



ΕΥΦΥΗΣ ΑΛΛΗΛΕΠΙΔΡΑΣΗ ΜΕ ΚΟΙΝΩΝΙΚΑ ΔΙΚΤΥΑ

ΕΡΓΑΣΙΑ 2018 - 2019



Π16036	ΙΩΑΝΝΙΔΗΣ ΠΑΝΑΓΙΩΤΗΣ
Π16097	ΝΙΚΑΣ ΔΙΟΝΥΣΙΟΣ
Π16112	ΠΑΡΑΒΑΝΤΗΣ ΑΘΑΝΑΣΙΟΣ

I. Περιεχόμενα

II. Επεξήγηση	2
III. Εκτέλεση.....	4
IV. Γραφήματα	6
A. Subgraphs.....	6
B. Betweenness Centrality.....	9
C. Closeness Centrality	12
D. Degree Centrality	15
E. Eigenvector Centrality.....	18
F. In Degree Centrality	21
G. Katz Centrality.....	24
H. Out Degree Centrality	27

II. Επεξήγηση

Η εργασία του μαθήματος που παραδίδουμε αναλύει σε βάθος το πακέτο δεδομένων StackOverflow temporal network, αναδεικνύοντας ουσιαστικές πληροφορίες και σχέσεις μεταξύ προγραμματιστών. Για την επίλυση της εργασίας χρησιμοποιήθηκαν οι εξής τεχνολογίες:

- Γλώσσα προγραμματισμού Python 3.7
- Βιβλιοθήκη ανάλυσης δεδομένων Pandas
<https://pandas.pydata.org/>
- Βιβλιοθήκη γραφημάτων NetworkX
<https://networkx.github.io/>
- Βιβλιοθήκη γραφικών παραστάσεων Matplotlib
<https://matplotlib.org/>

Επειδή τα δεδομένα που μας δίνονται περιέχουν εκατομμύρια γραμμές, αποφασίσαμε να δουλέψουμε πάνω στις πρώτες 1.000 καταχωρήσεις, δηλαδή σε ένα υποσύνολο όλων των συνεργασιών μεταξύ προγραμματιστών.

Αφού φορτώσουμε τα δεδομένα με τη βιβλιοθήκη Pandas υπολογίζουμε την ελάχιστη και μέγιστη χρονική στιγμή για το 1^ο ερώτημα. Προχωρώντας στο 2^ο ερώτημα, χωρίζουμε αυτή τη χρονική περίοδο σε N διαστήματα που καθορίζονται από τον χρήστη. Έχοντας υπ' όψη όλες τις χρονικές περιόδους βασιζόμενοι πάντα στους τύπους της εκφώνησης, στο 3^ο ερώτημα σχεδιάζουμε με τη βοήθεια της βιβλιοθήκης NetworkX όλα τα γραφήματα, όπως εξελίσσονται στη διάρκεια του χρόνου.

Τα γραφήματα του 3^{ου} ερωτήματος έρχονται να συμπληρώσουν το 4^ο ερώτημα, το οποίο υπολογίζει τα εξής για κάθε γράφημα: degree centrality, in-degree centrality, out-degree centrality, closeness centrality, betweenness centrality, eigenvector centrality και katz centrality. Τα ευρήματα των υπολογισμών αποθηκεύονται στον ίδιο φάκελο σε μορφή γραφημάτων χρησιμοποιώντας τη βιβλιοθήκη Matplotlib. Ο αριθμός των γραφημάτων έχει άμεση σχέση με τον αριθμό των χρονικών περιόδων και με τις συνεργασίες των προγραμματιστών.

Για το 5^ο ερώτημα, βρίσκουμε τις κοινές κορυφές και ακμές μεταξύ όλων των περιόδων. Για την εύρεση των κοινών κορυφών βρίσκουμε την τομή μεταξύ κάθε χρονικής περιόδου και αντίστοιχα για την εύρεση των ακμών, υπολογίζουμε τα ζευγάρια κόμβων που παραμένουν σταθερά στην εξέλιξη του χρόνου. Χρησιμοποιώντας τις πράξεις συνόλων της Python, οι υπολογισμοί αυτοί γίνονται εύκολα και γρήγορα.

Στο 6^ο ερώτημα, χρησιμοποιούμε ξανά τις χρήσιμες μεθόδους τους NetworkX για τα ζεύγη κόμβων στην εξέλιξη των γραφημάτων. Υπολογίζουμε τα εξής: graph distance, common neighbors, jaccard's coefficient, adamic / adar και preferential attachment. Αφού γίνουν οι υπολογισμοί των μετρικών αυτών, προχωράμε αμέσως στο 7^ο ερώτημα που καλούμαστε να μετρήσουμε την αποτελεσματικότητα της πρόγνωσης μελλοντικών ακμών για κάθε μετρική. Οι τιμές των παραμέτρων καθορίζονται από τον χρήστη και οι τιμές τους κυμαίνονται μεταξύ του 0 και 1.

Η πηγαίος κώδικας που απαντάει στα ερωτήματα της εργασίας βρίσκεται στο αρχείο main.py. Για την εκτέλεση: python main.py αφού εγκαταστήσουμε τις βιβλιοθήκες που χρησιμοποιούνται.

III. Εκτέλεση

Παρακάτω ακολουθούν στιγμιότυπα από την εκτέλεση του προγράμματος. Αρχικά, στην εικόνα 1, παρουσιάζονται τα αποτελέσματα που εμφανίζονται στη κονσόλα για τα ερωτήματα 1, 2, 3, 4, 5.

```
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018, 22:20:52) [MSC v.1916 32 bit (Intel)] on win32
In[2]: runfile('C:/Users/pioan/PycharmProjects/sotiropoulos/main.py', wdir='C:/Users/pioan/PycharmProjects/sotiropoulos')
== StackOverflow Dataset ==
Reading first 1000 rows...

-- 1st part -----
t_min: 1217567877
t_max: 1218036494

-- 2nd part -----

Enter partitioning value (it must be an integer): >? 3

-- 3rd part -----

Sub-graphs have been created!

-- 4th part -----
All centralities have been plotted and saved !!
All centralities have been plotted and saved !!
All centralities have been plotted and saved !!

-- 5th part -----
```

Εικόνα 1 - Αποτελέσματα ερωτημάτων 1, 2, 3, 4, 5 με N=3.

Στη συνέχεια βλέπουμε τα αποτελέσματα για τα ερωτήματα 6 και 7, εικόνα 2 και εικόνα 3.

```
-- 6th and 7th part -----
Give param_GD. Must be 0 < param_GD <= 1
>? 0.3
Give param_CN. Must be 0 < param_CN <= 1
>? 0.5
Give param_JC. Must be 0 < param_JC <= 1
>? 0.6
Give param_A. Must be 0 < param_A <= 1
>? 0.7
Give param_PA. Must be 0 < param_PA <= 1
>? 0.3
++ Result for set of node 0 +++++
The success rate for Graph Distance is: 0.6388804380894433 %
The success rate for Common Neighbors is: 0.2061855670103093 %
The success rate for Jaccard Coefficient is: 0.0 %
The success rate for Adamic Adar is: 0.0 %
The success rate for Preferential Attachment is: 0.6246450880181714 %
```

