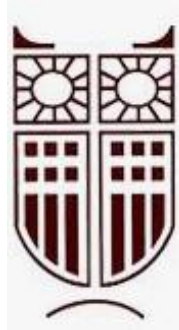


ΠΑΝΤΕΙΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΟΙΝΩΝΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES



ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ & ΠΕΡΙΦΕΡΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ «ΕΦΑΡΜΟΣΜΕΝΑ
ΟΙΚΟΝΟΜΙΚΑ ΚΑΙ ΔΙΟΙΚΗΣΗ»

Geolocated IP Addresses as a Source of Economic Data

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αθανάσιος – Γεώργιος Σχοινάς

Αθήνα 2022

Τριμελής Επιτροπή

Επίκουρος Καθηγητής: Γρηγόριος Σιουρούνης (Επιβλέπων)

Ομότιμος Καθηγητής: Σαράντης Λώλος

Αναπληρωτής Καθηγητής: Σταύρος Ντεγιαννάκης

Copyright ©Αθανάσιος Γεώργιος. Σχοινάς, 2022

All rights reserved. Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας διπλωματικής εργασίας εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της διπλωματικής εργασίας για κερδοσκοπικό σκοπό, πρέπει να απευθύνονται προς τον συγγραφέα.

Η έγκριση της διπλωματικής εργασίας από τον Πάντειον Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών, δεν δηλώνει αποδοχή των γνώμων του συγγραφέα.

*Η παρούσα διπλωματική εργασία αφιερώνεται στους γονείς μου και σε όσους
μου στάθηκαν και με ενθάρρυναν για την εκπόνηση της.*

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Γρηγόριο Σιουρούνη για την αμέριστη συμπαράσταση του, καθώς και στην εμπιστοσύνη που επέδειξε προς το πρόσωπο μου για την εκπόνηση της παρούσας διπλωματικής εργασίας.

Περίληψη

Το Διαδίκτυο (Internet) στις μέρες μας, διαδραματίζει έναν κυρίαρχο ρόλο, καθώς ένα μεγάλο μέρος της καθημερινότητας μας και του χρόνου μας, αξιοποιείται σε αυτό, είτε μέσω χρήσης ηλεκτρονικών υπολογιστών, είτε μέσω κινητών τηλεφώνων (smartphones), και άλλων διαφόρων ειδών ηλεκτρονικών συσκευών. Ο Γεωεντοπισμός IP (Geolocation), η ταυτοποίηση δηλαδή της τοποθεσίας μιας σύνδεσης-χρήστη μέσω της IP διεύθυνσης του, είναι ένας νέος τρόπος παρατήρησης και καταγραφής δεδομένων με σκοπό την αξιοποίηση τους σε διάφορους τομείς της κοινωνίας (οικονομία, υγεία, εκπαίδευση κ.α.) για την εξαγωγή χρήσιμων συμπερασμάτων. Στην παρούσα εργασία παρουσιάζεται ένα δείγμα του ερευνητικού έργου που έχει πραγματοποιηθεί πάνω σε αυτή τη νέα μέθοδο και μελετάται κατά πόσο αξιόπιστα δεδομένα παρέχει, ώστε στη συνέχεια να αξιοποιηθούν για περαιτέρω διερεύνηση. Στη μελέτη, παρατηρείται ότι ο Γεωεντοπισμός IP αποτελεί ένα σπουδαίο μέσο άντλησης δεδομένων οικονομικού χαρακτήρα, το οποίο όμως θα πρέπει να χρησιμοποιείται ως συμπληρωματικό μέσο πηγών δεδομένων.

Summary

Nowadays the Internet has played a crucial role in our daily routine, with the majority of people spending a lot of time on it through different electronic devices (personal computers, laptops, smart phones e.t.c). IP Geolocation is a new technique using people's IP address in order to find the exact location of them, in order to extract important information, useful for the study of the economy, health, education e.t.c. In this project it's examined a sample of the current research of IP Geolocation among different parts of a society such as the education on the Massive Open Online Courses (MOOC) and is examined the accuracy of this method as a source of extracting useful information. In this work, a sample of the research project that has been carried out on this new method is presented and it is studied how reliable data it provides, so that it can then be used for further investigation. In this project, it is observed that IP Geolocation is an important means of extracting data of an economic nature, which should, however, be used as a complementary means of data sources.

Λέξεις- Κλειδιά

Διαδίκτυο, IP Γεωεντοπισμός, Βάσεις Δεδομένων IP, Διαδικτυακή Εκπαίδευση, Αξιοπιστία Γεωεντοπισμού IP, GNSS, GPS.

Περιεχόμενα

Περίληψη	5
Summary	5
Κατάλογος Γραφημάτων	8
Κατάλογος Εικόνων.....	9
Εισαγωγή	10
Κεφάλαιο Πρώτο: Το Διαδίκτυο και ο Γεωεντοπισμός.....	11
1.1 Ορισμός IP	11
1.2 Ο IP Γεωεντοπισμός	13
1.3 Στιγμιότυπα τοποθεσίας IP	15
Κεφάλαιο Δεύτερο: Βάσεις Δεδομένων Γεωεντοπισμού και Αξιοπιστία	16
2.1 Οι Βάσεις Δεδομένων των IP	16
2.2 Η Βάση Γεωεντοπισμού της MaxMind	17
2.3 Η Ανάλυση της Διαδικασίας Εντοπισμού IP	18
2.4 Η Ακρίβεια Γεωεντοπισμού IP Διευθύνσεων	19
2.5 Η Βάση Δεδομένων της Google Maps Geocoding Application Programming Interface	20
2.6 Υπολογισμός των Δεδομένων.....	20
2.7 Αποτελέσματα Μέτρησης Στιγμιότυπων.....	22
Κεφάλαιο Τρίτο: Ο IP Γεωεντοπισμός στη Διαδικτυακή Εκπαίδευση	23
3.1 IP και Τηλεκπαίδευση.....	23
3.2 Μέθοδος Ανάλυσης Μελέτης.	25
3.3 Η Διαδικτυακή Εκπαίδευση (ΔΕ)	26
3.4 Γεωγραφικοί Δείκτες Κοινωνικοοικονομικής Κατάστασης.....	29
3.5 Συλλογή των Δεδομένων	29
3.6 Δείκτης Κοινωνικής Εξουθένωσης Distressed Community Index	31
3.7 Απόδοση των Μετρήσεων	31
3.8 Προσέγγιση Επεξεργασίας των Δεδομένων	32
3.9 Αποτελέσματα των Δεδομένων	34
3.10 Οικονομική Συσχέτιση των Δεδομένων	36
3.11 Παρατηρήσεις Ευρημάτων Ανάλυσης.....	38
3.12 Εξέταση Ακρίβειας των Μετρήσεων του Geolocation.....	39
3.13 Προσέγγιση Διαδικασίας Γεωεντοπισμού	41
3.14 Υπολογισμός Σφάλματος Γεωεντοπισμού.....	42
3.15 Ανάλυση Παρατηρήσεων Οικονομικού Χαρακτήρα.....	42

3.16 Αξιολόγηση Μετρήσεων Αξιοπιστίας	42
3.17 Συνεκτίμηση των DCI δεικτών	45
3.18 Σύνοψη των Παρατηρήσεων.....	46
Κεφάλαιο Τέταρτο: Ο Γεωεντοπισμός IP βασισμένο σε συστήματα GNSS (Global Navigation Surveillance Systems)	47
4.1 IP Γεωεντοπισμός και τα GNSS	47
4.2 Τα Συστήματα GNSS.....	48
4.3 Το GPS-based IP Geolocation	48
4.4 Η Βάση Δεδομένων Unacast.....	48
4.5 Η Βάση Ookla Speedtest.....	49
4.6 Δεδομένα Σύγκρισης IP	49
4.7 Αποτελέσματα των Μετρήσεων	50
4.8 Επίδραση Γεωεντοπισμού σε Επίπεδο Καταναλωτή.....	53
Κεφάλαιο Πέμπτο	55
5.1 Επίδραση IP Γεωεντοπισμού στο Διαδίκτυο	55
5.2 Πιθανοί Τρόποι Βελτίωσης Αξιοπιστίας	57
5.3 Ανασκόπηση της Εργασίας.....	57
5.4 Συμπεράσματα	58
ΒΙΒΛΙΟΓΡΑΦΙΑ	59

Κατάλογος Γραφημάτων

Γράφημα 1. Ιστόγραμμα κατανομής συμμετοχών σε ΔΕ των πανεπιστημίων Harvard / MITx ανά εκατομμύριο κατοίκων. © Ganelin κ.α.,2019.	34
Γράφημα 2. Ιστόγραμμα μαθημάτων από τις καταχωρήσεις TK στα 1,1 εκατομμύρια δεδομένων. © Ganelin κ.α.,2019.	35
Γράφημα 3. Μπλοκς από TK και κατά κεφαλήν συμμετοχές σε ΔΕ. Βαθμίδες από 0 (εξουθενωμένοι οικονομικά) έως 4 (ευημερείς). Για κάθε βαθμίδα το μπλοκ απεικονίζει το 25%, 50% και 75% των δεδομένων. Οι αγκύλες απεικονίζουν τις τιμές εντός εμβέλειας, ενώ οι κουκίδες τις γραμμές εκτός εμβέλειας. © Ganelin κ.α.,2019.	36
Γράφημα 4. Απεικόνιση των eigenvalues μετά από εφαρμογή μεθόδου ανάλυσης παραγόντων για όλα τα δεδομένα συνολικά. Επικρατεί ένας παράγοντας. © Ganelin. κ.α.,2019.	37
Γράφημα 5. Παράγοντας φορτίου και του λογάριθμου των συμμετοχών σε ΔΕ. © Ganelin κ.α.,2019.....	38
Γράφημα 6. Ιστόγραμμα απεικόνισης των κατανομών δεδομένων χρηστών ανά κατηγορία DCI, της MaxMind και πραγματικών τοποθεσιών. © Ganelin κ.α. 2019.	46
Γράφημα 7. Διαφορά απόστασης της ενδιάμεσης τιμής των υποδικτύων στις σταθερές συνδέσεις στην πάροδο δύο μηνών (Αυγούστου-Οκτωβρίου 2020). © Saxons κ.α.,2021.	53
Γράφημα 8. Ποσοστό Διατήρησης IP διεύθυνσης συσκευής σε σταθερή σύνδεση. © Saxon κ.α.,2021.	54
Γράφημα 9. . Ποσοστό μεταβολής του λογαρίθμου ενδιάμεσης τιμής νοικοκυριών της MaxMind σε σχέση με πραγματικά δεδομένα απογραφής. © Saxons κ.α.,2021.	55

Κατάλογος Εικόνων

Εικόνα 1. Χάρτης κατανομής συμμετοχών σε ΔΕ από τη MaxMind την περίοδο 2012-2018. © Ganelin κ.α.,2019.....	34
Εικόνα 2.Γραφήματα σύγκρισης του αριθμού χρηστών σε ΔΕ ανά ΤΚ με βάση την MaxMind και τις πραγματικές τοποθεσίες, ταξινομημένα ανάλογα το οικονομικό τους προφίλ, την περιοχή και την πυκνότητα πληθυσμού. © Ganelin κ.α.,2019.	43
Εικόνα 3.Γραφήματα σφαλμάτων εντοπισμού της MaxMind ερμηνευμένο σε λογαριθμική κλίμακα. © Ganelin κ.α.2019.	44
Εικόνα 4. Σφάλμα εντοπισμού τοποθεσίας στην πόλη του Σικάγο της Unacast και Ookla σε σχέση με τα αντίστοιχα δεδομένα από IP2LOCATION και MaxMind. ©Saxon κ.α, 2021.	50
Εικόνα 5. Αθροιστική Συνάρτηση Κατανομής σφάλματος γεωεντοπισμού ανά πόλη και πάροχο δικτύου με βάση τις βάσεις MaxMind IP2Location. © Saxon κ.α, 2021.	51

Εισαγωγή

Το διαδίκτυο (Internet), είναι ένα παγκόσμιο δίκτυο υπολογιστών που διασυνδέει υπολογιστές-τερματικά (κινητά τηλέφωνα, tablets laptops κ.α.) σε όλο τον κόσμο. Παρατηρώντας κάποιος το διαδίκτυο και συνειδητοποιώντας το μέγεθος του, γεννιέται εύλογα το ερώτημα και εγείρεται ο προβληματισμός του συντονισμού και καθορισμού του τρόπου επικοινωνίας όλων αυτών των συσκευών στο διαδίκτυο. Η απάντηση στον προβληματισμό αυτόν είναι τα πρωτόκολλα επικοινωνίας τα οποία έχουν καθοριστεί και διευθετούν τον τρόπο με τον οποίο οι διάφοροι χρήστες μπορούν να ανταλλάσσουν πακέτα δεδομένων με σκοπό την επικοινωνία τους. Ένα από αυτά τα πρωτόκολλα και πολύ βασικό είναι το Internet Protocol (IP). Το πρωτόκολλο IP καθορίζει τη μορφή των πακέτων που στέλνονται και λαμβάνονται μεταξύ των τερματικών μηχανημάτων. Ως βασική αρχή για την επίτευξη της παραπάνω ζεύξης χρησιμοποιεί διευθύνσεις, τις λεγόμενες IP διευθύνσεις (Kurose,2008).

Συνεπώς, από τη φύση του το διαδίκτυο αποδεικνύει πόσο πολύπλοκο είναι αλλά παράλληλα και πόσο χρήσιμο εργαλείο για το κάθε άτομο ξεχωριστά. Καθημερινά μέσω του διαδικτύου ανταλλάσσονται εκατομμύρια πακέτα δεδομένων και τεράστιοι όγκοι πληροφοριών, οι οποίοι έχουν διευκολύνει την καθημερινότητα μας και έχουν εξοικονομήσει πολύ χρόνο και χρήμα.

Μια από τις πληροφορίες που μπορούν να αντληθούν από το διαδίκτυο, είναι και η τοποθεσία του κάθε χρήστη ξεχωριστά. Μέσα από τη διεύθυνση IP, δύναται να εντοπιστεί η τοποθεσία του κάθε χρήστη. Ως εκ τούτου διάφοροι οργανισμοί και εταιρίες, μπορούν να εξάγουν χρήσιμα συμπεράσματα και να διαμορφώσουν ένα προφίλ του χρήστη, ανάλογα με τις προτιμήσεις του και τις τάσεις του κατά την περιήγηση στο διαδίκτυο. Με αυτόν τον τρόπο είναι δυνατό να σχηματίσει μια εικόνα της κοινωνικής και οικονομικής κατάστασης για την κάθε περιοχή ξεχωριστά. Το γεγονός αυτό οδηγεί στη δημιουργία ενός ολόκληρου κλάδου που επικεντρώνεται στον εντοπισμό της τοποθεσίας IP διευθύνσεων.

Μελετώντας όμως την ανωτέρω διαδικασία, παρατηρείται ότι δημιουργούνται ερωτήματα και προβληματισμοί στο κατά πόσον αξιόπιστα μπορεί να είναι οι πληροφορίες τοποθεσίας που παρέχονται από τους διάφορους οργανισμούς. Συναφώς, το μεγαλύτερο ποσοστό της ερευνητικής κοινότητας, τείνει να εστιάζει περισσότερο στην αξιοπιστία των παρεχόμενων δεδομένων και όχι τόσο στην πληροφορία που αυτά

παρέχουν. Στην παρούσα εργασία περιγράφονται κάποιες από τις έρευνες που έχουν γίνει ως προς την εξέταση εγκυρότητας των δεδομένων τοποθεσίας IP, με βάση το υφιστάμενο μελετητικό έργο και περιγράφονται πιθανοί τρόποι βελτίωσης της αξιοπιστίας της τοποθεσίας, με σκοπό τη βέλτιστη χρήση των πληροφοριών.

Κεφάλαιο Πρώτο: Το Διαδίκτυο και ο Γεωεντοπισμός

1.1 Ορισμός IP

Η διεύθυνση IP αποτελεί μια αριθμητική ετικέτα (label) η οποία εκπροσωπεί μια συσκευή η οποία είναι συνδεδεμένη στο διαδίκτυο. Σκοπός της IP είναι η ταυτοποίηση του εξοπλισμού δικτύου (συσκευής) και ο εντοπισμός της ως προς την τοποθεσία της. Υπάρχουν δύο ειδών διευθύνσεις IP: οι IPv4 version, οι οποίες αποτελούνται από συνολικά 32 bit ψηφίων για να παρουσιάσουν μια διεύθυνση και οι IPv6, οι οποίες χρησιμοποιούν 64 bits ψηφίων για να αναπαραστήσουν μια διεύθυνση. Ως bit νοείται η στοιχειώδης μονάδα μέτρησης στην επιστήμη των υπολογιστών και αναπαρίσταται από το δυαδικό σύστημα 0 και 1. Συνεπώς, μια διεύθυνση IPv4 είναι της μορφής «192.15.8.64», ενώ μια διεύθυνση IPv6 είναι της μορφής: «2a02:1388:208d:b2ff:58ff:c30e:e773:94cd». Στο άμεσο μέλλον το σύνολο των διευθύνσεων θα αποτελείται ως επί το πλείστον από διευθύνσεις IPv6, λόγω κορεσμού του αριθμού των συσκευών που μπορούν να συνδεθούν και να περιηγηθούν στο διαδίκτυο, καθώς οι διευθύνσεις IPv6 διευρύνουν επιπρόσθετα τη διαθεσιμότητα διευθύνσεων, γεγονός που οφείλεται στο μεγαλύτερο μέγεθος της ετικέτας τους.

Η διεύθυνση IP ανατίθεται σε μια συσκευή από τον δρομολογητή (router) του οικιακού ή δημόσιου δικτύου και αποσκοπεί στην ταυτοποίηση της συσκευής κατά την περιήγηση της στο διαδίκτυο. Σε ένα τοπικό δίκτυο ο δρομολογητής προσδίδει προσωπικές διευθύνσεις IP στις εκάστοτε συσκευές. Στη συνέχεια ο πάροχος του δικτύου προσδίδει στο δρομολογητή μια δημόσια διεύθυνση IP η οποία χρησιμεύει για την σύνδεση του δρομολογητή και κατ' επέκταση της οικίας ή του χώρου που βρίσκεται ο δρομολογητής στο διαδίκτυο.

Όσον αφορά την απόδοση των διευθύνσεων IP σε παγκόσμιο επίπεδο, έχει οριστεί ως υπεύθυνη αρχή η «Internet Assigned Numbers Authority» (IANA) καθώς επίσης και οι 5 «Regional Internet Registries» (RIRs) οι οποίοι είναι υπεύθυνοι για την απόδοση διευθύνσεων σε τοπικό επίπεδο, όπως είναι οι πάροχοι δικτύου (Internet Service Providers). Ενδεικτικά οι IPv4 διευθύνσεις διανέμονται από την IANA στα

RIRs σε 5 πακέτα των 16,8 εκ. διευθύνσεων ανά πακέτο, οι οποίες έχουν εξαντληθεί από το 2011 πλην όμως κάποιων περιοχών όπως στην Αφρική όπου υπάρχουν ελεύθερες ετικέτες μέχρι σήμερα (IPv4 Address Report, 2022).

Ουσιαστικά η διεύθυνση IP διασφαλίζει την αναγνώριση ενός χρήστη-συσσκευής μέσα στο διαδίκτυο με σκοπό τη δημιουργία επαφής με έναν άλλο χρήστη-συσσκευή για τη μετέπειτα ανταλλαγή δεδομένων. Συνεπώς κάθε ιστοσελίδα που αναζητάται στο διαδίκτυο, αντιστοιχίζεται σε μια IP διεύθυνση. Για παράδειγμα, σε μια περιήγηση (web browsing), το όνομα χρησιμεύει για να ταυτοποιηθεί το αντικείμενο που αναζητείται, ενώ η διεύθυνση απευθύνεται στο που βρίσκεται το αντικείμενο. Η διαδρομή (route) του πακέτου δεδομένων καθορίζει το πώς θα επιτευχθεί η ζεύξη. Συνεπώς η επικεφαλίδα κάθε διεύθυνσης IP πακέτου, εμπεριέχει τη διεύθυνση του αποστολέα και τον προορισμό του πακέτου (Internet Protocol, 1981).

Υπάρχουν IP διευθύνσεις, οι οποίες αποδίδονται σε ένα διακομιστή είτε μόνιμα ανεξαρτήτως του πόσες φορές κάποιος συνδέεται στο δίκτυο (static IP address), είτε δυναμικά όπου στην ουσία κάθε φορά που συνδέεται στο δίκτυο, του αποδίδεται εκ νέου μια διεύθυνση (Dynamic IP address). Το πρωτόκολλο στο οποίο στηρίζεται η δυναμική διευθυνσιοδότηση ονομάζεται Dynamic Host Configuration Protocol (DHCP) και περιγράφει επί της ουσίας τη διαδικασία απόδοσης δυναμικών διευθύνσεων IP. Η πιο διαδεδομένη μορφή διευθυνσιοδότησης είναι η δυναμική, καθώς με αυτόν τον τρόπο αποφεύγεται ο επιπλέον χρόνος επεξεργασίας και ελέγχου ενός χρήστη ως προς την αρχική του διεύθυνση, ενώ παράλληλα αποφορτίζεται το δίκτυο από τις επιπρόσθετες διαδικασίες ανάθεσης στατικών διευθύνσεων στους χρήστες. Επιπλέον, με το DHCP πολλοί χρήστες μπορούν να έχουν την ίδια διεύθυνση IP όταν δεν είναι συνδεδεμένοι ταυτόχρονα στο διαδίκτυο, εξοικονομώντας με αυτόν τον τρόπο πόρους του δικτύου και βελτιώνοντας θέματα χωρητικότητας.

Επιπρόσθετα, σε περιπτώσεις εξοπλισμού και συσκευών που είναι υπεύθυνα για τη διαμόρφωση της δομής του δικτύου, χρησιμοποιούνται τεχνικές στατικής απόδοσης διευθύνσεων IP, καθώς αποτελούν εξοπλισμό διαμόρφωσης του δικτύου και η διαδικασία αυτή διευκολύνει τη δομή. Όμως, κάποιες φορές εμφανίζονται προβλήματα διπλής διευθυνσιοδότησης, όπου μια διεύθυνση IP αντιστοιχεί σε δύο ή και παραπάνω συσκευές. Σε αυτήν την περίπτωση διακόπτεται αυτόματα η λειτουργία

στους εμπλεκόμενους χρήστες και αφού ενημερωθεί ο πάροχος, δίδεται αυτόματα εκ νέου νέα διευθυνσιοδότηση (IP Address Conflicts, 2022).

1.2 Ο IP Γεωεντοπισμός

Η πιο ευρέως διαδεδομένη μορφή εύρεσης της τοποθεσίας της IP (IP geolocation), αποτελείται από τη δημιουργία μιας βάσης δεδομένων ώστε να αντιστοιχίσουν τη σύνδεση της διεύθυνσης IP ενός χρήστη με την πραγματική του τοποθεσία. Πολλές από αυτές τις βάσεις δεδομένων χρησιμοποιούνται από εταιρίες και ιστοσελίδες προσπαθώντας να εξάγουν διάφορα συμπεράσματα. Σε πολλές περιπτώσεις όμως η προσπάθεια εντοπισμού μιας IP μπορεί να οδηγήσει σε εσφαλμένα συμπεράσματα. Παρακάτω γίνεται ανάλυση της διαδικασίας εντοπισμού IP και της δυνατότητας ερμηνείας τυχόν αποτελεσμάτων για την παρατήρηση διαφόρων κοινωνικών και οικονομικών δεδομένων.

Για την υλοποίηση των ανωτέρω, γίνεται η παράθεση και μελέτη των στοιχείων που μπορούν να αντληθούν από τα δεδομένα των IP. Ένας κλάδος ο οποίος αρχικά μελετάται είναι η τηλεκπαίδευση και ο ουσιώδης ρόλος που παίζει στις μέρες μας η εξ' αποστάσεως εκπαίδευση. Οι τηλεκπαιδευτές, χωρίς κάποια οικονομική εστίαση, χρησιμοποιούν κάλλιστα τη μέθοδο εντοπισμού θέσης μέσω της IP, με σκοπό τη χρήση για άντληση δεδομένων σε επίπεδο ηπείρου (Kizilcec & Halawa, 2015), χώρας (Martinez, 2014,) και πόλης (Diver and Martinez, 2015). Επιπρόσθετα πολλοί συγγραφείς δεν αναφέρονται στην ακριβή μέθοδο που χρησιμοποιούν καθώς δεν είναι πάντοτε αξιόπιστη, εγείροντας αμφιβολίες και προβληματισμούς.

