

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
ΑΝΑΛΥΣΗ ΣΥΝΘΕΤΩΝ ΓΡΑΦΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΓΡΑΦΩΝ ΓΝΩΣΗΣ ΜΕ
ΧΡΗΣΗ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ ΝΕΟ4J

Υποψήφια: Ανδριάννα Πασατσιφλικιώτη, Π2019173
Επιβλέπων: Βασίλειος Καρυώτης, Αναπληρωτής Καθηγητής

Κέρκυρα

2024

Επιβλέπων

Βασίλειος Καρυώτης, Αναπληρωτής Καθηγητής
Ιόνιο Πανεπιστήμιο, Τμήμα Πληροφορικής

Τριμελής Επιτροπή

Βασίλειος Καρυώτης, Αναπληρωτής Καθηγητής
Ιόνιο Πανεπιστήμιο, Τμήμα Πληροφορικής

Παναγιώτης Κουρουθανάσης, Καθηγητής
Ιόνιο Πανεπιστήμιο, Τμήμα Πληροφορικής

Αριστείδης Βραχάτης, Επίκουρος Καθηγητής
Ιόνιο Πανεπιστήμιο, Τμήμα Πληροφορικής

Υπεύθυνη Δήλωση

Με το παρόν κείμενο βεβαιώνω ότι ο κάτωθι υπογράφων είμαι συγγραφέας της παρούσης πτυχιακής εργασίας, η οποία εκπονήθηκε στο πλαίσιο των απαιτήσεων του προγράμματος σπουδών του Τμήματος Ψηφιακών Μέσων & Επικοινωνίας του Ιόνιου Πανεπιστήμιου και η οποία παραδόθηκε, μετά από έγκριση του επιβλέποντας καθηγητή μου, σε έντυπη και ψηφιακή μορφή στη Γραμματεία του Τμήματος. Επίσης δηλώνω πως κάθε πηγή που χρησιμοποίησα (βιβλιογραφία, αρθρογραφία, δικτυογραφία), για την υποστήριξη των υποθέσεων της μελέτης και της ερευνάς μου, είναι πλήρως συμβατή με τα ακολουθούμενα επιστημονικά πρότυπα και, επιπλέον, αναφέρεται ρητά, υπό μορφή αναφοράς-παραπομπής, σε όλο το φάσμα κειμένων της παρούσης εργασίας. Το αυτό ισχύει για τη χρήση δευτερογενών δεδομένων (πινάκων, διαγραμμάτων και εικόνων), ιδεών και λέξεων, τα οποία και αναφέρονται είτε ακριβώς όπως υπάρχουν στις πηγές είτε μεθερμηνεύονται από εμένα.

Επώνυμο	Πασατσιφλικιώτη
Όνομα	Ανδριάνα
Αριθμός μητρώου	Π2019173
Ημερομηνία	2023-2024
Υπογραφή	

Περίληψη

Στην σύγχρονη κοινωνία, όπου όλα προοδεύουν, υπάρχει υψηλή ανάγκη θωράκισής της με νέες τεχνολογίες, για την επιτυχή αντιμετώπιση απρόσμενων προβλημάτων. Αρχικά, ένας από τους σημαντικότερους τομείς παγκοσμίως που πρέπει να διασφαλιστεί, είναι ο τομέας της υγείας. Ένα σπουδαίο παράδειγμα, που αποτέλεσε και την εκ κινητήριο δύναμη αυτής της ανάγκης είναι η πρόσφατη παγκόσμια πανδημία της νέας νόσου του κορονοϊού 2019 (COVID-19). Αυτή η νόσος επέφερε αμέτρητες απώλειες ανθρώπινων ζωών, καθώς η καταπολέμηση της διήρκεσε αρκετό χρονικό διάστημα. Κατά αυτό το διάστημα γιατροί, ερευνητές και γενικότερα οι αρμόδιες αρχές έκαναν υπεράνθρωπες προσπάθειες συλλογής δεδομένων και αποτελεσμάτων, ώστε να καταφέρουν να επιφέρουν την βέλτιστη λύση. Έτσι, βλέπουμε ότι πληροφορίες που θα μπορούσαν να εξαχθούν από Δεδομένα σχετικά με τον COVID-19 από τη μικροβιολογία, τη βιολογία, τα κλινικά ιδρύματα, τη φαρμακευτική βιομηχανία, γίνονται ιδιαίτερα απαιτητικές λόγω του μεγάλου όγκου δεδομένων, για την κατανόηση της μετάδοσης, της θεραπείας και της πρόληψης. Αυτό θα μπορούσε να αντιμετωπιστεί με την χρήση μεθόδων ανάλυσης και μίας γραφικής παράστασης, όπου θα είναι συλλεγμένα όλα τα δεδομένα. Επιπρόσθετα, ένας ακόμα τομέας, του οποίου η γρήγορη παροχή αντιμετώπισης είναι αναγκαίο να εξασφαλιστεί, είναι ο οικονομικός τομέας. Δυστυχώς, στις μέρες μας αντιμετωπίζουμε νομιμοποιήσεις εσόδων από παράνομες δραστηριότητες, ασφαλιστικές απάτες και άλλων ομοίων ύποπτων δραστηριοτήτων. Όλο αυτό το φαινόμενο ταραξεί τις οικονομίες των χωρών και αποδιοργανώνει την σύστασή τους. Σε αυτήν την περίπτωση αρμόζει ως τρόπος επίλυσης του προβλήματος η χρήση γραφών γνώσης. Αυτό συμβαίνει, διότι τα γραφήματα γνώσης μπορούν να αντιπροσωπεύουν εταιρείες, οι οποίες κατέχουν σύνθετες μετοχές και πληροφορίες σχετικά με επενδύσεις, εξαγορές, τυχόν χρεοκοπίες πραγματικών οικονομικών δεδομένων. Οι κόμβοι των γραφημάτων αυτών θα αντιπροσωπεύουν πελάτες και έμπορους και οι άκρες θα είναι συναλλαγές μεταξύ πελατών, πληροφορίες κατοικίας πελατών και γενικότερα θα έχουν πληροφορίες που θα σχετίζονται με τους πελάτες. Αυτή η απεικόνιση θα μπορούσε να οδηγήσει σε μια σημαντική πηγή γνώσεων, που στην πάροδο του χρόνου θα βοηθήσει στην αποφυγή λαθών και την μείωση της εξαπάτησης.

Η πτυχιακή αυτή έχει ως αντικείμενο έρευνας την εύρεση της σημαντικότητας χρήσης γραφημάτων γνώσης σε βιολογικά και εν δυνάμει σε οικονομικά προβλήματα. Συγκεκριμένα, μέσω της έρευνας αυτής, επιθυμητή είναι η γνωστοποίηση της έννοιας και της σημασίας των γραφημάτων γνώσης σε

συνδυασμό με δεδομένα και οι διάφοροι τρόποι με τους οποίους μπορούν να συνεισφέρουν στην ταχύτερη εύρεση λύσεων σε προβλήματα βιολογικού και οικονομικού χαρακτήρα.

Η συγκεκριμένη πτυχιακή περιλαμβάνει 5 κεφάλαια. Αναλυτικότητα στο 1^ο κεφάλαιο γίνεται μια εισαγωγή, στην οποία περιγράφεται ολοκληρωμένα η έννοια των γραφημάτων γνώσης, οι διάφοροι τύποι που υπάρχουν και μία σύντομη σύγκριση αυτών. Στο 2^ο κεφάλαιο γίνεται μία εκτενής αναφορά στη βιβλιογραφική ανασκόπηση των ερευνών και των συμπερασμάτων τους, που μελετήθηκαν. Στο 3^ο κεφάλαιο εξηγείται η έρευνα και η διαδικασία ολοκλήρωσης της. Στο 4^ο κεφάλαιο επεξηγείται λεπτομερώς ο Γράφος Γνώσης που δημιουργήθηκε, τα δεδομένα που χρησιμοποιήθηκαν και τα ερωτήματα (Queries) που έγιναν προς την βάση δεδομένων. Στο 5^ο κεφάλαιο αναλύονται τα αποτελέσματα από τα Queries. Στο 6^ο κεφάλαιο αναγράφονται τα συμπεράσματα και το πως αυτά μπορούν να βοηθήσουν στην αντιμετώπιση προβλημάτων υγειονομικού τομέα μέσω Γράφων Γνώσης. Στο 6^ο κεφάλαιο υπάρχει το γενικό συμπέρασμα της έρευνας, καθώς και προτάσεις για μελλοντική έρευνα. Τέλος, υπάρχει η βιβλιογραφική ανασκόπηση που μελετήθηκε.

Λέξεις κλειδιά: Γράφος Γνώσης, δεδομένα, ιατρικό ιστορικό, Covid-19, ασθενείς.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κύριο Βασίλειο Καρυώτη, για το χρόνο που διέθεσε δίνοντάς μου χρήσιμες συμβουλές και για την καθοδήγηση που μου προσέφερε, ώστε να επιτευχθεί η ολοκλήρωση της πτυχιακής μου εργασίας. Στο ίδιο πλαίσιο ευγνωμοσύνης, θα ήθελα να ευχαριστήσω όλες τις καθηγήτριες και τους καθηγητές του Τμήματος για τη συμβολή τους στην επιστημονική και τεχνολογική μου εκπαίδευση καθ' όλη τη διάρκεια των σπουδών μου. Οφείλω επίσης ένα μεγάλο ευχαριστώ σε όλους εκείνους που συνέβαλαν γενικότερα στην συγκρότηση της εργασίας μου και πιο συγκεκριμένα στο ερευνητικό εργαστήριο ISDLab του Τμήματος, καθώς χρησιμοποίησα κάποια από τα δεδομένα τους για την έρευνά μου. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την οικονομική τους υποστήριξη και την πίστη που έχουν σε μένα και τις δυνατότητές μου, καθώς και τους συγγενείς και τους φίλους για την ηθική τους υποστήριξη σε όλο το διάστημα των σπουδών μου.

Κέρκυρα, 2024

Ανδριάνα Πασατσιφλικιώτη

Πίνακας περιεχομένων

Περίληψη	4
Ευχαριστίες.....	6
Κατάλογος Σχημάτων/Εικόνων	9
Κεφάλαιο 1 ^ο : Εισαγωγή	12
Εισαγωγή	12
1.1 Γράφοι Γνώσης	12
1.2 Γενικά χαρακτηριστικά	13
1.3 Εισαγωγή στην θεωρία γραφημάτων	15
Κεφάλαιο 2 ^ο : Βιβλιογραφική Ανασκόπηση	25
Εισαγωγή	25
2.1 Γραφική Αναπαράσταση Βάσεων Δεδομένων (Graph Databases).	25
2.2 Το σύστημα διαχείρισης βάσεων δεδομένων γραφημάτων Neo4j.	27
2.3 Έρευνα: Κατασκευή γραφημάτων γνώσης στον βιοϊατρικό κλάδο.	27
2.4 Έρευνα: Εφαρμογές στην υγειονομική περίθαλψη με χρήση την ανάλυση γράφων.....	29
2.5 Έρευνα: Ο οικονομικός σχεδιασμός και η τεχνική υλοποίηση με βάση γράφους γνώσης.	30
2.6 Έρευνα: Νευροσυμβολική αναπαράσταση με βάση γράφους γνώσης με βιολογικά δεδομένα.	31
2.7 Έρευνα: Ανάλυση γράφων γνώσης που αφορούν οικονομικά, τράπεζες και χρηματοοικονομικά.	33
2.8 Έρευνα: Μελέτη γράφων γνώσης που αφορούν τα χρηματοοικονομικά με χρήση υβριδικής λογικής.....	34
2.9 Έρευνα: Ειδικοί γράφοι γνώσης για την υγεία.	36
2.10 Έρευνα: Εξατομικευμένο Γράφημα Γνώσεων Υγείας.	38
2.11 Έρευνα: Γραφήματα γνώσης και εξηγήσιμη τεχνητή νοημοσύνη στην Υγεία.	39
2.12 Έρευνα: Κατασκευή γραφήματος γνώσης υγειονομικής περίθαλψης: Μια συστηματική ανασκόπηση της τελευταίας τεχνολογίας, των ανοιχτών θεμάτων και των ευκαιριών.	41
2.13 Έρευνα: Πρόταση νήματος ενισχυμένου γραφήματος γνώσης στα Φόρουμ Υγείας.....	42
2.14 Έρευνα: Γράφημα Γνώσης Σημαιολογικής Υγείας: Σημαιολογική Ένταξη του Ετερογενείς Ιατρικές Γνώσεις και Υπηρεσίες.	44
2.15 Έρευνα: Εκμάθηση Γραφήματος Γνώσης Υγείας από Ηλεκτρονικά Ιατρικά Μητρώα. ...	45

2.16 Ενσωματώσεις γραφήματος γνώσης για πρόβλεψη επανένταξης στη Μονάδα Εντατικής Θεραπείας(ΜΕΘ).....	48
Κεφάλαιο 3 °: Ανάλυση δεδομένων με γράφους γνώσης.....	50
Εισαγωγή	50
3.1 Προσέγγιση.....	50
3.2 Σκοπός.....	51
3.3 Διαδικασία διεκπεραίωσης έρευνας	51
Κεφάλαιο 4°: Γράφος γνώσης έρευνας και δεδομένα	52
Εισαγωγή	52
4.1 Δημιουργία Γράφου Γνώσης.....	52
4.4.1 Διαδικασία σχεδιασμού.....	52
4.4.2 Προβλήματα υλοποίησης.....	66
4.4.3 Τελική μορφή του Γράφου Γνώσης	66
4.2 Τα Δεδομένα	67
4.3 Ερωτήματα προς την βάση δεδομένων (Queries)	68
Κεφάλαιο 5°: Αποτελέσματα έρευνας.....	90
Εισαγωγή:	90
5.1 Πίνακας ερωτήσεων και αποτελεσμάτων έρευνας.....	90
Κεφάλαιο 6°: Συμπεράσματα και ιδέες για μελλοντική έρευνα	93
6.1 Συμπεράσματα	93
6.2 Ιδέες για μελλοντική έρευνα	95
Βιβλιογραφία:	97

Κατάλογος Σχημάτων/Εικόνων

ΕΙΚΟΝΑ 1 ΤΟ ΣΧΗΜΑ ΣΤΑ ΑΡΙΣΤΕΡΑ ΑΝΤΙΠΡΟΣΩΠΕΥΕΙ ΤΙΣ 7 ΓΕΦΥΡΕΣ ΤΟΥ ΚΟΝΙΓΣΒΕΡΓ ΚΑΙ ΤΟ ΣΧΗΜΑ ΣΤΑ ΔΕΞΙΑ ΕΙΝΑΙ ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ ΑΝΤΙΣΤΟΙΧΟΥ ΓΡΑΦΗΜΑΤΟΣ EULER. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015)	13
ΕΙΚΟΝΑ 2 ΕΔΩ ΑΠΕΙΚΟΝΙΖΕΤΑΙ ΕΝΑΣ ΓΡΑΦΟΣ G, ΟΠΟΥ ΕΧΕΙ 2 ΚΟΜΒΟΥΣ ΤΟΝ Χ ΚΑΙ Υ ΚΑΙ ΔΥΟ ΑΚΜΕΣ ΑΡΙΘΜΗΜΕΝΕΣ, ΤΗΝ 1 ΚΑΙ 2. ΤΟ ΣΥΝΟΛΟ ΚΟΜΒΩΝ ΕΙΝΑΙ ΤΟ $V(G)=(X, Y)$ ΚΑΙ ΤΟ ΣΥΝΟΛΟ ΤΩΝ ΑΚΜΩΝ ΕΙΝΑΙ ΤΟ $E(G)$ ΣΤΟ ΟΠΟΙΟ ΑΝΗΚΟΥΝ ΟΙ ΑΚΜΕΣ $\{X, Y\}$ ΚΑΙ $\{Y, X\}$. ΟΙ ΔΥΟ ΑΥΤΕΣ ΑΚΜΕΣ ΕΙΝΑΙ ΠΑΡΑΛΛΗΛΕΣ, ΚΑΘΩΣ ΠΡΟΣΠΙΠΤΟΥΝ ΣΤΟ ΙΔΙΟ ΖΕΥΓΟΣ ΚΟΜΒΩΝ.....	14
ΕΙΚΟΝΑ 3 ΣΤΟ ΣΧΗΜΑ ΑΥΤΟ ΦΑΙΝΕΤΑΙ ΕΝΑΣ ΚΟΜΒΟΣ, Ο Χ ΚΑΙ ΜΙΑ ΑΚΜΗ Η ΧΧ, Η ΟΠΟΙΑ ΘΕΩΡΕΙΤΑΙ ΒΡΟΧΟΣ, ΚΑΘΩΣ ΞΕΚΙΝΑΕΙ ΚΑΙ ΤΕΛΕΙΩΝΕΙ ΣΤΟΝ ΙΔΙΟ ΚΟΜΒΟ.	15
ΕΙΚΟΝΑ 4 ΔΙΑΦΟΡΕΣ ΑΝΑΠΑΡΑΣΤΑΣΕΙΣ ΕΝΟΣ ΜΗ-ΚΑΤΕΥΘΥΝΟΜΕΝΟΥ ΓΡΑΦΗΜΑΤΟΣ G. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015)	16
ΕΙΚΟΝΑ 5 ΔΙΑΦΟΡΕΣ ΑΝΑΠΑΡΑΣΤΑΣΕΙΣ ΕΝΟΣ ΚΑΤΕΥΘΥΝΟΜΕΝΟΥ ΓΡΑΦΗΜΑΤΟΣ G., (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015)	17
ΕΙΚΟΝΑ 6 ΤΡΙΕΙΣ ΑΝΑΠΑΡΑΣΤΑΣΕΙΣ ΕΝΟΣ ΓΡΑΦΗΜΑΤΟΣ HAMILTON. , (ΝΙΚΟΛΟΠΟΥΛΟΣ, 2017-2018)	17
ΕΙΚΟΝΑ 7 ΤΑ ΣΥΓΚΕΚΡΙΜΕΝΟ ΓΡΑΦΗΜΑ ΕΧΕΙ Κ ΚΟΜΒΟΥΣ, ΟΙ ΟΠΟΙΟΙ ΕΙΝΑΙ ΑΝΕΞΑΡΤΗΤΟΙ ΜΕΤΑΞΥ ΤΟΥΣ, ΔΗΛΑΔΗ ΑΝΑ ΔΥΟ ΔΕΝ ΕΧΟΥΝ ΚΑΜΙΑ ΣΧΕΣΗ ΜΕΤΑΞΥ ΤΟΥΣ ΚΑΙ ΔΗΜΙΟΥΡΓΟΥΝ ΕΝΑ ΑΝΕΞΑΡΤΗΤΟ ΓΡΑΦΗΜΑ. , (ΝΙΚΟΛΟΠΟΥΛΟΣ , ΘΕΩΡΙΑ ΓΡΑΦΗΜΑΤΩΝ ΕΝΟΤΗΤΑ 9 ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΠΡΟΒΛΗΜΑΤΩΝ, 2016-2017)	18
ΕΙΚΟΝΑ 8 ΠΛΗΡΕΣ ΓΡΑΦΗΜΑ G., (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015)	18
ΕΙΚΟΝΑ 9 ΨΕΥΔΟΓΡΑΦΗΜΑ G ΜΕ ΒΡΟΧΟΥΣ. , (ΝΙΚΟΛΟΠΟΥΛΟΣ, 2017-2018).....	19
ΕΙΚΟΝΑ 10 ΠΟΛΥΓΡΑΦΗΜΑ G., (ΝΙΚΟΛΟΠΟΥΛΟΣ, 2017-2018).....	19
ΕΙΚΟΝΑ 11 ΑΠΛΟ ΓΡΑΦΗΜΑ G. , (ΝΙΚΟΛΟΠΟΥΛΟΣ, 2017-2018).....	20
ΕΙΚΟΝΑ 12 ΔΥΟ ΕΜΒΑΡΑ ΓΡΑΦΗΜΑΤΑ, ΤΑ ΟΠΟΙΑ ΤΟ G1 ΕΙΝΑΙ ΜΗ-ΚΑΤΕΥΘΥΝΟΜΕΝΟ ΚΑΙ ΤΟ G2 ΕΙΝΑΙ ΚΑΤΕΥΘΥΝΟΜΕΝΟ. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015)	20
ΕΙΚΟΝΑ 13 ΔΙΜΕΡΕΣ ΓΡΑΦΗΜΑ G. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015)	21
ΕΙΚΟΝΑ 14 ΠΛΗΡΕΣ ΔΙΜΕΡΕΣ ΓΡΑΦΗΜΑ G. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015)	22
ΕΙΚΟΝΑ 15 ΤΟ ΑΡΙΣΤΕΡΟ ΣΧΗΜΑ ΕΙΝΑΙ ΕΝΑ ΜΗ-ΚΑΤΕΥΘΥΝΟΜΕΝΟ ΓΡΑΦΗΜΑ G ΚΑΙ ΤΟ ΔΕΞΙ ΣΧΗΜΑ ΕΙΝΑΙ ΤΟ ΑΝΤΙΣΤΟΙΧΟ L-ΓΡΑΦΗΜΑ ΤΟΥ. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015).....	22
ΕΙΚΟΝΑ 16 ΤΟ ΑΡΙΣΤΕΡΟ ΣΧΗΜΑ ΕΙΝΑΙ ΕΝΑ ΓΡΑΦΗΜΑ ΚΑΙ ΤΟ ΔΕΞΙ ΣΧΗΜΑ ΕΙΝΑΙ ΤΟ ΑΝΤΙΣΤΟΙΧΟ ΕΠΙΠΕΔΟ ΓΡΑΦΗΜΑ ΤΟΥ. (ΕΧΩ ΕΝΑ ΚΟΜΜΑΤΙ ΤΗΣ ΕΡΕΥΝΑΣ ΚΑΙ ΔΕΝ ΕΧΕΙ ΣΤΟΙΧΕΙΑ ΓΙΑ ΝΑ ΜΠΟΡΕΣΩ ΝΑ ΒΑΛΩ ΒΙΒΛΙΟΓΡΑΦΙΑ)	23
ΕΙΚΟΝΑ 17 ΣΤΑ ΑΡΙΣΤΕΡΑ ΑΠΕΙΚΟΝΙΖΕΤΑΙ ΕΝΑΣ ΓΡΑΦΟΣ ΚΑΙ ΣΤΑ ΔΕΞΙΑ ΥΠΑΡΧΕΙ Ο ΠΙΝΑΚΑΣ ΓΕΙΤΝΙΑΣΗΣ ΤΟΥ ΓΡΑΦΟΥ. , (ΑΝΑΠΑΡΑΣΤΑΣΗ ΓΡΑΦΗΜΑΤΩΝ, 2015).....	24
ΕΙΚΟΝΑ 18 ΑΠΕΙΚΟΝΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ EXCEL, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΟΝ ΚΟΜΒΟ PATIENT.....	55
ΕΙΚΟΝΑ 19 ΑΠΕΙΚΟΝΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ CSV UTF-8, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΟΝ ΚΟΜΒΟ PATIENT.....	56
ΕΙΚΟΝΑ 20 ΠΡΩΤΟ ΜΕΡΟΣ ΑΠΕΙΚΟΝΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ EXCEL, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΟΝ ΚΟΜΒΟ DISEASES. .56	
ΕΙΚΟΝΑ 21 ΔΕΥΤΕΡΟ ΜΕΡΟΣ ΑΠΕΙΚΟΝΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ EXCEL, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΟΝ ΚΟΜΒΟ DISEASES.	57
ΕΙΚΟΝΑ 22 ΠΡΩΤΟ ΜΕΡΟΣ ΑΠΕΙΚΟΝΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ CSV UTF-8, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΟΝ ΚΟΜΒΟ DISEASES.	57
ΕΙΚΟΝΑ 23 ΔΕΥΤΕΡΟ ΜΕΡΟΣ ΑΠΕΙΚΟΝΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ CSV UTF-8, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΟΝ ΚΟΜΒΟ DISEASES.....	58
ΕΙΚΟΝΑ 24 ΑΠΕΙΚΟΝΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ EXCEL, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΟΝ ΚΟΜΒΟ DATES.	58
ΕΙΚΟΝΑ 25 ΑΠΕΙΚΟΝΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ CSV UTF-8, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΟΝ ΚΟΜΒΟ DATES.	59

