



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ  
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ  
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

# Περιορισμός της παραπληροφόρησης στις πλατφόρμες κοινωνικής δικτύωσης

Διπλωματική Εργασία

Ηλιάνα Μαρία Ξύγκου  
Επιβλέπων καθηγητής: Συμεών Παπαβασιλείου

Αθήνα, Ιούνιος 2022

# Διάρθρωση παρουσίασης

- Εισαγωγικές έννοιες
- Βιβλιογραφική ανασκόπηση
- Πρόβλημα Cautious Misinformation Minimization (CMM)
- Δυσκολία προβλήματος
- Αλγόριθμοι επίλυσης
- Πειράματα
- Συμπεράσματα

# Εισαγωγικές έννοιες

- Ένα κοινωνικό δίκτυο αναπαρίσταται μαθηματικά μέσω ενός γράφου  $G = (V, E)$ , όπου  $V$  το σύνολο των κόμβων που αντιστοιχούν στους χρήστες και  $E$  το σύνολο των ακμών που αντιστοιχούν στις συνδέσεις μεταξύ των χρηστών.
- Ως Σύνθετα Δίκτυα νοούνται αυτά που συναντώνται πλέον κατά κόρον στην καθημερινότητα με τη μορφή δικτύων επικοινωνιών, κοινωνικών δικτύων, βιολογικών δικτύων, κ.α., και μοντελοποιούνται μαθηματικά με τεχνητά δίκτυα.
- Τα κοινωνικά δίκτυα μπορούν να προσομοιωθούν με μεγάλη ακρίβεια από τα Δίκτυα Ελεύθερης Κλίμακας (Scale-Free networks).

# Χαρακτηριστικά Δικτύων Ελεύθερης Κλίμακας

- Ο βαθμός κόμβου ακολουθεί την κατανομή νόμου-δύναμης (power-law distribution) – λίγοι κόμβοι συνδέονται με πολλούς κόμβους ενώ πολλοί κόμβοι με λίγους.
- Το μέσο μήκος μονοπατιού είναι μικρό – η απόσταση μεταξύ δύο κόμβων είναι μικρή.
- Ο συντελεστής ομαδοποίησης είναι σχετικά υψηλός – οι γείτονες των κόμβων δημιουργούν κλίκες μεταξύ τους.

# Κεντρικότητες κόμβων

Δείχνουν πόσο σημαντικός είναι ένας κόμβος στο δίκτυο.  
Οι σημαντικότερες εξ αυτών είναι:

- Κεντρικότητα βαθμού:  $C_D(u) = \frac{\deg(u)}{|V| - 1}$
- Κεντρικότητα εγγύτητας:  $C_P(u) = \frac{|V| - 1}{\sum_{v \in V, v \neq u} d(u, v)}$
- Κεντρικότητα ενδιαμεσικότητας:  $C_B(u) = \frac{2 \cdot \sum_{s \neq u \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}}}{(|V| - 1)(|V| - 2)}$
- Κεντρικότητα ιδιοδιανύσματος:  $v(u) = \frac{1}{\lambda} \sum_{k \in V} A[u, k] v(k)$

# Κεντρικότητα ενδιαμεσικότητας ακμής

Δείχνει πόσο σημαντική είναι μια ακμή στη διάδοση μιας πληροφορίας εντός ενός δικτύου, αν αυτή γίνεται μέσω των συντομότερων μονοπατιών μεταξύ των κόμβων.

$$C_B(e) = \frac{2 \cdot \sum_{s \neq t} \frac{\sigma_{st}(e)}{\sigma_{st}}}{|V|(|V| - 1)}$$

# Διάδοση της πληροφορίας

Κυριότερα μοντέλα διάδοσης της πληροφορίας:

- Independent Cascade (IC): Θεωρείται ένα σύνολο αρχικών ενστερνιστών  $A$  (seed set). Θεωρώντας διακριτά χρονικά βήματα  $t \in \mathbb{Z}$  σε κάθε χρονική στιγμή  $t$  κάθε κόμβος  $v$  που γίνεται για πρώτη φορά ενεργός επιχειρεί για μία και μοναδική φορά να ενεργοποιήσει κάθε εξερχόμενο γείτονά του  $w$  με πιθανότητα επιτυχίας  $p(v, w)$ . Η διαδικασία σταματάει τη χρονική στιγμή  $t'$ , κατά την οποία κανένας κόμβος δε μεταβαίνει σε ενεργή κατάσταση.
- Linear Threshold (LT): Θεωρείται ένα σύνολο αρχικών ενστερνιστών  $A$  (seed set) και μια τυχαία σύμφωνα με την ομοιόμορφη κατανομή στο διάστημα  $[0,1]$  επιλογή κατωφλίων  $\theta(u)$ :  $u \in V$ . Θεωρώντας διακριτά χρονικά βήματα  $t \in \mathbb{Z}$  σε κάθε χρονική στιγμή  $t$  κάθε κόμβος  $u$  ενεργοποιείται αν ισχύει  $\sum_{w \in N_{in}(u)} b(w, u) \geq \theta(u)$ . Η διαδικασία σταματάει τη χρονική στιγμή  $t'$ , κατά την οποία κανένας κόμβος δε μεταβαίνει σε ενεργή κατάσταση.

