ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΠΜΣ «Πληροφορική και Τηλεματική»

ΑΝΑΛΥΣΗ ΚΟΙΝΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

ΕΜΜΑΝΟΥΗΛ ΜΟΡΦΙΑΔΑΚΗΣ itp21110
ΙΩΑΝΝΗΣ ΤΖΙΟΓΚΑΣ itp20135

Εισαγωγή

Το αντικείμενο της εργασίας αποτελεί η μελέτη του κοινωνικού δικτύου του Facebook. Το Dataset που επιλέξαμε είναι το https://snap.stanford.edu/data/ego-Facebook.html. Θα το μελετήσουμε περιγραφικά και θα εμβαθύνουμε στον γράφο βγάζοντας συμπεράσματα.Στα κοινωνικά δίκτυα, ο κόμβος αναπαριστά έναν χρήστη ενώ οι ακμές τις συνδέσεις του με άλλους χρήστες, οι οποίοι αποκαλούνται “φίλοι”. Για να μελετήσουμε καλύτερα τον γράφο μας, συνθέσαμε και έναν τυχαίο, με ίδιο αριθμό κόμβων και παραπλήσιο αριθμό ακμών για να παρατηρήσουμε και να συγκρίνουμε τα αποτελέσματα.

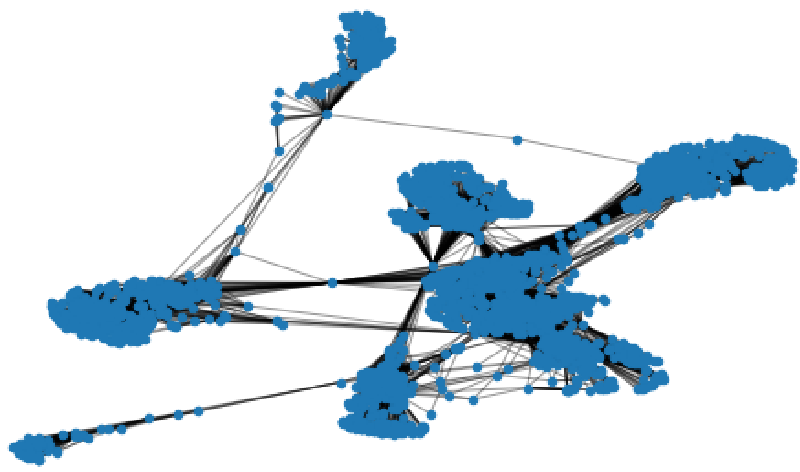
Για τα πειράματα και τα αποτελέσματα που θα παρουσιάσουμε χρησιμοποιήθηκε η Python και κυρίως η βιβλιοθήκη NetworkX αλλά και άλλες όπως η Pandas και η Scipy.Ο γράφος G είναι ο γράφος του Facebook και ο V είναι ο τυχαίος μας γράφος.

Περιγραφή των γράφων

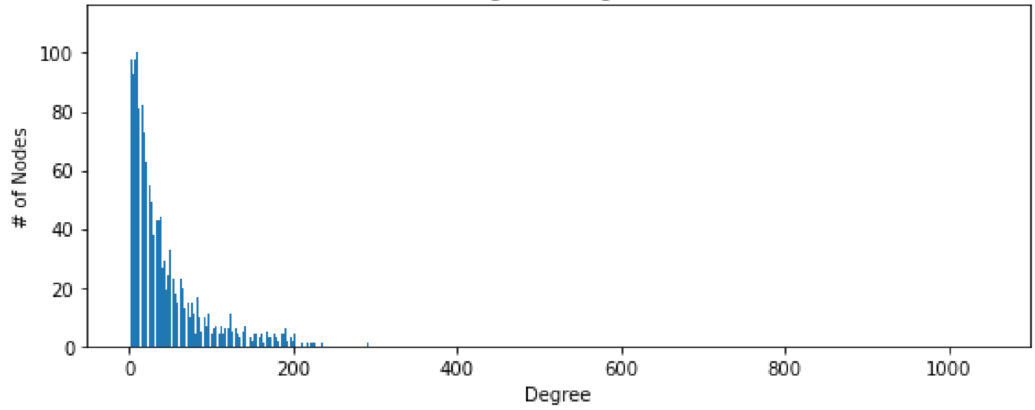
Γράφος G

Ο γράφος G αποτελείται από 4039 κόμβους και 88234 ακμές. Παρακάτω βλέπουμε τον γράφο και την κατανομή των βαθμών των κόμβων.

Connected components of G



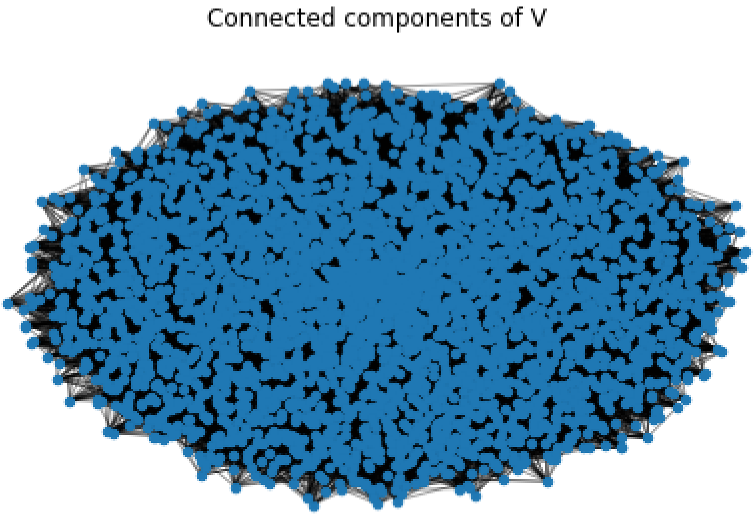
Degree histogram

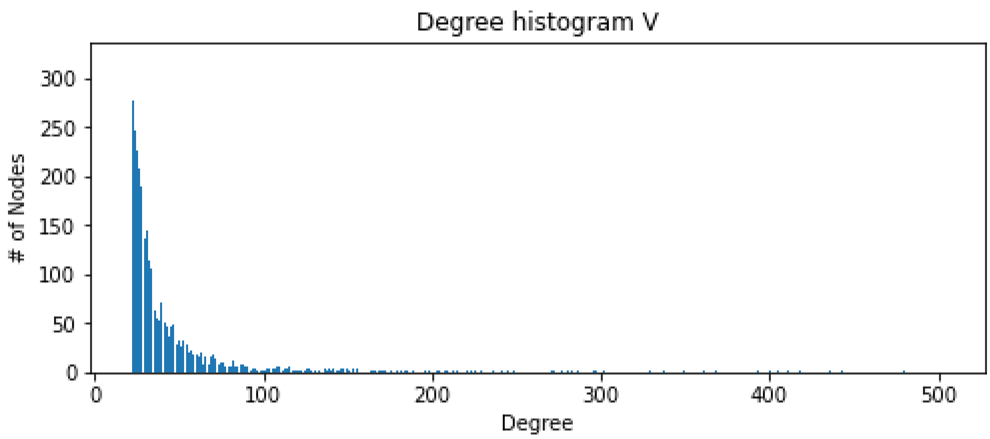


Με μια πρώτη ματιά, θα έλεγε κανείς ότι παρατηρούνται στον γράφο να δημιουργούνται κάποιες κοινότητες. Συγκεκριμένα οκτώ, αλλά αυτό προς το παρόν αποτελεί απλά μια εκτίμηση. Θα το μελετήσουμε διεξοδικά στα παρακάτω κεφάλαια. Επίσης στην κατανομή των ακμών των κόμβων θα έλεγε κανείς ότι παρατηρείται να ακολουθεί την Power Law κατανομή.
Διάμετρος του γράφου

Γράφος V

Ο γράφος G αποτελείται από 4039 κόμβους και ακμές. Παρακάτω βλέπουμε τον γράφο και την κατανομή των βαθμών των κόμβων. Είναι ένα δίκτυο ελεύθερης κλίμακας Barabashi Albert το οποίο ακολουθεί την κατανομή powerlaw.