ΕΙΚΟΝΑ 26 ΣΤΗΝ ΣΥΓΚΕΚΡΙΜΕΝΗ ΕΙΚΟΝΑ ΑΠΕΙΚΟΝΙΖΟΝΤΑΙ ΟΙ ΚΟΜΒΟΙ ΤΟΥ ΓΡΑΦΟΥ ΓΝΩΣΗΣ ΚΑΙ ΤΑ ΓΝΩΡΙΣΜΑΤΑ ΤΟΥΣ. ΣΤΟ ΔΕΞΙ ΜΕΡΟΣ ΤΗΣ ΕΙΚΟΝΑΣ ΦΑΙΝΟΝΤΑΙ ΤΑ ΓΝΩΡΙΣΜΑΤΑ ΕΝΟΣ ΑΠΟ ΤΟΥΣ ΚΟΜΒΟΥΣ ΠΟΥ ΑΝΗΚΟΥΝ ΣΤΟ LABEL PATIENT.	59
ΕΙΚΟΝΑ 27 ΑΠΕΙΚΟΝΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ EXCEL, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΤΩΝ ΣΧΕΣΕΩΝ ΜΕΤΑΞΥ ΤΩΝ ΚΟΜΒΩΝ.	60
ΕΙΚΟΝΑ 28 ΑΠΕΙΚΟΝΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΗ CSV UTF-8, ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΤΩΝ ΣΧΕΣΕΩΝ ΜΕΤΑΞΥ ΤΩΝ ΚΟΜΒΩΝ.	61
ΕΙΚΟΝΑ 29 ΣΤΗΝ ΣΥΓΚΕΚΡΙΜΕΝΗ ΕΙΚΟΝΑ ΑΠΕΙΚΟΝΙΖΕΤΑΙ Η ΤΕΛΙΚΗ ΜΟΡΦΗ ΤΟΥ ΓΡΑΦΟΥ ΓΝΩΣΗΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΕ ΣΤΗΝ ΕΡΕΥΝΑ.	61
ΕΙΚΟΝΑ 30 ΣΤΗ ΣΥΓΚΕΚΡΙΜΕΝΗ ΕΙΚΟΝΑ ΠΑΡΟΥΣΙΑΖΕΤΑΙ Ο ΓΡΑΦΟΣ ΓΝΩΣΗΣ ΠΟΥ ΚΑΤΑΣΚΕΥΑΣΤΗΚΕ ΓΙΑ ΤΗΝ ΜΕΛΕΤΗ.	67
ΕΙΚΟΝΑ 31 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, ΠΩΣ ΟΙ ΠΕΡΙΣΣΟΤΕΡΟΙ ΑΣΘΕΝΕΙΣ ΠΟΥ ΚΑΤΕΛΗΞΑΝ ΝΟΣΟΥΣΑΝ ΜΟΝΟ ΑΠΟ ΤΟΝ COVID-19 ΚΑΙ ΠΩΣ ΗΤΑΝ ΥΓΙΕΙΣ.	69
ΕΙΚΟΝΑ 32 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, ΠΩΣ ΟΤΑΝ ΟΙ ΑΣΘΕΝΕΙΣ ΝΟΣΟΥΝ ΑΠΟ ΤΟΝ COVID-19, ΕΙΝΑΙ ΠΟΛΥ ΕΠΙΚΙΝΔΥΝΟ ΑΝ ΝΟΣΟΥΝ ΚΑΙ ΑΠΟ ΑΣΘΕΝΕΙΣ ΤΗΣ ΚΑΡΔΙΑΣ.	70
ΕΙΚΟΝΑ 33 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, ΠΩΣ ΟΤΑΝ ΟΙ ΑΣΘΕΝΕΙΣ ΝΟΣΟΥΝ ΑΠΟ ΤΟΝ COVID-19, ΔΙΑΤΡΕΧΟΥΝ ΛΙΓΟΤΕΡΟ ΚΙΝΔΥΝΟ ΑΝ ΕΜΦΑΝΙΖΟΥΝ ΣΥΜΠΤΩΜΑΤΑ ΚΑΙ ΕΧΟΥΝ ΜΗ ΦΥΣΙΟΛΟΓΙΚΑ ΚΛΙΝΙΚΑ ΚΑΙ ΕΡΓΑΣΤΗΡΙΑΚΑ ΕΥΡΗΜΑΤΑ ΑΠΟ ΟΤΙ ΑΝ ΕΧΟΥΝ ΚΑΠΟΙΑ ΑΛΛΗ ΑΣΘΕΝΕΙΑ.	71
ΕΙΚΟΝΑ 34 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, ΠΩΣ ΤΑ ΠΕΡΙΣΣΟΤΕΡΑ ΚΡΟΥΣΜΑΤΑ ΘΑΝΑΤΩΝ ΜΕΤΑΞΥ ΤΟΥ COVID-19 MULTIPLE CAUSE ΚΑΙ ΤΟΥ COVID-19 UNDERLYING CAUSE, ΕΧΟΥΝ ΠΡΟΕΛΘΕΙ ΑΠΟ ΤΟΝ COVID-19 UNDERLYING CAUSE.	72
ΕΙΚΟΝΑ 35 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, ΟΛΕΣ ΟΙ ΠΛΗΡΟΦΟΡΙΕΣ ΤΩΝ ΑΣΘΕΝΩΝ ΠΟΥ ΤΟ ΦΥΛΟ ΤΟΥΣ ΕΙΝΑΙ ΑΡΣΕΝΙΚΟ.	73
ΕΙΚΟΝΑ 36 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΜΟΝΟ ΤΟΥΣ ΑΣΘΕΝΕΙΣ ΠΟΥ ΤΟ ΦΥΛΟ ΤΟΥΣ ΗΤΑΝ ΑΡΣΕΝΙΚΟ.	74
ΕΙΚΟΝΑ 37 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΜΟΝΟ ΤΟΥΣ ΑΣΘΕΝΕΙΣ ΠΟΥ ΤΟ ΦΥΛΟ ΤΟΥΣ ΗΤΑΝ ΘΗΛΥΚΟ.	74
ΕΙΚΟΝΑ 38 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΚΑΙ ΤΩΝ ΔΥΟ ΦΥΛΩΝ.	75
ΕΙΚΟΝΑ 39 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΠΟΥ ΚΑΤΟΙΚΟΥΝ ΣΤΙΣ ΗΝΩΜΕΝΕΣ ΠΟΛΙΤΕΙΕΣ.	76
ΕΙΚΟΝΑ 40 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΠΟΥ ΚΑΤΟΙΚΟΥΝ ΣΕ ΟΛΕΣ ΤΙΣ ΗΠΕΙΡΟΥΣ.	76
ΕΙΚΟΝΑ 41 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΠΟΥ ΚΑΤΟΙΚΟΥΝ ΣΤΗΝ ΙΣΠΑΝΙΑ.	77
ΕΙΚΟΝΑ 42 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΠΟΥ ΜΕΝΟΥΝ ΣΤΗΝ ΑΜΕΡΙΚΗ, ΣΤΗΝ ΙΝΔΙΑ Η ΣΤΗΝ ΑΛΑΣΚΑ.	77
ΕΙΚΟΝΑ 43 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΠΟΥ ΚΑΤΟΙΚΟΥΝ ΣΤΗΝ ΑΣΙΑ.	78
ΕΙΚΟΝΑ 44 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΠΟΥ ΑΝΗΚΟΥΝ ΣΤΗΝ ΜΑΥΡΗ ΦΥΛΗ.	78
ΕΙΚΟΝΑ 45 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΠΟΥ ΜΕΝΟΥΝ ΣΤΗΝ ΧΑΒΑΗ Η ΓΕΝΙΚΑ ΝΗΣΙΩΤΕΣ ΤΟΥ ΕΙΡΗΝΙΚΟΥ ΩΚΕΑΝΟΥ.	79
ΕΙΚΟΝΑ 46 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΠΟΥ ΑΝΗΚΟΥΝ ΣΤΗΝ ΛΕΥΚΗ ΦΥΛΗ.	79
ΕΙΚΟΝΑ 47 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΟΛΩΝ ΤΩΝ ΗΛΙΚΙΩΝ.	80
ΕΙΚΟΝΑ 48 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΤΩΝ ΟΠΟΙΩΝ Η ΗΛΙΚΙΑ ΕΙΝΑΙ ΚΑΤΩ ΤΟΥ ΕΝΟΣ ΕΤΟΥΣ.	81
ΕΙΚΟΝΑ 49 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΤΩΝ ΟΠΟΙΩΝ Η ΗΛΙΚΙΑ ΕΙΝΑΙ ΜΕΧΡΙ ΤΟ 170 ΕΤΟΣ.	81
ΕΙΚΟΝΑ 50 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΤΩΝ ΟΠΟΙΩΝ Η ΗΛΙΚΙΑ ΕΙΝΑΙ ΜΕΤΑΞΥ 1-4 ΕΤΩΝ.	82
ΕΙΚΟΝΑ 51 ΣΤΗΝ ΠΑΡΑΠΑΝΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΝΟΣ QUERY ΣΤΟ ΝΕΟ4J, Ο ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΘΑΝΑΤΩΝ ΠΟΥ ΑΦΟΡΑ ΑΣΘΕΝΕΙΣ ΤΩΝ ΟΠΟΙΩΝ Η ΗΛΙΚΙΑ ΕΙΝΑΙ ΜΕΤΑΞΥ 5-14 ΕΤΩΝ.	82

[illegible]

Κεφάλαιο 1^ο: Εισαγωγή

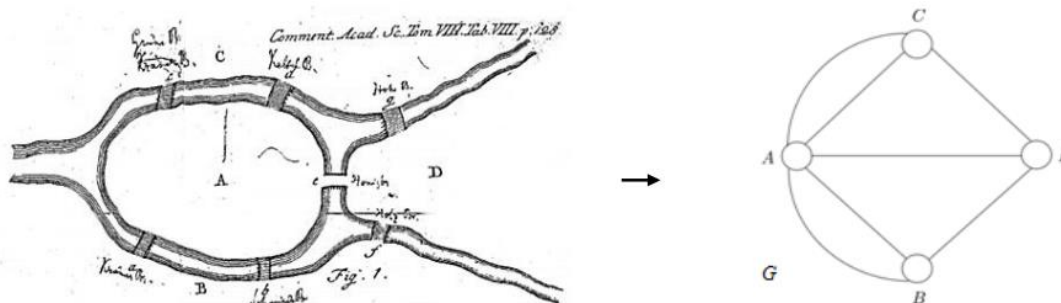
Εισαγωγή

Κύριο θέμα της πτυχιακής είναι η έννοια των γράφων γνώσης και ειδικότερα η εφαρμογή τους στον ιατρικό και οικονομικό τομέα. Για τον λόγο αυτό, πρώτο μέλημα είναι η ολοκληρωμένη παρουσίαση της έννοιάς τους.

1.1 Γράφοι Γνώσης

Η θεωρία των γράφων πρωτοεμφανίστηκε τον 18ο αιώνα, όταν εμφανίστηκε το πρόβλημα των 7 γεφυρών του Königsberg και ζητήθηκε η επίλυσή του. Αναλυτικότερα, στην περιοχή του Königsberg (σημερινό Kaliningrad) διέρχεται ο ποταμός Pregel, ο οποίος διαπερνά την πόλη και τη διαχωρίζει σε τέσσερα τμήματα. Αυτό, δυσκόλευε την διέλευση των πολιτών από την μία μεριά της πόλης στην άλλη. Έτσι, για να διευκολυνθεί η μετακίνηση των πολιτών, ο ποταμός Pregel συνδεόταν με 7 γέφυρες. Ωστόσο, δημιουργήθηκε το πρόβλημα εύρεσης μίας διαδρομής που να ξενικά από μία χερσαία περιοχή της πόλης, έστω A, να επισκέπτεται τις περιοχές B, C και D, και να καταλήγει πάλι στην περιοχή A, διασχίζοντας και τις 7 γέφυρες με τον περιορισμό ότι κάθε γέφυρα θα διασχιστεί ακριβώς μία φορά. (Εικόνα 1) Έπειτα από πολλές προσπάθειες, οι οποίες δεν απέφεραν αποτελέσματα, πολλοί κατέληξαν στο συμπέρασμα πως δεν υπήρχε τέτοια διαδρομή. Έτσι, το πρόβλημα της εύρεσης μίας τέτοιας διαδρομής πήρε το όνομα το Πρόβλημα των Γεφυρών του Königsberg και επικρατούσε μέχρι το 1736, όπου ο Ελβετός μαθηματικός Leonard Euler απέδωσε μία μαθηματική λύση για τη μη ύπαρξη λύσης. Η στρατηγική της λύσης ήταν να ονομάσει τις χερσαίες περιοχές της πόλης A, B, C και D, να μοντελοποιήσει το πρόβλημα των γεφυρών του Königsberg στο γράφημα της Εικόνας 1, αντιστοιχίζοντας τους κόμβους του γραφήματος στις χερσαίες περιοχές και τις ακμές στις 7 γέφυρες. Μετά τη μοντελοποίηση αυτή, γεννήθηκε ένα νέο πρόβλημα για την περίπτωση εάν το γράφημα περιέχει διαδρομή που να ξεκινά από έναν αρχικό κόμβο, να τερματίζει στον ίδιο κόμβο και να περιέχει όλες τις ακμές του μία μόνο φορά. Και σε αυτήν την περίπτωση ο Euler απέδειξε ότι στο γράφο δεν υπάρχει τέτοια διαδρομή και, επομένως, απέδειξε ότι το διάσημο πρόβλημα των 7 γεφυρών του

Konigsberg δεν έχει λύση. Παρόλο που η απόδειξη της λύσης του προβλήματος των γεφυρών του Konigsberg από τον Euler τεκμηριώθηκε ως ανεπαρκής, ο Euler απέδειξε γενικότερα πότε ένας γράφος έχει μία τέτοια διαδρομή, η οποία ονομάστηκε και διαδρομή Euler, ενώ ο γράφος ονομάστηκε γράφος Euler προς τιμήν του. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΟΣ, 2015) Επομένως, κάθε γνώση μπορεί να επεξεργαστεί και να δημιουργηθεί κατάλληλος γράφος για αυτήν. Αυτό οδηγεί σε έναν γράφο γνώσης, ο οποίος εμπεριέχει όλες τις απαραίτητες πληροφορίες και τις διασυνδέσεις τους. Έτσι, καθιστά την εύρεση της πληροφορίας ευκολότερη. Στην συνέχεια της πτυχιακής, θα γίνει αναφορά σε ειδικούς γράφους για την αντιμετώπιση προβλημάτων.



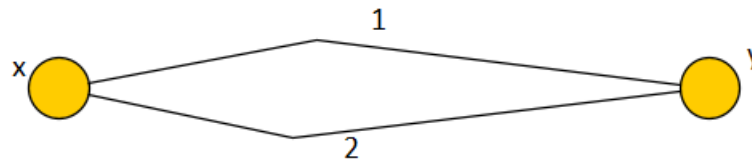
Εικόνα 1 Το σχήμα στα αριστερά αντιπροσωπεύει τις 7 γέφυρες του Konigsberg και το σχήμα στα δεξιά είναι παρουσίαση του αντίστοιχου γραφήματος Euler. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΟΣ, 2015)

1.2 Γενικά χαρακτηριστικά

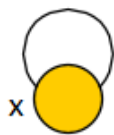
Η απάντηση στην ερώτηση τί είναι ένα γράφημα είναι η διατύπωση της ορολογίας του. Έτσι, η απάντηση είναι πως ένα γράφημα (graph) είναι ένα μαθηματικό ή συνδυαστικό αντικείμενο (mathematical or combinatorial object), το οποίο έχει την ιδιότητα να αναπαρασταθεί με εικόνες ή εικονογραφημένη αναπαράσταση (pictorial representation). Συγκεκριμένα, ένα γράφημα G είναι ένα ζεύγος συνόλων (V, E) , όπου V είναι ένα πεπερασμένο μη-κενό σύνολο η

στοιχείων και E ένα πεπερασμένο σύνολο m μη-διατεταγμένων ζευγών με στοιχεία του συνόλου V . (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

Σημαντικό είναι επίσης, να αναφερθεί από τι αποτελείται ένα γράφημα. Ένα γράφημα, λοιπόν, αποτελείται από τα στοιχεία του συνόλου, V που ονομάζονται κόμβοι (vertices) και E που ονομάζονται ακμές (edges). Τόσο οι κόμβοι, όσο και οι ακμές έχουν σύνολα, όπου ονομάζονται V σύνολο κόμβων (vertex set) και E το σύνολο ακμών (edge set) αντίστοιχα. Τα σύνολα V και E συμβολίζονται $V(G)$ και $E(G)$. Οι ακμές ενός γραφήματος $G = (V, E)$ συμβολίζονται $\{x, y\}$, (x, y) ή xy , όπου x, y , οι κόμβοι του γραφήματος. Επίσης, η ακμή xx που ανήκει στο $E(G)$ ονομάζεται βρόχος (loop). Στην εικόνα 3 παρουσιάζεται μία τέτοια ακμή. Δύο ή περισσότερες ακμές xy που ανήκουν στο $E(G)$ οι οποίες προσπίπτουν στο ίδιο ζεύγος κόμβων ονομάζονται παράλληλες ακμές (parallel edges). Στην Εικόνα 2 παρουσιάζονται παράλληλες ακμές. (Robinson, Webber, & Eifrem, 2015)



Εικόνα 2 Εδώ απεικονίζεται ένας γράφος G , όπου έχει 2 κόμβους τον x και y και δύο ακμές αριθμημένες, την 1 και 2. Το σύνολο κόμβων είναι το $V(G)=\{x, y\}$ και το σύνολο των ακμών είναι το $E(G)$ στο οποίο ανήκουν οι ακμές $\{x, y\}$ και $\{y, x\}$. Οι δύο αυτές ακμές είναι παράλληλες, καθώς προσπίπτουν στο ίδιο ζεύγος κόμβων.



Εικόνα 3 Στο σχήμα αυτό φαίνεται ένας κόμβος, ο x και μία ακμή η xx , η οποία θεωρείται βρόχος, καθώς ξεκινάει και τελειώνει στον ίδιο κόμβο.

Επομένως, ένα γράφημα $G = (V, E)$ είναι ένα μαθηματικό αντικείμενο που ορίζεται αυστηρά με δύο σύνολα: το σύνολο κόμβων $V(G)$ και το σύνολο ακμών $E(G)$.

Για παράδειγμα, έστω ότι έχουμε τα δύο ακόλουθα σύνολα:

$$V(G) = \{x, y, z, c, v, d\}$$

$$E(G) = \{xy, xz, zc, cv, ud\}$$

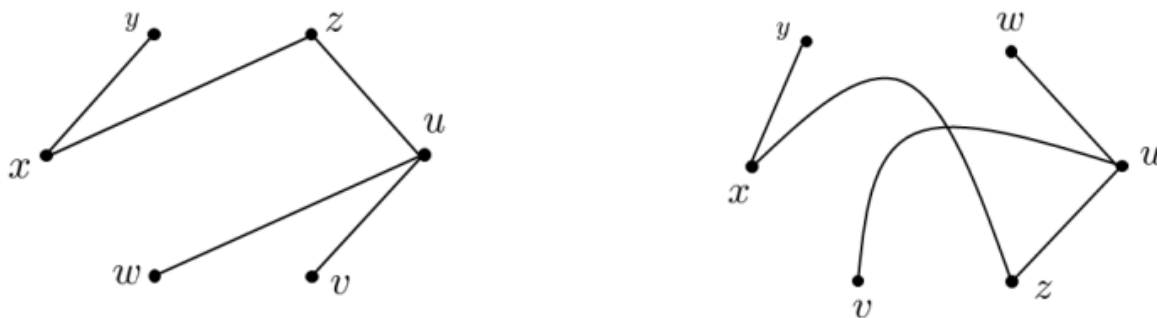
με βάση αυτά τα δύο σύνολα ορίζουμε ένα γράφημα G το οποίο απαρτίζεται από $n = 6$ και 5 ακμές.

Επιπρόσθετα, έχουμε τον πληθικό αριθμό των συνόλων $V(G)$ και $E(G)$ ενός γραφήματος G , που σημαίνει πως έχουμε το πλήθος των κόμβων και το πλήθος των ακμών του, το οποίο ονομάζεται τάξη (order) και μέγεθος (size) του γραφήματος G . Τέλος, συμβολίζεται με n η τάξη ενός γραφήματος και με m το μέγεθός του, εκτός εάν δηλώνεται διαφορετικά.

1.3 Εισαγωγή στην θεωρία γραφημάτων

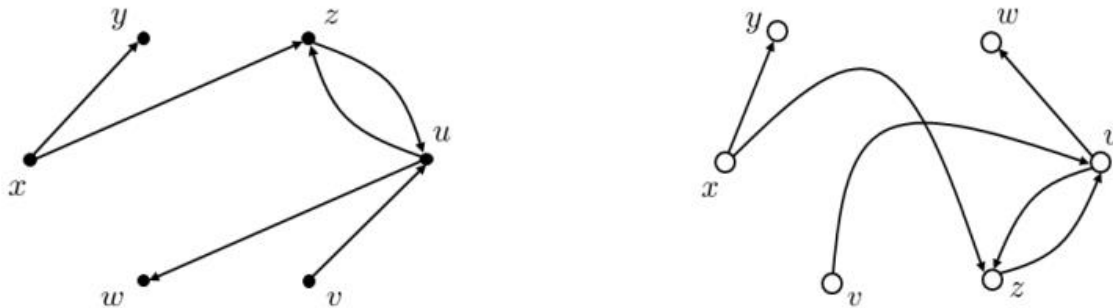
Οι τύποι των γραφημάτων γνώσης ποικίλουν και είναι εξίσου αναγκαίοι για την επίλυση προβλημάτων και την ανάκτηση πληροφοριών και γνώσης. Παρακάτω αναγράφονται οι κύριοι τύποι γραφημάτων.

Μη-κατευθυνόμενα γραφήματα: Υπάρχει μία θεώρηση για το σύνολο E του ζεύγους (V, E) , η οποία αναφέρει πως ένα σύνολο μη-διατεταγμένων ζευγών κόμβων ορίζει έναν τύπο γραφημάτων που ονομάζουμε μη-κατευθυνόμενα γραφήματα (undirected graphs). Παρακάτω ακολουθεί η Εικόνα 1.2, όπου παρουσιάζονται διάφορες αναπαραστάσεις ενός μη-κατευθυνόμενου γραφήματος G . Σε ένα μη-κατευθυνόμενο γράφημα G ισχύει $xy = yx$ για κάθε x, y που ανήκει στο $V(G)$. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)



Εικόνα 4 Διάφορες αναπαραστάσεις ενός μη-κατευθυνόμενου γραφήματος G . , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

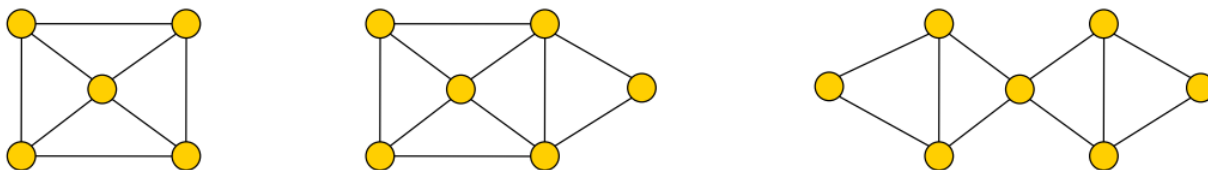
Κατευθυνόμενα γραφήματα: Η θεωρία των κατευθυνόμενων γραφημάτων γνωστοποιεί πως ένα σύνολο ακμών E , σε συνδυασμό με ένα σύνολο διατεταγμένων ζευγών κόμβων, δημιουργεί το ζεύγος συνόλων (V, E) , το οποίο ορίζει τον τύπο γραφημάτων που ονομάζουμε κατευθυνόμενα γραφήματα (directed graphs). Στην Εικόνα 1.3 παρουσιάζονται δύο διαφορετικές απεικονίσεις του ίδιου κατευθυνόμενου γραφήματος G . Θα συμβολίζονται με $V(G)$ και $E(G)$ τα σύνολα κόμβων και ακμών, αντίστοιχα, ενός κατευθυνόμενου γραφήματος G . Σημειώνουμε ότι, επειδή τα στοιχεία του $E(G)$ ενός κατευθυνόμενου γραφήματος G ορίστηκαν ως διατεταγμένα ζεύγη κόμβων, ισχύει ότι το xy είναι διαφορετικό από το yx για κάθε x, y που ανήκει στο $V(G)$. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)



Εικόνα 5 Διάφορες αναπαραστάσεις ενός κατευθυνόμενου γραφήματος G , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

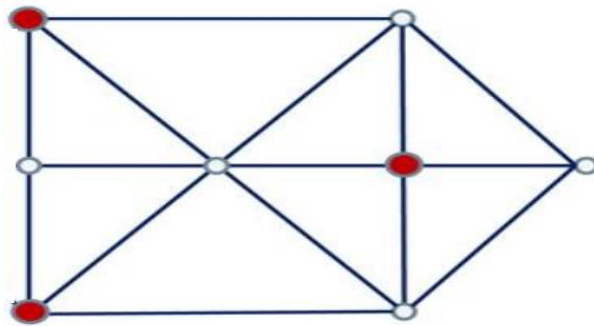
Διαφορές ανάμεσα σε μη-κατευθυνόμενα και κατευθυνόμενα γραφήματα G : Κύρια διαφορά ανάμεσα σε αυτούς τους παρόμοιους τύπους γραφημάτων είναι η αναπαράσταση τους. Όπως μπορείτε να δείτε και από τα Σχήματα 1.2 και 1.3, η αναπαράσταση ενός κατευθυνόμενου γραφήματος G , σε σχέση με ενός μη-κατευθυνόμενου, διαφοροποιείται μόνον ως προς τις ακμές, όπου χρησιμοποιούνται τόξα (arcs), αντί για γραμμές (lines), για να γίνεται εμφανής η διάταξη των κόμβων μίας κατευθυνόμενης ακμής xy ή (x, y) . Δηλαδή, μέσω της αναπαράστασης ενός κατευθυνόμενου γραφήματος φαίνεται το τόξο, καθώς και η αρχή του, από που αρχίζει στον κόμβο x , αλλά και το πέρας του, που τελειώνει στον κόμβο y .