# Επιρροή ενός κόμβου

- Δείχνει πόσο ισχυρός είναι ένας κόμβος στο δίκτυο λαμβάνοντας υπόψη τη διάδοση μιας πληροφορίας.
- Ορίζεται ως το αναμενόμενο πλήθος κόμβων που θα βρίσκονται εν τέλει σε ενεργή κατάσταση δεδομένου ότι ο κόμβος  $u$  είναι ο αρχικός ενστερνιστής της ιδέας, και μαθηματικά συμβολίζεται ως  $\sigma(u, G)$ .
- Ο ακριβής υπολογισμός της επιρροής είναι #P-Δύσκολος υπό τα μοντέλα IC και LT.



# Βιβλιογραφική ανασκόπηση

Η αντιμετώπιση της παραπληροφόρησης χωρίζεται σε 2 στάδια:

- Εντοπισμός της ψευδούς είδησης
- Ανακοπή της διάδοσης της ψευδούς είδησης

Στον εντοπισμό της ψευδούς είδησης επιστρατεύονται εργαλεία της Μηχανικής Μάθησης.

Στον περιορισμό της εξάπλωσής της χρησιμοποιούνται αλγόριθμοι αφαίρεσης ακμών ή κόμβων, ή εκκίνησης διάδοσης της αντίστοιχης αληθούς είδησης.

# Μέθοδοι αφαίρεσης κόμβων

Στόχος είναι η εύρεση ενός συνόλου κόμβων περιορισμένης πληθικότητας η αφαίρεση των οποίων αποφέρει την ελαχιστοποίηση της διάδοσης της παραπληροφόρησης.

- Στατικές μέθοδοι: Η αφαίρεση των κόμβων γίνεται στην αρχή χωρίς να λαμβάνεται υπόψη η χρονική εξέλιξη της διάδοσης της πληροφορίας. Είναι υπολογιστικά αποδοτικές, αλλά στερούνται ακρίβειας.
- Δυναμικές μέθοδοι: Η αφαίρεση των κόμβων γίνεται σε χρονικά βήματα προσαρμοζόμενη στον τρόπο εξάπλωσης της πληροφορίας. Είναι πιο αποτελεσματικές από τις στατικές μεθόδους, αλλά υπολογιστικά κοστοβόρες.

# Μέθοδοι αφαίρεσης ακμών

Στόχος είναι η εύρεση ενός συνόλου ακμών περιορισμένης πληθικότητας η αφαίρεση των οποίων αποφέρει την ελαχιστοποίηση της διάδοσης της παραπληροφόρησης.

- Μέθοδοι με άγνωστο σύνολο αρχικών ενστερνιστών: Η αφαίρεση των ακμών αποσκοπεί στον περιορισμό της ροής της πληροφορίας στο δίκτυο. Δεν είναι αποτελεσματικές, αλλά έχουν γενική εφαρμογή.
- Μέθοδοι με γνωστό σύνολο αρχικών ενστερνιστών: Η αφαίρεση των ακμών αποσκοπεί στον περιορισμό της διάδοσης της πληροφορίας όπως αυτή αρχικά διαδίδεται από ένα γνωστό σύνολο κόμβων. Είναι αποτελεσματικότερες από τις παραπάνω μεθόδους, αλλά απαιτούν τον ακριβή και ταχύ εντοπισμό των αρχικών υποστηρικτών της είδησης.

# Μέθοδοι διαφώτισης

Στόχος αποτελεί η επιλογή ενός συνόλου κόμβων που θα εκκινήσουν τη διάδοση της αληθούς πληροφορίας ώστε να περιορισθεί το πλήθος των κόμβων που θα υιοθετήσουν την αντίστοιχη ψευδή πληροφορία.

- Μέθοδοι εκστρατείας: Οι κόμβοι επιλέγονται αποσκοπώντας στη μείωση της διάδοσης της παραπληροφόρησης λαμβάνοντας υπόψη είτε την τοπολογία του δικτύου είτε σε συνδυασμό με τη δομή του γράφου, την ατομική συμπεριφορά των χρηστών (προτιμήσεις, προσωπικό όφελος, τοποθεσία, κ.α.). Η δεύτερη περίπτωση είναι πιο αποτελεσματική, αλλά απαιτεί πληροφορίες σχετικά με τους χρήστες που μπορεί να μην είναι διαθέσιμες.
- Μέθοδοι προστασίας: Οι κόμβοι επιλέγονται αποσκοπώντας στην αποτροπή της υιοθέτησης της ψευδούς πληροφορίας από τουλάχιστον ένα συγκεκριμένο ποσοστό χρηστών.