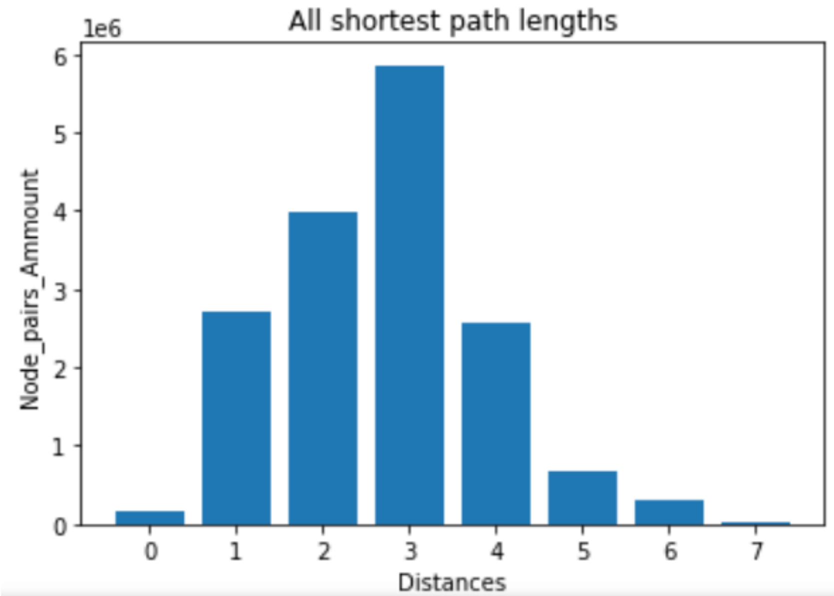


Με μια πρώτη ματιά, δεν φαίνεται να δημιουργούνται κοινότητες πέρα από μία που συνιστά όλο τον γράφο. Θα το μελετήσουμε διεξοδικά στα παρακάτω κεφάλαια. Επίσης στην κατανομή των ακμών των κόμβων θα έλεγε κανείς ότι παρατηρείται να ακολουθεί την power law κατανομή, μένει να το επιβεβαιώσουμε με κάποια good fit tests.

Μετρικές

α. Μήκος μικρότερων μονοπατιών

Θα μετρήσουμε το μήκος του μικρότερου μονοπατιού που ενώνει κάθε κόμβο u με κάθε κόμβο v σε κάθε γράφο. Παρακάτω βλέπουμε την συνολική εικόνα για τον G:



Παρατηρούμε ότι η πιο συνήθης απόσταση μεταξύ δύο κόμβων είναι το 3. Οι συνολικοί συνδυασμοί ελαχίστων μονοπατιών μεταξύ των κόμβων είναι NxN. Στην περίπτωση μας είναι 4039x4039=16.313.521 συνδυασμοί. Παρακάτω βλέπουμε και τον αναλυτικό πίνακα στον οποίο δεν είναι τα μήκη μονοπατιού

Μήκος μικρότερου μονοπατιού	Αριθμός συνδυασμών κόμβων u,v
0	4.039
1	176.468
2	2.716.134
3	3.981.852
4	5.861.560
5	2.565.170
6	677.214
7	315.464
8	15.620
Σύνολο	16.313.521

Παρατηρούμε πάνω στον γράφο ότι υπάρχουν shortest paths με μήκος 0 και μάλιστα 4039. Αυτό είναι απολύτως φυσιολογικό διότι έχουμε 4039 κόμβους και είναι το shortest path από τον εαυτό τους.

Παρακάτω βλέπουμε την συνολική εικόνα για τον V:

Μήκος μικρότερου μονοπατιού	Αριθμός συνδυασμών κόμβων u,v
0	4.039
1	176.748
2	7.532.836
3	8.599.894
4	4
Σύνολο	16.313.521

Στον τυχαίο γράφο παρατηρούμε εντελώς διαφορετική συμπεριφορά με τον γράφο που μελετάμε. Επίσης αυτό που κάνει εντύπωση είναι ότι οι συνδυασμοί κόμβων που έχουν απόσταση 1 είναι σχεδόν ίδιοι. Στον τυχαίο γράφο οι αποστάσεις είναι πολύ μικρότερες και η συνήθης απόσταση μεταξύ δύο κόμβων είναι 2 ή 3.

β. Eccentricity

Εκκεντρικότητα ενός κόμβου v είναι η μεγαλύτερη απόσταση(shortest path) από κάθε άλλο κόμβο u του γράφου. Με απλά λόγια η μεγαλύτερη απόσταση των shortest path του κόμβου v.

Παρακάτω βλέπουμε τα eccentricity των κόμβων για κάθε γράφο

Eccentricity κόμβων (G)	Αριθμός συνδυασμών κόμβων u,v
4	1
5	112
6	2.579
7	1.150
8	197
Σύνολο	4.039

Eccentricity κόμβων (V)	Αριθμός συνδυασμών κόμβων u,v
3	4.035
4	4
Σύνολο	4.039

Άλλο ένα δείγμα το πόσο ομοιόμορφα είναι κατανεμημένα οι κόμβοι του τυχαίου γράφου ανάλογα και με την δημοφιλία τους καθώς και τις μικρές αποστάσεις που έχουν οι κόμβοι μεταξύ τους, σε αντίθεση με τον γράφο G.

γ. Closeness centrality

Είναι το κλάσμα του αριθμού των κόμβων προς το άθροισμα όλων των μικρότερων μονοπατιών του v με όλους τους υπόλοιπους κόμβους u του γράφου.

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$$

Στον γράφο G το μεγαλύτερο Closeness centrality το έχει ο κόμβος 107 με C(107)=0,46. Ακολουθούν οι κόμβοι 58,428,563,1684,171,348,483,414 και 376 με τον τελευταίο C(376)=0,37.

Στον γράφο V το μεγαλύτερο Closeness centrality το έχει ο κόμβος 27 με C(27)=0,53. Ακολουθούν οι κόμβοι 24,0,31,26,23,25,29,34 και 28 με τον τελευταίο C(28)=0,52.

Μας επιβεβαιώνει λοιπόν το closeness centrality την ομοιομορφία μεταξύ των ισχυρών κόμβων εντός του γράφου, στον τυχαίο γράφο η διαφορά του μεγαλύτερου Closeness centrality είναι κατα 0,01 διαφοροποιημένη από τον 10ο μεγαλύτερο centrality ενώ στον γράφο G η διαφορά τους είναι εννιαπλασια, δηλαδή 0,09

δ.Betweenness centrality

Υπολογίζουμε ποσοστό των shortest paths περνάνε από έναν κόμβο

Δηλαδή κατα κάποιων τρόπο βλέπουμε την συμβολή των κόμβων στη μετάδοση

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

όπου V είναι ένα σύνολο από ακμές, σ(s,t) είναι ο αριθμός όλων των shortest paths και σ(s,t|v) είναι ο αριθμός των shortest paths που περνάνε από έναν κόμβο v.

Στον γράφο G το μεγαλύτερο betweenness centrality το έχει ο κόμβος 107 με Cb(107)=48%. Ακολουθούν οι κόμβοι 1684,3437,1912,1085,0,698,567,58 και 428 με τον τελευταίο C(428)=6%. Βλέπουμε δύο από τους δέκα με το μεγαλύτερο Closeness Centrality να έχουν και από τα μεγαλύτερα ποσοστά betweenness centrality.

Στον γράφο V το μεγαλύτερο betweenness centrality το έχει ο κόμβος 27 με Cb(27)=2,5%. Ακολουθούν οι κόμβοι 0,24,31,26,23,25,29,34 και 28 με τον τελευταίο C(428)=1,4%. Βλέπουμε δύο από τους δέκα με το μεγαλύτερο Closeness Centrality να έχουν και από τα μεγαλύτερα ποσοστά betweenness centrality. Εδώ παρατηρούμε πως οι σχεδόν οι ίδιοι κόμβοι που έχουν το μεγαλύτερο Closeness centrality έχουν και το ίδιο betweenness centrality.

ε.Eigenvector centrality
Υπολογίζει την κεντρικότητα ενός κόμβου βασισμένο στην κεντρικότητα των γειτόνων του κόμβου.

$$Ax = \lambda x$$

Όπου A είναι ο πίνακας γειτνίασης με ιδιοτιμή λ.

Ουσιαστικά γίνεται προσδιορισμός της αξίας κάθε κόμβου ο οποίος είναι ανάλογος της αξίας των γειτόνων του κόμβου

ζ.Katz centrality
Και ο katz υπολογίζει την κεντρικότητα ενός κόμβου όπως ο Eigenvector.

$$x_i = \alpha \sum_j A_{ij} x_j + \beta,$$

η.Edge betweenness centrality
Μοιάζει με το betweenness centrality με την μόνη διαφορά ότι αντί για την κεντρικότητα ενός κόμβου υπολογίζεται η κεντρικότητα μιας ακμής σύμφωνα με τα shortest paths.