Εικόνα 2 - param_GD=0.3, param_CN=0.5, param_JC=0.6, param_A=0.7, param_PA=0.3 για το 1^ο γράφημα που προκύπτει.

```

Give param_GD. Must be 0 < param_GD <= 1
>? 0.9
Give param_CN. Must be 0 < param_CN <= 1
>? 0.6
Give param_JC. Must be 0 < param_JC <= 1
>? 0.4
Give param_A. Must be 0 < param_A <= 1
>? 0.5
Give param_PA. Must be 0 < param_PA <= 1
>? 0.3
++ Result for set of node 1 ++++++
The success rate for Graph Distance is: 0.2548853016142736 %
The success rate for Common Neighbors is: 0.8116883116883116 %
The success rate for Jaccard Coefficient is: 0.9900990099009901 %
The success rate for Adamic Adar is: 1.7804154302670623 %
The success rate for Preferential Attachment is: 0.6491945179129598 %

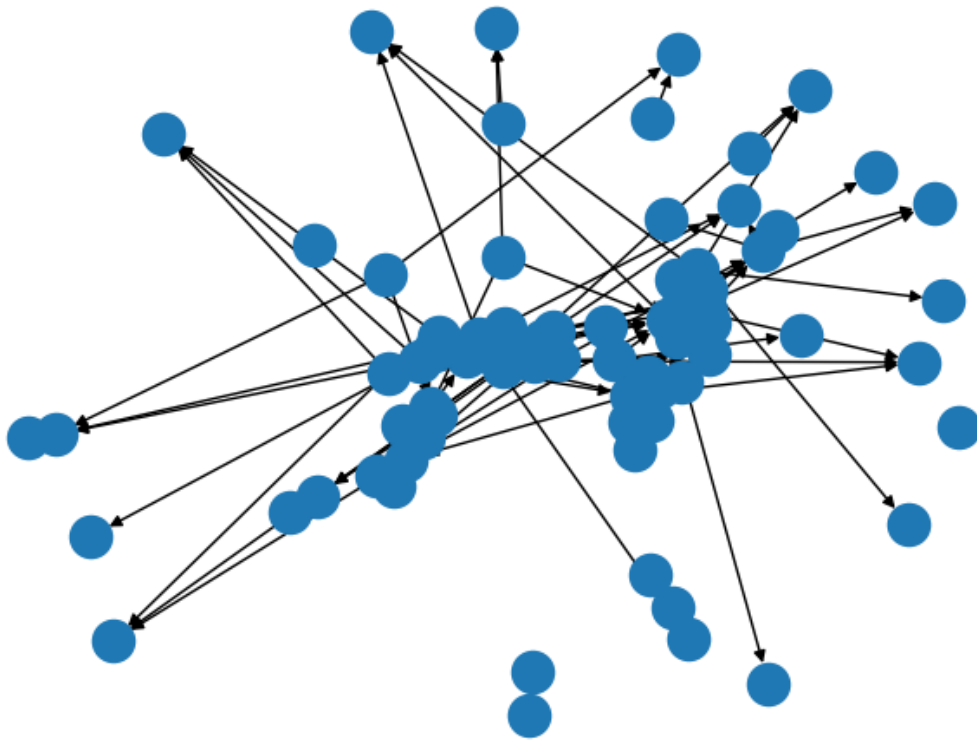
```

Εικόνα 3 - param_GD=0.9, param_CN=0.6, param_JC=0.4, param_A=0.5, param_PA=0.3
για το 2^ο γράφημα που προκύπτει.

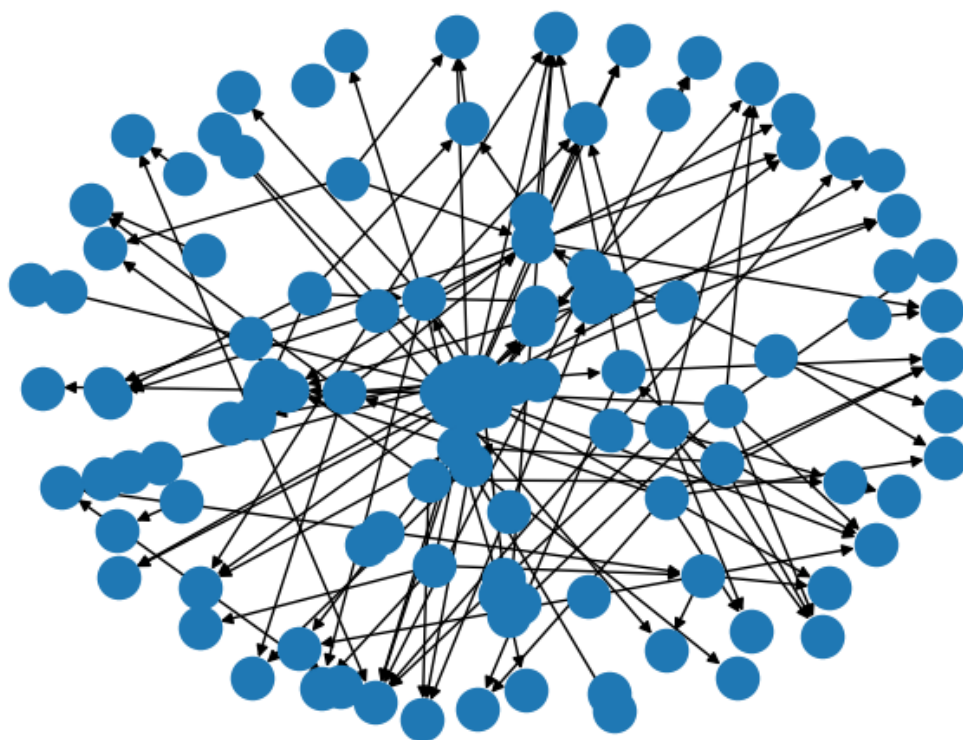
IV. Γραφήματα

Τα παρακάτω γραφήματα είναι αποτέλεσμα της εκτέλεσης του 4^{ου} ερωτήματος για N=3 χρονικές περιόδους.