# Πρόβλημα Cautious Misinformation Minimization (CMM)

- Έστω κοινωνικό δίκτυο αναπαριστώμενο από απλό κατευθυνόμενο γράφο  $G = (V, E, w)$ , με  $w: V^2 \rightarrow [0,1]$ , όπου  $w(u, v)$  η πιθανότητα διάδοσης μιας πληροφορίας από τον κόμβο  $u$  στον κόμβο  $v$ .
- Έστω  $I_T$  και  $I_F$  οι κλάσεις πληροφορίας αληθούς και ψευδούς περιεχομένου αντίστοιχα σχετικές με τη θεματική κατηγορία  $I$ . Θεωρείται ότι οι διαδιδόμενες πληροφορίες δεν είναι μεταξύ τους ανταγωνιστικές, αλλά ανεξάρτητες.
- Έστω  $e: V \rightarrow [0,1]$  η μετρική της εξειδίκευσης ενός χρήστη σχετικά με τη θεματική κατηγορία της διαδιδόμενης πληροφορίας, όπου οι τιμές 0 ή 1 δηλώνουν την πλήρη άγνοια ή την πλήρη αυθεντία αντίστοιχα.

# Μοντέλα διάδοσης

- Μοντέλο IC:

Λαμβάνεται υπόψη η εξειδίκευση  $e$  ενός χρήστη ως εξής:

Έστω  $X_u$  τυχαία μεταβλητή τέτοια ώστε:  $X_u = \begin{cases} 1 & \text{αν ο χρήστης } u \text{ υιοθετήσει την είδηση.} \\ 0 & \text{αλλιώς.} \end{cases}$

Τότε,  $p_T(u) = Pr\{X_u = 1 | i \in I_T\} = \max\{e(u), 1 - e(u)\}$ ,  $p_F(u) = Pr\{X_u = 1 | i \in I_F\} = 1 - e(u)$

και η πιθανότητα ενεργοποίησης ενός κόμβου  $u$  τη χρονική στιγμή  $t + 1$  από ένα για πρώτη φορά τη χρονική στιγμή  $t$  ενεργό εισερχόμενο γείτονα  $v \in N_{in}(u)$  υπολογίζεται ως  $p_C(v, u) = w(v, u) \cdot p_C(u)$ ,  $C \in \{T, F\}$ .

Έτσι, μπορούν προσομοιωθούν οι διαδόσεις ειδήσεων που ανήκουν στις κλάσεις  $I_T$  και  $I_F$  ορίζοντας δύο νέους γράφους  $G_T = (V, E, p_T)$  και  $G_F = (V, E, p_F)$  αντίστοιχα και εφαρμόζοντας τις αρχές του κλασικού μοντέλου IC.

- Μοντέλο LT:

Δε λαμβάνεται υπόψη η εξειδίκευση ενός χρήστη, με αποτέλεσμα η εξέλιξη της εξάπλωσης μιας πληροφορίας στο δίκτυο να είναι ως προς τη δυναμική ίδια ανεξάρτητα από το αν ανήκει στην κλάση  $I_T$  ή στην κλάση  $I_F$ .

- Μοντέλο Deterministic LT (DLT):

Διαφέρει από το παραπάνω πιθανοτικό μοντέλο LT λόγω της ντετερμινιστικής επιλογής των κατωφλίων  $\theta$  των κόμβων. Λαμβάνοντας υπόψη την εξειδίκευση  $e$  ενός χρήστη τα κατώφλια για την αποδοχή της πληροφορίας  $i \in I_T$  ή  $i \in I_F$  διαμορφώνονται αντίστοιχα ως:

$$\theta_T(u) = \min\{e(u), 1 - e(u)\}, \theta_F(u) = e(u)$$

Έτσι, μπορούν προσομοιωθούν οι διαδόσεις ειδήσεων που ανήκουν στις κλάσεις  $I_T$  και  $I_F$  ορίζοντας δύο νέους γράφους  $G_T = (V, E, w)$  με κατώφλι  $\theta_T(u)$ ,  $\forall u \in V$  και  $G_F = (V, E, w)$  με κατώφλι  $\theta_F(u)$ ,  $\forall u \in V$  αντίστοιχα, και εφαρμόζοντας τις αρχές διάδοσης του κλασικού μοντέλου LT.