Δηλαδή κατα κάποιων τρόπο βλέπουμε την συμβολή των ακμών στη μετάδοση

$$c_B(e) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)}$$

θ.Clustering coefficient
Ο συντελεστής συσταδοποίησης για τους κόμβους, για γράφους χωρίς βάρη, η συσταδοποίηση ενός κόμβου ν είναι το κλάσμα όλων των πιθανών τριγώνων μέσω αυτού του κόμβου.

$$c_u = \frac{2T(u)}{deg(u)(deg(u) - 1)}$$

Δηλαδή μπορούμε να δούμε κατα κάποιο τρόπο πόσες παρέες έχει κάποιος.

T(u) είναι ο αριθμός των τριγώνων μέσω του κόμβου u και deg(u) είναι ο βαθμός του κόμβου u.

ι.Triangles
Υπολογίζει τον αριθμό των τριγώνων που σχηματίζονται έχοντας ακμή στον κόμβο u.

Στον γράφο G το μεγαλύτερο πλήθος τριγώνων το έχει ο κόμβος 1912. Ακολουθούν οι κόμβοι 107,2347,2266,2206,2543,2233,2464,2218 και 2142 με το μεγαλύτερο αριθμό να είναι ο 30025.

Στον γράφο V το μεγαλύτερο πλήθος τριγώνων το έχει ο κόμβος 27. Ακολουθούν οι κόμβοι 31,24,0,23,25,26,29,30 και 28 με το μεγαλύτερο αριθμό να είναι ο 4529.

Και εδώ παρατηρούμε ότι οι ισχυρότεροι κόμβοι δημιουργούν το μεγαλύτερο πλήθος τριγώνων στον τυχαίο γράφο.

Link Prediction

Για το Link Prediction χρησιμοποιήσαμε έξι συναρτήσεις συνδυαστικά μεταξύ τους:

- Κοινοί γείτονες
- Jaccard index
- Resource allocation index
- Adamic/Adar index
- Preferential Attachment index
- Clustering Coefficient

Σαν παραμέτρους στα μοντέλα που δημιουργήσαμε χρησιμοποιήσαμε τις παραπάνω μεθόδους για να βρούμε την ομοιότητα των κόμβων. Κανονικοποιήσαμε όλες τις τιμές των παραμέτρων για να έχουν το ίδιο διάστημα. Σαν μετρική για το πόσο αξιόπιστο είναι τα μοντέλα μας χρησιμοποιήσαμε το accuracy.

$$\text{Accuracy} \qquad \frac{TP + TN}{TP + FP + TN + FN}$$

Κάθε μοντέλο το εκπαιδεύσαμε πέντε φορές κάνοντας διαφορετικό ανακάτεμα στα δεδομένα μας και από τις πέντε εκπαιδεύσεις που πραγματοποιήσαμε βγάλαμε σαν score ένα average accuracy.

α.Κοινοί γείτονες

Η λογική είναι απλή, εφόσον δύο άτομα έχουν πολλούς κοινούς φίλους τότε είναι πολύ πιθανό να γίνουν φίλοι και μεταξύ τους.

$$s(x, y) = |\Gamma_x \cap \Gamma_y|$$

Όπου Γχ οι γείτονες το κόμβου χ και Γy οι γείτονες του κόμβου y.

β.Jaccard index

Είναι η αναλογία των κοινών γειτόνων των δύο κόμβων προς το σύνολο των γειτόνων των δύο κόμβων.

$$s(x,y) = \frac{|\Gamma_x \cap \Gamma_y|}{|\Gamma_x \cup \Gamma_y|}$$

Ουσιαστικά πόσο δημοφιλή είναι ο x στις επαφές του y και ο y στις επαφές του x

γ.Resource allocation index
Η λογική του resource allocation index είναι ότι παίρνει σαν παράμετρο και τον αριθμό των γειτόνων των κοινών γειτόνων. Για παράδειγμα αν έχουν κοινό φίλο δύο κόμβοι κάποιο διάσημο που έχει πολλές συνδέσεις, αυτό δεν σημαίνει ότι θα γίνουν απαραίτητα φίλοι. Επηρεάζει αλλά σε μικρότερο βαθμό(κοινό χαρακτηριστικό ότι ακούνε την ίδια μουσική, αλλά αυτό δεν είναι το μοναδικό κριτήριο).

$$s(x,y) = \sum_{z \in \Gamma_x \cap \Gamma_y} \frac{1}{|\Gamma_z|}$$

δ.Adamic/Adar index
Παρόμοια λογική(επιβάλλει κάποιου είδους ποινή αν οι κοινοί φίλοι έχουν πολλούς γείτονες) με του resource allocation αλλά λογαριθμίζουμε τους γείτονες των κοινών φίλων

ε.Preferential Attachment index
Σε scale free δίκτυα όσοι έχουν τους πολλούς φίλους τήνουν να αποκτούν όλο και περισσότερους (λειτουργεί άπληστα). Με αυτόν τον τρόπο δύο διάσημοι έχουν μεγαλύτερη πιθανότητα να αποκτήσουν όλο και περισσότερους φίλους.

$$s(x,y) = |\Gamma_x||\Gamma_y|$$

ζ.Clustering Coefficient_pair(u,v)
Clustering_Coefficient(u)*clustering_coefficient(v)

Αποτελέσματα γράφου G

- Κοινοί γείτονες (CN)
- Jaccard index (JI)
- Resource allocation index (RAI)
- Adamic/Adar index (AAI)
- Preferential Attachment index (PAI)

- Clustering Coefficient (CC)

Μέθοδοι	Average Accuracy
Όλες	97,15 %
JI,AAI,CC	97,17%
JI,AAI,CC,PAI	97,15%
CN,AAI,CC	97,20%
CN,JI,AAI,CC	97,26%
CN,JI	97,18%
CN,JI,PAI	97,10%

Παρατηρήσαμε ότι τα καλύτερα αποτελέσματα τα πήραμε με το μοντέλο με τις τέσσερεις μεταβλητές που είχε τους κοινούς γείτονες, jaccard index,Adamic/Adar index,CC. Μας έδωσαν καλύτερα αποτελέσματα ακόμη και όταν χρησιμοποιήσαμε όλες τις συναρτήσεις σαν παραμέτρους στο μοντέλο μας. Πολλές φορές όταν χρησιμοποιούμε πανομοιότυπες παραμέτρους δε μας δίνουν την προστιθέμενη αξία που επιθυμούμε και πολλές φορές μας επιστρέφουν και χειρότερα αποτελέσματα, κατα κάποιο τρόπο μπερδεύουν τον αλγόριθμο που χρησιμοποιούμε για το training. Για επέκταση της εργασίας καλύτερα θα ήταν να συγκρίναμε τις μεταβλητές ανά δύο μεταξύ τους τι συσχέτιση έχουν και αν κάποιες είχαν μεγάλη συσχέτιση να διαλέγαμε πιθανόν αυτή που έχει το μικρότερο p-value.

Αποτελέσματα γράφου V

- Κοινοί γείτονες (CN)
- Jaccard index (JI)
- Resource allocation index (RAI)
- Adamic/Adar index (AAI)
- Preferential Attachment index (PAI)
- Clustering Coefficient (CC)

Μέθοδοι	Average Accuracy
Όλες	58,65%
JI,AAI,CC	50,09%
JI,AAI,CC,PAI	53,38%
CN,AAI,CC	50,72%

CN,JI,AAI,CC	47,92%
CN,JI	46,48%
CN,JI,PAI	51,76%

Στον τυχαίο γράφο τα καλύτερα αποτελέσματα τα είχε το μοντέλο που χρησιμοποιήσαμε όλες τις παραμέτρους. Τα χαμηλά ποσοστά accuracy που είχε ο γράφος V οφείλονταν ότι οι ακμές είχαν δημιουργηθεί με τυχαίο αλλά ομοιόμορφο τρόπο και πολλοί κόμβοι παρουσίαζαν όμοια χαρακτηριστικά. Συμπεραίνουμε ότι έναν γράφο που προέβλεψε το μοντέλο μας ως link και στην πραγματικότητα δεν ήταν, μελλοντικά όντως θα υπάρξει σύνδεση.