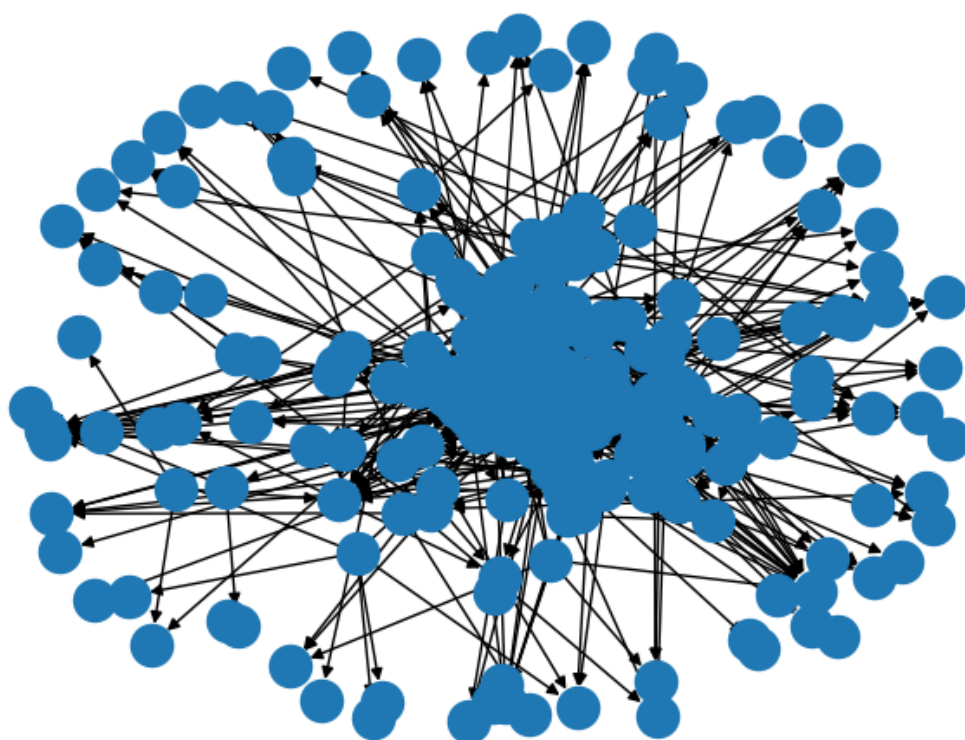
A. Subgraphs



Εικόνα 4 - Υπογράφημα 1

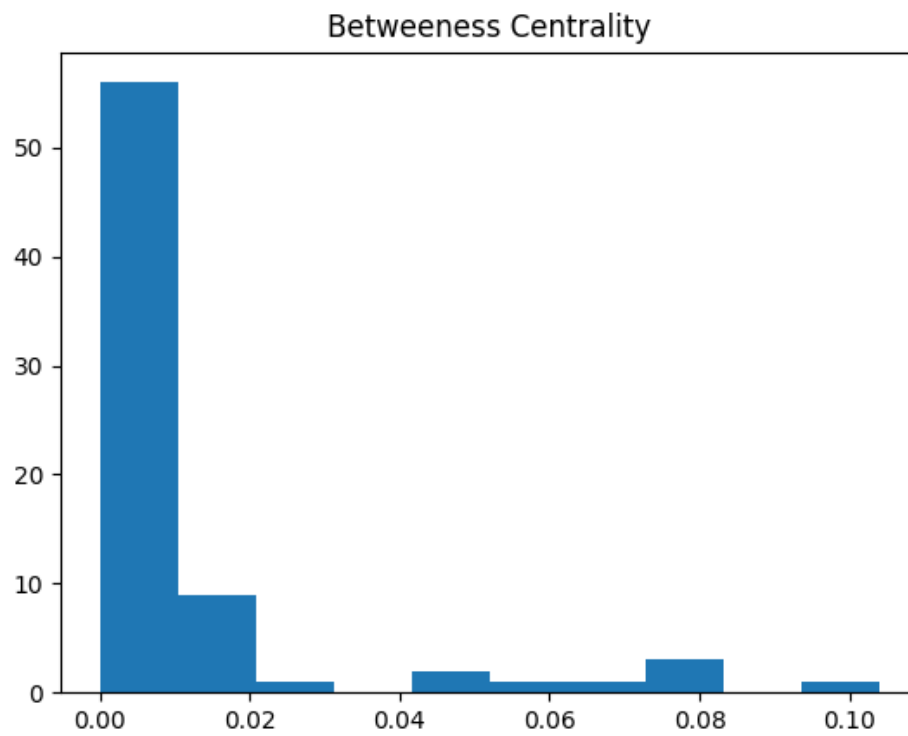


Εικόνα 5 - Υπογράφημα 2

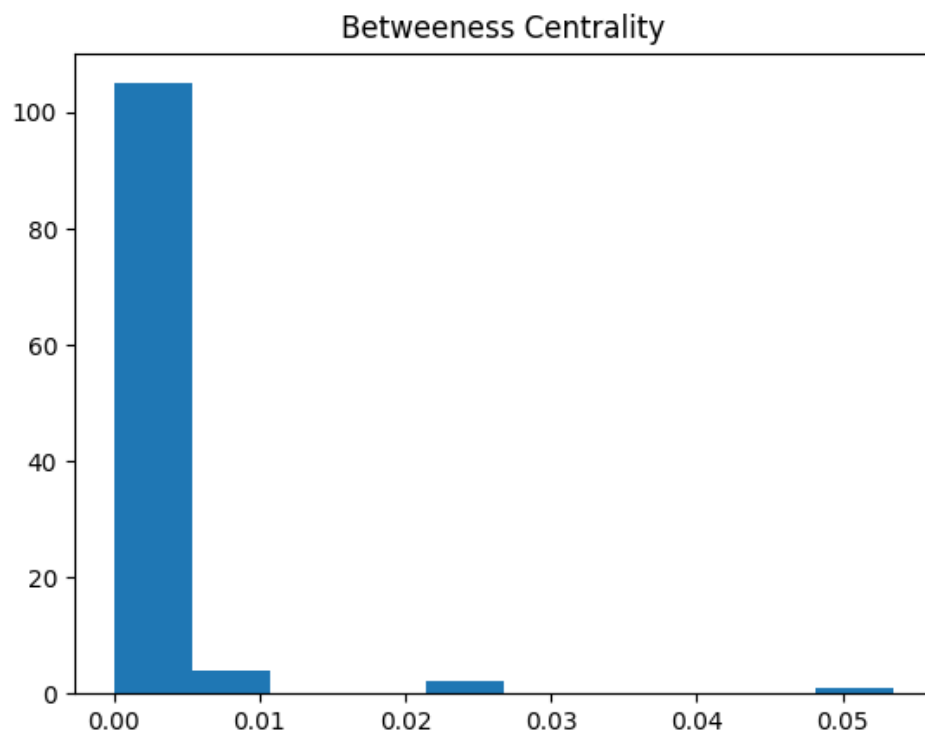


Εικόνα 6 - Υπογράφημα 3

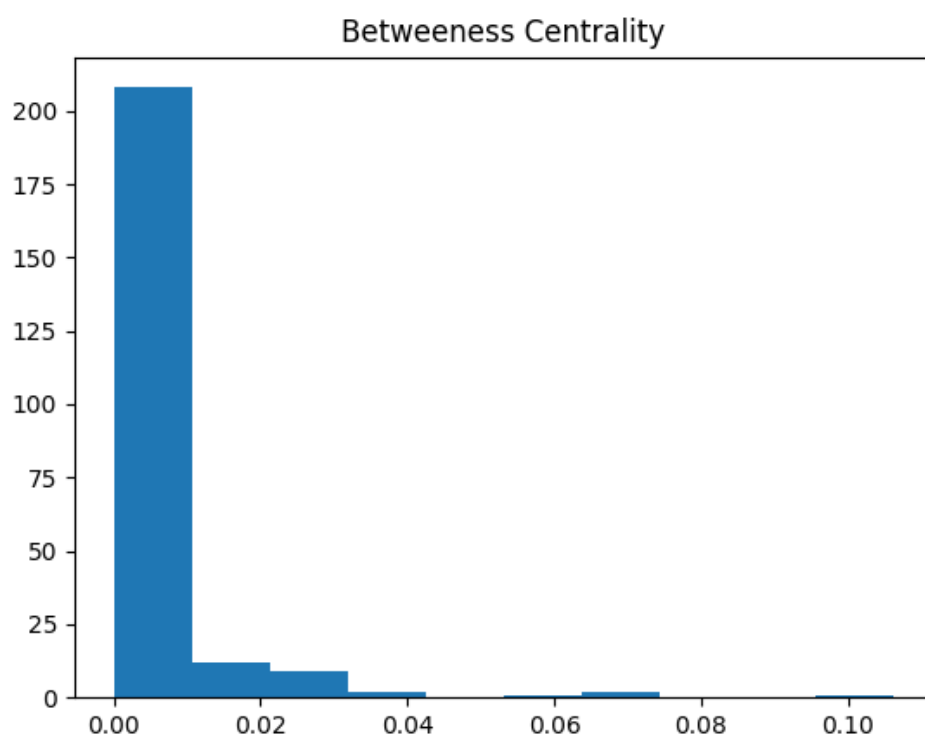
B. Betweenness Centrality



Εικόνα 7 - Betweenness Centrality για το 1 από τα 3 γραφήματα

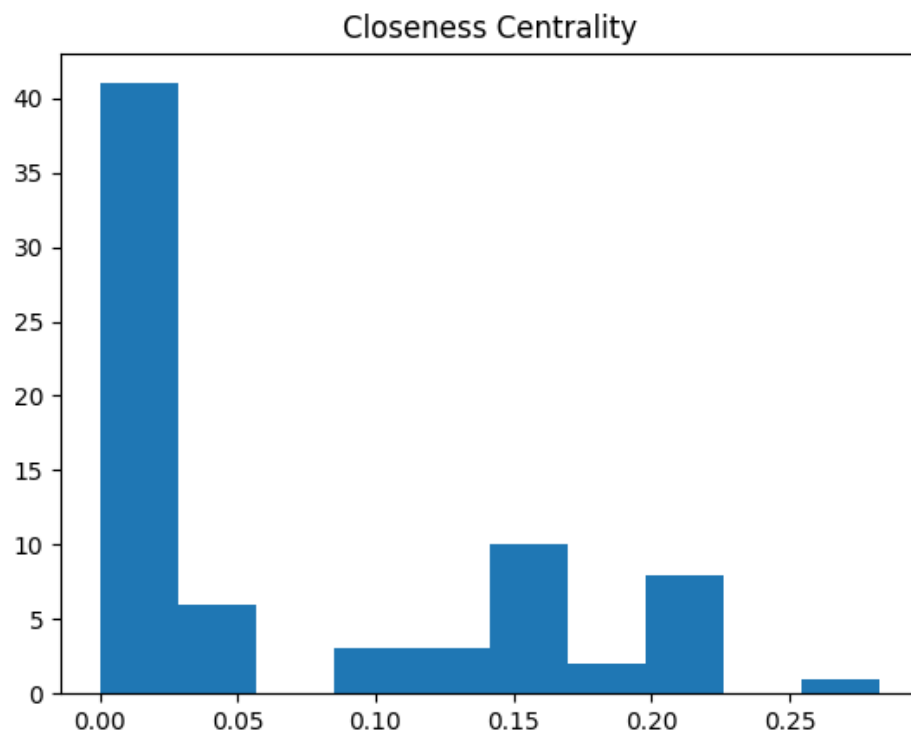