Γράφημα Hamilton: Ο ορισμός του συγκεκριμένου γραφήματος διατυπώνει με σαφήνεια πως ένα γράφημα G ονομάζεται γράφημα Hamilton (Hamiltonian graph), εάν το G έχει ένα κύκλο Hamilton. Επίσης, το γράφημα G ονομάζεται ανιχνεύσιμο (traceable), εάν έχει μία διαδρομή Hamilton. Στην Εικόνα 1.4 φαίνονται 3 αναπαραστάσεις ενός γραφήματος Hamilton. (Νικολόπουλος, 2017-2018)



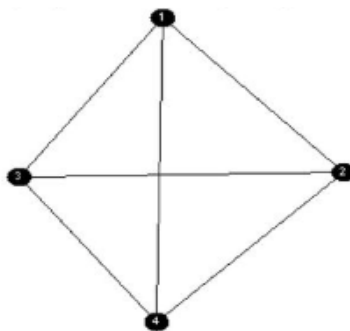
Εικόνα 6 Τρεις αναπαραστάσεις ενός γραφήματος Hamilton. , (Νικολόπουλος, 2017-2018)

Ανεξάρτητο ή ευσταθές γράφημα: Νωρίτερα, στην υπό-ενότητα 1.2 Γενικά χαρακτηριστικά, αναφερθήκαμε στο μέγεθος και την τάξη ενός γραφήματος. Αυτά τα ειδικά χαρακτηριστικά βοηθάνε στον προσδιορισμό ενός γραφήματος G . Αναλυτικότερα, για το μέγεθος ενός γραφήματος G τάξης n ισχύει ότι το m είναι μεγαλύτερο ή ίσο με το 0 και μικρότερο ή ίσο με το $n(n-1)/2$. Επομένως, εάν $m=0$, τότε το γράφημα ονομάζεται ανεξάρτητο ή ευσταθές (independent or stable). Στην Εικόνα 1.5 φαίνεται ένα ανεξάρτητο γράφημα. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)



Εικόνα 7 Τα συγκεκριμένο γράφημα έχει k κόμβους, οι οποίοι είναι ανεξάρτητοι μεταξύ τους, δηλαδή ανά δύο δεν έχουν καμία σχέση μεταξύ τους και δημιουργούν ένα ανεξάρτητο γράφημα. , (Νικολόπουλος , Θεωρία Γραφημάτων Ενότητα 9 Μετασχηματισμοί Υπολογιστικών Προβλημάτων, 2016-2017)

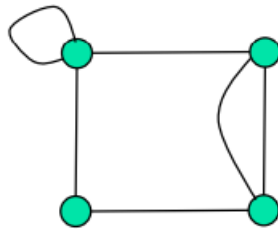
Πλήρες γράφημα: Ένα γράφημα ονομάζεται πλήρες όταν για το μέγεθός του ισχύει ότι $m = n(n-1)/2$, ονομάζεται πλήρες (complete). Επίσης, για ένα πλήρες γράφημα ισχύει ότι το πλήθος των ακμών του γραφήματός του με n κορυφές είναι $n(n-1)/2$. Στην Εικόνα 1.6 απεικονίζεται ένα πλήρες γράφημα. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)



Εικόνα 8 Πλήρες γράφημα G , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

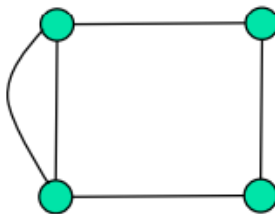
Τετριμμένο γράφημα: Ένα γράφημα G με έναν κόμβο και χωρίς ακμές ονομάζεται τετριμμένο (trivial). (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

Ψευδογράφημα: Ένα γράφημα με βρόχους ονομάζεται ψευδογράφημα (pseudograph). Στην Εικόνα 1.7 παρουσιάζεται η αναπαράσταση από ένα ψευδογράφημα. (Νικολόπουλος, 2017-2018)



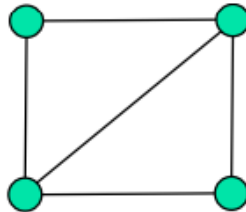
Εικόνα 9 Ψευδογράφημα G με βρόχους. , (Νικολόπουλος, 2017-2018)

Πολυγράφημα: Ένα γράφημα G με παράλληλες ακμές ονομάζεται πολυγράφημα (multigraph). Στην Εικόνα 1.8 φαίνεται ένα πολυγράφημα. (Νικολόπουλος, 2017-2018)



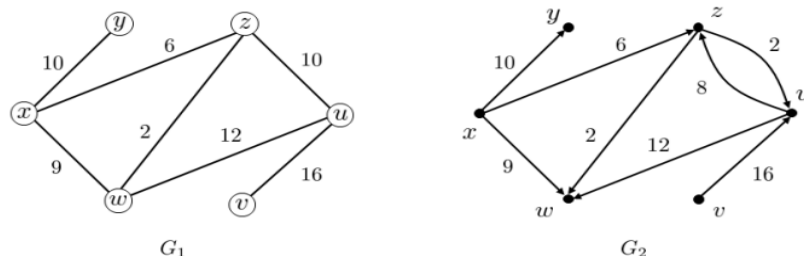
Εικόνα 10 Πολυγράφημα G ., (Νικολόπουλος, 2017-2018)

Απλό γράφημα: Ένα γράφημα, το οποίο δεν περιέχει βρόχους και παράλληλες ακμές, το ονομάζουμε απλό (simple). Στην Εικόνα 1.9 παρουσιάζεται ένα απλό γράφημα. (Νικολόπουλος, 2017-2018)



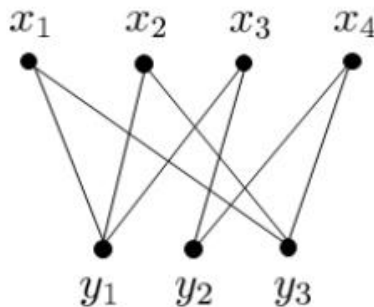
Εικόνα 11 Απλό γράφημα G . , (Νικολόπουλος, 2017-2018)

Έμβαρο γράφημα: Ένα γράφημα $G = (V, E)$ ονομάζεται έμβαρο (weighted graph), εάν σε κάθε ακμή e , που ανήκει στο E του G , έχει ανατεθεί ένας πραγματικός αριθμός $w(e)$. Το $w(e)$ ονομάζεται βάρος της ακμής e . Ένα έμβαρο γράφημα έχει δύο τύπους, δηλαδή μπορεί να είναι μη-κατευθυνόμενο γράφημα ή κατευθυνόμενο γράφημα. Στην Εικόνα 1.10 έχουμε 2 έμβαρα γραφήματα από τα οποία το ένα είναι μη-κατευθυνόμενο, ενώ τα άλλο κατευθυνόμενο. Στο πρακτικό επίπεδο, για παράδειγμα σε μία εφαρμογή, τα βάρη ενός έμβαρου γραφήματος θα μπορούσαν να αντιπροσωπεύουν, τα έξοδα, τις αποστάσεις, τους χρόνους, τις ικανότητες ή ακόμα και τους βαθμούς φιλίας στα κοινωνικά δίκτυα. Μία ακόμα εφαρμογή των έμβαρων γραφημάτων είναι τα δίκτυα (networks). (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

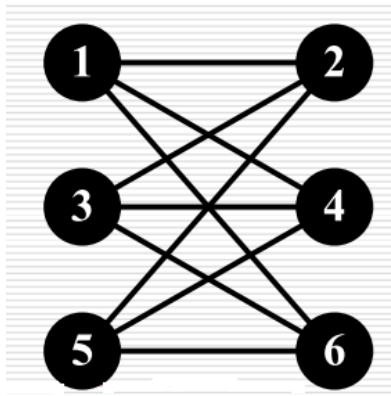


Εικόνα 12 Δύο έμβαρα γραφήματα, τα οποία το G_1 είναι μη-κατευθυνόμενο και το G_2 είναι κατευθυνόμενο. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

Διμερές γράφημα και Πλήρες Διμερές γράφημα: Ένα μη-κατευθυνόμενο γράφημα $G = (V, E)$ ονομάζεται διμερές, εάν το σύνολο των κόμβων του V μπορεί να διαμεριστεί σε δύο ευσταθή σύνολα $V = S_1 + S_2$ (το 1 και 2 στα ποδαράκια) και, επομένως, κάθε ακμή του γραφήματος έχει το ένα άκρο στο S_1 και το άλλο στο S_2 . Συχνά θα χρησιμοποιούμε το συμβολισμό $G = (S_1, S_2, E)$, όταν θέλουμε να δώσουμε έμφαση στη διαμέριση. Θα λέμε ότι οι κόμβοι x που ανήκει στο S_i και y που ανήκει στο S_j είναι ίδιας ισοτιμίας (same parity), εάν $i = j$ και αντίθετης ισοτιμίας (opposite parity), εάν i είναι διαφορετικό από το j . Στην Εικόνα 1.11 φαίνεται ένα παράδειγμα ενός διμερούς γραφήματος. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015) Επίσης, έχουμε και το πλήρες διμερές γράφημα. Αυτό το γράφημα το συγκροτεί ένα διμερές γράφημα $G = (S_1, S_2, E)$, εάν για κάθε κόμβο x , που ανήκει στο S_1 , και κάθε κόμβο y , που ανήκει στο S_2 , ισχύει xy ανήκει στο E . Προφανώς, το πλήθος m των ακμών ενός πλήρους διμερούς γραφήματος είναι $m = |S_1| \cdot |S_2|$. Στην Εικόνα 1.12 απεικονίζεται ένα πλήρες διμερές γράφημα. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015) Τέλος, σημαντικό είναι να αναφερθεί πως οι γράφοι γνώσης συχνά είναι διμερή γραφήματα, επομένως η πλήρη κατανόηση του παραπάνω ορισμού είναι απαραίτητη.

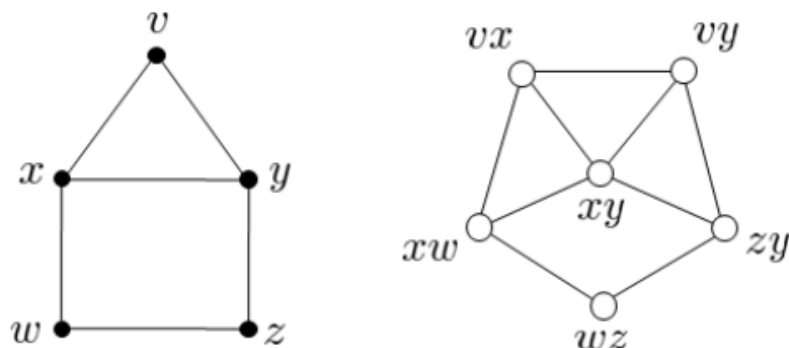


Εικόνα 13 Διμερές γράφημα G , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΛΗΣ, 2015)



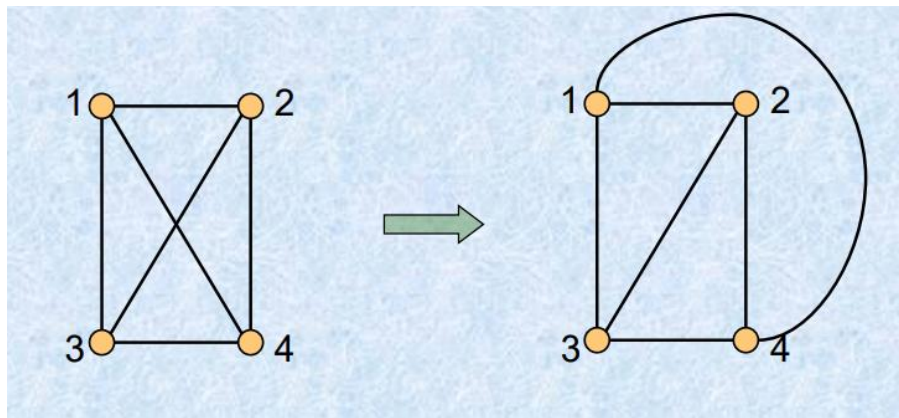
Εικόνα 14 Πλήρες διμερές γράφημα G . , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

Γραμμικό γράφημα ή L-γράφημα: Έστω $G = (S_1, S_2, E)$ ένα μη-κατευθυνόμενο γράφημα τάξης n και μεγέθους m . Το γραμμικό γράφημα $L(G)$ (ή L-γράφημα) του G είναι ένα γράφημα τάξης m με κόμβους $v_1, v_2, \dots, v_m$, οι οποίοι αντιστοιχούν στις m ακμές $e_1, e_2, \dots, e_m$ του G και ακμές $v_i v_j$, για τις οποίες ισχύει: οι αντίστοιχες ακμές e_i και e_j των κόμβων v_i και v_j πρόσκεινται στο ίδιο κόμβο στο γράφημα G . Στην Εικόνα 1.13 φαίνεται ένα γράφημα και το L-γράφημά του.



Εικόνα 15 Το αριστερό σχήμα είναι ένα μη-κατευθυνόμενο γράφημα G και το δεξί σχήμα είναι το αντίστοιχο L-γράφημά του. , (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

Επίπεδο γράφημα: Ένα γράφημα λέγεται επίπεδο (planar), εάν μπορεί να σχεδιασθεί στο επίπεδο χωρίς τεμνόμενες ακμές. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015) Στην Εικόνα 1.14 απεικονίζεται ένα γράφημα και το πως μπορεί να σχεδιαστεί, ώστε να είναι το αντίστοιχο επίπεδο γράφημα.



Εικόνα 16 Το αριστερό σχήμα είναι ένα γράφημα και το δεξί σχήμα είναι το αντίστοιχο επίπεδο γράφημά του. (έχω ένα κομμάτι της έρευνας και δεν έχει στοιχεία για να μπορέσω να βάλω βιβλιογραφία)

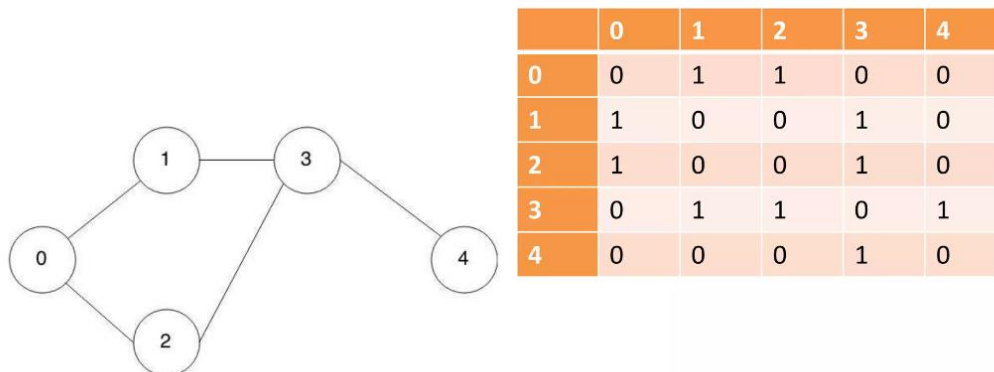
K-κομβικά-χρωματίσιμο γράφημα: Ένα γράφημα χωρίς βρόχους G ονομάζεται k -κομβικά-χρωματίσιμο (k -vertex-colorable) ή, ισοδύναμα, k -χρωματίσιμο (k -colorable), εάν οι κόμβοι του μπορούν να χρωματισθούν με k χρώματα, έτσι ώστε δύο κόμβοι που συνδέονται με ακμή να μην έχουν το ίδιο χρώμα. (ΓΕΩΡΓΙΑΔΗΣ, ΝΙΚΟΛΟΠΟΥΛΟΣ, & ΠΑΛΗΟΣ, 2015)

K-ακμικά-χρωματίσιμο γράφημα: Ένα γράφημα G ονομάζεται k -ακμικά-χρωματίσιμο (k -edge-colorable), εάν οι ακμές του μπορούν να χρωματισθούν με k χρώματα, έτσι ώστε δύο ακμές προσπίπτουσες στον ίδιο κόμβο να μην έχουν το ίδιο χρώμα.

Τριγωνικό γράφημα: Εάν είναι τέλει σχήμα απαλοιφής, τότε το γράφημα G είναι τριγωνικό, άλλως το G δεν είναι τριγωνικό.

Συμπερασματικά, γίνεται αντιληπτή η σημαντικότητα ύπαρξης των γραφημάτων γνώσης, καθώς και η χρησιμότητά τους στην αντιμετώπιση διάφορων προβλημάτων. Δύο κύριοι κλάδοι στους οποίους βλέπουμε μεγάλη ανάγκη εμπλουτισμού τους με τα γραφήματα γνώσης είναι ο βιολογικός και ο οικονομικός.

Είναι σημαντικό, επίσης, να αναφερθεί ο όρος του πίνακα γειτνίασης, ο οποίος είναι άμεσα συνδεδεμένος με τους γράφους. ο πίνακας γειτνίασης, λοιπόν, είναι ένας τετραγωνικός πίνακας που αναπαριστά έναν γράφο. Στην Εικόνα 17 δίνεται ένα παράδειγμα.



Εικόνα 17 Στα αριστερά απεικονίζεται ένας γράφος και στα δεξιά υπάρχει ο πίνακας γειτνίασης του γράφου. ,
(Αναπαράσταση Γραφημάτων, 2015)

Κεφάλαιο 2^ο: Βιβλιογραφική Ανασκόπηση

Εισαγωγή

Για την συγγραφή της πτυχιακής αυτής, προηγήθηκε έρευνα στη χρήση των γραφημάτων γνώσης σε βιολογικό και οικονομικό επίπεδο και μελετήθηκε ο τρόπος προσέγγισης και οι λύσεις του προβλήματος της κάθε έρευνας με την χρήση γράφων γνώσης. Παρακάτω αναφέρεται το πρόβλημα που καλείται να αντιμετωπίσει η κάθε έρευνα, ο ειδικός γράφος γνώσης που χρησιμοποιείται για την επίλυση του αντίστοιχου προβλήματος, καθώς και οι εφαρμογές αυτού, που σχετίζονται με το θέμα της συγκεκριμένης πτυχιακής.

2.1 Γραφική Αναπαράσταση Βάσεων Δεδομένων (Graph Databases).

Στις μέρες μας, θεωρείται πως υπέρτατο αγαθό είναι η πληροφορία λόγω της ραγδαίας ανάπτυξης της πληροφορικής. Ο ορισμός της πληροφορίας περιλαμβάνει ειδήσεις, γεγονότα και γενικότερα δεδομένα, τα οποία πληθαίνουν καθημερινά. Όλες αυτές οι πληροφορίες αποτελούν έναν τεράστιο όγκο δεδομένων, που πρέπει να επεξεργαστούν και να ταξινομηθούν κατάλληλα, ώστε να αποφευχθούν προβλήματα. Αναλυτικότερα, κάποια από αυτά τα προβλήματα μπορεί να αναφέρονται σε:

- πλεονασμό των δεδομένων, που συμβαίνει σε περίπτωση επανάληψης των ίδιων δεδομένων σε αρχεία διαφορετικών εφαρμογών,
- ασυνέπεια των δεδομένων, που είναι πιθανό όταν υπάρχει ήδη πλεονασμός και απαιτείται η διόρθωση πολλών αρχείων, αλλά είναι δυνατή η διόρθωση μόνο ενός από αυτά,
- αδυναμία μερισμού δεδομένων, που προσδιορίζει το μερισμό δεδομένων, δίνει τη δυνατότητα για κοινή χρήση των στοιχείων και η ύπαρξη της αδυναμίας αυτής δημιουργεί καθυστέρηση στη λήψη αποφάσεων και

- αδυναμία προτυποποίησης, που έχει να κάνει με την ανομοιομορφία, με την διαφορετική αναπαράσταση και οργάνωση των δεδομένων στα αρχεία και οδηγεί σε προβλήματα προσαρμογής και ανταλλαγή δεδομένων μεταξύ συστημάτων.

Λύση στα παραπάνω προβλήματα επιφέρει η ανακάλυψη των Βάσεων Δεδομένων. Σημαντικό ρόλο είχε η χρήση των ηλεκτρονικών υπολογιστών και κατά επέκταση η ηλεκτρονική καταχώρηση και επεξεργασία δεδομένων τη δεκαετία του '70. Επίσης, απαραίτητη ήταν η συνένωση των αρχείων, καθώς και η σωστή οργάνωσή τους. Μέσα από αυτά δημιουργήθηκαν οι Βάσεις Δεδομένων (Data Bases). Αναλυτικότερα, μια Βάση Δεδομένων είναι ένα σύνολο οργανωμένων αρχείων, τα οποία είναι συνδεδεμένα μεταξύ τους με λογικές σχέσεις. Μέσα από τις σχέσεις αυτές μπορούν να χρησιμοποιούνται δεδομένα από πολλές εφαρμογές και από πολλούς χρήστες ταυτόχρονα. Αυτό δίνει τη δυνατότητα μετάδοσης της πληροφορίας αποτελεσματικά και ταχύτατα. Επιπρόσθετα, τα συστήματα βάσεων δεδομένων χρησιμοποιούνται, ώστε να αποθηκεύεται και να επεξεργάζεται αυτός ο τεράστιος όγκος των πληροφοριών, που αναφέρθηκε παραπάνω. Με αυτόν τον τρόπο, αντικαθίστανται οι πίνακες αποθήκευσης δεδομένων. Τέλος, τα ερωτήματα πολλαπλών βημάτων εκτελούνται με βελτιστοποιημένο σχεδιασμό ερωτημάτων, σε αντίθεση με τις σχεσιακές βάσεις δεδομένων που απαιτούν αργές, δαπανηρές λειτουργίες σύνδεσης, γεγονός που παρακινεί τη χρήση βάσεων δεδομένων γραφημάτων. Συμπερασματικά, η βάση δεδομένων είναι ένα δυναμικό σύστημα αποθήκευσης και διαχείρισης δεδομένων, που ανάλογα με την χρήση του και το αντικείμενο στο οποίο αναφέρεται προσαρμόζεται κατάλληλα.

Προχωρώντας από την θεωρία στην πράξη, μία βάση δεδομένων χρησιμοποιεί τη θεωρία γραφημάτων για την αποθήκευση, χαρτογράφηση και διερεύνηση σχέσεων μεταξύ των δεδομένων. Συγκεκριμένα, δημιουργείται μία συλλογή από άκρα και κόμβους. Κάθε κόμβος αντιπροσωπεύει μια οντότητα και ορίζεται από ένα μοναδικό αναγνωριστικό, ενώ κάθε άκρο αντιπροσωπεύει μία σχέση μεταξύ δύο κόμβων και έχει τις ιδιότητες των κόμβων που δηλώνονται ως κλειδί-τιμή. Μια άκρη ορίζεται επίσης από ένα μοναδικό αναγνωριστικό και έχει έναν κόμβο έναρξης και ένα τέλος, καθώς και ένα σύνολο ιδιοτήτων. Μία βάση δεδομένων γραφημάτων, λοιπόν, είναι απαραίτητη για την ανάλυση των σχέσεων μεταξύ των δεδομένων και έχει ως επακόλουθο την εξόρυξη δεδομένων. Στην

συγκεκριμένη πτυχιακή, χρησιμοποιήθηκε το σύστημα διαχείρισης βάσεων δεδομένων γραφημάτων Neo4j.

2.2 Το σύστημα διαχείρισης βάσεων δεδομένων γραφημάτων Neo4j.

Το Neo4j αποτελεί ένα σύστημα διαχείρισης βάσεων δεδομένων γραφημάτων, που αναπτύχθηκε από τη Neo4j, Inc. Στο Neo4j αποθηκεύονται στοιχεία σε μορφή κόμβων και η σχέσεις που έχουν τα στοιχεία μεταξύ τους παρουσιάζονται με άκρες. Το Neo4j χαρακτηρίζεται από υψηλή ταχύτητα, ασφάλεια, ακεραιότητα και αποδοτικότητα, στοιχεία που το κάνουν πρώτη επιλογή για τις επιχειρήσεις ανάμεσα σε πολλές βάσεις δεδομένων γραφημάτων. Αυτό συμβαίνει, διότι μπορεί να φέρει εις πέρας δοκιμασίες και απαιτήσεις που κατέχουν μεγάλο όγκο δεδομένων και μπορεί να απαντήσει σε σύνθετες επιχειρηματικές ερωτήσεις, που έχουν βαθιές πληροφορίες από συνδεδεμένα δεδομένα. Επίσης, μπορεί να υποστηρίξει διάφορες επιλογές, όπως cloud, υβριδική, multi-cloud και μέσα από αυτό προσφέρονται πολλές δυνατότητες. Ένα ακόμη χαρακτηριστικό, είναι η εύκολη οριζόντια κλιμάκωση του Neo4j. Αυτό συνδέεται με τις επιλογές που προαναφέρθηκαν, καθώς παρέχει μια αρχιτεκτονική, που προσαρμόζεται ανάλογα με τις ανάγκες του χρήστη και των δεδομένων. Η αυτόνομη ομαδοποίηση δίνει παραπάνω εφόδια στην συγκεκριμένη βάση δεδομένων, αφού επιτρέπει στον χρήστη να κλιμακώνει οριζόντια τα δεδομένα και έτσι, απαιτείται λιγότερη χειροκίνητη δουλειά. Σημαντικό είναι να τονιστεί το ότι η ιδιότητα Cypher Parallel Runtime παρέχει την ιδιότητα της βελτιστοποιημένης εκτέλεσης αναλυτικών ερωτημάτων, που αφορούν μεγάλα τμήματα.

2.3 Έρευνα: Κατασκευή γραφημάτων γνώσης στον βιοϊατρικό κλάδο.

Θέμα μελέτης της συγκεκριμένης έρευνας ήταν η εύρεση λύσης στο ερώτημα πώς θα μπορούσαν οι μέθοδοι μηχανικής εκμάθησης να συμβάλουν σε προβλέψεις που αφορούν γονιδιωματικούς, φαρμακευτικούς και κλινικούς τομείς. Αναλυτικότερα, γίνεται αναφορά στην αναπαραστατική μάθηση και τον τρόπο χρήσης της, ώστε να δημιουργηθούν προγνωστικά.

Η αναπαραστατική μάθηση είναι μία διαδικασία κατά την οποία αναπαριστώντα γραφήματα γνώσης σε χώρο χαμηλών διαστάσεων, καθώς και ο τρόπος με τον οποίο δεδομένα υψηλών διαστάσεων μπορούν να χαρτογραφηθούν σε γραφήματα χαμηλών διαστάσεων. Στόχος αυτής της

διαδικασίας είναι η εύρεση και διατήρηση της δομής ενός γραφήματος γνώσης, διαφορετικό ανάλογα με το κάθε πρόβλημα, και ταυτόχρονα η δημιουργία της αντίστοιχης αναπαράστασης του. Αυτή η αναπαράσταση έχει την δυνατότητα να χρησιμοποιηθεί σε μεθόδους μηχανικής μάθησης για την δημιουργία προγνωστικών. Μέσα από την αναπαράσταση γραφημάτων γνώσης σε χώρο χαμηλών διαστάσεων, αναγνωρίζεται εύκολα η ουσία των γραφημάτων γνώσης ως πυκνά διανύσματα. Έτσι, τα διανύσματα αυτά είναι οι κόμβοι των γραφημάτων. Αυτό που είναι σημαντικό να αναφερθεί είναι πως σε αυτήν την περίπτωση υπάρχει μεγάλη ανάγκη συνεχούς πληροφόρησης για τον τρόπο σύνδεσης και επικοινωνίας των κόμβων μεταξύ τους. Ωστόσο, αυτό δεν συμβαίνει με τις ακμές, οι οποίες μπορούν να ξανά χρησιμοποιηθούν οι ίδιες. Μόλις ο χώρος αυτός σχηματιστεί είναι εφικτές οι τεχνικές μηχανικής μάθησης.

