται κάποια ενέργεια, τόσο περισσότερο μειώνεται η πιθανότητα να επηρεαστεί κάποιος γειτονικός χρήστης. Τέλος, πάλι από την μελέτη του γράφου του κοινωνικού δικτύου η έρευνα [1] πρωτοπορεί, υπολογίζοντας και την αρνητική επιρροή που μπορεί να ασκήσει ένας χρήστης, καθώς η επιρροή του λειτουργεί αντίστροφα σε χρήστες που δεν έχουν σχέση εμπιστοσύνης.

Η δεύτερη, και ίσως πιο διαδεδομένη, μέθοδος υπολογισμού της δημοτικότητας στα κοινωνικά δίκτυα, βασίζεται στην ανάλυση των στατιστικών στοιχείων τους. Πέρα από τα ξεχωριστά χαρακτηριστικά που έχει το κάθε ένα, υπάρχουν κάποια κοινά στοιχεία που διέπουν όλα τα κοινωνικά δίκτυα με παρόμοιο περιεχόμενο. Για παράδειγμα στα κοινωνικά δίκτυα ανταλλαγής απόψεων, όπως είναι το Twitter ή τα κοινωνικά δίκτυα τύπου blog, ο υπολογισμός της επιρροής γίνεται εξετάζοντας κάθε δημοσίευση ξεχωριστά [2] [3] [4] [7] [8]. Τα στοιχεία της κάθε δημοσίευσης που λαμβάνονται υπόψιν είναι ο αριθμός των αναδημοσιεύσεων, ο αριθμός των “like”, ο αριθμός των αναφορών σε άλλες πηγές (Outlinks) και ο αριθμός των φορών που αποτελεί πηγή ή εξεταζόμενη δημοσίευση (Inlinks) καθώς και η λεκτική έκτασή της, θεωρώντας συνήθως ότι όσο εκτενέστερη είναι τόσο πιο αυθεντική είναι και μπορεί να επηρεάσει τους υπόλοιπους χρήστες. Παρόλο που τα παραπάνω κριτήρια μπορεί να θεωρηθεί ότι εφαρμόζονται μόνο σε κοινωνικά δίκτυα λεκτικού περιεχομένου (Twitter, blogs, etc.), με συγκεκριμένες αναπροσαρμογές μπορούν να υπολογίσουν την δημοτικότητα σε διαφορετικού τύπου κοινωνικά δίκτυα.

Χαρακτηριστική είναι η μελέτη [5] όπου γίνεται ανάλυση ενός ακαδημαϊκού κοινωνικού δικτύου. Ένα ακαδημαϊκό κοινωνικό δίκτυο αναπαριστά τις σχέσεις που έχουν μεταξύ τους οι διάφορες επιστημονικές δημοσιεύσεις είτε σε επίπεδο θεματολογίας είτε σε επίπεδο συγγραφικής ομάδας. Έτσι λοιπόν, οι αναφορές σε άλλες πηγές (Outlinks) αποτελούν τη βιβλιογραφία μίας δημοσίευσης και αντίστοιχα ο αριθμός των φορών που η εξεταζόμενη δημοσίευση βρίσκεται σε βιβλιογραφία άλλης έρευνα αποτελεί τον αριθμό Inlinks που την χαρακτηρίζει.

Μία πρωτότυπη και γενικευμένη εξέταση της δημοτικότητας γίνεται στην δημοσίευση [7] όπου συνδυάζονται τα στοιχεία κοινωνικού δικτύου αλλά και περιεχομένου ώστε να εξαχθεί καλύτερο αποτέλεσμα. Ειδικότερα η μελέτη γίνεται πάνω στο κοινωνικό δίκτυο Flickr, το οποίο έχει ως περιεχόμενο τις φωτογραφίες που δημοσιεύουν οι χρήστες του. Από πλευράς κοινωνικού δικτύου τα στοιχεία που εξετάζονται είναι ο αριθμός των «φίλων» του χρήστη, ο αριθμός των κατηγοριών που προσθέτει στην φωτογραφία του (hashtags) αλλά και η περιγραφή που της προσθέτει, η οποία εξετάζεται παρόμοια με την λεκτική έκταση που αναφέραμε παραπάνω. Όσον αφορά τη μελέτη του περιεχομένου της εικόνας λαμβάνονται υπόψιν φωτογραφικά τεχνικά χαρακτηριστικά όπως είναι η υφή των χρωμάτων, το φόντο της, ο αριθμός των εμφανιζόμενων αντικειμένων κ.α. τα οποία μεταφράζονται σε αριθμούς και σε συνδυασμό με τα παραπάνω κριτήρια κοινωνικού δικτύου γίνεται ο τελικός υπολογισμός της δημοτικότητας μίας φωτογραφίας. Το μοντέλο της συγκεκριμένης έρευνας εξελίσσει τη μελέτη υπολογισμού της επιρροής καθώς παρέχει και τη δυνατότητα πρόβλεψης επιρροής χωρίς να έχει γίνει δημοσίευση μίας φωτογραφίας.

3. ΑΝΑΛΥΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΣΥΣΤΗΜΑΤΩΝ ΕΠΙΡΡΟΗΣ ΓΙΑ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ

Στο κεφάλαιο αυτό περιγράφουμε ένα σύνολο συστημάτων επιρροής που συναντάμε στη βιβλιογραφία και συγκεκριμένα στις εξής εργασίες: (α) Εύρεση Επιδραστικών Κόμβων σε ένα Δίκτυο Εμπιστοσύνης [1], (β) Εύρεση Επιδραστικών Μπλόγερ (Bloggers) σε ένα “Κοινωνικό Ιστολόγιο” [2], (γ) Μηχανισμοί Φήμης σε Κοινωνικά Δίκτυα – Η περίπτωση ενός Συστήματος Υπολογισμού Επιρροής στο Twitter [3], (δ) Εφαρμογή Μετρικών Κοινωνικών Δικτύων σε ένα Συνεργατικό Μοντέλο Εμπιστοσύνης για τη Δημιουργία Προσωποποιημένων Συστάσεων Χρηστών [4], (ε) Εύρεση Επιδραστικών Χρηστών σε Κοινωνικά Δίκτυα [5], (στ) Εκμάθηση Πιθανοτήτων Επιρροής στα Κοινωνικά Δίκτυα [6], (ζ) Στοιχεία καθορισμού μιας εικόνας ως δημοφιλούς [7] και (η) Μέτρηση Επιρροής στο Twitter [8]. Για κάθε ένα από τα συστήματα αυτά παραθέτουμε μια σύντομη περιγραφή που περιλαμβάνει την ακολουθούμενη προσέγγιση και τα σχετικά συμπεράσματα των συγγραφέων.

3.1 Εύρεση Επιδραστικών Κόμβων σε ένα Δίκτυο Εμπιστοσύνης [1]

Η έννοια της επιρροής από αυτό το σύστημα, αναλύεται ως προς τις ενέργειες που κάνει ένας χρήστης, επηρεαζόμενος από τους γειτονικούς του χρήστες, μελετάει τον τρόπο με τον οποίο γίνεται η διάχυση της επιρροής στους χρήστες του δικτύου και ως στόχο έχει την εύρεση των κόμβων που ασκούν τη μέγιστη επιρροή στο δίκτυο. Το κοινωνικό δίκτυο απεικονίζεται ως ένας γράφος του οποίου οι κόμβοι αναπαριστούν τους χρήστες του και οι ακμές του αναπαριστούν τις σχέσεις εμπιστοσύνης μεταξύ τους. Οι σχέσεις εμπιστοσύνης μπορεί να είναι θετικές αλλά και αρνητικές με αποτέλεσμα να υπάρχει η ανάγκη του υπολογισμού δύο πιθανοτήτων, της πιθανότητας επηρεασμού του χρήστη και η αντίθετη. Το προτεινόμενο μοντέλο (Trust-Influential Node Miner) λειτουργεί σε δύο στάδια, όπως θα δούμε παρακάτω.

Στο πρώτο στάδιο υπολογίζονται οι πιθανότητες επιρροής του κάθε κόμβου ξεχωριστά. Ο υπολογισμός των πιθανοτήτων αυτών εξάγεται από δύο πίνακες, ο ένας (πίνακας A) περιέχει πληροφορία για τη σχέση εμπιστοσύνης μεταξύ των χρηστών και ο δεύτερος (πίνακας B) περιέχει πληροφορίες για τις ενέργειες που κάνουν οι χρήστες. Ο πίνακας A περιέχει τριάδες από δεδομένα (u,v,trust), όπου u, v είναι μέλη του κοινωνικού δικτύου, το στοιχείο trust δείχνει τη σχέση εμπιστοσύνης μεταξύ των χρηστών (1 σε περίπτωση που υπάρχει εμπιστοσύνη μεταξύ τους και -1 στην αντίθετη περίπτωση) και από τα δεδομένα αυτά σχεδιάζεται ο γράφος του κοινωνικού δικτύου. Κάθε γραμμή του πίνακα αντιπροσωπεύει μία ακμή του γράφου καθώς μπορεί να προκαλέσει επιρροή μεταξύ των κόμβων (είτε θετική είτε αρνητική).

Ο πίνακας B περιέχει τριάδες από δεδομένα (Action,User,Time) και περιγράφει τις πράξεις των χρηστών καθώς και την χρονική στιγμή που πραγματοποιήθηκαν. Από τον πίνακα B γίνεται ο υπολογισμός των πιθανοτήτων επιρροής ελέγχοντας το σύνολο των αριθμό των πράξεων ενός χρήστη που έχουν συμβεί, αφότου οι ίδιες πράξεις έχουν συμβεί από τους έμπιστους χρήστες του κοινωνικού δικτύου. Αντίστοιχα υπολογίζεται και η πιθανότητα μη-επιρροής, ελέγχοντας το σύνολο των πράξεων που δεν έχουν γίνει από ένα χρήστη, αφότου οι ίδιες πράξεις έχουν συμβεί από τους μη-έμπιστους γειτονικούς χρήστες. Ο υπολογιστικός τύπος των παραπάνω πιθανοτήτων είναι το πηλίκο των παραπάνω αναφερθέντων

συνόλων, προς το συνολικό αριθμό πράξεων του χρήστη. Με τον υπολογισμό των πιθανοτήτων επιρροής ολοκληρώνεται το πρώτο στάδιο του συστήματος.

Στο δεύτερο στάδιο του συστήματος γίνεται η επιλογή των χρηστών που ασκούν τη μεγαλύτερη επιρροή στο κοινωνικό δίκτυο εφαρμόζοντας τον αλγόριθμο mineSeedLS. Ο αλγόριθμος δέχεται σαν είσοδο το γράφο του κοινωνικού δικτύου και τις πιθανότητες επιρροής που υπολογίσαμε στο προηγούμενο στάδιο καθώς και τον αριθμό των χρηστών-κόμβων που θέλουμε να εξάγουμε. Σε αυτό το σημείο είναι αναγκαίος ο υπολογισμός της συνδυασμένης πιθανότητας επιρροής ενός χρήστη, η οποία υπολογίζεται συνδυάζοντας τις πιθανότητες επιρροής των γειτονικών κόμβων του χρήστη, που έχουν κάνει ήδη την ίδια πράξη (είναι δηλαδή ενεργοί χρήστες). Ο αλγόριθμος του συστήματος, για κάθε κόμβο ξεχωριστά, υπολογίζει την συνολική διασπορά επιρροής και προχωράει αναλόγως. Ειδικότερα αν ο υπό εξέταση κόμβος αυξάνει την διασπορά επιρροής τότε προστίθεται στο σύνολο των κόμβων επιρροής ενώ στην αντίθετη περίπτωση αφαιρείται. Επίσης υπάρχει και η περίπτωση αντικατάστασης ενώ κόμβου που βρίσκεται στο σύνολο των κόμβων επιρροής από κάποιο νέο κόμβο που αυξάνει την διασπορά επιρροής. Η εκτέλεση του αλγόριθμου θα τερματιστεί όταν ελεγχθούν όλοι οι κόμβοι λαμβάνοντας υπόψιν τον αριθμό των κόμβων που θέλουμε να εξάγουμε αλλά και το ποιοι από τους κόμβους του δικτύου προκαλούν τη μεγαλύτερη επιρροή στους γειτονικούς κόμβους.

3.2. Εύρεση Επιδραστικών Μπλόγгер (Bloggers) σε ένα “Κοινωνικό Ιστολόγιο” [2]

Η μελέτη αυτή εστιάζει στα κοινωνικά δίκτυα τύπου blog και ως στόχο έχει την εύρεση των χρηστών που ασκούν τη μεγαλύτερη επιρροή στους υπολοίπους. Από τα δύο είδη blog που υπάρχουν, τα προσωπικά, στα οποία μόνο ο ιδιοκτήτης του blog μπορεί να ξεκινήσει συζητήσεις και οι υπόλοιποι χρήστες μπορούν μόνο να σχολιάσουν, και τα κοινωνικά, στα οποία ο κάθε χρήστης ξεκινάει δικές του συζητήσεις και σχολιασμούς, η μελέτη ασχολείται μόνο με την δεύτερη κατηγορία καθώς η χρήστες τους έχουν πιο ενεργή συμμετοχή και υπάρχει μεγαλύτερο ενδιαφέρον για την εύρεση των πιο πειστικών – έμπιστων χρηστών. Το προτεινόμενο μοντέλο λαμβάνει υπόψιν διάφορα στατιστικά στοιχεία για τα σχόλια των χρηστών, τους βαθμολογεί αναλόγως και στο τέλος εξάγει τους πιο έμπιστους.

