

# SOCIAL NETWORKS ΚΑΙ ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΑΣ

Εργασία στο μάθημα Ανάκτησης  
Πληροφορίας

Γεωργίου Νίκη (ΑΜ: 6282)

Κουλεντιανού Μαρία Νεκταρία (ΑΜ: 6310)

# Περιεχόμενα:

Εισαγωγή .....	σελ.3
1. Το Φαινόμενο του Μικρόκοσμου .....	σελ.5
2. Το Πείραμα του Milgram .....	σελ.7
3. Δίκτυα .....	σελ.11
3.1 Δίκτυα που Υποστηρίζουν την Αποτελεσματική Αναζήτηση .....	σελ.11
3.2 Ανάπτυξη Δικτύου .....	σελ.14
3.3 Πλοήγηση Δικτύου .....	σελ.16
3.4 Ιεραρχικά Μοντέλα Δικτύου .....	σελ.17
4. Ανάκτηση Πληροφορίας .....	σελ.20
4.1 Μοντέλο Ανάκτησης Πληροφορίας .....	σελ.21
4.2 Ανάκτηση Πληροφοριών στις Folksonomies .....	σελ.23
4.3 Αναζήτηση και Κατάταξη Χρησιμοποιώντας Προσαρμοσμένο PageRank.....	σελ.24
Συμπεράσματα .....	σελ.25
Βιβλιογραφία .....	σελ.27

## Εισαγωγή:

Η ανάκτηση πληροφορίας ή αλλιώς Information Retrieval, είναι ένα από τα επιστημονικά πεδία της πληροφορικής το οποίο μελετά την αποδοτική αναζήτηση πληροφοριών και κάθε τύπου δεδομένων μέσα σε έγγραφα, όπως επίσης και με την αναζήτηση σε βάσεις δεδομένων και στον Παγκόσμιο Ιστό. Η ανάκτηση πληροφοριών στηρίζεται στη θεωρία των βάσεων δεδομένων, σε κατάλληλα υπολογιστικά συστήματα και σε μαθηματικές μεθόδους της τεχνητής νοημοσύνης, ενώ εφαρμόζεται στην επιστήμη πληροφόρησης και στον Παγκόσμιο Ιστό με μηχανές αναζήτησης.

Κατά τη διαδικασία ανάκτησης πληροφοριών ένας χρήστης εισάγει ένα ερώτημα στο σύστημα, δηλαδή μια αίτηση αναζήτησης σε κάποια βάση δεδομένων, όπως για παράδειγμα η αναζήτηση ενός αλφαριθμητικού σε κάποια μηχανή αναζήτησης του ιστού. Η αναζήτηση θα βγάλει πολλά αντικείμενα ως αποτελέσματα στο ερώτημα του χρήστη τα οποία μπορεί να ταιριάζουν στο ερώτημα με διαφορετικούς βαθμούς σχετικότητας. Ένα αντικείμενο είναι μια οντότητα που αντιπροσωπεύεται από κάποιες πληροφορίες σε μία βάση δεδομένων. Για να απαντηθεί το ερώτημα του χρήστη, γίνεται αναζήτηση ανάμεσα στα στοιχεία της βάσης για πιθανές απαντήσεις. Τα δεδομένα μπορεί να είναι έγγραφα κειμένων, εικόνες, αρχεία ήχου ή βίντεο. Συνήθως τα περισσότερα συστήματα ανάκτησης πληροφοριών υπολογίζουν μια αριθμητική βαθμολογία για το πόσο καλά κάθε αντικείμενο στη βάση δεδομένων ταιριάζει με το ερώτημα και κατατάσσει τα αντικείμενα σύμφωνα με την τιμή αυτή. Τα αντικείμενα που βρίσκονται στη κορυφή της κατάταξης εμφανίζονται στο χρήστη. Η διαδικασία αυτή μπορεί να επαναλαμβάνεται, αν ο χρήστης επιθυμεί να βελτιώσει το ερώτημα.

Ο τεράστιος όγκος δεδομένων μαζί με την ανάγκη για αξιοποίηση της πληροφορίας αυτής δημιουργούν την ανάγκη νέων τεχνολογιών και μεθόδων εξόρυξης της. Η ραγδαία ανάπτυξη της τεχνολογίας και των μεθόδων εξαγωγής μεγάλου όγκου δεδομένων, οδηγεί ταυτόχρονα στην αύξηση των αναγκών για την εύρεση βελτιωμένων τεχνικών, όσον αφορά τον χρόνο ανταπόκρισης αλλά και την απόδοση.

Τα κοινωνικά δίκτυα, τα οποία χρησιμοποιούνται από εκατομμύρια ανθρώπους καθημερινά για επικοινωνία, ένταξη σε διάφορες κοινότητες που έχουν ίδια ενδιαφέροντα με αυτούς, διασκέδαση, οργάνωση, έκφραση απόψεων κ.α., είναι το μέσω το οποίο θα μελετήσουμε για την ανάκτηση των πληροφοριών και την μετέπειτα αποθήκευσή τους σε μια βάση δεδομένων. Η πληροφορία που θα εξάγεται κάθε φορά θα ικανοποιεί κάποιους περιορισμούς, για παράδειγμα θα αφορά συγκεκριμένες γεωγραφικές θέσεις ή θα αφορά μια προκαθορισμένη

χρονική περίοδο. Κατά αυτόν τον τρόπο με το ερώτημα του χρήστη σε μία μηχανή αναζήτησης τα αποτελέσματα, αντικείμενα, θα είναι εξαγόμενα από κοινωνικά δίκτυα που σχετίζονται με το ερώτημα. Έτσι πετυχαίνουμε μικρό χρόνο ανταπόκρισης και μεγάλη απόδοση στη διαδικασία ανάκτησης πληροφορίας.

## 1. Το φαινόμενο του μικρόκοσμου

Η σύνδεση των ανθρώπων με, διαφόρων ειδών, αλυσίδες γνωριμιών είναι η αρχή που συνιστά το φαινόμενο του μικρόκοσμου. Το τελευταίο είναι ο χώρος πειραματικής μελέτης στις κοινωνικές επιστήμες. Σε πρόσφατη εργασία των Watts και Strogatz διατυπώθηκε με αναμφισβήτητα στοιχεία ότι το φαινόμενο αυτό είναι διαδεδομένο σε μια σειρά δικτύων που προκύπτουν από τη φύση και την τεχνολογία, επομένως αποτελεί ένα βασικό συστατικό στην εξέλιξη του Παγκόσμιου Ιστού.

Παρ' όλα αυτά υπάρχοντα μοντέλα δικτύου φαίνεται να είναι ανεπαρκή για να ερμηνεύσουν την εντυπωσιακή αλγοριθμική συνιστώσα των αρχικών διαπιστώσεων του Stanley Milgram, όπου τη δεκαετία του 1960 μέσα από το έργο του προσπάθησε να αποδείξει ότι η ομάδα ατόμων που χρησιμοποιούν τοπικές πληροφορίες είναι πολύ αποτελεσματική στη πραγματική κατασκευή βραχέων διαδρομών μεταξύ δύο σημείων σε ένα κοινωνικό δίκτυο. Τα σύντομα μονοπάτια των πρόσφατων προτεινόμενων δικτύων δεν αρκούν, διότι έχει αποδειχθεί ότι κανένας αποκεντρωμένος αλγόριθμος – ο οποίος να λειτουργεί μόνο με τοπικές πληροφορίες – που να μπορεί να κατασκευάσει σύντομα μονοπάτια σε αυτά τα δίκτυα με μη αμελητέα πιθανότητα. Έχει οριστεί μια οικογένεια μοντέλων δικτύου που αποτελεί ουσιαστικά τη γενίκευση του μοντέλου Watts – Strogatz και δείχνει ότι για ένα από αυτά τα μοντέλα υπάρχει ένας αποκεντρωμένος αλγόριθμος ικανός να βρει σύντομες διαδρομές με μεγάλη πιθανότητα.

ΤΟ ΦΑΙΝΟΜΕΝΟ ΤΟΥ ΜΙΚΡΟΚΟΣΜΟΥ: Ένα κοινωνικό δίκτυο παρουσιάζει το φαινόμενο του μικρόκοσμου αν δυο άτομα στο δίκτυο είναι πιθανό να συνδεθούν μέσω μιας σύντομης ακολουθίας ενδιάμεσων γνωριμιών. Συχνά συναντάμε έναν ξένο και ανακαλύπτουμε ότι έχουμε μια κοινή γνωριμία, αυτό αποτελούσε μία ανέκδοτη παρατήρηση και από τότε έχει εξελιχθεί σε σημαντικό τομέα σπουδών στις κοινωνικές επιστήμες. Πρόσφατες εργασίες δείχνουν ότι το φαινόμενο είναι διαδεδομένο στα δίκτυα που προκύπτουν από τη φύση και τη τεχνολογία και αποτελεί ένα βασικό συστατικό στη δομική εξέλιξη του Παγκόσμιου Ιστού. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΟΥ ΦΑΙΝΟΜΕΝΟΥ: Η εμπειρική επικύρωση του φαινομένου έχει οδηγήσει σε ένα αναπόφευκτο αναλυτικό έργο που αποσκοπεί στην απάντηση της ακόλουθης γενικής ερώτησης:

Γιατί πρέπει να υπάρχουν μικρές αλυσίδες γνωστών που να συνδέουν μεταξύ τους  
αυθαίρετα ζεύγη ξένων;

Η βασική εξήγηση στο θέμα αυτό, μέσω πειραμάτων του Milgram, είναι ότι τυχαία δίκτυα έχουν μικρή διάμετρο. Αν για παράδειγμα τα Α και Β είναι δύο άτομα με έναν κοινό φίλο, είναι πολύ πιθανό ότι οι ίδιοι είναι φίλοι. Ταυτόχρονα ένα δίκτυο

γνωριμιών που είναι πολύ συσσωρευμένο δεν θα έχει τη χαμηλή διάμετρο που έδειξαν τα πειράματα του Milgram.