# Ορισμός του προβλήματος CMM

- Υπό το μοντέλο LT:

Δοθέντων ενός δικτύου  $G = (V, E, w)$ , των δύο κλάσεων πληροφορίας  $I_T, I_F$  και των αντίστοιχων τρόπων διάδοσής τους υπό το μοντέλο LT, τα σύνολα των αρχικών ενστερνιστών της κλάσης  $I_T$  και της κλάσης  $I_F$ ,  $S_T$  και  $S_F$  αντίστοιχα, και ενός θετικού ακέραιου αριθμού  $k$ , στόχος είναι η εύρεση ενός συνόλου ακμών  $E' \subseteq E, |E'| \leq k$  των οποίων η αφαίρεση θα επιφέρει τη μέγιστη δυνατή μείωση της διάδοσης της κλάσης  $I_F$  με την ταυτόχρονη ελάχιστη δυνατή μείωση της διάδοσης της κλάσης  $I_T$ . Μαθηματικά εκφράζεται ως:

$$E' = \arg \min_{E', E' \subseteq E, |E'| \leq k} \left\{ \sum_{a \in S_T} \sigma(a, (V, E)) - \sum_{a \in S_T} \sigma(a, (V, E \setminus E')) + \sum_{b \in S_F} \sigma(b, (V, E \setminus E')) \right\}$$

, με αντικειμενική συνάρτηση  $f_{LT}(E') = \sum_{a \in S_T} \sigma(a, (V, E)) - \sum_{a \in S_T} \sigma(a, (V, E \setminus E')) + \sum_{b \in S_F} \sigma(b, (V, E \setminus E'))$

- Υπό τα μοντέλα IC και DLT:

Δοθέντων ενός δικτύου  $G = (V, E, w)$ , των δύο κλάσεων πληροφορίας  $I_T, I_F$  και των αντίστοιχων τρόπων διάδοσής τους υπό το μοντέλο IC ή το μοντέλο DLT, τα σύνολα των αρχικών ενστερνιστών της κλάσης  $I_T$  και της κλάσης  $I_F$ ,  $S_T$  και  $S_F$  αντίστοιχα, και ενός θετικού ακέραιου αριθμού  $k$ , στόχος είναι η εύρεση ενός συνόλου ακμών  $E' \subseteq E, |E'| \leq k$  των οποίων η αφαίρεση θα επιφέρει τη μέγιστη δυνατή μείωση της διάδοσης της κλάσης  $I_F$  με την ταυτόχρονη ελάχιστη δυνατή μείωση της διάδοσης της κλάσης  $I_T$ . Μαθηματικά εκφράζεται ως:

$$E' = \arg \min_{E', E' \subseteq E, |E'| \leq k} \left\{ \sigma(S_T, (V, E)) - \sigma(S_T, (V, E \setminus E')) + \sigma(S_F, (V, E \setminus E')) \right\}$$

, με αντικειμενική συνάρτηση  $f_{IC/DLT}(E') = \sigma(S_T, (V, E)) - \sigma(S_T, (V, E \setminus E')) + \sigma(S_F, (V, E \setminus E'))$



# Δυσκολία του προβλήματος CMM

Το πρόβλημα CMM είναι NP-Hard υπό και τα 3 μοντέλα:

- Υπό τα μοντέλα IC και DLT γίνεται μια σειρά αναγωγών από NP-Complete προβλήματα,  $Max\ Cut \leq_P Max\ Directed\ Cut \leq_P Max\ Bisection \leq_P Bisection\ Width \leq_P CMM$ .
- Υπό το μοντέλο LT ορίζεται στιγμιότυπο του προβλήματος CMM με  $S_T = \emptyset$ . Τότε, η αντικειμενική συνάρτηση απλοποιείται ως:

$$f_{LT}(E') = \sum_{b \in S_F} \sigma(b, (V, E \setminus E'))$$

και ορίζεται ισοδύναμη συνάρτηση, η οποία είναι μη φθίνουσα, submodular και μη αρνητική:

$$h(E') = \sum_{b \in S_F} \sigma(b, (V, E)) - \sum_{b \in S_F} \sigma(b, (V, E \setminus E'))$$

Το πρόβλημα CMM ζητά την μεγιστοποίηση αυτής της συνάρτησης υπό περιορισμό στην πληθικότητα της λύσης, πρόβλημα το οποίο είναι NP-Hard.

# Αλγόριθμοι επίλυσης

- Υπό το μοντέλο LT :

Άπληστος επαναληπτικός αλγόριθμος με το πολύ  $k$  επαναλήψεις.