Η πληροφορία της τοποθεσίας της IP είναι πολύ σημαντική σε πολλά ερευνητικά και επιστημονικά πεδία. Αρχικά βελτιώνει την κατανόηση της διασποράς της χρήσης το δικτύου Internet και των χαρακτηριστικών του, βοηθώντας επίσης και στην εξερεύνηση πληροφοριών όπου χωρίς αυτήν την μέθοδο θα ήταν αδύνατο να ανακαλυφθούν. Ένα τέτοιο παράδειγμα θεωρείται η επίδραση της τοποθεσίας των εξυπηρετητών (servers) των παρόχων δικτύου στην αποτελεσματικότητα και αποδοτικότητα των παρεχόμενων πληροφοριών και υπηρεσιών.

Ο IP Γεωεντοπισμός, είναι μια διαδικασία προσπάθειας εύρεσης της ακριβούς τοποθεσίας του χρήστη με βάση τη διεύθυνση IP του. Για τους ερευνητές είναι μια εφικτή και με χαμηλό κόστος διαδικασία από την οποία μπορεί κάποιος να εξάγει δεδομένα και συμπεράσματα αναφορικά με το περιβάλλον και τις επιθυμίες του

χρήστη. Εδώ αξίζει να σημειωθεί ότι παρόλο το χαμηλό κόστος της εν λόγω διαδικασίας, ενέχουν αρκετοί κίνδυνοι ως προς την αξιοπιστία των δεδομένων. Ο κυριότερος λόγος έγκειται στο γεγονός ότι σε πολλές περιπτώσεις οι πάροχοι δικτύου πραγματοποιούν αναδρομολόγηση των διευθύνσεων με αποτέλεσμα να δημιουργούνται σφάλματα ως προς την ακριβή τοποθεσία του κάθε χρήστη. Παρακάτω γίνεται αναφορά στα υπόψη προγράμματα και ιστοσελίδες οι οποίες υπάρχουν και παρέχουν υπηρεσίες γεωεντοπισμό των IP (Ganelin D,2019).

Τα τελευταία χρόνια, υπηρεσίες γεωεντοπισμού έχουν γίνει απαραίτητο και αναπόσπαστο κομμάτι σε πολλά και ευρύ πεδία της κοινωνίας και της οικονομίας. Αν και ο τελικός χρήστης (πελάτης) τις περισσότερες φορές δεν είναι ενήμερος, πολλές ιστοσελίδες χρησιμοποιούν τα δεδομένα γεωεντοπισμού των επισκεπτών τους, με σκοπό για παράδειγμα τη στοχευμένη διαφήμιση, την προβολή τοπικών ειδήσεων και τη συμμόρφωση των παρεχόμενων υπηρεσιών της εκάστοτε ιστοσελίδας στην κείμενη νομοθεσία της χώρας προβολής.

Ένα εξίσου σημαντικό και απαραίτητο στοιχείο που καθιστούν αναγκαίο το γεωεντοπισμό, είναι οι περιπτώσεις κυβερνοεπίθεσης και προσπάθειας «μόλυνσης» των ηλεκτρονικών υπολογιστών με κακόβουλο λογισμικό. Στις περιπτώσεις αυτές ο γεωεντοπισμός των διευθύνσεων που στοχεύουν στη μόλυνση και κακόβουλη επίθεση, διευκολύνει και βοηθά στην εύρεση περαιτέρω στοιχείων των επιτιθέμενων χρηστών. Εκτιμάται ότι οι στρατιωτικές και κυβερνητικές υπηρεσίες, χρησιμοποιούν τεχνικές γεωεντοπισμού για να ανιχνεύσουν την πηγή ύποπτων δραστηριοτήτων για τρομοκρατία μέσω διαδικτύου, καθώς επίσης και για να εντοπίσουν τα εχθρικά κέντρα επιχειρήσεων (Shavitt & Zilberman, 2011).

Ένα εξίσου ευρύ πεδίο στο οποίο ο γεωεντοπισμός βρίσκει εφαρμογή, είναι τα τραπεζικά και πληροφοριακά συστήματα. Στις περιπτώσεις όπου γίνεται χρήση προσωπικών κωδικών για τη μεταφορά χρηματικών ποσών, είναι πολύ πιθανό να επιδιωχθούν επιθέσεις από επιτήδειους «hackers», οι οποίοι προσπαθούν μέσω διαδικασιών «phishing» και άλλων μορφών απάτης, να αποσπάσουν δεδομένα και να επιχειρήσουν να κλέψουν χρήματα ή να πραγματοποιήσουν μεταφορά μετοχών, να κάνουν αγορές μέσω διαδικτύου και άλλες παράνομες δραστηριότητες. Σε αυτές τις περιπτώσεις ο γεωεντοπισμός μπορεί να χρησιμεύσει δίνοντας δεδομένα από περιοχές που εμφανίζονται υψηλά ποσοστά «phishing» και επίθεσης με κακόβουλο λογισμικό.

Τα δεδομένα αυτά μπορούν να αποδοθούν στις τράπεζες οι οποίες στη συνέχεια, είτε μπλοκάρουν τις διευθύνσεις IP που προέρχονται από τις «επικίνδυνες» περιοχές, είτε διασταυρώνουν την τοποθεσία του χρήστη τη δεδομένη χρονική στιγμή, με την τοποθεσία που έχει δηλώσει στην τράπεζα και αυτόματα προχωρούν στο «μπλοκάρισμα» των συναλλαγών σε κάποια ύποπτη κίνηση.

Ο Μηχανισμός Internet Engineering Task Force (IETF) έχει παράλληλα αναπτύξει διαδικασίες για τον καθορισμό προτύπων για τις υπηρεσίες γεωεντοπισμού και εκτάκτων κλήσεων μέσω της ομάδας εργασίας IETF Geographic Privacy (GEOPRIV), με σκοπό την εφαρμογή πρωτοκόλλων στις διαδικασίες γεωεντοπισμού καθώς και την προστασία της ιδιωτικότητας στο γεωεντοπισμό (Shavitt & Zilberman, 2011).

Ένας εξίσου σημαντικός τομέας στον οποίο ο γεωεντοπισμός IP έχει παίζει σημαντικό ρόλο, είναι η επιστήμη. Όσον αφορά το διαδίκτυο, βελτιώνει την χαρτογράφηση και τη «δόμηση» του δικτύου και βοηθά στην εξαγωγή χρήσιμων συμπερασμάτων αναφορικά με τις επιπτώσεις που μπορεί για παράδειγμα να έχει η τοποθεσία ενός «server», ενός παρόχου ως προς τους αντίστοιχους «servers» άλλων παρόχων και τις εν δυνάμει αλληλεπιδράσεις αυτών.

1.3 Στιγμιότυπα τοποθεσίας IP

Καθορίζοντας τη φυσική τοποθεσία των χρηστών, επιτυγχάνεται η εξαγωγή χρήσιμων συμπερασμάτων αναφορικά με τη διαφήμιση τον καθορισμό του περιεχομένου που επιθυμούν οι χρήστες να δουν στην ιστοσελίδα, τη διαμόρφωση της γλώσσας καθώς και θέματα ασφαλείας και προστασίας της ιστοσελίδας από απειλές και κακόβουλα λογισμικά. Από την άλλη πλευρά το διαδίκτυο και η δομή του πολλές φορές αποτυγχάνει να βρει τη σωστή τοποθεσία ενός χρήστη καθώς μπορεί οι ίδιοι να μην επιθυμούν την προβολή των στοιχείων τους και με κατάλληλες τεχνικές και μέσα να «παραπλανούν» τα δεδομένα τους (Gouel M., Vermeulen, K., Fourmaux, O., Friedman T. & Beverly, 2021). Στην έρευνα που πραγματοποιήθηκε από τους (Gouel κ.ά., 2021), εξετάστηκαν οι IP διευθύνσεις, οι τοποθεσίες τους, τα χαρακτηριστικά τους καθώς επίσης και ο βαθμός στον οποίο μια στιγμιαία καταγραφή δεδομένων IP, επηρεάζει την εγκυρότητα των παρατηρήσεων. Πολλοί ερευνητές για την εξαγωγή συμπερασμάτων χρησιμοποιούν στιγμιότυπα από δεδομένα τοποθεσιών διευθύνσεων IP. Συνεπώς είναι πολύ πιθανό την ημέρα που έγινε η καταγραφή του στιγμιότυπου, με την ημέρα που

εξάχθηκαν τα συμπεράσματα να διαφέρουν, με κίνδυνο τη μη εγκυρότητα των αποτελεσμάτων. Αυτή η περίπτωση είναι πιο πιθανή να συμβεί σε κάποια μακροχρόνια έρευνα κατά την οποία είναι δύσκολη η συλλογή «συγχρονισμένων» δεδομένων. Από την έρευνα των (Gouel κ.ά., 2021) παρατηρήθηκε ότι τα στιγμιότυπα δεδομένων διευθύνσεων, αν και διαφέρουν σε διαφορετικές χρονικές στιγμές, σε καθημερινή βάση δεν παρουσιάζεται κάποια ριζική αλλαγή η οποία να επηρεάζει την εξαγωγή συμπερασμάτων και ευρημάτων.

Για παράδειγμα σε μια τρίμηνη καταγραφή δεδομένων διευθύνσεων IP, παρατηρήθηκε μεταβολή έως και 22% σε διευθύνσεις οι οποίες αλλάζουν τοποθεσία σε ακτίνα μέχρι και 40 χλμ, ενώ οι μεταβολές σε διευθύνσεις άνω των 40 χλμ ανέρχονται σε 18%. Ως εκ τούτου στις περισσότερες μελέτες που έχουν πραγματοποιηθεί, η αναφορά στις ημερομηνίες λήψης των δεδομένων εντοπισμού θέσης των διευθύνσεων IP παραλείπεται.

Κεφάλαιο Δεύτερο: Βάσεις Δεδομένων Γεωεντοπισμού και Αξιοπιστία

2.1 Οι Βάσεις Δεδομένων των IP

Σε πρώτη φάση για να γίνει αντιληπτή η κατανόηση των χαρακτηριστικών των βάσεων δεδομένων των IP διευθύνσεων, θα πρέπει να καθοριστούν κάποια στοιχεία όπως είναι ο ορισμός των μετρήσεων της επιμονής, της επικράτησης της κάλυψης και της απόστασης. Στη μελέτη των Gouel κ.α., 2021, παρατίθενται δεδομένα ανάλυσης δέκα ετών, αντλούμενα από την επικρατέστερη και πιο συχνά χρησιμοποιούμενη βάση δεδομένων, τη MaxMind. Συγκεκριμένα η μελέτη επικεντρώνεται στα εξής:

- Στην έρευνα που καθορίζει το βαθμό στον οποίο τα υπάρχοντα συστήματα και η βιβλιογραφία του διαδικτύου χρησιμοποιούν δεδομένα εντοπισμού τοποθεσίας των IP διευθύνσεων και πόσο άμεσα εξαρτώνται από αυτά.
- Στον καθορισμό των μετρήσεων ώστε να γίνει κατανοητή η σταθερότητα του διαδικτύου και η συμπεριφορά των βάσεων δεδομένων γεωεντοπισμού IP διευθύνσεων.
- Στη μελέτη των δεδομένων γεωεντοπισμού των IP διευθύνσεων και στην εξαγωγή κρίσιμων συμπερασμάτων.
- Στη μελέτη με πρότερη παρατήρηση των δεδομένων βασισμένων στην τοποθεσία των διευθύνσεων IP, στην οποία παρατηρείται ότι τα δεδομένα

διαφέρουν ανάλογα την εκάστοτε βάση δεδομένων διευθύνσεων που χρησιμοποιούνται.

- Στις συστάσεις για τη σημασία των δεδομένων τοποθεσίας των IP για μελλοντικές έρευνες.

2.2 Η Βάση Γεωεντοπισμού της MaxMind

Όσον αφορά τη MaxMind ως βάση δεδομένων, πρόκειται για μια εμπορική οντότητα που εξειδικεύεται στην εύρεση τοποθεσίας δεδομένων IP διευθύνσεων και παροχή αντίστοιχων άλλων υπηρεσιών. Η MaxMind παρέχει 2 βάσεις δεδομένων, τη GeoLite η οποία παρέχεται ως υπηρεσία δωρεάν και τη GeoIP η οποία απαιτεί συνδρομή. Οι περισσότερες ακαδημαϊκές μελέτες χρησιμοποιούν τη GeoLite. Η GeoLite είναι διαθέσιμη σε μορφή στιγμιότυπων (snapshots) τα οποία συνήθως ενημερώνονται σε εβδομαδιαία βάση και είναι διαθέσιμα για λήψη. Τα στιγμιότυπα αυτά περιλαμβάνουν διευθύνσεις IP με αντίστοιχα δεδομένα μεταξύ των οποίων είναι και η τοποθεσία τους. Συνεπώς, κάθε IP συνοδεύεται από την πόλη, τη χώρα, το γεωγραφικό μήκος και πλάτος του χρήστη καθώς και την ακρίβεια της τοποθεσίας της σε κλίμακα χιλιομέτρων (σφάλμα θέσης). Στην εργασία των (Gouel κ.ά., 2021), έγινε η μελέτη των διευθύνσεων τύπου IPv4, βασισμένη στη βάση δεδομένων GeoLite της MaxMind.

Για την εξαγωγή συμπερασμάτων πραγματοποιήθηκε αρχικά μια καταγραφή της βιβλιογραφίας στην οποία αναγράφεται και μνημονεύεται η MaxMind. Στη συνέχεια οι (Gouel κ.ά., 2021) διαχώρισαν τις περιπτώσεις όπου γίνεται απλά αναφορά στη βάση και στις περιπτώσεις όπου εξάγονταν συμπεράσματα. Από τις αναφορές που εξάγονταν συμπεράσματα προχώρησαν στο διαχωρισμό των περιπτώσεων IP δεδομένων με κριτήριο το εάν πρόκειται για τελικό χρήστη ή για κάποιον εξυπηρετητή (server) ή ακόμη και κάποιον δρομολογητή (router). Επισημαίνεται ότι η MaxMind είναι η δημοφιλέστερη βάση δεδομένων εντοπισμού τοποθεσίας IP διευθύνσεων. Παρόλα αυτά στις περισσότερες έρευνες που έχουν γίνει, τα δεδομένα είναι αρκετά «εύθραυστα» λόγω έλλειψης λήψεων στιγμιότυπων και μεγάλης χρονικής διαφοράς μεταξύ των δεδομένων που έχουν συλλεχθεί και των στιγμιότυπων από τη βάση δεδομένων.

Στην έρευνα των (Gouel κ.ά., 2021) έγινε αρχικά η λήψη διαφορετικών στιγμιότυπων από τη βάση δεδομένων με διευθύνσεις IP και συγκρίθηκε η διαφορά στη δυναμική των διαφορετικών στιγμιότυπων. Με αυτόν τον τρόπο επιχειρήθηκε μια

βαθύτερη κατανόηση της μεταβολής των δεδομένων των διευθύνσεων. Για τη διευκόλυνση των μετρήσεων έγινε η υπόθεση ότι οι διευθύνσεις εμπεριέχουν το πρόθεμα της τοποθεσίας του κάθε χρήστη. Επιπρόσθετα τα δεδομένα των στιγμιότυπων ήταν ομοιόμορφα κατανεμημένα στο χώρο με διαφορετικά στιγμιότυπα σε διαφορετικές χρονικές στιγμές.

2.3 Η Ανάλυση της Διαδικασίας Εντοπισμού IP

Σε αυτό το σημείο γίνεται κάποια ανάλυση της διαδικασίας εντοπισμού των διευθύνσεων IP. Αρχικά όπως προαναφέρθηκε, οποιαδήποτε συσκευή είτε τερματικός υπολογιστής, είτε server είτε δρομολογητής, για να έχει πρόσβαση στο διαδίκτυο και να μπορεί να αναγνωρίζεται από τις υπόλοιπες συσκευές, θα πρέπει να διαθέτει μια διεύθυνση IP. Ένα σύνολο διευθύνσεων IP (μπλοκ IP), είναι ένα πακέτο από διευθύνσεις IP, οι οποίες ανήκουν στον ίδιο πάροχο δικτύου (Internet Service Provider ISP). Συνεπώς το κάθε μπλοκ υποδηλώνει ότι οι διευθύνσεις που περιέχει προέρχονται θεωρητικά από την ίδια τοποθεσία. Το να προβλεφθεί ορθά η τοποθεσία ενός χρήστη με βάση την IP του είναι αρκετά περίπλοκη και δύσκολη διαδικασία. Πολλές φορές οι ίδιοι οι πάροχοι αναδιανέμουν τις IP αρκετά συχνά, ειδικά όταν η συσκευή αποσυνδέεται και επανασυνδέεται στο διαδίκτυο, με αποτέλεσμα την αδυναμία εύρεσης μιας έγκυρης τοποθεσίας. Πέραν όμως αυτού, υπάρχουν και άλλες τεχνικές που μπορούν να αλλοιώσουν τα δεδομένα και να οδηγήσουν σε εσφαλμένες πληροφορίες για την τοποθεσία μιας IP. Μερικές από αυτές είναι η χρήση εικονικών ιδιωτικών δικτύων, τα λεγόμενα Virtual Private Networks (VPN). Με τα VPNs παρακάμπτονται τα τείχη προστασίας (firewalls) των τερματικών και οι χρήστες συνδέονται στο διαδίκτυο μέσω άλλων server προστατεύοντας την αρχική τους θέση και προβάλλοντας στο διαδίκτυο μια «ψεύτικη» τοποθεσία (Muir κ.α. 2006).

Οι βάσεις δεδομένων που παρέχουν πληροφορίες σχετικά με την τοποθεσία μιας διεύθυνσης IP, στην ουσία συνδέουν ένα μπλοκ IP με τη φυσική τοποθεσία του όπως είναι η χώρα, η πόλη, ο TK ή οι εκτιμώμενες συντεταγμένες (γεωγραφικό μήκος και πλάτος). Σε κάθε περίπτωση όμως υπάρχει πιθανότητα εσφαλμένης πληροφορίας την οποία δεν μπορούν να διασταυρώσουν και να εντοπίσουν.

Δεδομένα για την τοποθεσία μιας IP ενός server, μπορούν για παράδειγμα να εξαχθούν και από το όνομα του Domain του Server (www.ddd.gr), ή από έναν χρήστη από την ώρα ζώνης ή τη γλώσσα που έχει επιλέξει κατά την είσοδο του σε κάποια

ιστοσελίδα. Σε αυτές τις περιπτώσεις υπάρχει εξίσου ο κίνδυνος εσφαλμένης πληροφόρησης ή παραπλάνησης. Οι περισσότερες βάσεις δεδομένων που ασχολούνται με τις τοποθεσίες των IP, αγοράζουν από τους παρόχους τις πληροφορίες των διευθύνσεων και στην περίπτωση που αυτό δεν είναι εφικτό, παραπλανούν τους παρόχους ώστε εκείνοι να τις αποκαλύψουν.

2.4 Η Ακρίβεια Γεωεντοπισμού IP Διευθύνσεων

Υπάρχουν αρκετές μελέτες που εξετάζουν κατά πόσο οι δεδομένες διευθύνσεις IP είναι έγκυρες ως προς τις πληροφορίες τοποθεσίας τους. Οι Shavitt και Zilberman το 2011, εξετάσανε ένα σύνολο διευθύνσεων των οποίων οι τοποθεσίες τους ήταν γνωστές και συγκρίνανε την εγκυρότητα των αποτελεσμάτων με τη μέθοδο γεωεντοπισμού. Από την έρευνα τους διαπιστώθηκαν σφάλματα στις τοποθεσίες των IP πολλών χιλιομέτρων, με αποτέλεσμα πολλές IP να απεικονίζονται σε εσφαλμένη χώρα. Συγκεκριμένα η βάση δεδομένων της MaxMind απέδωσε ως τοποθεσία της πλειοψηφίας των διευθύνσεων, την περιοχή των ΗΠΑ και πιο εξειδικευμένα της Ουάσιγκτον. Ως εκ τούτου και οι ίδιοι κατέληξαν πως η διαδικασία γεωεντοπισμού είναι μειωμένης αξιοπιστίας καθώς σε πολλές περιπτώσεις εξάγονται λανθασμένα συμπεράσματα (Shavitt κ.α., 2011).

Επιπρόσθετα οι Poesse κ.α. παίρνοντας ως δεδομένα τις φυσικές διευθύνσεις των IP τις οποίες είχαν προμηθευτεί από έναν ευρωπαϊκό πάροχο δικτύου, παρατήρησαν ότι η απόκλιση στην πραγματική τοποθεσία IP και στην διεύθυνση μέσω γεωεντοπισμού, είχε μια διαφορά της τάξεως μερικών εκατοντάδων, έως μερικών χιλιάδων χιλιομέτρων και σε κάθε περίπτωση η ακρίβεια μπορούσε να επιτευχθεί σε επίπεδο χώρας και όχι σε επίπεδο πόλης (Poesse κ.α., 2011).

Με παρόμοιο τρόπο οι Gueye και Uhlig το 2008 σύγκριναν βάσεις δεδομένων IP διευθύνσεων με βάση τη χρονική υστέρηση για τη διάδοση των δεδομένων στο διαδίκτυο. Από την έρευνα τους, διαπιστώθηκε ότι η ένδειξη φυσικής τοποθεσίας για ένα σύνολο διευθύνσεων IP εμφάνιζε αποκλίσεις της τάξεως των 500 χιλιομέτρων, με αποτέλεσμα η παρεχόμενη πληροφορία τοποθεσίας να είναι μειωμένης εμπιστοσύνης. Συνεπώς πρότειναν ως λύση, οι εκάστοτε βάσεις δεδομένων να αναγράφουν την πιθανότητα σφάλματος και την έλλειψη ακρίβειας τοποθεσίας προκειμένου να αποφευχθούν παρανοήσεις.

2.5 Η Βάση Δεδομένων της Google Maps Geocoding Application Programming Interface

Το Google Maps Geocoding Application Programming Interface (API), παρέχει τη δυνατότητα να αποδίδει αναλυτικές πληροφορίες και λεπτομέρειες για μια ελλιπή ταχυδρομική διεύθυνση. Με δεδομένα στην είσοδο, τα στοιχεία μιας οδού, μπορεί να ταυτοποιήσει τη διεύθυνση και να αποδώσει πληροφορίες όπως τον ΤΚ, το γεωγραφικό μήκος και πλάτος μιας τοποθεσίας, καθώς και πληροφορίες για την ακρίβεια του σφάλματος της διεύθυνσης. Παρόλα αυτά και όπως είναι φυσικό, η συγκεκριμένη εφαρμογή παρουσιάζει ορισμένες φορές σφάλματα στην ακρίβεια, με την αδυναμία απόδοσης ακριβούς τοποθεσίας και έγκυρης διεύθυνσης στο 100% των περιπτώσεων (Google Developers, 2021).

Επιπλέον, έχει παρατηρηθεί το γεγονός ότι σε οδούς με παρόμοια ονόματα, παρουσιάζονται σφάλματα ταυτοποίησης και επομένως η εφαρμογή αδυνατεί κάποιες φορές να αποδώσει έγκυρη διεύθυνση. Τέλος σε περιπτώσεις που έχουν δοθεί αριθμοί διαμερισμάτων, η εφαρμογή μπορεί να τα αναγνωρίσει ως αριθμοί οδού και να αποδώσει εσφαλμένη πληροφορία.

Σε γενικές γραμμές, στην πλειοψηφία των καταγραφών της, παρατηρείται έγκυρη απόδοση των αποτελεσμάτων, χαρακτηρίζοντας το Geocoding API ως έγκυρο εργαλείο ταυτοποίησης διευθύνσεων για τη συνέχιση διεξαγωγής της έρευνας.

2.6 Υπολογισμός των Δεδομένων

Στην εργασία των Gouel κ.ά., για την εύρεση της διαφοράς απόστασης σε κάθε στιγμιότυπο για την ίδια διεύθυνση IP, χρησιμοποιήθηκε ο μέσος όρος των λογαρίθμων. Συγκεκριμένα ο υπολογισμός της απόστασης των στιγμιότυπων υπολογίστηκε από τον ακόλουθο τύπο:

$$\text{distdiff}(M_i, M_j) = (\text{mean}(\{\log_{10}(\text{dist}(a, M_i, M_j)) \mid x \in AM_i \cup AM_j\})) \quad (1)$$

όπου, M_i, M_j αποτελούν τους πίνακες στιγμιότυπων σε διαφορετικές χρονικές στιγμές, ενώ το a αποτελεί τη διεύθυνση IP που εμφανίζεται και στα δύο στιγμιότυπα.