Καταρχήν για να θεωρηθεί κάποιος χρήστης πειστικός θα πρέπει να έχει γράψει τουλάχιστον ένα πειστικό σχόλιο. Για κάθε σχόλιο ενός χρήστη υπολογίζεται το ποσό (score) πειστικότητας του σχολίου και το μεγαλύτερο ποσό πειστικότητας από όλα τα σχόλια του χρήστη αντιστοιχεί στον Δείκτη Πειστικότητάς του (iIndex). Ταξινομώντας κατά φθίνουσα σειρά τους Δείκτες Πειστικότητας των χρηστών μπορούμε εύκολα να εξάγουμε τους πιο πειστικούς από αυτούς και να τους ξεχωρίσουμε. Παρακάτω θα περιγράψουμε τον τρόπο με τον οποίο γίνεται ο υπολογισμός του ποσού πειστικότητας του κάθε σχολίου ξεχωριστά.

Τα στοιχεία που είναι απαραίτητα για τον υπολογισμό του ποσού πειστικότητας παρουσιάζονται παρακάτω και είναι σημαντικό να σημειώσουμε ότι πρέπει να υπολογίζονται συνδυαστικά και όχι μεμονωμένα:

- Αναγνώριση : Η αναγνώριση ενός σχολίου μπορεί να υπολογισθεί από τον αριθμό i των φορών που έχει αναφερθεί σε άλλα σχόλια (αριθμός inlinks). Όσες πιο πολλές

φορές έχει αναφερθεί το σχόλιο, τόσο πιο έμπιστο θεωρείται από τους υπόλοιπους χρήστες του δικτύου.

- **Παραγόμενη δραστηριότητα:** Ο αριθμός των σχολίων και των νέων συζητήσεων που μπορεί να δημιουργήσει ένα post αποτελεί βασικό κριτήριο για το score επιρροής του post. Ένας μεγάλος αριθμός σχολίων (συμβολίζεται με το γράμμα γ) αποδεικνύει ότι πολλοί χρήστες αφιέρωσαν χρόνο να διαβάσουν αλλά και να σχολιάσουν το post αυτό και ως αποτέλεσμα μπορεί να θεωρηθεί ως έμπιστο post και κατ' επέκταση έμπιστος ο χρήστης που το δημιούργησε.
- **Αυθεντικότητα:** Η πρωτοτυπία ενός post του προσδίδει κύρος και μπορεί να καταστήσει των ιδιοκτήτη του ως πιστικό χρήστη. Το ποσό πρωτοτυπίας ενός post, αντιστοιχεί στον αριθμό αναφορών σε άλλα post (outlinks θ). Ειδικότερα όσες περισσότερες αναφορές σε άλλα post υπάρχουν στο κείμενο, τόσο περισσότερο βασίζεται σε περιεχόμενο άλλου χρήστη άρα είναι λιγότερο αυθεντικό και δε μπορεί να ασκήσει μεγάλη επιρροή στους υπόλοιπους χρήστες.
- **Ευγλωττία:** Ο τρόπος γραφής ενός σχολίου και κυρίως η έκτασή του παίζουν σημαντικό ρόλο για την επιρροή που μπορεί να ασκήσει. Σαν στατιστικό στοιχείο είναι δύσκολο να εκφραστεί με νούμερο, και στη μελέτη αυτή ορίζεται σαν το μήκος που έχει (αριθμός λέξεων και γραμμάτων) και συμβολίζεται ως λ .

Για τον υπολογισμό του ποσού επιρροής του κάθε post θεωρούνται ως βασικότερα από τα παραπάνω στοιχεία οι αριθμοί των inlinks και των outlinks. Ειδικότερα ορίζεται το ποσό InfluenceFlow(p) του post p το οποίο υπολογίζει τη διαφορά της επιρροής όλων των inlinks από τη συνολική διαφορά της επιρροής όλων των outlinks. Ο αριθμός των σχολίων και των νέων συζητήσεων που δημιουργούνται γύρω από το post p , αθροίζονται με το InfluenceFlow(p) και νέο σύνολο προστίθεται σαν συντελεστής το ποσοστό ευγλωττίας που περιγράψαμε νωρίτερα. Αξίζει να σημειωθεί ότι χρησιμοποιήθηκε το προσθετικό μοντέλο γιατί μπορεί να αναδείξει τη συνδυασμένη αξία του κάθε δεδομένου. Επίσης να σημειωθεί ότι κάθε δεδομένο χρησιμοποιείται και με ξεχωριστό δείκτη βαρύτητας, αναλόγως με το πιο στοιχείο θέλουμε να θεωρήσουμε σημαντικότερο. Επαναλαμβάνοντας την παραπάνω διαδικασία για κάθε σχόλιο ενός χρήστη, μπορούμε να βρούμε το σχόλιο με το μεγαλύτερο score, το οποίο και αποτελεί τον δείκτη επιρροής του χρήστη.

Τα αποτελέσματα του μοντέλου σε διάφορα πειράματα καταλήγουν στο ότι δεν είναι αναγκαίο για έναν έμπιστο χρήστη να είναι και ενεργός, καθώς μπορεί να έχει αραιή συμμετοχή στο blog, αλλά τα σχόλια του να επηρεάζουν πολλούς χρήστες. Έτσι χωρίζει τους χρήστες σε ενεργούς και έμπιστους, ενεργούς και μη έμπιστους, μη ενεργούς και έμπιστους, μη ενεργούς και μη έμπιστους και νέους χρήστες.

3.3 Μηχανισμοί Φήμης σε Κοινωνικά Δίκτυα – Η περίπτωση ενός Συστήματος Υπολογισμού Επιρροής στο Twitter [3]

Η έρευνα αυτή εστιάζει στο κοινωνικό δίκτυο Twitter και ως στόχο έχει την εύρεση των χρηστών που ασκούν τη μεγαλύτερη επιρροή αλλά και των θεματικών ενοτήτων που επηρεάζουν περισσότερο τους χρήστες του. Το Twitter είναι ένα είδος microblogging κοινωνικό δίκτυο στο οποίο οι χρήστες του δημοσιεύουν σε μορφή κειμένου μικρής σχετικά έκτασης (μέγιστη έκταση 140 χαρακτήρες) σε συνδυασμό με φωτογραφίες ή βίντεο τις απόψεις τους για διάφορα θέματα. Για τον υπολογισμό της επιρροής ορίζεται το «Ποσό

Επιρροής-Influence score $I(t)$ » και ο υπολογισμός αυτού του μεγέθους χρειάζεται σα δεδομένα κάποια χαρακτηριστικά του κάθε σχολίου-tweet.

Τα τέσσερα βασικά στοιχεία που χρησιμοποιούνται για τον υπολογισμό της επιρροής ενός tweet t , είναι ο αριθμός i των inlinks, δηλαδή ο αριθμός των φορών που άλλα tweets αναφέρονται στο t (στην ορολογία του Twitter είναι γνωστό ως retweet), ο αριθμός γ των φορών που άρεσε το t κάποιο χρήστη (στο Twitter απεικονίζεται με μία εικόνα καρδιάς), ο αριθμός θ των outlinks, δηλαδή ο αριθμός των φορών που το t αναφέρεται σε άλλα tweets, που δείχνει την καινοτομία και την αυθεντικότητα ενός tweet (όσο μεγαλύτερος είναι ο αριθμός θ τόσο λιγότερο αυθεντικό είναι το tweet) και τέλος ο αριθμός λ ο οποίος δηλώνει το μέγεθος του tweet και θεωρείται ότι όσο μεγαλύτερο είναι το λ τόσο μεγαλύτερη επιρροή μπορεί να ασκήσει το t . Αυτά τα τέσσερα στοιχεία χρησιμοποιούνται αρχικά για τον υπολογισμό του $I(t)$ για κάθε t ξεχωριστά και στη συνέχεια συνδυάζονται για την εύρεση των πιο σημαντικών χρηστών και θεματικών περιοχών. Στον μαθηματικό τύπο που υπολογίζεται το $I(t)$, αθροίζεται ο αριθμός γ με την διαφορά του i από το θ και το παραπάνω αποτέλεσμα έχει σα συντελεστή (δηλαδή έχει σαν πολλαπλασιαστή) τον αριθμό λ . Αξίζει να σημειωθεί ότι κάθε ένα από τα παραπάνω στοιχεία χρησιμοποιείται στην εξίσωση με ξεχωριστό βάρος-συντελεστή ώστε αναλόγως με το που θέλουμε να εστιάσουμε, το μοντέλο να δίνει μεγαλύτερη έμφαση στο στοιχείο που θεωρείται βασικότερο.

Αφού έχουν υπολογιστεί τα $I(t)$ για τα tweets που μας ενδιαφέρουν σειρά έχει ο υπολογισμός της επιρροής ενός χρήστη και μίας θεματικής ενότητας. Για τον υπολογισμό και των δύο στοιχείων ορίζεται ένα κατώφλι κ το οποίο ξεχωρίζει τα απλά tweets από τα tweets που μπορούν να ασκήσουν επιρροή, αν και μόνο αν το $I(t)$ είναι μεγαλύτερο από το κ . Για υπολογισμό του συνολικού ποσού επιρροής ενός χρήστη ($InfluenceScoreofUser(u)$) πολλαπλασιάζεται ο αριθμός των tweets που ασκούν επιρροή με το άθροισμα των $I(t)$ αυτών και το γινόμενο τους διαιρείται με τον συνολικό αριθμό tweets του χρήστη. Αντίστοιχα για τον υπολογισμό επιρροής μίας θεματικής ενότητας ($InfluenceofScoreofHashtag(h)$) πολλαπλασιάζεται ο αριθμός των tweets που ασκούν επιρροή με το άθροισμα των $I(t)$ αυτών και το γινόμενο τους διαιρείται με τον συνολικό αριθμό tweets στα οποία γίνεται αναφορά της συγκεκριμένης θεματικής ενότητας (hashtag). Ο παραπάνω μηχανισμός ονομάζεται Twitter Influence Computer (TIC) και εντοπίζει τους χρήστες του κοινωνικού δικτύου που τραβούν την περισσότερη προσοχή

3.4 Εφαρμογή Μετρικών Κοινωνικών Δικτύων σε ένα Συνεργατικό Μοντέλο Εμπιστοσύνης για τη Δημιουργία Προσωποποιημένων Συστάσεων Χρηστών [4]

Η έρευνα αυτή ξεκινάει με την μεταφορά του κοινωνικού δικτύου σε γράφο και θεωρεί ότι οι κόμβοι του συμβολίζουν τους χρήστες του δικτύου και οι ακμές του τις σχέσεις μεταξύ τους. Χρησιμοποιεί δύο ειδών δεδομένα για τον υπολογισμό της επιρροής, τις γενικευμένες μετρήσεις, που υπολογίζονται αντλώντας πληροφορίες από το γράφο του κοινωνικού δικτύου και τις τοπικές συνεργατικές μετρήσεις που δέχονται σαν είσοδο τιμές που συμβολίζουν την εμπιστοσύνη που έχει ένας χρήστης για τους γύρω του.

Οι γενικευμένες μετρήσεις κινούνται σε δύο βασικούς άξονες, την κεντρικότητα και το κύρος. Η κεντρικότητα χωρίζεται σε τρεις βασικές υποκατηγορίες:

- Βαθμός κεντρικότητας $Gd(i)$: υπολογίζεται σύμφωνα με το βαθμό του κόμβου i , δηλαδή των αριθμό των ακμών που συνδέονται με αυτόν.

- Εγγύτητα κεντρικότητας $G_c(i)$: το ποσό αυτό υποδηλώνει το πόσο εύκολα ο χρήστης i αλληλοεπιδρά με τους υπόλοιπους χρήστες του κοινωνικού δικτύου. Ειδικότερα όσο μικρότερη απόσταση έχει ο i από τους υπόλοιπους χρήστες τόσο πιο κεντρικά είναι τοποθετημένος στο γράφο και τόσο μεγαλύτερη επιρροή μπορεί να ασκήσει.
- Μεταξύ τους κεντρικότητα $G_b(i)$: σε περίπτωση που ο i βρίσκεται μεταξύ δύο μη γειτονικών κόμβων, τότε αποτελεί συνδετικό κρίκο μεταξύ τους και μπορεί να επηρεάσει τις επιλογές τους. Όσες περισσότερες φορές συμβαίνει αυτό για τον χρήστη i , τόσο πιο σημαντική θέση έχει ο χρήστης στο κοινωνικό δίκτυο.