Επιπρόσθετα, μπορεί να γίνει ομαδοποίηση μέσω αναλύσεων και να προκύψουν οι κατηγορίες: παραγοντοποίηση μήτρας, μεταφραστική απόσταση και μοντέλα νευρωνικών δικτύων. (Nicholson & Greene, 2020). Στην κατηγορία παραγοντοποίηση μήτρας χρησιμοποιήθηκε ο αλγόριθμος SVD (έναν τετραγωνικός πίνακας γειτνίασης, που τα στοιχεία του είναι κόμβοι και η καταχώρηση είναι μία αναπαράσταση των ακμών των κόμβων). Με τη βοήθεια αυτού του αλγόριθμου, απεικονίστηκαν γραφήματα γνώσης σε χώρο χαμηλών διαστάσεων. Η διαδικασία αυτή αποτελείται από κάποια βήματα. Αρχικά, πρέπει να διατηρηθεί η δομή των χαμηλών διαστάσεων, ώστε τα στοιχεία να είναι οι κόμβοι και οι ακμές, η απόσταση μεταξύ των κόμβων αυτών. Έπειτα, θα πρέπει να υπάρχει ένας πίνακας γειτνίασης, ένας πίνακας βαθμών, ο οποίος προκύπτει από την αναπαράσταση γραφημάτων και ένας πίνακας μοιρών (διαγώνιος πίνακας, όπου κάθε καταχώρηση αντιπροσωπεύει τον αριθμό των άκρων που συνδέονται με τον κόμβο). Στην δεύτερη κατηγορία, τα μοντέλα μεταφραστικής απόστασης, αντιπροσωπεύουν τις ακμές του γραφήματος γνώσης ως γραμμικοί μετασχηματισμοί. (Nicholson & Greene, 2020) Μέσα από τους μετασχηματισμούς αυτούς δημιουργούνται εξισώσεις, που μιμούνται τις εικονικές μεταφράσεις διανυσμάτων λέξεων, με το μοντέλο word2vec. Η τελευταία κατηγορία, τα νευρωνικά δίκτυα, είναι μία κατηγορία μοντέλων μηχανικής μάθησης, που είναι συνδεδεμένη με την έννοια των βιολογικών νευρωνικών δικτύων. Τις παραπάνω 3 κατηγορίες φαίνεται πως μπορούμε να τις συνδυάσουμε και να έχουμε ένα θεμιτό αποτέλεσμα. Τα γραφήματα γνώσης που μπορούν να προκύψουν, λέγεται πως έχουν εφαρμοστεί σε πολλές βιοϊατρικές προκλήσεις, που κυμαίνονται από τον προσδιορισμό των λειτουργιών των πρωτεϊνών έως την επισήμανση των καρκινικών γονιδίων έως τη σύσταση ασφαλέστερων φαρμάκων για τους ασθενείς. (Nicholson & Greene, 2020)

Τέλος, γίνεται αναφορά στις Multi-omic εφαρμογές, οι οποίες χρησιμοποιούν γραφήματα γνώσης, με σκοπό τη μελέτη του γονιδιώματος. Συγκεκριμένα, κατασκευάστηκε ένα δίκτυο miRNA-Ασθένειας, χρησιμοποιώντας τον άνθρωπο ως Βάση δεδομένων MicroRNA Disease (HMDD). Δημιούργησε μια μήτρα γειτνίασης με τις σειρές που αντιπροσωπεύουν το miRNA και τις στήλες που αντιπροσωπεύουν ασθένειες. Με αυτόν τον τρόπο, η μήτρα αποσυντέθηκε σε μικρούς ορθογώνιους πίνακες, χρησιμοποιώντας SVD, και στη συνέχεια αυτοί οι μικροί πίνακες χρησιμοποιήθηκαν για τον υπολογισμό των βαθμολογιών ομοιότητας μεταξύ miRNA και ασθενειών. Επίσης, οι Yang χρησιμοποίησαν node2vec για την δημιουργία ενός συστήματος συστάσεων. Με αυτό το σύστημα γίνονται επιτυχείς συσχετίσεις μεταξύ γονιδίων και συμπτωμάτων ασθενειών. Με βάση αυτά, κατασκευάστηκε ένα γράφημα γνώσης συμπτωμάτων γονιδιακής νόσου. Επιπλέον, σε άλλες προσεγγίσεις έχουν χρησιμοποιήσει μοντέλα γραφημάτων για την εκτέλεση κλινικών εργασιών με στόχο την εύρεση της σωστής σύστασης φαρμάκου και διαγνώσεις ασθενών. Επομένως, με αυτήν την έρευνα φαίνεται πως τα γραφήματα γνώσης έχουν δείξει αρκετά ελπιδοφόρα αποτελέσματα και μεθόδους αντιμετώπισης προβλημάτων στον τομέα της υγείας.

2.4 Έρευνα: Εφαρμογές στην υγειονομική περίθαλψη με χρήση την ανάλυση γράφων.

Αυτή η έρευνα έλαβε χώρα με αφορμή την πρόσφατη παγκόσμια πανδημία γνωστή ως νόσος του κορονοϊού 2019 (COVID-19). Η πανδημία αυτή σάρωσε τις ηπείρους ολόκληρης της υδρόγειου και δημιούργησε τόσο προβλήματα στην λειτουργία των υγειονομικών κέντρων όσο και ραγδαία πτώση του συνολικού πληθυσμού. Στην έρευνα αυτή γίνεται γνωστή η εξόρυξη πληροφορίας βάση γραφημάτων στην βιοϊατρική, δηλαδή στην ενσωμάτωση δικτύου, την κατασκευή γραφημάτων γνώσης και την ακολουθία συμπερασμάτων, τα μοντέλα γενετικών γραφημάτων και τέλος η ερμηνεία, η ασφάλεια και η προκατάληψη γύρω από τα γραφήματα. Είναι γνωστό, πως για την αντιμετώπιση του προβλήματος χρειάζονται πληροφορίες για αυτό. Έτσι, συγκεντρώθηκαν πληροφορίες με δεδομένα σχετικά με τον COVID-19 από τα κλινικά ιδρύματα, τη φαρμακευτική βιομηχανία, τη βιοϊατρική, καθώς και από άλλα παρεμφερή μέρη. Αυτή η διαδικασία περισυλλογής δεδομένων είναι ιδιαίτερα απαραίτητη και απαιτητική για την κατανόηση της μετάδοσης, της θεραπείας και της πρόληψης της κάθε νόσου. Και σε αυτήν την έρευνα έγινε χρήση γραφημάτων. Αναλυτικότερα, χρησιμοποιήθηκε ένας αλγόριθμος, ο οποίος βασίζεται σε γραφήματα και με βάση την έρευνα

αποτελεί δημοφιλές εργαλείο για την ανάλυση διασυνδεδεμένων οντοτήτων δεδομένων. Ο συγκεκριμένος αλγόριθμος, λοιπόν, μπορεί να μοντελοποιήσει τόσο τα χαρακτηριστικά δεδομένων όσο και της σχέσεις ταυτόχρονα. Αυτό έρχεται σε αντίθεση με της μεθοδολογίες που αντιμετωπίζουν τα δεδομένα ως ανεξάρτητα το ένα από το άλλο και πανομοιότυπα κατανεμημένα. Αυτή η διαφορά είναι που αναδεικνύει την έρευνα και την κάνει μοναδική, καθώς η υγειονομική περίθαλψη αποτελείται από πολύπλοκα, πολύ ερωτογενή και ανόμοια δεδομένα. (Wang, Cui, Pei, Song, & Zang, 2020)

2.5 Έρευνα: Ο οικονομικός σχεδιασμός και η τεχνική υλοποίηση με βάση γράφους γνώσης.

Η ακόλουθη έρευνα, αναφέρεται σε προβλήματα του οικονομικού τομέα και σε τρόπους αντιμετώπισης αυτών των προβλημάτων. Συγκεκριμένα, τονίζεται η ανάγκη χρήσης γράφων γνώσης με γνώμονα την ορθολογική λήψη αποφάσεων. Αναλυτικότερα, στον σχεδιασμό γράφων γνώσης όσο πιο σωστή είναι η λήψη μίας απόφασης, τόσο πιο βέλτιστη θα είναι η λύση. Η έρευνα αυτή, προτείνει να γεφυρωθεί ο αμφίδρομος σχεδιασμός των γράφων, που υπάρχει λόγω του σχεδιασμού με βάση την οικονομία και της εφαρμογής της τεχνολογίας. Αυτό θα μεγιστοποιήσει την αποδοτικότητα και θα επιφέρει ραγδαία ανάπτυξη στην οικονομία. Αποδεικνύεται, επίσης, πως αυτό μπορεί να επιτευχθεί με βάση το γράφημα δεδομένων και τη γνώση γραφικής παράστασης. (Shao, et al., 2017) Με αυτόν τον τρόπο, έγινε ανάλυση των προβλημάτων της συγκεκριμένης έρευνας, που είχαν αρνητικό αντίκτυπο σχετικά με τις δραστηριότητες ανάπτυξης λογισμικού(απαιτήσεις ανάλυσης, σύνοψης και σχεδιασμού λεπτομερειών). Επιπλέον, προτείνεται η βελτίωση της αξιοπιστίας και της ευρωστίας του συστήματος, κάνοντας διαχείριση δεδομένων και επαναχρησιμοποίηση των πληροφοριών που ήδη κατέχονται. Σημαντικό είναι να αναφερθεί, πως η ανάπτυξη λογισμικού, θεωρείται το κλειδί του συνόλου σχεδιασμού του συστήματος. (Shao, et al., 2017) Για όσα αναφέρθηκαν παραπάνω υπάρχει μία διαδικασία, ώστε να γίνει η υλοποίηση. Αρχικά, γίνεται η αξιολόγηση των διαφόρων περιορισμών, που επηρεάζουν το έργο, ώστε να γίνει ο σχεδιασμός. Έπειτα, αναπτύσσεται ένα ασύρματο σύστημα δικτύου αισθητήρων, αναλύεται συνολικά δομή και διαδικασία ανάπτυξης του συστήματος σε δεδομένα, σε επίπεδο πληροφόρησης και σε επίπεδο γνώσης, αντίστοιχα. Πρέπει να γίνει γνωστό ότι η ανάλυση απαιτήσεων και η απόκτηση απαιτήσεων πραγματοποιούνται ταυτόχρονα, όπως τονίζεται στην έρευνα. Σε περίπτωση λάθους κατά τον έλεγχο

πληροφοριών ή και απώλεια πληροφοριών, φαίνεται πως το επακόλουθο θα είναι η καθυστέρηση ή μέχρι και η αποτυχία των δραστηριοτήτων μοντελοποίησης. Επομένως, αναμενόμενο είναι και το ότι το χρονικό διάστημα τέλεσης της διαδικασίας, θα υπερβεί τον αναμενόμενο χρόνο. Αυτό συμβαίνει, καθώς, υπάρχει σχέση εξάρτησης. Προχωρώντας στο πρακτικό κομμάτι, η διαδικασία σχεδιασμού λεπτομερειών λογισμικού, γίνεται με Unified Modeling Language (UML) από της προγραμματιστές. Σε αυτήν την περίπτωση το γράφημα γνώσης κατέχει φυσική σημασιολογία και μπορεί να περιέχει πιο ολοκληρωμένες πληροφορίες. Τέλος, προτείνεται η εμβάθυνση εύρεσης λύσης, η οποία θα έχει ως χαρακτηριστικό τον γνώμονα της αξίας, δηλαδή την υπηρεσία μοτίβων αξίας (SVB), με στόχο την αντιμετώπισης της οικονομικής ανάλυσης για την υλοποίηση της πληροφορικής.

2.6 Έρευνα: Νευροσυμβολική αναπαράσταση με βάση γράφους γνώσης με βιολογικά δεδομένα.

Σε αυτήν την έρευνα, όπως και στην έρευνα που είναι στην υπό ενότητα 2.2, βλέπουμε πως εξετάστηκαν οι τεχνικές των γραφημάτων γνώσης σε βιολογικά δεδομένα. Συγκεκριμένα, αναπτύχθηκε μια νέα μέθοδο για την εκμάθηση χαρακτηριστικών σε γραφήματα βιολογικής γνώσης. Στόχος της έρευνας ήταν να βρεθεί μία μέθοδος, η οποία θα μπορούσε να ενσωματώσει κόμβους, που κωδικοποιούν σχετικές πληροφορίες με χρήση γραφημάτων γνώσης. Επιπρόσθετα, γίνεται μία λεπτομερής αναφορά στην μέθοδο, που χρησιμοποίησαν οι ερευνητές για να μπορέσουν να επιτύχουν τον στόχο τους, η οποία φαίνεται πως συνδυάζει συμβολικές μεθόδους αναπαράστασης γνώσης. Αναφέρεται πως για να μπορέσουν να συνδυάσουν αυτές τις μεθόδους, χρησιμοποίησαν συμβολική λογική και αυτοματοποιημένη συλλογιστική, με νευρωνικά δίκτυα. Είναι αξιοσημείωτο να σημειωθεί πως η μέθοδος που αναγράφεται στην έρευνα, μπορεί να εφαρμοστεί σε οποιοδήποτε γράφημα βιολογικής γνώσης. Αυτό είναι ιδιαίτερα σημαντικό, καθώς με αυτόν τον τρόπο θα αυξηθεί ο αριθμός των βάσεων γνώσης, που λαμβάνουν χώρο στη βιολογία. Επιπλέον, γίνεται αναφορά στα Συνδεδεμένα Δεδομένα και λέγεται πως μπορούν να επιτρέψουν τη δια λειτουργικότητα μεταξύ πολλών βάσεων δεδομένων. Η διαδικασία για να συμβεί, χρειάζεται απλώς να επαναχρησιμοποιηθούν τα αναγνωριστικά και να γίνουν ερώτημα σε no-SQL γλώσσες, όπως η SPARQL που μπορεί να εκτελέσει καταναμεμημένα ερωτήματα σε πολλαπλές βάσεις δεδομένων. Ένα δεύτερο σημαντικό συστατικό των εφαρμογών του Σημασιολογικού Ιστού στις βιοεπιστήμες υπήρξε

η ανάπτυξη και χρήση οντολογιών. Η έρευνα εξηγεί πως οι οντολογίες χρησιμοποιούνται πλέον ευρέως σε βιολογικά σύνολα δεδομένων για τον σχολιασμό και την παροχή μεταδεδομένων. Ενώ, οι περισσότερες από αυτές τις μεθόδους έχουν αναπτυχθεί για γραφήματα με ετικέτες εκτός άκρων, αυτό έρχεται σε αντίθεση με τα Συνδεδεμένα δεδομένα, τα οποία έχουν ετικέτες άκρων που αντιπροσωπεύουν τον τύπο σχέσης μεταξύ των οντοτήτων των γραφημάτων. Υπάρχουν και μέθοδοι, οι οποίες έχουν επεκταθεί για να ενσωματώσουν ετικέτες άκρων. (Alshahrani, et al., 2017) Στην έρευνα αναγράφεται πως ανέπτυξαν μια μέθοδο για να αξιοποιήσουν το σημασιολογικό επίπεδο σε γραφήματα προεξοχής γνώσης, όπως ο Σημασιολογικός Ιστός ή τα Wikidata. Αυτό το κατάφεραν, συνδυάζοντας αυτοματοποιημένο συλλογισμό πάνω από οντολογίες και εκμάθηση χαρακτηριστικών με ουδέτερα δίκτυα, για τη δημιουργία διανυσματικών αναπαραστάσεων κόμβων σε αυτά γραφήματα (ενσωματώσεις κόμβων). Εξηγείται, επίσης, πως ένα γράφημα γνώσης είναι μια αναπαράσταση οντοτήτων, που βασίζεται σε γράφημα που αφορά τον κόσμο και τις αλληλεπιδράσεις του και πως τα γραφήματα γνώσης χρησιμοποιούνται ευρέως για τη διευκόλυνση και τη βελτίωση της αναζήτησης. Παρουσιάζεται και το απαγωγικά κλειστό γράφημα, το οποίο έχει το πλεονέκτημα, όχι μόνο να περιλαμβάνει συναγόμενη γνώση, που δεν υπάρχει ρητά στο γράφημα αξιωμάτων, αλλά να ληφθούν υπόψη και οι αναπαραστάσεις τους. Δίνεται το παράδειγμα, για έναν ισχυρισμό ότι ένα γονίδιο g έχει μια συνάρτηση F (όπου F είναι μια κλάση στο GO), όλες οι υπερκλάσεις του F στο GO θα προστεθούν ως σχολιασμοί στο g . Μέσα από όλα αυτά η έρευνα καταλήγει στο συμπέρασμα πως οι συμβολικές αναπαραστάσεις στη βιολογία, βασισμένες σε Συνδεδεμένα Δεδομένα και οντολογίες, βασίζονται σε επίσημες γλώσσες όπως η Ontology Web Language (OWL) και η Resource Description Framework (RDF), και χρησιμοποιούνται συμβολικά στη συναγωγή. Το είδος των συμπερασμάτων που γίνονται σε αυτή τη γνώση καθορίζεται, είτε επίσημα στη γλώσσα αναπαράστασης της γνώσης, είτε παράγονται από μη-αυτοματοποιημένους κανόνες συμπερασμάτων που ισχύουν σε μια συγκεκριμένη βάση δεδομένων, εφαρμογή ή ερώτημα. (Alshahrani, et al., 2017)

2.7 Έρευνα: Ανάλυση γράφων γνώσης που αφορούν οικονομικά, τράπεζες και χρηματοοικονομικά.

Τα έσοδα μέσω παράνομων δραστηριοτήτων, ασφαλιστικής απάτης και άλλων σχετικών ύποπτων δραστηριοτήτων, αποτέλεσμα μοχλό εκκίνησης εύρεσης λύσης αυτών των φαινομένων και ως επακόλουθο την διεξαγωγή αυτής της έρευνας. Συγκεκριμένα, μελετήθηκε το πως θα μπορούσαν τα γραφήματα γνώσης, να βοηθήσουν στον εντοπισμό της ενέργειας νομιμοποίησης εσόδων από παράνομες δραστηριότητες, ασφαλιστικές απάτες και γενικότερα τέτοιου είδους παράνομες. Σημαντικό ρόλο είχε το γεγονός ότι τα γραφήματα γνώσης είναι δυναμικά όσον αφορά την χρήση τους, καθώς μπορούν να αντιπροσωπεύουν εταιρείες δομημένες σε σύνθετες μετοχές, καθώς και πληροφορίες σχετικά με επενδύσεις, εξαγορές και χρεοκοπίες. Χρησιμοποιήθηκαν γραφήματα γνώσης πραγματικών οικονομικών δεδομένων όπου κόμβοι είναι πελάτης, έμπορος, κτίριο και οι άκρες μπορεί να είναι συναλλαγές μεταξύ πελατών και πληροφορίες κατοικίας για τους πελάτες. Μέσα από αυτό υπήρξε ένας προβληματισμός, διότι πολλές από τις πηγές γνώσης αναπτύχθηκαν ανεξάρτητα η μία από την άλλη. Η μόνη λύση αυτού του προβλήματος για να προχωρήσει η έρευνα ήταν η σύντηξη αυτών των πηγών σε εξαρτημένα γραφήματα γνώσης. Αυτή η κίνηση έγινε, καθώς θα είχε ως αποτέλεσμα μια σημαντικά πλουσιότερη πηγή γνώσεων, της οποίας θα μπορούσαν να βελτιώσουν την απόδοση σε υπάρχουσες εφαρμογές. Σε αυτή την εργασία, μελετήθηκαν οι ιδιαίτερες σχέσεις της ολοκλήρωσης των γραφημάτων γνώσης σε συνδυασμό με ανάλυση των στατιστικών και γραφικών, θεωρητικών ιδιοτήτων. Αναλυτικότερα, μελετήθηκαν ιδιότητες ολοκληρωμένης γνώσης γραφημάτων, συνδυάζοντας υπάρχοντα γραφήματα γνώσης σε τομείς της οικονομίας, των τραπεζών και των χρηματοοικονομικών. Επιπρόσθετα, παρουσιάστηκαν επίσημα μοντέλα, που αποσκοπούν στον καθορισμό σαφούς κοινής σημασίας για τον χρηματοπιστωτικό κλάδο. Ακολούθησε μία σύγκριση μεταξύ οντολογιών, της οντολογίας Financial Regulation (FRO), της Οντολογίας Τραπεζικού Κανονισμού (BRO) και της Insurance Οντολογία Κανονισμού¹(IRO), το Fund Ontology. Η οντολογία Financial Regulation, έχει χρησιμοποιηθεί ως ανώτερη σε επίπεδο, βασική οντολογία για οντολογίες. (Wang, 2022) Με την σύγκριση αυτή, έγινε αντιληπτό πως η Οντολογία Τραπεζικού Κανονισμού (BRO) είναι πολύ μεγαλύτερη και χρησιμοποιεί δύο βιομηχανικά πρότυπα, συγκεκριμένα το FIBO και το Legal Knowledge Interchange Format(LKIF). (Wang, 2022) Επίσης, πάνω της δημιουργήθηκε η οντολογία Financial Regulation Ontology(FRO). Με βάση αυτά, το συμπέρασμα της έρευνας ήταν πως πολλά γραφήματα γνώσης αναπτύσσονται από τις τράπεζες και δεν είναι ανοιχτού κώδικα, κάτι το οποίο θα δημιουργεί

προβλήματα, και πως όλα τα γραφήματα γνώσης μετατράπηκαν πρώτα σε μορφή Turtle και στη συνέχεια χρησιμοποιήθηκαν το RDFpro για τη διαδικασία ολοκλήρωσης. Το RDFpro είναι μια εργαλειοθήκη ανοιχτού κώδικα, προσανατολισμένη στη ροή για την επεξεργασία των τριπλών RDF. Για την έρευνα χρησιμοποιήθηκε το RDFpro. Στη συνέχεια, όλα τα αρχεία μετατράπηκαν στη μορφή HDT τους για περαιτέρω πειράματα. Το γράφημα της ολοκληρωμένης γνώσης αποτελείται από 1.778.755 μοναδικές τριάδες (άκρες) και 610.866 κόμβους. Για λόγους ταχύτητας, κατά τη μελέτη των ιδιοτήτων των γραφημάτων γνώσης, απαραίτητη ήταν η χρήση αρχείων σε μορφή HDT. Αναφέρεται, επίσης, πως ένα ισχυρά συνδεδεμένο στοιχείο, το SCC, ενός κατευθυνόμενου γραφήματος είναι ένα μέγιστο υπογράφημα. Σε αυτό εξηγείται πως υπάρχει μια διαδρομή μεταξύ όλων των ζευγών κορυφών. Ένα ασθενώς συνδεδεμένο στοιχείο, το WCC, είναι ένα υπογράφημα του αρχικού γραφήματος όπου όλες οι κορυφές συνδέονται μεταξύ τους με κάποιο μονοπάτι, αγνοώντας την κατεύθυνση των άκρων. Συμπερασματικά, οι σύνδεσμοι ταυτότητας είναι σχέσεις μεταξύ οντοτήτων, που θεωρείται πως είναι πανομοιότυπο και προορίζονται να αναφέρονται στις ίδιες οντότητες του πραγματικού κόσμου. Τέλος, η έρευνα αυτή βοήθησε στην κατανόηση:

1. πως το ολοκληρωμένο γράφημα γνώσης μπορεί να χρησιμοποιηθεί για την αξιολόγηση της διαλειτουργικότητας των δεδομένων,
2. πως ο εμπλουτισμός των χαρακτηριστικών οντοτήτων μπορεί να αυξήσει την ακρίβεια του σχεδίου αναγνώριση με χρήση μηχανικής μάθησης για την ανίχνευση εξαγορών, ξεπλύματος μαύρου χρήματος, ασφαλιστικής απάτης, πλαστογράφησης και
3. πως τα γραφήματα γνώσης μπορούν να χρησιμοποιηθούν για τη βελτίωση της ποιότητας των αναφορών ύποπτης δραστηριότητας. (Wang, 2022)

2.8 Έρευνα: Μελέτη γράφων γνώσης που αφορούν τα χρηματοοικονομικά με χρήση υβριδικής λογικής.