Υπολογίζοντας το μέσο των λογαρίθμων στην εξίσωση (1), μειώνεται η πιθανότητα ύπαρξης μεγάλων διαφορών στις τιμές απόστασης σε ακραίες τιμές (πχ στις τάξεις των 20,000 χλμ) και με αυτόν τον τρόπο δεν «υπερκαλύπτονται» οι συγκριτικά μικρότερες αποστάσεις (πχ της τάξης των 100 χλμ). Επισημαίνεται ότι

πάνω από το 50% των διευθύνσεων παρουσίασε μηδενική μεταβολή στην απόσταση, συνεπώς ο ενδιαμέσος παρείχε σημαντική πληροφορία για την μηδενική μεταβολή.

Στη συνέχεια και με σκοπό την εμβάθυνση στην εγκυρότητα των δεδομένων οι (Gouel κ.ά., 2021) προχώρησαν στην απόδοση κάποιων εννοιών και ορισμών όπως είναι η διαφορά κάλυψης μεταξύ των στιγμιότυπων, η επικράτηση και η ακρίβεια των τοποθεσιών και η διαφορά κατανομής των αποστάσεων μεταξύ των στιγμιότυπων. Συγκεκριμένα:

Για την κάλυψη της διαφοράς ενός συνόλου στιγμιότυπων γεωεντοπισμού ορίστηκε η Τζακαρτιανή (Jacard) απόσταση:

$$\text{Covdif}(M1, ..., Mn) = \frac{|U M_I - \cap M_I|}{|U M_I|} \quad (2)$$

Εάν για παράδειγμα η διαφορά κάλυψης είναι 0,5 και πάνω, μετά από την παρέλευση ενός έτους, αυτό σημαίνει ότι το 50% των IP διευθύνσεων δεν εμφανίζεται σε τουλάχιστον ένα στιγμιότυπο για αυτό το έτος. Συνεπώς όσο πιο μικρό είναι αυτό το μέγεθος, τόσο πιο σταθερά είναι τα δεδομένα των στιγμιότυπων και κατ' επέκταση και η τοποθεσία τους.

Όσον αφορά την επικράτηση μιας διεύθυνσης IP σε μια τοποθεσία, αυτή καθορίζεται με το ποσοστό στο οποίο σε διαφορετικές χρονικές στιγμές η διεύθυνση IP παρουσίαζε την ίδια τοποθεσία. Για παράδειγμα μπορεί σε 8 διαφορετικά στιγμιότυπα, μια συγκεκριμένη IP να εμφανίζεται σε 3 από τα 8 στιγμιότυπα με διαφορετική τοποθεσία. Συνεπώς η επικράτηση της διεύθυνσης αυτής είναι: 3/8. Συναφώς γίνεται εύκολα αντιληπτό ότι τιμές της επικράτησης σε ποσοστά 1/8 και 2/8 θεωρούνται χαμηλές, ενώ αναλογίες από 4/8 και πάνω συνεπάγονται υψηλές τιμές επικράτησης της τοποθεσίας για μια διεύθυνση IP (Gouel κ.α., 2021).

Η επιμονή μιας διεύθυνσης να διατηρήσει την τοποθεσία της, εκφράζεται με την πιθανότητα σε κάθε χρονική στιγμή η τοποθεσία της συγκεκριμένης διεύθυνσης IP να παραμένει σταθερή. Η τιμή της επιμονής κυμαίνεται μεταξύ 0 και 1, με το 0 να υποδηλώνει ότι η τοποθεσία μιας διεύθυνσης IP είναι διαφορετική σε κάποιο στιγμιότυπο και το 1 ότι η τοποθεσία της διεύθυνσης παραμένει σταθερή (Gouel κ.ά., 2021).

Η απόσταση, η επικράτηση και η επιμονή σαν μεγέθη, χαρακτηρίζουν τη δυναμική από ένα διακριτό σημείο των διευθύνσεων IP. Οι (Gouel κ.ά., 2021) στη συνέχεια εξέτασαν τη μέγιστη απόσταση μεταξύ των τοποθεσιών των δεδομένων διευθύνσεων IP. Με αυτόν εξάχθηκαν χρήσιμα συμπεράσματα όπως, για παράδειγμα το γεγονός ότι μπορεί να παρατηρηθεί η διαφορά στην απόσταση τοποθεσίας μιας συγκεκριμένης διεύθυνσης IP σε μια χώρα. Επιπλέον διαπιστώθηκε ότι ανάλογα τη χώρα που ανήκει η τοποθεσία της διεύθυνσης IP, παρουσιάζεται και αντίστοιχη διαφορά στην απόσταση της τοποθεσίας της (μικρότερη-μεγαλύτερη).

Από την έρευνα των (Gouel κ.ά., 2021) συλλέχθηκαν 214 στιγμιότυπα σε βάθος δεκαετίας από τον Ιανουάριο του 2010 μέχρι το Δεκέμβριο του 2019, με κάθε ένα στιγμιότυπο να έχει ληφθεί τουλάχιστον μια φορά το μήνα. Για την απλότητα των υπολογισμών έγινε κανονικοποίηση των αποτελεσμάτων. Επιπλέον στα πλαίσια απλοποίησης των δεδομένων δημιουργήθηκε ένα ιστόγραμμα με την κάθε διεύθυνση IP να αναγράφει σε κάθε στιγμιότυπο της, τη χώρα την πόλη και συντεταγμένες της.

2.7 Αποτελέσματα Μέτρησης Στιγμιότυπων

Από τις ανωτέρω μετρήσεις απόστασης παρατηρήθηκε ότι το 98% των διευθύνσεων IP εμφάνιζαν σταθερή τη χώρα προέλευσης. Επισημαίνεται ότι υπήρχαν αποκλίσεις σε χώρες οι οποίες δεν έχουν μεγάλη γεωγραφική έκταση όπως είναι το Κόσοβο και κάποια μικρά νησιά. Επιπρόσθετα για το Ηνωμένο Βασίλειο παρουσιάστηκε μικρότερη ακρίβεια στην τοποθεσία της τάξης του 8% των διευθύνσεων.

Από τα ανωτέρω συνεπάγεται ότι σχεδόν όλες οι διευθύνσεις IP είχαν μια μοναδική τοποθεσία όσον αφορά τη χώρα. Όμως όσον αφορά την ακριβή τοποθεσία μέσα στην ίδια τη χώρα, εκεί παρουσιάστηκαν διαφορές. Εάν χρησιμοποιηθεί η MaxMind ως βάση δεδομένων για τον γεωεντοπισμό θα πρέπει να γίνει με προσοχή, καθώς τα αποτελέσματα της δεν είναι πάντοτε έγκυρα για κάποιες χώρες, αφού μέσα στο χρόνο υπάρχει η πιθανότητα η τοποθεσία της χώρας, για μια διεύθυνση IP, να μεταβληθεί.

Σε επίπεδο πόλεων η μελέτη έδειξε μια ακρίβεια της τάξης του 38% με αποτέλεσμα τα δεδομένα να μην είναι αξιόπιστα για όλα τα είδη διευθύνσεων IP. Επισημαίνεται ότι κάποιες πόλεις εμφάνισαν μεγαλύτερη ακρίβεια, αλλά αυτό κατά πάσα πιθανότητα, οφείλεται στο γεγονός ότι βρίσκονται σε χώρες με ελλιπή δεδομένα (π.χ. Αφρική, Κορέα), με αποτέλεσμα να μην είναι απόλυτα έγκυρες οι μετρήσεις. Σε

αυτό το σημείο αξίζει να σημειωθεί ότι στις περιπτώσεις όπου η MaxMind διαθέτει ελλιπή δεδομένα για μια περιοχή, τότε υπάρχει ο κίνδυνος η τοποθεσία της IP να μην μεταβληθεί και τα δεδομένα να διαστρεβλωθούν.

Αναφορικά με την απόσταση των τοποθεσιών μιας ίδιας διεύθυνσης IP, συμπεραίνεται ότι τα πρώτα χρόνια παρατήρησης των δεδομένων υπήρχαν μεγαλύτερες διαφορές. Ενδεικτικά, το 2012 το 40% των IP παρουσίαζαν μεταβολή της μέγιστης απόστασης για πάνω από 1000km, ενώ το 2018 το 42% των IP διευθύνσεων είχαν μεταβολή στην απόσταση τους πάνω από 40 χλμ και το 2019, το ποσοστό έπεσε στο 22% (Gouel κ.ά., 2021). Συνεπώς προκύπτει ότι αν και στην πλειονότητα των διευθύνσεων η διαφορά απόστασης είναι μικρότερη από 40 χλμ, εμβαθύνοντας στη μελέτη εμφανίστηκε μεγάλη διασπορά των μετρήσεων των δεδομένων.

Από τη δεκαετή παρατήρηση των μεταβολών των τοποθεσιών των IP, είναι δύσκολο να εξαχθεί κάποια χαρακτηριστική τάση ώστε να βγουν ασφαλή συμπεράσματα για το ρυθμό μεταβολής της τοποθεσίας ορισμένων διευθύνσεων. Αναφορικά με την έννοια της κάλυψης, οι (Gouel κ.ά., 2021) χρησιμοποιώντας τους πίνακες της Border Gateway Protocol (BGP) από το Route Views, συνδέοντας τους ταυτόχρονα με τα αντίστοιχα στιγμιότυπα της MaxMind, παρατήρησαν ότι η συνολική κάλυψη διευθύνσεων IP είναι συνεχής στην πάροδο του χρόνου αλλά εξαρτάται από το είδος της IP. Συγκεκριμένα η κάλυψη των διευθύνσεων απλών χρηστών άγγιζε το 90%, ενώ το 37% ανήκε σε IP υποδομών (εξυπηρετητές) και το 41% σε δρομολογητές (routers). Συνεπώς μια διεύθυνση IP μπορεί να δέχεται μεγάλες μεταβολές στην τοποθεσία της κατά την πάροδο του χρόνου.

Σε κάθε περίπτωση η προσπάθεια εντοπισμού της θέσης ενός τερματικού υπολογιστή είναι απαραίτητη για την εξαγωγή χρήσιμων συμπερασμάτων. Παρόλα αυτά διαφαίνεται ότι η ακρίβεια των δεδομένων δεν μπορεί να θεωρηθεί υψηλή, καθότι σε βάθος χρόνου, παρατηρούνται μεταβολές στην κάλυψη την διατήρηση και την επικράτηση των διευθύνσεων IP.

Κεφάλαιο Τρίτο: Ο IP Γεωεντοπισμός στη Διαδικτυακή Εκπαίδευση

3.1 IP και Τηλεκπαίδευση

Στο κεφάλαιο αυτό επισημαίνεται και αξιολογείται η σημαντικότητα του Γεωεντοπισμού IP στον τομέα της εκπαίδευσης. Για το σκοπό αυτό γίνεται

ανασκόπηση στα υπάρχοντα μελετητικά έργα, εστιάζοντας περισσότερο στη μελέτη των Daniela Ganelin και Isaac Chuang, 2019, της οποίας τα ευρήματα θεωρούνται επίκαιρα και διαφωτιστικά όσον αφορά τον τομέα της εκπαίδευσης στον Γεωεντοπισμό IP.

Στην υπόψη μελέτη χρησιμοποιήθηκαν δεδομένα από τις πλατφόρμες των πανεπιστημίων Massachusetts Institute of Technology (MIT) και Harvard, με σκοπό την παροχή δωρεάν εκπαιδευτικών προγραμμάτων σε χρήστες από όλο τον κόσμο. Μέσω των δωρεάν διαδικτυακών προγραμμάτων εκπαίδευσης ανοίγει μια νέα εποχή στον τομέα της εκπαίδευσης, καθώς διευρύνονται οι ορίζοντες και σε άτομα με περιορισμένες οικονομικές δυνατότητες. Με αυτόν τον τρόπο δίδεται η ευκαιρία σε άτομα από όλο τον κόσμο να παρακολουθήσουν εκπαιδεύσεις και να διευρύνουν τους ορίζοντες τους χωρίς ιδιαίτερο κόστος, παρά μόνο την εξασφάλιση ύπαρξης σύνδεσης στο διαδίκτυο.

Συγκεκριμένα στις Ηνωμένες Πολιτείες, παρατηρείται μεγάλη διασπορά στο μορφωτικό επίπεδο ανά περιοχή και πολιτεία. Η ιδέα της παροχής εκπαιδευτικών προγραμμάτων μέσω διαδικτύου αποσκοπεί στην απόδοση κινήτρων στις χαμηλότερα σε μορφωτικό επίπεδο περιοχές, για να εκμεταλλευτούν την ευκαιρία που τους δίνεται και να προσπαθήσουν να βελτιώσουν την ποιότητα εκπαίδευσης τους χωρίς κάποιο επιπλέον κόστος. Συναφώς δίνεται η ευκαιρία στις περιοχές αυτές να αντιμετωπίζουν προβλήματα χαμηλής στελέχωσης και ανεργίας, λόγω έλλειψης εκπαίδευσης, πετυχαίνοντας με αυτόν τον τρόπο τη βελτίωση, τόσο της ποιότητας της εργασίας, όσο και της ποιότητας ζωής (Ganelin κ.α, 2019).

Αξίζει να σημειωθεί ότι η μελέτη των οικονομικών στοιχείων των συμμετεχόντων στα προγράμματα διαδικτυακής εκπαίδευσης είναι δύσκολο να πραγματοποιηθεί. Σε πολλές περιπτώσεις τα δεδομένα αντλούνται από τους ίδιους τους χρήστες οι οποίοι συμπληρώνουν την αίτηση συμμετοχής και περιγράφουν το ακαδημαϊκό τους υπόβαθρο και περιοχή στην οποία δραστηριοποιούνται. Συνεπώς η διαδικασία εντοπισμού των οικονομικών στοιχείων των ενδιαφερομένων που συλλέχθηκαν στην παρούσα εργασία δεν θεωρείται τόσο αξιόπιστη.

Στην μελέτη των Ganelin κ.α., για τον εντοπισμό των διευθύνσεων IP χρησιμοποιήθηκαν βάσεις δεδομένων με διευθύνσεις IP, οι οποίες επιχειρούν να εκτιμήσουν την πραγματική τοποθεσία του χρήστη. Ο εντοπισμός της IP μέσω αυτής

της διαδικασίας είναι μια γρήγορη και με άμεσα αποτελέσματα μέθοδος, η οποία αποδίδει στους ερευνητές τα απαιτούμενα συμπεράσματα με σχετικά χαμηλό κόστος. Η ίδια μέθοδος χρησιμοποιείται και σε άλλες περιπτώσεις όπου απαιτούνται να εξαχθούν συμπεράσματα ανάλογα με την περιοχή και το πεδίο εφαρμογής μιας τεχνικής σε επίπεδο συνοικίας, περιοχής, πόλης, ή ακόμη και χώρας.

Σε αυτό το σημείο αξίζει να αναφερθεί το γεγονός της αξιοπιστίας των πληροφοριών αυτών. Πολλές φορές οι διευθύνσεις IP δεν ανταποκρίνονται στις πραγματικές διευθύνσεις των χρηστών. Για παράδειγμα πολλές φορές οι ίδιοι πάροχοι διαδικτύου τροποποιούν και αλλάζουν τη διεύθυνση IP ενός χρήστη στα πλαίσια εξοικονόμησης πόρων και βελτίωσης παρεχόμενης υπηρεσίας. Ενδεικτικό είναι το γεγονός ότι σε περιοχές με ανεπτυγμένο βιοτικό επίπεδο τα δεδομένα είναι πιο αξιόπιστα, καθότι υπάρχει καλύτερο δίκτυο σύνδεσης (Ganelin κ.α,2019).

3.2 Μέθοδος Ανάλυσης Μελέτης.

Στο πρώτο μέρος της εργασίας έγινε εστίαση στα εξής σημεία:

- Στο ποσοστό κατανομής των συμμετεχόντων στις εκπαιδεύσεις ανά κάτοικο των ΗΠΑ.
- Στη συσχέτιση του ποσοστού των Ταχυδρομικών Κωδικών (TK) των χρηστών που συμμετέχουν στις εκπαιδεύσεις και των TK των αντίστοιχων οικονομικών περιουσιών.
- Στην εξέταση για το αν οι εμπλεκόμενοι TK είναι μονοδιάστατο μέγεθος και αν από αυτό το χαρακτηριστικό μπορεί να εξαχθεί κάποιο συμπέρασμα για επιπλέον οικονομικούς παράγοντες που επηρεάζουν τη συμμετοχή στις διαδικτυακές εκπαιδεύσεις.

Στο δεύτερο μέρος της εργασίας, μελετήθηκε η ακρίβεια και η μεροληψία του εντοπισμού θέσης μέσω IP, χρησιμοποιώντας πραγματικά δεδομένα τα οποία είχαν συμπληρώσει οι ίδιοι οι χρήστες των εκπαιδεύσεων. Μέσα από αυτή τη διαδικασία εξετάστηκε:

- Η διαφορά, μέσω σύγκρισης, των πραγματικών δεδομένων και των εκτιμήσεων μέσω της διαδικασίας εντοπισμού τοποθεσίας IP.
- Η μελέτη του σφάλματος στο γεωεντοπισμό και πως εκείνο ποικίλει ανάλογα τον TK και τις οικονομικές δυνατότητες της περιοχής.

- Η συνολική διαφορά στις κατανομές του οικονομικού υπόβαθρου μεταξύ των πραγματικών δεδομένων τοποθεσίας και των αντίστοιχων διευθύνσεων IP.

Για την καταγραφή των TK των περιοχών ανά διαδικτυακή διάλεξη, πραγματοποιήθηκε ψυχομετρική μέτρηση κατά την οποία ο κάθε TK, αντιστοιχούσε σε έναν «υποψήφιο», ενώ η διάλεξη ήταν το «διαγώνισμα». Επί της ουσίας εφαρμόστηκε περιγραφική στατιστική και έγινε χαρτογράφηση των δεδομένων, για να βρεθούν πεπατημένες στις πληροφορίες και να εξεταστούν οι αλληλοσχετίσεις μεταξύ των TK που χρησιμοποιούνταν και των οικονομικών ιδιοτήτων τους. Για τη διασφάλιση της μονοδιάστασης των δεδομένων καθώς και της αλληλοσυσχέτισης τους, χρησιμοποιήθηκε η μέθοδος ανάλυσης παραγόντων, ενώ για τον έλεγχο της αξιοπιστίας τους έγινε χρήση της κλασικής θεωρίας.

Στην υπάρχουσα μελέτη, οι Ganelin κ.α, έλαβαν δεδομένα από τις πραγματικές διευθύνσεις που δήλωσαν οι χρήστες τους, μέσα από τους TK των περιοχών τους και σύγκριναν τις πληροφορίες αυτές με τους TK που εμφάνιζε η MaxMind από τη βάση δεδομένων για την κάθε IP που συνδεόταν. Στη συνέχεια υπολογίστηκε η απόσταση των δυο TK και εξετάστηκε η κατανομή των αποστάσεων η οποία βασίστηκε στις πραγματικές κατοικίες (ευημερία, περιοχή, πληθυσμός και πυκνότητα περιοχής) με βάση τον πραγματικό TK. Επισημαίνεται ότι, η συγκεκριμένη μελέτη θεωρείται πρωτοπόρα και αποτελεί μια από τις ελάχιστες προσπάθειες που επιδιώκει να εξετάσει τις οικονομικές πληροφορίες και δεδομένα που μπορούν να εξαχθούν από την τοποθεσία της IP διεύθυνσης ενός χρήστη. Το γεγονός αυτό, υποδηλώνει πόσο νέα τεχνολογία θεωρείται ο Γεωεντοπισμός IP ως τεχνική για την άντληση στοιχείων στον τομέα της οικονομίας.

3.3 Η Διαδικτυακή Εκπαίδευση (ΔΕ)

Έχουν γίνει αρκετές έρευνες οι οποίες φτάνουν στο συμπέρασμα ότι η ΔΕ και τα διαδικτυακά σεμινάρια τείνουν να απαρτίζονται από συμμετέχοντες σε χώρες με υψηλό κοινωνικοοικονομικό υπόβαθρο. Συγκεκριμένα, υπάρχουν αναλύσεις που εστιάζουν στη συμμετοχή στη ΔΕ σε επίπεδο χώρας. Οι Christensen Gayle, Steinmetz, Alcorn, Bennett, Woods και Ezekiel Emmanuel, σε έρευνα που πραγματοποίησαν ανάμεσα στους συμμετέχοντες του πανεπιστημίου της Πενσυλβάνια των ΗΠΑ σε επιμορφωτικά σεμινάρια στη εκπαιδευτική πλατφόρμα Coursera, παρατήρησαν ότι η μεγαλύτερη πλειοψηφία εξ αυτών προέρχονταν από αναπτυγμένες χώρες και συγκεκριμένα τις

ΗΠΑ (Christensen κ.α., 2013). Επιπρόσθετα, οι Chuang και Ho μελετώντας τις διαδικτυακές πλατφόρμες των πανεπιστημίων του Χάρβαρντ και MIT, παρατήρησαν μια εκθετική σχέση ανάμεσα στο δείκτη ανθρώπινης ανάπτυξης (Human Development Index) και στον αριθμό των πτυχίων που κατέχουν οι κάτοικοι σε συσχέτιση με τον πληθυσμό (Chuang και Ho, 2016). Παρόμοια αποτελέσματα εμφάνισαν και οι μελέτες των Kizilcec, Piech and Schneider, 2017, όπου παρατηρήθηκε όμοια συσχέτιση του Human Development Index (HDI) και του ποσοστού ολοκλήρωσης σεμιναρίων στην πλατφόρμα του Coursera (Kizilcec κ.α., 2017). Συνολικά το 73% των συμμετεχόντων σε ΔΕ αυτού του επιπέδου διέθετε ανώτερη εκπαίδευση. Συγκεκριμένα στην πλατφόρμα του HarvardX/MITx, το 73% διέθετε εκπαιδευτικό υπόβαθρο υψηλού επιπέδου, ενώ αντίστοιχα οι συμμετέχοντες του πανεπιστημίου της Πενσυλβάνια (ΠΠ) που κατείχαν ανώτερη εκπαίδευση αποτελούσαν το 79% (Christensen κ.α., 2013). Στην εν λόγω μελέτη έγινε επίσης σύγκριση των κατανομών των χωρών σύμφωνα με τις αναφορές και τον εντοπισμό της τοποθεσίας με βάση την IP τους. Η μελέτη αρχικά είχε σκοπό τον έλεγχο της εγκυρότητας των δεδομένων των συμμετεχόντων της έρευνας και δευτερευόντως την ακρίβεια της τοποθεσίας της IP. Από την μελέτη παρατηρήθηκε ότι η διαφορά του αριθμού των μαθητών με βάση τα δεδομένα της έρευνας τους και του αντίστοιχου αριθμού με βάση την IP τους, εμφάνιζαν διαφορά μικρότερη από 1,56%.

Επιπρόσθετα από την ανωτέρω μελέτη παρατηρήθηκε ότι οι συμμετέχοντες στο πανεπιστήμιο της Πενσυλβάνια είναι πιο πιθανόν να έχουν πτυχίο κολεγίου, σε σχέση με τους αντίστοιχους συμμετέχοντες άλλων χωρών (Emanuel, 2013).

Πέραν των ανωτέρω, υπάρχουν έρευνες οι οποίες εστιάζουν στην κατανομή του εκπαιδευτικού υποβάθρου των συμμετεχόντων στις χώρες τους. Για παράδειγμα οι Ινδοί συμμετέχοντες στα μαθήματα της πλατφόρμας Coursera του ΠΠ, προέρχονταν από ευημερείς πόλεις της χώρας, οι οποίες αποτελούσαν ένα μικρό ποσοστό του συνολικού πληθυσμού της Ινδίας (Alcorn, Christensen και Kapur, 2015). Η ανωτέρω παρατήρηση έγινε με τη χρήση εντοπισμού της τοποθεσίας IP των χρηστών στη ΔΕ και επιβεβαιώνει τη θετική συσχέτιση του βιοτικού επιπέδου των κατοίκων με την εκπαίδευση.