Το κύρος ενός χρήστη υπολογίζεται με βάση τον αριθμό των αναφορών του σε άλλους χρήστες, αλλά και των αναφορών που δέχεται από άλλους χρήστες. Πιο συγκεκριμένα το ποσό αυθεντίας $G_a(i)$ υποδηλώνει τον αριθμό των φορών που ο i έχει αναφερθεί από άλλους χρήστες και το ποσό διάδοσης $G_h(i)$ υποδηλώνει τον αριθμό των χρηστών στους οποίους έχει αναφερθεί ο i . Είναι σημαντικό να αναφερθεί ότι τα δύο ποσά αυτά είναι ρητά συνδεδεμένα μεταξύ τους καθώς ένας χρήστης με υψηλό ποσό αυθεντίας αναφέρεται σε αξιοσημείωτα κανάλια διάδοσης και αντίστοιχα τα κανάλια διάδοσης με υψηλό ποσό διάδοσης αναφέρουν υψηλού επιπέδου αυθεντίες. Το άθροισμα όλων των παραπάνω γενικευμένων μετρήσεων αποτελεί τη συνολική γενικευμένη επιρροή $G_l(i)$ του χρήστη i προς τους υπόλοιπους χρήστες του δικτύου.

Οι τοπικές συνεργατικές μετρήσεις έχουν διαφορετικό τρόπο υπολογισμού, καθώς υπολογίζονται ανά ζευγάρια χρηστών λαμβάνοντας υπόψιν τις άμεσες και έμμεσες σχέσεις εμπιστοσύνης μεταξύ των χρηστών. Ως ένα παράδειγμα διαχωρισμού των δύο αυτών ποσών, μπορούμε να πούμε πως άμεση εμπιστοσύνη είναι η σχέση «φιλίας» μεταξύ δύο χρηστών, ενώ έμμεση εμπιστοσύνη είναι απλά ένα σχόλιο σε κάποια δημοσίευση. Φυσικά η επιλογή των κριτηρίων για τον ορισμό της άμεσης και έμμεσης επιρροής, διαφέρουν ανάλογα με το κοινωνικό δίκτυο. Το μοντέλο υπολογισμού δέχεται ως δεδομένο το χρόνο και έτσι λαμβάνει υπόψιν τις αλλαγές που μπορούν να συμβούν στις σχέσεις των χρηστών σε βάθος χρόνου. Οι τοπικές συνεργατικές μετρήσεις αποτελούνται από τα παρακάτω μεγέθη:

- Local score $LS(i,j)$: υπολογίζεται από τους παράγοντες άμεσης και έμμεσης επιρροής, όπου καθένα από τα δύο μεγέθη έχει τον αντίστοιχο δείκτη βαρύτητας, ο οποίος εξαρτάται από το είδος του κοινωνικού δικτύου που εξετάζεται αλλά και τον τρόπο προσέγγισης της μελέτης (μεγαλύτερη βαρύτητα σε ένα από τα δύο μεγέθη)
- Local accumulative score $LAS(i,j)$: υπολογίζεται το LS για συγκεκριμένη χρονική περίοδο, αθροίζοντας τα ποσά των LS για κάθε χρονική στιγμή της επιλεγμένης περιόδου. Το μέγεθος της χρονικής περιόδου, αντιπροσωπεύει τη μνήμη του συστήματος και με αυτό τον τρόπο υπολογίζεται και το στοιχείο της ανανέωσης, καθώς οι έμμεση και άμεση επιρροή μεταξύ των χρηστών μπορεί να μεταβληθεί απότομα αναλόγως τη μορφή του κοινωνικού δικτύου.
- Collaborative local score $CLS(i,j)$: αποτελεί τη συνολική τιμή της τοπικής συνεργατικής μέτρησης ενός κόμβου i καθώς αθροίζει το άμεσο τοπικό ποσό επιρροής μεταξύ των χρηστών i, j με το έμμεσο τοπικό ποσό επιρροής που ασκούν οι έμπιστοι χρήστες του i στον j . Σημειώνουμε ότι ο παραπάνω υπολογισμός γίνεται σε βάθος χρόνο και με τη χρήση ειδικών δεικτών βαρύτητας αναλόγως με το σκοπό του εκάστοτε πειράματος.

Χρησιμοποιώντας τα παραπάνω στοιχεία, το προτεινόμενο μοντέλο έχει ως στόχο την παραγωγή εξατομικευμένων προτάσεων στους χρήστες του κοινωνικού δικτύου, είτε για ενδιαφέροντες χρήστες είτε για ενδιαφέρων περιεχόμενο. Διαχωρίζει την επιρροή σε τρεις τύπους: την ατομική επιρροή του χρήστη από τον ίδιο του τον εαυτό, την επιρροή από το κοντινό περιβάλλον του και την επιρροή από τους πιο έμπιστους χρήστες του κοινωνικού δικτύου (και οι τρεις τύποι επιρροής υπολογίζονται από τις προαναφερθέντες μετρήσεις). Στην συνέχεια υπολογίζεται η συνολική επιρροή για κάθε χρήστη και δημιουργείται μια κατάταξη με το ποσό επιρροής των χρηστών, η οποία χρησιμοποιείται για την παροχή εξατομικευμένων προτάσεων στον χρήστη, ακόμα και στην περίπτωση των νέων χρηστών που δεν έχουν προλάβει να πραγματοποιήσουν μεγάλη αλληλεπίδραση στο κοινωνικό δίκτυο. Τέλος, αξίζει να σημειωθεί ότι το συγκεκριμένο μοντέλο μπορεί να εφαρμοστεί σε όλων των ειδών τα κοινωνικά δίκτυα.

3.5 Εύρεση Επιδραστικών Χρηστών σε Κοινωνικά Δίκτυα [5]

Η συγκεκριμένη μελέτη, στοχεύει στον υπολογισμό της επιρροής σε δύο ειδών κοινωνικά δίκτυα: στα κοινωνικά δίκτυα που λαμβάνουν υπόψιν την τοποθεσία του χρήστη και στα ακαδημαϊκά κοινωνικά δίκτυα εξετάζοντας τους συγγραφείς και τα συγγραφικά τους έργα.

Ο γράφος των κοινωνικών δικτύων που περιέχουν πληροφορίες για την γεωγραφική περιοχή του χρήστη, αποτελείται συνήθως από δύο είδη κόμβων (κόμβους που αντιπροσωπεύουν χρήστες και κόμβους που αντιπροσωπεύουν τοποθεσίες) και ακμές που αντιπροσωπεύουν τις σχέσεις μεταξύ τους. Η μελέτη αυτή περιγράφει μία διαφορετική οπτική καθώς οι κόμβοι του γράφου απεικονίζουν μόνο χρήστες, με τις ακμές να είναι κατευθυνόμενες και με συντελεστή, ο οποίος υπολογίζεται από το πηλίκο του αριθμού των τοποθεσιών που έχει επισκεφθεί ένας χρήστης, προς τον αριθμό των τοποθεσιών που έχουν επισκεφθεί και οι δύο γειτονικοί χρήστες. Για τον υπολογισμό της επιρροής γίνεται χρήση του ήδη υπάρχοντος *IP* αλγόριθμου. Ο αλγόριθμος αυτός εξετάζοντας το γράφο, αποδίδει σε κάθε χρήστη ένα ποσό για την επιρροή που ασκεί και ένα ποσό για την παθητικότητά του, δηλαδή την κατάσταση στην οποία ο χρήστης δέχεται πληροφορίες αλλά δεν τις μεταδίδει στους υπόλοιπους χρήστες του κοινωνικού δικτύου. Ο αλγόριθμος που προτείνεται από αυτή τη μελέτη, αποτελεί μια τροποποίηση του *IP* αλγόριθμου, ώστε να λαμβάνει υπόψιν και τις γεωγραφικού περιεχομένου πληροφορίες του δικτύου.

Στα τα ακαδημαϊκά κοινωνικά δίκτυα η έννοια της επιρροής εξετάζεται είτε ως προς τη σημαντικότητα του συγγραφέα είτε ως προς τη σημαντικότητα του θέματος. Ειδικότερα ένα θέμα είναι περισσότερο σημαντικό εάν είναι επίκαιρο και ασχολούνται με αυτό πολλοί συγγραφείς. Οι γράφοι των ακαδημαϊκών κοινωνικών δικτύων αναπαριστούν είτε τους συγγραφείς (όπου οι ακμές αναφέρονται σε papers που έχουν συμβάλει και οι δυο) είτε τις δημοσιεύσεις (όπου οι ακμές αναφέρονται σε παραπομπές μεταξύ τους). Η μέθοδος υπολογισμού της επιρροής βασίζεται στον *Pagerank* αλγόριθμο, ο οποίος γενικότερα εφαρμόζεται σε διαδικασίες κατάταξης ιστοσελίδων και είναι βασισμένος στον τρόπο εφαρμογής των παραπομπών των ακαδημαϊκών δημοσιεύσεων. Από την εφαρμογή του παραπάνω αλγόριθμου επιλέχθηκαν οι δέκα δημοσιεύσεις με το μεγαλύτερο αποτέλεσμα και συγκρίθηκαν, ως προς τους συγγραφείς τους, με δημοσιεύσεις που κέρδισαν επιστημονικά βραβεία ώστε να ξεχωρίσουν τους συγγραφείς με τη μεγαλύτερη επιρροή στην επιστημονική κοινότητα. Η μελέτη επίσης εξελίσσει την έρευνα της, υπολογίζοντας το μελλοντικό αποτέλεσμα του *Pagerank* αλγόριθμου για συγκεκριμένη χρονιά με βάση παλαιότερα

αποτελέσματα και επίσης αναλογικά προσπαθεί να προβλέψει τον αριθμό των λήψεων της δημοσίευσης. Για τον υπολογισμό των παραπάνω μελλοντικών δεδομένων είναι αναγκαίες οι παρακάτω κατηγορίες πληροφοριών:

- Απόλυτα αποτελέσματα: Περιλαμβάνει τα αποτελέσματα του Pagerank αλγόριθμου κάθε δημοσίευσης για τα τελευταία πέντε χρόνια.
- Αποτέλεσμα διαφοροποίησης: Δείχνει την εξέλιξη του αποτελέσματος του Pagerank ανά χρονιά.
- Πληροφορίες συγγραφέα: Περιλαμβάνει α) τη μέση τιμή του Pagerank αποτελέσματος των δημοσιεύσεων που έχουν τουλάχιστον ένα κοινό συγγραφέα και β) τη μέγιστη τιμή του Pagerank αποτελέσματος του ίδιου συνόλου δημοσιεύσεων.
- Ηλικία: Αποτελεί τη διαφορά του έτους έκδοσης της δημοσίευσης από τη σημερινή ημερομηνία.
- Κείμενο: Αναφέρεται στη συχνότητα εμφάνισης των 100 πιο δημοφιλή λημμάτων στην περίληψη (abstract) της δημοσίευσης.

Εφαρμόζοντας τον αλγόριθμο Initialized Gradient Boosting Regression Trees (IGBRT) στα παραπάνω δεδομένα γίνεται η πρόβλεψη των παραπομπών και των λήψεων των δημοσιεύσεων.

3.6 Εκμάθηση Πιθανοτήτων Επιρροής στα Κοινωνικά Δίκτυα [6]

Στα κοινωνικά δίκτυα υπάρχουν τρεις βασικοί λόγοι από τους οποίους μπορεί να επηρεαστεί κάποιος χρήστης και να κάνει μία πράξη. Πρώτων μπορεί να επηρεαστεί από φίλους, δεύτερων από κάποιο γεγονός και τρίτων υπάρχει η πιθανότητα ο χρήστης να είναι πολύ ενεργός και να μην επηρεάζεται από κάπου για τις πράξεις που κάνει. Η έρευνα αυτή επικεντρώνεται στον υπολογισμό της πιθανότητας επιρροής από γειτονικούς χρήστες, ένα ποσό το οποίο θεωρείται ως δεδομένο στα περισσότερα μοντέλα υπολογισμού επιρροής. Για τον υπολογισμό της πιθανότητας επιρροής χρειάζεται ένας πίνακας πληροφοριών ο οποίος περιέχει τρεις στήλες (Χρήστης, Πράξη, Χρόνος) όπου κάθε γραμμή του πίνακα δείχνει ότι ο χρήστης έκανε τη συγκεκριμένη πράξη σε συγκεκριμένο χρόνο και ένα γράφο του κοινωνικού δικτύου, όπου οι κόμβοι του είναι οι χρήστες του και οι ακμές οι σχέσεις εμπιστοσύνης μεταξύ τους.