Community detection

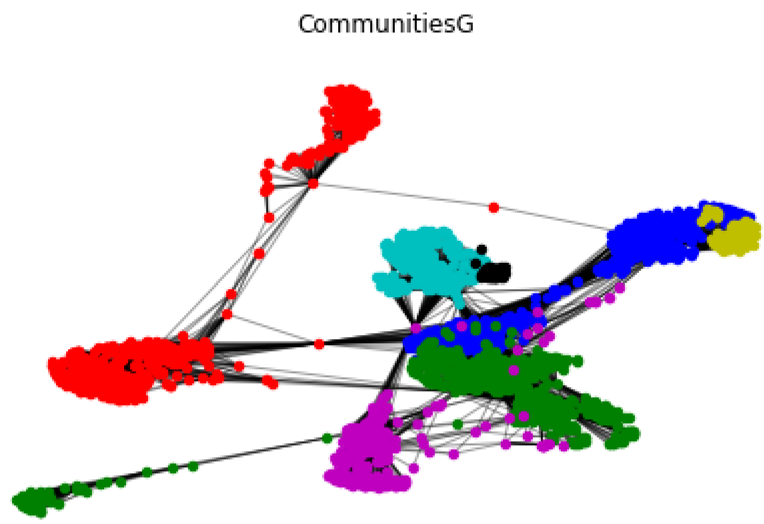
Greedy Modularity Communities

Γινεται ταξινόμηση του edge betweeness και αφαιρούνται οι πρώτες k ακμές.

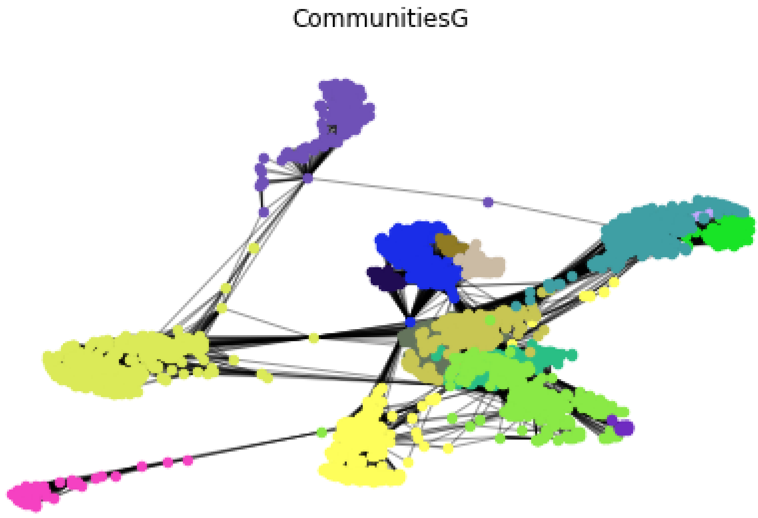
Louvain Method

Αρχικά κάθε κόμβος αποτελεί μια κοινότητα μονος του.Μετά κάθε κόμβο τον εισάγω στην κοινότητα με το μεγαλύτερο modularity μέχρις ότου να μην υπάρχει καλύτερη ανάθεση για κανέναν κόμβο.

Σε 1η φάση για τον πραγματικό γράφο του facebook δοκιμάσαμε να βρούμε τα communities με τον greedy modularity algorithm
Επειδή οπτικα φαίνεται πως υπάρχουν 7 communities ορίσαμε στο k την τιμή 7

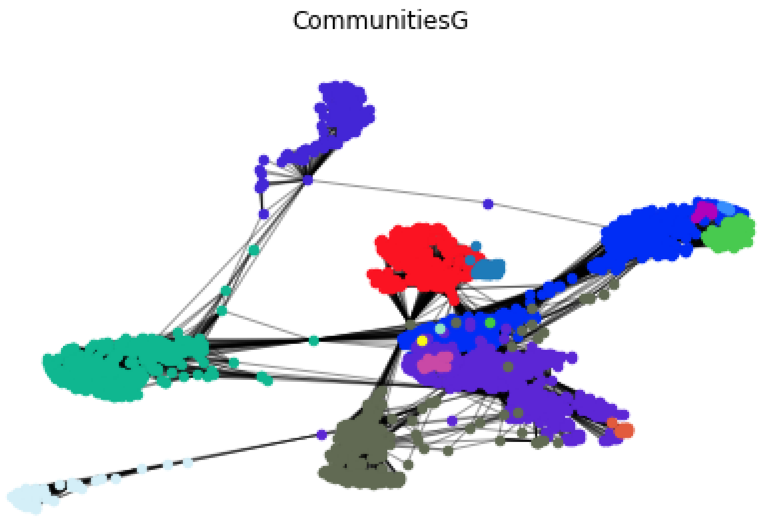


Οπως φαίνεται σε έναν βαθμό φαίνονται τα communities και μάλιστα πολύ καλά
Επειτα δοκιμάσαμε και με luvain.



Λαμβάνοντας υπόψη τα 2 προηγούμενα community_outputs Η louvain πετυχαίνει πολύ καλύτερο community detection στον συγκεκριμένο γράφο

Στη συνέχεια ξανά εκτελέσαμε τον greedy αλγόριθμο με τον αριθμο των communities που βρήκαμε παραπάνω(luvain)

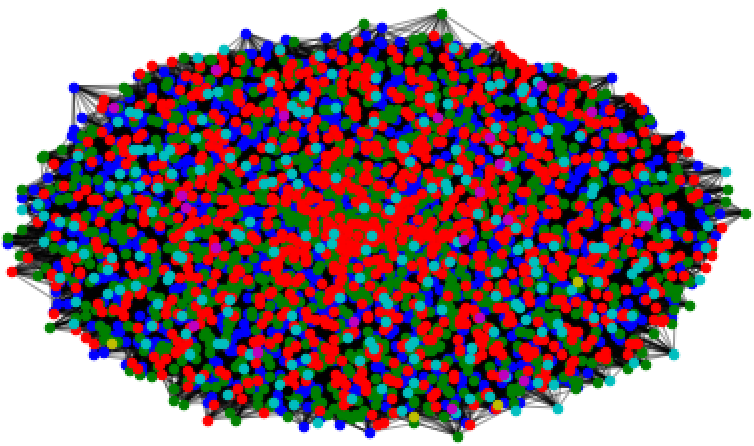


Συγκρίνοντας τα 3 παραπάνω καλύτερο community detection achieved by louvain.

Αυτό το υποψιαζόμαστε επειδή η luvain ομαδοποιεί τους κόμβους με βάση το modularity.

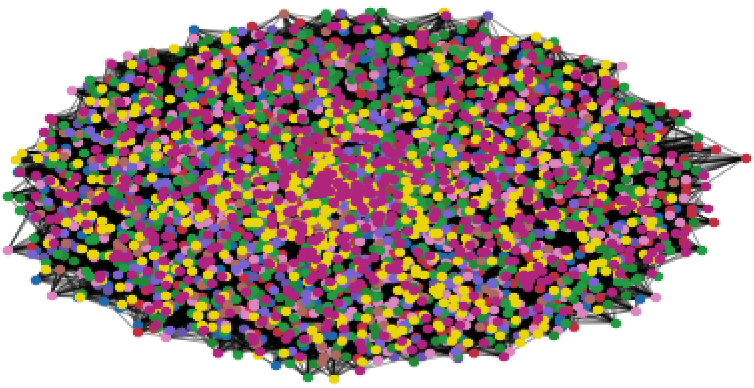
Τους ίδιους αλγοριθμους χρησιμοποιήσαμε και για τον random graph V με την ίδια σειρά

CommunitiesV



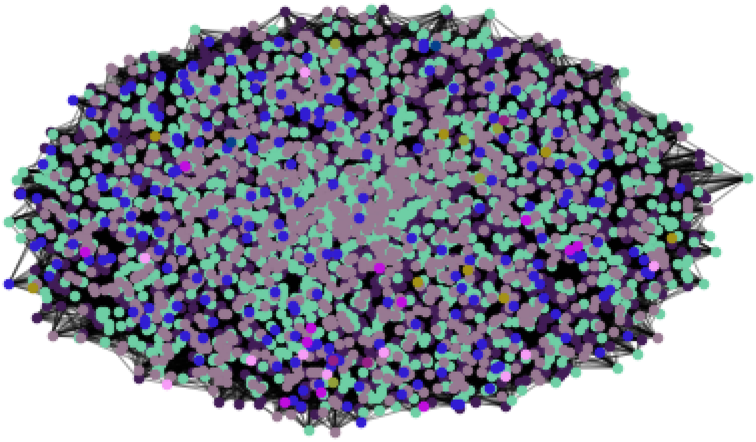
Με 7 communities και greedy_modularity έχουμε αυτό.
Οπως φαίνεται εδώ τα communities είναι σαν να σχηματίζουν σχέδια.
Αυτό εξηγείται διότι ο γράφος είναι random μαζί και albert barabasi.
Επίσης οπτικά έχουμε 1 community