Εικόνα 8 - Betweenness Centrality για το 2 από τα 3 γραφήματα

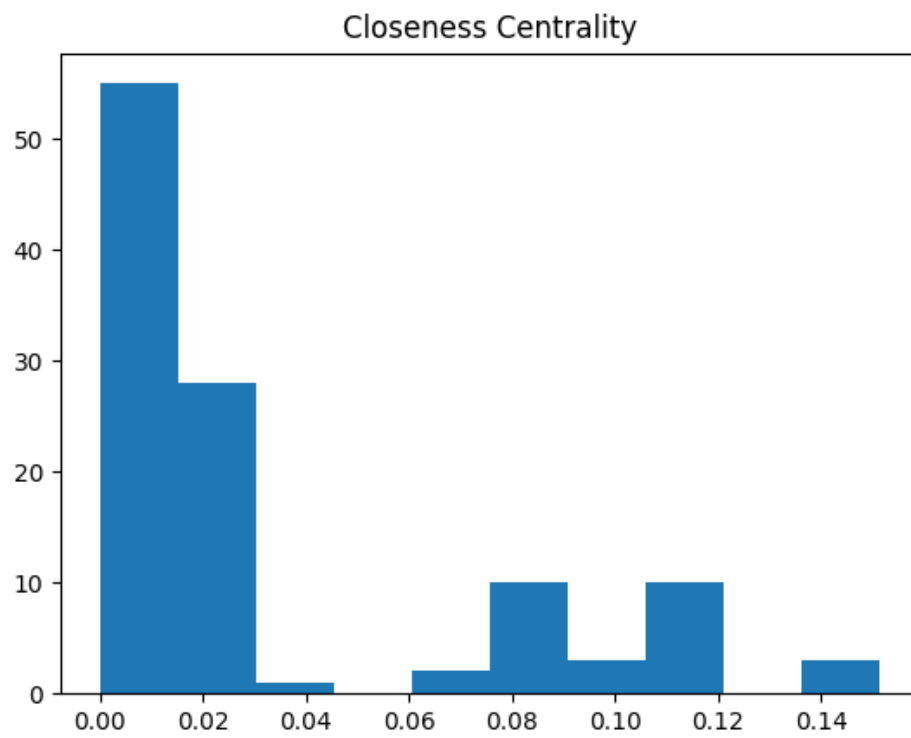


Εικόνα 9 - Betweenness Centrality για το 3 από τα 3 γραφήματα

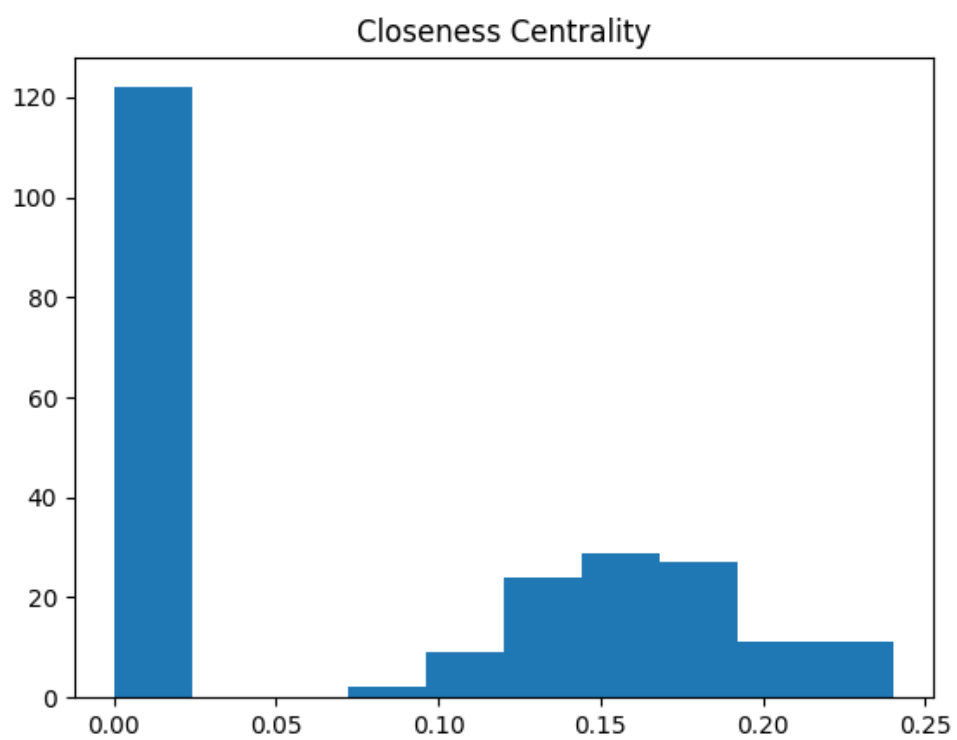
C. Closeness Centrality



Εικόνα 10 - Closeness Centrality για το 1 από τα 3 γραφήματα

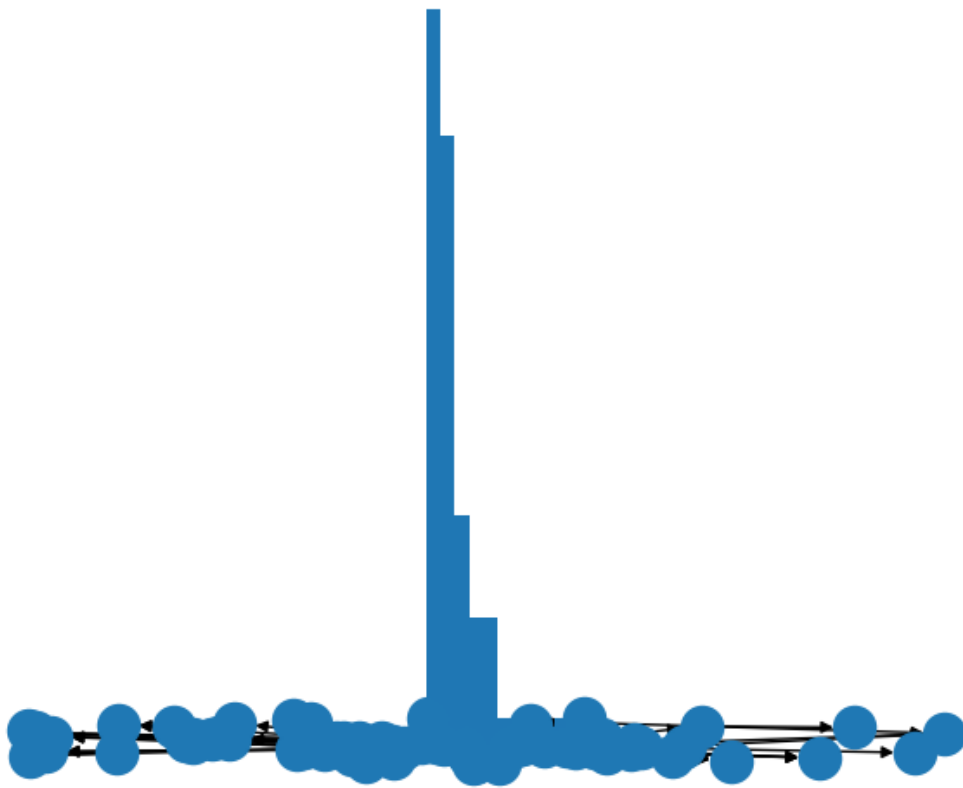


Εικόνα 9 - Closeness Centrality για το 2 από τα 3 γραφήματα

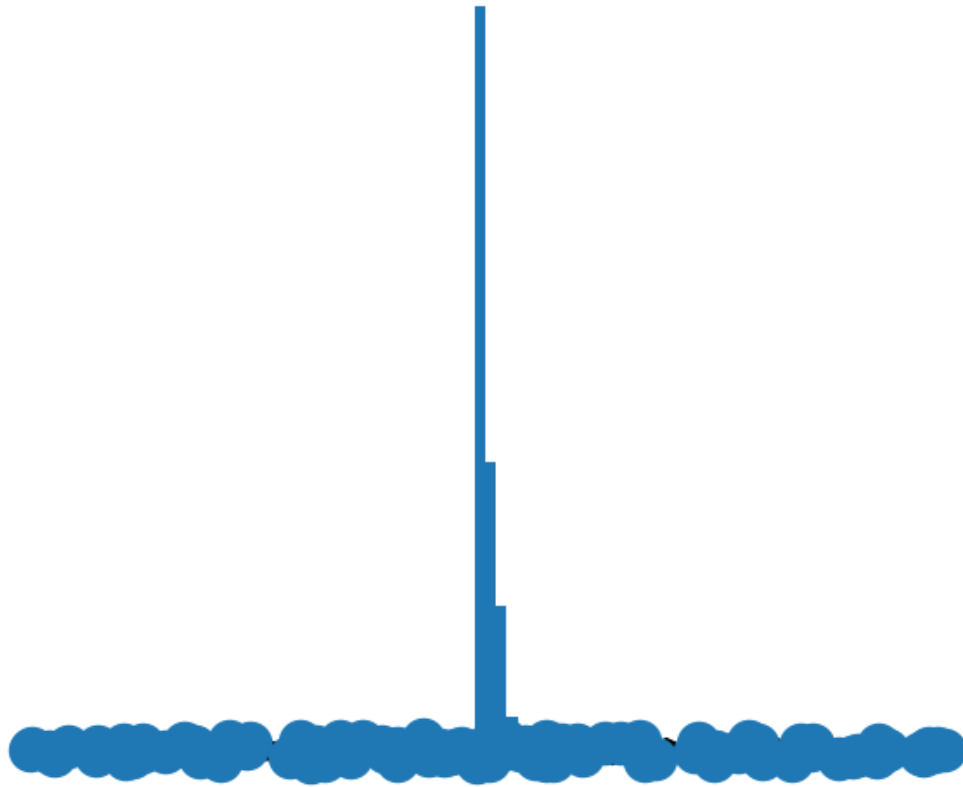