Πρόσφατα, προτάθηκε από τους Watts και Strogatz ένα πρότυπο για το φαινόμενο του μικρόκοσμου με βάση μια τάξη τυχαίων δικτύων που παρεμβάλλουν μεταξύ των «τοπικών» και των «μακρινών» επαφών. Αρχικά μελέτησαν ένα πλέγμα επανασυνδεδεμένου δακτυλίου, το οποίο κατασκευάστηκε ως εξής. Ένα άτομο ξεκινά με ένα σύνολο  $V$  από  $n$  σημεία διατεταγμένα ομοιόμορφα σε έναν κύκλο και συνδέει κάθε σημείο με μια άκρη σε καθένα από τους πλησιέστερους γείτονές του, για μια μικρή σταθερά  $k$ . Αυτές είναι οι «τοπικές» επαφές στο δίκτυο. Στη συνέχεια ένα άλλο άτομο εισάγει ένα μικρό αριθμό ακμών στις οποίες επιλέγονται τα τελικά σημεία ομοιόμορφα – τυχαία από το  $V$ . Αυτές είναι οι επαφές «μεγάλου βεληνεκούς» ή αλλιώς οι «μακρινές» επαφές. Οι Watts και Strogatz υποστήριξαν ότι ένα τέτοιο μοντέλο συμπεριλαμβάνει δύο κρίσιμες παραμέτρους των κοινωνικών δικτύων: υπάρχει μια απλή υποκείμενη δομή που εξηγεί τη παρουσία των περισσότερων άκρων, αλλά μερικές ακμές παράγονται από τυχαία διαδικασία που δε σέβεται αυτή τη δομή. Τα δίκτυα τους έτσι έχουν μικρή διάμετρο (όπως τα ομοιόμορφα τυχαία δίκτυα), αλλά έχουν την ιδιότητα πολλοί από τους γείτονες ενός κόμβου  $u$  είναι οι ίδιοι γείτονες (σε αντίθεση με τα ομοιόμορφα τυχαία δίκτυα). Έδειξαν έτσι ότι ένας αριθμός φυσικά αναδυόμενων δικτύων παρουσιάζουν αυτό το ζεύγος ιδιοτήτων. Η προσέγγιση τους εφαρμόστηκε στην ανάλυση του γράφου υπερσύνδεσης του Παγκόσμιου Ιστού.

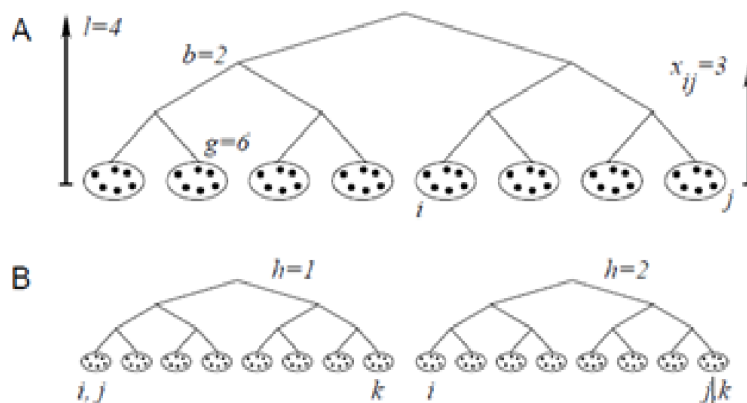
## 2. Το πείραμα του Milgram

Στα τέλη της δεκαετίας του 1960, οι Travers και Milgram διεξήγαγαν ένα πείραμα στο οποίο τυχαία επιλεγμένα άτομα στη Βοστώνη, στη Μασαχουσέτη, στην Ομάχα και στη Νεμπράσκα, κλήθηκαν να κατευθύνουν επιστολές σε ένα άτομο-στόχο στη Βοστώνη. Το κάθε άτομο στέλνει την επιστολή του σε μια μόνο γνωριμία του την οποία έκρινε ότι ήταν πιο κοντά από τον εαυτό του, στόχος. Το ίδιο έκαναν και οι επόμενοι παραλήπτες. Το μέσο μήκος των επακόλουθων αλυσίδων γνωριμίας για τις επιστολές που τελικά έφθασαν στο στόχο ήταν περίπου 6.

Αυτό αποκαλύπτει ότι όχι μόνο σύντομα μονοπάτια υπάρχουν μεταξύ των ατόμων σε μεγάλο κοινωνικό- επίπεδο δίκτυο, αλλά ότι και απλοί άνθρωποι μπορούν να βρουν αυτά τα μονοπάτια. Οι άνθρωποι γνωρίζουν ποιοι είναι οι φίλοι τους, μπορεί να γνωρίζουν επίσης ποιοι είναι μερικοί από τους φίλους των φίλων τους. Αλλά κανείς δεν γνωρίζει την ταυτότητα ολόκληρης της αλυσίδας των ατόμων μεταξύ τους και ενός αυθαίρετου στόχου. Η ιδιότητα του να είναι σε θέση να βρει έναν στόχο γρήγορα, το οποίο ονομάζουμε δυνατότητα αναζήτησης, έχει αποδειχθεί έχει συγκεκριμένες κατηγορίες δικτύων που είτε διαθέτουν ορισμένο κλάσμα κόμβων ή είναι χτισμένα σε υποκείμενο γεωμετρικό πλέγμα που ενεργεί ως πληρεξούσιος για τον "κοινωνικό χώρο". Κανείς από αυτούς τους τύπους δικτύου, ωστόσο, δεν είναι ικανοποιητικός μοντέλο της κοινωνίας.

Υπάρχει μοντέλο για ένα κοινωνικό δίκτυο το οποίο βασίζεται σε κοινωνικές δομές και δίνει μια εξήγηση για το φαινόμενο της αναζήτησης. Το μοντέλο ακολουθείτε από έξι ισχυρισμούς για τα κοινωνικά δίκτυα:

1. Τα άτομα στα κοινωνικά δίκτυα δεν είναι προικισμένα μόνο με δεσμούς δικτύου, αλλά και με ταυτότητες. Με τον όρο ταυτότητα εννοούμε διάφορα σύνολα χαρακτήρων, χαρακτηριστικά που αποδίδουν στον εαυτό τους και στους άλλους λόγω της σύνδεσης τους με τους άλλους και της συμμετοχής τους σε Κοινωνικές Ομάδες. Ο όρος όμιλος αναφέρεται σε οποιαδήποτε συλλογή ατόμων με τα οποία έχει οριστεί ένα καλά καθορισμένο σύνολο με κοινά χαρακτηριστικά.



2.

Τα άτομα διαχωρίζουν τον κόσμο ιεραρχικά σε μια σειρά από στρώματα, δηλαδή διαδικασία διάκρισης με διαίρεση, όπου το ανώτατο επίπεδο αντιστοιχεί σε ολόκληρο τον κόσμο και κάθε διαδοχικά βαθύτερη στρώση αντιπροσωπεύει μια γνωστική διαίρεση σε έναν μεγαλύτερο αριθμό όλο και πιο συγκεκριμένων ομάδων. Αυτή η διαδικασία διάκρισης, μπορεί να επιδιωχθεί μέχρι το επίπεδο των ατόμων, οπότε κάθε άτομο συνδέεται μοναδικά με τη δική του ομάδα. Για τους σκοπούς της αναγνώρισης, ωστόσο, οι άνθρωποι δε το κάνουν συνήθως, αλλά τερματίζουν τη διαδικασία σε επίπεδο όπου το αντίστοιχο μέγεθος της ομάδας  $g$ , όπου  $g$  το μέγεθος της ομάδας, γίνεται γνωστό.

(π.χ. Τα ακαδημαϊκά τμήματα είναι μερικές φορές αρκετά μικρά για να λειτουργήσουν ως ενιαία ομάδα, αλλά τείνουν να χωριστούν σε εξειδικευμένες υποομάδες καθώς μεγαλώνουν)

Ορίζουμε την ομοιότητα  $x_{ij}$  μεταξύ ατόμων  $i$  και  $j$  ως το ύψος του χαμηλότερου κοινού προγόνου τους στην προκύπτουσα ιεραρχία, ρυθμίζοντας το  $x_{ij} = 1$  αν τα  $i$  και  $j$  ανήκουν στην ίδια ομάδα. Η ιεραρχία χαρακτηρίζεται πλήρως από το βάθος  $l$  και τη σταθερή αναλογία διακλάδωσης  $b$ .

3. Η ένταξη στην ομάδα, εκτός από τον ορισμό της ατομικής ταυτότητας, αποτελεί πρωταρχική βάση για κοινωνική αλληλεπίδραση και ως εκ τούτου γνωριμία. Έτσι, η πιθανότητα γνωριμίας μεταξύ των ατόμων  $i$  και  $j$  μειώνεται με τη μείωση της ομοιότητας των ομάδων στις οποίες ανήκουν αντίστοιχα. Αυτό το μοντέλο επιλέγεται διαλέγοντας ένα άτομο  $i$  τυχαία και μια απόσταση σύνδεσης  $x$  με τη πιθανότητα  $p(x) = c \exp\{-ax\}$  (όπου  $a$  μια συντονισμένη παράμετρος και  $c$  μια σταθερά κανονικοποίησης). Στη συνέχεια επιλέγεται ένας δεύτερος κόμβος  $j$ , ομοιόμορφα μεταξύ όλων των κόμβων που είναι σε απόσταση  $x$  από το  $i$ , συνεχίζεται αυτή η διαδικασία μέχρι να δημιουργηθεί ένα δίκτυο στο οποίο τα άτομα έχουν έναν μέσο αριθμό φίλων  $z$ .