CommunitiesV



Όπως φαίνεται με luvain method έχουμε τα παραπάνω communities.
Εδώ τα πράγματα είναι περίπου το ίδιο με πρίν.
Οπως φαίνεται η luvain δεν δουλεύει καλά σε small world graphs

CommunitiesV

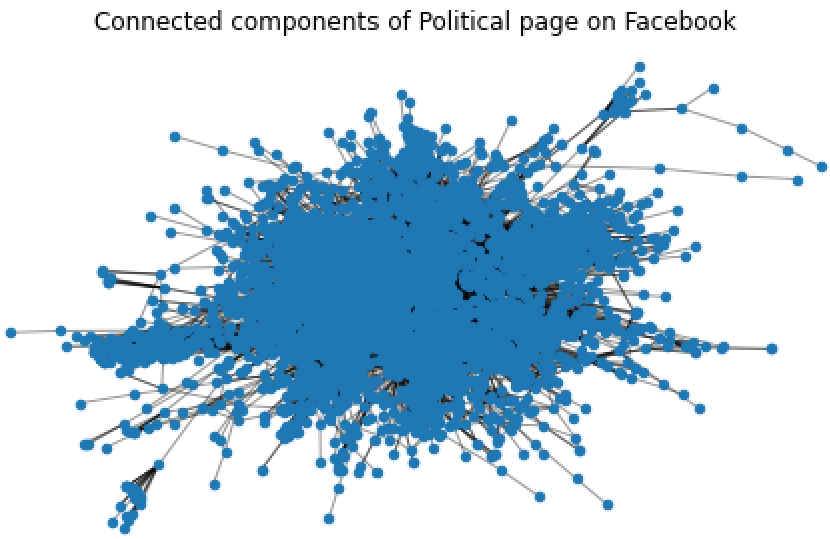


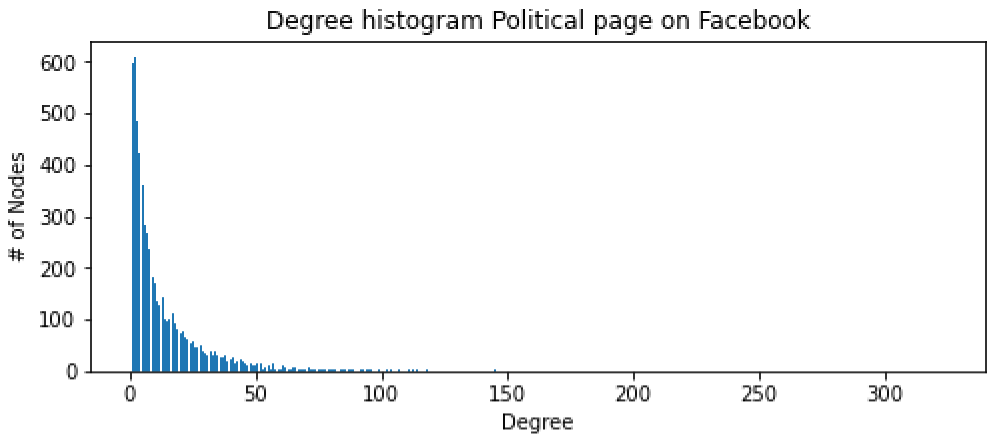
Τέλος προσπαθούμε να βρούμε communities με greedy_modulrity ορίζοντας ως k τον αριθμό των communities που βρήκαμε απ τον luvain παραπάνω.

Οπως φαίνεται το τελευταίο community_detection στον random_graph ήταν το καλύτερο.

Απ ότι φαίνεται δεν εντοπίστηκαν ικανοποιητικά τα communities στις 3 μεθόδους για τον random_graph.
Αυτό υπάρχει περίπτωση να συμβαίνει επειδή μάλλον δεν μπορούν οι αλγόριθμοι να δουλέψουν καλά όταν υπάρχει μόνο μια κοινότητα(οπτικά) σε έναν γράφο

Influence maximization
Στο Dataset του Facebook υπήρχε μια κοινότητα με κόμβους οι οποίοι παρουσίασαν την σχέση μεταξύ χρηστών σε μια σελίδα στο Facebook που αφορά πολιτικές συζητήσεις. Το Dataset όπως και το δείγμα το αρχικό φαίνεται να ακολουθεί την Power law κατανομή. Σκοπός μας είναι να βρούμε την μέγιστη επίδραση που μπορούν να φέρουν k κόμβοι εντός του δικτύου έτσι ώστε να χρησιμοποιηθούν για προεκλογική εκστρατεία ή να δοκιμάσουμε να επηρεάσουμε όλες τις κινότητες ενός δικτύου. Το group με τις πολιτικές συζητήσεις έχει 5886 κόμβους και είναι το παρακάτω.





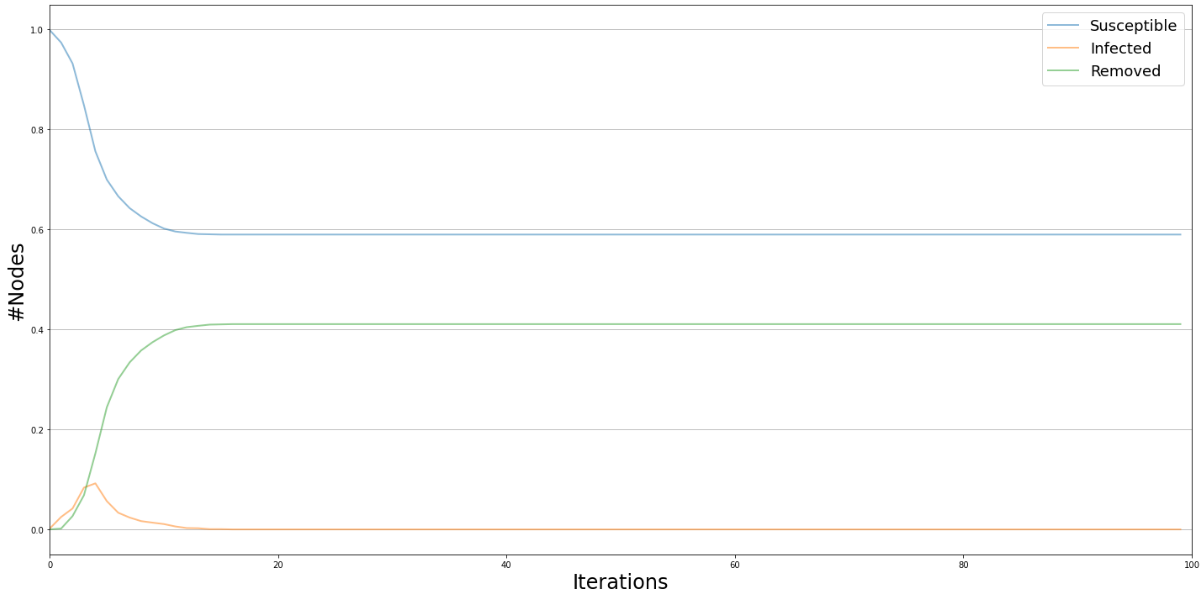
Θα δοκιμάσουμε να δούμε πως μπορεί να μεταδοθεί η πληροφορία εντός του αρχικού μας γράφου που έχει πολλές κοινότητες και ακολουθεί την powerlaw καθώς και εντός του γράφου που απεικονίζει μια κοινίτητα. Για αρχή επιλέξαμε τους 10 κόμβους που έχουν το μεγαλύτερο betweenness centrality, degree, closeness centrality και δοκιμάσαμε τον pagerank. Χρησιμοποιήσαμε το μοντέλο Independent Cascades.
Ουσιαστικά λειτουργεί με τον παρακάτω τρόπο:

Το μοντέλο ξεκινάει με ένα set κόμβων που είναι οι ενεργοί μας κόμβοι την χρονική στιγμή t0 που ξεκινάει η προσομοίωση.