Εικόνα 11 - Closeness Centrality για το 3 από τα 3 γραφήματα

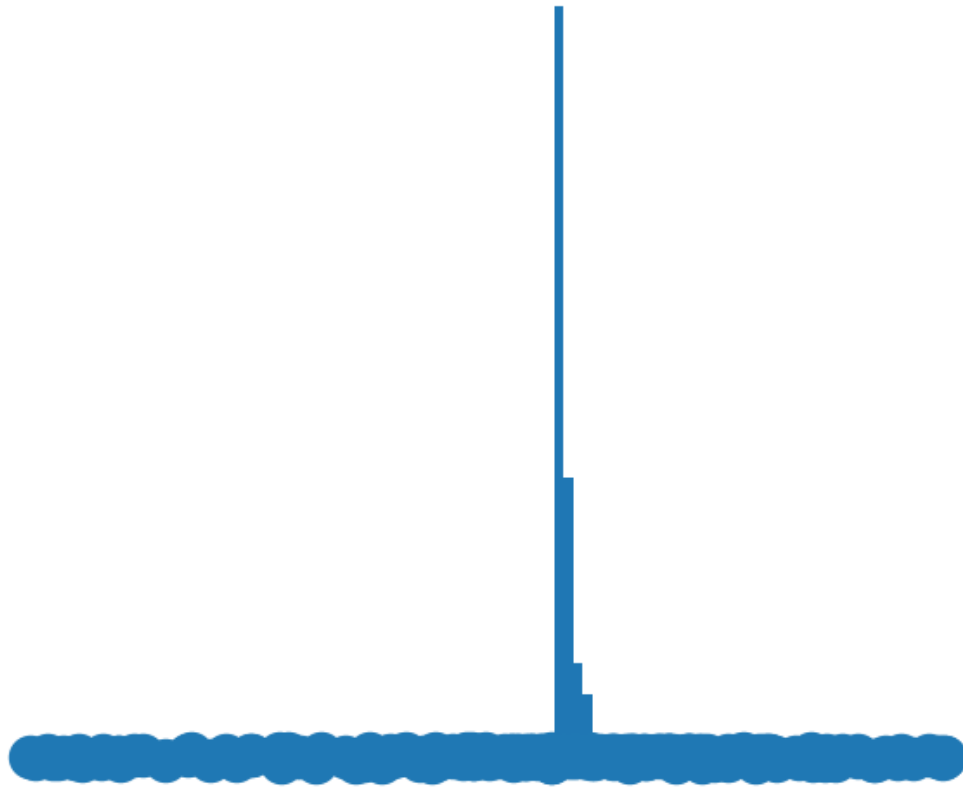
D. Degree Centrality



Εικόνα 12 - Degree Centrality για το 1 από τα 3 γραφήματα

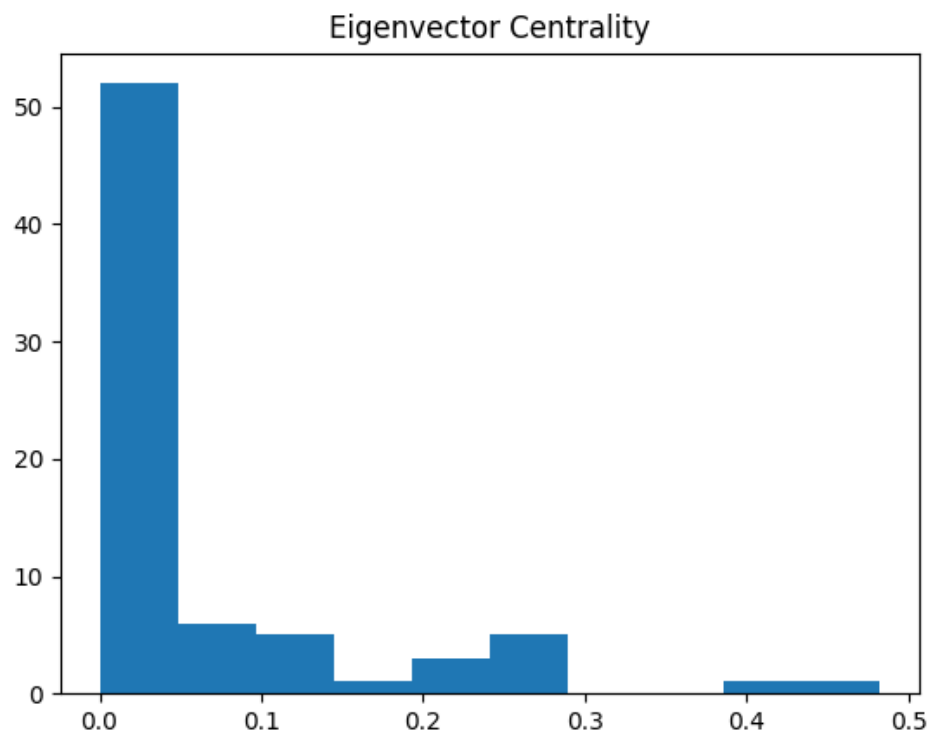


Εικόνα 13 - Degree Centrality για το 2 από τα 3 γραφήματα

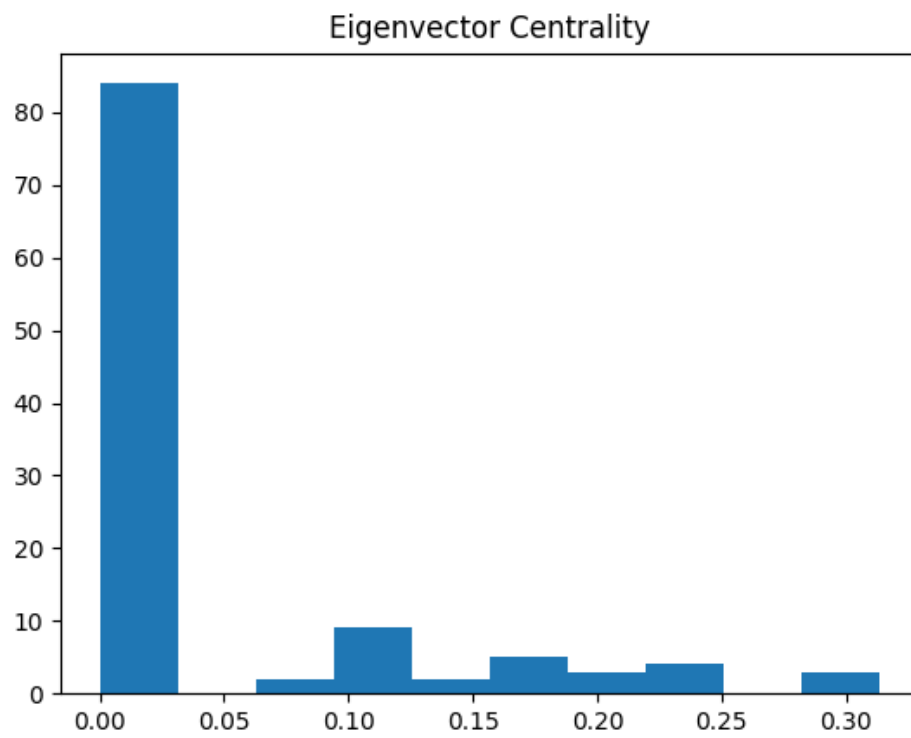


Εικόνα 14 - Degree Centrality για το 3 από τα 3 γραφήματα