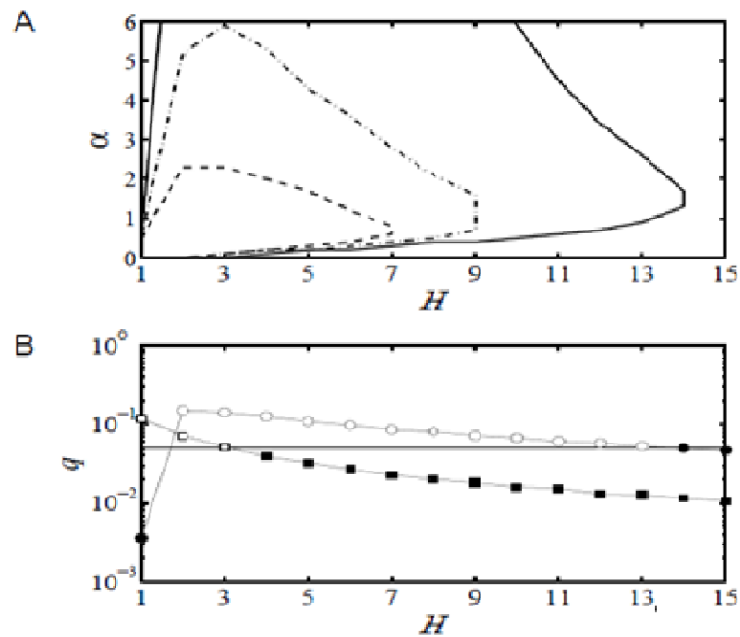
Όταν  $e^{-a} \ll 1$ , όλες οι συνδέσεις θα είναι συντομότερες και τα άτομα θα συνδέονται μόνο με μέλη της δικής τους ομάδας στο κάτω επίπεδο, εξασφαλίζοντας έτσι έναν ομοφυλόφιλο κόσμο, χωρίς στερεότυπα.

Αντίθετα, όταν  $e^{-a} = b$ , κάθε άτομο είναι εξίσου πιθανό να αλληλοεπιδράσει με οποιοδήποτε άλλο, αποδίδοντας ένα ομοιόμορφο τυχαίο γράφημα.

4. Τα άτομα συγκεντρώνουν ιεραρχικά τον κοινωνικό κόσμο σε πολλές κατηγορίες (π.χ. γεωγραφικά και κατά το επάγγελμα), οι οποίες θεωρούνται ανεξάρτητες, καθώς η εγκυρότητα της μιας δεν εγγυάται και της άλλης. Στο μοντέλο εκπροσωπείται κάθε τέτοια κοινωνική διάσταση από μια ανεξάρτητα κατανεμημένη ιεραρχία. Η ταυτότητα ενός κόμβου ορίζεται στη συνέχεια ως διαστασιολογικός συντελεστής  $\vec{u}_i$ ,  $\vec{u}_i^h$ , όπου είναι η θέση του κόμβου  $i$  στην ιεραρχία ή τη διάσταση.

Σε κάθε κόμβο  $i$  εκχωρείται τυχαία μια συντεταγμένη σε κάθε μία από τις διαστάσεις  $H$ , και στη συνέχεια κατανέμονται φίλοι όπως περιγράφεται παραπάνω, όπου τώρα επιλέγει τυχαία μια διάσταση  $h$  που θα χρησιμοποιήσει για κάθε ισοπαλία. Όταν  $H=1$  και  $e^{-a} \ll 1$  η πυκνότητα των δεσμών δικτύου πρέπει να συμμορφώνεται με τον περιορισμό  $z < g$ .

5. Με βάση την αντιληπτή ομοιότητα τους με άλλους κόμβους, τα άτομα κατασκευάζουν ένα μέτρο της «κοινωνικής απόστασης», το οποίο ορίζουμε ως την ελάχιστη υπέρμετρη απόσταση από όλες της διαστάσεις μεταξύ δύο κόμβων  $i$  και  $j$ . Αυτή η ελάχιστη μέτρηση καταγράφει τη διαισθητική αντίληψη ότι η εγγύτητα σε μία μόνο διάσταση αρκεί για να υποδηλώσει την υπαγωγή (π.χ. γεωγραφικά και εθνολογικά μακρινοί ερευνητές που συνεργάζονται στο ίδιο έργο). Μια συνέπεια που απεικονίζεται στο σχήμα Β, είναι ότι η κοινωνική απόσταση παραβιάζει την τριγωνική ανισότητα, επομένως δεν είναι μια πραγματική μετρική απόσταση επειδή τα άτομα  $i$  και  $j$  μπορεί να είναι κοντά στη διάσταση  $h_1$  και τα άτομα  $j$  και  $k$  μπορούν να είναι κλειστά στη διάσταση  $h_2$ , αλλά τα  $i$  και  $k$  μπορούν να διαχωριστούν πολύ και στις δύο διαστάσεις.
6. Τα άτομα διαβιβάζουν ένα μήνυμα σε έναν μόνο γείτονα, δίνοντας μόνο τοπικές πληροφορίες σχετικά με το δίκτυο.



Υποθέτουμε:

- Κάθε κόμβος γνωρίζει μόνο το δικό του συντεταγμένο διάνυσμα
- Τους συντεταγμένους φορείς των άμεσων γειτόνων του δικτύου του
- Το διάνυσμα συντεταγμένων ενός συγκεκριμένου στόχου ατομικό

Αλλά αλλιώς αγνοεί τις ταυτότητες ή τους δεσμούς δικτύου κόμβων πέρα από τον άμεσο κύκλο γνωριμιών του.

Τα άτομα έχουν δύο είδη μερικών πληροφοριών: την κοινωνική απόσταση (η οποία μπορεί να μετρηθεί σε παγκόσμια κλίμακα, αλλά δεν είναι μια πραγματική απόσταση). Παρόλο που κανένα είδος πληροφορίας από μόνο του δεν επαρκεί για την εκτέλεση αποτελεσματικών αναζητήσεων, εδώ δείχνουμε ότι ένας απλός αλγόριθμος που συνδυάζει τη γνώση των δεσμών δικτύου και της κοινωνικής ταυτότητας μπορεί να επιτύχει να κατευθύνει τα μηνύματα με αποτελεσματικότητα. Ο αλγόριθμος που εφαρμόστηκε είναι ο ίδιος άπληστος αλγόριθμος που πρότεινε ο Milgram: κάθε μέλος  $i$  μιας αλυσίδας μηνυμάτων διαβιβάζει το μήνυμα στον γείτονά του  $j$  ο οποίος αντιλαμβάνεται ότι είναι πιο κοντά στον στόχο  $t$  από την άποψη της κοινωνικής απόστασης που είναι, στη γειτονιά του δικτύου  $i$ .

### 3. Δίκτυα

#### 3.1 Δίκτυα που υποστηρίζουν την αποτελεσματική αναζήτηση

Ένα μοντέλο ενός «πλοηγήσιμου» δικτύου αποτελείται από μερικά βασικά χαρακτηριστικά. Θα πρέπει να περιέχει μικρές διαδρομές μεταξύ όλων – των περισσότερων – ζευγών κόμβων. Για να είναι μη-τετριμμένη η δομή του θα πρέπει να είναι μισή γνωστή και μισή άγνωστη στους κόμβους του. Με αυτό το τρόπο οι πληροφορίες σχετικά με τα γνωστά σημεία μπορούν να χρησιμοποιηθούν για τη κατασκευή διαδρομών που χρησιμοποιούν και τα άγνωστα. Έτσι ήταν προφανές τι συνέβαινε στα πειράματα του Milgram. Οι συμμετέχοντες χρησιμοποιώντας τις πληροφορίες που είχαν στη διάθεσή τους εκτιμούσαν ποιοί από τους γνωστούς τους θα οδηγούσαν στο μικρότερο μονοπάτι μέσα από το πλήρες κοινωνικό δίκτυο. Με γνώμονα αυτές τις παρατηρήσεις στρεφόμαστε προς το έργο των Strogatz και Watts, το οποίο προτείνει ένα μοντέλο «δικτύων μικρόκοσμου» που ενσωματώνει πολύ σύντομα αυτά τα χαρακτηριστικά. Μια απλή παραλλαγή του μοντέλου τους θα μπορούσε να περιγραφεί ως εξής: Ξεκινώντας με ένα π-διάστατο πλέγμα, στο οποίο οι κόμβοι συνδέονται μόνο με τους πλησιέστερους γείτονές τους. Προσθέτοντας, τότε,  $k$  κατευθυνόμενες συνδέσεις μεγάλης απόστασης από κάθε κόμβο  $v$ , για μια σταθερά  $k$ . Το τελικό σημείο κάθε συνδέσμου επιλέγεται ομοιόμορφα τυχαία. Τα αποτελέσματα της θεωρίας των τυχαίων γραφημάτων μπορούν να χρησιμοποιηθούν για να δείξουν ότι με μεγάλη πιθανότητα θα υπάρχουν σύντομες διαδρομές που θα συνδέουν όλα τα ζεύγη κόμβων. Παράλληλα το δίκτυο θα διατηρήσει τοπικά μια δομή που μοιάζει με πλέγμα. Κριτήριό μας για την «σμίκρυνση» των διαδρομών είναι ότι θα πρέπει να υπάρχουν διαδρομές των οποίων το μήκος να οριοθετείται από ένα πολυώνυμο  $\log n$ , όπου  $n$  είναι ο αριθμός των κόμβων. Θα αναφερόμαστε σε τέτοιες λειτουργίες ως πολυλογαριθμικές.