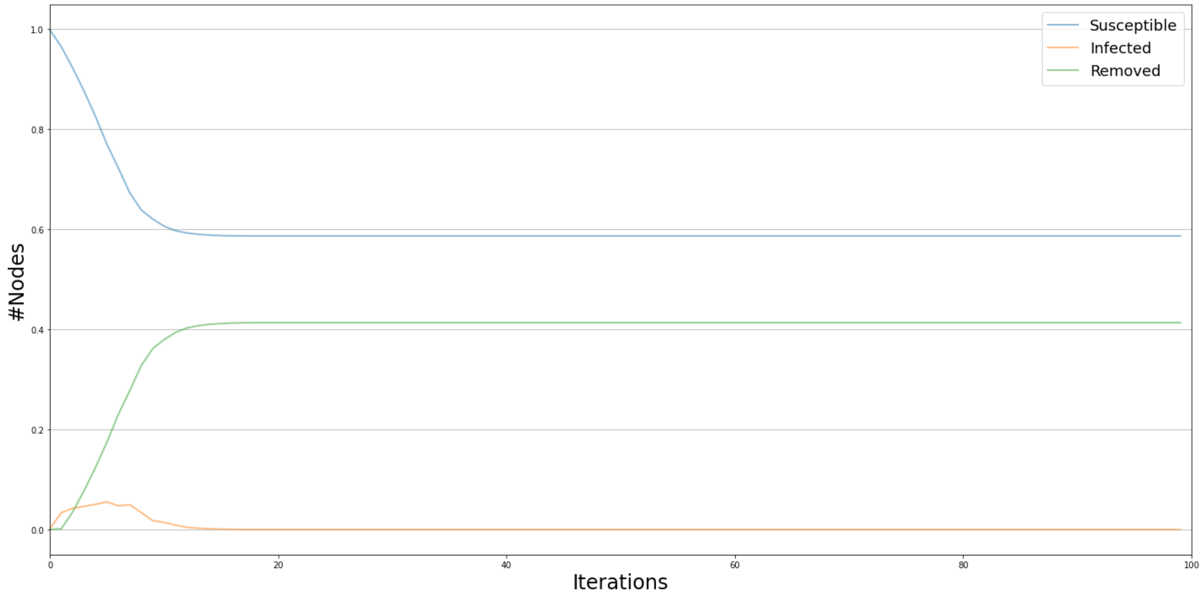
- Όταν ένας κόμβος γίνεται ενεργός στο βήμα t, του δίνεται μια μοναδική ευκαιρία να επιδράσει και να ενεργοποιήσει κάθε ανενεργό γείτονα του w εκείνη την χρονική στιγμή, το πετυχαίνει με πιθανότητα p(v,w).
- Εάν ο w έχει πολλούς νέους ενεργούς κόμβους σαν γείτονες, οι προσπάθειες επίδρασης στον w πραγματοποιούνται σε τυχαία σειρά.
- Εάν κάποιος ενεργός κόμβος v επιτύχει να επηρεάσει τον w, τότε ο w θα γίνει actrive την χρονική στιγμή t+1, σε κάθε άλλη περίπτωση ο κόμβος v δεν θα μπορεί να δοκιμάσει εκ νέου σε κάποια άλλη στιγμή να επηρεάσει τον w.
- Όλη η παραπάνω διαδικασία γίνεται μέχρις ότου να μην μπορεί να γίνουν άλλες ενεργοποιήσεις στο δίκτυο.
- Κάθε κόμβος που γίνεται active θεωρούμε στο μοντέλο μας ότι δε μπορεί να επηρεαστεί και να γίνει inactive εκ νέου.

Αυτά που βγάλαμε σαν αποτελέσματα είναι τα παρακάτω:

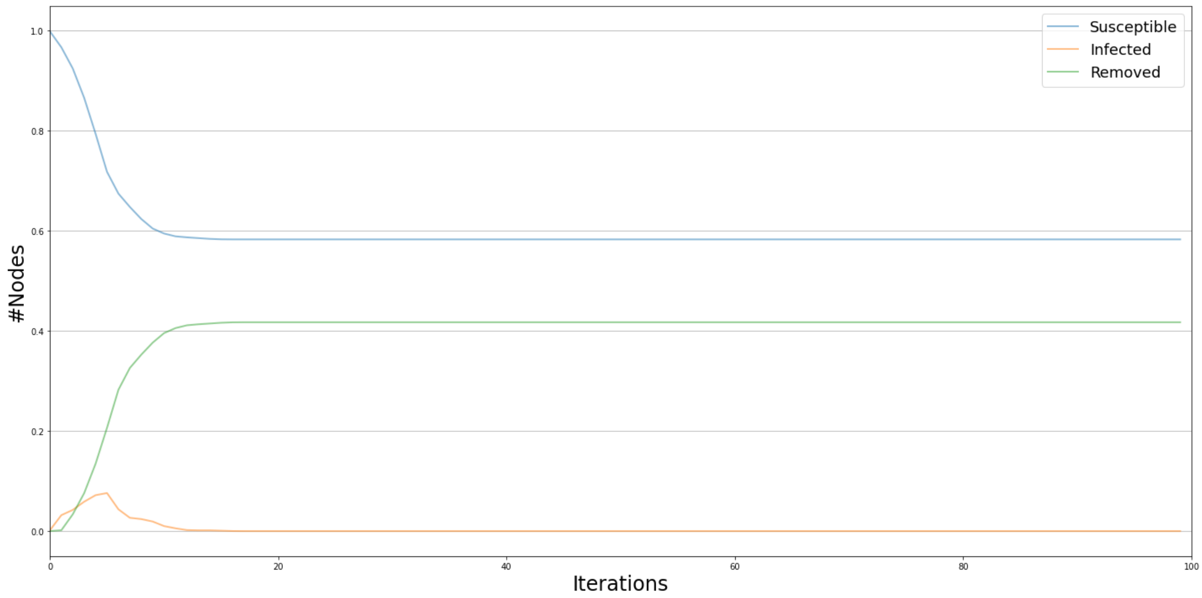
α. Betweenness Centrality (Αρχικοί κόμβοι 10)



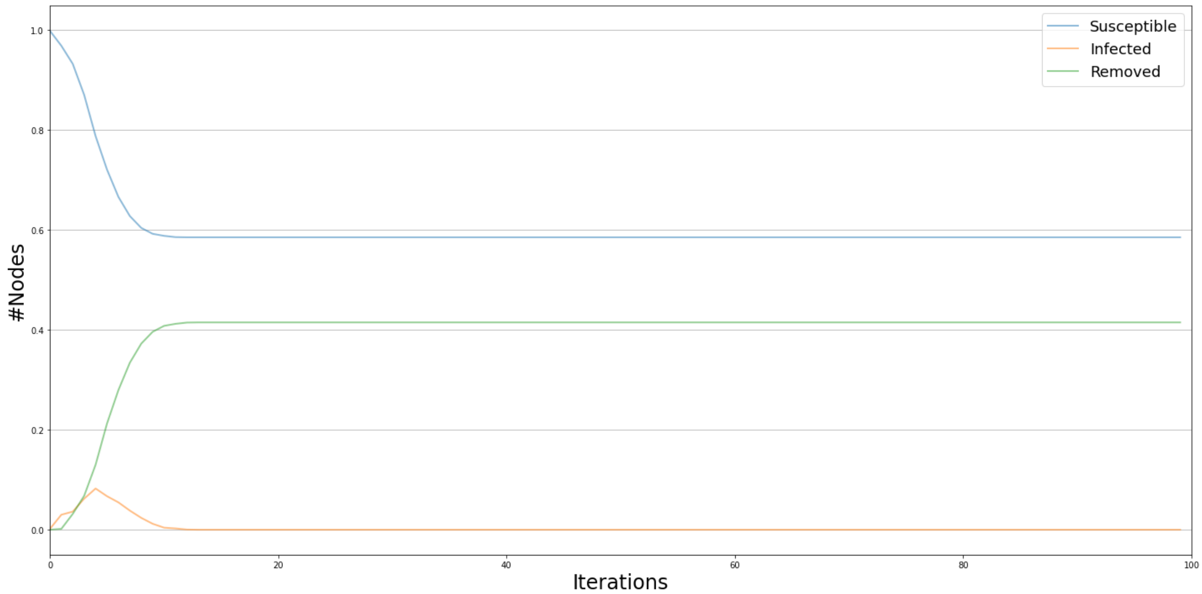
β.Βαθμός κόμβου (Αρχικοί κόμβοι 10)



γ. Eccentricity (Αρχικοί κόμβοι 10)