Μια εξίσου σημαντική έρευνα που πραγματοποιήθηκε στις ηλεκτρονικές πλατφόρμες εκπαίδευσης HarvardX και MITx, στην οποία μέσω των φυσικών

διευθύνσεων που είχαν δηλωθεί από τους συμμετέχοντες, οι μελετητές διαχώρισαν τις περιοχές σε μικρότερα τμήματα σε επίπεδα γειτονιάς για να εξαχθούν λεπτομερέστερα συμπεράσματα. Από τη μελέτη διαπιστώθηκε ότι στις ΗΠΑ οι συμμετέχοντες σε προγράμματα ΔΕ προέρχονταν από ευημερείς, ανώτερης εκπαίδευσης και πυκνοκατοικημένες περιοχές (Hansen κ.α.,2015).

Επιπρόσθετα, από έρευνα που πραγματοποιήθηκε από τους Zhenghao, Chen, Brandon Alcorn, Gayle Christensen, Nicholas Eriksson, Daphne Koller και Ezekiel J. Emanuel, 2015, παρατηρήθηκε ότι οι συμμετέχοντες σε σεμινάρια της Coursera, οι οποίοι προέρχονταν από βαθμίδα ανώτερης εκπαίδευσης, ναι μεν επωφελήθηκαν από το σεμινάριο, αλλά δεν διέκριναν κάποια ιδιαίτερη βελτίωση στο εργασιακό τους αντικείμενο. Αντιθέτως, οι συμμετέχοντες που προέρχονταν από κατώτερες βαθμίδες εκπαίδευσης, ανέφεραν ότι με την ολοκλήρωση των σεμιναρίων επωφελήθηκαν, είτε με αύξηση στο μισθό τους, είτε με προαγωγή στην εργασία τους, ενώ ακόμη σε μερικές περιπτώσεις οι συμμετέχοντες ανέφεραν ότι τους εμφανίστηκαν νέες προοπτικές και ευκαιρίες καριέρας (Zhenghao κ.α.,2015).

Πέραν των παραπάνω μελετών, αναφορικά με τις ΔΕ, ο εντοπισμός της τοποθεσίας της IP χρησιμοποιείται και σε λοιπούς διεπιστημονικούς κλάδους, όπου εγείρονται εξίσου ανησυχίες και προβληματισμοί για την ακρίβεια και εγκυρότητα τους. Οι State, Bogdan, Ingmar Weber και Emilio Zagheni, 2013, μελέτησαν σε διεθνές επίπεδο τη μετανάστευση των χρηστών της ηλεκτρονικής πλατφόρμας Yahoo και των οικονομικών παραγόντων που οφείλονται για αυτήν. Τα αποτελέσματα τους παρουσίασαν αξιόπιστες πληροφορίες όσον αφορά τον εντοπισμό της θέσης IP σε επίπεδο χώρας, αλλά ήταν λιγότερο αξιόπιστες σε επίπεδο πόλης (State κ.α, 2013). Τα εν λόγω αποτελέσματα είχαν λάβει δικλίδες ασφαλείας εξαιρώντας διευθύνσεις με ελλιπή δεδομένα θέσης (ιστορικό των δεδομένων IP). Παρόμοια μελέτη διεξήχθη από τους Zagheni & Weber, οι οποίοι μελέτησαν τη μετανάστευση με βάση τις διευθύνσεις e-mail των χρηστών (Zagheni κ.α.,2012). Οι Mielke & Chen χρησιμοποίησαν τον εντοπισμό τοποθεσίας της IP για την ιχνηλάτηση των θέσεων Τζιχαντιστών μέσω των ιστοσελίδων τους σε επίπεδο πόλης (Mielke κ.α.,2008).

Γενικότερα, ο εντοπισμός τοποθεσίας μιας IP χρησιμοποιείται πολλές φορές και από οργανισμούς για να εντοπίσουν τους συνδρομητές τους (Goodell & Syverson, 2007), αλλά και από εταιρίες ειδήσεων για να επιλέξουν να προβάλουν τις τοπικές

ειδήσεις στην εκάστοτε IP που θα επισκεφτεί την ιστοσελίδα (Shavitt κ.α. 2011). Επιπλέον πολλές οικονομικές εταιρίες με το γεωεντοπισμό πετυχαίνουν να «μπλοκάρουν» τις απάτες (Curran κ.α.,2011). Οι έμποροι επίσης αξιοποιούν τα δεδομένα για να καθορίσουν την τιμή (Valentino DeVries, Jennifer, Jeremy Singer-Vine, and Ashkan Soltani,2012) ή για να προσδιορίσουν τα επίπεδα φόρων των ηλεκτρονικών αγορών (Svantenson, 2019). Τέλος η αστυνομία χρησιμοποιεί το γεωεντοπισμό για να καταπολεμήσει το έγκλημα εντοπίζοντας εγκληματίες και για να αποτρέπει τις απάτες ενισχύοντας την κυβερνοάμυνα (Shavitt κ.α. 2011).

3.4 Γεωγραφικοί Δείκτες Κοινωνικοοικονομικής Κατάστασης

Αναφορικά με την ταυτοποίηση των οικονομικών στοιχείων και τη μετέπειτα διασταύρωση με τις διευθύνσεις IP τους, έχουν πραγματοποιηθεί μελέτες όπου έχουν στόχο την παρατήρηση των δεδομένων κοινωνικοοικονομικής κατάστασης των πολιτών μιας χώρας και της τοποθεσίας τους. Οι Geronimus, Arline & Bound, το 1998 διαπίστωσαν ασυμφωνία μεταξύ των TK και της κοινωνικοοικονομικής κατάστασης των πολιτών, με βάση το εισόδημα τους, καθώς υπήρχε 89% μεγαλύτερη διακύμανση του εισοδήματος των πολιτών εντός του ίδιου TK, σε σχέση με το συνολικό πληθυσμό της χώρας. Συνεπώς τα δεδομένα οικονομικού χαρακτήρα μιας IP με μόνο κριτήριο καθαρά την τοποθεσία της, δεν μπορούν να θεωρηθούν αξιόπιστα για την εξαγωγή έμπιστων και ασφαλών συμπερασμάτων (Geronimus κ.α., 1998).

3.5 Συλλογή των Δεδομένων

Για την συλλογή των δεδομένων που χρησιμοποιήθηκαν στη μελέτη που διεξήχθη από την Ganelin κ.α. το 2019, ακολουθήθηκε η μια συνδυαστική διαδικασία. Ως αρχική πηγή δεδομένων χρησιμοποιήθηκε η διαδικτυακή πλατφόρμα παροχής σεμιναρίων και εκπαιδεύσεων EDx. Σε πρώτο στάδιο συλλέχθηκαν οι IP διευθύνσεις των χρηστών που είχαν δηλώσει συμμετοχή σε κάποια εκπαίδευση. Στη συνέχεια καταγράφηκαν οι ταχυδρομικές διευθύνσεις που είχαν δηλωθεί από τους χρήστες κατά τη συμπλήρωση της αρχικής φόρμας εγγραφής. Με χρήση της βάσης δεδομένων MaxMind πραγματοποιήθηκε εντοπισμός των TK των χρηστών με βάση την IP τους και στη συνέχεια με την εφαρμογή Google Maps Geocoding API αντιστοιχίστηκαν οι ταχυδρομικές διευθύνσεις τους, με τους εκάστοτε TK των περιοχών. Κατόπιν αυτών έγινε σύγκριση των στοιχείων με το δείκτη κοινωνικής εξουθένωσης, με βάση τους TK των περιοχών και εξάχθηκαν τα αντίστοιχα συμπεράσματα.

Για τη συλλογή δεδομένων της εργασίας χρησιμοποιήθηκαν από την πλατφόρμα EDx πληροφορίες αναφορικά με τις εγγραφές των συμμετεχόντων των πανεπιστημίων του Harvard και MIT από το 2012 μέχρι και το 2018. Κάθε υποψήφιος - συμμετέχων σε πρόγραμμα ΔΕ αποτελούσε και μια καταγραφή (διεύθυνση IP). Στην περίπτωση που ένας υποψήφιος παρακολουθούσε και δεύτερο πρόγραμμα ΔΕ, αυτόματα προσμετρούνταν σαν δεύτερη καταγραφή. Κάθε καταγραφή υποψηφίου περιελάμβανε τα ακόλουθα:

- Πληροφορίες αναφορικά με το πρόγραμμα ΔΕ, όπως ο τίτλος προγράμματος, η ημερομηνία διεξαγωγής, ο φορέας υλοποίησης του προγράμματος, η θεματολογία και ο ελάχιστος απαιτούμενος βαθμός επιτυχίας.
- Αυτοματοποιημένες πληροφορίες του χρήστη, όπως η ταυτότητα του χρήστη, το e-mail του, οι παράμετροι των ρυθμίσεων του προγράμματος περιήγησης που χρησιμοποιεί για την εκπαίδευση.
- Δημογραφικά στοιχεία του χρήστη, σε περίπτωση που τα δηλώσει ο ίδιος, όπως η ημερομηνία γέννησης του, το φύλο και το επίπεδο εκπαίδευσης του.
- Η διεύθυνση IP του χρήστη και συγκεκριμένα η πιο συχνή ταχυδρομική διεύθυνση με την οποία συμμετέχει στο πρόγραμμα.
- Συνοπτικές πληροφορίες αναφορικά με την εμπλοκή του υποψηφίου με το πρόγραμμα, όπως στοιχεία για το εάν ολοκλήρωσε το πρόγραμμα, το βαθμό επίδοσης του και εάν έχει ολοκληρώσει τη διαδικασία ταυτοποίησης.
- Επιπρόσθετες πληροφορίες του χρήστη, όπως ο αριθμός παύσεων του προγράμματος, εάν επέλεξε την επιλογή εμφάνισης απάντησης σε προβλήματα και εάν υπήρχε σχολιασμός κατά τη διάρκεια της εκπαίδευσης.

Συνολικά συγκεντρώθηκαν 135.000 χρήστες οι οποίοι συμπλήρωσαν σε φόρμα συμμετοχής τις ταχυδρομικές τους διευθύνσεις. Επισημαίνεται ότι σε αρκετές περιπτώσεις υπήρχαν ελλείψεις δεδομένων (απουσία πόλης ή περιοχής) και εσφαλμένα e-mails. Για τις περιπτώσεις αυτές αν και αναφέρονται παρακάτω κάποια εργαλεία που διευκολύνουν την περαιτέρω ανάλυση, για λόγους εγκυρότητας παραλήφθηκαν από τη μελέτη.

Αναφορικά με τα δεδομένα που ανασύρθηκαν από τη MaxMind τονίζεται ότι για την ακρίβεια και εγκυρότητα των δεδομένων η βάση επιτυγχάνει 90% ακρίβεια με εκτιμώμενο σφάλμα εντοπισμού την ακτίνα των 100 χλμ.

3.6 Δείκτης Κοινωνικής Εξουθένωσης Distressed Community Index

Το Economic Innovation Group (EIG) δημοσιεύει πίνακες με την οικονομική «υγεία» για πάνω από 25.000 TK των ΗΠΑ (Economic Group, 2018). Κάθε TK περιλαμβάνει πληθυσμιακά δεδομένα όπως την αναλογία διαφόρων εθνικοτήτων ανά TK, την πυκνότητα του πληθυσμού σε μια κλίμακα από «Πολλή Χαμηλή» μέχρι «Πολύ Υψηλή» και επτά οικονομικά μεγέθη όπου περιγράφουν: την αναλογία ενηλίκων χωρίς απολυτήριο λυκείου, ποσοστό διαθεσιμότητας οικιών, ποσοστό ανεργίας, ποσοστό φτώχειας, λόγος μέσου εισοδήματος προς το μέσο εισόδημα του κράτους, ποσοστό μεταβολής της ανεργίας από το 2012 στο 2016 και το ποσοστό αλλαγής του αριθμού των επιχειρήσεων από το 2012 έως το 2016.

Ο κάθε πίνακας περιλαμβάνει τον δείκτη κοινωνικής εξουθένωσης (ΔΚΕ) για τον εκάστοτε TK. Ο υπολογισμός του γίνεται με βάση την ταξινόμηση των επτά παραπάνω οικονομικών μεγεθών, υπολογίζοντας έναν μέσο όρο στον κάθε TK. Στη συνέχεια γίνεται ομοιόμορφη κατανομή των δεδομένων δημιουργώντας μια κλίμακα από 0 έως 100, στην οποία από το 0 έως 20 η βιωσιμότητα της περιοχής με το συγκεκριμένο TK χαρακτηρίζεται ως «ευημερής», από 20 έως 40 χαρακτηρίζεται ως «άνετη», από 40 έως 60 χαρακτηρίζεται «μέτρια», από 60 έως 80 χαρακτηρίζεται «επισφαλής» και από 80-100 η περιοχή θεωρείται «εξουθενωμένη».

3.7 Απόδοση των Μετρήσεων

Σε πρώτη φάση διερευνήθηκε η εμπλοκή των μαθημάτων του HarvardX και MITx στους διαφορετικούς TK, αντιστοιχίζοντας αυτούς με τις εκάστοτε IP. Αρχικά μελετήθηκε η σχέση μεταξύ του κατά κεφαλήν ποσοστού χρήσης ανά TK και των διαφόρων οικονομικών παραγόντων τους. Για την επίτευξη αυτού χρησιμοποιήθηκαν τα ευρήματα των (Hansen κ.α., 2015b) αξιολογώντας έναν αξιόπιστο δείκτη οικονομικής ευημερίας με δείγμα όλους τους χρήστες της πλατφόρμας EDx για τα έξι έτη. Παράλληλα στην υπό ανάλυση μελέτη έγινε και καταγραφή του δείκτη που δείχνει την τάση για συμμετοχή σε ΔΕ ανά TK. Με τον δείκτη αυτό γίνεται πιο εμφανής η συσχέτιση της εκπαίδευσης σε ΔΕ και των οικονομικών παραγόντων και μειώνονται τα ποσοστά εμφάνισης ψευδών δεδομένων ή «θορύβου», όπως για παράδειγμα θα

μπορούσε να είναι η υψηλή συμμετοχή ατόμων σε ΔΕ σε έναν ΤΚ, λόγω ύπαρξης πανεπιστημίων στο συγκεκριμένο ΤΚ. Για τον υπολογισμό των δεδομένων χρησιμοποιήθηκε η μέθοδος Ανάλυσης Παραγόντων (Factor Analysis) με δεδομένα τους ΤΚ και ως μεταβλητές τις συμμετοχές των υποψηφίων σε προγράμματα ΔΕ. Θέτοντας ως δεδομένα γεωγραφικές περιοχές, εμφανίζονται προβλήματα όπως η πιθανή εξάρτηση μεταξύ γειτονικών περιοχών (Ganelin κ.α.,2019).

Στη μελέτη τους οι (Ganelin κ.α.,2019) χειρίστηκαν την εμπλοκή των ΔΕ χρησιμοποιώντας τον αριθμό συμμετεχόντων, την ολοκλήρωση των σπουδών και τις πιστοποιήσεις τους για κάθε ΤΚ. Αν και τα δύο τελευταία ταυτίζονται, πολλοί σπουδαστές ολοκλήρωσαν την παρακολούθηση των ΔΕ αλλά δεν προχώρησαν στη διαδικασία απόδοσης πιστοποίησης καθώς δεν είχαν ολοκληρώσει τις τελικές εξετάσεις των προγραμμάτων. Τα πιστοποιητικά παρακολούθησης και απόδοσης πτυχίων διέφεραν από την κάθε εκπαίδευση. Ως εκ τούτου είναι δύσκολο να καθοριστεί το ποσοστό περάτωσης εκπαίδευσης καθώς σε πολλές ΔΕ, δεν υπήρχε ξεκάθαρος αριθμός συμμετεχόντων σε συγκεκριμένους ΤΚ.

Συγκεκριμένα επιδιώχθηκε ο προσδιορισμός της κατανομής των ΔΕ και των συσχετιζόμενων οικονομικών παραγόντων τους λαμβάνοντας υπόψη τα κάτωθι:

- Τη μελέτη της κατανομής των ποσοστών συμμετοχών ανά πρόγραμμα εκπαίδευσης, ολοκλήρωσης και απόδοσης πιστοποιητικού για τους ΤΚ των ΗΠΑ.
- Τον προσδιορισμό της συσχέτισης των ΤΚ που έχουν δηλωθεί για τις ΔΕ και των οικονομικών στοιχείων των ΤΚ αυτών.
- Την εξέταση εάν η εμπλοκή συγκεκριμένων ΤΚ με τις ΔΕ είναι μονοδιάστατη, ώστε αυτό το χαρακτηριστικό να δύναται να χρησιμοποιείται ως δείκτης για τη μελέτη περαιτέρω οικονομικών στοιχείων και την πρόβλεψη τους σε δεύτερη φάση.

3.8 Προσέγγιση Επεξεργασίας των Δεδομένων

Για την υλοποίηση των ανωτέρω, πραγματοποιήθηκε ο συνδυασμός των πληροφοριών από τρεις διαφορετικές βάσεις. Στη συνέχεια, έγινε χαρτογράφηση των δεδομένων με χρήση των βιβλιοθηκών των γλωσσών προγραμματισμού Python Panda, Geo Panda, Matplotlib και PySAL και υλοποιήθηκε μια ποσοτική ανάλυση των δεδομένων, δημιουργώντας γραφήματα μέσω Python και STATA.

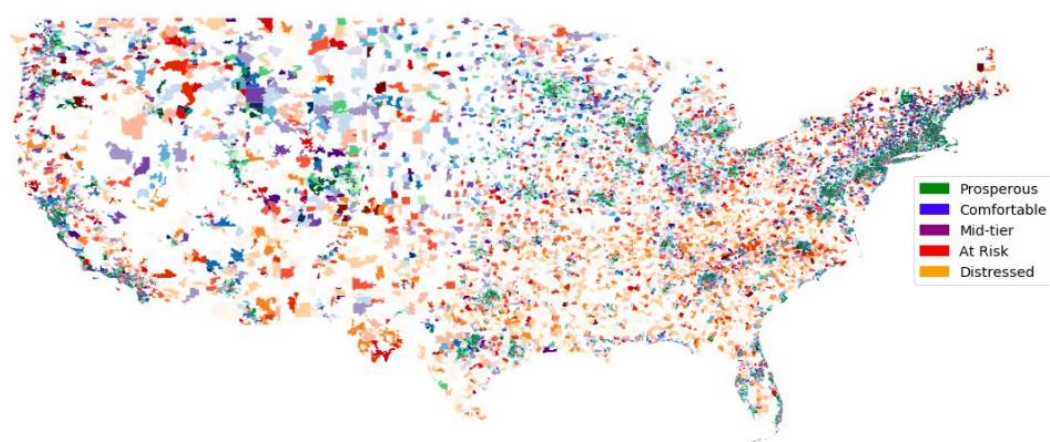
Αναλυτικότερα με τη μέθοδο Factor Analysis, μελετήθηκε η κάθε ΔΕ ανά ΤΚ, μοντελοποιώντας με αυτόν τον τρόπο την εμπλοκή του κάθε προγράμματος εκπαίδευσης με τον εκάστοτε ΤΚ. Επί της ουσίας έγινε εξομοίωση υπό τη μορφή «εξετάσεων», όπου οι διαγωνιζόμενοι ήταν οι ΤΚ και οι ΔΕ ήταν οι εξεταστές. Η βαθμολογία που σχηματίστηκε ανά ΤΚ για συγκεκριμένη ΔΕ, ήταν ο αριθμός των συμμετεχόντων (είτε περάτωση, είτε πιστοποίηση) ανά εκατομμύριο κατοίκων ανά ΤΚ. Συγκεκριμένα οι (Ganelin κ.α.,20) ακολούθησαν την εξής θεωρία. Έστω p ο αριθμός των εκπαιδευτικών προγραμμάτων. Κάθε σημείο δεδομένων r_{jk} είναι ο αριθμός των συμμετεχόντων για το j πρόγραμμα ΔΕ στον k ΤΚ ανά εκατομμύριο συμμετεχόντων. Έστω u_j η μέση τιμή των ΤΚ για το κάθε πρόγραμμα εκπαίδευσης j . Με τη μέθοδο Factor Analysis προστίθεται m παράγοντες υπολογίζοντας έναν παράγοντα φορτίου l_{ij} για κάθε παράγοντα i σε κάθε πρόγραμμα j και έναν παράγοντα επίδοσης (score) s_{ik} για κάθε παράγοντα i σε κάθε k ΤΚ. Συνεπώς η εξίσωση υπολογισμού των δεδομένων είναι η ακόλουθη:

$$r_{jk} = u_j + \sum_{i=1}^m l_{ij}s_{ik} + e_{jk} , \quad (3)$$

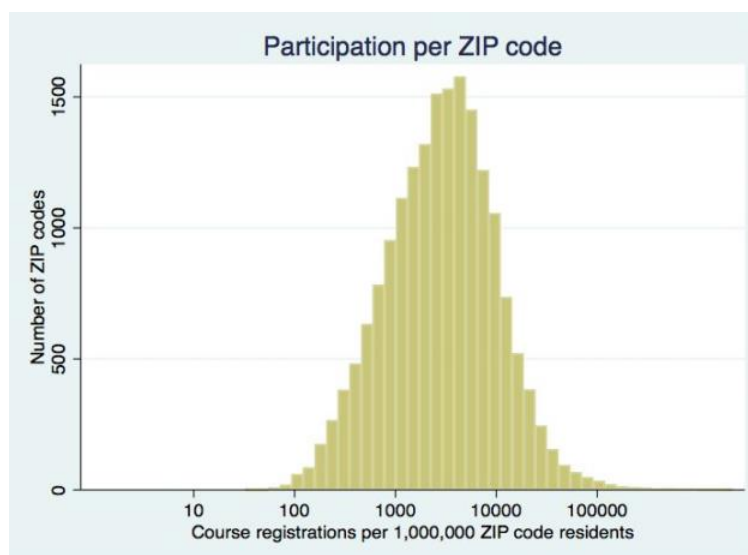
όπου e_{jk} είναι ο παράγοντας σφαλμάτων - καταλοίπων. Για την ακρίβεια και αξιοπιστία των αποτελεσμάτων χρησιμοποιήθηκε η συνάρτηση Cronbach's alpha. Στη συνέχεια, υπολογίστηκε η συσχέτιση μεταξύ των παραγόντων φορτίου και επίδοσης με τις άλλες μεταβλητές, έπειτα από εφαρμογή μετασχηματισμών που επιχειρήθηκαν για τη εξέταση αλληλεπίδρασης των δεδομένων. Συγκεκριμένα οι Ganelin κ.α., για τους εκάστοτε ΤΚ, αναζήτησαν την αλληλεπίδραση των επιδόσεων των πρώτων παραγόντων, του αριθμού ολοκλήρωσης και πιστοποίησης ανά πληθυσμό και ανά αριθμό συμμετεχόντων, του δείκτη DCI, των παραγόντων επτά στοιχείων του δείκτη DCI, του ποσοστού μειοψηφίας των κατοίκων και του πληθυσμού και πυκνότητας πληθυσμού. Για τα προγράμματα εκπαίδευσης, ερευνήθηκε η σχέση μεταξύ των παραγόντων φορτίου αριθμού συμμετοχών, περάτωσης και απόδοσης πιστοποιητικού, καθώς επίσης και του ποσοστού συμμετεχόντων που περάτωσαν και πιστοποιήθηκαν. Η συσχέτιση που καταγράφηκε από την έρευνα για την Cronbach's alpha κυμάνθηκε στο $|r| > 0.2$. Το μέγεθος αυτό για το συντελεστή αξιοπιστίας Cronbach's alpha δεν θεωρείται και τόσο αποδεκτό καθότι στη στατιστική, τιμές μικρότερες από 0,5 θεωρούνται μη αποδεκτές καθότι οι υπό μελέτη μεταβλητές δεν αλληλοεπιδρούν.

3.9 Αποτελέσματα των Δεδομένων

Από τη ψυχομετρική ανάλυση χρησιμοποιήθηκαν δεδομένα από τη MaxMind και υπολογίστηκαν οι TK για τους οποίους υπήρχαν δεδομένα και οι δείκτες οικονομικής εξουθένωσης των περιοχών. Συνεπώς συλλέχθηκαν 2,7 εκατομμύρια συμμετοχών καταναμημένα σε πάνω από 18.000 TK και 576 εκπαιδευτικά προγράμματα, πολλά εκ των οποίων ήταν συμπληρωματικά των ήδη υπαρχόντων εκπαιδευτικών προγραμμάτων σε διαφορετικά εξάμηνα. Στην εικόνα1 απεικονίζεται η κατανομή των ΔΕ ανά TK στις ΗΠΑ με βάση τη MaxMind ενώ στο γράφημα 2 απεικονίζεται το αντίστοιχο ιστόγραμμα για τις ΔΕ ανά TK (Ganelin κ.α.,2019).