Αυτό το μοντέλο δικτύου ως μια υπέρθεση ενός πλέγματος και ένα σύνολο συνδέσμων μεγάλης εμβέλειας, μελετά τη συμπεριφορά ενός αλγόριθμου αποκεντρωμένης αναζήτησης. Ο αλγόριθμος γνωρίζει τη δομή του πλέγματος. Ξεκινά από ένα κόμβο  $s$  με στόχο τις συντεταγμένες ενός κόμβου  $t$ . Διασχίζει διαδοχικά τους συνδέσμους του δικτύου έτσι ώστε να επιτευχθεί ο στόχος όσο το δυνατόν γρηγορότερα, αλλά δε γνωρίζει τους συνδέσμους μεγάλης εμβέλειας από τους κόμβους που δεν έχει επισκευθεί ακόμα. Εκτός από τη κίνηση προς τα εμπρός μέσω κατευθυνόμενων συνδέσεων, ο αλγόριθμος μπορεί να ταξιδέψει αντίστροφα σε κάθε σύνδεσμο που έχει ήδη ακολουθήσει στην προς τα εμπρός κατεύθυνση. Αυτό του επιτρέπει να κάνει πίσω όταν δε θέλει να συνεχίσει την εξερεύνηση από τον τρέχοντα κόμβο. Θα μπορούσαμε να κατανοήσουμε τη

λειτουργία αυτή αν το σκεφτούμε σαν να πατά κάποιος το «κουμπί πίσω» σε ένα πρόγραμμα περιήγησης στο Internet – ή να επιστρέφει το γράμμα στον προηγούμενο κάτοχό του στα πειράματα του Milgram, με οδηγίες ότι αυτός ή αυτή θα πρέπει να δοκιμάσει κάποιον άλλο. Λέμε ότι ο αλγόριθμος έχει χρόνο αναζήτησης  $Y(n)$  εάν σε ένα τυχαία παραγόμενο δίκτυο  $n$  κόμβων με  $s$  και  $t$  επιλεγμένα ομοιόμορφα τυχαία, φτάνει στο στόχο  $t$  σε το πολύ  $Y(n)$  βήματα με πιθανότητα τουλάχιστον  $1-\varepsilon(n)$ , για μια συνάρτηση  $\varepsilon()$  που πηγαίνει στο 0 με  $n$ .

Το πρώτο αποτέλεσμα είναι ότι κανένας αποκεντρωμένος αλγόριθμος δεν μπορεί να επιτύχει έναν πολυλογαριθμικό χρόνο αναζήτησης σε αυτό το μοντέλο δικτύου – παρόλο που πιθανότατα υπάρχουν μονοπάτια πολυλογαριθμικού μήκους που συνδέουν όλα τα ζεύγη κόμβων. Ωστόσο, αν γενικεύσουμε το μοντέλο ελαφρώς, τότε μπορεί να υποστηρίξει την αποτελεσματική αναζήτηση. Ειδικά όταν κατασκευάζουμε έναν σύνδεσμο μεγάλης απόστασης  $(v,w)$  από το  $v$ , δεν επιλέγουμε  $w$  ομοιόμορφα τυχαία. Αντίθετα, το επιλέγουμε με πιθανότητα ανάλογη με  $d^{-\alpha}$ , όπου  $d$  είναι η απόσταση πλέγματος από τον  $v$  στον  $w$ . Με αυτό το τρόπο οι σύνδεσμοι μεγάλης απόστασης συσχετίζονται με τη γεωμετρία του πλέγματος. Αν το  $\alpha$  είναι ίσο με το  $\rho$ , όπου  $\rho$  η διάσταση του υποκείμενου πλέγματος, τότε ένας αποκεντρωμένος άπληστος αλγόριθμος επιτυγχάνει χρόνο αναζήτησης ανάλογο με  $\log_2 n$  και για οποιαδήποτε άλλη τιμή του  $\alpha$  δεν υπάρχει αποκεντρωμένος αλγόριθμος με πολυλογαριθμικό χρόνο αναζήτησης.

Πρόσφατες εργασίες των Zhang, Goel και Govindan έδειξαν ότι η κατανομή των συνδέσεων που σχετίζονται με τη βέλτιστη τιμή της  $\alpha$  μπορεί να οδηγήσει σε βελτιωμένη απόδοση για αποκεντρωμένη αναζήτηση στο σύστημα Freenet peer-to-peer. Οι Adamic, Lukose, Puniyani και Huberman έχουν θεώρηση πρόσφατα μια παραλλαγή του προβλήματος της αποκεντρωμένης αναζήτησης σε ένα δίκτυο που ουσιαστικά δεν έχει γνωστή βασική δομή. Ωστόσο, όταν ο αριθμός των συνδέσεων που εμπλέκονται σε κόμβους ακολουθεί μία κατανομή νόμου εξουσίας, τότε μια στρατηγική αναζητήσεων που ερευνά κόμβους υψηλού βαθμού μπορεί να είναι αποτελεσματική. Έχουν εφαρμόσει τα αποτελέσματά τους στο σύστημα Gnutella, το οποίο παρουσιάζει μια τέτοια δομή. Σε συνεργασία με του Kempe και Demers, μελετήσαμε πως οι διανομές που είναι αντίστροφα πολυώνυμα στην απόσταση μεταξύ κόμβων μπορούν να χρησιμοποιηθούν στο σχεδιασμό πρωτοκόλλων κουτσομπολιού για τη διάδοση πληροφοριών σε ένα δίκτυο επικοινωνούντων παραγόντων.

Στόχος του παρόντος εγγράφου είναι να εξετάσει γενικότερα το πρόβλημα της αποκεντρωμένης αναζήτησης σε δίκτυα με μερική πληροφόρηση σχετικά με τη βασική δομή. Ενώ ένα πλέγμα δημιουργεί μια φυσική ραχοκοκαλιά του δικτύου, θα θέλαμε να καταλάβουμε σε πιο βαθμό οι αρχές που διέπουν την

αποτελεσματική αποκεντρωμένη αναζήτηση μπορούν να αφαιρεθούν από μια ιεραρχική δομή παρόμοια με πλέγμα. Αρχικά εξετάζουμε τα δίκτυα που παράγονται από μια ιεραρχική δομή και δείχνουν ότι μπορούν να ληφθούν ποιοτικά παρόμοια αποτελέσματα. Στη συνέχεια συζητάμε για ένα γενικό μοντέλο δομών ομάδας, το οποίο μπορεί να θεωρηθεί ως ταυτόχρονη γενίκευση των ζωνών και των ιεραρχιών.

Αναφερόμαστε στο  $k$ , τον αριθμό των αναβολών ανά κόμβο, ως το βαθμό του μοντέλου. Οι τεχνικές λεπτομέρειες των αποτελεσμάτων μας – τόσο στις δηλώσεις των αποτελεσμάτων όσο και στις αποδείξεις – είναι απλούστερες αν επιτρέψουμε στο βαθμό του μοντέλου να είναι πολυλογαριθμικός παρά σταθερός.

.

## 3.2 Ανάπτυξη Δικτύου

Διάφορα μοντέλα έχουν προταθεί για την ερμηνεία της κατανομής του νόμου για το βαθμό της ιστοσελίδας. Εδώ γίνεται εστίαση σε γενετικά μοντέλα, τα οποία προσπαθούν να εξηγήσουν την τοπολογία του Ιστού με βάση τη συμπεριφορά μεμονωμένων συγγραφέων που συνδέουν νέες σελίδες με υπάρχουσες. Τα πιο γενετικά μοντέλα βασίζονται σε προτιμησιακό προσάρτημα, οπότε ένας κόμβος κάθε φορά προστίθεται στο δίκτυο με  $m$  νέα άκρα σε υπάρχοντες κόμβους που έχουν επιλεγεί σύμφωνα με ορισμένες η κατανομή πιθανότητας, τυπικά μια συνάρτηση κάποιου χαρακτηριστικού των υφιστάμενων κόμβων. Το πιο γνωστό μοντέλο προνομιακής προσκόλλησης είναι το μοντέλο Barabasi-Albert (BA), όπου ένας κόμβος  $i$  λαμβάνει μια νέα ακμή με πιθανότητα ανάλογη με την τρέχουσα το μοντέλο BA παράγει τοπολογίες δικτύου με διανομές βαθμού εξουσίας, αλλά βασίζεται στην μη ρεαλιστική υπόθεση ότι οι συγγραφείς του διαδικτύου έχουν πλήρη παγκόσμια γνώση σχετικά με το πτυχίο Ιστού. Σύμφωνα με το μοντέλο BA, οι σύνδεσμοι παράγονται σύμφωνα με τη δημοτικότητα μόνο και, κατά συνέπεια, οι παλαιότεροι κόμβοι είναι εκείνοι με τον υψηλότερο βαθμό.

Για να δώσουμε στους νεότερους κόμβους την ευκαιρία να συναγωνιστούν για συνδέσεις, μερικές επεκτάσεις του μοντέλου BA χρησιμοποιούν  $Pr(i)$  και  $(i)k(i)$ , όπου η  $(i)$  είναι η καταλληλότητα της σελίδας  $i$ . Αυτά τα μοντέλα εξακολουθούν να παρέχουν κατανομές βαθμού ισχύος, αλλά μετά από αρκετό χρόνο οι σελίδες που κερδίζουν είναι εκείνες με την υψηλότερη φυσική κατάσταση. Μια άλλη κατηγορία παραλλαγών που επιτρέπουν σε νέες σελίδες να ανταγωνίζονται για συνδέσμους βασίζεται στη σύνδεση σε έναν κόμβο με βάση το βαθμό του με την πιθανότητα ή σε έναν ομοιόμορφα επιλεγμένο κόμβο. Ένα τέτοιο μοντέλο μίγματος δημιουργεί δίκτυα που μπορούν να χωρέσουν όχι μόνο στην κατανομή του βαθμού εξουσίας στο νόμο ολόκληρου του Ιστού, αλλά και στην κατανομή των υποσυνόλων ιστοσελίδων όπως πανεπιστήμια, εταιρείες ή εφημερίδες. Δυστυχώς, όλα αυτά τα μοντέλα εξακολουθούν να υπάρχουν και βασίζονται στην παγκόσμια γνώση του πτυχίου. Επιπλέον, δεν καταφέρνουν να συλλάβουν τις γνωστικές διαδικασίες που οδηγούν τους συγγραφείς να επιλέγουν τις σελίδες που συνδέουν. Ούτε το παγκόσμιο μέτρο καταλληλότητας ούτε η ομοιόμορφη διανομή ταιριάζουν με την ετερογενή φύση του Ιστού και τα συμφέροντα των συγγραφέων.