Όπως η έρευνα που είδαμε στην υπό ενότητα 2.5, έτσι και η έρευνα για την οποία θα αναφερθούμε σε αυτήν την υπό ενότητα, διεξάχθηκε για τραπεζική εποπτεία, αξιολόγηση πιστοληπτικής ικανότητας, καταπολέμηση της νομιμοποίησης εσόδων από παράνομες δραστηριότητες, ανίχνευση ασφαλιστικής απάτης και οικονομική και στατιστική έρευνα. Αρχικά, γίνεται μία γνωριμία με την Warded Datalog+/-,

η οποία είναι μια γλώσσα Kubernetes Resource Recommender(KRR) που βασίζεται σε Datalog και εγγυάται την αποφασιστικότητα και την ικανότητα επεξεργασίας του οντολογικού συλλογισμού. Αυτό γίνεται χάρις στις ευνοϊκές θεωρητικές ιδιότητες που κατέχει. Επίσης, παρουσιάζεται το σύστημα συλλογισμού Vadalog, που εκμεταλλεύεται το Warded Datalog+/- για να παρέχει πρακτική εφαρμογή διαφορετικών συλλογιστικών εργασιών, μέσω βασικών ελέγχων ισομορφισμού. (Vlad, Vahdati, Nayyeri, Bellomarini, & Sallinger, 2022) Ένας προβληματισμός που δημιουργήθηκε σε αυτό το σημείο της έρευνας είναι, πως αυτά που αναφέρθηκαν παραπάνω, μπορεί να είναι απαγορευτικά σε χωρική και χρονική πολυπλοκότητα στον οικονομικό και χρηματοοικονομικό τομέα. Και οι δύο αυτοί τομείς χαρακτηρίζονται από δεδομένα ακραίας κλίμακας καταστήματα και πολύπλοκη δυναμική των κοινωνικών δικτύων, που υπάρχουν σε αυτούς. Έτσι, οι χρηματοοικονομικές γνώσεις καθίστανται απαραίτητες για την αντιμετώπιση προβλημάτων. Στην έρευνα παρατίθεται πως τα παραπάνω μπορούν, να μοντελοποιηθούν και να αποτυπωθούν, χρησιμοποιώντας γραφήματα γνώσης. Αυτά τα γραφήματα είναι κεντρικά αντικείμενα στα εταιρικά οικονομικά και χρησιμοποιούνται από κεντρικές τράπεζες, χρηματοπιστωτικές αρχές και εθνικές στατιστικές υπηρεσίες για τραπεζική εποπτεία, αξιολόγηση πιστοληπτικής ικανότητας, καταπολέμηση της νομιμοποίησης εσόδων από παράνομες δραστηριότητες, ανίχνευση ασφαλιστικής απάτης και οικονομική και στατιστική έρευνα. (Vlad, Vahdati, Nayyeri, Bellomarini, & Sallinger, 2022) Καλύτερη κατανόηση αυτών επιτυγχάνεται μέσω του παραδείγματος που δίνεται στην έρευνα. Το παράδειγμα είναι το εξής: στο Warded Datalog± υπάρχει ο περιορισμός στον έλεγχο ισομορφισμού, δηλαδή όταν δύο γεγονότα έχουν το ίδιο κατηγορήμα, τις ίδιες σταθερές στις ίδιες θέσεις και υπάρχει διχογνωμία μεταξύ των επισημασμένων μηδενικών, δηλαδή είναι ισόμορφο, ακριβώς ένα από αυτά δημιουργείται στη διαδικασία, χωρίς απώλεια πληροφοριών για την απάντηση ερωτήματος. (Vlad, Vahdati, Nayyeri, Bellomarini, & Sallinger, 2022) Όμως, λέγεται πως στον τυπικό ισομορφισμό ο έλεγχος θα μπορούσε να είναι αναποτελεσματικός, λόγω της μη πρακτικής απομνημόνευσης όλων των παραγόμενων γεγονότων, που σε πολλές περιπτώσεις απαγορευτικό στο χώρο και στο χρόνο. Έπειτα, στην έρευνα προσμετρήθηκαν οι προσπάθειες για να περιοριστεί ο χώρος αναζήτησης και αυτό να έχει ως αποτέλεσμα την βελτίωση της μνήμης. Και σε αυτήν την περίπτωση λύση έδωσε το Warded Datalog±, που μπορεί να περιορίσει την αναζήτηση ισομορφικών αντιγράφων σε συγκεκριμένα τμήματα του γραφήματος παραγωγής καταδίωξης των παράνομων ενεργειών. (Vlad, Vahdati, Nayyeri, Bellomarini, & Sallinger, 2022) Συμπεραίνουμε ότι η γλώσσα Warded Datalog+/-, μπορεί να είναι η λύση

προβλημάτων, όπως δεδομένα ακραίας κλίμακας καταστάματα και πολύπλοκη δυναμική, που υπάρχουν συχνά στον οικονομικό τομέα.

2.9 Έρευνα: Ειδικοί γράφοι γνώσης για την υγεία.

Με βάση πρόσφατες μελέτες έχει γίνει γνωστό πως υπάρχει η ανάγκη για μια ποικιλία αποτελεσματικών λύσεων για τη διατήρηση ενός υγιεινού τρόπου ζωής, όπου είναι και το κύριο θέμα της έρευνας αυτής. Ένα παράδειγμα, το οποίο αναφέρεται στην έρευνα είναι το Harifork4, που είναι ένα ηλεκτρονικό πιρούνι που βοηθά και καθιστά ακριβής την παρακολούθηση της διατροφής. Κύριο μέλημα της έρευνας ήταν ένας τρόπος κατά τον οποίο όταν θα υπάρχει ένας ασθενής με συγκεκριμένα προβλήματα και συμπτώματα, δηλαδή με συγκεκριμένα δεδομένα και πως εισχωρώντας αυτά τα δεδομένα στην μηχανή αναζήτησης Google, αυτή θα έχει την δυνατότητα να δίνει σαν αποτέλεσμα, σαν έξοδο, τον ειδικό γιατρό που θα έχει εξειδικευμένες γνώσεις πάνω στην αρρώστια του ασθενούς. Αρχικά, γίνεται μία αναφορά σε διάφορες εφαρμογές όπως το "Personalized Preventive Coach" και το "Digital Health Advisor" που χρησιμοποιούνται για την παρακολούθηση και την ερμηνεία των δεδομένων υγείας και ευεξίας που λαμβάνουν, ανάλογα με τον χρήστη. Ακόμα, γνωστοποιείται πως το περιεχόμενο αυτών των εφαρμογών είναι Ασύρματα Δίκτυα Περιοχής Σώματος (WBAN), ιατρικό Διαδίκτυο of Things (mIoT), m-health, e-health και Ambient Assisted Living5 (AAL). (Gyrard, Gaur, Shekarpour, Thirunarayan, & Sheth, 2018) Είναι, επίσης, σημαντικό να αναφερθεί για την καλύτερη κατανόηση του τρόπου λειτουργίας αυτών των εφαρμογών, πως υπολογίζουν και τους εσωτερικούς και τους εξωτερικούς παράγοντες. Κάποιοι από τους εσωτερικούς παράγοντες που χρησιμοποιούν είναι οι φθινοί αισθητήρες με σκοπό τη συλλογή ακατέργαστων δεδομένων, που παράγονται από το IoT συσκευές, όπως συσκευή ζυγαριάς, ένα μοντέλο τομέα, όπως οντολογία άσθματος ή παχυσαρκίας, για τη δομή και την αφαίρεση των δεδομένων και έναν συλλογιστικό μηχανισμό για την εξαγωγή γνώσεων και συστάσεις, όπως η χρήση Δείκτης Μάζας Σώματος (ΔΜΣ) και περιβαλλοντικών δεδομένων με μηχανή συμπερασμάτων βάσει κανόνων για την παροχή σχετικών πληροφοριών. Κάποιοι από τους εξωτερικούς παράγοντες είναι το Context-Awareness, που αναφέρεται στη χρήση εξωτερικών δεδομένων που μπορούν να επηρεάσουν την κατάσταση του χρήστη. Για παράδειγμα, οι συσκευές IoT μπορούν να χρησιμοποιηθούν για την παρακολούθηση του περιβάλλοντος στο οποίο είναι ο χρήστης. Επίσης, στην έρευνα αναφέρεται ο τρόπος ερμηνείας των δεδομένων IoT, ο οποίος είναι: χρησιμοποιώντας

ένα μοντέλο υποβάθρου για αφαίρεση μπορεί να παρέχει συνείδηση στους γιατρούς και τα κλινικά πρωτόκολλα θα πρέπει να τα λαμβάνουν υπόψη για τον προσδιορισμό της κατάστασης του ασθενούς. Σε αυτό το σημείο λαμβάνουν χώρο τα γραφήματα γνώσης, που σε αυτήν την έρευνα είναι τα Google Healthcare, τα οποία είναι μη αυτόματα επιμελημένα γράφημα γνώσεων υγείας. Αυτό σημαίνει πως ενσωματώνουν το ICD-9 και το UMLS9 μαζί με την πιθανολογική μηχανική μάθηση και τον ιατρό, ο οποίος μπορεί να προσφέρει υποστήριξη για την παροχή σχετικών πληροφοριών κατά την αναζήτηση χρήστη. Παρατίθεται το παράδειγμα για το σύνολο δεδομένων παχυσαρκίας, όπου αποτελείται από δεδομένα που παράγονται από το IoT συσκευές όπως ζυγαριά, χάπι και μπουκάλι νερό και Fitbit για τη λήψη παραμέτρων όπως βάρος, κατανάλωση φαρμάκων, καρδιακός ρυθμός και δραστηριότητα ύπνου. Επιπλέον, το σύνολο δεδομένων Parkinson¹⁶ από το Kaggle αποτελείται από δεδομένα αισθητήρων κινητής τηλεφωνίας από επιταχυνσιόμετρο, πυξίδα και μικρόφωνο στο smartphone για να συνθέσει ο ασθενής τις πληροφορίες για τα συμπτώματά του, όπως ασταθές περπάτημα, ελλείψεις ισορροπεί, έχει πτώση και έχει μπερδεμένη ομιλία. Ο σχεδιασμός του Personalized Health Knowledge Graph (PHKG) είναι, επίσης, κρίσιμος για την επίτευξη των Digitalized Personalized Health coaches, όπου σε αυτήν την περίπτωση θα μπορούσαν να χρησιμοποιηθούν chatbots για να βοηθήσουν τους γιατρούς και τους ασθενείς. Σημειώνεται το ότι ο σχεδιασμός του Personalized Health Knowledge Graph είναι δύσκολος, γιατί απαιτεί σημασιολογικό ενσωμάτωση ετερογενών δεδομένων από παρόχους υγειονομικής περίθαλψης, συσκευές IoT και τον Ιστό, λαμβάνοντας υπόψη το πλαίσιο και το προσωπικό ιστορικό για τη συλλογιστική και την εξαγωγή συμπερασμάτων, δηλαδή υψηλού επιπέδου αφαιρέσεις και αποτελεσματικές πρακτικές ιδέες. (Gyrard, Gaur, Shekarpour, Thirunarayan, & Sheth, 2018) Η έρευνα υποστηρίζει πως το Personalized Health Knowledge Graph μπορεί να χρησιμεύσει ως βάση βοήθειας των γιατρών, ώστε να κατανοήσουν τα συμπτώματα, να υποθέσουν και να εξηγήσουν τόσο την ασθένεια όσο και την εξέλιξη της, και τέλος να είναι σε θέση να διαμορφώσουν το ανάλογο σχέδιο διαχείρισης και θεραπείας της ασθένειας. Επομένως, μέσα από αυτήν την έρευνα συμπεραίνουμε ότι το Personalized Health Knowledge Graph, είναι η αποτελεσματική λύση για τη διατήρηση της υγείας και την γρηγορότερη, πιο στοχευόμενη διάγνωση και αντιμετώπιση μίας ασθένειας.

2.10 Έρευνα: Εξατομικευμένο Γράφημα Γνώσεων Υγείας.

Κύριο θέμα της συγκεκριμένης έρευνας αποτελούν οι εφαρμογές που υπάρχουν για την υγεία και οι προκλήσεις διαχείρισης αυτών. Υποστηρίζεται πως οι εφαρμογές αυτές έχουν ελλειπείς γνώσεις που αφορούν την εξατομικευμένη γνώση για τους ασθενείς. Το γεγονός αυτό δημιουργεί προβλήματα διαχείρισης και αντιμετώπισης των ασθενειών. Επιπρόσθετα, γίνεται αναφορά στην δημιουργία "Personalized Coach for Healthcare" εφαρμογής, που αποτελεί λύση σε αυτά τα προβλήματα. Έτσι, θα είναι δυνατή η διαχείριση χρόνιων ασθενειών και ο εμπλουτισμός των φακέλων των ασθενών με την κατάσταση της υγείας τους και της θεραπείας τους. Ωστόσο, για την ανάπτυξη της εφαρμογής ΡΗΚΓ, είναι απαραίτητη η συγκέντρωση γνώσεων από διάφορες ετερογενείς πηγές, όπως συσκευές Internet of Things (IoT), κλινικές σημειώσεις και Ηλεκτρονικά Ιατρικά Αρχεία (EMR). Λύση φέρει ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ), υπολογίζοντας τον αριθμό των ανθρώπων που πάσχουν από άσθμα, παχυσαρκία και νόσο του Πάρκινσον σε 235 εκατομμύρια, 650 εκατομμύρια και 4–6 εκατομμύρια ανθρώπους αντίστοιχα (*Gyrard, Gaur, Shekarpour, Thirunarayan, & Sheth, 2018*). Αυτή η διαδικασία γίνεται, καθώς οι χρόνιες ασθένειες (όπως η παχυσαρκία και το άσθμα) είναι πολυπαραγοντικές ασθένειες, που με το πέρασμα των χρόνων έχουν γίνει στις μέρες μας επιδημίες. Έχει αποδειχθεί μέσα από μελέτες πως υπάρχει ανάγκη για μια ποικιλία αποτελεσματικών λύσεων για τη διατήρηση ενός υγιεινού τρόπου ζωής. Δίνεται το παράδειγμα του Harifork4, που είναι ένα ηλεκτρονικό πιρούνι και βοηθά στην παρακολούθηση της διατροφής. Το Harifork4 έχει διπλό ρόλο, καθώς βοηθάει στην διατροφή και αποτελεί τη βάση για αναδυόμενες εφαρμογές όπως το "Personalized Preventive Coach" και το "Digital Health". Επιπρόσθετα, υπάρχουν τομείς όπως τα ασύρματα δίκτυα περιοχών σώματος, το ιατρικό Διαδίκτυο των πραγμάτων (mIoT), το m-health, η ηλεκτρονική υγεία και η υποβοηθούμενη διαβίωση περιβάλλοντος (AAL). Οι τομείς αυτοί είναι σημαντικοί, αφού χρησιμοποιούν φθηνούς αισθητήρες για τη συλλογή ακατέργαστων δεδομένων που δημιουργούνται από τις συσκευές IoT, ένα μοντέλο τομέα για τη δομή και την αφαίρεση των δεδομένων και ένας συλλογιστικός μηχανισμός για την εξαγωγή γνώσεων και συστάσεων (*Gyrard, Gaur, Shekarpour, Thirunarayan, & Sheth, 2018*). Σημαντικό είναι να τονιστεί η συσχέτιση του Context-Awareness με τα συστήματα υγείας. Το Context-Awareness αναφέρεται στη χρήση εξωτερικών δεδομένων, που μπορούν να επηρεάσουν την κλινική κατάσταση του χρήστη και να μεταβάλλουν την υγεία του. Αυτό συμβαίνει, διότι οι συσκευές IoT μπορούν να χρησιμοποιηθούν για την παρακολούθηση του περιβάλλοντος. Κάθε ασθενής μπορεί να αντιδράσει διαφορετικά όταν

εκτίθεται σε διαφορετικούς περιβαλλοντικούς παράγοντες, επομένως απαραίτητη είναι η εξατομικευμένη προσέγγιση στη θεραπεία κάθε ασθενούς. Τα κλινικά έγγραφα κάθε ασθενούς είναι δημογραφικές πληροφορίες, παρατηρήσεις κλινικού ιατρού, εργαστηριακές εξετάσεις και δεδομένα από κλινικές επισκέψεις. Συμπερασματικά, η εξατομίκευση που προσφέρουν εφαρμογές, σε συνδυασμό με προγνωστικά αναλυτικά στοιχεία, καθιστά δυνατή τη λήψη πρακτικών πληροφοριών για τον κάθε ασθενή ξεχωριστά, βοηθώντας με αυτόν τον τρόπο στην γρήγορη και δραστική καταπολέμηση ασθενειών. (Gyrard, Gaur, Shekarpour, Thirunarayan, & Sheth, 2018)

2.11 Έρευνα: Γραφήματα γνώσης και εξηγήσιμη τεχνητή νοημοσύνη στην Υγεία.

Η Τεχνητή Νοημοσύνη (ΧΑΙ) χαρακτηρίζεται από εμπιστευτικότητα και διαφάνεια, χαρακτηριστικά που απαιτούνται στην υγειονομική περίθαλψη και μπορούν να βοηθήσουν στη διαδικασία λήψης αποφάσεων. Αρχικά, χάρις την Τεχνητή Νοημοσύνη είναι δυνατή η χρήση γραφημάτων γνώσης για την επεξήγηση των δεδομένων και την εξαγωγή χαρακτηριστικών και σχέσεων. Στη συγκεκριμένη έρευνα, παρατίθεται ο ρόλος των γραφημάτων γνώσης σε μοντέλα ΧΑΙ στην υγειονομική περίθαλψη, λαμβάνοντας υπόψη μία ανασκόπηση τελευταίας τεχνολογίας. Η ανασκόπηση αυτή υποστηρίζει, ότι τα γραφήματα γνώσης έχουν χρησιμοποιηθεί για την επεξήγηση και την ανίχνευση παραπληροφόρησης στην υγειονομική περίθαλψη (Rajabi & Kafaie , 2022). Αυτό επιφέρει την αποφυγή πρόκλησης παρενεργειών από την χορήγηση ανεπιθύμητων φαρμάκων στους ασθενείς και αλληλεπιδράσεων φαρμάκων. Προκειμένου να μειωθεί το χάσμα γνώσης, που υπάρχει μεταξύ των ειδικών στον τομέα της υγείας, φαίνεται πως απαραίτητος είναι ο εμπλουτισμός των μοντέλων, που χρησιμοποιούνται, με τη τεχνητή νοημοσύνη. Στη συνέχεια, γίνεται μία πιο αναλυτική αναφορά στα συστήματα Τεχνητής Νοημοσύνης (AI), υπογραμμίζοντας πως καθιστούν την επεξεργασία περίπλοκων εργασιών γρηγορότερη και αυτόματη. Τα δεδομένα υγειονομικής περίθαλψης κατατάσσονται σε τρεις κατηγορίες: δεδομένα omics, κλινικά δεδομένα και δεδομένα αισθητήρων (Rajabi & Kafaie , 2022). Δηλαδή, τα δεδομένα Omics ασχολούνται με δεδομένα μεγάλης κλίμακας και υψηλών διαστάσεων, τα οποία υποδεικνύουν το σύνολο κάποιου είδους όπως η γονιδιωματική, η πρωτεϊνική και η μεταγραφική (Rajabi & Kafaie , 2022). Έπειτα, τα κλινικά δεδομένα εμπεριέχουν ηλεκτρονικά αρχεία υγείας ασθενών, όπου συλλέγονται κατά την διάρκεια της συνεχούς φροντίδας.

Ο ρόλος αυτών είναι ως η κύρια πηγή για τις περισσότερες ιατρικές και υγειονομικές έρευνες. Τέλος, τα δεδομένα αισθητήρων αντιπροσωπεύουν δεδομένα, τα οποία συλλέγονται από πολυάριθμες φορητές και ασύρματες συσκευές αισθητήρων (όπως παλμική οξυμετρία, συσκευές παρακολούθησης καρδιακών παλμών). Αυτό έρχεται σε αντίθεση με τον παλαιότερο τρόπο επεξεργασίας, που γινόταν χειροκίνητα από ειδικούς ιατρούς. Ο σχεδιασμός ενός συστήματος AI (ΧΑΙ) παρέχει επισκόπηση του ίδιου του συστήματος AI και εκπαιδεύει τους χρήστες του. Αυτό κάνει τα συστήματα AI πιο κατανοητά και εύκολα διαχειρίσιμα από τους ανθρώπους που εργάζονται στον τομέα της υγείας. Με αυτόν τον τρόπο, οι αποφάσεις που καλούνται να πάρουν οι ιατροί, σχετικά με την θεραπεία μίας ασθένειας, είναι πιο ακριβείς, καθώς τα συμπεράσματα που διεξάγονται είναι εύκολα αναγνωρίσιμα. Τα παραδοσιακά μοντέλα εκσυγχρονίζονται με τα χαρακτηριστικά της εμπιστοσύνης και της εχεμύθειας σε κλινικά περιβάλλοντα. Η διαδικασία που ακολουθείτε από την αρχή της εξέτασης ενός ασθενούς έως την εξαγωγή της θεραπείας, στην οποία θα υποβληθεί, είναι:

1. εξαγωγή οντοτήτων/σχέσεων: οι αλληλεπιδράσεις των φαρμάκων που υπάρχουν σε κλινικές σημειώσεις με τη μορφή ελεύθερου κειμένου στην υγειονομική περίθαλψη, μετατρέπονται σε πληροφορίες με μια δομημένη μορφή, όπως μια ονομαστική οντότητα. Τα γραφήματα γνώσης αναπαριστούν τις οντότητες αυτές με μεθόδους αναγνώρισης ονομασμένων οντοτήτων και τις αντιστοιχίζουν σε λεξιλόγια χρησιμοποιώντας τεχνικές κανονικοποίησης ονομασμένων οντοτήτων. Διεξάγονται σχέσεις μεταξύ των οντοτήτων, όπου η σημασιολογική σχέση εξάγεται συνήθως μεταξύ δύο οντοτήτων.
2. εμπλουτισμός: στη συνέχεια, τα γραφήματα γνώσης χρησιμοποιούνται για τον εμπλουτισμό των συνόλων δεδομένων με πληροφορίες. Σημαντικό είναι να αναφερθεί πως όσο περισσότερη είναι η γνώση που υπάρχει στο AI, τόσο πιο αξιόπιστα θα είναι τα συμπεράσματα του μοντέλου μηχανικής εκμάθησης.
3. συμπεράσματα και συλλογισμός: τα γραφήματα γνώσης μέσα από τη συλλογιστική εξαγωγή, καταλήγουν σε καινούρια συμπεράσματα και επομένως σε νέες γνώσεις. Αναλυτικότερα, η συλλογιστική σε ένα γράφημα γνώσης αποτελεί προσέγγιση, η οποία είναι βασισμένη σε στοιχεία εύκολα ερμηνεύσιμα από τους κλινικούς γιατρούς.
4. επεξηγήσεις και απεικονίσεις: τέλος, τα μοντέλα ΧΑΙ παρέχουν στους επαγγελματίες του τομέα της υγείας εξηγήσεις, με σκοπό την διεξαγωγή σαφών αποτελεσμάτων (Rajabi & Kafaie , 2022).

Μπορούμε να καταλήξουμε στο συμπέρασμα, πως τα γραφήματα γνώσης είναι ευρέως χρησιμοποιημένα για να εξηγήσουν τις διάφορες αλληλεπιδράσεις που μπορεί να επιφέρει η ταυτόχρονη χορήγηση πολλών φαρμάκων. Το γεγονός αυτό, προσφέρει αποτελεσματική αποφυγή τυχόν λαθών και μείωση πρόκλησης περαιτέρω προβλημάτων στους ασθενείς. Επίσης, ο συνδυασμός γραφημάτων γνώσης με μοντέλα μηχανικής μάθησης, μπορεί να παρέχει περισσότερες πληροφορίες για να γίνουν κατανοητά τα μοντέλα, που βασίζονται σε τεχνητή νοημοσύνη (*Rajabi & Kafaie , 2022*).

2.12 Έρευνα: Κατασκευή γραφήματος γνώσης υγειονομικής περίθαλψης:

Μια συστηματική ανασκόπηση της τελευταίας τεχνολογίας, των ανοιχτών θεμάτων και των ευκαιριών.

Λόγος σύνταξης αυτής της έρευνας αποτέλεσε η ανάδειξη της σημαντικότητας ταξινόμησης των γραφημάτων γνώσης στον τομέα της υγειονομικής περίθαλψης. Στις μέρες μας, κάνει την εμφάνισή του μεγάλος όγκος δεδομένων υγειονομικής περίθαλψης και δημιουργείται η ανάγκη εύρεσης λύσεων σε προβλήματα ανάλυσης δεδομένων με αποτελεσματικό τρόπο. Γνωστό είναι το γεγονός, ότι τα γραφήματα γνώσης είναι ιδιαίτερα χρήσιμα σε αυτόν τον τομέα. Επίσης, κατέχουν μία σειρά εφαρμογών υγειονομικής περίθαλψης, που προσφέρει καλύτερη αναπαράσταση δεδομένων. Ωστόσο, η ύπαρξη έλλειψης αντιπροσωπευτικής ταξινόμησης κατασκευής των γραφημάτων γνώσης, επιφέρει ανεπαρκείς αποτελέσματα. Στην συγκεκριμένη έρευνα παρουσιάζονται για πρώτη φορά μια ολοκληρωμένη ταξινόμηση και μια πανοραμική άποψη της κατασκευής γραφημάτων γνώσης, που αντιστοιχούν στην υγειονομική περίθαλψη. Επίσης, γίνεται εξονυχιστική εξέταση των σημερινών τεχνικών αιχμής που προέρχονται από ακαδημαϊκές εργασίες και σχετίζονται με διάφορα πλαίσια υγειονομικής περίθαλψης. Αυτές οι τεχνικές χωρίζονται σε δύο κατηγορίες: ως προς τις μεθόδους, που χρησιμοποιούνται για την εξαγωγή γνώσης, ως προς τους τύπους της βάσης γνώσης και των πηγών ενσωματωμένων πρωτόκολλων αξιολόγησης (*Bilal , et al., 2023*). Προχωρώντας στην διατύπωση μεθοδολογίας της έρευνας, αξίζει να σημειωθεί πως μελετά τις πρόσφατες προσεγγίσεις κατασκευής γραφημάτων γνώσης, που σχετίζονται με εφαρμογές υγειονομικής περίθαλψης. Έγινε αναζήτηση γνώσης σε άρθρα που δημοσιεύτηκαν την τελευταία πενταετία, δηλαδή από το έτος 2018 μέχρι και το 2022. Στα άρθρα αυτά, χρησιμοποιήθηκε το πλαίσιο PRISMA για να καθοδηγήσει αυτήν τη συστηματική ανασκόπηση. Στο πρώτο στάδιο της έρευνας επιλέχθηκαν περίπου 560 άρθρα από

διάφορες βάσεις δεδομένων(Elsevier, ACM Digital Library, Multidisciplinary Digital Publishing Institute (MDPI), IEEE Xplore Digital Library, Google Scholar) (Bilal , et al., 2023). Στη συνέχεια, εντοπίστηκαν επιπλέον 83 άρθρα και προστέθηκαν στην ανασκόπηση. Αναλυτικότερα, πολλά από τα άρθρα συζητούσαν προσεγγίσεις για ενσωματώσεις γραφημάτων γνώσης, που εφαρμόζονται σε υπάρχοντα γραφήματα γνώσης. Έτσι, δεν έγιναν προσπάθειες ανάλυσης της ανάγκης κατασκευής νέων γραφημάτων γνώσης στην υγειονομική περίθαλψη. Μια άλλη οπτική γωνία παρουσιάζεται στα άρθρα, αναφέροντας την κατασκευή γραφημάτων γνώσης για άλλους τομείς γνώσης, ώστε να υπάρχει μία ολοκληρωμένη άποψη. Ωστόσο, στα άρθρα αυτά αναφέρεται η «υγειονομική περίθαλψη» ως παράδειγμα της δημοτικότητας των γραφημάτων γνώσης για την αντιμετώπιση βιομηχανικών εφαρμογών. Μέσω της έρευνας γίνεται αντιληπτό ότι ο τεράστιος όγκος δεδομένων υγειονομικής περίθαλψης, που συλλέγεται είναι σε ποικίλες μορφές και αφορά ένα ευρύ φάσμα θεμάτων αποτελεί μια κρίσιμη πρόκληση για τους αναλυτές. Συμπερασματικά, τα γραφήματα γνώσης δίνουν την δυνατότητα μιας αποτελεσματικής λύσης σε αυτή την πρόκληση, καθώς προσφέρουν νέες δυνατότητες στις μηχανές, που με την χρήση τους κατανοούν τα νοήματα, μειώνοντας σε μεγάλο βαθμό το σημασιολογικό χάσμα μεταξύ αυτών και των ανθρώπων. Έτσι, τα γραφήματα γνώσης στον κλάδο της υγειονομικής περίθαλψης βοηθούν στην αντιμετώπιση διαφορετικών ζητημάτων υγειονομικής περίθαλψης.