Για τον υπολογισμό της πιθανότητας επιρροής μεταξύ δύο κόμβων u , v υπάρχουν τέσσερις διαφορετικές μαθηματικές προσεγγίσεις και εξηγούνται παρακάτω:

- Bernouli distribution: Κάθε προσπάθεια επιρροής αναμεταξύ των χρηστών συσχετίζεται με μία πράξη, άρα λαμβάνεται υπόψιν το σύνολο των πράξεων και η πιθανότητα επιρροής δίνεται από το παρακάτω πηλίκο: (πράξεις του χρήστη u που έχουν επηρεαστεί από τον v) / (συνολικές πράξεις του χρήστη u)
- Jaccard Index: Χρησιμοποιείται κυρίως για την εύρεση όμοιων στοιχείων ανάμεσα σε σύνολα και η πιθανότητα επιρροής δίνεται από το παρακάτω πηλίκο: (πράξεις του χρήστη u που έχουν επηρεαστεί από τον v) / (πράξεις που κάνει είτε ο u είτε ο v)
- Bernouli distribution Partial Credit: ομοίως με παραπάνω συνυπολογίζοντας την επιρροή των γειτονικών κόμβων

- Jaccard Index Partial Credit: ομοίως με παραπάνω συνυπολογίζοντας την επιρροή των γειτονικών κόμβων

Οι προαναφερθέντες τρόποι υπολογισμού της πιθανότητας επιρροής εφαρμόζονται σε τρία προτεινόμενα μοντέλα τα οποία ως κύριο στοιχείο διαφοροποίησης μεταξύ τους έχουν το στοιχείο του χρόνου και παρουσιάζονται παρακάτω:

- Στατικό μοντέλο – Static model: Αυτό το μοντέλο είναι ανεξάρτητο του χρόνου, δηλαδή οποιαδήποτε στιγμή ένας χρήστης u προσπαθήσει να επηρεάσει έναν χρήστη v , η τιμή της πιθανότητας επιρροής παραμένει σταθερή και υπολογίζεται με μία από τις παραπάνω μεθόδους.
- Μοντέλο Συνεχούς χρόνου – Continuous Time model (CT): Στην πραγματικότητα η διάχυση επιρροής στα κοινωνικά δίκτυα επηρεάζεται άμεσα από το χρόνο. Ένας χρήστης έχει τη μέγιστη πιθανότητα να επηρεάσει τους γειτονικούς του κόμβους για μία πράξη, την αμέσως επόμενη στιγμή από την στιγμή που την πραγματοποιεί. Αμέσως μετά ξεκινά να μειώνεται εκθετικά η πιθανότητα αντιστρόφως ανάλογα με τον χρόνο, δηλαδή όσο αυξάνεται ο χρόνος από τη στιγμή ενεργοποίησης τόσο μειώνεται η πιθανότητα επιρροής. Επίσης στα μοντέλα συνεχούς χρόνου η συνδυαστική πιθανότητα επιρροής αλλάζει δραστικά όταν κάποιος νέος χρήστης ενεργοποιείται, καθώς αποκτάει τη μέγιστη τιμή η πιθανότητα επιρροής των γειτονικών του κόμβων και αυτό σημαίνει ότι πρέπει να υπολογιστεί ξανά από την αρχή.
- Μοντέλο διακριτού χρόνου – Discrete Time model (DT): Επειδή το μοντέλο συνεχούς χρόνου αλλάζει συνεχώς και ο υπολογισμός των πιθανοτήτων κοστίζει πολύ σε χρόνο προτείνεται το ενδιάμεσο μοντέλο διακριτού χρόνου. Στο μοντέλο αυτό η πιθανότητα επιρροής ενός χρήστη παραμένει σταθερή για συγκεκριμένο χρονικό διάστημα από τη στιγμή ενεργοποίησής του και μετά το πέρας του διαστήματος αυτού μηδενίζεται.

Για τον υπολογισμό των παραμέτρων των παραπάνω μοντέλων, ο προτεινόμενος αλγόριθμος χρειάζεται ως είσοδο το γράφο του κοινωνικού δικτύου, όπου οι κόμβοι του αντιπροσωπεύουν τους χρήστες και οι ακμές τις κοινωνικές σχέσεις μεταξύ τους και τον πίνακα ενεργειών, του οποίου οι γραμμές είναι της μορφής (u, a, t) και δείχνουν ότι τη χρονική στιγμή t , ο χρήστης u , έκανε την πράξη a . Στις ακμές του γράφου αντιστοιχούν τιμές οι οποίες δείχνουν τη χρονική στιγμή που οι δύο χρήστες έγιναν γείτονες στο κοινωνικό δίκτυο και ότι ο πίνακας ενεργειών είναι ταξινομημένος πρώτα κατά πράξη και στη συνέχεια κατά τη χρονική στιγμή που έγινε η πράξη, πράγμα το οποίο βοηθάει τον αλγόριθμο να εκτελείτε για κάθε πράξη ξεχωριστά και να εξοικονομεί χρόνο, το οποίο είναι πολύ σημαντικό καθώς οι πίνακες ενεργειών πραγματικών κοινωνικών δικτύων έχουν πολύ μεγάλο μέγεθος. Ο αλγόριθμος εξετάζοντας μία-μία τις γραμμές του πίνακα ενεργειών και λαμβάνοντας υπόψιν τη χρονολογική σειρά των πράξεων (χρονική στιγμή έναρξης κοινωνικού δεσμού και χρονική στιγμή πράξης της ενέργειας), αυξάνει τους αντίστοιχους μετρητές και υπολογίζει τις πράξεις που έχει κάνει ένας χρήστης, τις κοινές πράξεις που έχουν οι χρήστες μεταξύ τους (και τα δύο προηγούμενα μεγέθη είναι απαραίτητα για τον υπολογισμό της πιθανότητας επιρροής) καθώς και το μέσω χρόνο που χρειάστηκε ώστε να επηρεαστεί ο γειτονικός χρήστης. Τα αποτελέσματα του αλγορίθμου κατατάσσονται όπου φαίνονται οι πιθανότητες επιρροής ανάμεσα στους χρήστες και οι χρονικές περιόδους που διαμεσολάβησαν ώστε να διαρρεύσει μία ενέργεια στους γειτονικούς χρήστες. Τέλος, αξίζει να σημειωθεί ότι για τον υπολογισμό των *Bernoulli distribution* και *Jaccard Index* σε μοντέλα σταθερού και συνεχούς χρόνου ο

αλγόριθμος χρειάζεται μόνο έναν έλεγχο (scan) του πίνακα ενεργειών ενώ δεύτερος έλεγχος θα χρειαστεί για τους *Partial Credit* υπολογισμούς στα διακριτού χρόνο μοντέλα.

3.7 Στοιχεία καθορισμού μιας εικόνας ως δημοφιλούς [7]

Η συγκεκριμένη μελέτη συνδυάζει το οπτικό περιεχόμενο της εικόνας με στοιχεία του κοινωνικού δικτύου στο οποίο έχει δημοσιευθεί η εικόνα με σκοπό τον υπολογισμό της δημοτικότητας που μπορεί να έχει η συγκεκριμένη εικόνα, πριν τη δουν οι χρήστες. Η μελέτη χρησιμοποιεί ως δεδομένα 2.3 εκατομμύρια φωτογραφίες από το κοινωνικό δίκτυο Flickr.

Ξεκινώντας με την ανάλυση του οπτικού περιεχομένου της εικόνας, εξετάζονται απλά στοιχεία ανθρώπινης παρατήρησης, όπως είναι η ένταση και η ποικιλία των χρωμάτων το οποία αποδεικνύεται ότι έχουν αμελητέα συμβολή στην αύξηση της δημοτικότητας της εικόνας και για αυτό σχεδόν παραλείπονται. Παρόλα αυτά, αναλύοντας το ιστόγραμμα χρωμάτων των εικόνων βγαίνει το συμπέρασμα ότι τα φανταχτερά χρώματα που προσελκύουν τη ματιά και την προσοχή του χρήστη, βοηθούν ώστε η εικόνα να έχει περισσότερες προβολές. Προχωρώντας σε πιο εξειδικευμένα στοιχεία παρατήρησης των εικόνων, αναλύονται χαρακτηριστικά όπως είναι το φόντο, στοιχείο που μπορεί να βοηθήσει το χρήστη να κατανοήσει πιο γρήγορα το περιεχόμενο της εικόνας, η οπτική υφή, η ομάδα των αποχρώσεων, πράγμα το οποίο βοηθάει στους υπολογισμούς καθώς οι αποχρώσεις είναι άπειρες (η συγκεκριμένη μελέτη χρησιμοποιεί το μοντέλο 50 αποχρώσεων), τα βαθμωτά ερεθίσματα των οπτικών νεύρων, τα οποία υπόκεινται σε υπολογιστική επεξεργασία σύμφωνα με τη μεθοδολογία του *Histogram of Oriented Gradient* (μέθοδος αναγνώρισης αντικειμένων σε εικόνες) καθώς και deep learning αλγόριθμοι, οι οποίοι καταφέρνουν απεικονίσεις εικόνων βασιζόμενοι στους νευρώνες του εγκέφαλου. Η ακρίβεια της πρόβλεψης δημοτικότητας αποκτάει τη μέγιστη τιμή της όταν υπάρχει συνδυασμός όλων των παραπάνω στοιχείων και έχοντας ως ελάχιστη τιμή το 0.3 δείχνει ότι η ανάλυση του οπτικού περιεχομένου παίζει σημαντικό ρόλο στην πρόβλεψη της δημοτικότητας και μπορεί να συμπληρώσει τα στοιχεία του κοινωνικού δικτύου. Τέλος, εξετάζεται και ο τρόπος με τον οποίο σχετίζονται με την δημοτικότητα της εικόνας ο αριθμός και το είδος των αντικειμένων που περιέχει, και υλοποιείται ένας ταξινομητής με το είδος των αντικειμένων που προσφέρουν μικρή ή μεγάλη επιρροή στη δημοτικότητα της εικόνας.

Παράλληλα, τα στοιχεία που αντλούνται από το κοινωνικό δίκτυο μπορούν να βοηθήσουν στον υπολογισμό της δημοτικότητας. Τα χαρακτηριστικά που χρησιμοποιούνται για τον υπολογισμό είναι ο συνολικός αριθμός όλων των προβολών των εικόνων τους χρήστη, ο αριθμός των φωτογραφιών που έχει διαμοιράσει ο χρήστης, ο αριθμός των επαφών που έχει δημιουργήσει κοινωνικό δεσμό φιλίας, ο αριθμός των ομάδων στις οποίες ανήκει καθώς και ο μέσος αριθμός των μελών των ομάδων που ανήκει και τέλος το είδος του λογαριασμού του χρήστη, αν είναι απλός ή εξελιγμένος. Μεγάλη και σημαντική βοήθεια στη διάδοση της εικόνας και αύξηση της δημοτικότητάς της παίζουν η περιγραφή και οι "σημειώσεις" άλλων χρηστών (tags) σε αυτή καθώς αυξάνει τις εμφανίσεις της στις αναζητήσεις άλλων χρηστών.

Η έρευνα καταλήγει στο συμπέρασμα ότι και τα δύο είδη στοιχείων είναι απαραίτητα για τον υπολογισμό της δημοτικότητας των εικόνων καθώς αναφέρονται σε τελείως ξεχωριστά χαρακτηριστικά της.

3.8 Μέτρηση Επιρροής στο Twitter [8]

Η μελέτη αυτή επικεντρώνεται αποκλειστικά στο κοινωνικό δίκτυο Twitter και στους τρόπους υπολογισμού της επιρροής των χρηστών του. Για να θεωρήσουμε ότι το περιεχόμενο που δημοσιοποιεί ένας χρήστης A του Twitter ασκεί επιρροή σε ένα χρήστη B, αρκεί ο B είτε να αναδημοσιεύσει το πρωτότυπο περιεχόμενο του χρήστη A είτε να σχολιάσει δημόσια το μήνυμα ώστε να έρθει σε επαφή με τον χρήστη A. Συνεπώς, η επιρροή στο Twitter καθορίζεται από τη συμπεριφορά του χρήστη στις νέες δημοσιεύσεις των γειτονικών του χρηστών, για την οποία μπορεί να ισχύουν διάφορες πιθανότητες: Ο χρήστης A μπορεί να αποδέχεται δημόσια τη γνώμη του χρήστη B, είτε προσδοκώντας μία αλληλεπίδραση με αυτόν (αναδημοσίευση ή σχόλιο), είτε λόγω της θέσης του B στο κοινωνικό δίκτυο (πολλοί ακόλουθοι, διασημότητα). Επίσης, ο A μπορεί να συμφωνεί τόσο δημόσια όσο και ιδιωτικά με τον B και τις πράξεις του και να προβαίνει στη συνέχεια και αυτός σε διάφορες πράξεις, όπως συστάσεις, και retweets. Αντίθετα σε μείωση της επιρροής μπορεί να οδηγήσει η παράβλεψη ή η διαφωνία μεταξύ των χρηστών για το περιεχόμενο της εκάστοτε δημοσίευσης,

Τα βασικότερα στοιχεία στον υπολογισμό της επιρροής είναι ο αριθμός των ακολούθων (*followers*), ο αριθμός των αναδημοσιεύσεων (*retweets*) που έχει ένας χρήστης καθώς και ο αριθμός των αναφορών (*mentions*) σε ένα χρήστη, τα οποία συνδυάζονται ανάλογα με το είδος της επιρροής που χρειάζεται να υπολογιστεί. Αναλυτικότερα παρακάτω παρουσιάζονται οι δείκτες που χρησιμοποιώντας τα παραπάνω στοιχεία εξάγουν ενδιαφέροντα αποτελέσματα για την επιρροή που ασκεί ένας χρήστης:

- Follower/Following Ratio (r_f): συγκρίνει τον αριθμό των ακολούθων ενός χρήστη A σε σχέση με τους χρήστες που ακολουθεί ο ίδιος ο A. Όσο μεγαλύτερος είναι ο δείκτης αυτός, τόσο περισσότεροι χρήστες δείχνουν ενδιαφέρον για τις δημοσιεύσεις του χρήστη A.
- (A's Tweets that lead to action)/(total number of A's tweets) = Retweet and Mention Ratio (r_{RT}): Συγκρίνει τον αριθμό των tweets που οδήγησαν σε κάποια πράξη (είτε retweet είτε συζήτηση ανάμεσα στους χρήστες) με το συνολικό αριθμό των δημοσιεύσεων του χρήστη A. Αυτός ο δείκτης χρησιμοποιείται όταν θέλουμε να υπολογίσουμε την επιρροή που δημιουργείται με βάση το περιεχόμενο της δημοσίευσης.
- (Number of retweets)/(number of followers) = Interactor Ratio (r_i): Συγκρίνει τους χρήστες που αναδημοσιεύουν μία ανάρτηση του χρήστη A σε σχέση με το συνολικό αριθμό που τον ακολουθούν.