Το μοντέλο αντιγραφής εφαρμόζει μια πλούσια διαδικασία που ισοδυναμεί με το μοντέλο BA, αλλά χωρίς να απαιτείται ρητή παγκόσμια γνώση των βαθμών κόμβων. Για κάθε νέο κόμβο, ένας πρώην πρωτότυπος κόμβος  $i$  επιλέγεται για πρώτη φορά τυχαία. Στη συνέχεια, κάθε σύνδεσμος  $j$  από τον νέο κόμβο είτε συνδέεται με έναν ομοιόμορφα επιλεγμένο κόμβο με πιθανότητα, είτε με τον στόχο του  $i$ . Οι κόμβοι υψηλότερου βαθμού ευνοούνται αυτόματα από τον μηχανισμό αντιγραφής,

παράγοντας τον ίδιο νόμο εξουσίας βαθμούς ως το μοντέλο BA. Ο κόμβος του πρωτότυπου θα μπορούσε να αντιστοιχεί σε μια σχετική σελίδα γνωστή στον δημιουργό, διαμορφώνοντας έτσι ένα είδος τοπικής διαδικασίας επιλογής.

Ένα γενετικό μοντέλο που χρησιμοποιεί τοπικές πληροφορίες ακόμη πιο ρητά προτάθηκε πρόσφατα για τα δέντρα. Στους κόμβους δίδονται τυχαίες συντεταγμένες στο τετράγωνο μονάδων. Στη συνέχεια, οι νέοι κόμβοι συνδέονται με υπάρχοντες με βάση έναν γραμμικό συνδυασμό ζεύξης και γεωγραφικής απόκλισης. Σε αυτό το μοντέλο ένας νέος κόμβος  $j$  συνδέεται αιτιοκρατικά με τον κόμβο  $\arg \min_i (\varphi r_{ij} + g_i)$ , όπου  $r_{ij}$  είναι η Euclidean απόσταση και  $g_i$  είναι ένα μέτρο του  $i$  στο δέντρο. Δείχνεται ότι για τις κρίσιμες τιμές του  $\varphi$  (όπου  $\varphi \sim$  είναι μια σταθερά και  $N$  είναι ο αριθμός των κόμβων), αυτό το μοντέλο αποδίδει μια κατανομή ισχύος βαθμού ισχύος. Η ευκλείδεια απόσταση εδώ είναι ανεξάρτητη από την τοπολογία της σύνδεσης και επομένως θα μπορούσε να χρησιμοποιηθεί για να μοντελοποιήσει τους παράγοντες που οδηγούν τους συγγραφείς να συνδέουν τις σελίδες τους με άλλες σελίδες βασισμένες, για παράδειγμα, στο περιεχόμενο και όχι μόνο στη δημοτικότητα. Ωστόσο, το μοντέλο γραμμικού συνδυασμού απαιτεί ακόμα παγκόσμια γνώση της δομής του δέντρου και η ομοιόμορφη κατανομή των κόμβων στο τετράγωνο μονάδων ως γεωγραφικό μοντέλο δεν καταγράφει μια ρεαλιστική τοπολογία περιεχομένου.

### 3.3 Πλοήγηση Δικτύου

Οι σχέσεις μεταξύ της τοπολογίας του συνδέσμου Web και των εννοιών της σημασιολογικής ομοιότητας που απορρέουν από το περιεχόμενο της σελίδας ή τη θεματική ταξινόμηση έχουν σημαντικές εφαρμογές για το σχεδιασμό πιο αποτελεσματικών εργαλείων αναζήτησης. Οι ανιχνευτές με γνώμονα το θέμα θεωρούνται όλο και περισσότερο ως ένας τρόπος αντιμετώπισης των περιορισμών κλιμάκωσης των σημερινών καθολικών μηχανών αναζήτησης, διανέμοντας τη διαδικασία ανίχνευσης σε χρήστες, ερωτήματα ή ακόμα και υπολογιστές-πελάτες. Το πλαίσιο που διατίθεται σε αυτά τα προγράμματα ανίχνευσης μπορεί να καθοδηγήσει την πλοήγηση συνδέσμων με στόχο την αποτελεσματική εύρεση πολύ σχετικών σελίδων στόχων. Με την ανάγκη να βρεθούν άγνωστες σελίδες στόχου, μας ενδιαφέρει μόνο η αποκεντρωμένη ανίχνευση αλγόριθμοι, οι οποίοι μπορούν να χρησιμοποιήσουν μόνο πληροφορίες διαθέσιμες τοπικά σχετικά με μια σελίδα και τη γειτονιά της. Ξεκινώντας από κάποια ιστοσελίδα προέλευσης, στοχεύουμε να επισκεφτούμε μια σελίδα προορισμού με την επίσκεψη  $I < N$  σελίδες, όπου  $N$  είναι το μέγεθος του Ιστού, αρκετά δισεκατομμύρια σελίδες.

Δεδομένου ότι ο ιστός είναι ένα μικρό παγκόσμιο δίκτυο γνωρίζουμε ότι η διάμετρος του ζυγίζεται λογαριθμικά με  $N$ , ή ακριβέστερα μικρού μήκους διαδρομή.

$$I' \sim \frac{\log N}{\log \log N}$$

Εάν οι μόνες τοπικές διαθέσιμες πληροφορίες αφορούν τον βαθμό συνδέσμου υπερκειμένου κάθε κόμβου και των γειτόνων του, τότε απλοί άπληστοι αλγόριθμοι που πάντοτε επιλέγουν τον γείτονα με τον υψηλότερο βαθμό οδηγούν σε μονοπάτια των οποίων το μήκος, ζυγίζει λογαριθμικός ( $I' \sim \log N$ ). Ωστόσο, ένας πραγματικός ανιχνευτής ιστού θα πρέπει να επισκεφθεί όλους τους γείτονες μιας σελίδας για να καθορίσει το πτυχίο τους. Λόγω της κατανομής βαθμού εξουσίας, η μετάβαση σε κόμβους υψηλού βαθμού οδηγεί στην κάλυψη των περισσότερων τους κόμβους σε ένα μικρό αριθμό βημάτων.

Οι Kleinberg και Watts. χαρακτήρισαν ορισμένες κατηγορίες δικτύων των οποίων η τοπολογία συνδέσεων εξαρτάται από εξωτερικές γεωγραφικές ή ιεραρχικές δομές, που δείχνει ότι σε αυτές τις ειδικές περιπτώσεις οι αλγόριθμοι πλοήγησης είναι εγγυημένοι ότι υπάρχουν με πολυγραμματογραφικά όρια.



### 3.4 Ιεραρχικά Μοντέλα Δικτύου

Σε πολλές ρυθμίσεις οι κόμβοι αντιπροσωπεύουν αντικείμενα που μπορούν να ταξινομηθούν σύμφωνα με μία ιεραρχία ή ταξινόμηση και οι κόμβοι είναι πιθανότερο να σχηματίσουν συνδέσμους αν ανήκουν στο ίδιο μικρό υπό-δέντρο στην ιεραρχία, δείχνοντας ότι είναι πιο στενά συνδεδεμένοι.

Γι' αν κατασκευάσουμε ένα μοντέλο δικτύου με αυτή την ιδέα παρουσιάζουμε την ιεραρχία χρησιμοποιώντας ένα πλήρες  $b$ -ary  $T$  δέντρο, όπου το  $b$  είναι μια σταθερά. Έστω το  $V$  το σύνολο των φύλλων του  $T$ . Χρησιμοποιούμε το  $n$  για να δηλώσουμε το μέγεθος του  $V$ , και για τα δύο φύλλα  $v$  και  $w$ , χρησιμοποιούμε  $h(v, w)$  για να δηλώσουμε το ύψος του ελάχιστου κοινού προγόνου των  $v$  και  $w$  στο  $T$ . Δίνεται επίσης μια μονότονη, μη αυξανόμενη συνάρτηση  $f()$  που θα καθορίσει τις πιθανότητες σύνδεσης. Για κάθε κόμβο  $v \in V$  δημιουργούμε μια τυχαία σύνδεση στο  $w$  με πιθανότητα ανάλογη προς το  $f(h(v, w))$ . Με άλλα λόγια η πιθανότητα επιλογής  $w$  είναι ίση με  $f(h(v, w)) / \sum_{x \neq v} f(h(v, x))$ . Δημιουργούμε  $k$  συνδέσεις από κάθε κόμβο  $v$  κατά αυτό το τρόπο επιλέγοντας το τελικό σημείο  $w$  κάθε φορά ανεξάρτητα και με δυνατότητα επανάληψης. Αυτό οδηγεί σε ένα γράφημα  $G$  στο σύνολο  $V$ .

Για αυτήν την ανάλυση έστω ότι ο βαθμός out-degree είναι  $k = c \log 2n$ , για μια σταθερά  $c$ . Είναι σημαντικό να σημειωθεί ότι το δέντρο  $T$  χρησιμοποιείται μόνο στη διαδικασία παραγωγής του  $G$ . Στο  $G$  δεν εμφανίζονται ούτε οι άκρες ούτε οι κόμβοι του δέντρου  $T$  που δεν έχει φύλλα. Χρησιμοποιώντας τον όρο «κόμβος» χωρίς καμία βαθμολόγηση, αναφερόμαστε στους κόμβους του  $G$  και συνεπώς τα φύλλα του  $T$ , και τον όρο «εσωτερικός κόμβος» για να αναφερθούμε σε κόμβους του  $T$  που δεν έχουν φύλλα.