2.13 Έρευνα: Πρόταση νήματος ενισχυμένου γραφήματος γνώσης στα Φόρουμ Υγείας.

Στην συγκεκριμένη έρευνα, μελετήθηκαν οι μέθοδοι σύστασης νημάτων υγείας. Οι μέθοδοι αυτοί είναι σημαντικοί με βάση την μελέτη, καθώς προτείνουν τα περισσότερα σχετικά υπάρχοντα νήματα για έναν χρήστη και βασίζονται στη μοντελοποίηση του περιεχομένου της ανάρτησης με στόχο την ανάκτηση σχετικών απαντήσεων (Limeng & Dongwon , 2022). Απαραίτητο είναι να τονιστεί, πως ορισμένες από αυτές τις αναρτήσεις, που γράφτηκαν από χρήστες με διαφορετικές κλινικές παθήσεις, μπορεί να χαρακτηρίζονται ως λεξικά παρόμοιες ή ως μη σχετιζόμενες ασθένειες. Αυτό μπορεί να συμβεί σε παθήσεις όπως στηθάγχη και οστεοπόρωση, που μπορεί να έχουν τα ίδια συμπτώματα ή πλάτη και πόνος. Έτσι, πρέπει να εξετάζονται οι συνδέσεις μεταξύ χρηστών και νημάτων και οι περιγραφές των συμπτωμάτων και των κλινικών τους καταστάσεων. Η ανάγκη ύπαρξης ενός

μοντέλου που εμφανίζει σωστά αποτελέσματα σε αναζητήσεις χρηστών γύρω από το θέμα της υγείας τους, είναι αποτέλεσμα της ευρείας χρήσης των μέσων κοινωνικής δικτύωσης. Στις μέρες μας, η δημιουργία, η κοινή χρήση περιεχομένου και η αναζήτηση πληροφοριών για την υγεία στα μέσα κοινωνικής δικτύωσης, έχει ραγδαία ανάπτυξη, καθώς τα μέσα κοινωνικής δικτύωσης παρέχουν πρόσβαση σε τεράστια γνώση και ενθαρρύνουν τη λήψη αποφάσεων από τους χρήστες, πριν την επίσκεψη στον γιατρό. Γίνεται αναφορά στο Pew Research Center, το οποίο υποστηρίζει μετά από μία έρευνα παρακολούθησης της υγείας, πως η αναζήτηση πληροφοριών για θέματα υγείας είναι μία από τις κορυφαίες διαδικτυακές δραστηριότητες. Στην μελέτη αυτή, όσον αφορά το πρόβλημα που προαναφέρθηκε με τη σύσταση νήματος στο διαδικτυακό φόρουμ υγειονομικής περίθαλψης, προτείνεται ένα βελτιωμένο γράφημα γνώσεων Μοντέλο Σύστασης Νημάτων (KETCH). Αυτό αξιοποιεί το γράφημα γνώσης στα νευρωνικά δίκτυα για τη μοντελοποίηση των αλληλεπιδράσεων που υπάρχουν μεταξύ των χρηστών και των νημάτων. Επίσης, χρησιμοποιεί στρατηγική, η οποία αφορά τη μετάδοση μηνυμάτων με συγκέντρωση πληροφοριών στο δίκτυο και αποτελεί την πρώτη μέθοδο για τη μοντελοποίηση αλληλεπιδράσεων υψηλής τάξης μεταξύ των χρηστών και των νημάτων μέσα διαδικτυακά φόρουμ υγείας. Κατά την διάρκεια της μελέτης, γίνεται μία εισαγωγή ενός μηχανισμού, που είναι ενισχυμένος με γνώσεις για την καταγραφή λανθανόντων καταστάσεων. Για να υπάρχει μία ολοκληρωμένη οπτική γωνία γίνεται χρήση αυτής της μεθόδου στην πρόβλεψη των παρενεργειών που προκαλούν τα φάρμακα. Με αυτόν τον τρόπο, αποδεικνύεται ότι το KETCH παρέχει τη δυνατότητα να συμπληρώσει το γράφημα της ιατρικής γνώσης. Τονίζεται πως η σύγκριση με τα καλύτερα αποτελέσματα επτά ανταγωνιστικών μεθόδων, από άποψη του MRR, το KETCH υπερέχει όλων των μεθόδων κατά τουλάχιστον 0,125 στο Δεδομένα MedHelp, 0,048 στο σύνολο δεδομένων ασθενών και 0,092 στο σύνολο δεδομένων HealthBoards, αντίστοιχα (Limeng & Dongwon, 2022). Στη συνέχεια, παρατέθηκε το σετ δεδομένων που χρησιμοποιήθηκε από τους ερευνητές και αναφέρθηκαν τα τρία πειράματα που έκαναν. Τα πειράματα αυτά έγιναν για να απαντηθούν οι παρακάτω βασικές ερωτήσεις:

- EQ1: Είναι το KETCH σε θέση να βελτιώσει την απόδοση σύστασης νημάτων ενσωματώνοντας το γράφημα ιατρικής γνώσης;
- EQ2: Πόσο αποτελεσματικό είναι το γράφημα γνώσης και η γνώση της γνώσης προσοχή, αντίστοιχα, στη βελτίωση της σύστασης νήματος εκτέλεση;
- EQ3: Μπορεί το KETCH να παρέχει αξιόπιστες και υψηλής ποιότητας πληροφορίες στους χρήστες;

Μπορούμε να καταλήξουμε στο συμπέρασμα ότι, τα γραφήματα γνώσης αποτελούν λύση στο πρόβλημα της λανθασμένης μετάδοσης μηνυμάτων για λήψη αποφάσεων χρήστη σε επίπεδο νήματος και μεταγενέστερο αποτελεί το KETCH, καθώς χρησιμοποιεί πληροφορίες από ένα γραφήματα γνώσης ιατρικής προσέγγισης και προσφέρει αξιόπιστες πληροφορίες.

2.14 Έρευνα: Γράφημα Γνώσης Σημασιολογικής Υγείας: Σημασιολογική Ένταξη του Ετερογενείς Ιατρικές Γνώσεις και Υπηρεσίες.

Σε αυτήν την έρευνα, θέμα μελέτης ήταν η κειμενική ιατρική γνώση (TMK), η οποία έχει πρωταγωνιστικό ρόλο στα συστήματα υγειονομικής περίθαλψης (HIS) στην παροχή γνώσης και λήψη αποφάσεων στον τομέα της υγείας. Στις μέρες μας, η συνεχόμενη ψηφιοποίηση της ιατρικής βιβλιογραφίας και η επέκταση της βιοϊατρικής γνώσης σε συνδυασμό με τη ταχεία εξάπλωση της ιεραρχικής διαδικτυακής υγειονομικής περίθαλψης αυξάνουν την κειμενική ιατρική γνώση (TMK) (Longxiang , et al., 2017). Έτσι, η ανάγκη αντιμετώπισης ετερογενών TMK απαιτούσε λύση. Με την οργάνωση της κειμενικής ιατρικής γνώσης που θα επιφέρει η λύση, μπορούν να ενσωματωθούν όλες οι σχετικές πληροφορίες και έπειτα να παρέχουν χρήσιμες επεξεργασμένες πληροφορίες στους χρήστες. Στην μελέτη εξηγείται πως οι συμβατικές μέθοδοι χρησιμοποιούν ετερογενή γνώση με την αντιστοίχιση των λέξεων-κλειδιών. Σημαντικό ρόλο έχουν και οι βάσεις δεδομένων, που διαχειρίζονται και ενημερώνονται με μη αυτόματο τρόπο. Αυτός είναι ο λόγος που δεν είναι σε θέση να αντιμετωπίσουν τον πολλαπλασιασμό της TMK. Σε αυτό το σημείο γίνεται αναφορά στα γραφήματα γνώσης και στον τρόπο με τον οποίο μπορούν να βοηθήσουν. Με την πάροδο του χρόνου η επέκταση του Παγκόσμιου Ιστού, επιφέρει δυνατότητες ενοποίησης χάρις τις τεχνολογίες σημασιολογικού Ιστού. Επίσης, η αναζήτηση στο πολυάριθμο ετερογενές περιεχόμενο του Ιστού γίνεται πιο αποτελεσματική. Ωστόσο, με την οργάνωση του περιεχομένου του Ιστού σε εννοιολογικά γραφήματα γνώσης, χρησιμοποιώντας οντολογίες και Πλαίσιο Περιγραφής Πόρων(RDF), οι τεχνολογίες σημασιολογικού Ιστού, δίνεται η δυνατότητα στον ιστότοπο, να κατανοήσει και να επεξεργαστεί την ανθρώπινη γνώση. Μέσα από αυτό, δίνεται ένα αποτελεσματικό πλαίσιο ερωτημάτων και συλλογισμών για τα τεράστια ετερογενή περιεχόμενα του Ιστού (Longxiang , et al., 2017). Επιπλέον, παρατίθενται παραδείγματα γραφημάτων γνώσης και αποδεικνύεται η σπουδαιότητά τους, όπως:

- Γράφημα γνώσης της Google,
- Το DBPedia και
- Το YAGO.

Και οι τρεις κατηγορίες περιέχουν γραφήματα γνώσης, των οποίων τα χαρακτηριστικά τους είναι τεχνικές σημασιολογικού ιστού που τα καθιστούν ιδανική επιλογή για να ανταποκριθούν σε απαιτήσεις ετερογενών ΤΜΚ. Επιπρόσθετα, σε αυτό το άρθρο προτείνεται μία νέα προσέγγιση για την οργάνωση των ΤΜΚ σε εννοιολογικά γραφήματα γνώσης. Αναλυτικότερα, η προσέγγιση αφορά:

- Ένα μοντέλο για την ολοκλήρωση των ετερογενών κειμενικής ιατρικής γνώσης με δεδομένα υγείας, που έχει την ικανότητα να υποστηρίξει σημασιολογική αναζήτηση,
- Ένα αυτόματο πλαίσιο ανάκτησης γνώσης, ώστε να μετατραπεί το κείμενο γνώσης σε μηχανικά αναγνώσιμη μορφή, ώστε να δημιουργηθεί ένα Γράφημα Γνώσης Σημασιολογικής Υγείας και
- Έναν αλγόριθμο για το κλάδεμα του ανούσιου συμπεράσματα πάνω από το γράφημα γνώσης (Longxiang , et al., 2017).

Τέλος, γίνεται αναφορά σε πείραμα, το οποίο αποδεικνύει ότι ο αλγόριθμός βελτιώνει την απόδοση των αποτελεσμάτων. Συμπερασματικά, είναι κατανοητό πως η τεράστια ποσότητα ΤΜΚ, δίνει την ικανότητα διαμοιρασμού και αξιοποίησης πληροφοριών, με την προϋπόθεση πως θα γίνει χρήση ενός ειδικού μοντέλου στα δεδομένα υγειονομικής περίθαλψης. Επίσης, αποδίδεται η σημαντικότητα χρήσης των γραφημάτων γνώσης για την επίτευξη της οργάνωσης του ΤΜΚ, καθώς αποδίδουν δομές δεδομένων για τα δεδομένα και κατέχουν σημασιολογικές αντιστοιχίσεις μεταξύ των ΤΜΚ και της ιατρικής γνώσης.

2.15 Έρευνα: Εκμάθηση Γραφήματος Γνώσης Υγείας από Ηλεκτρονικά Ιατρικά Μητρώα.

Στο άρθρο αυτό ερευνάται μία αυτοματοποιημένη διαδικασία εκμάθησης βάσεων γνώσης υψηλής ποιότητας, οι οποίες συσχετίζονται με ασθένειες και συμπτώματα από ηλεκτρονικά ιατρικά αρχεία. Τα συστήματα υποστήριξης κλινικών αποφάσεων εδραιώνονται όλο και περισσότερο στον τομέα της

ιατρικής, καθώς η ανάγκη ταχείας διάγνωσης και εύρεσης θεραπείας είναι πλέον αναγκαία. Επίσης, οι υπάρχουσες πλατφόρμες είναι σχεδιασμένες έτσι ώστε να λειτουργούν σε βάσεις γνώσης με μη αυτόματο τρόπο. Ωστόσο, χρησιμοποιούνται αυτοματοποιημένα εργαλεία για την υποστήριξη ιατρικών διαγνωστικών συλλογισμών από ασθενείς. Έχει παρατηρηθεί η εντατική αναζήτηση πληροφοριών σχετικά με μία ασθένεια από χρήστες του διαδικτύου και κλινικούς ιατρούς με λέξεις κλειδιά τα συμπτώματα. Αυτό είναι συχνό φαινόμενο στην αντιμετώπιση δύσκολων περιπτώσεων και στην αποφυγή άσκοπης εστίασης σε πιθανές διαγνώσεις, που μπορεί να μην είναι ακριβείς. Η έρευνα κάνει μία ιστορική αναφορά, λέγοντας πως τα μοντέλα που χρησιμοποιήθηκαν από τα συστήματα διαγνωστικής συλλογιστικής καθορίζονταν χειροκίνητα, γεγονός που απαιτεί τεράστιο χρόνο και προσπάθεια από ειδικούς (Rotmensch, Halpern, Tlimal, Horng, & Sontag, 2017). Με την εξέλιξη της τεχνολογίας έκαναν την εμφάνισή τους τα δεδομένα EMR. Τονίζεται, το γεγονός ότι τα δεδομένα EMR(Electronic Medical Records) είναι δύσκολο να ερμηνευτούν για τέσσερις βασικούς λόγους:

1. Πρώτον, οι σημειώσεις του γιατρού και της νοσηλευτικής έχουν λιγότερο επίσημο ύφος από αυτό των παραδοσιακών σχολικών βιβλίων και αυτό καθιστά δύσκολο τον σταθερό εντοπισμό αναφορών ασθενειών και συμπτωμάτων,
2. Δεύτερον, τα σχολικά βιβλία και τα περιοδικά παρουσιάζουν συχνά απλοποιημένες περιπτώσεις που αναμεταδίδουν μόνο τα πιο τυπικά συμπτώματα. Αυτό προκαλεί προβλήματα και παραπλανήσεις σχετικά με την ασθένεια που έχει κάποιος, ενώ τα δεδομένα EMR παρουσιάζουν τους πραγματικούς ασθενείς με όλες τις συν νοσηρότητες, τους συσχετικούς παράγοντες και τις αποχρώσεις που τους καθιστούν άτομα και έτσι υπάρχει μία πιο ολοκληρωμένη εικόνα του ασθενούς,
3. Τρίτον, είναι γνωστό ότι τα σχολικά βιβλία αναφέρουν τις σχέσεις μεταξύ ασθενειών και συμπτωμάτων με δηλωτικό τρόπο, όμως, οι συσχετίσεις μεταξύ ασθενειών και συμπτωμάτων στο EMR είναι στατιστικές. Αυτό γίνεται ώστε να είναι εύκολη η σύγχυση της συσχέτισης με την αιτιότητα και
4. Τέταρτον, ο τρόπος με τον οποίο καταγράφονται οι παρατηρήσεις στο EMR φιλτράρεται μέσω της διαδικασίας λήψης αποφάσεων του ιατρού και με αυτόν τον τρόπο αποφεύγονται λάθη λόγω ελλειπών ή μη σχετικών πληροφοριών.

Όπως φαίνεται τα δεδομένα EMR είναι πιο δύσκολο να εργαστούν για τους παραπάνω λόγους, ωστόσο έχουν πλεονέκτημα. Τα δεδομένα αυτά είναι πιο κοντά στην πραγματική πρακτική της ιατρικής από ότι είναι οι εξιδανικευμένες πληροφορίες που παρουσιάζονται σε σχολικά βιβλία και τα πε-

ριοδικά. Για παράδειγμα, η εκμάθηση από τα EMR παρέχει την ευκαιρία να ανακαλύψετε νέες σχέσεις μεταξύ ασθενειών και συμπτωμάτων, που προηγουμένως δεν είχαν αναγνωριστεί. Επιπλέον, η εκμάθηση ενός γραφήματος υποψήφιων αιτιατών σχέσεων(ασθενειών-συμπτωμάτων) από EMR μας φέρνει ένα βήμα πιο κοντά στα μοντέλα μάθησης. Αυτό είναι πολύ σημαντικό, καθώς τα μοντέλα μάθησης πραγματοποιούν διαγνωστικά συμπεράσματα απευθείας από τα πραγματικά δεδομένα που παράγονται συνεχώς από το σύστημα υγειονομικής περίθαλψης (Rotmensch, Halpern, Tlimal, Horng, & Sontag, 2017). Στη συνέχεια, δίνεται το παράδειγμα ότι από την μία δαπανήθηκαν περίπου δεκαπέντε έτη για τη δημιουργία της βάσης γνώσεων Internist-1/QMR για την ιατρική (Rotmensch, Halpern, Tlimal, Horng, & Sontag, 2017), από την άλλη οι χειροκίνητες προδιαγραφές έκαναν αυτά τα μοντέλα εξαιρετικά εύθραυστα και δύσκολο να προσαρμοστούν σε νέες ασθένειες και κλινικές ρυθμίσεις, που μπορεί να πρόκυπταν. Επιπρόσθετα, τέτοια γραφήματα παρέχουν αξία από μόνα τους. Αυτό μπορεί να γίνει εύκολα αντιληπτό δεδομένου ότι οι μηχανές αναζήτησης ιστού γενικής χρήσης είναι από τις πιο συχνά συμβουλευτικές πηγές όσον αφορά τις ιατρικές πληροφορίες, όπως αυτά που παρέχονται από την Google, χρησιμοποιώντας το γράφημα γνώσεων υγείας τους, έχουν τεράστιες δυνατότητες επιρροής (Rotmensch, Halpern, Tlimal, Horng, & Sontag, 2017). Η μελέτη που πραγματοποιήθηκε στην συγκεκριμένη έρευνα ήταν αναδρομική παρατήρηση με την χρήση δεδομένων από ηλεκτρονικά ιατρικά αρχεία για την δημιουργία γραφήματος γνώσης(ασθενειών-συμπτωμάτων) και έγινε μία αξιολόγηση των υποψήφιων γραφημάτων γνώσεων με ένα εκτενές και χειροκίνητο γράφημα γνώσεων, που παρέχεται από την Google (Rotmensch, Halpern, Tlimal, Horng, & Sontag, 2017). Τα δεδομένα που χρησιμοποιήθηκαν ήταν δομημένα και μη δομημένα και προήλθαν από θετικές αναφορές ασθενειών και συμπτωμάτων. Τα δομημένα δεδομένα αποτελούνταν από κωδικούς διάγνωσης ICD-9 (Διεθνής Ταξινόμηση Νοσημάτων). Τα μη δομημένα δεδομένα αποτελούνταν από κύρια καταγγελία, Αξιολόγηση Διαλογής, Σημειώσεις Νοσηλευτικής και σχόλια MD (Rotmensch, Halpern, Tlimal, Horng, & Sontag, 2017). Επίσης, το σύνολο των ασθενειών και των συμπτωμάτων που εξετάστηκαν στην μελέτη, επιλέχθηκαν από το GHKG για να δημιουργηθεί μια βάση για μεταγενέστερη σύγκριση. Μέσα από την έρευνα αυτή, φαίνεται πως η δημιουργία ενός γραφήματος γνώσεων υγείας υψηλής ποιότητας απευθείας από ηλεκτρονικά ιατρικά αρχεία είναι δυνατή. Δίνεται η δυνατότητα δημιουργίας γραφημάτων από EMR σε οποιονδήποτε αριθμό τομέων γρήγορα και χωρίς προηγούμενη γνώση, χάρις την μέθοδο αυτόματης κατασκευής γραφημάτων γνώσης. Μπορεί να διεξαχθεί το συμπέρασμα ότι η χρήση δεδομένων EMR για την δημιουργία γραφημάτων γνώσης, καθώς και η

χρήση των ίδιων των γραφημάτων γνώσης στον τομέα της ιατρικής είναι σωστές επιλογές, αφού τα αποτελέσματα είναι τα βέλτιστα.

2.16 Ενσωματώσεις γραφήματος γνώσης για πρόβλεψη επανένταξης στη Μονάδα Εντατικής Θεραπείας(MΕΘ).

Θέμα μελέτης της έρευνας αυτής αποτελούν οι επανεντάξεις στη Μονάδα Εντατικής Θεραπείας (ΜΕΘ) και ο τρόπος με τον οποίο οι οντολογίες και τα γραφήματα γνώσης μπορούν να βοηθήσουν στα προβλήματα που δημιουργούνται. Αρχικά, σημαντικό είναι να τονιστεί το γεγονός ότι με τις επανεντάξεις στη ΜΕΘ εμφανίζονται προβλήματα, όπως κίνδυνος για την υγεία των ασθενών με αυξημένα ποσοστά θνησιμότητας και γενική επιδείνωση της υγείας, καθώς και οικονομική επιβάρυνση για τις εγκαταστάσεις υγειονομικής περίθαλψης. Επίσης, πρέπει να αντιμετωπιστεί και το χάσμα που υπάρχει μεταξύ δεδομένων και επιστημονικού πλαισίου, το οποίο οφείλεται στη χρήση μεθόδων μηχανικής μάθησης που εφαρμόζονται στην πρόβλεψη του κινδύνου επανεισδοχής στη ΜΕΘ (*Carvalho, Oliveira, & Pesquita, 2023*). Η υγειονομική περίθαλψη καθοδηγείται περισσότερο από δεδομένα με την εισαγωγή των Ηλεκτρονικών Μητρώων Υγείας(EHR) και η χρήση μεθόδων μηχανικής μάθησης είναι αναγκαία. Ωστόσο, με την χρήση της δεν λαμβάνεται υπόψιν το νόημα και οι σχέσεις των δεδομένων. Λόγω αυτής της κατάστασης οι διεργασίες γίνονται τυφλά πάνω στα κλινικά δεδομένα, χωρίς να λαμβάνουν υπόψιν την επιστημονική γνώση και το πλαίσιο. Έπειτα, παρατίθεται η μεθοδολογία αντιμετώπισης των προβλημάτων που προτείνει η έρευνα. Αναλυτικότερα, παρουσιάζεται ένα Γράφημα Γνώσης, το οποίο συγκροτήθηκε από δεδομένα ΗΜΥ με σημασιολογικούς σχολιασμούς σε οντολογίες. Αναγράφονται 3 κύρια στοιχεία στην μελέτη τα οποία είναι:

- (1) συλλογή και προ επεξεργασία δεδομένων,
- (2) δημιουργία σημασιολογικών χαρακτηριστικών και
- (3) πρόβλεψη επανεισδοχής.

Τα δεδομένα που συλλέγονται αποτελούν όλες τις πληροφορίες της ΜΕΘ που απαιτούνται για τα μοντέλα ML και τον σημασιολογικό εμπλουτισμό από τη βάση δεδομένων MIMIC-III. Εξηγείται πως το

MIMIC-III είναι μια εκτεταμένη, ελεύθερα διαθέσιμη βάση δεδομένων που διαθέτει αρχεία που σχετίζονται με 53.423 διαφορετικές εισαγωγές στο νοσοκομείο ενηλίκων ασθενών. Αυτοί οι ασθενείς ανήκουν στην ηλικιακή ομάδα των 16 ετών και άνω και έχουν το κοινό ότι παρέμειναν στις μονάδες εντατικής θεραπείας του Beth Israel Deaconess Medical Center στα έτη 2001 και 2012 (*Carvalho, Oliveira, & Pesquita, 2023*). Επιπρόσθετα, αναφέρεται η μεθοδολογία παραμονή ενός ασθενούς στη ΜΕΘ. Η παραμονή αντιπροσωπεύεται από ενσωματώσεις Γραφήματος Γνώσης, οι οποίες είναι κατάλληλα αναγνωρίσιμες από τα οποία μοντέλα μηχανικής μάθησης και αφορούν την πρόβλεψη επανένταξης στη ΜΕΘ 30 ημερών (*Carvalho, Oliveira, & Pesquita, 2023*). Η μεθοδολογία παραμονής προσφέρει πλεονεκτήματα, καθώς εμπλουτίζει το σύνολο δεδομένων MIMIC-III με εξατομικευμένους σχολιασμούς στον ασθενή σε διάφορες βιοϊατρικές οντολογίες, επομένως, δημιουργεί ένα Γράφημα Γνώσης που ορίζει δεδομένα ασθενών με βιοϊατρικές οντολογίες, σχεδιάζει ένα προγνωστικό μοντέλο κινδύνου επανένταξης στη ΜΕΘ που χρησιμοποιεί ενσωματώσεις Γραφημάτων Γνώσης και τέλος, προτείνει μία παραλλαγή του προγνωστικού μοντέλου που στοχεύει σε διαφορετικά χρονικά σημεία κατά τη διάρκεια μιας παραμονής στη ΜΕΘ για καλύτερη διαχείριση. Με βάση την παραπάνω έρευνα αποδεικνύεται πως η εφαρμογή της νέας προσέγγισης μπορεί να βοηθήσει τους κλινικούς ιατρούς, να πάρουν την βέλτιστη απόφαση. Αυτό θα συμβαίνει, διότι θα είναι σε θέση να κρίνουν ορθότερα εάν ένας ασθενής μπορεί να πάρει εξιτήριο ή εάν δεν είναι απαραίτητη η επανένταξή του στην Μονάδα Εντατικής Θεραπείας. Έτσι, θα αποφεύγονται άσκοπες παραμονές ασθενών και η λειτουργία των ΜΕΘ θα είναι καλύτερη. Συμπερασματικά, τα προβλήματα που εμφανίζονται λόγω της αναγκαιότητας ύπαρξης κλινικών δεδομένων μπορούν να αντιμετωπιστούν με τον σημασιολογικό σχολιασμό και την ενσωμάτωση Γραφημάτων Γνώσης και έτσι, να παρέχουν την βέλτιστη λήψη αποφάσεων σχετικά με την επανένταξη στη ΜΕΘ, βοηθώντας στις ακριβείς προβλέψεις κινδύνου.