Εικόνα 1. Χάρτης κατανομής συμμετοχών σε ΔΕ από τη MaxMind την περίοδο 2012-2018. © Ganelin κ.α.,2019.

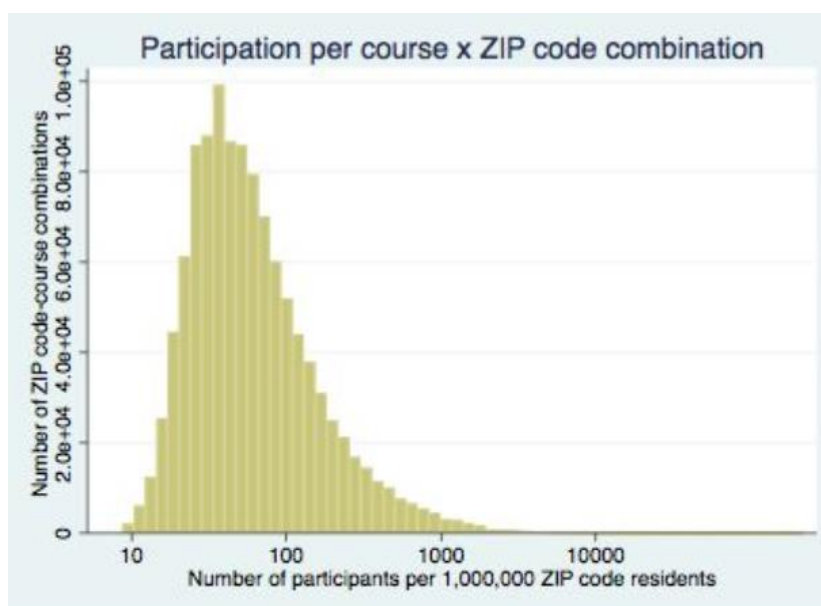


Γράφημα 1. Ιστόγραμμα κατανομής συμμετοχών σε ΔΕ των πανεπιστημίων Harvard / MITx ανά εκατομμύριο κατοίκων. © Ganelin κ.α.,2019.

Επισημαίνεται ότι για τους TK που έχουν εξαιρεθεί από τις μετρήσεις λόγω μη ταυτοποιημένης καταγραφής, αυτοί κυμάνθηκαν από 2,2 εκατομμύρια έως 32

εκατομμύρια συμμετοχές ανά εκατομμύριο πληθυσμού με μια ενδιάμεση τιμή 2.970 και μέσο 7.380. Το γεγονός αυτό υποδηλώνει ότι πολλοί χρήστες παρακολουθούσαν ή είχαν καταχωρηθεί ταυτόχρονα σε διαφορετικές ΔΕ, με αποτέλεσμα ο αριθμός των προγραμμάτων εκπαίδευσης να είναι μεγαλύτερος από τον πληθυσμό. Για το λόγο αυτό στις περιπτώσεις όπου τα προγράμματα ήταν περισσότερα από τους συμμετέχοντες θεωρήθηκαν «ύποπτες» και αναξιόπιστες για την εξαγωγή συμπερασμάτων. Επιπρόσθετα από την εικόνα 1 και κυρίως από το γράφημα 1 παρατηρήθηκε ότι οι περιοχές που εμφάνιζαν αριθμό συμμετοχών 900.000 ανά εκατομμύριο κατοίκων, ήταν περιοχές με πολλά πανεπιστήμια και ως εκ τούτου χαρακτηρίστηκαν ως κέντρα τεχνολογίας (tech hub) όντας ανάμεσά τους το Cambridge, το Boston, το St.Louis, κ.α. (Ganelin κ.α.,2019).

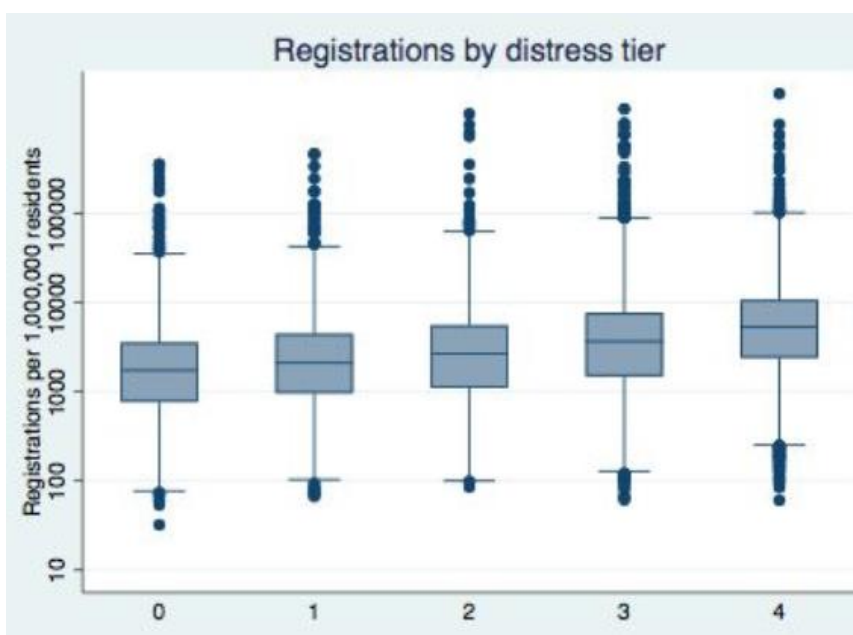
Επιπρόσθετα από την παραπάνω μελέτη, παρατηρήθηκε ότι στις περιοχές με λίγους κατοίκους ανά TK, τα δεδομένα ήταν πιο ευμετάβλητα. Από τους 18.000 TK ο αριθμός συμμετοχών σε ΔΕ κυμαίνονταν από 510 μέχρι 115.000 με ενδιάμεση τιμή 10.860 και μέση τιμή 16.535. Συνολικά, για πάνω από 10,5 εκατομμύρια πιθανών συνδυασμών ενός προγράμματος εκπαίδευσης και ενός TK, μόνο τα 1,1 εκατομμύρια απεικονίζονταν σε δεδομένα. Στο γράφημα 2 απεικονίζεται η κατανομή συμμετοχών ανάμεσα στο 1,1 εκατομμύριο ζευγαριών. Σε αυτή τη βάση οι εγγραφές κυμάνθηκαν από 8,7 έως 220.000 για προγράμματα εκπαίδευσης ανά εκατομμύριο κατοίκων ανά TK, με ενδιάμεση τιμή 52 και μέση τιμή 121.



Γράφημα 2. Ιστόγραμμα μαθημάτων από τις καταχωρήσεις TK στα 1,1 εκατομμύρια δεδομένων. © Ganelin κ.α.,2019.

3.10 Οικονομική Συσχέτιση των Δεδομένων

Όπως ήταν αναμενόμενο και έχει παρατηρηθεί και σε πρότερες μελέτες, οι πιο ευημερείς ΤΚ εμφάνισαν τις πιο πολλές καταχωρήσεις. Στο γράφημα 3 απεικονίζεται το σχεδιάγραμμα όπου ερμηνεύει όσον αφορά τους ΤΚ, τη θετική συσχέτιση του λογάριθμου των κατά κεφαλήν καταχωρήσεων σε ΔΕ και την πυκνότητα του πληθυσμού, με συντελεστή $r = 0.36$, ενώ με το μέσο εισόδημα ο συντελεστής διαμορφώνεται στο $r = 0.24$. Αντιθέτως, παρουσιάστηκε αρνητική συσχέτιση σε αναλογία με τους ενηλίκους που δεν κατείχαν απολυτήριο λυκείου, με $r = -0.27$, DCI $r = -2.6$ και το ποσοστό ανεργίας $r = -0.25$.

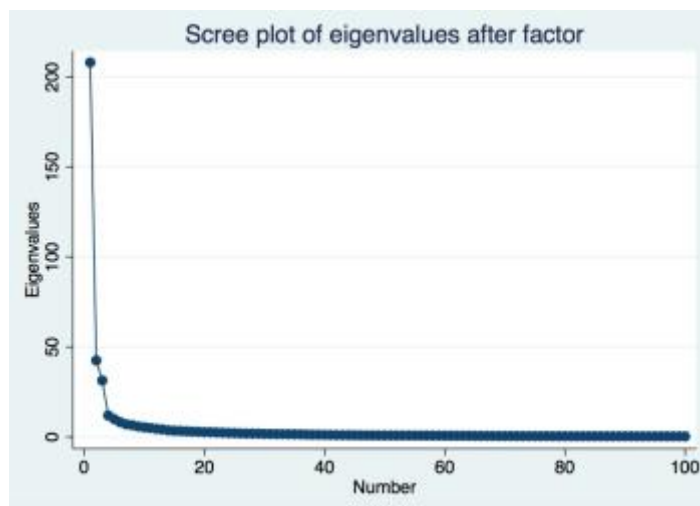


Γράφημα 3. Μπλοκς από ΤΚ και κατά κεφαλήν συμμετοχές σε ΔΕ. Βαθμίδες από 0 (εξουθενωμένοι οικονομικά) έως 4 (ευημερείς). Για κάθε βαθμίδα το μπλοκ απεικονίζει το 25%, 50% και 75% των δεδομένων. Οι αγκύλες απεικονίζουν τις τιμές εντός εμβέλειας, ενώ οι κουκίδες τις γραμμές εκτός εμβέλειας. © Ganelin κ.α., 2019.

Από το Γράφημα 3 παρατηρείται ότι η συσχέτιση μεταξύ ευημερίας και αναλογίας συμμετεχόντων που λαμβάνουν πιστοποίηση, δεν είναι ξεκάθαρη, γεγονός που οφείλεται στο ότι το 4% των συμμετεχόντων δεν έλαβαν πιστοποίηση. Συνεπώς παρατηρήθηκε αρνητική συσχέτιση του λογαρίθμου του ποσοστού ολοκλήρωσης και του λογαρίθμου του πληθυσμού με $r = -0.2$. Επιπρόσθετα, ο αντίστοιχος λόγος του λογαρίθμου του ποσοστού απόδοσης πιστοποιητικού είναι $r = -0.21$. Είναι αξιοσημείωτο να αναφερθεί ότι ο λογάριθμος ολοκλήρωσης ΔΕ ανά συμμετέχοντα και ο λογάριθμος πληθυσμού παρουσίασε $r = -0.53$. Αντίστοιχα ο λογάριθμος ολοκλήρωσης ΔΕ με το λογάριθμο της πυκνότητας πληθυσμού παρουσίασε συσχέτιση

με $r = -0.3$, ενώ με το ποσοστό μειονότητας κατοίκων παρουσίασε συσχέτιση $r = -0.23$. Ως εκ τούτου, παρατηρήθηκε ότι πολίτες από πυκνοκατοικημένες περιοχές, είναι περισσότερο πιθανό να εγγραφούν σε ΔΕ, αλλά είναι λιγότερο πιθανό να ολοκληρώσουν την εκπαίδευση και να λάβουν πιστοποίηση. Από το γεγονός αυτό δύναται να εκμαιευτεί το συμπέρασμα ότι στις πυκνοκατοικημένες περιοχές οι ΔΕ ήταν πιο διαδεδομένες, όμως πολλοί από τους συμμετέχοντες, όντας εξοικειωμένοι με αυτή τη μορφή εκπαίδευσης, δεν ενδιαφέρονταν για τη λήψη πιστοποιητικών παρακολούθησης. Από την άλλη πλευρά, οι κάτοικοι αγροτικών περιοχών ήταν εκείνοι που επιδίωκαν τη λήψη πιστοποιητικού.

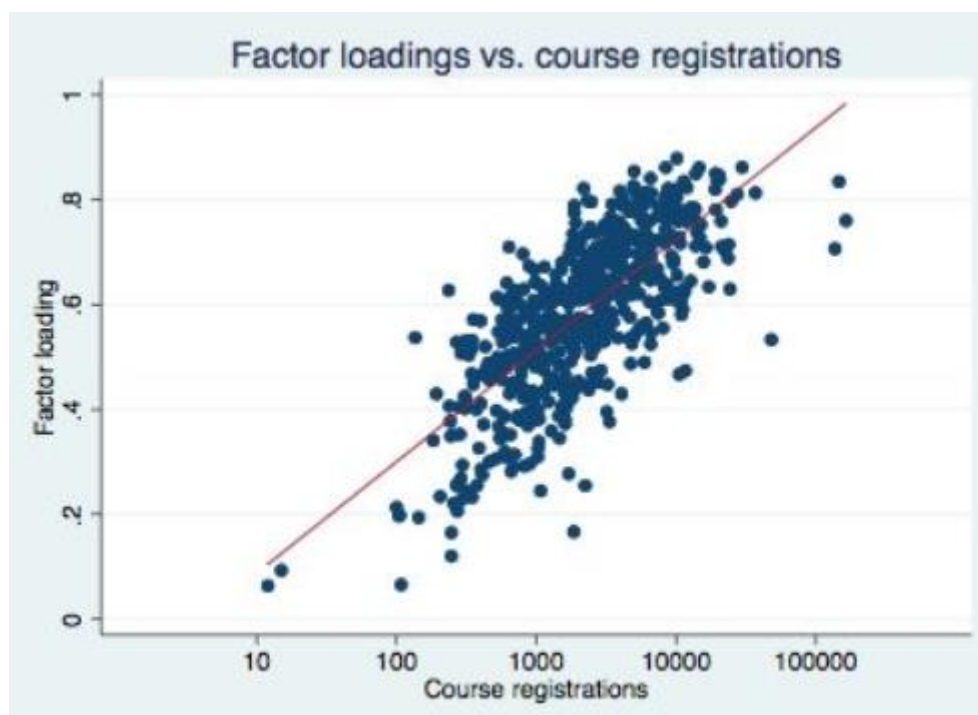
Για την καλύτερη ερμηνεία οικονομικών μεγεθών σε σχέση με τους ΤΚ, εφαρμόστηκε η μέθοδος Factor Analysis στους ΤΚ οι οποίοι ερμηνεύτηκαν ως άτομα, οι ΔΕ ως αντικείμενα και οι εγγραφές ανά εκατομμύριο ερμηνεύτηκαν ως επιδόσεις. Από την μελέτη των Ganelin κ.α. η Cronbach' a έδειξε αξιοπιστία των δεδομένων της τάξης του 0,98, γεγονός που υποδηλώνει αξιόπιστα αποτελέσματα. Επιπρόσθετα η Ανάλυση Παραγόντων ανέδειξε τη μονοδιάστατη απεικόνιση της καθώς ένας και μόνο παράγοντας παρουσίασε υψηλή διακύμανση των μεταβλητών, όπως φαίνεται και στο γράφημα 4.



Γράφημα 4. Απεικόνιση των eigenvalues μετά από εφαρμογή μεθόδου ανάλυσης παραγόντων για όλα τα δεδομένα συνολικά. Επικρατεί ένας παράγοντας. © Ganelin. κ.α., 2019.

Η μονοδιάστατη απεικόνιση των δεδομένων επαληθεύεται και από το Γράφημα 5. Στο γράφημα αυτό εμφανίζεται η υψηλή συσχέτιση του πρώτου παράγοντα φορτίου με το λογάριθμο των συμμετοχών σε προγράμματα εκπαίδευσης. Συνεπώς σε όσα περισσότερα προγράμματα ΔΕ συμμετείχε ένα άτομο, τόσες περισσότερες

πληροφορίες αντλούνταν για τους ΤΚ. Συνολικά με τη χρήση ενός και μόνο παράγοντα μπορούν να ερμηνευτούν το 46% της διακύμανσης των μεταβλητών.



Γράφημα 5. Παράγοντας φορτίου και του λογάριθμου των συμμετοχών σε ΔΕ. © Ganelin κ.α.,2019.

Επιχειρώντας να προσδιοριστούν επιπλέον στοιχεία για τους ΤΚ και τα μαθήματα, η μελέτη επεκτάθηκε και στην καταγραφή του πρώτου παράγοντα με το λογάριθμο των κατά κεφαλήν εγγραφών σε ΔΕ. Μεταξύ αυτών των δύο μεγεθών, παρουσιάστηκε υψηλή συσχέτιση. Συνεκτιμώντας την υψηλή συσχέτιση με τον παράγοντα φορτίου και των οικονομικών μεταβλητών, παρατηρήθηκε ότι η συμπεριφορά αυτών των μεγεθών είναι όμοια. Λαμβάνοντας υπόψη και τη μονοδιάστατη φύση των μεγεθών, συμπεραίνεται ότι η μέθοδος ανάλυσης Παραγόντων δεν προσδίδει επιπρόσθετη πληροφορία πέραν του υπολογισμού των ΤΚ και των συμμετοχών σε ΔΕ ως ανεξάρτητα μεγέθη.

3.11 Παρατηρήσεις Ευρημάτων Ανάλυσης

Με την εφαρμογή της μεθόδου ανάλυσης παραγόντων για την μελέτη της συμπεριφοράς της συμμετοχής σε προγράμματα ΔΕ ανά ΤΚ διαπιστώθηκε ότι αν μελετηθούν τα ποσοστά συμμετοχής, ολοκλήρωσης και πιστοποίησης σε ΔΕ, παρουσιάζεται ελάχιστη διαφορά, ως προς τις οικονομικές μεταβλητές, όσον αφορά τα δεδομένα ανά ποσοστό πληθυσμού και σε αναλογία με τους ΤΚ (Ganelin κ.α.,2019). Από τη σύγκριση των οικονομικών μεγεθών που ήταν ήδη γνωστά και σε σύγκριση με

τα αντίστοιχα δεδομένα καταχωρήσεων σε προγράμματα ΔΕ, δεν παρατηρήθηκε ανά πληθυσμό σημαντική διαφοροποίηση των οικονομικών μεταβλητών, με αποτέλεσμα τα δεδομένα της μελέτης να μην να προσφέρουν κάποια δυνατότητα πρόβλεψης των οικονομικών μεταβλητών ανά ΤΚ. Παρόλα αυτά, η καταγραφή των στοιχείων της μελέτης των Ganelin κ.α., μπορούν να χρησιμοποιηθούν σε μελλοντική μελέτη νέων ΤΚ των οποίων υπάρχουν ελλειπείς ή και καθόλου πληροφορίες οικονομικών στοιχείων. Η υψηλή συσχέτιση, ο βαθμός μονοδιάστασης των δεδομένων και η συσχέτιση των παραγόντων φορτίου (factor loadings) και επιδόσεων των μετρήσεων των εγγραφών, δηλώνουν ότι τα ατομικά προγράμματα εκπαίδευσης τείνουν να συμπεριφερθούν όπως και η γεωγραφική τους ταξινόμηση. Συγκεκριμένα, κάποιες ΔΕ εμφανίζονται περισσότερο σε συγκεκριμένες περιοχές, δηλώνοντας ότι αναγνωρίζονται και επιλέγονται από συμμετέχοντες με όμοια οικονομικά κριτήρια. Το γεγονός όμως αυτό, αν και ερμηνεύει μια τάση, δεν επικρατεί στη μελέτη, καθώς σε μεγάλο αριθμό ΔΕ δεν αναγράφεται ο ΤΚ των συμμετεχόντων.

Στην προσπάθεια εύρεσης πεπατημένης οι Ganelin κ.α. ομαδοποίησαν τους ΤΚ στις περιπτώσεις όπου ήταν ελλειπείς. Από την ομαδοποίηση των δεδομένων παρουσιάστηκε μια ελαφριά τάση όπου οι ΔΕ επιμόρφωσης στελεχών και κλασικών μελετών, ήταν πιο ελκυστικές και παρουσίαζαν αυξημένη ζήτηση σε πιο δυσχερείς οικονομικά περιοχές. Αντίθετα στις πιο ευημερείς περιοχές, παρουσιάστηκε αυξημένη τάση συμμετοχής σε προγράμματα που αφορούσαν επιστήμες υπολογιστών (Ganelin, κ.α., 2019). Εν τέλει παρατηρήθηκε υψηλή συσχέτιση μεταξύ των οικονομικών παραγόντων των ΤΚ (factors) και των καταχωρήσεων στα αντίστοιχα προγράμματα ΔΕ.

3.12 Εξέταση Ακρίβειας των Μετρήσεων του Geolocation

Ερμηνεύοντας τα παραπάνω δεδομένα, σε αυτήν την ενότητα, επιχειρείται να εξεταστεί η ακρίβεια των διευθύνσεων IP ως προς την τοποθεσία τους. Επισημαίνεται ότι για τη MaxMind θεωρείται σε φυσιολογικά πλαίσια η ύπαρξη ανακρίβειας της τοποθεσίας, καθώς η διαδικασία γεωεντοπισμού, πρόκειται ως επί το πλείστο για μια «θορυβώδη» διαδικασία με χαμηλή ακρίβεια. Για παράδειγμα η MaxMind, έχοντας εντοπίσει την τοποθεσία ενός χρήστη σε μια περιοχή με ακτίνα 20 μιλίων, προσδίδει μια τυχαία τοποθεσία μέσα σε αυτήν την εμβέλεια. Για τη μελέτη τους οι Ganelin κ.α., θεώρησαν αμελητέο αυτό το σφάλμα, υπολογίζοντας την τοποθεσία που προσδιόριζε η MaxMind ως ακριβή.

Παρόλα αυτά, παραμένει ένας άλλος σημαντικός παράγοντας ο οποίος πρέπει να συνεκτιμηθεί. Μέσω της MaxMind, η ακρίβεια της τοποθεσίας δύναται να επηρεαστεί από τους οικονομικούς παράγοντες. Σε αγροτικές περιοχές που είναι πιο αραιοκατοικημένες, ο γεωεντοπισμός εμφανίζει μεγαλύτερα σφάλματα από ότι στις αστικές περιοχές, γεγονός που οφείλεται καθαρά στην ύπαρξη πιο σύγχρονων υποδομών όσον αφορά το κομμάτι επικοινωνιών (δίκτυα επικοινωνιών, οπτικές ίνες, κλπ). Από τους Shavitt & Zilberman έχει επίσης μελετηθεί η ακρίβεια γεωεντοπισμού της MaxMind, αναφέροντας λανθασμένες εκτιμήσεις της βάσης όσον αφορά την τοποθεσία χρηστών που βρίσκονταν στις ΗΠΑ και στην πολιτεία της Ουάσιγκτον (Shavitt, Zilberman, 2011).

Αξίζει να σημειωθεί ότι οι ανωτέρω εσφαλμένες ενδείξεις, μειώνουν την βαρύτητα της ανάλυσης που περιεγράφηκε στην προηγούμενη ενότητα, καθώς τα δεδομένα γεωεντοπισμού ενδέχεται να είναι ανακριβή. Συνεπώς, παρόμοιες μελέτες εξαγωγής συμπερασμάτων οικονομικού χαρακτήρα, που έχουν πραγματοποιηθεί με βάση το γεωεντοπισμό, εγκυμονούν κινδύνους εσφαλμένων συμπερασμάτων (Alcorn, & Christensen, 2015). Ως εκ τούτου, πρόκειται για μια νέα μέθοδο εξαγωγής διαφόρων ειδών πληροφοριών, με ιδιαίτερα χαμηλό κόστος, τα οποία μπορούν να χρησιμοποιηθούν για περεταίρω μελέτη. Παρόλα αυτά απαιτείται επιπρόσθετη μελέτη για την επαλήθευση ακρίβειας της μεθόδου του γεωεντοπισμού.

Μια αρχική προσέγγιση της επαλήθευσης των πραγματικών δεδομένων τοποθεσίας, είναι η διασταύρωση των στοιχείων TK αντιστοιχίζοντας με την κάθε ταχυδρομική διεύθυνση του συμμετέχοντα σε ΔΕ. Η αντιστοιχία όμως αυτή, δεν είναι πάντα εφικτή, καθώς δεν δηλώνεται από όλους τους υποψηφίους η ταχυδρομική διεύθυνση. Στις περιπτώσεις όμως που αυτό είναι δυνατό, εξάγονται πληροφορίες όπως:

- Επαλήθευση των δεδομένων TK (από τις φυσικές διευθύνσεις) και του ποσοστού συμμετοχής σε προγράμματα εκπαιδεύσεων ανά περιοχή.
- Έλεγχος και επαλήθευση των σφαλμάτων που εμφανίζεται στη διαφορά της διαδικασίας εντοπισμού θέσης, συγκρίνοντας τα δεδομένα από πραγματικές διευθύνσεις και από την IP του κάθε χρήστη.
- Διαφοροποίηση του οικονομικού υπόβαθρου των δεδομένων της MaxMind και των πραγματικών δεδομένων των TK.