Αναφερόμαστε στη διαδικασία που παράγει το  $G$  ως ιεραρχικό μοντέλο με εκθέτη  $\alpha$  εάν η συνάρτηση  $f(h)$  αυξάνεται ασυμπτωτικά όπως η  $b^{-\alpha h}$ :

$$\lim_{h \rightarrow \infty} \left( \frac{f(h)}{b^{-\alpha' h}} \right) = 0, \text{ όπου } \alpha' < \alpha \text{ και } \lim_{h \rightarrow \infty} \left( \frac{b^{-\alpha'' h}}{f(h)} \right) = 0, \text{ όπου } \alpha'' > \alpha.$$

Υπάρχουν πολλές φυσικές ερμηνείες για ένα ιεραρχικό πρότυπο δικτύου. Το ένα είναι από την άποψη του Παγκόσμιου Ιστού, όπου  $T$  είναι ιεραρχία θεμάτων, όπως η [www.yahoo.gr](http://www.yahoo.gr). Κάθε φύλλο του  $T$  αντιστοιχεί σε μια ιστοσελίδα και η διαδρομή του από τη ρίζα προσδιορίζει ολόένα και πιο λεπτομερή περιγραφή του θέματος της σελίδας. Έτσι ένα συγκεκριμένο φύλλο μπορεί να σχετίζεται με την «Επιστήμη / Επιστήμη των Υπολογιστών / Αλγόριθμοι ή με την Τέχνη / Μουσική / Όπερα». Οι πιθανότητες σύνδεσης έχουν τότε ένα απλό νόημα- βασίζονται δηλαδή στην απόσταση μεταξύ των θεμάτων των σελίδων, όπως μετράται από το ύψος του ελάχιστου κοινού προγόνου τους στην ιεραρχία θεμάτων. Μία σελίδα σχετικά με

την «Επιστήμη / Επιστήμη των Υπολογιστών / Αλγόριθμοι», μπορεί να είναι πιο πιθανό να συνδεθεί σε μια σελίδα με την «Επιστήμη / Επιστήμη των Υπολογιστών / Μάθηση» των Μηχανών παρά σε μία στην «Τέχνη / Μουσική / Όπερα». Φυσικά το μοντέλο αποτελεί μια ισχυρή απλοποίηση αφού οι δομές του θέματος δεν είναι πλήρως ιεραρχικές και σίγουρα δεν έχουν ομοιόμορφη διακλάδωση και βάθος. Αξιοσημείωτο είναι το γεγονός ότι ορισμένα πρόσφατα μοντέλα για τη σύνδεσης της δομής του Ιστού, καθώς και άλλες σχεσιακές δομές εξέτασαν τους διαφορετικούς τρόπους με τους οποίους οι ομοιότητες στο περιεχόμενο μπορούν να επηρεάσουν τη πιθανότητα σύνδεσης.

Μία άλλη ερμηνεία του ιεραρχικού μοντέλου είναι από την άποψη του αρχικού πειράματος του Milgram. Οι μελέτες που πραγματοποιήθηκαν από τους Killworth και Bernand έδειξαν ότι κατά την επιλογή ενός παραλήπτη της επιστολής, οι συμμετέχοντες καθοδηγούσαν συντριπτικά ένα από τα δύο κριτήρια:

- Ομοιότητα με τον γεωγραφικό στόχο
- Ομοιότητα με το στόχο στη κατοχή.

Αν κάποιος βλέπει το πλέγμα να σχηματίζει ένα απλό μοντέλο γεωγραφικών παραγόντων, το ιεραρχικό πρότυπο μπορεί να ερμηνευτεί κατά παρόμοιο τρόπο ως μια ιεραρχία θεμάτων των επαγγελμάτων, με άτομα στα φύλλα. Έτσι, για παράδειγμα, τα επαγγέλματα του «τραπεζίτη» και του «μεσίτη μετοχών» μπορεί να ανήκουν στο ίδιο μικρό υπό-δέντρο, δεδομένου ότι ο στόχος σε ένα από τα πειράματα του Milgram ήταν χρηματιστής, μπορεί να είναι μια καλή στρατηγική για την αποστολή της επιστολής σε έναν τραπεζίτη. Οι Watts, Dodds, και Newman έχουν μελετήσει πρόσφατα ιεραρχικές δομές για τη μοντελοποίηση του πειράματος του Milgram στα κοινωνικά δίκτυα.

Ας εξετάσουμε το πρόβλημα αναζήτησης σε ένα γράφημα  $G$  που παράγεται από ένα ιεραρχικό μοντέλο: Ένας αποκεντρωμένος αλγόριθμος έχει γνώση του δέντρου  $T$  και γνωρίζει τη θέση ενός φύλλου-στόχου που πρέπει να φτάσει. Ωστόσο, μαθαίνει μόνο τη δομή του  $G$  καθώς επισκέπτεται κόμβους. Ο εκθέτης  $\alpha$  καθορίζει τον τρόπο με τον οποίο συνδέονται οι δομές των  $G$  και  $T$ . Πως αυτό επιδρά στη πλεύση του  $G$ ; Στην ανάλυση του μοντέλου πλέγματος που εξετάσαμε προηγουμένως, η βασική ιδιότητα του βέλτιστου εκθέτη ήταν ότι από οποιοδήποτε σημείο υπήρχε εύλογη πιθανότητα μιας μακρινής σύνδεσης να μειώσει την απόσταση προς το στόχο κατά το ήμισυ. Χρησιμοποιώντας μια παρόμοια ιδέα: όταν  $\alpha = 1$ , υπάρχει πάντα μια λογική πιθανότητα να βρεθεί μια μακρινή σύνδεση σε ένα μικρότερο υπό-δέντρο που περιείχε το στόχο. Όπως αναφέρθηκε παραπάνω, εστιάζουμε στην περίπτωση του πολυλογαριθμικού out-degree, με τη περίπτωση του συνεχούς out-degree να αναβάλλεται αργότερα.

### ΘΕΩΡΗΜΑ 2.1

**α)** Υπάρχει ένα ιεραρχικό μοντέλο με εκθέτη  $\alpha = 1$  και πολυλογαριθμικό out-degree στο οποίο ένας αποκεντρωμένος αλγόριθμος μπορεί να επιτύχει τον χρόνο αναζήτησης  $O(\log n)$

**β)** Για κάθε  $\alpha \neq 1$ , δεν υπάρχει ιεραρχικό μοντέλο με εκθέτη  $\alpha$  και πολυλογαριθμικό εξωτερικό βαθμό out-degree, στο οποίο ένας αποκεντρωμένος αλγόριθμος μπορεί να επιτύχει πολυλογαριθμικό χρόνο αναζήτησης.

## 4. Ανάκτηση Πληροφορίας

Τα πασίγνωστα μοντέλα ΑΠ είναι ο Boolean, ο διανυσματικός χώρος, ο ασαφής και η απεικόνιση. Αυτά έχουν μελετηθεί λεπτομερώς και υλοποιούνται για πειραματισμούς και εμπορικούς σκοπούς. Παρ' όλ' αυτά, οι γνωστοί περιορισμοί αυτών των μοντέλων έχουν προκαλέσει ερευνητές να προτείνουν νέα μοντέλα. Ένα τέτοιο μοντέλο είναι το λογικό μοντέλο είναι το λογικό μοντέλο για την ΑΠ. Τα τελευταία χρόνια έγιναν αρκετές προσπάθειες για τον ορισμό μιας λογικής για την ανάκτηση της πληροφορίας κατά μήκος της λεγόμενης λογικής προσέγγισης. Τα λογικά ΑΠ μοντέλα μελετήθηκαν για να παρέχουν μια πλούσια και ομοιόμορφη αναπαράσταση της πληροφορίας και της σημασιολογίας της, με στόχο τη βελτίωση της αποτελεσματικής ανάλυσης. Σύμφωνα με αυτές τις λογικές μελέτες, υπάρχουν επίσης έρευνες που ασχολούνται με την απεικόνιση στην ΑΠ. Η ιδέα αυτή προτάθηκε ρητά και εφαρμόστηκε κυρίως για την επίλυση του προβλήματος της αβεβαιότητας στην ΑΠ. Επιτρέπεται ένας πιο σύνθετος ορισμός της συνάφειας από ότι σε άλλα μοντέλα. Χρησιμοποιώντας παρόμοια προσέγγιση, το μοντέλο αυτό βασίζεται στη λογική-αβεβαιότητα και τη θεωρία των πιθανοτήτων για την ενίσχυση της ΑΠ χρησιμοποιώντας τις προσεγγίσεις εξαγωγής κοινωνικών δικτύων, όπου η έξοδος του συστήματος ΑΠ βασίζεται στη βαθμολογία της συνάφειας, ώστε τα έγγραφα να μπορούν να ταξινομηθούν ανάλογα με τη συνάφεια του ερωτήματος, αλλά και τα μελλοντικά μοντέλα θα ενσωματώνουν τις περιγραφές με ΑΠ και κατά προτίμηση θα ενσωματώνουν την πιθανοτική φύση των αναθέσεων της σημασιολογίας.

Η ανάκτηση πληροφοριών είναι μια τεχνολογία γνώσης που αφορά την αποτελεσματική ανάκτηση πληροφοριών για την μετέπειτα χρήση τους. Η ΑΠ συνήθως περιλαμβάνει την αναζήτηση μη δομημένων ή ημιδομημένων πληροφοριών. Η πρώτη αναφέρεται στο περιεχόμενο του αδόμητου κειμένου (γραπτό ή προφορικό), των εικόνων, των βίντεο και των ήχων. Το τελευταίο αναφέρεται σε σαφώς καθορισμένα μεταδεδομένα που επισυνάπτονται ειδικά στις ιστοσελίδες. Όπως και σε άλλες τεχνολογίες έτσι και η ΑΠ μπορεί επίσης να διατυπωθεί χρησιμοποιώντας τη λογική. Η λογική αιτία είναι η ουσία των καθοριστικών παραδειγμάτων και για την κατανόηση των φαινομένων στο χώρο των γεγονότων. Στη κλασική λογική, όπου τα συμπεράσματα συσχετίζονται συχνά με λογικές συνέπειες: μια ιστοσελίδα  $\omega$  είναι σχετική με ένα ερώτημα  $q$  εάν υποδηλώνει το ερώτημα ή με άλλα λόγια αν το ερώτημα μπορεί να συναχθεί από την ιστοσελίδα, δηλαδή εάν  $\omega \Rightarrow q$  είναι αληθές. Ένα καλό παράδειγμα του ερωτήματος για μια ιστοσελίδα είναι η εισαγωγή βασικών όρων και η αντιστοίχισή τους με τους όρους με τους οποίους οι ιστοσελίδες είναι ευρετηριασμένες.