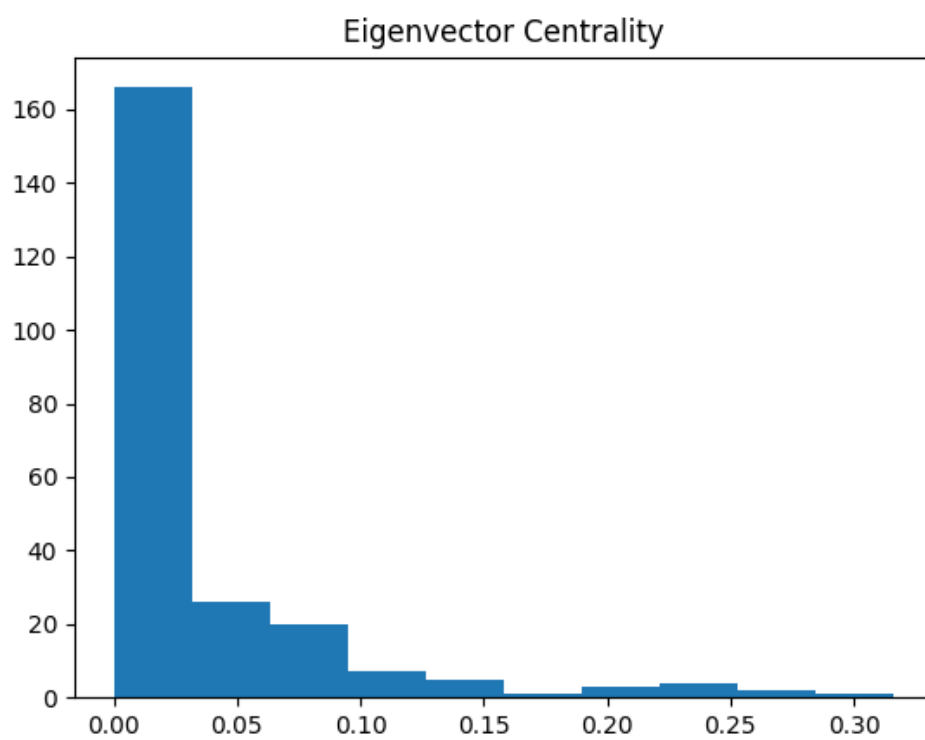
E. Eigenvector Centrality



Εικόνα 15 - Eigenvector Centrality για το 1 από τα 3 γραφήματα

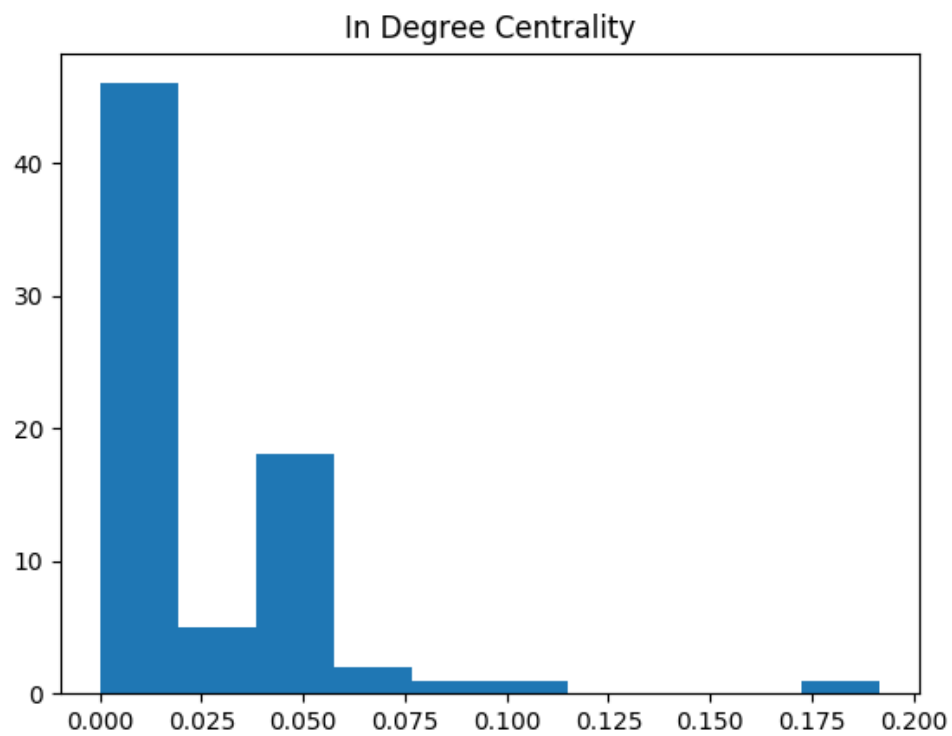


Εικόνα 16 - Eigenvector Centrality για το 2 από τα 3 γραφήματα

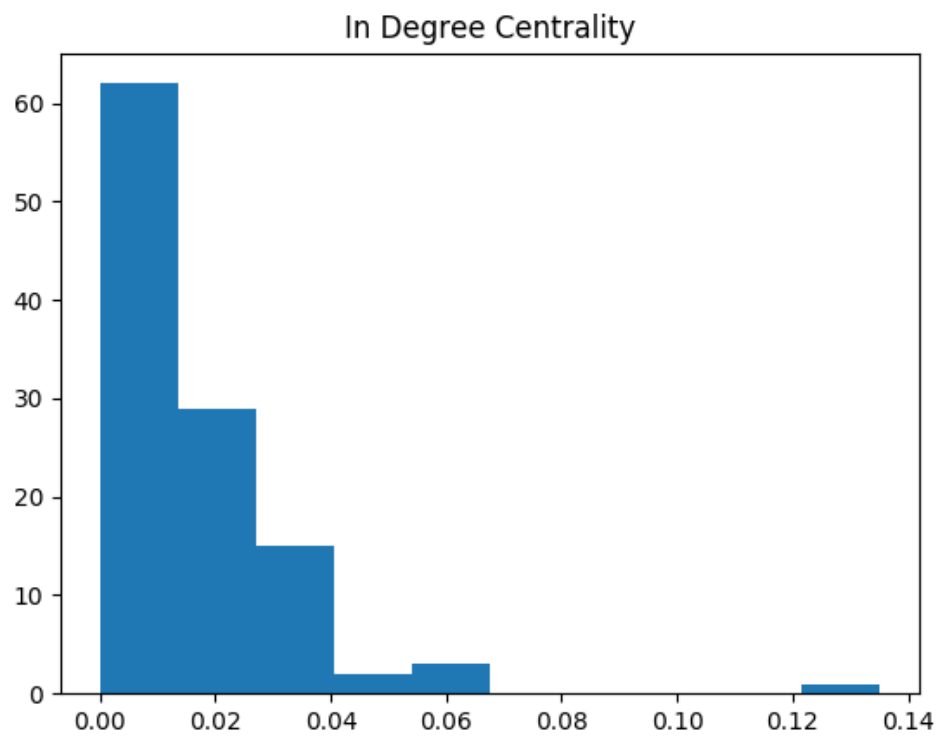


Εικόνα 17 - Eigenvector Centrality για το 3 από τα 3 γραφήματα

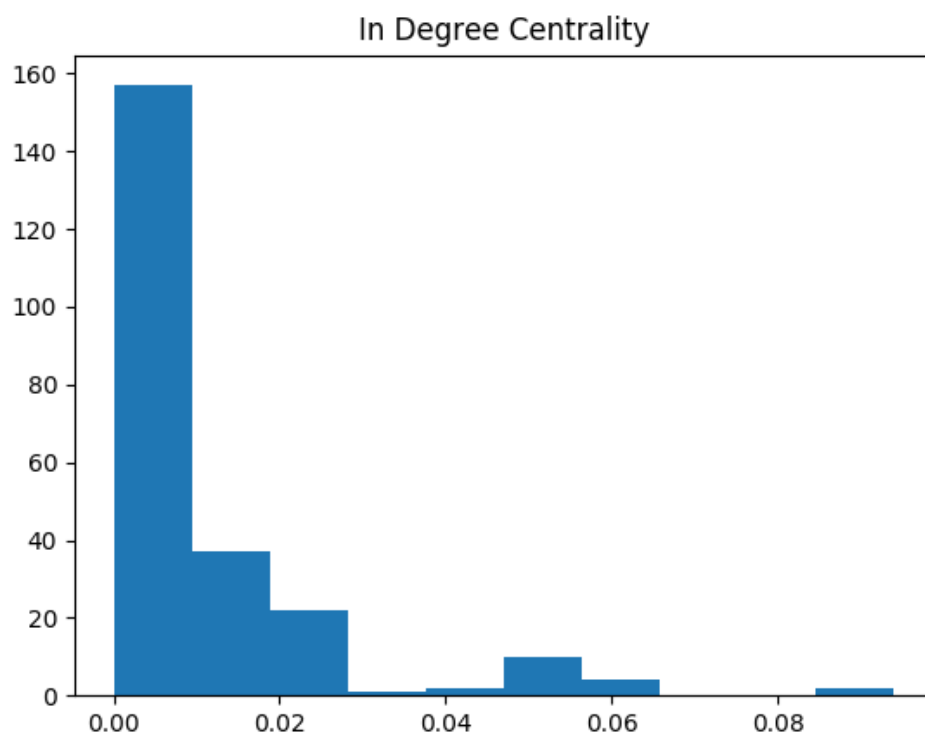
F. In Degree Centrality