Ο καθένας από τους παραπάνω δείκτες μπορεί να οδηγήσει σε διαφορετική κατάσταση χρηστών, δηλαδή ανάλογα με το στοιχείο που μας ενδιαφέρει περισσότερο αλλάζει η ταξινόμηση των χρηστών. Παραδείγματος χάριν, μόνος του ο αριθμός των ακολούθων ενός χρήστη δεν αποκαλύπτει και την επιρροή που ασκεί πάνω τους ειδικά για το κοινωνικό δίκτυο *Twitter*, το οποίο βασίζεται κυρίως στο περιεχόμενο της δημοσίευσης. Για αυτό το λόγο παρουσιάζεται ο δείκτης *SNP* (*Social Network Potential*) ο οποίος συνδυάζει πληροφορίες και για το περιεχόμενο των αναρτήσεων αλλά και για τις διαπροσωπικές σχέσεις μεταξύ των χρηστών, χρησιμοποιώντας τους r_{RT} και r_i δείκτες που αναλύσαμε παραπάνω. Ο *SNP* δείκτης υπολογίζεται διαιρώντας το άθροισμα των r_{RT} και r_i δια του δύο.

4. ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΥΠΟ ΕΞΕΤΑΣΗ ΣΥΣΤΗΜΑΤΩΝ ΕΠΙΡΡΟΗΣ

Το κεφάλαιο αυτό παρουσιάζει μία γραφική απεικόνιση σε μορφή πίνακα της ταξινόμησης συστημάτων υπολογισμού επιρροής, η οποία βασίζεται στα κριτήρια που παρουσιάσαμε σε προηγούμενη ενότητα (Κεφάλαιο 2.1.1). Η ανάλυση των συστημάτων σύμφωνα με τα ίδια κριτήρια μας βοηθάει στον προσδιορισμό ομοιοτήτων και διαφορών αναμεταξύ τους και επίσης μπορεί να οδηγήσει εύκολα σε ομαδοποίηση και διαχωρισμό αυτών. Οι στήλες του πίνακα αναφέρονται στα συστήματα υπολογισμού της επιρροής που έχουμε μελετήσει και αναλύσει σε προηγούμενη ενότητα (Κεφάλαιο 3) και οι γραμμές αναφέρονται στα κριτήρια με βάση τα οποία προχωράμε στην ανάλυση. Παρακάτω παρατίθεται ο πίνακας της ταξινόμησης:

Πίνακας 1. Ταξινόμηση μηχανισμών υπολογισμού επιρροής

	Discovering Influential Nodes from Trust Network [1]	Identifying the influential bloggers in a community [2]	Reputation Mechanisms in online Social Networks - Influence reputation system in Twitter [3]	Application of Social Networking Metrics for generating personalized user recommendations [4]	Finding Influencers in Social Networks [5]	Learning Influence Probabilities In Social Networks [6]	What Makes an Image Popular [7]	Measuring Influence On Twitter - SNP [8]
Reputation System Description	Influence probabilities are estimated based on trust relationships and action logs	Identifies the influential users according to criteria that each blog post of the user has. Also studies the time period that a user is active and influential.	Influence score for tweets, users and specific topics (hashtags), using social activity information from S/N Twitter	(a) calculates score by user's position on the SN Graph (b) calculates user's opinion score by the user's reviews to articles and other users	Identifies in (a) influential users in S/N such as Foursquare and Twitter focused on location based criteria (b) influential papers in Academic Citation networks	Calculates influence of a user and his influence probability for static and continuous time model and also predicts the time that the user will get activated	Popularity prediction of an image in the Flickr S/N, by aggregating both user's social media criteria and image content characteristics.	Takes into consideration various Twitter statistics that calculate different types of Influence and combines them in one model
Target	User	User and Content	User and Content	User	Users (in location based S/Ns) and content in Academic citation networks	User	User and Content	User and Content
Scope	Global	Global	Global	Combination of Personalized and Global	Global	Global	Global	Global
Information Used	Trust relationships and action logs	Number of inlinks, outlinks, comments or new threads that the post started, length of the post.	Number of inlinks (retweets), outlinks, likes and the length (number of characters) of the tweet.	Information Regarding User position on SN Graph and Metrics about users each other opinions	(a) Social graph's node statistics (average path length, node's degree) (b) Several paper attributes that serve Pagerank algorithm	Action log about user's interactions and graph of the SN	Social media metrics (number of friends, photos, likes) and image content features (colors, digital analysis features, objects)	Number of followers, tweets, retweets, mentions, comments
Reasoning	Influence probabilities based on action log and trust relationships. Suggested algorithm discovers the influential nodes	Combines all the above characteristics and uses a mathematical formula to get an Influence score.	Calculates: (a) Influence Scores for tweets based on a social activity characteristics regarding this tweet, (b) Influence scores for each user and topic (according of their tweets scores). Then, according to specific thresholds, it chooses the most influential users or hashtags.	Sums all SN Graph position information with the opinion scores and each value weighted	Already implemented algorithms used the above mentioned inputs, which are: HITS, Pagerank, Influence Passitivity	Action log and social graph which give information about users interactions are used for estimation of the probability of influence between two users using different probability metrics	Machine learning techniques are used for image popularity prediction based on three different types of large datasets. Prediction is done for image content features, for social network metrics and then for their combination.	Adds conversation-oriented and content-oriented values separately and divides the result by two (2)
Display Method	Ranked list of users	Table rank visualization for both Influence score and Active-Inactive Users calculation	Web interface that includes Rank table and charts with statistics.	LeaderBoard	Table Rank visualization	Influence Matrix and also ROC Curves	(a) x,y axis system where x: ground truth popularity and y: predicted popularity (b) Heatmap of regions that makes an image popular	LeaderBoard

5. ΔΙΑΘΕΣΙΜΑ ΕΡΓΑΛΕΙΑ ΥΠΟΛΟΓΙΣΜΟΥ ΕΠΙΡΡΟΗΣ ΣΕ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ

Η επιρροή και η δημοτικότητα ενός χρήστη στα κοινωνικά δίκτυα αποτελεί πολύ σημαντική πληροφορία για τον ίδιο τον χρήστη και μπορεί να χρειαστεί σε διάφορες περιπτώσεις. Παραδείγματος χάριν ένας χρήστης με υψηλό ποσό επιρροής μπορεί να επιφέρει καλύτερα αποτελέσματα σε μία διαδικτυακή διαφημιστική εκστρατεία, με αποτέλεσμα να αποτελεί πλέον σημαντικό κριτήριο ακόμα και για την επιλογή υποψηφίων για εργασιακή απασχόληση σε κλάδους διαφημιστικού ενδιαφέροντος και όχι μόνο. Για αυτό το λόγο υπήρξε η ανάγκη δημιουργίας ενός κοινού τρόπου μέτρησης της επιρροής ώστε να γίνεται άμεση και ευδιάκριτη σύγκριση μεταξύ των χρηστών του εκάστοτε κοινωνικού δικτύου. Στο διαδίκτυο πλέον υπάρχουν πολλά εργαλεία υπολογισμού της επιρροής των χρηστών στα κοινωνικά δίκτυα, όπου καθένα από αυτά χρησιμοποιεί διαφορετικά κριτήρια και αλγόριθμους για τον υπολογισμό της τελικής τιμής επιρροής (score). Παρακάτω θα παρουσιάσουμε τα πιο διαδεδομένα και ολοκληρωμένα εργαλεία υπολογισμού επιρροής, τον τρόπο λειτουργίας τους, τα δεδομένα που δέχονται ως είσοδο, ποια αποτελέσματα παράγουν καθώς και τα κοινωνικά δίκτυα που λαμβάνουν υπόψιν. Τέλος, παραθέτουμε μία ταξινόμηση των μηχανισμών που έχουν παρουσιαστεί σύμφωνα με τα κριτήρια της ενότητας 2.1.1.

5.1 Klout – [10], [11], [12]

Ο μηχανισμός *Klout* αποτελούσε ένα από τα δημοφιλέστερα εργαλεία υπολογισμού της επιρροής όχι μόνο από ένα κοινωνικό δίκτυο, αλλά από το σύνολο αυτών στα οποία διατηρεί λογαριασμό ο χρήστης. Πρόσφατα εξαγοράστηκε την από εταιρία λογισμικού Lithium [13] όπως περιγράφεται παρακάτω. Στο μηχανισμό του Klout η έννοια της επιρροής αναφέρεται κυρίως στην ικανότητα μιας πράξης που κάνει ένας χρήστης να εκκινήσει δραστηριότητα σε άλλους χρήστες και τελικά να τους επηρεάσει. Ο αλγόριθμος υπολογισμού *Klout* μπορεί να συνδεθεί με εννέα διαφορετικά κοινωνικά δίκτυα και συνολικά αντλεί δεδομένα από περισσότερους από 3600 διαφορετικούς δείκτες. Παρακάτω παρουσιάζουμε τους πιο βασικούς από αυτούς ανά κοινωνικό δίκτυο:

- **Facebook:** Συνυπολογίζει τους αριθμούς των *Likes*, των σχόλιων, των δημοσιεύσεων καθώς και τον αριθμό των φίλων του χρήστη
- **Twitter:** Συνυπολογίζει τους αριθμούς των ακόλουθων, των αναδημοσιεύσεων, των αναφορών του χρήστη από άλλους χρήστες και τον αριθμό των λιστών στις οποίες ανήκει.
- **Instagram:** Συνυπολογίζει τους αριθμούς των *Likes*, των σχόλιων και των φωτογραφιών που έχει ανεβάσει ο χρήστης.
- **Google +:** Συνυπολογίζει τους αριθμούς των +1, των σχόλιων, των αναδημοσιεύσεων από το προφίλ του χρήστη μόνο.
- **LinkedIn:** Συνυπολογίζει τους αριθμούς των χρηστών με τους οποίους έχει συνδεθεί ο χρήστης, τον αριθμό των φορών που έχει προταθεί ο χρήστης από κάποιο άλλο χρήστη και τον αριθμό των σχόλιων που πραγματοποιεί ο χρήστης.

- Foursquare: Συνυπολογίζει τους αριθμούς των προτιμήσεων από άλλους χρήστες, των φίλων του χρήστη, των Check-ins καθώς και των τίτλων που έχει πάρει ο χρήστης (Mayorship).
- Wikipedia: Συνυπολογίζει τους αριθμούς των *inlinks* και *outlinks*.
- Youtube: Συνυπολογίζει φίλους, συνδρομητές, αριθμό προβολών,
- Klut: Συνυπολογίζει τον αριθμό των +K.