Σε κάθε επανάληψη αφαιρείται η ακμή που ικανοποιεί το κριτήριο:

$$e = (u, v) = \arg \max_{e \in E \setminus E_t} \{\Delta(e|E_t)\} = \arg \max_{e \in E \setminus E_t} \{f_{LT}(E_t) - f_{LT}(E_t \cup \{e\})\}$$

Σημείο ενδιαφέροντος: Πώς υπολογίζεται αποδοτικά και προσεγγιστικά η τιμή της συνάρτησης  $f_{LT}$ ;

Με χρήση των live-edge γράφων, οι οποίοι κατασκευάζονται ως εξής: Ανεξάρτητα,  $\forall v \in V$  επιλέγεται το πολύ μια εισερχόμενη ακμή  $(u, v)$  με πιθανότητα  $w(u, v)$ , ενώ δεν επιλέγεται καμία με πιθανότητα  $1 - \sum_{u: (u, v) \in E} w(u, v)$ . Έτσι, προκύπτει ο γράφος  $X = (V, E_X)$ , όπου  $E_X \subseteq E$  το σύνολο των επιλεγμένων ζωντανών (live) ακμών. Εδώ, η διάδοση ξεκινώντας από ένα αρχικό κόμβο  $u$  είναι ντετερμινιστική και αποτελείται από όλα τα μονοπάτια που αρχίζουν από τον κόμβο  $u$  και διασχίζουν live ακμές. Έστω  $r(u, X)$  το πλήθος των κόμβων που συμμετέχουν σε αυτά τα μονοπάτια, τότε:

$$\sigma(u, G) = \mathbb{E}_X[r(u, X)] = \sum_{X \in \mathcal{X}_G} Pr[X|G] \cdot r(u, X)$$

Προσεγγιστικά χρησιμοποιείται ο τύπος:

$$\sigma(u, G) \approx \bar{\sigma}(u, G) = \frac{1}{|\mathcal{X}_S|} \cdot \sum_{X_i \in \mathcal{X}_S} r(u, T_{X_i}^u)$$

, όπου  $X_S = \{X_i: 1 \leq i \leq x_S\} \subseteq X$  το σύνολο των δειγματοληπτημένων live-edge γράφων του γράφου  $G$  και  $T_{X_i}^u$  το δέντρο με ρίζα τον κόμβο  $u$  που προκύπτει με εκτέλεση BFS στο γράφο  $X_i$ .

- Υπό το μοντέλο DLT :

Άπληστος επαναληπτικός αλγόριθμος με το πολύ  $k$  επαναλήψεις.

Σε κάθε επανάληψη αφαιρείται η ακμή που ικανοποιεί το κριτήριο:

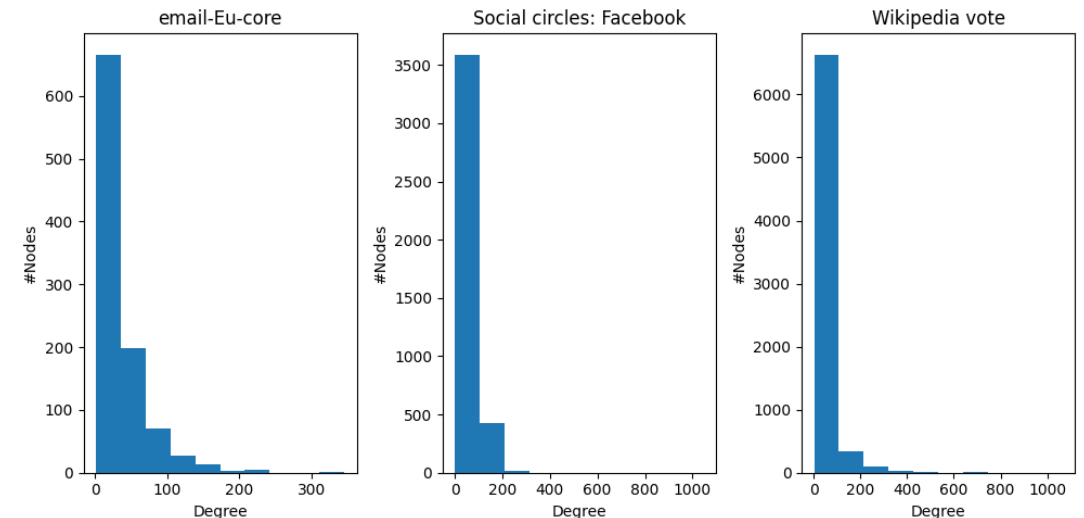
$$e = (u, v) = \arg \min_{e \in E \setminus E_t} \{f_{DLT}(E_t \cup \{e\})\}$$

Ο υπολογισμός της συνάρτησης  $f_{DLT}$  γίνεται ντετερμινιστικά σε γραμμικό χρόνο.