Στο κοινωνικό δίκτυο οι ερευνητές εκμεταλλεύονται τις κορυφαίες σχετικές ιστοσελίδες και τη σχέση μεταξύ των ιστοσελίδων για να προσδιορίσουν τις πιο

σχετικές οντότητες ή να δημιουργήσουν μια σχέση μεταξύ μιας οντότητας σε μια άλλη με βάση σχετικές ιστοσελίδες. Σημαίνει επίσης πόσο κοντά είναι μια ιστοσελίδα σε μια άλλη.

Θεώρημα: Οι ιστοσελίδες σχετίζονται με μια ανάγκη πληροφόρησης για το κοινωνικό δίκτυο εάν περιέχει τουλάχιστον μια πρόταση η οποία σχετίζεται με τη σχέση μεταξύ δύο οντοτήτων.

Το γεγονός doubleton είναι μια προσπάθεια για να συνδέονται δύο οντότητες όταν αναφέρονται συχνά στο ίδιο πλαίσιο, κυρίως η σχέση συγγραφέας – συνεργάτης στις ακαδημαϊκές τους εργασίες. Αυτή η προσπάθεια πρέπει να διακρίνεται από χαλαρότερες, όπως για παράδειγμα ο χώρος διανυσμάτων στον οποίο οι ιστοσελίδες κατατάσσονται σύμφωνα με ένα μέτρο ομοιότητας με το ερώτημα. Φυσικά, η δημιουργία του κοινωνικού δικτύου είναι βασισμένη σε μια θεραπεία ομοιότητας, κυρίως για να διαπιστωθούν λειτουργικά οι σχέσεις δύναμης με πολλούς τρόπους και τα μοντέλα ΑΠ που βασίζονται στην ομοιότητα δεν έχουν γενικά την θεωρητική ευρωστία των πιθανών μοντέλων. Ωστόσο, η ΑΠ είναι σε θέση να ψάξει αποτελεσματικά μέσω τεράστιων ποσοτήτων δεδομένων, επειδή δημιουργεί ευρετήρια από τις ιστοσελίδες (έγγραφα) στις οποίες υπάρχουν όλες οι ανάγκες πληροφόρησης που είναι πολύ δύσκολο να προσδιοριστούν εκ των προτέρων. Επιπλέον, τα ερωτήματα με τις λέξεις δεν εμφανίζονται πάντα σε σχετικές ιστοσελίδες. Έτσι, η επέκταση του ερωτήματος με συνώνυμο και συναφείς όρους είναι μια δημοφιλής εναλλακτική λύση που κυρίως ενισχύει την ανάκληση των αποτελεσμάτων της αναζήτησης. Επομένως, παρόλο που ένα μέτρο ομοιότητας δεν μπορεί να ερμηνευτεί άμεσα ως πιθανότητα, μπορούμε να χρησιμοποιήσουμε πιθανολογικά συμπεράσματα για να αντιμετωπίσουμε με αβεβαιότητα τη σχέση.

#### 4.1 Μοντέλο Ανάκτησης Πληροφορίας: Προοπτική Λήψης τους Μέσω Κοινωνικών Δικτύων

Το κοινωνικό δίκτυο αντικατοπτρίζει τη μετάβαση από τον ατομικισμό που είναι κοινός σε μια κοινωνική δομή ή από την ανταλλαγή πληροφοριών μεμονωμένα με την πληροφόρηση των σχέσεων στις οποίες ορίζονται οι θεμελιώδεις μονάδες, δηλαδή οι σχέσεις μεταξύ των οντοτήτων. Οι σχέσεις χαρακτηρίζονται από περιεχόμενο, κατεύθυνση και δύναμη. Το περιεχόμενο μιας σχέσης αναφέρεται στον πόρο που ανταλλάσσεται, μια σχέση μπορεί να κατευθύνεται ή να μη κατευθύνεται. Οι σχέσεις επίσης διαφέρουν στη δύναμη.

Θεώρημα: Εάν ένα κοινωνικό δίκτυο μπορεί να εξαχθεί από ιστοσελίδες (ή συλλογή εγγράφων), τότε οι ιστοσελίδες ως χώρος γεγονότων στο χώρο των σχέσεων είναι ένα δίκτυο εγγράφων που αντιπροσωπεύει τη συλλογή εγγράφων.

Σε ένα μοντέλο ΑΠ αυτό το θεώρημα παρέχει ότι  $\omega \Leftrightarrow t_a \wedge t_b$  στο χώρο της σχέσης  $\rho \in R$ , όπου  $\xi$  και  $\zeta$  (εξόρυξη κοινωνικού δικτύου) παράγουν  $\rho$ , δηλαδή πρώτα είναι σε δίκτυα ιστοσελίδων και το δεύτερο είναι σε δίκτυα ηθοποιών.

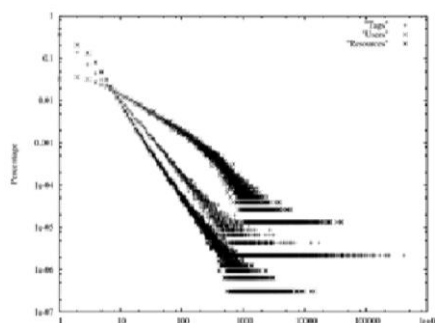
## 4.2 Ανάκτηση Πληροφοριών στις Folksonomies

Τα κοινωνικά εργαλεία σελιδοδεικτών αναδύονται γρήγορα στον Ιστό. Σε αυτά τα συστήματα οι χρήστες εγκαθιστούν ελαφρές εννοιολογικές δομές που ονομάζονται folksonomies. Ο λόγος για την άμεση επιτυχία τους είναι το γεγονός ότι δεν απαιτούνται ειδικές δεξιότητες για τη συμμετοχή. Προς το παρόν, ωστόσο, η υποστήριξη ανάκτησης πληροφοριών είναι περιορισμένη. Το FolkRank είναι ένα επίσημο μοντέλο και ένας νέος αλγόριθμος αναζήτησης για folksonomies και που εκμεταλλεύεται τη δομή της folksonomy. Ο προτεινόμενος αλγόριθμος εφαρμόζεται επίσης για την εύρεση κοινοτήτων εντός της folksonomy και χρησιμοποιείται για τη δομή των αποτελεσμάτων αναζήτησης. Όλα τα ευρήματα καταδεικνύονται σε ένα σύνολο δεδομένων μεγάλης κλίμακας.

Το ερευνητικό ερώτημα είναι πώς να παρέχουμε κατάλληλους μηχανισμούς κατάταξης, παρόμοιους με εκείνους που βασίζονται στη δομή γραφικών του διαδικτύου, αλλά τώρα εκμεταλλεύονται τη δομή των folksonomies αντ' αυτού. Προς το σκοπό αυτό, προτείνεται ένα επίσημο μοντέλο για folksonomies, και παρουσιάζεται ένας νέος αλγόριθμος, που ονομάζεται FolkRank, ο οποίος λαμβάνει υπόψη τη δομή της λαϊκονομίας για την ταξινόμηση των αιτημάτων αναζήτησης στα συστήματα που βασίζονται σε folksonomy. Ο αλγόριθμος θα χρησιμοποιηθεί για δύο σκοπούς: καθορισμός συνολικής κατάταξης και συγκεκριμένης κατάταξης σχετικής με θέματα.

Μια folksonomy περιγράφει τους χρήστες, τους πόρους και τις ετικέτες και την εκχώρηση από τον χρήστη των ετικετών στους πόρους. Παρουσιάζουμε εδώ έναν επίσημο ορισμό των folksonomies, η οποία είναι επίσης υποκείμενη στο σύστημα BibSonomy μας.

Οι χρήστες συνήθως περιγράφονται από το αναγνωριστικό χρήστη τους και οι ετικέτες μπορεί να είναι αυθαίρετες συμβολοσειρές. Αυτό που θεωρείται πόρος εξαρτάται από τον τύπο του συστήματος. Για παράδειγμα, στο del.icio.us, οι πόροι είναι οι διευθύνσεις URL και στο flickr οι πόροι είναι εικόνες. Από την άποψη της εφαρμογής, οι πόροι εκπροσωπούνται εσωτερικά από κάποια ταυτότητα.



Αριθμός εμφανίσεων TAS για ετικέτες, χρήστες, πόρους στο del.icio.us

### 4.3 Αναζήτηση και Κατάταξη Χρησιμοποιώντας Προσαρμοσμένο PageRank

Τα τρέχοντα εργαλεία folksonomy, όπως το del.icio.us, παρέχουν πολύ περιορισμένη υποστήριξη αναζήτησης εκτός από τη διεπαφή περιήγησής τους. Η αναζήτηση μπορεί να γίνει πάνω από το κείμενο των ετικετών και των περιγραφών πόρων, αλλά δεν γίνεται κατάταξη εκτός από την παραγγελία των επισκέψεων με αντίστροφη χρονολογική σειρά. Χρησιμοποιώντας την παραδοσιακή ανάκτηση πληροφοριών, τα περιεχόμενα του folksonomy μπορούν να αναζητηθούν με κείμενο. Ωστόσο, καθώς τα έγγραφα αποτελούνται μόνο από σύντομα αποσπάσματα κειμένου (συνήθως μια περιγραφή, π.χ. τον τίτλο της ιστοσελίδας και τις ίδιες τις ετικέτες), τα κανονικά συστήματα κατάταξης όπως το TF / IDF δεν είναι εφικτά.