Για την εξαγωγή ενός αντικειμενικού και ακριβούς ποσού επιρροής, έχουν ληφθεί υπόψη κάποιοι βασικοί παράγοντες. Ειδικότερα ο αλγόριθμος:

- Προσαρμόζεται άμεσα στις προσθαφαιρέσεις μελών στο γράφο του κοινωνικού δικτύου και να μπορεί να λειτουργήσει με εκατομμύρια από αυτούς (*User Scalability*).
- Έχει τη δυνατότητα υπολογισμού της επιρροής αθροιστικά από διαφορετικά κοινωνικά δίκτυα, υπολογίζοντας ξεχωριστά τις διαφορετικές συμπεριφορές και ενέργειες που περιλαμβάνει το καθένα από αυτά (*Network Scalability*).
- Έχει τη δυνατότητα να αναγνωρίζει τους διαφορετικούς τρόπους αλληλεπίδρασης των χρηστών μεταξύ τους από το γράφο του κοινωνικού δικτύου (*Interaction Graph*).
- Έχει τη δυνατότητα να υπολογίσει και να ξεχωρίσει δείκτες οι οποίοι αποδίδουν θετικά στην επιρροή του χρήστη μακροπρόθεσμα και είναι σταθεροί στο χρόνο σε αντίθεση με άλλους που αλλάζουν σε καθημερινή βάση (*Temporal factors*).
- Έχει τη δυνατότητα υπολογισμού της επιρροής και εκτός κοινωνικών δικτύων, κυρίως από δημοσιευμένα άρθρα (*Offline factors*).
- Μπορεί και υπολογίζει την ισχύ και το εύρος της επιρροής του κάθε χρήστη με βασικό κριτήριο τον αριθμό των αντιδράσεων που ξεκινούν οι πράξεις του (*Reach and Strength of Influence*)

Ο αλγόριθμος φιλτράρει και κατηγοριοποιεί τα δεδομένα που λαμβάνει ως προς το ποιος χρήστης αντέδρασε σε κάποια ενέργεια και τι θέση κατέχει αυτός στο κοινωνικό δίκτυο, τότε έκανε τη συγκεκριμένη ενέργεια σε σχέση με την εκάστοτε χρονική στιγμή, σε ποιο κοινωνικό δίκτυο αναφέρεται η ενέργεια καθώς και το είδος της αλληλεπίδρασης (*Like, Comment, Follow, etc*). Με αυτή τη μέθοδο η κάθε ενέργεια αποκτά ξεχωριστό βάρος υπολογισμού και βοηθάει τον μηχανισμό να έχει ακριβέστερα αποτελέσματα.

Στο παρελθόν ο αλγόριθμος αυτός λειτουργούσε σαν ξεχωριστό μηχανισμό υπολογισμού της επιρροής ενός χρήστη κοινωνικών δικτύων, αλλά μετά την εξαγορά του από την εταιρία λογισμικού *Lithium* καταργήθηκε η αυτόνομη λειτουργία του. Ειδικότερα ενώ παλαιότερα ο χρήστης μπορούσε να γραφτεί στην υπηρεσία και να εξαγάγει την τιμή επιρροής του (γνωστή και ως *Klout Score*), πλέον δεν είναι δυνατό. Παρόλα αυτά ο τρόπος λειτουργίας του αλγόριθμου συνέβαλε πολύ στην εξέλιξη των προϊόντων της εταιρίας *Lithium* στους τομείς τεχνητής νοημοσύνης και *machine learning*. Με αυτή την αλλαγή ο μηχανισμός *Klout*, μέσω του νέου τρόπου χρήσης του, βοηθάει εταιρίες να προσελκύουν πελάτες και να παράγουν καλύτερο και πιο στοχευμένο περιεχόμενο για αυτούς, με κοινό όμως στόχο να αποτελεί, όπως και στο παρελθόν, της αύξηση επιρροής στα ψηφιακά κοινωνικά δίκτυα.

5.2 Alexa [14], [15]

Ο μηχανισμός *Alexa* αποτελεί προϊόν της εταιρίας *Amazon* και έχει ως στόχο την κατάταξη των ιστοσελίδων σύμφωνα με την επισκεψιμότητά τους. Οι πηγές που αντλεί δεδομένα, είναι αρχικά οι χρήστες του διαδικτύου που χρησιμοποιούν την επιπρόσθετη εφαρμογή (*Add On*) του μηχανισμού *Alexa* στο πρόγραμμα περιήγησης στο διαδίκτυο (*Internet Browser*) καθώς και οι ιστοσελίδες που έχουν συμφωνήσει να εντάξουν στην δομή τους, τον συμπληρωματικό προγραμματιστικό κώδικα του μηχανισμού. Από τα παραπάνω βγαίνει το συμπέρασμα ότι ο μηχανισμός λαμβάνει πληροφορίες μόνο από χρήστες που εν γνώσει τους παρέχουν δεδομένα από τις περιηγήσεις τους στο διαδίκτυο και αγνοεί τις μη εξουσιοδοτημένες περιηγήσεις. Τα δεδομένα που δέχεται ο μηχανισμός για τον υπολογισμό της θέσης ταξινόμησης μιας ιστοσελίδας, είναι ο αριθμός των μοναδικών επισκεπτών της και ο αριθμός των προβολών της από χρήστες. Η θέση κατάταξης υπολογίζεται από δεδομένα των τελευταίων τριών μηνών, τα οποία ανανεώνονται καθημερινά ενώ παράλληλα δεν υπολογίζονται πολλαπλές επισκέψεις από ίδιο χρήστη σε μία ημέρα, ώστε να αποφευχθεί η παραπλάνηση του αλγόριθμου.

Ο μηχανισμός κατάταξης *Alexa* λειτουργεί και σε ιστότοπους τύπου Blog, όπου κάθε χρήστης διαχειρίζεται μία υποσελίδα. Αυτό σημαίνει ότι μπορεί να υπολογίσει τους πιο δημοφιλείς χρήστες του και να τους ταξινομήσει ανάλογα με την διαδικτυακή κίνηση της ιστοσελίδας τους. Αυτή η χρήση του μηχανισμού *Alexa* μπορεί να αποτελέσει ένα εργαλείο υπολογισμού επιρροής σε κοινωνικά δίκτυα, όπου οι χρήστες ανάλογα με την επιρροή που ασκούν σε αυτό, παίρνουν την αντίστοιχη θέση στην κατάταξη.

Το *Add-On* που εγκαθιστά ο χρήστης στο πρόγραμμα περιήγησης του διαδικτύου πέρα από τον γενικό αριθμό κατάταξης του ιστότοπου, παρέχει και άλλες επιπρόσθετες ενδιαφέρουσες πληροφορίες. Ειδικότερα παρέχει ένα δεύτερο αριθμό κατάταξης οποίος υπολογίζεται από τα τοπικά δεδομένα της χώρας στην οποία εδρεύει ο ιστότοπος καθώς και μια λίστα με τους πεντακόσιους (500) ιστότοπους με την υψηλότερη βαθμολογία της χώρας αυτής. Για κάθε ιστότοπο της λίστας παρέχονται πληροφορίες όπως ο μέσος χρόνος περιήγησης ενός χρήστη στον ιστότοπο αυτό, οι ημερίσιες προβολές του ιστότοπου ανά χρήστη καθώς και το ποσοστό των εμφανίσεων του ιστότοπου που προήλθαν από κάποια μηχανή αναζήτησης και όχι από κατευθείαν περιήγηση. Ένα ακόμα πολύ χρήσιμο στατιστικό που εμφανίζεται στην οθόνη του χρήστη είναι ο αριθμός των *inlinks* στο συγκεκριμένο ιστότοπο, δηλαδή τον αριθμό των φορών που άλλοι ιστότοποι αναφέρονται σε αυτόν. Επίσης προσφέρει διαγραμματική απεικόνιση της κίνησης που παρουσιάζει ο ιστότοπος σε βάθος ενός χρόνου, έναν πίνακα με τις πιο συχνές αναζητήσεις χρηστών, που οδήγησαν σε επίσκεψη καθώς και το χρόνο απόκρισης του ο οποίος συγκρίνεται κιόλας με τις υπόλοιπες ιστοσελίδες. Τέλος, παρέχεται στο χρήστη η δυνατότητα να ανατρέξει πίσω στο χρόνο και επιλέγοντας μία από τις προτεινόμενες ημερομηνίες του παρελθόντος, να επισκεφτεί τον ιστότοπο σε κάποια παλαιότερη έκδοσή του, λειτουργία η οποία είναι πολύ ενδιαφέρουσα.

Ο συνεχής έλεγχος της διαδικτυακής κίνησης που πραγματοποιεί ο μηχανισμός *Alexa* μπορεί να βοηθήσει σημαντικά τις επιχειρήσεις να αυξήσουν τη δημοτικότητά τους και να επεκτείνουν τις εμπορικές τους δραστηριότητες. Αξιοποιώντας ένας ιδιοκτήτης ενός ιστότοπου τον μηχανισμό αυτό μπορεί να εντοπίσει τον ανταγωνισμό, να έχει μία πολύ καλή απεικόνιση και σύγκριση στατιστικών στοιχείων για αυτόν και να λάβει συμβουλές

και προτάσεις για την αναβάθμιση του ιστότοπου ώστε να πάρει καλύτερη θέση στη γενική κατάταξη, άρα μεγαλύτερη επισκεψιμότητα. Μάλιστα μέσω των μηχανισμών ελέγχου που παρέχει η εφαρμογή μπορεί να καταλάβει πιθανά λάθη που οδηγούν σε χαμηλή βαθμολογία και να προτείνει λύσεις και τρόπους αντιμετώπισης. Πέρα λοιπόν από μηχανισμός κατάταξης ιστότοπων, ο *Alexa* παρέχει και προτάσεις βελτίωσης σε τυχόν προβληματικές ιστοσελίδες με χαμηλή βαθμολογία.

5.3 TweetAnalyzer [16]

Ο μηχανισμός *TweetAnalyzer* υπολογίζει το ποσό θετικότητας ενός χρήστη του κοινωνικού δικτύου Twitter με βάση τις τελευταίες 200-300 δημοσιεύσεις του χρήστη. Το συνολικό ποσό θετικότητας ενός χρήστη δίνεται από το μέσο όρο των επιμέρους δημοσιεύσεων του.

Ο υπολογισμός του ποσού θετικότητας της κάθε δημοσίευσης γίνεται με βάση τις λέξεις που βρίσκονται σε αυτή και της λίστας *AFINN* [17]. Η λίστα *AFINN* είναι μία λίστα λέξεων ή και φράσεων και αποτελεί ένα εργαλείο αξιολόγησης του περιεχόμενου τους ως προς τη θετικότητα ή αρνητικότητα ή χυδαιότητα που εκφράζουν. Περιέχει σχεδόν 2500 λέξεις και εκφράσεις του αγγλικού λεξιλογίου, στις οποίες αντιστοιχεί μία τιμή ανάλογα με το νόημά της ενώ το εύρος της τιμής αυτής κυμαίνεται από μείον πέντε (-5) έως πέντε (5). Ο καθορισμός της τιμής αυτής έχει προκύψει από την υποκειμενική γνώμη του δημιουργού της *Finn Årup Nielsen*. Αν το άθροισμα των λέξεων της δημοσίευσης είναι μεγαλύτερο του μηδενός, τότε η δημοσίευση θεωρείται θετική ενώ αντίθετα αν το άθροισμα είναι μικρότερο του μηδενός, η δημοσίευση θεωρείται αρνητική. Η τελική τιμή θετικότητας επιρροής του χρήστη υπολογίζεται από τον παρακάτω μαθηματικό τύπο:

$$\text{θετικότητα επιρροής} = \frac{\text{θετικές δημοσιεύσεις} - \text{αρνητικές δημοσιεύσεις}}{\text{συνολικό αριθμό δημοσιεύσεων}}$$

Η αναζήτηση που παρέχει ο ιστότοπος του παραπάνω μηχανισμού βασίζεται στο όνομα χρήστη του κοινωνικού δικτύου για τον οποίο θέλουμε να υπολογίσουμε το ποσό επιρροής του στους υπόλοιπους χρήστες του. Στην οθόνη εμφάνισης των αποτελεσμάτων εμφανίζεται ένας πίνακας που περιέχει πληροφορίες για τον χρήστη όπως, το συνολικό ποσό επιρροής του χρήστη (είτε θετικό είτε αρνητικό), τα ποσοστά των προσβλητικών και αδιάφορων δημοσιεύσεών του, το συνολικό αριθμό δημοσιεύσεων που λήφθηκαν υπόψιν, την πιο θετική, προσβλητική αλλά και αδιάφορη δημοσίευσή του (συμπεριλαμβανομένου και του κειμένου) καθώς και τα εικονίδια (*Emojis*) που χρησιμοποιεί συχνότερα. Επίσης κάτω από τον πίνακα εμφάνισης αποτελεσμάτων υπάρχει η λίστα με τους πιο δημοφιλείς χρήστες του κοινωνικού δικτύου με τις αντίστοιχες πληροφορίες.