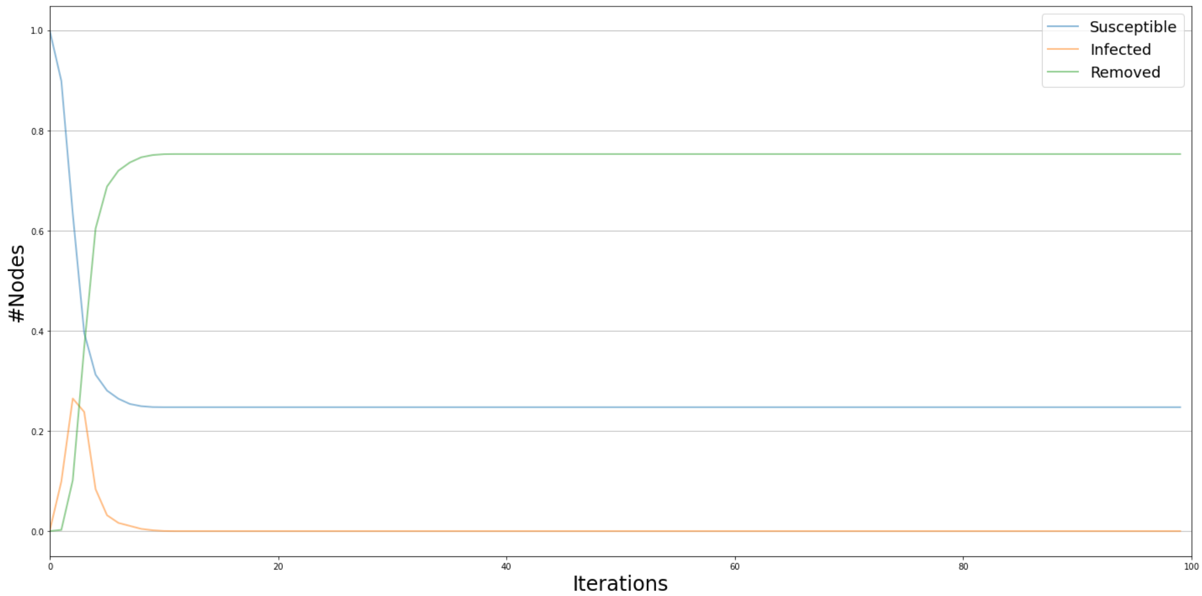


δ. Pagerank (Αρχικοί κόμβοι 10)

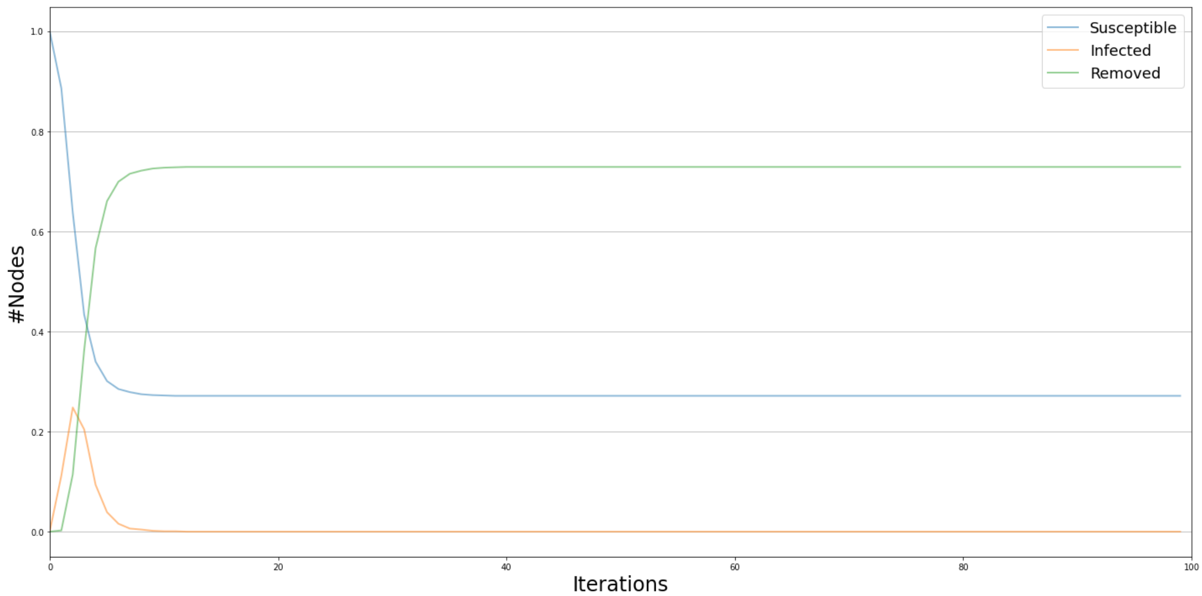


Από ότι παρατηρούμε στα παραπάνω γραφήματα, στο Dataset της σελίδας με τις πολιτικές συζητήσεις καταφέραμε να επηρεάσουμε το 41% περίπου του πλήθους της κοινότητας. Τα αποτελέσματα ήταν σχεδόν ίδια σε όλα τα seeds που επιλέξαμε σαν αρχικά ώστε να επηρεάσουν την κοινότητα. Δοκιμάσαμε και να μεγαλώσουμε τον αριθμό των seeds, όταν προσθέσαμε 700 seeds καταφέραμε να κάνουμε influence το 50% περίπου των κόμβων. Ουσιαστικά θα ήταν οικονομικά ασύμφορο να πληρώνει κάποιος 690 άτομα για να του δώσουν την επιρροή που του προσφέρουν 10 άτομα στο τετραπλάσιο.
Δοκιμάσαμε στον γράφο με τις κοινότητες που παρουσιάσαμε στην εργασία, τα αποτελέσματα που πήραμε τα βλέπουμε παρακάτω:

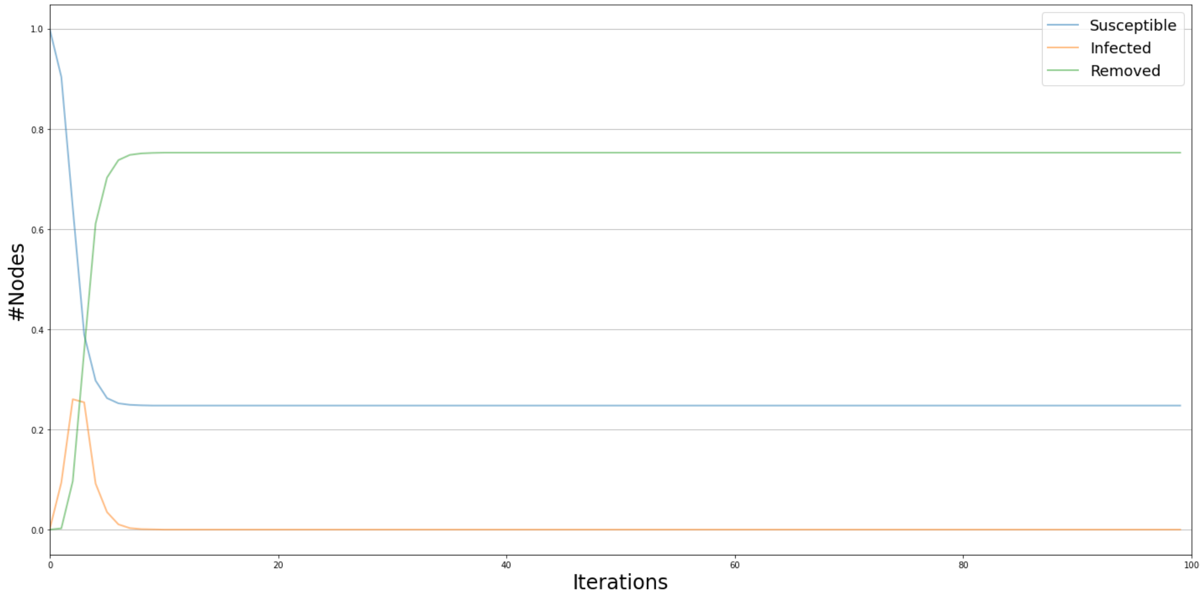
α. Betweenness Centrality (Αρχικοί κόμβοι 10)