# Πειράματα

- 3 πραγματικά δίκτυα κοινωνικής δικτύωσης:
  - email-Eu-core: ανταλλαγή ηλεκτρονικής αλληλογραφίας εντός ενός πανεπιστημίου.
  - Social circles: Facebook: λίστες φίλων στο Facebook.
  - Wikipedia vote: ψηφοφορίες για την ανάδειξη διαχειριστών της πλατφόρμας Wikipedia.
- Μπορούν να χαρακτηριστούν ως scale-free δίκτυα:

Δίκτυο	Γράφος	Κόμβοι	Ακμές	Μέσο μήκος μονοπατιού	Μέσος συντελεστής ομαδοποίησης
email-Eu-core	Κατευθυνόμενος	986	16064	2.5869	0.4071
Social circles: Facebook	Μη κατευθυνόμενος	4039	88234	3.6925	0.6055
Wikipedia vote	Κατευθυνόμενος	7115	100762	3.247	0.1409



- Συγκριτικές μέθοδοι:

- Random: Αφαίρεση  $k$  τυχαίων ακμών, οι οποίες έχουν ως αρχή ένα κόμβο που ανήκει στο σύνολο των αρχικών ενστερνιστών της ψευδούς πληροφορίας.
- Weighted: Αφαίρεση των  $k$  ακμών, οι οποίες έχουν ως αρχή ένα κόμβο που ανήκει στο σύνολο των αρχικών ενστερνιστών της ψευδούς πληροφορίας, με την υψηλότερη πιθανότητα διάδοσης της πληροφορίας,  $w(u, v)$ .
- DistanceDiff (μόνο υπό το μοντέλο LT): Αφαίρεση των  $k$  ακμών με τη μικρότερη διαφορά της απόστασης από το σύνολο των αρχικών ενστερνιστών μόνο της αληθούς πληροφορίας από την απόσταση από το σύνολο των αρχικών ενστερνιστών μόνο της ψευδούς πληροφορίας.

$$diff(u, v) = \min_{sf \in S_F \setminus S_T} \{d(sf, u)\} - \min_{st \in S_T \setminus S_F} \{d(st, u)\}$$

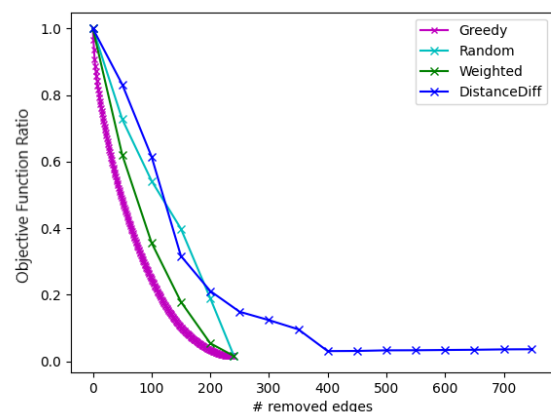
- EdgeBetweennessDiff (μόνο υπό το μοντέλο DLT): Αφαίρεση των  $k$  ακμών με τη μεγαλύτερη διαφορά της κεντρικότητας ενδιαμεσικότητας στον επαγόμενο γράφο διάδοσης της αληθούς πληροφορίας από την κεντρικότητα ενδιαμεσικότητας στον επαγόμενο γράφο διάδοσης της ψευδούς πληροφορίας.

$$diff(e) = C_{B,false}(e) - C_{B,true}(e)$$

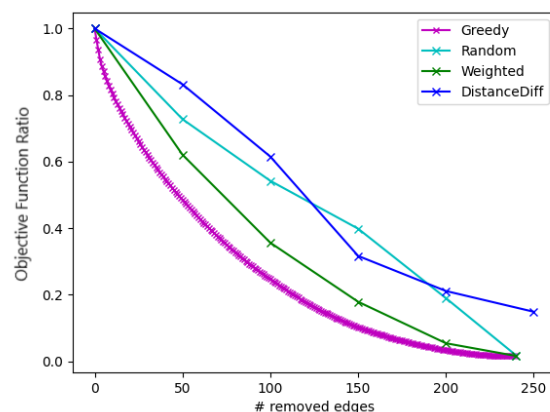
- Τα σύνολα  $S_T, S_F$  επιλέγονται τυχαία σύμφωνα με την ομοιόμορφη κατανομή με πληθικότητα  $|S_T| = |S_F| = \lceil 1\% \cdot |V| \rceil$ .
- Το  $k$  ορίζεται ως  $k = \lceil 3\% \cdot |E| \rceil$ .
- Υπό το μοντέλο LT, επιλέγεται  $x_S = 5000$ .
- Η απόδοση των αλγορίθμων ποσοτικοποιείται μετά την αφαίρεση  $k$  ακμών ως:

$$ratio(k) = \frac{f_{LT}(E')}{f_{LT}(\emptyset)}, \text{ με } |E'| = k \quad \text{ή} \quad ratio(k) = \frac{f_{DLT}(E')}{f_{DLT}(\emptyset)}, \text{ με } |E'| = k$$