Εικόνα 18 - In Degree Centrality για το 1 από τα 3 γραφήματα

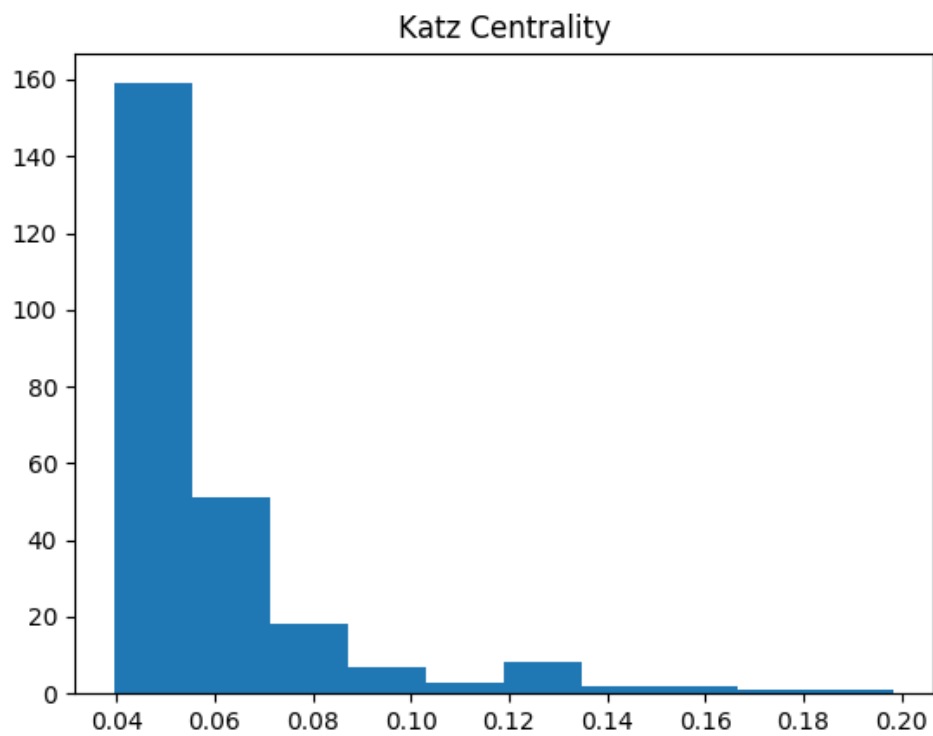


Εικόνα 19 - In Degree Centrality για το 2 από τα 3 γραφήματα

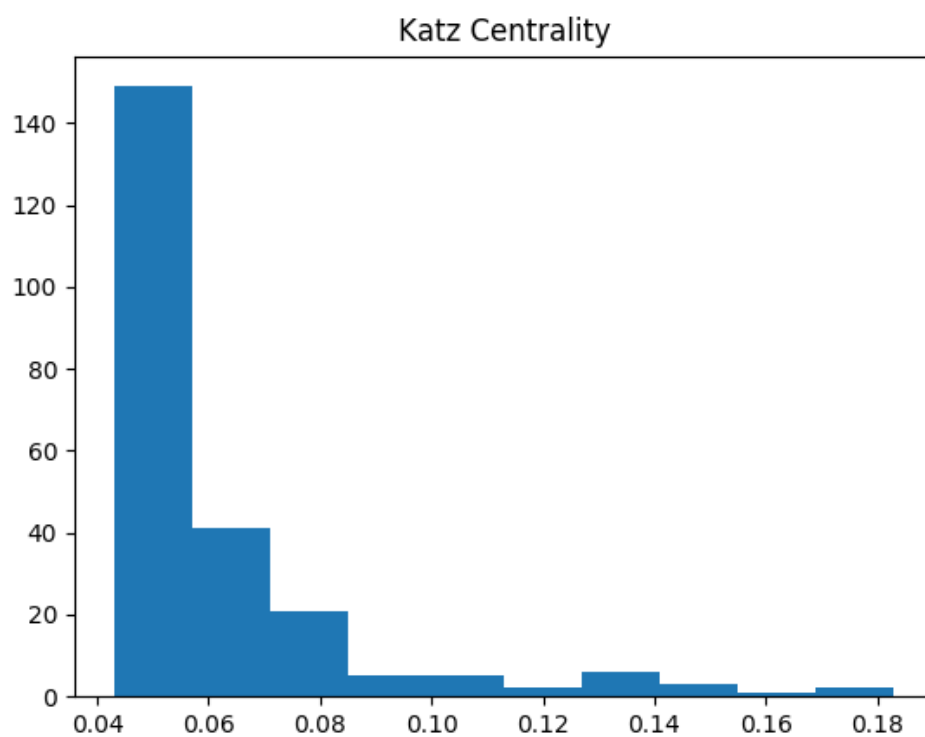


Εικόνα 20 - Katz Centrality για το 3 από τα 3 γραφήματα

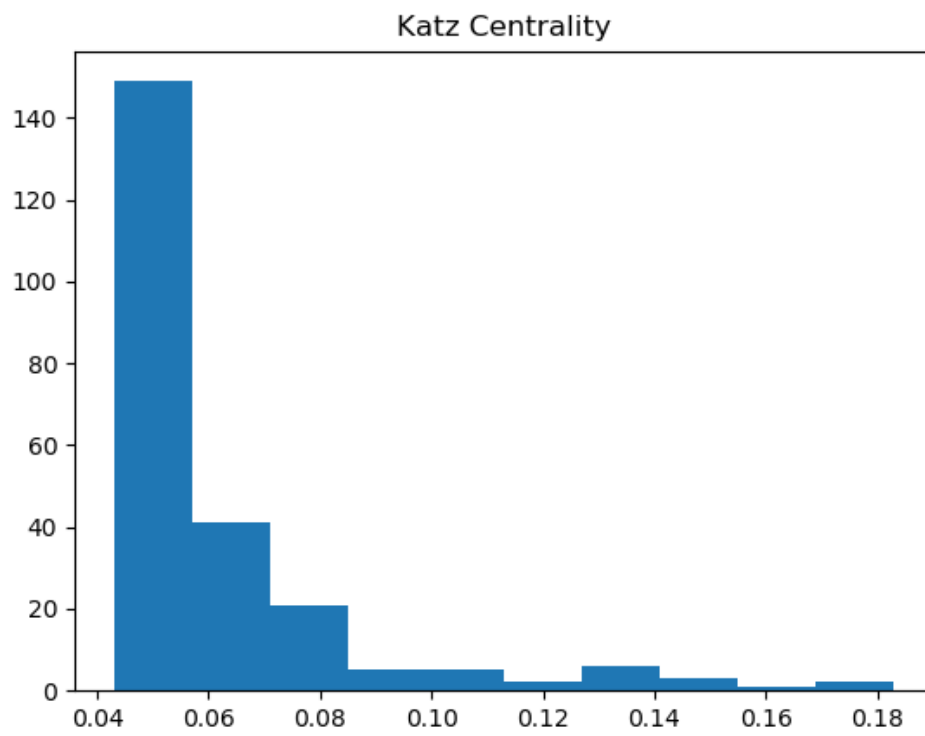
G. Katz Centrality



Εικόνα 21 - Katz Centrality για το 1 από τα 3 γραφήματα

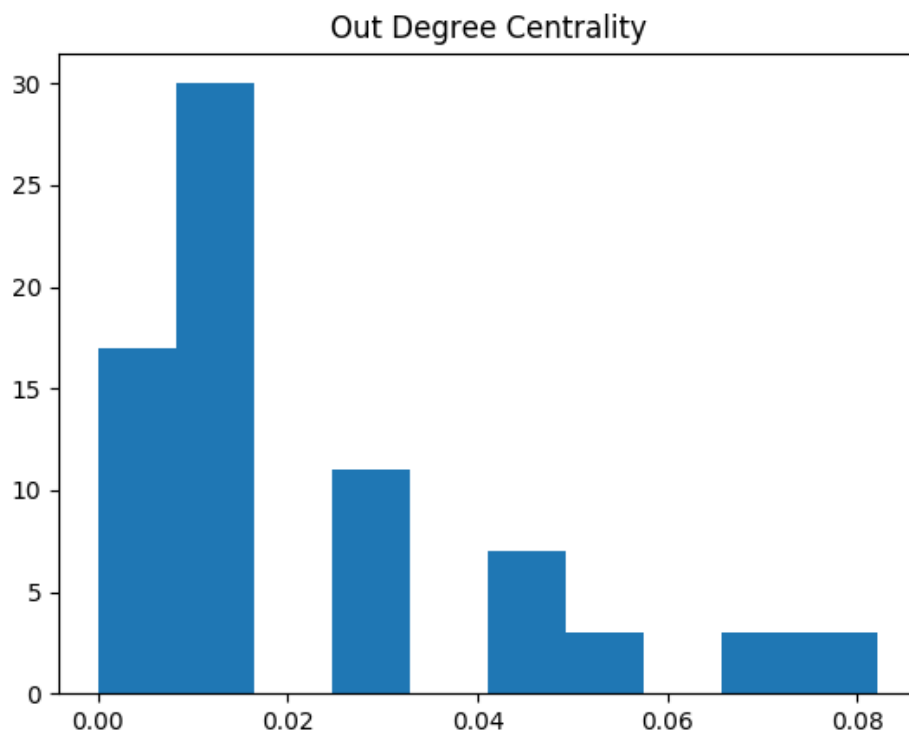