Μια folksonomy προκαλεί μια δομή γραφημάτων που θα το κάνουμε εκμεταλλευτείτε για την κατάταξη σε αυτή την ενότητα. Ο αλγόριθμος μας FolkRank είναι εμπνευσμένος από τον Αλγόριθμο PageRank. Η προσέγγιση διάδοσης του βάρους PageRank δεν μπορεί να εφαρμοστεί άμεσα στις folksonomies εξαιτίας της διαφορετικής φύσης των folksonomies σε σύγκριση με το γράφημα ιστού (μη κατευθυνόμενες τριαδικές υπερτάσεις αντί για κατευθυνόμενες δυαδικές ακμές). Στη συνέχεια θα συζητήσουμε πώς να ξεπεραστεί αυτό το πρόβλημα.

Εφαρμόζουμε το σύστημα κατάταξης βάρους στις folksonomies σε δύο βήματα. Κατ' αρχάς, μετατρέπουμε την υπεργράφημα μεταξύ των συνόλων χρηστών, ετικετών και πόρων σε ένα μη κατευθυνόμενο, σταθμισμένο, τριμερές γράφημα. Σε αυτό το γράφημα, εφαρμόζουμε μια έκδοση του PageRank που λαμβάνει υπόψη τα βάρη άκρων.

Μετατροπή της Folksonomy σε μη κατευθυνόμενο γράφημα. Αρχικά μετατρέπουμε τη  $F = (U, T, R, Y)$  σε ένα folksonomy μη κατευθυνόμενο τριμερές γράφημα  $GF = (V, E)$  ως εξής.

1. Το σύνολο  $V$  των κόμβων του γραφήματος αποτελείται από την αποσυνδεδεμένη ένωση των συνόλων ετικετών, οι χρήστες και οι πόροι:  $V = U \cup T \cup R$ .
2. Όλες οι συν-εμφάνσεις ετικετών και χρηστών, χρηστών και πόρων, ετικετών και πόρων καθίστανται μη κατευθυνόμενες, σταθμισμένες ακμές μεταξύ των αντίστοιχων κόμβων:  $E = \{\{u, t\}, \{t, r\}, \{u, r\} \mid (u, t, r) \in Y\}$ , κάθε ακμή  $\{u, t\}$  σταθμίζεται με  $|\{r \in R: (u, t, r) \in Y\}|$ , και κάθε άκρη  $\{u, r\}$  με  $|\{t \in T: (u, t, r) \in Y\}|$ .

Προσαρμοσμένη με Folksonomy PageRank. Η αρχική διατύπωση του PageRank αντικατοπτρίζει την ιδέα ότι μια σελίδα είναι σημαντική αν υπάρχουν πολλές σελίδες που συνδέονται με αυτήν και αν αυτές οι σελίδες είναι οι ίδιες σημαντικές. Η κατανομή των βαρών μπορεί έτσι να περιγράψει ως το σταθερό σημείο ενός σχεδίου διάβασης βάρους στο διάγραμμα ιστού. Αυτή η ιδέα επεκτάθηκε με παρόμοιο τρόπο με τα διμερή subgraphs του ιστού σε HITS και σε κατεύθυνση με κατεύθυνση προς τα κάτω. Χρησιμοποιούμε την



ίδια βασική αρχή για την κατάταξή μας στα folksonomies. Η βασική ιδέα είναι ότι ένας πόρος ο οποίος επισημαίνεται με σημαντικές ετικέτες από σημαντικούς χρήστες, καθίσταται ο ίδιος σημαντικός. Το ίδιο ισχύει, συμμετρικά, για τις ετικέτες και τους χρήστες. Έτσι έχουμε ένα γράφημα κορυφών που αλληλοενισχύονται μεταξύ τους με την εξάπλωση των βαρών τους.

Όπως το PageRank, χρησιμοποιείται το τυχαίο μοντέλο surfer, μια έννοια της σημασίας για ιστοσελίδες που βασίζεται στην ιδέα ότι ένας εξιδανικευμένος τυχαίος surfer ιστού ακολουθεί κανονικά υπερσύνδεσμους, αλλά από καιρό σε καιρό τυχαία μεταπηδά σε μια νέα ιστοσελίδα χωρίς να ακολουθήσει ένας σύνδεσμος για να αντισταθμίσει την απώλεια βάρους σε ιστοσελίδες χωρίς εξερχόμενες συνδέσεις. Αυτό οδηγεί στον ακόλουθο ορισμό της κατάταξης των κορυφών του γραφήματος καταχωρήσεις στο σταθερό σημείο  $w$  του υπολογισμού βάρους  $\vec{w} \leftarrow dA\vec{w} + (1 - d)\vec{p}$  (όπου  $w$  είναι ένας φορέας βάρους με μία είσοδο για κάθε σελίδα,  $A$  είναι η εκδοχή stochastic<sup>14</sup> σειράς της μήτρας γειτονίας του γραφήματος  $GF$  που ορίστηκε παραπάνω,  $p$  είναι η συνιστώσα τυχαίου surfer και  $d \in [0, 1]$  καθορίζει την επιρροή της  $p$ ). Στην αρχική PageRank, το  $p$  χρησιμοποιείται για να αντισταθμίσει την απώλεια βάρους σε ιστοσελίδες χωρίς εξερχόμενες συνδέσεις. Ωστόσο, για να υπολογίσει εξατομικευμένες σελίδες σελίδας, το  $p$  μπορεί να χρησιμοποιηθεί για να εκφράσει τις προτιμήσεις των χρηστών δίνοντας μεγαλύτερο βάρος στα συστατικά στοιχεία που αντιπροσωπεύουν τις προτιμώμενες ιστοσελίδες του χρήστη.

Χρησιμοποιούνται παρόμοια κίνητρα για την κατάταξη στο folksonomies. Το βασικό η ιδέα είναι ότι ένας πόρος ο οποίος επισημαίνεται με σημαντικές ετικέτες από σημαντικούς χρήστες καθίσταται ο ίδιος σημαντικός. Το ίδιο ισχύει και για τις ετικέτες και τους χρήστες, συμμετρικά, έτσι έχουμε ένα τριμερές γράφημα στο οποίο οι κορυφές αλληλοενισχύονται μεταξύ τους με την εξάπλωση των βαρών τους. Επίσημα, κατανέμεται το βάρος :

$$\vec{w} \leftarrow \alpha\vec{w} + \beta A\vec{w} + \gamma\vec{p}$$

όπου  $A$  είναι η σειρά στοχαστική εκδοχή της μήτρας γειτνίασης  $G_F$ ,  $p$  είναι μια προτίμηση  $\alpha, \beta, \gamma \in [0, 1]$  είναι σταθερές με  $\alpha + \beta + \gamma = 1$ . Η σταθερά  $\alpha$  προορίζεται να ρυθμίζει την ταχύτητα σύγκλισης, ενώ η αναλογία μεταξύ  $\beta$  και  $\gamma$  ελέγχει την επιρροή του διανύσματος προτιμήσεων.

## Συμπεράσματα

Τα αποτελέσματα έχουν το γενικό στόχο του προσδιορισμού ποιοτικών ιδιοτήτων των δικτύων που καθιστούν τη δρομολόγηση με τοπικές πληροφορίες δυνάμενες να μεταφερθούν και αποτελούν ένα μοντέλο για τη συλλογιστική σχετικά με τα αποτελεσματικά προγράμματα δρομολόγηση σε τέτοια δίκτυα. Το γενικότερο συμπέρασμα για τα δίκτυα του μικρόκοσμου στο οποίο καταλήγουμε είναι ότι η συσχέτιση μεταξύ της τοπικής διάρθρωσης και των συνδέσεων μεγάλης εμβέλειας παρέχει βασικές συμβουλές για την εύρεση διαδρομών μέσω του δικτύου. Όταν η συσχέτιση αυτή βρίσκεται κοντά σε ένα κρίσιμο κατώτατο όριο η δομή των συνδέσεων μεγάλης εμβέλειας αποτελεί έναν τύπο “κλίσης” που επιτρέπει στα άτομα να καθοδηγήσουν ένα μήνυμα αποτελεσματικά προς ένα στόχο. Καθώς η συσχέτιση πέφτει κάτω από αυτή τη κρίσιμη τιμή και το κοινωνικό δίκτυο γίνεται πιο ομοιογενές, τα στοιχεία αυτά αρχίζουν να εξαφανίζονται. Στο όριο, όταν οι συνδέσεις μεγάλων αποστάσεων παράγονται ομοιόμορφα τυχαία, το πρότυπο που αναφέραμε περιγράφει ένα κόσμο στον οποίο υπάρχουν μικρές αλυσίδες, αλλά άτομα που αντιμετωπίζουν μια αντιπροσανατολιστική σειρά κοινωνικών επαφών δεν μπορούν να τα εντοπίσουν. Επιπλέον, οι ενισχυμένες εγκαταστάσεις αναζήτησης είναι ζωτικής σημασίας για την αναδυόμενη σημασιολογία μέσα στα συστήματα που βασίζονται στη folksonomies. Παρουσιάσαμε ένα επίσημο μοντέλο για τις folksonomies, τον αλγόριθμο κατάταξης FolkRank που λαμβάνει υπόψη τη δομή των folksonomies και τα αποτελέσματα.

## Βιβλιογραφία

- [1] Watts, Duncan J., Peter Sheridan Dodds, and Mark EJ Newman. "Identity and search in social networks." *science* 296.5571 (2002): 1302-1305.
- [2] Menczer, Filippo. "Growing and navigating the small world web by local content." *Proceedings of the National Academy of Sciences* 99.22 (2002): 14014-14019.
- [3] Kleinberg, Jon M. "Small-world phenomena and the dynamics of information." *Advances in neural information processing systems*. 2002.
- [4] Kleinberg, Jon. *The small-world phenomenon: An algorithmic perspective*. Cornell University, 1999.
- [5] Nasution, Mahyuddin KM, and Shahrul Azman Noah. "Information retrieval model: A social network extraction perspective." *arXiv preprint arXiv:1207.3583*(2012).
- [6] Hotho, Andreas, et al. "Information retrieval in folksonomies: Search and ranking." *European Semantic Web conference*. Springer, Berlin, Heidelberg, 2006.
- [7] Langville, Amy N., and Carl D. Meyer. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.