3.13 Προσέγγιση Διαδικασίας Γεωεντοπισμού

Σε αρχικό στάδιο έγινε μια εκκαθάριση των καταχωρήσεων που είχε γίνει εγγραφή εκτός ΗΠΑ και αυτών που παρουσίαζαν ελλιπή δεδομένα ως προς την ταχυδρομική τους διεύθυνση. Με αυτόν τον τρόπο από την έρευνα των Ganelin κ.α. εξαιρέθηκαν 119.000 καταχωρήσεις. Μέσω της Python Library έγινε ένας αρχικός διαχωρισμός των δεδομένων εξαιρώντας τις διευθύνσεις με εσφαλμένα στοιχεία (Deng κ.α., 2014). Επιπρόσθετα για κάθε μια καταχώρηση αναλύθηκε η ταχυδρομική διεύθυνση, η πόλη και ο TK της περιοχής.

Στη συνέχεια με χρήση της εφαρμογής Google Maps Geocoding API, προσδιορίστηκε το γεωγραφικό μήκος και πλάτος για κάθε μια καταχώρηση, στο βαθμό που αυτό είναι εφικτό. Η διαδικασία αυτή υλοποιείται μέσω της εξής διεργασίας:

- Χρήση των δεδομένων για τα οποία οι συντεταγμένες είναι ακριβείς και σύμφωνα με τα πρότυπα της Google χαρακτηρισμένα ως “range interpolated” (Google developers, 2019).
- Αποκλεισμός των διευθύνσεων που είναι καταχωρημένες και αναφέρονται εκτός των ΗΠΑ. Στη μελέτη των Ganelin κ.α. οι εν λόγω διευθύνσεις αφορούσαν το 29% των περιπτώσεων.
- Χρήση των TK και των συντεταγμένων στις περιπτώσεις που υπάρχει ταυτοποίηση της τοποθεσίας.
- Στις περιπτώσεις όπου υπάρχει διαφορά στους TK της ανάλυσης (Python) και στα αποτελέσματα της Google, χρησιμοποιείται ο TK της ανάλυσης. Στην υπό ανάλυση μελέτη το ποσοστό των περιπτώσεων αυτών ήταν 3%.

Με βάση τα παραπάνω καταγράφηκαν συνολικά 79.000 TK και κατ’ επέκταση 79.000 συμμετοχές σε προγράμματα διαδικτυακών εκπαιδεύσεων. Επιπρόσθετα από τις ανωτέρω καταγραφές, οι 3.000 εξ αυτών είχαν εγγραφεί με εταιρικό προφίλ, συνεπώς δεν ήταν δυνατή η άντληση δεδομένων των πραγματικών τους TK. Για το λόγο αυτό εξαιρέθηκαν από τη μελέτη. Τέλος στην ανάλυση των δεδομένων τους, οι Ganelin κ.α χρησιμοποίησαν τους TK για τους οποίους διέθεταν γεωγραφικές συντεταγμένες, ήτοι 73.000 καταγραφές.

3.14 Υπολογισμός Σφάλματος Γεωεντοπισμού

Για τον υπολογισμό ψευδών δεδομένων, πραγματοποιήθηκε διασταύρωση των στοιχείων με τη βιβλιοθήκη της Shapely, για όσες διευθύνσεις διέθεταν TK. Η Shapely είναι μια βάση δεδομένων που διαθέτει τη βιβλιοθήκη της Python και χρησιμοποιείται για τη γεωμετρική μέτρηση αντικειμένων. Μέσω της διαδικασίας υπολογισμού της ευκλείδειας απόστασης των πραγματικών συντεταγμένων της καταχώρησης και του κοντινότερου σημείου της IP διεύθυνσης της ίδιας καταχώρησης, η οποία βρέθηκε από τη MaxMind, υπολογίζεται το σφάλμα των δεδομένων (Gillies, 2019.). Συνεπώς, στην περίπτωση που οι δύο συντεταγμένες ταυτίζονται (Google API και MaxMind), τότε η απόσταση είναι μηδενική. Σε περίπτωση που υπάρχει διαφορά, η απόσταση διαφοροποιείται και παραμετροποιείται με βάση έναν πίνακα. Για παράδειγμα σε απόκλιση μιας μοίρας, η απόσταση ερμηνεύεται σε απόσταση 50 μιλίων.

3.15 Ανάλυση Παρατηρήσεων Οικονομικού Χαρακτήρα

Για την ταξινόμηση των TK ανάλογα του οικονομικού υπόβαθρου των υποψηφίων, οι Ganelin κ.α. χρησιμοποίησαν τα δεδομένα του δείκτη οικονομικής εξουθένωσης για τους συμμετέχοντες σε ΔΕ στις ΗΠΑ. Από την αντιστοιχία αυτή, υπολογίστηκε ότι το 28% των TK χαρακτηρίστηκε ως ευημερείς, το 22% οικονομικά άνετοι, το 18% ως μικρομεσαίοι, 17% ως οριακά οικονομικά και το 15% ως οικονομικά εξουθενωμένοι.

Για την σύγκριση των ορθών δεδομένων τοποθεσίας και των αντίστοιχων που προέκυψαν από την εύρεση της τοποθεσίας της IP, χρησιμοποιήθηκε περιγραφική (descriptive) ανάλυση των δεδομένων, ενώ στη συνέχεια χρησιμοποιήθηκε το μοντέλο της γραμμικής παλινδρόμησης, για την παρατήρηση των διακυμάνσεων των αλληλεπιδράσεων μεταξύ των ιδιοτήτων των TK, την ακρίβεια του γεωεντοπισμού και την ταυτοποίηση των TK.

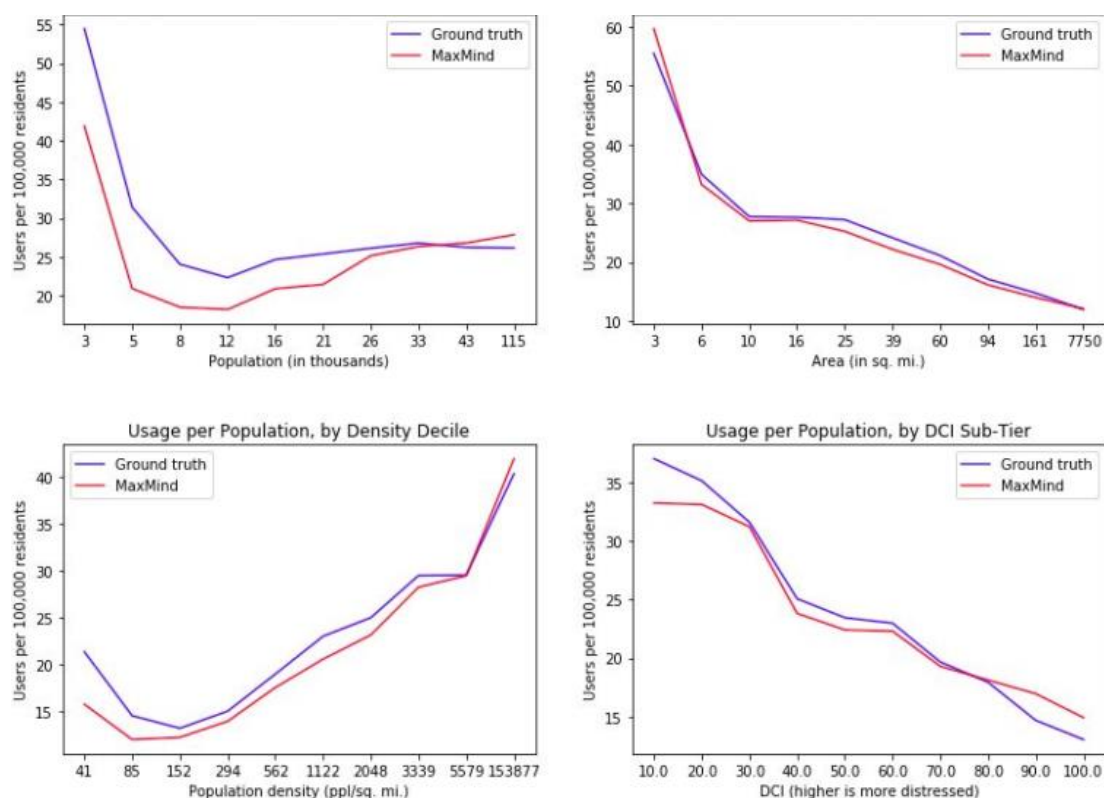
3.16 Αξιολόγηση Μετρήσεων Αξιοπιστίας

Από την αντιστοίχιση των δεδομένων προέκυψαν τα ακόλουθα δεδομένα. Με βάση τις τοποθεσίες από τα δεδομένα που εισήγαγαν οι χρήστες, οι 7 από τις 10 περιοχές με τις περισσότερες κατά κεφαλήν συμμετοχές σε προγράμματα εκπαίδευσης, ήταν από την πολιτεία της Μασαχουσέτης. Το γεγονός αυτό δεν μας εκπλήσσει, καθώς στην περιοχή αυτή βρίσκονται πολλά πανεπιστήμια και εκπαιδευτικά ιδρύματα. Συνεπώς είναι λογικό επακόλουθο η υψηλή αναλογία συμμετοχών σε ΔΕ ανά κάτοικο (18,2 συμμετοχές ανά 1000 κατοίκους).

Όσον αφορά τα δεδομένα που συλλέχθηκαν από τη MaxMind, προέκυψε ότι από τις 10 πρώτες περιοχές με τις περισσότερες συμμετοχές, μόνο οι 4 φαίνονταν να είναι από την πολιτεία της Μασαχουσέτης. Οι δύο περιοχές με τις περισσότερες συμμετοχές εμφανίζονταν στο Ιλινόις, ενώ η πρώτη εμφανιζόταν στο Σικάγο με ποσοστό συμμετοχής ανά ΔΕ 12% του συνολικού πληθυσμού (Ganelin,2019).

Όσον αφορά την κατανομή των TK με βάση τις πραγματικές τοποθεσίες, παρατηρήθηκε ότι το 41% των συμμετεχόντων ήταν ευημερείς, το 23% άνετοι, 16% μικρομεσαίοι, 12% ήταν σε οριακή κατάσταση και 7% ήταν εξουθενωμένοι οικονομικά.

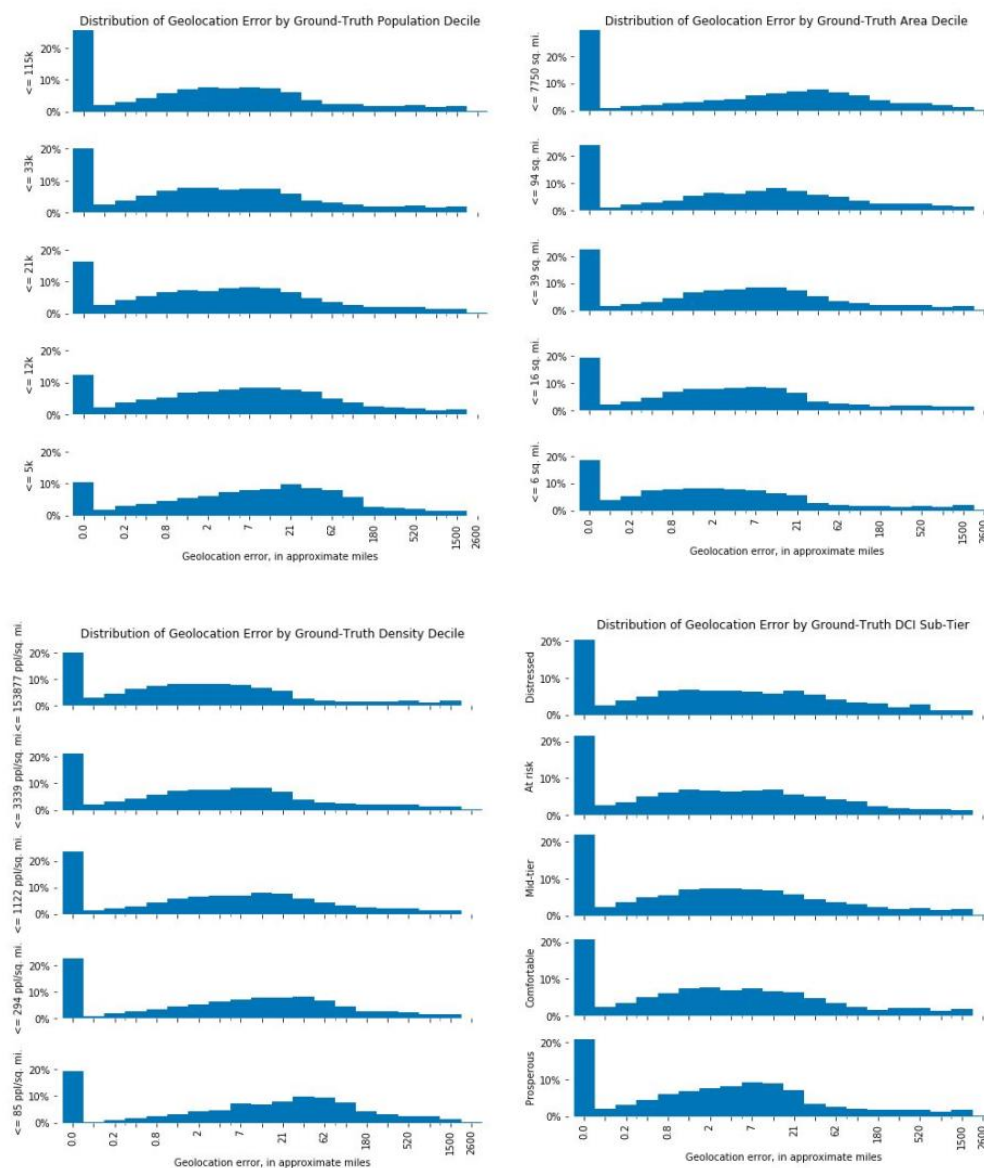
Στην εικόνα 2 απεικονίζονται τα αποτελέσματα της σύγκρισης της μελέτης των Ganelin κ.α. Η πλειοψηφία των περιπτώσεων περιέχει χρήστες οι οποίοι προέρχονταν από TK με μεγαλύτερη πυκνότητα πληθυσμού και χαμηλό δείκτη οικονομικής εξουθένωσης, ήτοι οικονομικά ευημερείς πολίτες.



Εικόνα 2.Γραφήματα σύγκρισης του αριθμού χρηστών σε ΔΕ ανά TK με βάση την MaxMind και τις πραγματικές τοποθεσίες, ταξινομημένα ανάλογα το οικονομικό τους προφίλ, την περιοχή και την πυκνότητα πληθυσμού. © Ganelin κ.α.,2019.

Συνεκτιμώντας και τα δεδομένα σφάλματος προέκυπτε ότι για το 18% των χρηστών της MaxMind και εκείνων που προέκυψαν από την πραγματική τοποθεσία

συμπίπτουν. Επισημαίνεται ότι στη MaxMind είναι λιγότερο πιθανό να εμφανίσει μεγάλο σφάλμα σε περιοχές με μεγάλη πυκνότητα αλλά είναι πιο πιθανό να γίνει λάθος όσον αφορά το δείκτη οικονομικής εξουθένωσης. Ως εκ τούτου η MaxMind, είναι πιο πιθανό να προβλέψει ορθά την τοποθεσία μιας διεύθυνσης IP σε μια πυκνοκατοικημένη περιοχή, ή αντίστοιχα να προβλέψει ένα μεγαλύτερο σφάλμα στην περίπτωση που αποτύχει να εντοπίσει ένα σύνολο διευθύνσεων IP από μια πολυπληθή περιοχή (Ganelin,2019). Αναλυτικά η μέτρηση του σφάλματος αναλύεται στην εικόνα 3 όπου μπορεί να διακριθεί ότι όσο μειώνεται η πυκνότητα και ο πληθυσμός, το σφάλμα εντοπισμού μεγαλώνει.

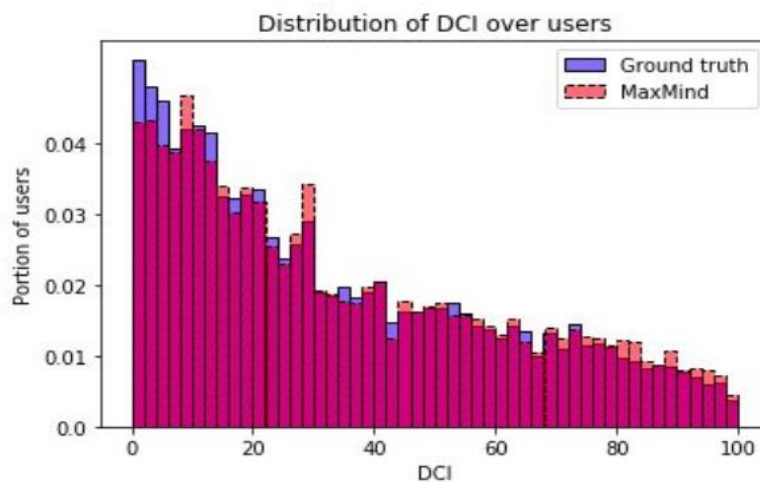


Εικόνα 3.Γραφήματα σφαλμάτων εντοπισμού της MaxMind ερμηνευμένο σε λογαριθμική κλίμακα. © Ganelin κ.α.2019.

Από τη μέθοδο γραμμικής παλινδρόμησης όπου επιχειρήθηκε να υπολογιστεί η πιθανότητα σωστής πρόβλεψης της τοποθεσίας ενός χρήστη από τη MaxMind, διαπιστώθηκε πως για τους συμμετέχοντες που κατοικούν σε ευημερείς περιοχές υπήρχε 59% πιθανότητα να γίνει ορθή πρόβλεψη της τοποθεσίας τους, ενώ ένας χρήστης από μια εξουθενωμένη οικονομικά περιοχή είχε 38% πιθανότητες ορθής πρόβλεψης.

3.17 Συνεκτίμηση των DCI δεικτών

Από τον υπολογισμό των ανωτέρω μεγεθών, μπορεί να εξαχθεί το συμπέρασμα ότι με τη διαδικασία εντοπισμού θέσης μέσω IP, παρατηρήθηκε «εσφαλμένη» εικόνα των υπό μελέτη μεγεθών. Συγκεκριμένα, οι Ganelin κ.α. μέσω των πραγματικών τοποθεσιών των χρηστών που συμμετείχαν σε ΔΕ αναλογούσαν 36,2 χρήστες ανά 100.000 πληθυσμού στην κατηγορία των ευημερών ανθρώπων και 14 χρήστες ανά 100.000 για τους οικονομικά εξουθενωμένους. Αντίστοιχα με τη μέθοδο εντοπισμού θέσης από την IP (MaxMind) ο αριθμός αυτός υπολογίζεται σε 33,2 για τους ευημερείς και 16,1 για τους οικονομικά εξουθενωμένους. Επιπρόσθετα από τις εξισώσεις παλινδρόμησης προέκυψε ότι μια αύξηση στην τιμή του δείκτη DCI κατά 10 μονάδες, συνεπαγόταν 0,47 λιγότερους χρήστες σύμφωνα με τα πραγματικά δεδομένα τοποθεσίας και 0,35 λιγότερους χρήστες σύμφωνα με τη MaxMind. Όπως φαίνεται και στην γράφημα 6, οι ανωτέρω παρατηρήσεις οδηγούν στην κατανομή της MaxMind να είναι λιγότερο κυρτή προς τα δεξιά σε σχέση με τα δεδομένα πραγματικής τοποθεσίας των χρηστών. Με αυτόν τον τρόπο η MaxMind απεικονίζει περισσότερους χρήστες από οικονομικά εξουθενωμένες περιοχές σε σχέση με τα δεδομένα πραγματικής τοποθεσίας (Ganelin,2019).



Με βάση τα παραπάνω προκύπτει ότι εάν επιλεγθεί ένας χρήστης τυχαία, η MaxMind θα δώσει κατά πάσα πιθανότητα ορθή εκτίμηση της τοποθεσίας του εάν ο χρήστης ανήκει στους ευημερείς κατοίκους και είναι πιθανότερο να σφάλει ως προς την τοποθεσία του χρήστη εάν αυτός ανήκει στους οικονομικά εξουθενωμένους κατοίκους. Επιπρόσθετα από το γράφημα 6, παρατηρείται ότι η MaxMind αν και είναι λιγότερο πιθανό να σφάλει για έναν χρήστη που ανήκει στην κατηγορία των ευημερών σε σχέση με κάποιον που ανήκει στους οικονομικά εξουθενωμένους, συνολικά εμφανίζει περισσότερα σφάλματα στους ευημερείς από ότι στους οικονομικά εξουθενωμένους. Το γεγονός αυτό επαληθεύει την αρχική εκτίμηση ότι η MaxMind και κατ' επέκταση όλες οι βάσεις δεδομένων που χρησιμοποιούνται για τον εντοπισμό IP διευθύνσεων, είναι πιο πιθανό να σφάλουν ως προς την ακρίβεια εντοπισμού τοποθεσιών που είναι οικονομικά ευημερείς. Ο λόγος για τον οποίο συμβαίνει κάτι τέτοιο οφείλεται στο ότι οι ευημερείς χρήστες που συμμετέχουν σε προγράμματα ΔΕ είναι πιο πιθανό να συμμετέχουν σε αυτά, συνεπώς το δείγμα μελέτης αυξάνεται και κατ' επέκταση αυξάνονται και οι πιθανότητες εμφάνισης σφαλμάτων.

3.18 Σύνοψη των Παρατηρήσεων

Με βάση τα όσα έχουν αναφερθεί παραπάνω, προκύπτει εύλογα το συμπέρασμα ότι λαμβάνοντας υπόψη αποκλειστικά και μόνο δεδομένα τοποθεσίας μέσω διευθύνσεων IP, είναι πιθανό να παρερμηνευθούν τα δεδομένα και να προκύψουν εσφαλμένες εκτιμήσεις. Συγκεκριμένα στις πιο εύρωστες οικονομικά περιοχές προκύπτουν αποτελέσματα αρκετά κοντά με την πραγματικότητα, όμως δεν συμβαίνει κάτι αντίστοιχο με τις λιγότερο οικονομικά ανεπτυγμένες περιοχές. Ως εκ τούτου οποιεσδήποτε μελέτες βασιστούν αποκλειστικά και μόνο σε δεδομένα IP διευθύνσεων για την εύρεση τοποθεσίας (σε επίπεδο TK), μπορεί να οδηγήσει σε ελλιπή συμπεράσματα.

Συναφώς, με βάση τα παραπάνω, παρατηρείται ότι η διαδικασία εντοπισμού μέσω IP διεύθυνσης δύναται να χρησιμοποιηθεί ως μέσο για έρευνα και μελέτη, στις περιπτώσεις όπου δεν υφίσταται η δυνατότητα καταγραφής της ακριβούς τοποθεσίας ή αν το κόστος εύρεσης της φυσικής διεύθυνσης είναι υπέρογκο. Παρόλα αυτά, πρέπει να επισημανθεί ότι τα δεδομένα που μπορούν να εξαχθούν από την τοποθεσία της IP, είναι επισφαλής και σε καμία περίπτωση δεν θεωρούνται έγκυρα για την καταγραφή και

συνέχιση περεταίρω ενεργειών. Με τον εντοπισμό θέσης μέσω IP διεύθυνσης, μπορεί να γίνει μια αρχική βολιδοσκόπηση της κατάστασης του προς παρατήρηση αντικειμένου, για να εκτιμηθεί μια κατάσταση και να δοθεί μια αρχική εικόνα στην παρατήρηση του εκάστοτε φαινομένου. Για την συνέχιση όμως της μελέτης και την καταγραφή στοιχείων όπου θα οδηγήσουν στην εξαγωγή συμπερασμάτων και στη λήψη απόφασης, θα πρέπει να γίνει καταγραφή των φυσικών διευθύνσεων των υπό μελέτη στοιχείων.