Κεφάλαιο 3^ο: Ανάλυση δεδομένων με γράφους γνώσης

Εισαγωγή

Ενδιαφέρον προκαλεί η εξέλιξη της τεχνολογίας και οι δυνατότητες που προσφέρει στην ανθρωπότητα. Η βελτίωση και η αντιμετώπιση των προβλημάτων στον τομέα υγείας, η οποία προκαλείται από τη χρήση της τεχνολογίας σε αυτόν, είναι δύο σημαντικά χαρακτηριστικά. Αναλυτικότερα, οι Γράφοι Γνώσης αποτελούν μία σπουδαία τεχνολογική καινοτομία και η χρήση τους μία άμεση λύση. Στη συγκεκριμένη έρευνα γίνεται έλεγχος και ανάλυση δεδομένων με τη χρήση Γράφων Γνώσης από ασθενείς που κατοικούν σε διάφορες περιοχές της υδρόγειου, δίνοντας έμφαση σε αυτούς που μένουν στις Ηνωμένες Πολιτείες. Η βασική ασθένεια που μελετάται είναι ο Covid-19 και η χρονική περίοδος καταγραφής των δεδομένων είναι τα έτη 2020-2023.

3.1 Προσέγγιση

Πυρήνα ενδιαφέροντος αποτελούν τα προνόμια που αποδίδει η τεχνολογία στον τομέα της υγείας. Έτσι, επιλογή του θέματος μελέτης ήταν η απόδειξη πως η τεχνολογία και συγκεκριμένα οι Γράφοι Γνώσης αποτελούν απαραίτητη προϋπόθεση στην καταπολέμηση των προβλημάτων υγείας. Σπουδαίο επίτευγμα είναι η επίλυση προβλημάτων αυτών, καθώς η υγεία είναι άμεσα συνδεδεμένη με την επιβίωση του ανθρώπου. Για να μπορέσει να αποδειχθεί η παραπάνω θεωρία, έπρεπε να γίνει ανάλυση βιολογικών δεδομένων. Για τον λόγο αυτό, προτιμήθηκαν δεδομένα που αφορούσαν κρούσματα από την νόσο Covid-19, που είναι η πιο πρόσφατη αιτία παγκόσμιας πανδημίας, επικράτησης πανικού και αμφισβήτησης για το μέλλον. Επιπλέον, για να υπάρχει μία αναλυτική και σφαιρική αντίληψη σχετικά με το θέμα μελέτης στα ιατρικά δεδομένα συμπεριλήφθηκαν και οι υπόλοιπες α-

σθένειες από τις οποίες μπορεί να υπέφερε ένας ασθενής. Τέλος, η χρονική περίοδος 2020-2023 ήταν αυτή που επιλέχθηκε για μελέτη των καταγραφών των δεδομένων αυτών.

3.2 Σκοπός

Σκοπός της πτυχιακής αποτέλεσε η απόδειξη της απαραίτητης χρήσης των Γράφων Γνώσης στον τομέα της υγείας και η ανάδειξη των προτερημάτων αυτής της χρήσης. Με την μελέτη αυτή διαφαίνεται πως η χρήση Γράφων Γνώσης επιλύει πολλά προβλήματα. Επίσης, οι Γράφοι Γνώσης μπορούν να προσαρμοστούν σε όλων των τύπων τα προβλήματα και να επιφέρουν κατάλληλα αποτελέσματα.

3.3 Διαδικασία διεκπεραίωσης έρευνας

Για να επιτευχθεί η αποπεράτωση της έρευνας χρειάστηκε η ακολουθία κάποιων βημάτων. Αρχικά, μελετήθηκαν κάποια σύνολα δεδομένων, επεξεργάστηκαν και από αυτά συγκεντρώθηκαν τα δεδομένα τα οποία και αποδείχθηκαν κατάλληλα για την έρευνα. Έπειτα, απαραίτητη ήταν μία γενική σχεδίαση του γράφου γνώσης, έτσι ώστε να υπάρχει μία ολοκληρωμένη εικόνα σχετικά με την τοποθέτηση των κόμβων, των σχέσεων, καθώς και των δεδομένων. Στη συνέχεια, δημιουργήθηκε στο Neo4j το τελικό σχήμα του Γράφου Γνώσης και αποθηκεύτηκαν τα δεδομένα σε αυτόν. Τέλος, παράχθηκαν τα queries που χρησιμοποιήθηκαν στον Γράφο Γνώσης για την έκβαση των συμπερασμάτων.

Στο 3^ο κεφάλαιο αναφέρθηκε επιγραμματικά η διαδικασία που έγινε, ώστε να ολοκληρωθεί η έρευνα. Προχωρώντας στο 4^ο κεφάλαιο αναγράφονται λεπτομερώς η δημιουργία του Γράφου Γνώσης και τα δεδομένα τα οποία αποθηκεύτηκαν σε αυτόν.

Κεφάλαιο 4^ο: Γράφος γνώσης έρευνας και δεδομένα

Εισαγωγή

Για την ολοκλήρωση της έρευνας απαραίτητη ήταν η ανάλυση βιολογικών και οικονομικών δεδομένων. Για την ανάλυση αυτή χρειάστηκε η κατασκευή Γράφου Γνώσης. Μέσω ενός συστήματος διαχείρισης βάσεων δεδομένων γραφημάτων, του Neo4j, έγινε η κατασκευή του Γράφου Γνώσης, η χρήση ερωτημάτων (queries) προς την βάση δεδομένων και τέλος, η συγκέντρωση συμπερασμάτων. Στο κεφάλαιο 3, ακολουθεί αναλυτική περιγραφή δημιουργίας του Γράφου Γνώσης, παρουσίαση των δεδομένων που χρησιμοποιήθηκαν σε αυτόν, καθώς και αναφορά των queries που παρατέθηκαν για την εξαγωγή συμπερασμάτων.

4.1 Δημιουργία Γράφου Γνώσης

4.4.1 Διαδικασία σχεδιασμού

Η υλοποίηση του Γράφου Γνώσης ξεκίνησε με την καταγραφή των στοιχείων που θα είχε. Επομένως, σημειώθηκαν οι κόμβοι του, τα γνωρίσματά τους και οι σχέσεις μεταξύ των κόμβων. Παρατηρήθηκε πως χρειάστηκαν κόμβοι οι οποίοι θα αντιπροσώπευαν τους ασθενείς, οι οποίοι είχαν προσβληθεί από τον Covid-19, τις επιπλέον ασθένειες, οι οποίες θα ήταν στο ιατρικό ιστορικό των ασθενών και τις ημερομηνίες που γινόντουσαν οι καταγραφές. Έτσι, δημιουργήθηκαν οι κόμβοι Patient, Diseases και Dates αντίστοιχα. Έπειτα, καταγράφηκαν τα γνωρίσματα των κόμβων αυτών:

- Για τον κόμβο Patient -> RaceorHispanicOriginGroup, State, Sex, AgeGroup και pid.
- Για τον κόμβο Diseases -> TotalDeaths, NaturalCause, Septicemia, Malignantneoplasms, Diabetesmellitus, Alzheimerdisease, Influenzaandpneumonia, Chroniclowerrespiratorydiseases, Otherdiseasesofrespiratorysystem, Nephritisnephroticsyndromeandnephrosis, Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified, Diseasesofheart, Cerebrovascular diseases, COVID19MultipleCause, COVID19UnderlyingCause και diid.
- Για τον κόμβο Dates -> StartWeek, EndWeek, MMWRWeek, MMWRYear και daid.

Επεξήγηση Γνωρισμάτων:

- RaceorHispanicOriginGroup: αφορά την Ομάδα φυλής ή την Ισπανική καταγωγή, το οποίο χρησιμοποιήθηκε για την μελέτη των ασθενών που είχαν ισπανική καταγωγή ή καταγωγή από φυλές όπως Ιθαγενείς Χαβάης ή νησιώτες του Ειρηνικού, Αμερικανός Ινδιάνος ή Ιθαγενής της Αλάσκας, Ασιάτες,
- State: αφορά, επίσης, καταγωγή με την διαφορά ότι αναφέρεται αποκλειστικά σε ασθενείς που έμεναν στις Ηνωμένες Πολιτείες. Αυτή η διάκριση μεταξύ των δύο χαρακτηριστικών (RaceorHispanicOriginGroup και State), γίνεται καθώς η έρευνα αυτή μελετά τους ασθενείς στις Ηνωμένες Πολιτείες, όπως προαναφέρθηκε στο κεφάλαιο 3, αλλά χρησιμοποιήθηκαν και δεδομένα για περιπτώσεις ασθενών άλλης καταγωγής για να μπορέσει να γίνει σύγκριση,
- Sex: σημαντικό στοιχείο ήταν το φύλο του ασθενούς,
- AgeGroup: μονάδα σύγκρισης αποτέλεσε η ηλικιακή ομάδα στην οποία ανήκε ο ασθενής,
- pid: χρησιμοποιήθηκε ως το id της κάθε πλειάδας του κόμβου Patient, δηλαδή το πρωτεύον κλειδί του,
- TotalDeaths: αναφέρεται στους συνολικούς θανάτους (φυσικούς ή από ασθένειες),
- NaturalCause: αναφέρεται στους συνολικούς θανάτους από φυσικά αίτια,
- Septicemia: μελέτη των κρουσμάτων της Σηψαιμίας,
- Malignantneoplasms: μελέτη των κρουσμάτων των Κακοήθων νεοπλασμάτων,
- Diabetesmellitus: μελέτη των κρουσμάτων του Σακχαρώδη διαβήτη,
- Alzheimerdisease: μελέτη των κρουσμάτων της νόσου του Αλτσχάιμερ,
- Influenzaandpneumonia: μελέτη κρουσμάτων στην κατάσταση ταυτόχρονης νοσηλείας για γρίπη και πνευμονία,
- Chroniclowerrespiratorydiseases: μελέτη των κρουσμάτων χρόνιων αναπνευστικών παθήσεων,
- Otherdiseasesofrespiratorysystem: μελέτη των κρουσμάτων από άλλες ασθένειες του αναπνευστικού συστήματος. Στο σημείο αυτό γίνεται ένας διαχωρισμός στα γνωρίσματα Chroniclowerrespiratorydiseases και Otherdiseasesofrespiratorysystem. Παρά το γεγονός ότι και τα δύο αναφέρονται σε παθήσεις του αναπνευστικού συστήματος, απαραίτητη ήταν αυτή η κατηγοριοποίηση, καθώς η ασθένεια Covid-19 αποδείχθηκε, ότι καταβάλλει ιδιαίτερα το αναπνευστικό σύστημα,
- Nephritisnephroticsyndromeandnephrosis: μελέτη των κρουσμάτων Νεφρωτικού σύνδρομου και νέφρωσης,

- Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified: μελέτη των κρουσμάτων που αφορούν σημάδια συμπτωμάτων και μη φυσιολογικά κλινικών και εργαστηριακών ευρημάτων,
- Diseasesofheart: μελέτη των κρουσμάτων σε παθήσεις της καρδιάς,
- Cerebrovasculardiseases: μελέτη των κρουσμάτων από εγκεφαλοαγγειακές παθήσεις,
- COVID19MultipleCause: μελέτη των κρουσμάτων από Covid-19. Το συγκεκριμένο γνώρισμα αναφέρεται στον όρο μακροχρόνιου COVID. Αναλυτικότερα, θα μπορούσε να εξηγηθεί ως μία παθολογική κατάσταση, στην οποία τα συμπτώματα εμφανίζονται ή επιμένουν μετά την περίοδο ανάρρωσης.
- COVID19UnderlyingCause: μελέτη των κρουσμάτων, επίσης, από Covid-19. Το γνώρισμα αυτό αναφέρεται στις περιπτώσεις θανάτων στις οποίες ο Covid-19 αποτελεί την υποκείμενη αιτία θανάτου. Η αιτία αυτή χαρακτηρίζεται ως υποκείμενη, διότι ένας θάνατος λόγω COVID-19 θα πρέπει να υπολογίζεται ανεξάρτητα από τις προ υπάρχουσες καταστάσεις υγείας του ασθενούς. Έτσι, θα πρέπει να καταγράφεται στο ιατρικό πιστοποιητικό αιτίας θανάτου ότι ο ασθενής νοσούσε και από COVID-19, που προκάλεσε ή συνέβαλε στο θάνατο. Το συγκεκριμένο γνώρισμα χρησιμοποιήθηκε για λόγους επιτήρησης των θανάτων, που προκύπτουν από μια πιθανή ή επιβεβαιωμένη περίπτωση COVID-19.
- diid: χρησιμοποιήθηκε ως το id της κάθε πλειάδας του κόμβου Diseases, δηλαδή το πρωτεύον κλειδί του,
- StartWeek: αναφέρεται στην εβδομάδα, η οποία ορίζεται για την συγκεκριμένη έρευνα ως η αφετηρία καταγραφής των κρουσμάτων,
- EndWeek: αναφέρεται στην εβδομάδα, η οποία ορίζεται για την συγκεκριμένη έρευνα ως το πέρας καταγραφής των κρουσμάτων,
- MMWRWeek: το γνώρισμα αυτό έχει τα στοιχεία (ημερομηνίες) της επιδημιολογικής εβδομάδας για την οποία η έκθεση νόσου του Εθνικού Συστήματος Επιτήρησης Ασθενειών Υποχρεωτικής Δήλωσης (NNDSS) ανατίθεται από το τοπικό ή κρατικό τμήμα υγείας για τους σκοπούς της αναφοράς και δημοσίευσης της επίπτωσης της νόσου. Στην συγκεκριμένη έρευνα αναφέρεται στον COVID-19.
- MMWRYear: το γνώρισμα αυτό έχει τα στοιχεία (ημερομηνίες) του επιδημιολογικού έτους για το οποίο η έκθεση νόσου του Εθνικού Συστήματος Επιτήρησης Ασθενειών Υποχρεωτικής Δήλωσης (NNDSS) ανατίθεται από το τοπικό ή κρατικό τμήμα υγείας για τους σκοπούς της ανα-

φοράς και δημοσίευσης της επίπτωσης της νόσου. Στην συγκεκριμένη έρευνα αναφέρεται στον COVID-19 και

- daid: χρησιμοποιήθηκε ως το id της κάθε πλειάδας του κόμβου Dates, δηλαδή το πρωτεύον κλειδί του.

Τα παραπάνω δεδομένα αποθηκεύτηκαν σε 3 διαφορετικά excel με τα ονόματα Patient, Diseases και Dates. Χρειάστηκαν 3 excel, διότι το κάθε ένα αντιστοιχεί στους κόμβους του Γράφου Γνώσης της μελέτης. Στην συνέχεια, τα excel αποθηκεύτηκαν με την μορφή CSV UTF-8, ώστε να είναι δυνατή η εισαγωγή των δεδομένων τους στο Neo4j και η επεξεργασία τους. Παρακάτω απεικονίζονται τα τρία excel, η μετατροπή τους σε αρχείο CSV UTF-8 και η μορφή του Γράφου Γνώσης μετά την δημιουργία των κόμβων και την εισαγωγή των δεδομένων.

	A	B	C	D	E
1	RaceorHispanicOriginGroup	State	Sex	AgeGroup	pid
2	All	United States	All Sexes	All Ages	1
3	All	United States	All Sexes	Under 1 year	2
4	All	United States	All Sexes	First 17 years	3
5	All	United States	All Sexes	1-4 years	4
6	All	United States	All Sexes	5-14 years	5
7	All	United States	All Sexes	15-24 years	6
8	All	United States	All Sexes	18-29 years	7
9	All	United States	All Sexes	25-34 years	8
10	All	United States	All Sexes	30-39 years	9
11	All	United States	All Sexes	35-44 years	10
12	All	United States	All Sexes	40-49 years	11
13	Hispanic	United States	All Sexes	45-54 years	12
14	Non-Hispanic American Indian or Alaska Native	United States	All Sexes	50-64 years	13
15	Non-Hispanic Asian	United States	All Sexes	55-64 years	14
16	Non-Hispanic Black	United States	All Sexes	65-74 years	15
17	Non-Hispanic More than one race	United States	All Sexes	75-84 years	16
18	Non-Hispanic Native Hawaiian or Other Pacific Islander	United States	All Sexes	85 years and over	17
19	Non-Hispanic White	United States	Male	All Ages	18
20	Total	United States	Male	Under 1 year	19
21	Unknown	United States	Male	First 17 years	20
22	Hispanic	United States	Male	1-4 years	21
23	Non-Hispanic American Indian or Alaska Native	United States	Male	5-14 years	22
24	Non-Hispanic Asian	United States	Male	15-24 years	23
25	Non-Hispanic Black	United States	Male	18-29 years	24
26	Non-Hispanic More than one race	United States	Male	25-34 years	25
27	Non-Hispanic Native Hawaiian or Other Pacific Islander	United States	Male	30-39 years	26
28	Non-Hispanic White	United States	Male	35-44 years	27
29	Total	United States	Male	40-49 years	28
30	Unknown	United States	Male	45-54 years	29

Εικόνα 18 Απεικόνιση δεδομένων σε μορφή excel, που αντιστοιχούν στον κόμβο Patient.

```

RaceorHispanicOriginGroup,State,Sex,AgeGroup,pid
All,United States,All Sexes,All Ages,1
All,United States,All Sexes,Under 1 year,2
All,United States,All Sexes,First 17 years,3
All,United States,All Sexes,1-4 years,4
All,United States,All Sexes,5-14 years,5
All,United States,All Sexes,15-24 years,6
All,United States,All Sexes,18-29 years,7
All,United States,All Sexes,25-34 years,8
All,United States,All Sexes,30-39 years,9
All,United States,All Sexes,35-44 years,10
All,United States,All Sexes,40-49 years,11
Hispanic,United States,All Sexes,45-54 years,12
Non-Hispanic American Indian or Alaska Native,United
States,All Sexes,50-64 years,13
Non-Hispanic Asian,United States,All Sexes,55-64 years,14
Non-Hispanic Black,United States,All Sexes,65-74 years,15
Non-Hispanic More than one race,United States,All Sexes,75-84
years,16
Non-Hispanic Native Hawaiian or Other Pacific Islander,United
States,All Sexes,85 years and over,17
Non-Hispanic White,United States,Male,All Ages,18
Total,United States,Male,Under 1 year,19
Unknown,United States,Male,First 17 years,20
Hispanic,United States,Male,1-4 years,21
Non-Hispanic American Indian or Alaska Native,United
States,Male,5-14 years,22
Non-Hispanic Asian,United States,Male,15-24 years,23
Non-Hispanic Black,United States,Male,18-29 years,24
Non-Hispanic More than one race,United States,Male,25-34
years,25
Non-Hispanic Native Hawaiian or Other Pacific Islander,United
States,Male,30-39 years,26
Non-Hispanic White,United States,Male,35-44 years,27
Total,United States,Male,40-49 years,28
Unknown,United States,Male,45-54 years,29

```

Εικόνα 19 Απεικόνιση δεδομένων σε μορφή csv utf-8, που αντιστοιχούν στον κόμβο Patient.

	A	B	C	D	E	F	G	H	I	J
1	TotalDeaths	NaturalCause	Septicemia	Malignantneoplasms	Diabetesmellitus	Alzheimerdisease	Influenzaandpneumonia	Chroniclowerrespiratorydiseases	Otherdiseasesofrespiratorysystem	Nephritisnephroticsyndromeandnephros
2	60179	55010	843	11567	1829	2537	1559	3501	1068	1094
3	60735	55755	861	11963	1942	2566	1528	3708	1035	1092
4	59363	54516	829	11701	1819	2491	1484	3526	992	1121
5	59162	54401	828	11879	1864	2517	1488	3403	979	1107
6	58834	54001	811	11963	1828	2480	1411	3313	981	1074
7	59491	54416	806	11704	1957	2515	1463	3413	975	1135
8	58815	53969	792	11807	1848	2537	1515	3477	979	1070
9	58912	53987	779	11780	1880	2515	1462	3454	968	1059
10	59334	54316	814	11790	1831	2519	1508	3460	1011	1092
11	59716	54401	811	11712	1868	2511	1608	3471	1002	1071
12	58672	53529	756	11568	1743	2445	1641	3390	993	1078
13	59220	54302	840	11732	1830	2514	1740	3384	1021	1105
14	63026	58238	846	11782	2044	2749	1783	3519	1064	1024
15	72294	67445	948	11592	2298	2871	1849	3538	1027	1038
16	79113	74012	833	11549	2354	2963	1613	3442	1017	1120
17	76809	71895	745	11210	2270	2899	1232	3195	916	1098
18	73909	68740	735	11367	2085	2804	1138	2993	887	987
19	69304	63931	739	11097	1935	2727	1009	2928	851	961
20	66814	61173	720	11014	1972	2498	864	2808	855	938
21	64479	58978	679	11269	1969	2432	857	2770	787	976
22	61642	56014	713	11116	1834	2417	804	2675	743	917
23	59682	53987	654	10903	1813	2261	751	2634	759	893
24	58908	52924	729	11083	1727	2302	701	2563	779	904
25	58034	52274	693	11130	1741	2326	700	2502	737	923
26	57999	52211	685	11157	1792	2362	756	2563	761	933
27	58510	52643	719	11356	1766	2288	722	2539	775	952
28	59837	53815	682	11298	1933	2362	658	2624	780	985
29	61962	55858	764	11333	1954	2472	739	2615	751	915
30	63177	57140	708	11373	1887	2501	767	2647	712	984

Εικόνα 20 Πρώτο μέρος απεικόνισης δεδομένων σε μορφή excel, που αντιστοιχούν στον κόμβο Diseases.

K	L	M	N	O	P
Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassi	Diseasesofheart	Cerebrovascular diseases	COVID19MultipleCause	COVID19UnderlyingCause	diid
630	14204	3110	0	0	1
649	13911	3189	1	1	2
612	13593	3256	2	1	3
638	13612	3185	3	2	4
616	13465	3084	0	0	5
600	14005	3056	4	3	6
613	13641	3089	6	2	7
608	13632	3083	6	6	8
676	13714	3126	9	9	9
655	13693	3097	38	34	10
642	13437	3165	60	55	11
618	13203	3066	588	550	12
629	13710	3067	3226	3043	13
689	14945	3165	10141	9638	14
664	15766	3191	16347	15528	15
668	14570	3204	17221	16305	16
675	13858	3058	15557	14659	17
548	13004	3042	13223	12437	18
587	13170	2849	11243	10463	19
563	12735	2958	9239	8497	20
569	12796	2827	7249	6593	21
576	12448	2851	6176	5596	22
554	12463	2791	5060	4519	23
587	12413	2838	4233	3730	24
611	12412	2923	3852	3347	25
575	12515	2959	3846	3333	26
593	12877	2832	4557	4050	27
668	13042	2921	5786	5236	28
666	12868	2982	7198	6530	29

Εικόνα 21 Δεύτερο μέρος απεικόνισης δεδομένων σε μορφή excel, που αντιστοιχούν στον κόμβο Diseases.