Εικόνα 22 - Katz Centrality για το 2 από τα 3 γραφήματα

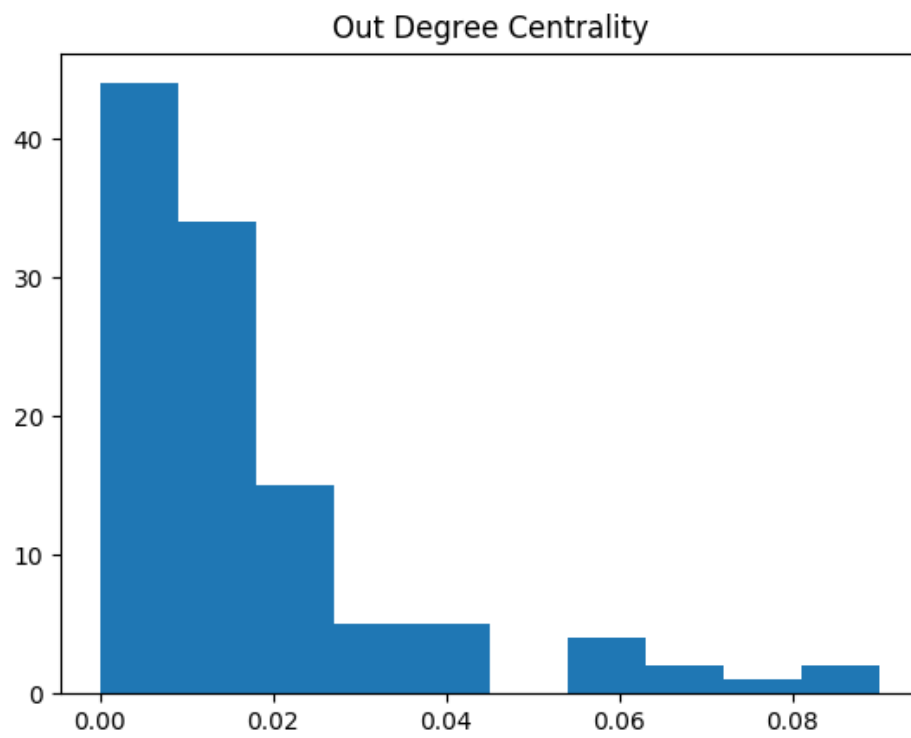


Εικόνα 23 - Katz Centrality για το 3 από τα 3 γραφήματα

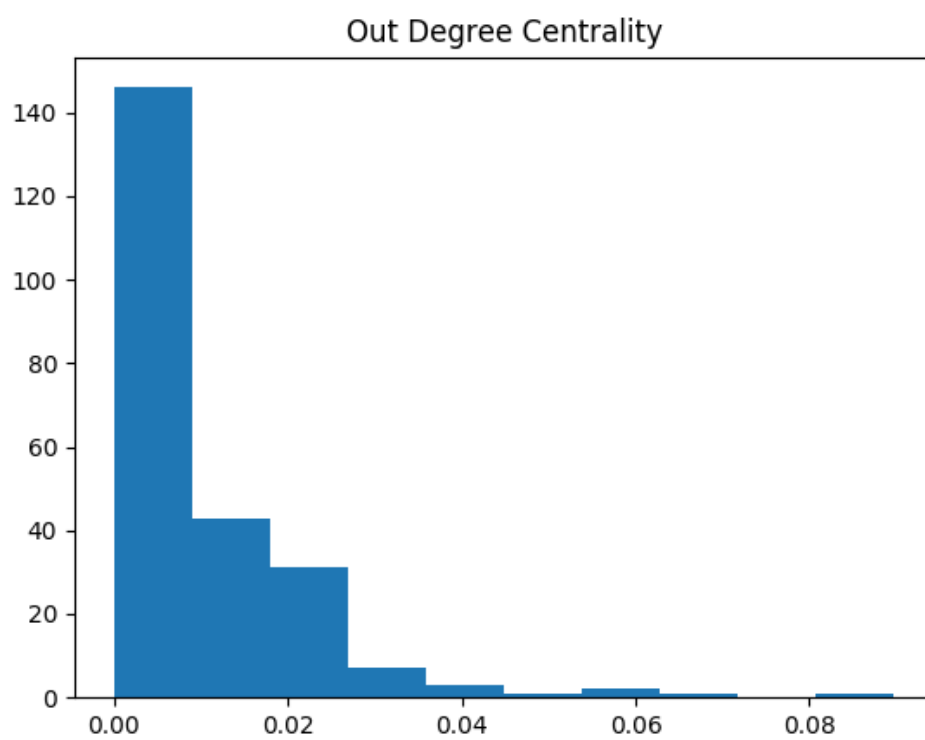
H. Out Degree Centrality



Εικόνα 24 - Out Degree Centrality για το 1 από τα 3 γραφήματα



Εικόνα 25 -4 Out Degree Centrality για το 2 από τα 3 γραφήματα



Εικόνα 26 -5 Out Degree Centrality για το 3 από τα 3 γραφήματα