Συνεπώς για την μελέτη και εκτίμηση οικονομικών στοιχείων, η αποκλειστική καταγραφή των δεδομένων τοποθεσίας μέσω IP, δεν ενδείκνυται, καθότι είναι επίφοβο ο κάθε χρήστης να χρησιμοποιεί ψευδή στοιχεία σύνδεσης ή ακόμα και ο πάροχος δικτύου να αλλοιώνει τα δεδομένα για διευκόλυνση του δικτύου. Για παράδειγμα στις χώρες με προηγμένες υποδομές στις δυνατότητες δικτύου, τα δεδομένα δύναται να είναι πιο ακριβή σε σχέση με χώρες που ο εξοπλισμός δικτύου δεν είναι τόσο εκσυγχρονισμένος (If-So Dynamic Content, 2022).

Σε κάθε περίπτωση για την καταγραφή και μελέτη επιπλέον οικονομικών μεγεθών και στοιχείων με βάση την τοποθεσία κάποιου χρήστη, η διαδικασία εντοπισμού μέσω IP, θα πρέπει να χρησιμοποιείται ως επικουρικό μέσο και όχι ως κύριο, καθότι τα ποσοστά ακρίβειας της πραγματικής τοποθεσίας είναι περιορισμένα. Ένας τρόπος με τον οποίο θα μπορούσε να βελτιωθεί η ακρίβεια των δεδομένων, είναι η παροχή πληροφοριών τοποθεσίας από τους παρόχους δικτύου. Για μια μελέτη η οποία χρησιμοποιεί δεδομένα αντλούμενα από τις βιβλιοθήκες των παρόχων, με τις διευθύνσεις των πελατών δύναται να βελτιώσει σημαντικά την ακρίβεια των δεδομένων και κατ' επέκταση την εγκυρότητα της εκάστοτε μελέτης.

Κεφάλαιο Τέταρτο: Ο Γεωεντοπισμός IP βασισμένο σε συστήματα GNSS (Global Navigation Surveillance Systems)

4.1 IP Γεωεντοπισμός και τα GNSS

Σε αυτό το σημείο της εργασίας, γίνεται η αναφορά και περιγραφή μιας διαφορετικής προσέγγισης εκτίμησης της ακρίβειας του γεωεντοπισμού IP διευθύνσεων. Η νέα αυτή προσέγγιση αποσκοπεί στον εντοπισμό της τοποθεσία ενός χρήστη με βάση τα συστήματα GNSS.

4.2 Τα Συστήματα GNSS

Τα συστήματα GNSS αναφέρονται σε ένα σύνολο δορυφόρων που σκοπό έχουν την παροχή πληροφοριών τοποθεσίας και δεδομένων χρόνου στους δέκτες τους. Τα συστήματα αυτά παρέχουν παγκόσμια κάλυψη και διαφοροποιούνται ανάλογα την εκάστοτε τεχνολογία. Τα κυριότερα συστήματα GNSS που χρησιμοποιούνται σήμερα είναι: το Global Positioning System (GPS), το Galileo, το GLONASS και το BEI DOU. (What is GNSS? | EU Agency for the Space Programme, 2022). Η απόδοση των συστημάτων αυτών βασίζεται στα εξής χαρακτηριστικά:

- Ακρίβεια
- Αξιοπιστία
- Απόδοση
- Διαθεσιμότητα

Για την βελτίωση της απόδοσης του συστήματος χρησιμοποιούνται επίγειοι σταθμοί οι οποίοι υπολογίζουν τη διαφορά σε σφάλματα του λαμβανόμενου σήματος και παρέχουν πληροφορίες διόρθωσης του σήματος για την κάλλιστη ακρίβεια. Για παράδειγμα το επίγειο σύστημα βελτίωσης του συστήματος GALILEO είναι το European Geostationary Navigation Overlay Service (EGNOS).

4.3 Το GPS-based IP Geolocation

Στη μελέτη των Saxon και Feamster, γίνεται η χρήση πληροφοριών δεδομένων τοποθεσίας IP διευθύνσεων των κινητών τηλεφώνων των χρηστών με χρήση αντίστοιχης εφαρμογής. Στην ανωτέρω μελέτη χρησιμοποιήθηκαν δύο βάσεις δεδομένων που λειτουργούν με βάση το στίγμα ενός χρήστη από το GPS για τον προσδιορισμό της θέσης και δύο συμβατικές βάσεις, που στηρίζονται στις «κλασικές» μεθόδους εντοπισμού που αναφέρθηκαν σε προηγούμενα κεφάλαια της παρούσας εργασίας.

4.4 Η Βάση Δεδομένων Unacast

Η Unacast είναι μια βάση η οποία χρησιμοποιεί τις τοποθεσίες των κινητών τηλεφώνων (smartphones) για την καταγραφή της τοποθεσίας τους. Στην ουσία συλλέγει δεδομένα από τις κεραίες των συσκευών και σε συνδυασμό με επιπρόσθετες πληροφορίες-δεδομένα από διάφορες εταιρίες, αποτυπώνουν την τοποθεσία ενός χρήστη-διεύθυνσης IP. Η παρεχόμενη πληροφορία επικυρώνεται μέσω μιας πολυσύνθετης διαδικασίας

επεξεργασίας των δεδομένων (Saxon κ.α., 2021). Από το σύνολο των διευθύνσεων IP που ελέγχεται και συλλέγεται από τη Unacast, το 99,6% έχουν τη μορφή IPv4.

Για την εκπόνηση της μελέτης χρησιμοποιήθηκαν δεδομένα σε μια ακτίνα εμβέλειας 40 μιλίων τριών κύριων πόλεων των ΗΠΑ: της Νέας Υόρκης, του Σικάγο και της Φιλαδέλφειας. Μέσα σε αυτό το μεγάλο δείγμα, εμπεριέχονται δεδομένα τόσο από το αστικό, όσο και από το αγροτικό περιβάλλον. Επιπρόσθετα τα δεδομένα που συλλέχθηκαν αφορούσαν δύο περιόδους: η πρώτη ήταν η περίοδος Αυγούστου-Οκτωβρίου 2020 και η δεύτερη ήταν η περίοδος Απριλίου του 2021. Η δεύτερη περίοδος συνέπιπτε με δεδομένα που αντλήθηκαν την ίδια περίοδο και από διαφορετικές βάσεις δεδομένων και βοηθούσαν στη συλλογή δεδομένων και την αξιολόγηση των αποτελεσμάτων. Συνολικά συλλέχθηκαν δεδομένα από 248 εκατομμύρια IP διευθύνσεις. Ο διάμεσος του πρώτου δείγματος εμφάνιζε ακρίβεια της τάξης των 17 μέτρων, ενώ του δεύτερου ακρίβεια της τάξης των 11 μέτρων (Saxon κ.α., 2021). Στην υπόψη μελέτη υπολογίστηκε και η χρήση κινητών συσκευών κατά τις νυχτερινές ώρες κατά την οποία παρατηρήθηκε ότι η αναλογία συσκευών ανά πληθυσμό αντιστοιχεί σε μία συσκευή ανά πέντε ανθρώπους.

4.5 Η Βάση Ookla Speedtest

Για τον εντοπισμό τοποθεσίας κινητών συσκευών, χρησιμοποιήθηκαν και τα δεδομένα από τη βάση δεδομένων της Ookla. Πρόκειται για μια βάση η οποία παρέχει στους χρήστες υπηρεσίες μέτρησης της ταχύτητας τους στο διαδίκτυο. Επιπρόσθετα η βάση για κάθε IP που καταχωρείται, εντοπίζει την τοποθεσία του χρήστη με βάση το GPS. Είναι μια βάση η οποία παρέχει πληροφορίες, με το μειονέκτημα όμως ότι τα δεδομένα είναι εμφανώς λιγότερα σε σχέση με τη UNACAST. Συγκεκριμένα για την περίοδο που αφορά στις μετρήσεις της μελέτης, από τα 4 εκατομμύρια IP που συλλέχθηκαν, μόνο οι 270 χιλιάδες συνέπιπταν, ενώ για την ασφάλεια των δεδομένων των χρηστών παρέχονταν πληροφορίες μόνο των πρώτων τριών bytes (1 byte ισούται με 8 bits δεδομένων) της IP (Saxon κ.α., 2021).

4.6 Δεδομένα Σύγκρισης IP

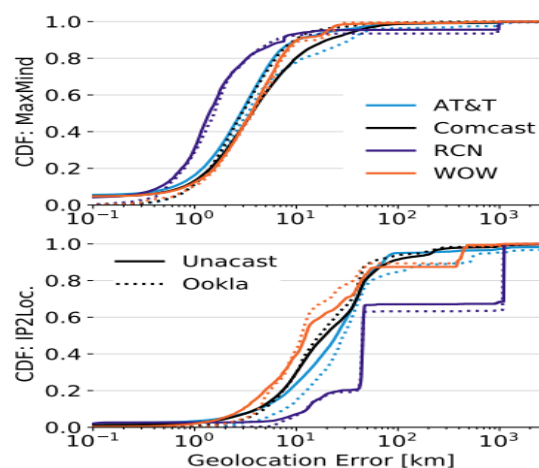
Για τη συλλογή των δεδομένων τοποθεσίας χρηστών και για τη σύγκριση τους με τις αντίστοιχες διευθύνσεις που εντοπίστηκαν με τις παραπάνω μεθόδους, χρησιμοποιήθηκαν οι βάσεις MaxMind και IP2Location, τόσο οι δωρεάν εκδόσεις τους, όσο και οι υπό πληρωμή εκδόσεις. Για την πραγματοποίηση της σύγκρισης και

την εξαγωγή συμπερασμάτων, τα δεδομένα των UNACAST και Ookla χρησιμοποιούνται ως οι πραγματικές τοποθεσίες (έχοντας δεδομένα τις συντεταγμένες από τους δέκτες GPS). Στη συνέχεια αυτές συγκρίθηκαν με τις τοποθεσίες των διευθύνσεων που προέκυψαν από τις «κλασικές» μεθόδους γεωεντοπισμού και ελέγχθηκε η ακρίβεια των μετρήσεων. Αξίζει να σημειωθεί ότι για τους κινητούς δέκτες υπάρχει μεγάλη πιθανότητα σφάλματος στην τοποθεσία του GPS, καθώς αν σε κάποια χρονική στιγμή που έχει συλλεχθεί η πληροφορία της τοποθεσίας χαθεί η σύνδεση στο διαδίκτυο, τότε όταν ο δέκτης που επανασυνδέεται στο διαδίκτυο, θα έχει καταγράψει εσφαλμένη τοποθεσία. Στη μελέτη των Saxon κ.α. το σφάλμα αυτό θεωρείται μη ανιχνεύσιμο, συνεπώς η μελέτη υλοποιείται με πάσα επιφύλαξη.

4.7 Αποτελέσματα των Μετρήσεων

Για τα κινητά σημεία διευθύνσεων IP χρησιμοποιήθηκαν δεδομένα από την εταιρία κινητής AT&T Mobility. Από τη σύγκριση των δεδομένων με τις αντίστοιχες διευθύνσεις της IP Geolocation και MaxMind, υπολογίστηκε μια διαφορά στην ενδιάμεση τιμή της τάξης του 2,5% ενώ η ακρίβεια της τοποθεσίας για τους κινητούς δέκτες ήταν μειωμένη με την αντίστοιχη των σταθερών δεκτών.

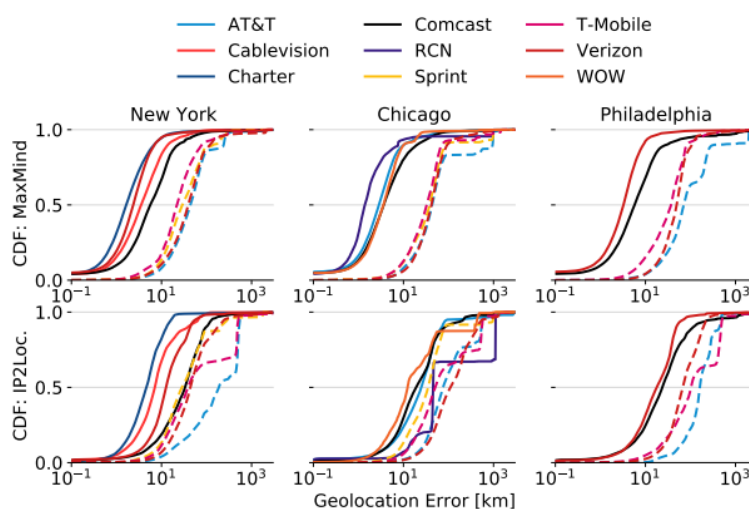
Όσον αφορά τα σταθερά σημεία στην εικόνα 4 απεικονίζονται τα γραφήματα των σφαλμάτων των δεδομένων και οι αποκλίσεις των τιμών για τις διάφορες εταιρίες τηλεπικοινωνιών για την πόλη του Σικάγο από το δείγμα της περιόδου Αυγούστου 2020. Από την εικόνα παρατηρείται ότι η MaxMind παρουσίαζε μικρότερο σφάλμα στις διευθύνσεις της Comcast σε σύγκριση με την Ookla, από τις αντίστοιχες μετρήσεις της Unacast, ενώ με τις αντίστοιχες διευθύνσεις των AT&T RCN και WOW οι αποκλίσεις είναι λίγο μεγαλύτερες.



Εικόνα 4. Σφάλμα εντοπισμού τοποθεσίας στην πόλη του Σικάγο της Unacast και Ookla σε σχέση με τα αντίστοιχα δεδομένα από IP2LOCATION και MaxMind. ©Saxon κ.α, 2021.

Επιπρόσθετα από την εικόνα 4, παρατηρείται πως για τα δεδομένα της Unacast το σφάλμα είναι αρκετά κοντά στο μηδέν. Ανάλογα τον εκάστοτε πάροχο, το σφάλμα αυτό κυμαίνεται από 1-2% για τα δεδομένα της IP2 Location και 4-5% για τα αντίστοιχα δεδομένα της MaxMind.

Όσον αφορά την εγκυρότητα των δεδομένων τοποθεσίας κινητής στην εικόνα 5 απεικονίζονται τα γραφήματα της ακρίβειας των μετρήσεων στις πόλεις της Νέας Υόρκης, του Σικάγο και της Φιλαδέλφειας για τους κυριότερους παρόχους δικτύου. Σε αυτές τις περιπτώσεις η ενδιαμέσος τιμή ακρίβειας για τις εταιρίες RCN, Comcast και MaxMind στην πόλη του Σικάγο βρέθηκε να είναι κάτω από 5 χλμ. Για κάθε πόλη η ακρίβεια που έχει υπολογιστεί είναι πολλή χαμηλή για τους κινητούς δέκτες. Συγκεκριμένα για την πόλη του Σικάγο η ακρίβεια τοποθεσίας ήταν πιο ακριβής για τις σταθερές συνδέσεις από ότι στις κινητές (Saxon κ.α., 2021). Επισημαίνεται ότι η ακρίβεια για τους κινητούς δέκτες παρουσίαζε διαφορετικό σφάλμα ανά εταιρία παρόχου. Ως εκ τούτου, αποδεικνύεται ότι ο εξοπλισμός διαδραματίζει σημαντικό ρόλο στην ακρίβεια εντοπισμού.



Εικόνα 5. Αθροιστική Συνάρτηση Κατανομής σφάλματος γεωεντοπισμού ανά πόλη και πάροχο δικτύου με βάση τις βάσεις MaxMind IP2Location. © Saxon κ.α., 2021.

Από τη μελέτη των Saxons κ.α. προέκυψε ότι το 66% των διευθύνσεων της Comcast στην πόλη της Νέας Υόρκης παρουσίαζε ακρίβεια 10 χλμ σε σχέση με τις συντεταγμένες που προέκυψαν από το GPS. Αντίστοιχα για την T-Mobile το 87% των διευθύνσεων ταυτοποιούνταν σε δύο περιοχές: τη Νέα Υόρκη και το Newark, ενώ μόνο το 18% παρουσίαζε ακρίβεια εντός 10 χλμ σε σύγκριση με την πραγματική τοποθεσία. Συνεπώς για την T-Mobile η ακρίβεια της εμβέλειας επιτυγχάνονταν σε ακτίνα 200 χλμ. Επιπρόσθετα, μεγάλη διαφορά στην ακρίβεια σφάλματος εντοπίστηκε για την

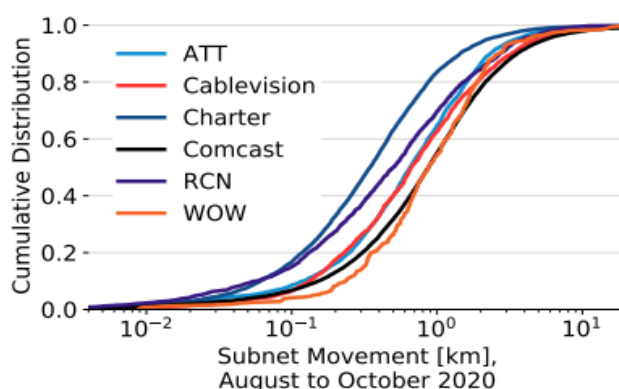
ΑΤ&Τ στις περιοχές της Νέας Υόρκης και της Φιλαδέλφειας. Ο λόγος για τον οποίο το σφάλμα ήταν τόσο μεγάλο για την ΑΤ&Τ στις περιοχές αυτές, οφείλεται στο γεγονός ότι η εταιρία παρείχε ως επί το πλείστο κινητές συνδέσεις με αποτέλεσμα να επιβεβαιώνεται για άλλη μια φορά η μειωμένη εμπιστοσύνη ακρίβειας για τις κινητές συνδέσεις. Ένας τρόπος με τον οποίο θα μπορούσε να εξαλειφθεί το σφάλμα αυτό, θα ήταν η χρήση σταθερών διευθύνσεων ΙΡ από τις κεραίες των εταιριών κινητής, γεγονός όμως που δεν υλοποιείται αυτή τη στιγμή και ως εκ τούτου δεν δύναται να αξιοποιηθούν τα δεδομένα της κινητής για περαιτέρω ανάλυση.

Για την πληρότητα της μελέτης οι Saxons κ.α. εξέτασαν και τη απόδοση ακρίβειας γεωεντοπισμού των διευθύνσεων που αφορούν τον τομέα της εκπαίδευσης και της επιχειρηματικότητας. Για το σκοπό αυτό συλλέχθηκαν δεδομένα από 10.000 φοιτητές πανεπιστημίων και τις εταιρίες της λίστας του Fortune 100 μέσω των πλατφορμών που χρησιμοποιούν για την είσοδο στην εκάστοτε δομή. Επισημαίνεται ότι το Fortune 100 αποτελεί μια λίστα εταιριών των ΗΠΑ και συγκεκριμένα τις 100 πιο επιτυχημένες εταιρίες δημοσίου ή ιδιωτικού τομέα, οι οποίες δημοσιεύονται στο περιοδικό Fortune. Η ένταξη των εταιριών στη λίστα καθορίζεται με βάση τα ετήσια έσοδα τους και η δημοσίευση στο περιοδικό γίνεται σε ετήσια βάση (Fortune 100 Definition, Requirements, and Top Companies, 2022).

Από τη μελέτη που πραγματοποιήθηκε, παρατηρήθηκε ότι τα εκπαιδευτικά ιδρύματα παρουσίασαν μεγαλύτερη ακρίβεια σε σχέση με τις διευθύνσεις των εταιριών, γεγονός το οποίο ερμηνεύεται φυσιολογικό, καθώς τα περισσότερα πανεπιστήμια διαθέτουν σταθερούς εξυπηρετητές (server) με αποτέλεσμα να επιτυγχάνεται μεγαλύτερη ακρίβεια. Συγκεκριμένα για την πλειοψηφία των πανεπιστημίων η ακρίβεια εντοπίστηκε σε ακτίνα 2 χλμ, πλην ορισμένων περιπτώσεων (πανεπιστήμιο του Σικάγο) όπου για το δείγμα του Αυγούστου του 2020 η εμβέλεια ακρίβειας αυξήθηκε στα 10 χλμ. Το γεγονός αυτό δικαιολογείται αν αναλογιστεί κανείς ότι επρόκειτο για καλοκαιρινή περίοδο όπου οι σχολές ήταν κλειστές, καθώς επίσης και στο γεγονός ότι την περίοδο εκείνη λόγω πανδημίας πολλοί σπουδαστές βρίσκονταν στις οικίες τους (Saxon κ.α., 2021).

Στη συνέχεια εξετάστηκε η ακρίβεια των διευθύνσεων, μελετώντας τα υποδίκτυα που δημιουργεί ο κάθε πάροχος για την καλύτερη διαχείριση των ΙΡ διευθύνσεων. Στο γράφημα 7 απεικονίζεται η απόσταση μεταξύ των ενδιάμεσων τιμών

των συντεταγμένων των υποδικτύων των παρόχων, συγκρινόμενα στις δυο διαφορετικές περιόδους. Επισημαίνεται ότι η ενδιάμεση τιμή έχει προκύψει από τις μέσες τιμές των συντεταγμένων χ (γεωγραφικό μήκος) και ψ (γεωγραφικό πλάτος) των IP διευθύνσεων του κάθε υποδικτύου. Για να οριστεί ένας τομέας ή μια περιοχή υποδίκτυο, θα πρέπει να εμπεριέχει τουλάχιστον 10 συσκευές, ήτοι 10 IP διευθύνσεις. Από το γράφημα 6, διαφαίνεται ότι υπάρχει μια ολίσθηση στην απόσταση της τάξης του ενός χιλιομέτρου. Συνεπώς από τα παραπάνω δύναται να θεωρηθεί ότι τα υποδίκτυα με την πάροδο του χρόνου παραμένουν σταθερά.

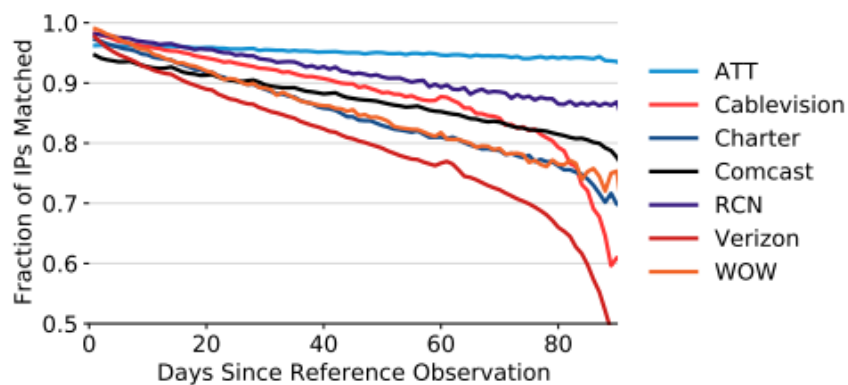


Γράφημα 7. Διαφορά απόστασης της ενδιάμεσης τιμής των υποδικτύων στις σταθερές συνδέσεις στην πάροδο δύο μηνών (Αυγούστου-Οκτωβρίου 2020). © Saxons κ.α., 2021.

4.8 Επίδραση Γεωεντοπισμού σε Επίπεδο Καταναλωτή

Με τη μέχρι στιγμής ανάλυση, προκύπτει ότι η ακρίβεια στον εντοπισμό της τοποθεσίας μιας διεύθυνσης IP, δεν μπορεί εύκολα να πέσει κάτω από τα 2 χλμ. Συνεπώς είναι αρκετά δύσκολο να εντοπιστεί η τοποθεσία ενός ατόμου με ακρίβεια. Ωστόσο εάν υπήρχε η δυνατότητα μια διεύθυνση IP να παραμένει σταθερή σε κάθε σύνδεση στο διαδίκτυο, θα ήταν και ευκολότερο, από τεχνικής άποψης, να διευκρινιστεί η τοποθεσία της. Για τους παρόχους όμως αυτό αποτελεί μια πρόκληση.

Στη μελέτη των Saxons κ.α. έγινε η προσπάθεια εύρεσης της χρονικής διάρκειας κατά την οποία μια διεύθυνση παραμένει σταθερή για ένα σπίτι. Για την υλοποίηση αυτού, ορίστηκε ένα μέγεθος ως η πιθανότητα επιστροφής μια συσκευής στην ίδια διεύθυνση IP που είχε, μετά από την παρέλευση χ ημερών. Για κάθε σύνδεση που γινόταν από μια συσκευή, εκλέχθηκε η διεύθυνση IP της, την πρώτη ημέρα ($\chi = 0$). Μετά από την παρέλευση $\chi = 1$ ημέρας επανελέγχθηκε η σύνδεση της συσκευής με την ίδια διεύθυνση. Με δεδομένα τα ανωτέρω, στη συνέχεια εξετάστηκε η πιθανότητα η συσκευή αυτή να έχει την ίδια IP μετά από π.χ. $\chi = 9$ ημέρες. Στο γράφημα 8 απεικονίζονται τα δεδομένα της καταγραφής μετά από την παρέλευση 90 ημερών.