Ο μηχανισμός *TweetAnalyzer* αποτελεί το προσωπικό έργο ενός ανθρώπου πράγμα το οποίο σημαίνει ότι είναι σχετικά απλός στην υλοποίηση και ενδεχόμενος να παρερμηνεύει μερικά δεδομένα και να μην υπολογίζει σωστά το τελικό αποτέλεσμα. Σε μερικές λέξεις ή εκφράσεις όπου το νόημά τους αλλάζει σύμφωνα με τα συμφραζόμενα, ο αλγόριθμος αδυνατεί να εντοπίσει το πραγματικό νόημα και παράγει λάθος αποτέλεσμα. Χαρακτηριστικά, ο ίδιος ο δημιουργός του μηχανισμού αναφέρει ότι δε χρησιμοποίησε

πιο εξελιγμένες προγραμματιστικές μεθόδους, προς χάριν απλότητας, γεγονός το οποίο προκαλεί κακή μετάφραση του περιεχομένου της δημοσίευσης. Για παράδειγμα ένα άρθρο το οποίο αναφέρεται σε βιασμό (rape) ή εμπόριο λευκής σαρκός (sex trafficking) θα χαρακτηριστεί από τον αλγόριθμο ως χυδαίο, ακόμα και αν απλά περιγράφει το πρόβλημα ή προτείνει λύσεις αντιμετώπισης. Παρόλα αυτά αποτελεί μία έξυπνη και γρήγορη λύση στο πρόβλημα υπολογισμού της επιρροής των χρηστών του κοινωνικού δικτύου *Twitter*.

5.4 TweetReach [18], [19]

Ο μηχανισμός *TweetReach* επικεντρώνεται στον υπολογισμό επιρροής στο κοινωνικό δίκτυο *Twitter*. Ως επιρροή μεταφράζεται η διάδοση που έχει μία δημοσίευση ή ένα θέμα συζήτησης και ο μηχανισμός υπολογίζει το μέγιστο δυνατό κοινό που μπορεί να αλληλοεπιδράσει με αυτό. Κατά τη χρήση του μηχανισμού, ο χρήστης μπορεί να πραγματοποιήσει αναζήτηση με βάση το όνομα χρήστη, την αναφορά (*hashtag*) ή μία οποιαδήποτε λέξη και ο αριθμός που θα υπολογιστεί είναι το "*Reach*" του στοιχείου για το οποίο έγινε η αναζήτηση. Ο αριθμός αυτός δηλώνει το πλήθος των μοναδικών χρηστών που έλαβαν το *tweet* για το οποίο έγινε η αναζήτηση και δεν υπολογίζει πολλαπλές φορές τους ίδιους χρήστες. Τα δεδομένα της δωρεάν έκδοσης του μηχανισμού μπορούν να επεξεργαστούν μέχρι και εκατό *Tweets* μέσα στο χρονικό διάστημα της τελευταίας εβδομάδας. Ωστόσο ο αριθμός *Reach* δεν είναι το μοναδικό αποτέλεσμα του μηχανισμού αναζήτησης.

Με κάθε αναζήτηση, ο μηχανισμός παραθέτει στα αποτελέσματά του διάφορες μετρήσεις σε έναν ευανάγνωστο πίνακα όπου σχεδόν κάθε μέτρηση απεικονίζεται με κάποιο σχεδιάγραμμα. Πέρα λοιπόν από τον αριθμό *Reach* που περιγράψαμε παραπάνω εμφανίζονται και τα εξής αποτελέσματα:

- ***Exposure***: Ο αριθμός των φορών που είναι πιθανό να εμφανιστεί το *tweet* από διαφορετικούς χρήστες.
- ***Activity***: Διαγραμματική απεικόνιση της δραστηριότητας που επικρατεί γύρω από ένα *tweet* με στοιχεία όπως πόσοι χρήστες το έχουν αναδημοσιεύσει, έχουν απαντήσει ή έχουν ξεκινήσει νέα συζήτηση πάνω στην ίδια θεματολογία.
- ***Contributors***: Εμφανίζεται λίστα κατάταξης με τους πιο ενεργητικούς χρήστες με βάση την αναζήτηση μας
- ***Retweet list***: Εμφάνιση λίστας με τα *tweet* που έχουν αναδημοσιευτεί τις περισσότερες φορές.

Ο μηχανισμός *TweetReach* είναι ένα εύχρηστο εργαλείο που μπορεί να φανεί πολύ χρήσιμο σε ηλεκτρονικές διαφημιστικές καμπάνιες καθώς μπορεί να χρησιμοποιηθεί σαν εργαλείο πρόβλεψης της απήχησης που μπορεί να επιτύχει.

5.5 UnionMetrics [20], [21]

Η εταιρία *UnionMetrics* και τα προϊόντα της εξειδικεύονται στον υπολογισμό στατιστικών μελετών πάνω στα κοινωνικά δίκτυα και τα επιμέρους στοιχεία αυτών. Αποτελεί σημαντικό εργαλείο των διαφημιστικών εταιριών και των αντίστοιχων τμημάτων και στα χέρια των κατάλληλων επαγγελματιών μπορεί να αποβεί καθοριστικό για κάθε διαφημιστική καμπάνια. Το κοινωνικό δίκτυο στο οποίο έχει την καλύτερη απόδοση είναι

το *Tumblr*, στο οποίο οι χρήστες του έχουν την ελευθερία να εκφραστούν με οποιαδήποτε μορφής δημοσίευση και μάλιστα το ίδιο το *Tumblr* έχει ανακηρύξει την *UnionMetrics* ως τον καλύτερο συνεργάτη του σε θέματα στατιστικής ανάλυσης και πρόβλεψης δημοσιεύσεων. Επεξεργάζεται καθημερινά πάνω από εκατό εκατομμύρια δημοσιεύσεις σε πραγματικό χρόνο. Η ανάλυση των δεδομένων παράγει πολλά διαφορετικά στοιχεία όπως είναι:

- η εμφάνιση της πιο αποδοτικής ώρας δημοσίευσης,
- ο μέσος όρος αύξησης ή μείωσης των αναδημοσιεύσεων όσον αφορά χρήστες ή λέξεις κλειδιά,
- η εύρεση των πιο αποδοτικών χρηστών, των χρηστών δηλαδή που ασκούν τη μεγαλύτερη επιρροή στους υπόλοιπους και φυσικά
- η εμφάνιση μέσω των αναδημοσιεύσεων του πόσο μακριά μπορεί να φτάσει μία δημοσίευση.

Η επιτυχία του μηχανισμού βασίζεται στην ακρίβεια των δεδομένων που δέχεται, στην ευκολία παραμετροποίησης από τη μεριά του χρήστη, την εξατομίκευση των αποτελεσμάτων και στην ιστορικότητα δεδομένων που διατηρεί πράγμα το οποίο οδηγεί σε πιο ακριβή και σωστά αποτελέσματα. Ο τεράστιος όγκος δεδομένων που παράγονται από τα κοινωνικά δίκτυα αποτελεί την είσοδο του μηχανισμού και μέσω της ανάλυσής τους καταφέρνει να εξάγει χρήσιμες πληροφορίες στο χρήστη. Η σωστή εξαγωγή και ομαδοποίηση των δεδομένων αποτελεί την τεχνική του μηχανισμού *Unionmetrics* ο οποίος ουσιαστικά δεν προσφέρει κάποια μελλοντική πρόγνωση αλλά, με την κατάλληλη μετάφραση των ήδη υπάρχων δεδομένων δίνει στο χρήστη πληροφορίες για προηγούμενες δημοσιεύσεις, οι οποίες τον βοηθούν να βελτιώσει τις επόμενες, είτε κάνοντας καλύτερο το περιεχόμενό τους, είτε στοχεύοντας σε πιο συγκεκριμένο κοινό.

Η παρουσίαση των στατιστικών μετρήσεων στο χρήστη γίνεται με τέτοιο τρόπο ώστε να είναι εμφανή και ξεκάθαρα τα αποτελέσματα, είτε με ποσοτικά διαγράμματα, είτε με πίνακες κατάταξης είτε με μορφή γράφου-δένδρου. Παραδείγματα απεικονίσεων, παρουσιάζονται παρακάτω:



Figure 1. Γραφική απεικόνιση πληροφοριών για τους χρήστες που αλληλοεπιδρούν στο περιεχόμενο των δημοσιεύσεων

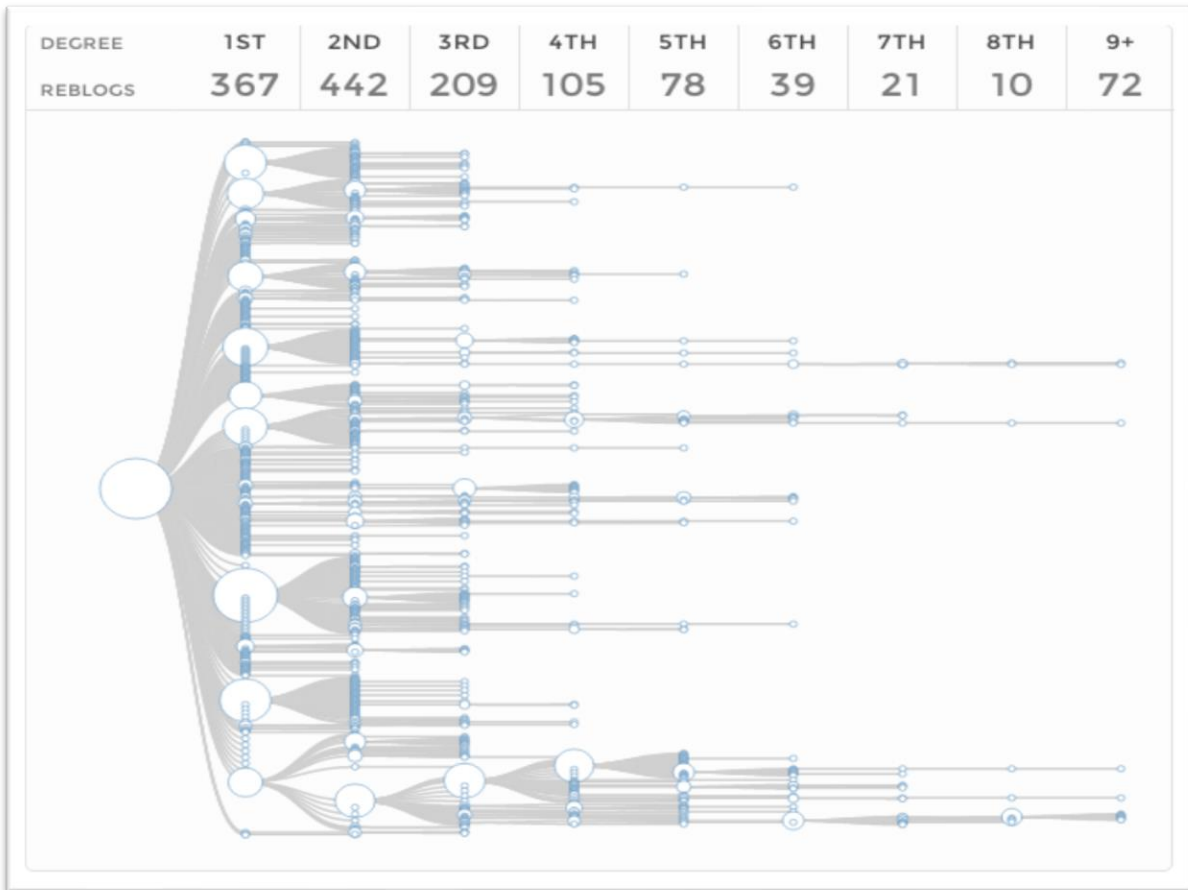


Figure 2. Γραφική απεικόνιση σε μορφή δένδρου σχετικά με τον αριθμό και το βαθμό αναδημοσιεύσεων μιας δημοσίευσης.

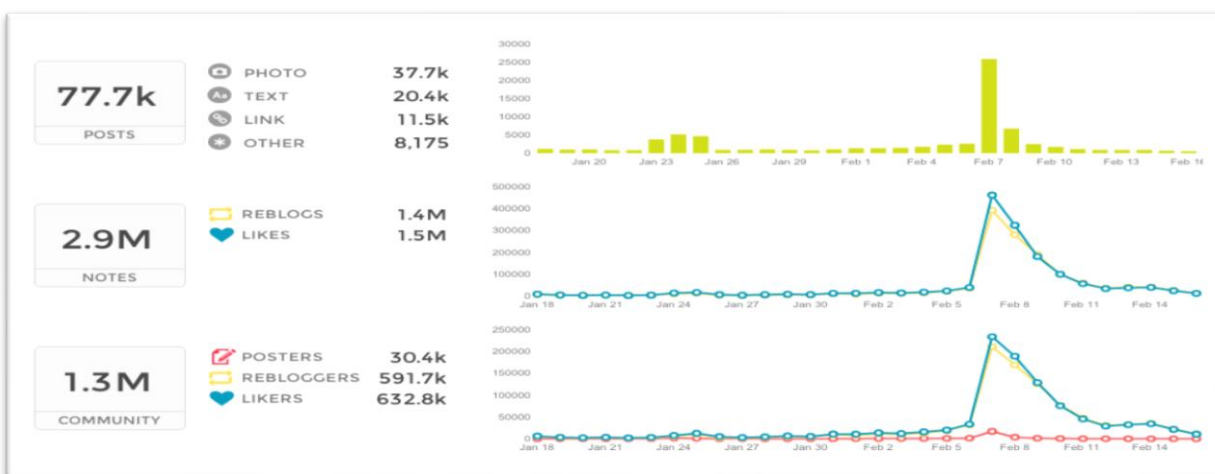


Figure 3. Γραφική απεικόνιση σχετικά με τη δραστηριότητα των δημοσιεύσεων ενός χρήστη σε βάθος χρόνου.