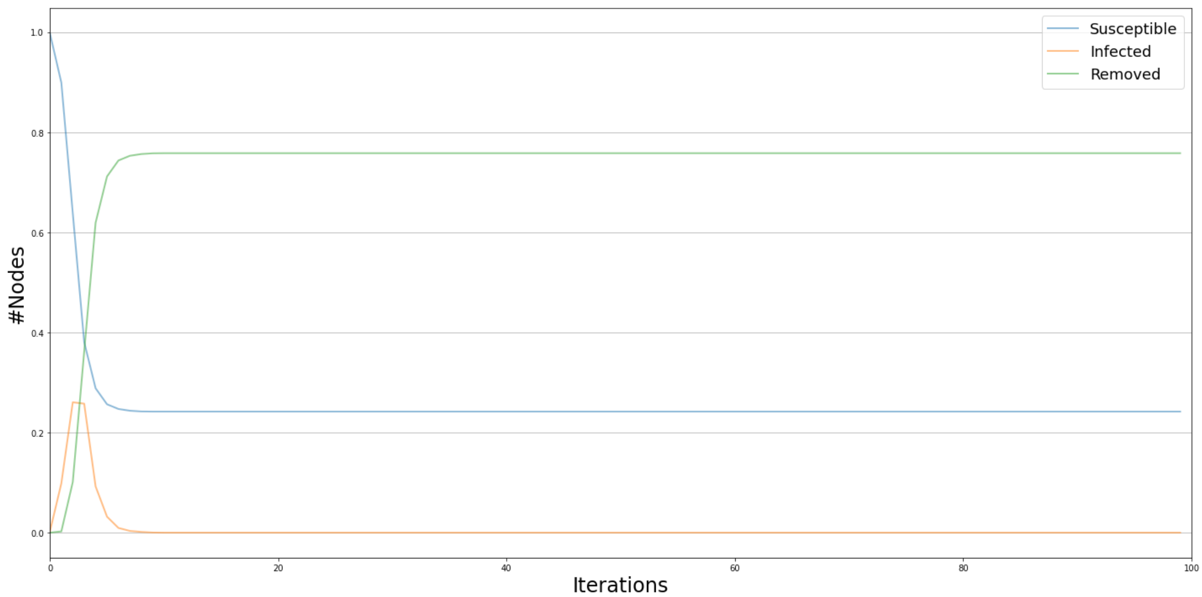
β.Βαθμός κόμβου (Αρχικοί κόμβοι 10)



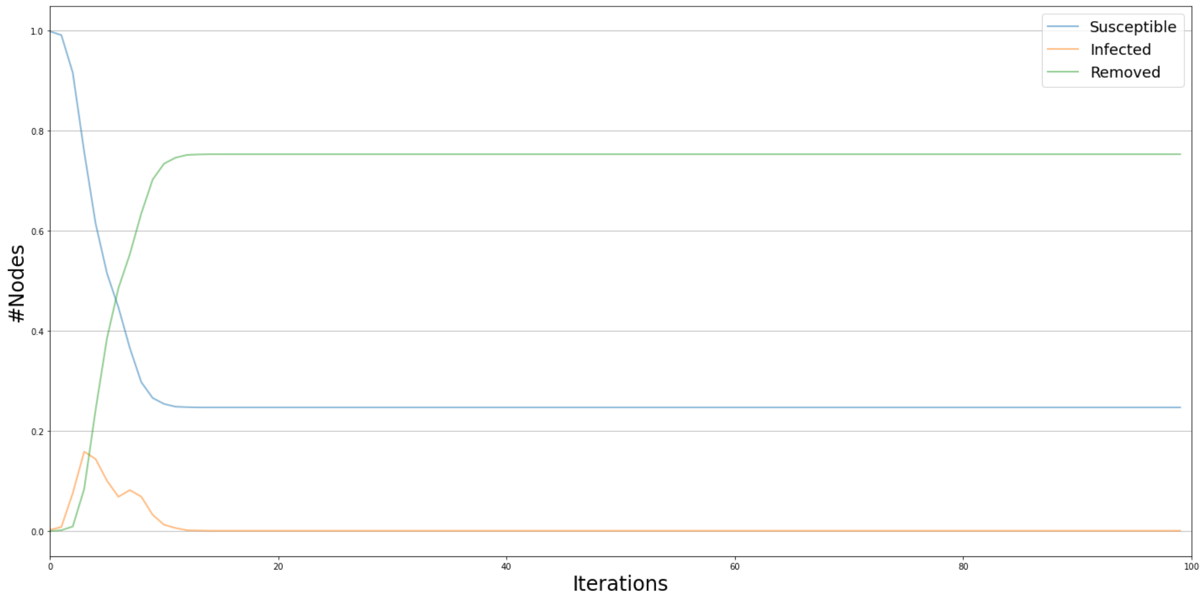
γ. Eccentricity (Αρχικοί κόμβοι 10)



δ. Pagerank (Αρχικοί κόμβοι 10)



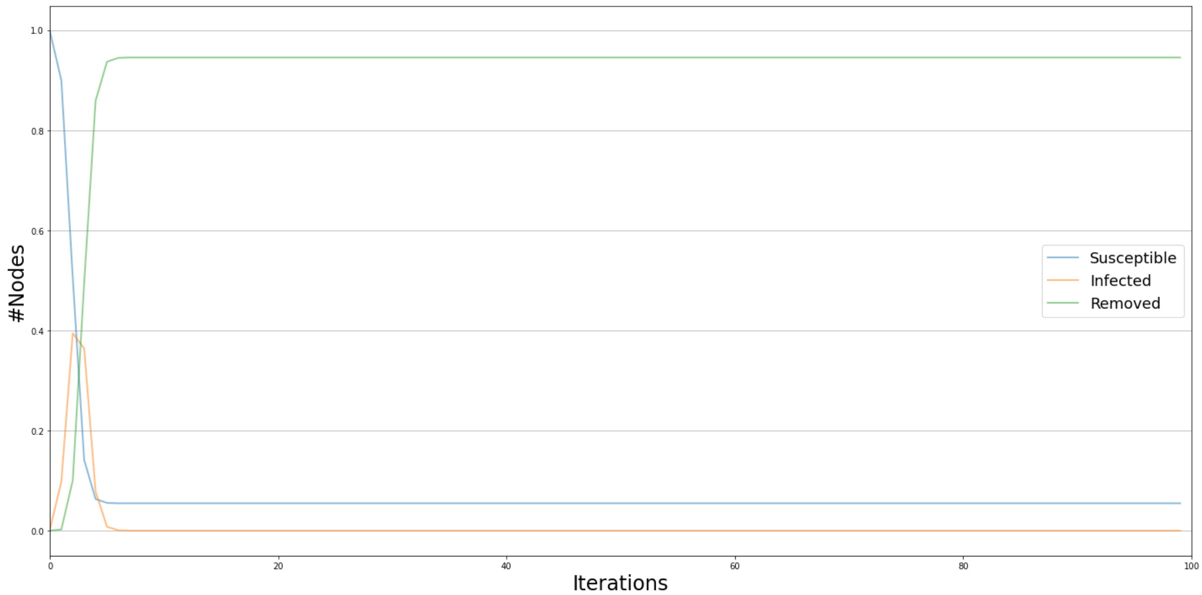
ε. Closeness centrality (Αρχικοί κόμβοι 10)



Στην περίπτωση του αρχικού μας γράφου τα αποτελέσματα ήταν πολύ καλύτερα. Κατάφεραν όλα τα seeds να επηρεάσουν το 77% περίπου όλου του δικτύου με παρόμοια αποτελέσματα ο καθένας. Θα περιμέναμε να βγάλουν διαφορετικά αποτελέσματα τα διαφορετικά seeds αλλά οι ιδιοι κόμβοι που ήταν και οι πιο κεντρικοί, αποτελούσαν και τον δίαυλο επικοινωνίας μεταξύ των διαφορετικών κοινοτήτων. Όποτε από την έρευνα μας είναι προτιμότερο να επιλέξουμε κεντρικούς κόμβους που αποτελούν τον δίαυλο μεταξύ των κοινοτήτων για να πετύχουμε τον στόχο μας που είναι η μέγιστη επιρροή.

Επέκταση εργασίας

Δοκιμάσαμε να κάνουμε και μεγιστοποίηση της επιρροής στον γράφο V τον τυχαίο μας. Τα αποτελέσματα ήταν τα παρακάτω.



Παρατηρήσαμε ότι μόλις σε λίγες επαναλήψεις είχαμε μεγάλη επιρροή στο 92% περίπου του δικτύου. Επίσης το παραπάνω βγήκε σε όλα τα seeds. Είτε προσθέταμε τους κόμβους με το μεγαλύτερο degree, είτε με το centrality. Για επέκταση της εργασίας ίσως θα άξιζε στον γράφο G να βρίσκαμε τους 20 κόμβους με το μεγαλύτερο degree, τους 20 με το μεγαλύτερο betweenness και closeness centrality και τους 20 με το μεγαλύτερο eccentricity. Από αυτούς λοιπόν όσους εμφανίζονταν σε όλα τα top 20 των μετρικών θα το χρησιμοποιήσουμε στα seeds και αν δεν βγάζαμε τα δέκα που επιθυμούσαμε, θα πηγαίναμε σε συνδυασμούς τριάδων,δυάδων.