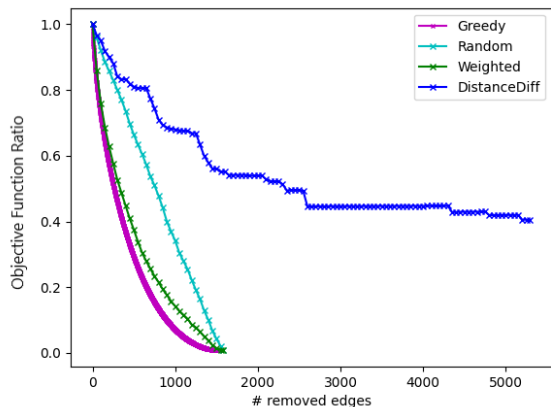
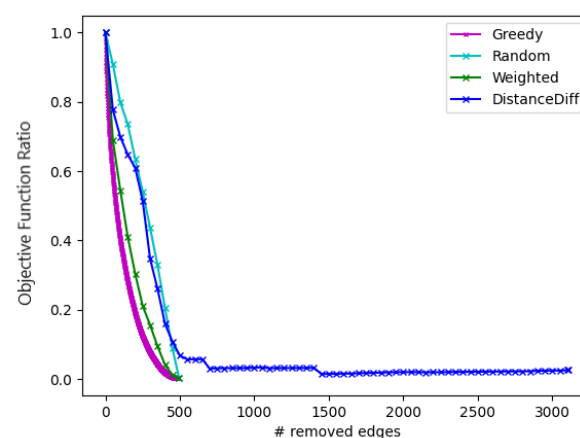
# Αποτελέσματα υπό το μοντέλο LT



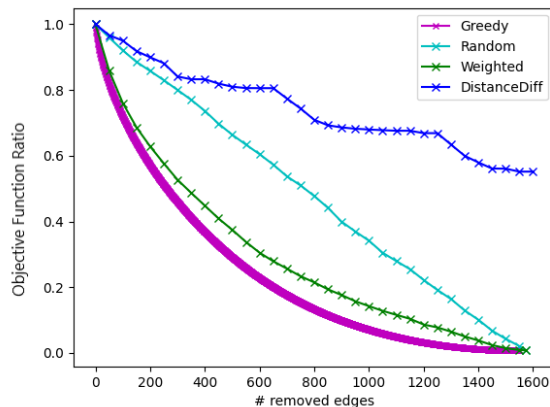
email-Eu-core



Wikipedia vote



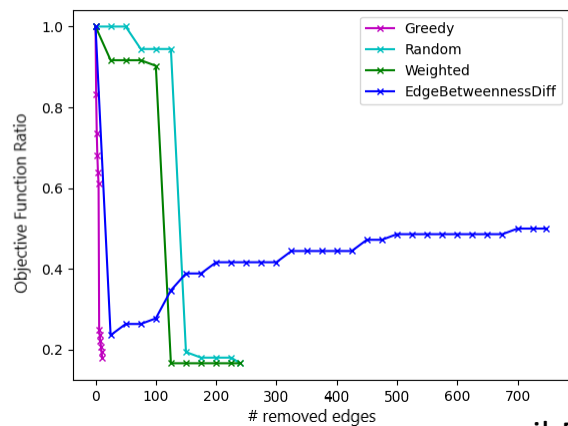
Social circles: Facebook



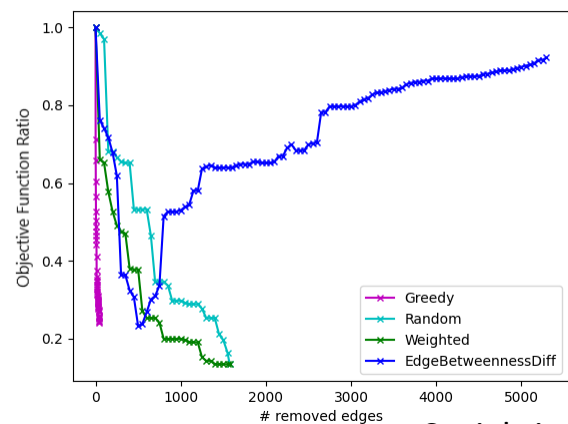
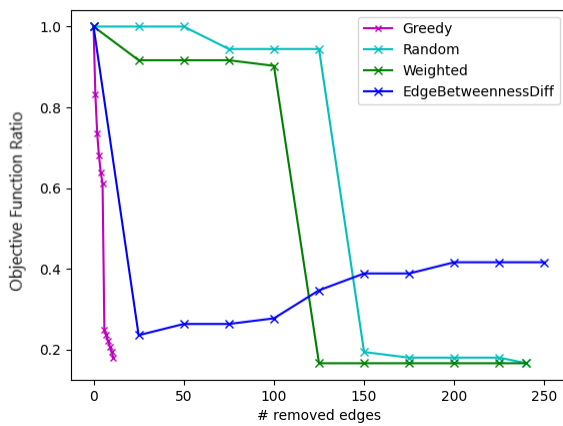
Η γραφική παράσταση στα αριστερά απεικονίζει τα πλήρη αποτελέσματα, ενώ αυτή στα δεξιά εστιάζει στο διάστημα του οριζόντιου άξονα μέχρι την τιμή του πλήθους των ακμών που αφαιρέθηκαν κατά την εφαρμογή των μεθόδων εκτός της “DistanceDiff”, με στόχο τη λεπτομερέστερη παρατήρηση της επίδρασής τους στην αντιμετώπιση του μελετούμενου προβλήματος.