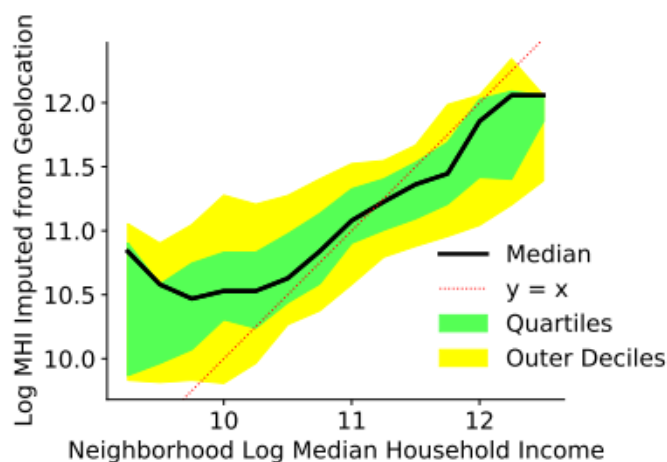
Γράφημα 8. Ποσοστό Διατήρησης IP διεύθυνσης συσκευής σε σταθερή σύνδεση. © Saxon κ.α., 2021.

Από τα ανωτέρω προκύπτει ότι σε διάστημα ενός μηνός το 90% των συσκευών διατήρησαν την ίδια διεύθυνση IP, ενώ σε διάστημα δύο μηνών περισσότερες από 75% των συσκευών διατήρησαν την αρχική διεύθυνση IP τους. Ως εκ τούτου και στις τρεις πόλεις της μελέτης διαφαίνεται η σταθερότητα στη διευθυνσιοδότηση για χρονικό διάστημα τουλάχιστον 60 ημερών.

Επιπρόσθετα στη μελέτη έγινε η προσπάθεια συσχέτισμού των δεδομένων τοποθεσίας διευθύνσεων IP, για την άντληση δημογραφικών στοιχείων ώστε να εξεταστεί κατά πόσο αξιόπιστο μέσο μπορεί να θεωρηθεί ο γεωεντοπισμός για την καταγραφή δεδομένων απογραφής του πληθυσμού. Από τη μελέτη των Saxons κ.α., διαπιστώθηκε αρχικά ότι υπήρχε μεγάλη απόκλιση των πληροφοριών όσον αφορά τις λιγότερο αναπτυγμένες περιοχές. Ο λόγος για τον οποίο συμβαίνει το φαινόμενο αυτό, οφείλεται στην έλλειψη ύπαρξης σταθερών συνδέσεων στο διαδίκτυο. Όπως αναφέρθηκε και παραπάνω οι σταθερές συνδέσεις παρουσίαζαν μεγαλύτερη ακρίβεια σε σχέση με τις κινητές, γεγονός που υποδηλώνει ότι στις περιοχές αυτές, οι κάτοικοι χρησιμοποιούσαν το διαδίκτυο μέσω των κινητών τους τηλεφώνων.

Συνεπώς, συμπεραίνεται ότι η ανάλυση από τα δεδομένα τοποθεσίας IP μιας συσκευής για την παρατήρηση δεδομένων πληθυσμού μιας περιοχής τείνει να αποτύχει στις περισσότερες περιπτώσεις. Η παραπάνω διαπίστωση απεικονίζεται και στο γράφημα 9, όπου γίνεται η σύγκριση του λογαρίθμου του εισοδήματος των νοικοκυριών στην Αμερική μεταξύ στοιχείων που συλλέχθηκαν κατά την απογραφή πληθυσμού και στοιχείων της βάσης δεδομένων της MaxMind. Από το γράφημα παρατηρείται η «υποτίμηση» των δεδομένων καθώς για παράδειγμα στις χαμηλότερες τιμές διαπιστώνονται μεγάλες αποκλίσεις μεταξύ των μεγεθών, ενώ στις μεγαλύτερες

υφίσταται και πάλι διαφορά αλλά σε μικρότερο βαθμό. Σε κάθε περίπτωση τα δεδομένα μπορούν να υποεκτιμήσουν την πραγματική κατάσταση.



Γράφημα 9. Ποσοστό μεταβολής του λογαρίθμου ενδιάμεσης τιμής νοικοκυριών της MaxMind σε σχέση με πραγματικά δεδομένα απογραφής. © Saxons κ.α., 2021.

Κεφάλαιο Πέμπτο

5.1 Επίδραση IP Γεωεντοπισμού στο Διαδίκτυο

Στο κεφάλαιο αυτό γίνεται μια γενικότερη συζήτηση λαμβάνοντας υπόψη τα προαναφερόμενα στις προηγούμενες ενότητες και παρουσιάζονται κάποια από τα συμπεράσματα που προκύπτουν ώστε να διαφωτίσουν το μελετητικό έργο για μελλοντικές έρευνες επί του θέματος.

Για την αρχική κατανόηση της σημασίας της παρούσας εργασίας, θα πρέπει καθοριστεί ο λόγος για τον οποίο μελετάται ο Γεωεντοπισμός IP. Λαμβάνοντας υπόψη τη σημασία που παίζει η γνώση σημαντικών πληροφοριών περί οικονομικών στοιχείων οποιασδήποτε οικονομίας, είναι αναγκαίο η προσπάθεια άντλησης δεδομένων οικονομικού χαρακτήρα από οποιαδήποτε πηγή. Με την πάροδο της τεχνολογίας, στα πλαίσια αυτής της αναζήτησης πληροφοριών, τέθηκαν και τα δεδομένα που μπορούν να συλλεχθούν από τον κάθε χρήστη που συνδέεται στο διαδίκτυο. Υπάρχουν διάφοροι τρόποι για να εξαχθούν πληροφορίες αλλά ο πιο «ανώδυνος» και λιγότερο δαπανηρός για τις εταιρίες, είναι η χρήση των εφαρμογών εύρεσης της τοποθεσίας ενός χρήστη μέσω της IP του.

Αν και μέσα από τις διάφορες εφαρμογές που υπάρχουν για το σκοπό αυτό αντλούνται σημαντικά δεδομένα, στις περισσότερες περιπτώσεις εγείρονται προβληματισμοί αναφορικά με την αξιοπιστία των παρεχόμενων πληροφοριών. Στην

πλειονηφία των περιπτώσεων όπου εταιρίες προσπάθησαν να αποκομίσουν λεπτομερέστερα στοιχεία ενός καταναλωτή, παρατηρήθηκε ότι τα δεδομένα που είχαν συλλεχθεί ήταν σε κάποιο βαθμό ανακριβή. Ως εκ τούτου ξεκίνησε μια προσπάθεια από πολλά εκπαιδευτικά ιδρύματα για την μέτρηση της ακρίβειας των πληροφοριών τοποθεσίας των IP διευθύνσεων.

Συνεπώς στον ακαδημαϊκό χώρο παρατηρήθηκε σημαντικότερη η ανάγκη για την προσπάθεια μέτρησης των σφαλμάτων ακρίβειας παρά για την άντληση στοιχείων οικονομικού χαρακτήρα για εξαγωγή συμπερασμάτων. Στην προσπάθεια αυτή των μελετών επιχειρείται να δοθεί απάντηση σε ένα βασικό ερώτημα: Μπορεί ο Γεωεντοπισμός IP να αποδώσει σαφή και έγκυρη εικόνα του διαδικτύου και των χρηστών του;

Για την απάντηση του παραπάνω ερωτήματος, τίγονται διάφορα θέματα όπως για παράδειγμα το γεγονός ότι κάποιος χρήστης μπορεί να χρησιμοποιεί το διαδίκτυο από διαφορετική τοποθεσία, από τη συνηθισμένη ή από υπολογιστή διαφορετικού χρήστη «παραπλανώντας» το διαδίκτυο. Επιπρόσθετα, όλες οι μελέτες που περιεγράφηκαν στην παρούσα εργασία, έχουν καταλήξει στο συμπέρασμα ότι στις πόλεις και γενικότερα στα αστικά κέντρα, η ακρίβεια είναι βελτιωμένη σε επίπεδο πόλης. Αυτό οφείλεται στο γεγονός ότι τα αστικά κέντρα χρησιμοποιούν τις περισσότερες φορές πιο σύγχρονο εξοπλισμό, πετυχαίνοντας ταυτόχρονα και ακρίβεια μέτρησης. Αντίστοιχα σε περιοχές με αραιωμένο πληθυσμό, δεν υπάρχει η ανάγκη για χρήση “state of the art” τεχνολογίας, όσον αφορά τη χωρητικότητα τερματικών, και κατ’ επέκταση παρατηρούνται μεγαλύτερα σφάλματα ακρίβειας.

Στο ίδιο πνεύμα με τα παραπάνω, σε διάφορες πόλεις παρατηρούνται αποκλίσεις μέσα στην ίδια την πόλη, αναφορικά πάντα με την ακρίβεια. Το γεγονός αυτό οφείλεται κυρίως ότι πολλοί χρήστες χρησιμοποιούν το διαδίκτυο μέσω κινητών τηλεφώνων, με αποτέλεσμα να αλλάζουν οι κεραίες που κάθε φορά συνδέονται (είτε λόγω μετακίνησης τους είτε όχι).

Ένα τελευταίο στοιχείο που εγείρει προβληματισμούς και αξίζει να αναφερθεί, είναι οι διευθύνσεις οι οποίες δεν δύναται να ανιχνευθούν από το σύστημα και πολλές φορές οι βάσεις τις ταξινομούν εσφαλμένα είτε με λάθος τοποθεσία, είτε με ανακριβή δεδομένα σύνδεσης. Σε όλες τις ανωτέρω μελέτες που σχολιάστηκαν, επισημάνθηκε η

πιθανότητα εσφαλμένης καταχώρησης μιας διεύθυνσης IP, «αλλοιώνοντας» με αυτόν τον τρόπο τα αποτελέσματα της μελέτης.

5.2 Πιθανοί Τρόποι Βελτίωσης Αξιοπιστίας

Όσον αφορά μια πιθανή λύση βελτίωσης της παρεχόμενης πληροφορίας τοποθεσίας από τις διευθύνσεις IP και τη μείωση των υπαρχόντων σφαλμάτων, θα μπορούσε να είναι η συλλογή των διευθύνσεων των χρηστών από τους παρόχους δικτύου (ISP). Οι πάροχοι δικτύων, διαθέτουν τα στοιχεία φυσικής διεύθυνσης των πελατών τους και γνωρίζουν τις διευθύνσεις IP που τους έχουν εξουσιοδοτήσει να χρησιμοποιούν. Συνεπώς θα μπορούσαν να παρέχουν με ακρίβεια τις πληροφορίες Γεωεντοπισμού. Για να εφαρμοστεί όμως κάτι τέτοιο, θα πρέπει να διασφαλιστεί ότι δεν ενέχει κάποιος κίνδυνος για τη διαρροή προσωπικών πληροφοριών των πελατών τους και σε κάθε περίπτωση, οι πελάτες – πολίτες, θα πρέπει να είναι ενήμεροι για την παροχή τυχόν προσωπικών τους δεδομένων.

Επιπλέον, η αναβάθμιση των υπαρχόντων υποδομών δικτύου, στις περιοχές όπου αυτός είναι απαραίτητος, δύναται να βελτιώσει κατά πολύ την ακρίβεια της τοποθεσίας μια IP. Όπως παρατηρήθηκε και στην παρούσα εργασία, στις περιοχές με πεπαλαιωμένο εξοπλισμό, διαπιστώθηκαν μεγαλύτερες αποκλίσεις σφαλμάτων, σε σχέση με πιο πυκνοκατοικημένες περιοχές, με μεγαλύτερες απαιτήσεις δικτύου και κατ' επέκταση με χρήση σύγχρονου εξοπλισμού.

5.3 Ανασκόπηση της Εργασίας

Στην παρούσα εργασία παρουσιάστηκε και εξετάστηκε το μελετητικό έργο που αφορά τα δεδομένα Γεωεντοπισμού IP διευθύνσεων και εξέτασης του τρόπου άντλησης χρησίμων δεδομένων οικονομικού χαρακτήρα.

Συγκεκριμένα η μελέτη εστιάστηκε στην περιγραφική ανάλυση του υπάρχοντος μελετητικού έργου. Αρχικά εξετάστηκε η ακρίβεια των δεδομένων τοποθεσίας, μέσω του γεωεντοπισμού των IP. Συγκρίνοντας στιγμιότυπα μεταξύ πραγματικών τοποθεσιών και δεδομένων διαφόρων βάσεων Γεωεντοπισμού. Στη συνέχεια περιεγράφηκε η σημασία του Γεωεντοπισμού στη Διαδικτυακή Εκπαίδευση και στην εξαγωγή διαφόρων κοινωνικών και οικονομικών πληροφοριών με βάση αυτήν. Από τη μελέτη τους παρατηρήθηκε ότι πράγματι υπάρχει μια τάση συμμετοχής ατόμων σε ΔΕ από περισσότερο αναπτυγμένες χώρες σε σύγκριση με τις αναπτυσσόμενες, όπως επίσης και από περισσότερο ευημερείς πολίτες απ' ότι οικονομικά εξουθενωμένους.

Παράλληλα προσπάθησε να γίνει μια χαρτογράφηση της συμπεριφοράς των κατοίκων ανά ΔΕ, χωρίς όμως αυτή να προσφέρει αξιόπιστα δεδομένα. Ως εκ τούτου, η εξαγωγή δεδομένων σε επίπεδο πολιτειών θεωρήθηκε επισφαλής, καθότι η εμβέλεια του σφάλματος της ακρίβειας δεν ήταν σταθερή. Για το σκοπό αυτό η παρούσα εργασία εξέτασε την ακρίβεια της τοποθεσίας.

Τέλος, όσον αφορά την αξιοπιστία της ακρίβειας, επιχειρήθηκε να σχολιαστεί το υπάρχον μελετητικό έργο που έχει πραγματοποιηθεί για τη σύγκριση δεδομένων τοποθεσίας IP με βάση τα συστήματα GNSS. Από την καταγραφή των παρατηρήσεων, διαπιστώθηκε ότι η ακρίβεια τοποθεσίας για τις κινητές συσκευές είναι αρκετά μειωμένη σε σύγκριση με τις σταθερές συνδέσεις, όπου η εμβέλεια της ακρίβειας της τοποθεσίας μιας IP, φτάνει τα 2 χιλιόμετρα.

5.4 Συμπεράσματα

Με βάση τα παραπάνω δύναται να οδηγηθεί κάποιος στην εύλογη παρατήρηση ότι τα δεδομένα γεωεντοπισμού IP μπορούν με σχετική ευκολία να μας δώσουν σαφή δεδομένα όσον αφορά την τοποθεσία σε επίπεδο χώρας και σε επίπεδο νομού. Αντιθέτως είναι αρκετά επισφαλής και ανακριβής σε επίπεδο πόλης καθώς οι αποστάσεις σφάλματος κυμαίνονται από 2 χλμ έως και 100 χλμ ανάλογα την εκάστοτε περίπτωση και βάσης δεδομένων που χρησιμοποιείται για την εξαγωγή των δεδομένων. Συνεπώς η χρήση της μεθόδου εντοπισμού θέσης ενός καταναλωτή με βάση την IP του ενέχει εσφαλμένα συμπεράσματα και δεν ενδείκνυται για εξαγωγή οικονομικών συμπερασμάτων ως επίσημη μορφή δεδομένων.

Ένα παράδειγμα που αναλύει τον παραπάνω συλλογισμό είναι το γεγονός ότι θα μπορούσαν να χρησιμοποιηθούν δεδομένα IP για τον εντοπισμό θέσης χρηστών που επισκέφτηκαν την ιστοσελίδα ενός ηλεκτρονικού καταστήματος, προκειμένου το ηλεκτρονικό κατάστημα να διαμορφώσει μια εικόνα του αγοραστικού κοινού στο οποίο απευθύνονται τα προϊόντα ή υπηρεσίες του. Το γεγονός όμως ότι η ακρίβεια του σφάλματος κυμαίνεται σε ένα εύρος 2-10 χλμ και σε κάποιες περιπτώσεις και πάνω από 20 χλμ, δεν δύναται να επιτύχει διαχωρισμό των αποτελεσμάτων σε επίπεδο συνοικιών. Αντιθέτως μπορεί να αποδώσει πληροφορίες σε επίπεδο νομού με αρκετά μεγάλη εγκυρότητα, έχοντας πάντα υπόψη την πιθανότητα σφάλματος. Ως πιθανό σφάλμα, μπορεί για παράδειγμα να θεωρηθεί η προσπάθεια «παραπλάνησης» των δεδομένων IP από διάφορους χρήστες είτε με χρήση δικτύων μασκών, είτε με διάφορες

τεχνικές χρήσης δικτύων VPN. Ως εκ τούτου, σε κάθε περίπτωση οι πληροφορίες είναι αξιόλογες σε αναλογία πάντοτε με το λόγο και την διεργασία στην οποία πρόκειται να αξιοποιηθούν.

Επιπρόσθετα με βάση τις υπάρχουσες μελέτες αναφορικά με τις IP, επισημαίνεται ο προβληματισμός που δημιουργείται ως προς την «ιδιωτικότητα» της διεύθυνσης. Σε ένα σπίτι, κάθε φορά που κάποιος χρήστης επισκέπτεται μια ιστοσελίδα, δεν διευκρινίζεται το φύλλο ή τα χαρακτηριστικά τα οποία αποτελεί. Κάθε δρομολογητής που συνδέεται στο διαδίκτυο χρησιμοποιεί μια διεύθυνση IP για την πλοήγηση του και στη συνέχεια χρησιμοποιεί ιδιωτικές διευθύνσεις για την εσωτερική δρομολόγηση των πακέτων – δεδομένων. Συνεπώς, σε μια οικογένεια, εάν τη μια στιγμή βρίσκεται σε μια ιστοσελίδα το ένα μέλος και ταυτόχρονα ένα άλλο μέλος της οικογένειας επισκέπτεται μια άλλη ιστοσελίδα, τότε επηρεάζεται η προσπάθεια «χαρτογράφησης» του προφίλ του αγοραστικού κοινού και δεν μπορούν με βεβαιότητα να αντληθούν πληροφορίες.

Ως εκ τούτου για την χρήση της νέας αυτής τεχνικής, ως εργαλείο άντλησης χρήσιμων στοιχείων οικονομικού χαρακτήρα, ο γεωεντοπισμός IP μπορεί να χρησιμοποιηθεί με ικανοποιητική ακρίβεια, για την εξαγωγή συμπερασμάτων σε επίπεδο Χώρας. Η άντληση πληροφοριών δύναται να πραγματοποιηθεί ακόμη και σε επίπεδο Νομών, με αρκετά αυξημένη όμως την πιθανότητα σφάλματος. Συνεπώς, ο Γεωεντοπισμός IP διευθύνσεων, σαφώς μπορεί να αποδώσει χρήσιμες πληροφορίες, σε κάθε περίπτωση όμως, θα πρέπει τα δεδομένα που αντλούνται να επεξεργάζονται και να αξιολογούνται ως συμπληρωματικά εργαλεία σε μια μελέτη και όχι τόσο ως κύριος μηχανισμός εξαγωγής συμπερασμάτων οικονομικού χαρακτήρα.

BIBΛΙΟΓΡΑΦΙΑ

Alcorn, Brandon, Gayle and Kapur, 2015. Higher Education and MOOC in India and the Global South. Change: The Magazine of Higher Learning 47 (3).

Christensen, Gayle, Andrew Steinmetz, Brandon Alcorn and Ezekiel Emanuel, 2013. The MOOC phenomenon: Who Takes Massive Open Online Courses and why? SSR Scholarly Paper ID 2350964. Rochester, NY

Chuang, Isaac & Andrew, 2016. Harvard MITx: Four Years of Open Online Courses. Rochester, SSRN, NY.

Curran, Kevin & Jonathan, 2011, Intergrating Geolocation into Electronic Finance Applications for Additional Security. International Journal of Electronic Finance.

Deng, Cathy & Forest, 2014. Usaddress 0.5.4 documentation.

Emanuel Ezekiel, 2013. Online Education: MOOC Taken by Educates Few. Nature 503.

Everything You Need to Know About IP Based Geolocation—If-So Dynamic Content, 2022. Ανακτήθηκε 15 Οκτώβριος 2022, από <https://www.if-so.com/geo-targeting/>.

Fortune 100 Definition, Requirements, and Top Companies. (χ.χ.). Ανακτήθηκε 3 Νοέμβριος 2022, από <https://www.investopedia.com/terms/f/fortune-100.asp>

Ganelin, D., & Chuang, I. (2019). IP Geolocation Underestimates Regressive Economic Patterns in MOOC Usage. Proceedings of the 2019 11th International Conference on Education Technology and Computers, 268–272. <https://doi.org/10.1145/3369255.3369301>

Geronimus, Arline & Bound, 1998. Use of Census Based Aggregate Variables to Proxy for Socioeconomic Group: Evidence from National Samples. American Journal of Epidemiology.

Goodell, Geoffrey & Syverson, 2007. The Right Place at the Right Time. Communications of the ACM.

Google Developers, 2019. Developer Guide/ Geocoding API.

Gueye, Bamba, Uhlig & Fedida, 2007. Investigating the Imprecision of IP Block-Based Geolocation.

Gouel, M., Vermeulen, K., Fourmaux, O., Friedman, T., & Beverly, R. (2021). Longitudinal Study of an IP Geolocation Database (arXiv:2107.03988). arXiv. <http://arxiv.org/abs/2107.03988>

Gillies, S. (2018). Shapely Documentation. 92.

Hansen, John D. & Justin Reich, 2015a. “Socioeconomic Status and MOOC Enrollment: Enriching Demographic Information with External Datasets.” In Proceedings of the Fifth Enriching International Conference on Learning Analytics and

Knowledge - LAK '15, 59–63. Poughkeepsie, New York: ACM Press.
<https://doi.org/10.1145/2723576.2723615>.

IPv4 Address Report, 2022. Ανακτήθηκε 21 Απριλίου 2022, από <https://ipv4.potaroo.net/>

Internet Protocol (Request for Comments RFC 791). (1981). Internet Engineering Task Force. <https://doi.org/10.17487/RFC0791>

IP Address Conflicts: What They Are and How to Fix Them. (χ.χ.). Lifewire. Ανακτήθηκε 21 Απρίλιος 2022, από <https://www.lifewire.com/what-is-ip-address-conflict-818381>.

Jim Kurose, Keith Ross, 2008. Computer Networking Atop-Down Approach, Boston.

Muir, James A, and P C van Oorschot. 2006. “Internet Geolocation and Evasion,” April, 22

Poese, Ingmar, Uhlig, Kaafar, Donnet & Gueye, 2011. “IP 84 Geolocation Databases: Unreliable?” *SIGCOMM Comput. Commun. Rev.* 41.

Saxon, J., & Feamster, N. (2021). GPS-Based Geolocation of Consumer IP Addresses (arXiv:2105.13389). arXiv. <http://arxiv.org/abs/2105.13389>.

Shavitt, Y., & Zilberman, N. (2011). A Geolocation Databases Study. *IEEE Journal on Selected Areas in Communications*, 29 (10), 2044–2056.
<https://doi.org/10.1109/JSAC.2011.111214>.

Soobader, M Le Clere, W Hadden, & B Maury. 2001. “Using Aggregate Geographic Data to Proxy Individual Socioeconomic Status: Does Size Matter?” *American Journal of Public Health* 91.

State, Bogdan, Ingmar Weber & Emilio Zagheni, 2013. “Studying Inter-National Mobility through IP Geolocation.” In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining - WSDM '13*, 265. Rome, Italy: ACM Press.

Valentino-DeVries, Jennifer, Jeremy Singer-Vine, and Ashkan Soltani. 2012. “Websites Vary Prices, Deals Based on Users’ Information.” *Wall Street Journal*, December 24, 2012, 85 sec. Tech

What is GNSS? | *EU Agency for the Space Programme*. 2022. Ανακτήθηκε 24 Οκτώβριος 2022, από <https://www.euspa.europa.eu/european-space/eu-space-programme/what-gnss>

Zhenghao, Chen, Brandon Alcorn, Gayle Christensen, Nicholas Eriksson, Daphne Koller, and Ezekiel J. Emanuel. 2015. “Who’s Benefiting from MOOCs, and Why.” *Harvard Business Review*, September 22, 2015