5.6 Ταξινόμηση των εργαλείων υπολογισμού επιρροής

Στον παρακάτω πίνακα παρουσιάζεται μία ταξινόμηση των εργαλείων υπολογισμού επιρροής σύμφωνα με τα κριτήρια που έχουμε περιγράψει στην ενότητα 2.1.1, μαζί με τα σχόλια της ανάλυσής μας.

Πίνακας 2 Ταξινόμηση των συστημάτων υπολογισμού επιρροής Klout και alexa

	klout.com	www.alexa.com
Reputation Tool Description	Gathers data from different social networks and combines the results into one reputation score	Browser add-on that monitors user's traffic and sorts web pages accordingly
Target	User	Content
Scope	Global	Global
Information Used	Over 3600 indexes from 9 different Social Networks	Browsing data from user and statistics from the cooperating web sites
Reasoning	The Reputation score is adjustable to the kind of each action performed, the completion time and the status of each participant. So each gets a different weight and the algorithm is more precise.	From the daily data received the algorithm calculates unique visitors, number of pageviews, average time spent on the web page and more.
Display Method	A display of the Klout Score including graphs and flow charts explaining the data that were calculated.	By clicking on the browser's add on, a window pop's up containing a table with the site's information and each cell leads to graphical representation environment (graphs, charts, tables, etc.)

Πίνακας 3. Ταξινόμηση των συστημάτων υπολογισμού επιρροής tweetanalyzer, tweetreach και unionmetrics

	tweetanalyzer.net	tweetreach.com	unionmetrics.com
Reputation Tool Description	calculates reputation from the past 200 tweets of a user by checking each post's word from the AFINN word list	Connects the influence of a user on twitter with the spread that his/her tweets have	focusing on Tumblr social network, the mechanism collects and analyze data of user's history
Target	Content	User	User and Content
Scope	Global	Global	Global
Information Used	The text of the tweets alongside with the AFINN word list	User's tweet history and activity on tweets from other users to calculate the influence number named "Reach"	Numerous posts on social networks including their related information (time, related users. Type of action, etc.)
Reasoning	Each tweet gets a total score from all the word it contains, and an average score is generated from the past 200 tweets of the user	Search on data and grouping the results	Data process to extract results
Display Method	The search results include the score and statistics of the user's tweets and also includes the most and least popular tweets of the user (best and worst score)	Plain display on a board of the calculated "Reach" number including side results such as tweet exposure and activity and top contributors also	Combination of charts and diagrams for each social network attribute that is displayed

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η μελέτη της επιρροής στα κοινωνικά δίκτυα αποτελεί ένα άκρως ενδιαφέρον πεδίο έρευνας το οποίο μπορεί να επιφέρει σημαντικά αποτελέσματα σε επιστημονικό επίπεδο αλλά μπορεί να έχει και πρακτικές εφαρμογές σε επαγγελματικές δραστηριότητες, καθώς μπορούν να επιφέρουν μεγάλες κέρδος, μειώνοντας σημαντικά το κόστος της εκάστοτε δραστηριότητας. Για παράδειγμα, είναι πολύ πιο εύκολο και οικονομικό για μια διαφημιστική εταιρία να προωθήσει τις καμπάνιες και τα προϊόντα της σε μία μικρή ομάδα ανθρώπων που ασκούν επιρροή σε ένα πολύ μεγάλο κοινό, παρά να απευθυνθεί απευθείας στο ευρύ κοινό. Η μελέτη μας επικεντρώθηκε στους μηχανισμούς υπολογισμού της επιρροής, στους μηχανισμούς αναγνώρισης των χρηστών και του περιεχομένου που ασκούν τη μεγαλύτερη επιρροή και τέλος πραγματοποιήσαμε μία ανάλυση για τα ήδη υπάρχοντα συστήματα αναγνώρισης και υπολογισμού της επιρροής. Για τα εξεταζόμενα συστήματα, πραγματοποιήθηκε συστηματική μελέτη των διαφόρων χαρακτηριστικών τους σύμφωνα με μια ταξινόμηση έμμεσων συστημάτων φήμης σε κοινωνικά δίκτυα, που συμπεριλαμβάνουν και τα συστήματα υπολογισμού φήμης.

Οι μηχανισμοί υπολογισμού επιρροής ανάλογα, με το σχεδιασμό τους, χωρίζονται κατά κύριο λόγο σε δύο είδη, σε αυτούς που χρησιμοποιούν στατιστικά στοιχεία του κοινωνικού δικτύου και σε αυτούς που χρησιμοποιούν το γράφο του κοινωνικού δικτύου. Μελετώντας τα στοιχεία του κοινωνικού δικτύου αποκτάμε μια πιο άμεση άποψη της επιρροής, καθώς εξετάζονται τα φανερά στοιχεία του κοινωνικού δικτύου (αριθμός like, inlinks, oulinks κτλ.) ενώ μεταφέροντας το κοινωνικό δίκτυο σε γράφο και μελετώντας τον, εξάγεται μια πιο θεωρητική οπτική της (εξετάζονται στοιχεία όπως κεντρικότητα κόμβου, μέγιστα μονοπάτια, αριθμός ακμών κτλ). Οι διάφορες μελέτες συνεισφέρουν ξεχωριστά στο ερευνητικό πεδίο, υπολογίζοντας εναλλακτικές όψεις του αντικείμενου της επιρροής. Κάποιες μελέτες προχωρούν σε υπολογισμό πιθανότητας επιρροής, είτε σε επίπεδο χρήστη είτε σε επίπεδο περιεχόμενου, άλλες στη μελλοντική επιρροή που μπορεί να ασκήσει ο χρήστης και άλλες στην μεταβολή της επιρροής σε βάθος χρόνου. Σύμφωνα με τη γνώμη μας τα πιο ακριβή αποτελέσματα μπορούν να εξαχθούν από συνδυαστικές μελέτες και αυτό γιατί πέρα από τη δημοτικότητα και την δημόσια απήχηση που μπορεί να έχει ένας χρήστης στο κοινωνικό δίκτυο, δεν μπορεί να παραμείνει για πολύ καιρό δημοφιλής και να ασκεί επιρροή αν το περιεχόμενο που παράγει ή προωθεί δεν είναι ποιοτικό και ενδιαφέρον.

Πολλές από τις παραπάνω εξεταζόμενες μελέτες έχουν παίξει καθοριστικό ρόλο στην υλοποίηση των συστημάτων υπολογισμού επιρροής που έχουμε παρουσιάσει στην εργασία μας. Παρόλα αυτά υπάρχουν και συστήματα [16] τα οποία έχουν εντελώς ξεχωριστή μέθοδο υπολογισμού, γεγονός που δηλώνει ότι οι οπτικές που αντιλαμβανόμαστε και υπολογίζουμε την επιρροή είναι πάρα πολλές και η μελέτη της επιρροής δε μπορεί να περιοριστεί σε στατιστικά και γράφους. Πιο γενικά, πέρα από τη μέθοδο υπολογισμού σημαντικό ρόλο παίζει και ο όγκος δεδομένων και ειδικότερα όσο περισσότερα δεδομένα υπάρχουν τόσο πιο αποδοτικός θα είναι ο υπολογισμός της επιρροής. Για αυτό το λόγο, τα συστήματα συνήθως χρησιμοποιούν πολλούς δείκτες και μάλιστα σε μερικές περιπτώσεις από διαφορετικά κοινωνικά δίκτυα, και με τον κατάλληλο συνδυασμό τους εξάγουν αποτελέσματα για τον χρήστη. Σημαντικό ρόλο, μετά από την κατάλληλη επεξεργασία των δεδομένων, αποτελεί και η εμφάνιση των αποτελεσμάτων. Με πιο συνηθισμένες μορφές απεικόνισης να αποτελούν την κατάταξη σε πίνακα και την

αποτύπωση των πληροφοριών σε διαγράμματα, τα συστήματα καταφέρνουν να οπτικοποιούν τα αποτελέσματα των υπολογισμών τους ώστε να τα αντιλαμβάνονται πιο εύκολα οι χρήστες τους.

Σαν γενικό συμπέρασμα καταλήγουμε ότι όσο πιο πολλά δεδομένα λαμβάνει υπόψιν ένα σύστημα υπολογισμού της επιρροής, τόσο πιο περίπλοκο και ταυτόχρονα αποδοτικότερο γίνεται. Άρα ο καλύτερος τρόπος επεξεργασίας και συγκομιδής των δεδομένων θα οδηγήσει σε ένα αποδοτικό και αξιόπιστο σύστημα υπολογισμού επιρροής. Ως συμπέρασμα της έρευνας της παρούσας εργασίας προκύπτει επίσης, ότι τα συστήματα υπολογισμού επιρροής στα κοινωνικά δίκτυα θα συνεχίζουν να αποτελούν ένα σημαντικό πεδίο έρευνας με πολλαπλές εφαρμογές, μιας και η ενασχόληση με τα κοινωνικά δίκτυα και την αξιοποίησή τους αυξάνεται διαρκώς.

ΑΝΑΦΟΡΕΣ

- [1] Ahmed, S. and Ezeife, C. I. 2013. Discovering influential nodes from trust network. In *Proc. of the 28th Annual ACM Symposium on Applied Computing (SAC '13)*, Coimbra, Portugal, March 18-22, 2013, pp. 121-128. ACM, New York, NY, USA.
- [2] Agarwal, N., Liu, H., Tang, L., and Yu, P. S. Identifying the influential bloggers in a community, In *Proc. of the International Conference on Web Search and Web Data Mining (Palo Alto, California, USA, February 11 - 12, 2008)*. WSDM '08. ACM, New York, NY, 207-218.
- [3] Koutrouli, E., Kanellopoulos, G., and Tsalgatidou, A. 2016. Reputation Mechanisms in on-line Social Networks – The case of an Influence Estimation System in Twitter, accepted for publication in *South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (IEEE SEEDA-CECNSM 2016)*, ACM.
- [4] Varlamis, I., Eirinaki, M., Louta, M.: Application of social network metrics to a trust-aware collaborative model for generating personalized user recommendations. In: *The Influence of Technology on Social Network Analysis and Mining*. LNSN, vol. 6, pp. 49–74. Springer (2013)
- [5] Carolina Bento. Finding influencers in social networks, MSc thesis Instituto Superior Técnico - Lisbon Tech/ INESC-ID, work in progress
- [6] Amit Goyal, Francesco Bonchi, and Laks V.S. Lakshmanan. 2010. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining (WSDM '10)*. ACM, New York, NY, USA, 241-250. DOI=<http://dx.doi.org/10.1145/1718487.1718518>
- [7] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. 2014. What makes an image popular?. In *Proceedings of the 23rd international conference on World wide web (WWW '14)*. ACM, New York, NY, USA, 867-876. DOI: <http://dx.doi.org/10.1145/2566486.2567996>
- [8] Isabel Anger and Christian Kittl. 2011. Measuring influence on Twitter. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW '11)*, Stefanie Lindstaedt and Michael Granitzer (Eds.). ACM, New York, NY, USA, , Article 31 , 4 pages. DOI=<http://dx.doi.org/10.1145/2024288.2024326>
- [9] <https://www.yelp.com> [Τελευταία προσπέλαση 30/10/18]
- [10] <https://en.wikipedia.org/wiki/Klout> [Τελευταία προσπέλαση 30/10/18]
- [11] <https://www.socialmediatoday.com/social-business/adhutchinson/2015-10-31/wanna-know-how-your-klout-score-really-calculated> [Τελευταία προσπέλαση 30/10/18]
- [12] <https://www.lithium.com/pdfs/datasheets/Lithium-Company-Overview.pdf> [Τελευταία προσπέλαση 30/10/18]
- [13] <https://www.lithium.com/products/klout> [Τελευταία προσπέλαση 30/10/18]
- [14] <https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined> [Τελευταία προσπέλαση 30/10/18]
- [15] <https://www.alexa.com/siteinfo> [Τελευταία προσπέλαση 30/10/18]
- [16] <https://www.curiousgnu.com/tweetanalyzer> [Τελευταία προσπέλαση 30/10/18]
- [17] http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010 [Τελευταία προσπέλαση 30/10/18]
- [18] <https://tweetreach.com/about/2011/07/what-is-reach-and-why-does-it-matter/> [Τελευταία προσπέλαση 30/10/18]
- [19] <https://unionmetrics.zendesk.com/hc/en-us/articles/201201556-Understanding-the-TweetReach-snapshot-report> [Τελευταία προσπέλαση 30/10/18]
- [20] <https://unionmetrics.com/use-cases/optimize-content/> [Τελευταία προσπέλαση 30/10/18]
- [21] <https://unionmetrics.com/product/> [Τελευταία προσπέλαση 30/10/18]