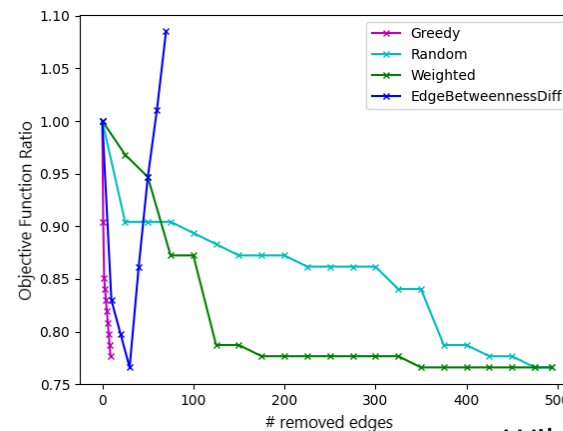
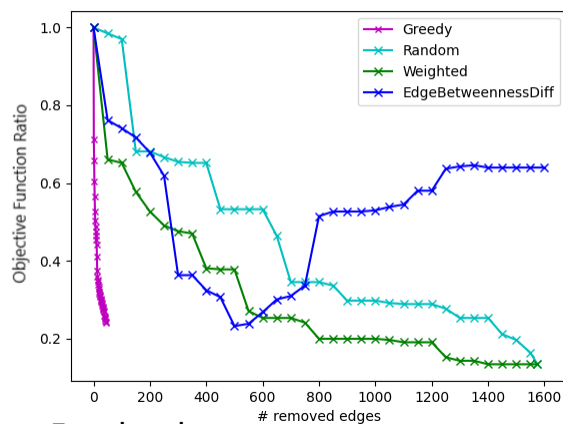
# Αποτελέσματα υπό το μοντέλο DLT



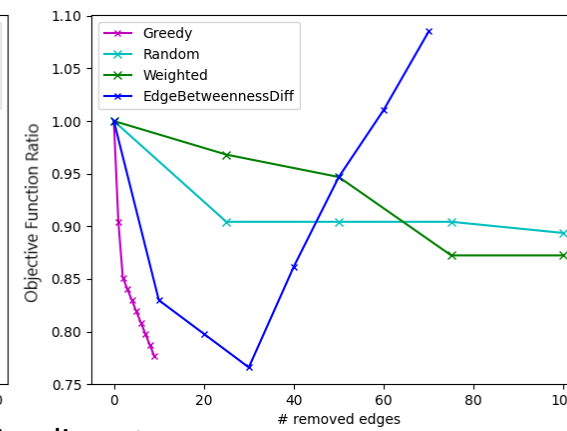
email-Eu-core



Social circles: Facebook



Wikipedia vote



# Συμπεράσματα

- Οι προτεινόμενες άπληστες μέθοδοι υπερέχουν όλων των υπολοίπων, αφού μειώνεται σημαντικά η αντικειμενική συνάρτηση με την αφαίρεση λιγότερων ακμών.
- Οι μέθοδοι DistanceDiff και EdgeBetweennessDiff που αξιοποιούν κυρίως την τοπολογία του δικτύου δεν είναι ιδιαίτερα αποτελεσματικές.
- Μελλοντικά φαίνεται ενδιαφέρον να:
  - Μελετηθεί πειραματικά και το μοντέλο IC (στα πλαίσια της εργασίας ο υπολογιστικός φόρτος ήταν απαγορευτικός).
  - Γίνει εκμάθηση των παραμέτρων των μοντέλων διάδοσης αντί να ανατίθενται τυχαίες τιμές.
  - Να μελετηθούν και δυναμικά μεταβαλλόμενα δίκτυα πέραν των στατικών.
  - Να βρεθεί κατάλληλο κριτήριο παύσης της μεθόδου EdgeBetweennessDiff υπό το μοντέλο DLT, η οποία είναι αποδοτική όταν αφαιρούνται μόνο λίγες ακμές.

Σας ευχαριστώ για την προσοχή σας!