TotalDeaths,NaturalCause,Septicemia ,Malignantneoplasms ,Diabetes mellitus ,Alzheimerdisease ,Influenzaandpneumonia ,Chroniclowerrespiratorydiseases ,Otherdiseasesofrespiratorysystem ,Nephritisnephroticsyndromeandnephrosis ,Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified ,Diseasesofheart ,Cerebrovascular diseases ,COVID19MultipleCause ,COVID19UnderlyingCause ,diid
60179,55010,843,11567,1829,2537,1559,3501,1068,1094,630,14204,3110,0,0,1
60735,55755,861,11963,1942,2566,1528,3708,1035,1092,649,13911,3189,1,1,2
59363,54516,829,11701,1819,2491,1484,3526,992,1121,612,13593,3256,2,1,3
59162,54401,828,11879,1864,2517,1488,3403,979,1107,638,13612,3185,3,2,4
58834,54001,811,11963,1828,2480,1411,3313,981,1074,616,13465,3084,0,0,5
59491,54416,806,11704,1957,2515,1463,3413,975,1135,600,14005,3056,4,3,6
58815,53969,792,11807,1848,2537,1515,3477,979,1070,613,13641,3089,6,2,7
58912,53987,779,11780,1880,2515,1462,3454,968,1059,608,13632,3083,6,6,8
59334,54316,814,11790,1831,2519,1508,3460,1011,1092,676,13714,3126,9,9,9
59716,54401,811,11712,1868,2511,1608,3471,1002,1071,655,13693,3097,38,34,10
58672,53529,756,11568,1743,2445,1641,3390,993,1078,642,13437,3165,60,55,11
59220,54302,840,11732,1830,2514,1740,3384,1021,1105,618,13203,3066,588,550,12
63026,58238,846,11782,2044,2749,1783,3519,1064,1024,629,13710,3067,3226,3043,13
72294,67445,948,11592,2298,2871,1849,3538,1027,1038,689,14945,3165,10141,9638,14
79113,74012,833,11549,2354,2963,1613,3442,1017,1120,664,15766,3191,16347,15528,15
76809,71895,745,11210,2270,2899,1232,3195,916,1098,668,14570,3204,17221,16305,16

Εικόνα 22 Πρώτο μέρος απεικόνισης δεδομένων σε μορφή csv utf-8, που αντιστοιχούν στον κόμβο Diseases.

```

73909,68740,735,11367,2085,2804,1138,2993,887,987,675,13858,3058,
15557,14659,17
69304,63931,739,11097,1935,2727,1009,2928,851,961,548,13004,3042,
13223,12437,18
66814,61173,720,11014,1972,2498,864,2808,855,938,587,13170,2849,1
1243,10463,19
64479,58978,679,11269,1969,2432,857,2770,787,976,563,12735,2958,9
239,8497,20
61642,56014,713,11116,1834,2417,804,2675,743,917,569,12796,2827,7
249,6593,21
59682,53987,654,10903,1813,2261,751,2634,759,893,576,12448,2851,6
176,5596,22
58908,52924,729,11083,1727,2302,701,2563,779,904,554,12463,2791,5
060,4519,23
58034,52274,693,11130,1741,2326,700,2502,737,923,587,12413,2838,4
233,3730,24
57999,52211,685,11157,1792,2362,756,2563,761,933,611,12412,2923,3
852,3347,25
58510,52643,719,11356,1766,2288,722,2539,775,952,575,12515,2959,3
846,3333,26
59837,53815,682,11298,1933,2362,658,2624,780,985,593,12877,2832,4
557,4050,27
61962,55858,764,11333,1954,2472,739,2615,751,915,668,13042,2921,5
786,5236,28
63177,57140,708,11373,1887,2501,767,2647,712,984,666,12868,2982,7
198,6530,29

```

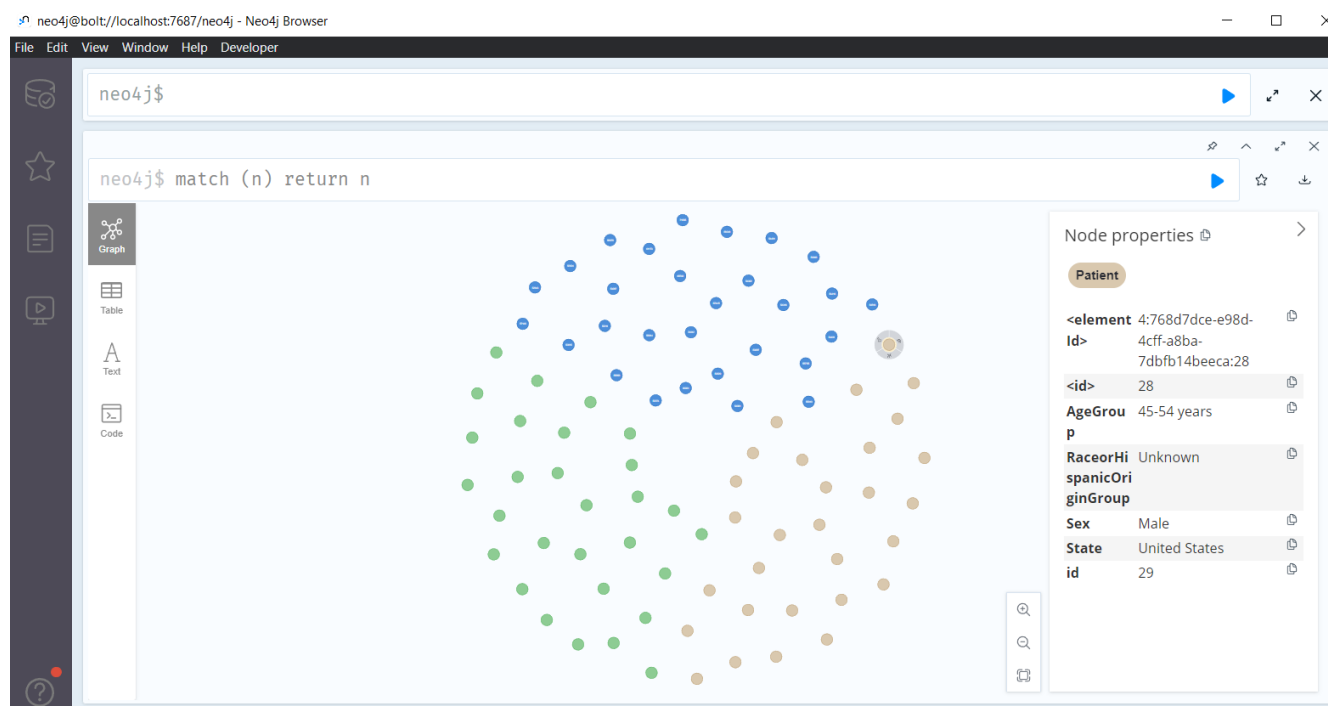
Εικόνα 23 Δεύτερο μέρος απεικόνισης δεδομένων σε μορφή csv utf-8, που αντιστοιχούν στον κόμβο Diseases.

	A	B	C	D	E
1	StartWeek	EndWeek	MMWRWeek	MMWRYear	daid
2	01/04/2020	06/24/2023	1	2020	1
3	01/04/2020	06/24/2023	2	2020	2
4	01/04/2020	06/24/2023	3	2020	3
5	01/04/2020	06/24/2023	4	2020	4
6	01/04/2020	06/24/2023	5	2020	5
7	01/04/2020	06/24/2023	6	2020	6
8	01/04/2020	06/24/2023	7	2020	7
9	01/04/2020	06/24/2023	8	2020	8
10	01/04/2020	06/24/2023	9	2020	9
11	01/04/2020	06/24/2023	10	2020	10
12	01/04/2020	06/24/2023	11	2020	11
13	01/04/2020	06/24/2023	12	2020	12
14	01/04/2020	06/24/2023	13	2020	13
15	01/04/2020	06/24/2023	14	2020	14
16	01/04/2020	06/24/2023	15	2020	15
17	01/04/2020	06/24/2023	16	2020	16
18	01/04/2020	06/24/2023	17	2020	17
19	01/04/2020	06/24/2023	18	2020	18
20	01/04/2020	06/24/2023	19	2020	19
21	01/04/2020	06/24/2023	20	2020	20
22	01/04/2020	06/24/2023	21	2020	21
23	01/04/2020	06/24/2023	22	2020	22
24	01/04/2020	06/24/2023	23	2020	23
25	01/04/2020	06/24/2023	24	2020	24
26	01/04/2020	06/24/2023	25	2020	25
27	01/04/2020	06/24/2023	26	2020	26
28	01/04/2020	06/24/2023	27	2020	27
29	01/04/2020	06/24/2023	28	2020	28
30	01/04/2020	06/24/2023	29	2020	29

Εικόνα 24 Απεικόνιση δεδομένων σε μορφή excel, που αντιστοιχούν στον κόμβο Dates.

```
StartWeek,EndWeek,MMWRWeek,MMWYear,daid
01/04/2020,06/24/2023,1,2020,1
01/04/2020,06/24/2023,2,2020,2
01/04/2020,06/24/2023,3,2020,3
01/04/2020,06/24/2023,4,2020,4
01/04/2020,06/24/2023,5,2020,5
01/04/2020,06/24/2023,6,2020,6
01/04/2020,06/24/2023,7,2020,7
01/04/2020,06/24/2023,8,2020,8
01/04/2020,06/24/2023,9,2020,9
01/04/2020,06/24/2023,10,2020,10
01/04/2020,06/24/2023,11,2020,11
01/04/2020,06/24/2023,12,2020,12
01/04/2020,06/24/2023,13,2020,13
01/04/2020,06/24/2023,14,2020,14
01/04/2020,06/24/2023,15,2020,15
01/04/2020,06/24/2023,16,2020,16
01/04/2020,06/24/2023,17,2020,17
01/04/2020,06/24/2023,18,2020,18
01/04/2020,06/24/2023,19,2020,19
01/04/2020,06/24/2023,20,2020,20
01/04/2020,06/24/2023,21,2020,21
01/04/2020,06/24/2023,22,2020,22
01/04/2020,06/24/2023,23,2020,23
01/04/2020,06/24/2023,24,2020,24
01/04/2020,06/24/2023,25,2020,25
01/04/2020,06/24/2023,26,2020,26
01/04/2020,06/24/2023,27,2020,27
01/04/2020,06/24/2023,28,2020,28
01/04/2020,06/24/2023,29,2020,29
```

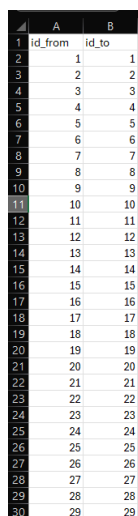
Εικόνα 25 Απεικόνιση δεδομένων σε μορφή csv utf-8, που αντιστοιχούν στον κόμβο Dates.



Εικόνα 26 Στην συγκεκριμένη εικόνα απεικονίζονται οι κόμβοι του Γράφου Γνώσης και τα γνωρίσματά τους. Στο δεξί μέρος της εικόνας φαίνονται τα γνωρίσματα ενός από τους κόμβους που ανήκουν στο label Patient.

Στη συνέχεια, μετά την δημιουργία των κόμβων χρειάστηκε να προστεθούν οι κατάλληλες σχέσεις που τους συνδέουν. Οι σχέσεις αυτές αναπτύχθηκαν μεταξύ των κόμβων Patient και Diseases με την ονομασία DIED και των κόμβων Diseases και Dates με την ονομασία DATE. Η σκέψη πίσω από τις συνδέσεις ήταν:

- μεταξύ των κόμβων Patient και Diseases: οι ασθενείς, που αντιπροσωπεύουν οι κόμβοι με το label Patient, πεθαίνουν από τις ασθένειες που έχουν, που οι ασθένειες αυτές αντιπροσωπεύονται από κόμβους με το label Diseases. Επομένως, έχουμε την σχέση DIED που συνδέει τους ασθενείς με τις ασθένειες τους και
- μεταξύ των κόμβων Diseases και Dates: στην έρευνα χρησιμοποιήθηκαν καταγραφές των κρουσμάτων από ασθένειες, που αντιπροσωπεύονται από κόμβους με το label Diseases και οι καταγραφές αυτές έγιναν σε συγκεκριμένες ημερομηνίες, οι οποίες αποθηκεύτηκαν στους κόμβους με το label Dates. Έτσι, έχουμε τη σχέση DATE που συνδέει τα κρούσματα και την ημερομηνία την οποία εμφανίστηκαν.
- Επίσης, απαραίτητη ήταν η δημιουργία ενός excel και η μετατροπή του σε μορφή CSV UTF-8, το οποίο θα είχε τις σχέσεις μεταξύ των κόμβων. Το συγκεκριμένο excel χρησιμοποιήθηκε για να μπορέσει να γίνει η σύνδεση και ονομάστηκε Relationships. Τέλος, το αρχείο έγινε import στη βάση δεδομένων. Παρακάτω απεικονίζεται το excel, η μετατροπή του σε αρχείο CSV UTF-8 και η μορφή του Γράφου Γνώσης μετά την δημιουργία των σχέσεων.



	A	B
	id_from	id_to
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	28	28
30	29	29

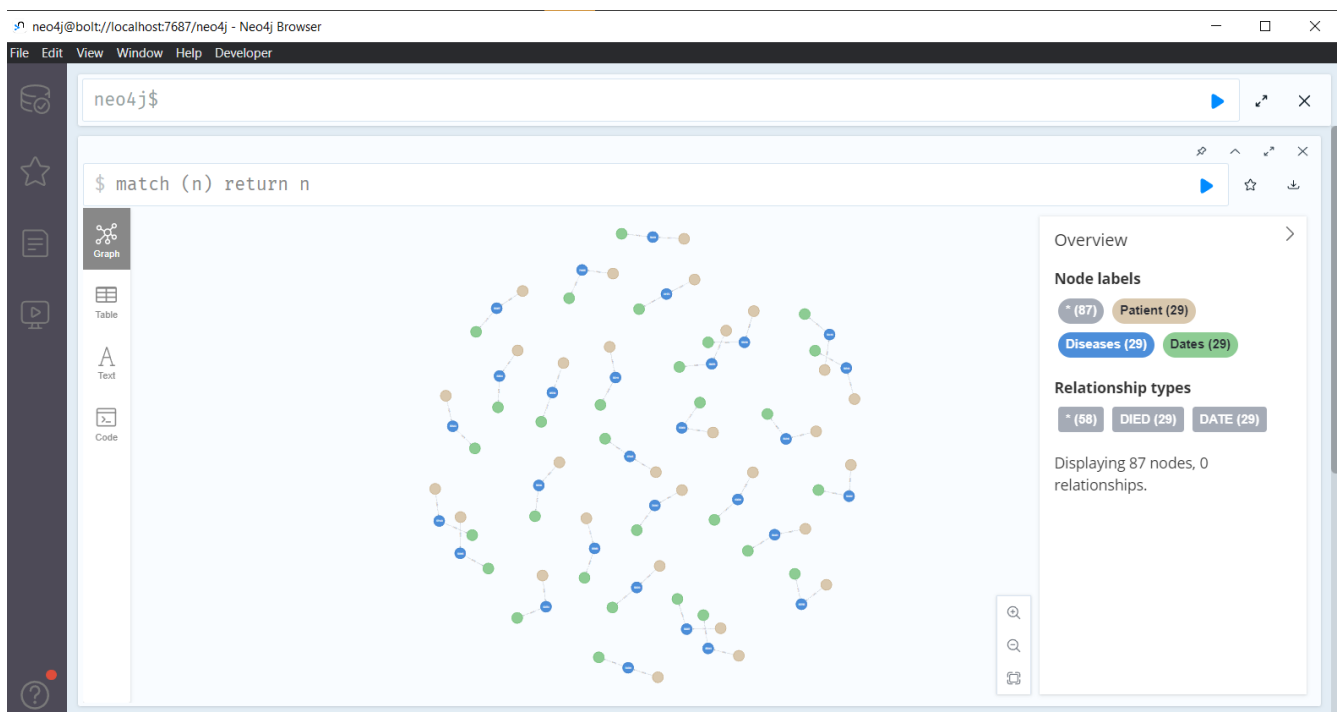
Εικόνα 27 Απεικόνιση δεδομένων σε μορφή excel, που αντιστοιχούν στην δημιουργία των σχέσεων μεταξύ των κόμβων.

```

id_from,id_to
1,1
2,2
3,3
4,4
5,5
6,6
7,7
8,8
9,9
10,10
11,11
12,12
13,13
14,14
15,15
16,16
17,17
18,18
19,19
20,20
21,21
22,22
23,23
24,24
25,25
26,26
27,27
28,28
29,29

```

Εικόνα 28 Απεικόνιση δεδομένων σε μορφή csv utf-8, που αντιστοιχούν στην δημιουργία των σχέσεων μεταξύ των κόμβων.



Εικόνα 29 Στην συγκεκριμένη εικόνα απεικονίζεται η τελική μορφή του Γράφου Γνώσης που χρησιμοποιήθηκε στην έρευνα.

Ο κώδικας που χρησιμοποιήθηκε για την δημιουργία του Γράφου Γνώσης αναγράφεται παρακάτω:

- Για την προσθήκη των κόμβων, των σχέσεων τους και των δεδομένων τους

```
LOAD CSV WITH HEADERS FROM "file:///Patient.csv" AS row
CREATE (:Patient {RaceorHispanicOriginGroup: row.RaceorHispanicOriginGroup,
State:row.State, Sex:row.Sex, AgeGroup:row.AgeGroup, id:row.pid});
```

```
LOAD CSV WITH HEADERS from 'file:///Diseases.csv' as diseases create (:Diseases {TotalDeaths:
diseases.TotalDeaths, NaturalCause:diseases.NaturalCause, Septicemia:diseases.Septicemia,
Malignantneoplasms:diseases.Malignantneoplasms, Diabetesmelli-
tus:diseases.Diabetesmellitus, Alzheimerdisease:diseases.Alzheimerdisease, Influenzaandpneumonia:diseases.Influenzaandpneumonia,
Chroniclowerrespiratorydiseases:diseases.Chroniclowerrespiratorydiseases, Otherdis-
easesofrespiratorysystem:diseases.Otherdiseasesofrespiratorysystem,
Nephritisnephroticsyndromeandnephrosis:diseases.Nephritisnephroticsyndromeandnephrosis,
Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelse-
whereclassi-
fied:diseases.Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified,
Diseasesofheart:diseases.Diseasesofheart, Cerebrovasculardiseas-
es:diseases.Cerebrovasculardiseases, COVID19MultipleCause:diseases.COVID19MultipleCause,
COVID19UnderlyingCause:diseases.COVID19UnderlyingCause, id:diseases.diid});
```

```
LOAD CSV WITH HEADERS FROM "file:///Dates.csv" AS dates
CREATE (:Dates {StartWeek: dates.StartWeek, EndWeek:dates.EndWeek, MMWR-
Week:dates.MMWRWeek, MMWRYear:dates.MMWRYear, id: dates.daid}) ;
```

```
LOAD CSV WITH HEADERS FROM "file:///Relationships.csv" AS row
//look up the two nodes we want to connect up
MATCH (p1:Patient {id:row.id_from}), (p2:Diseases {id:row.id_to})
//now create a relationship between them
CREATE (p1)-[:DIED]->(p2);
```

```
LOAD CSV WITH HEADERS FROM "file:///Relationships.csv" AS row
MATCH (d1:Diseases {id:row.id_from}), (d2:Dates {id:row.id_to})
```

//now create a relationship between them

CREATE (d1)-[:DATE]->(d2);

- Για την προσθήκη των δεδομένων που δεν μπόρεσαν να γίνουν import, τα οποία properties στο Patient και στο Diseases

```
LOAD CSV FROM "file:///Patient.csv" AS rec
```

```
MERGE (n:Patient {id:rec[4]})
```

```
ON MATCH SET n.RaceorHispanicOrigin= rec[0], n.AgeGroup= rec[3]
```

```
LOAD CSV FROM "file:///Diseases.csv" AS rec
```

```
MERGE (n:Diseases {id:rec[15]})
```

```
ON MATCH SET n.Septicemia= rec[2], n.Malignantneoplasms= rec[3], n.Diabetesmellitus=
```

```
rec[4], n.Alzheimerdisease= rec[5], n.Influenzaandpneumonia= rec[6],
```

```
n.Chroniclowerrespiratorydiseases= rec[7], n.Otherdiseasesofrespiratorysystem= rec[8],
```

```
n.Nephritisnephroticsyndromeandnephrosis= rec[9],
```

```
n.Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified= rec[10],
```

```
n.Diseasesofheart= rec[11],
```

```
n.Cerebrovasculardiseases= rec[12], n.COVID19MultipleCause= rec[13],
```

```
n.COVID19UnderlyingCause= rec[14]
```

- Διαγραφή του έξτρα κόμβου που είχε δημιουργηθεί αλλά δεν ήταν χρήσιμος
match (n) where id(n)=88 delete n

- Για την απάντηση του πρώτου ερωτήματος

```
Query: MATCH (d:Diseases)
```

```
RETURN max(d.NaturalCause), max(d.Septicemia), max(d.Malignantneoplasms),
```

```
max(d.Diabetesmellitus), max(d.Alzheimerdisease), max(d.Influenzaandpneumonia),
```

```
max(d.Chroniclowerrespiratorydiseases), max(d.Otherdiseasesofrespiratorysystem),
```

```
max(d.Nephritisnephroticsyndromeandnephrosis),
```

```
max(d.Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified),
```

```
max(d.Diseasesofheart), max(d.Cerebrovasculardiseases)
```

- Για την απάντηση του δεύτερου ερωτήματος

Query: MATCH (d:Diseases)

RETURN max(d.Septicemia), max(d.Malignantneoplasms), max(d.Diabetesmellitus),
max(d.Alzheimerdisease), max(d.Influenzaandpneumonia),
max(d.Chroniclowerrespiratorydiseases), max(d.Otherdiseasesofrespiratorysystem),
max(d.Nephritisnephroticsyndromeandnephrosis),
max(d.Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified),
max(d.Diseasesofheart), max(d.Cerebrovascular diseases)

- Για την απάντηση του τρίτου ερωτήματος

Query: MATCH (d:Diseases)

RETURN max(d.Septicemia), max(d.Malignantneoplasms), max(d.Diabetesmellitus),
max(d.Alzheimerdisease), max(d.Influenzaandpneumonia),
max(d.Chroniclowerrespiratorydiseases), max(d.Otherdiseasesofrespiratorysystem),
max(d.Nephritisnephroticsyndromeandnephrosis),
max(d.Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified),
max(d.Cerebrovascular diseases)

- Για την απάντηση του τέταρτου ερωτήματος

Query: MATCH (d:Diseases)

RETURN max(d.COVID19UnderlyingCause), max(d.COVID19MultipleCause)

- Για την απάντηση του πέμπτου ερωτήματος

Query: MATCH (p:Patient)-[:DIED]->(d:Diseases), (d)-[:DATE]->(da:Dates) WHERE p.Sex ="Male"
RETURN p, d, da ;

- Για την απάντηση του έκτου ερωτήματος

Query1: MATCH (n:Patient) WHERE (n.Sex='Male') RETURN count(n) as count

Query2: MATCH (n:Patient) WHERE (n.Sex='Female') RETURN count(n) as count

Query3: MATCH (n:Patient) WHERE (n.Sex='All Sexes') RETURN count(n) as count

➤ Για την απάντηση του έβδομου ερωτήματος

Query1: MATCH (n:Patient) WHERE (n.State='United States') RETURN count(n) as count

Query2: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='All') RETURN count(n) as count

Query3: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Hispanic') RETURN count(n) as count

Query4: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic American Indian or Alaska Native') RETURN count(n) as count

Query5: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic Asian') RETURN count(n) as count

Query6: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic Black') RETURN count(n) as count

Query7: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic Native Hawaiian or Other Pacific Islander') RETURN count(n) as count

Query8: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic White') RETURN count(n) as count

➤ Για την απάντηση του όγδοου ερωτήματος

Query1: MATCH (n:Patient) WHERE (n.AgeGroup='All Ages') RETURN count(n) as count

Query2: MATCH (n:Patient) WHERE (n.AgeGroup='Under 1 year') RETURN count(n) as count

Query3: MATCH (n:Patient) WHERE (n.AgeGroup='First 17 years') RETURN count(n) as count

Query4: MATCH (n:Patient) WHERE (n.AgeGroup='1-4 years') RETURN count(n) as count

Query5: MATCH (n:Patient) WHERE (n.AgeGroup='5-14 years') RETURN count(n) as count

Query6: MATCH (n:Patient) WHERE (n.AgeGroup='15-24 years') RETURN count(n) as count

Query7: MATCH (n:Patient) WHERE (n.AgeGroup='18-29 years') RETURN count(n) as count

Query8: MATCH (n:Patient) WHERE (n.AgeGroup='30-39 years') RETURN count(n) as count

Query9: MATCH (n:Patient) WHERE (n.AgeGroup='35-44 years') RETURN count(n) as count

Query10: MATCH (n:Patient) WHERE (n.AgeGroup='40-49 years') RETURN count(n) as count

Query11: MATCH (n:Patient) WHERE (n.AgeGroup='45-54 years') RETURN count(n) as count

Query12: MATCH (n:Patient) WHERE (n.AgeGroup='50-64 years') RETURN count(n) as count

Query13: MATCH (n:Patient) WHERE (n.AgeGroup='55-64 years') RETURN count(n) as count

Query14: MATCH (n:Patient) WHERE (n.AgeGroup='65-74 years') RETURN count(n) as count

Query15: MATCH (n:Patient) WHERE (n.AgeGroup='75-84 years') RETURN count(n) as count

Query16: MATCH (n:Patient) WHERE (n.AgeGroup='85 years and over years') RETURN count(n) as count

Query17: MATCH (n:Patient) WHERE (n.AgeGroup='25-34 years') RETURN count(n) as count

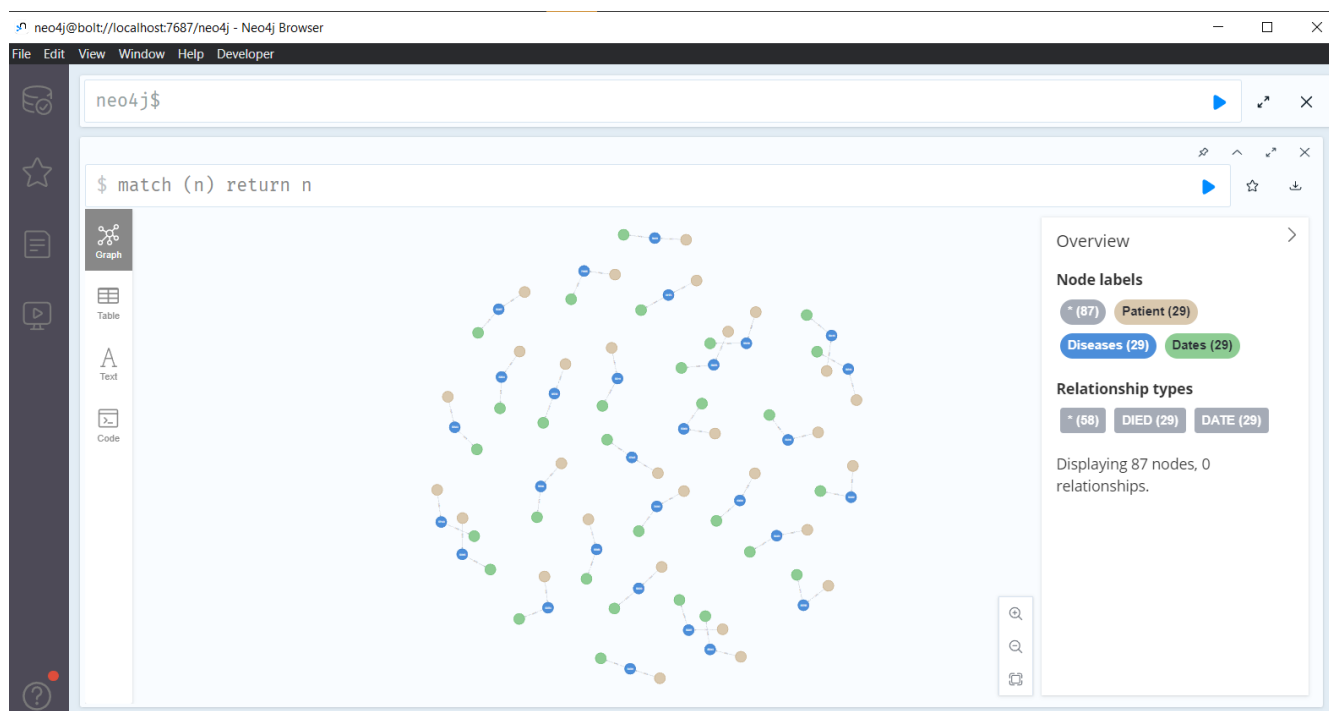
4.4.2 Προβλήματα υλοποίησης

Η διαδικασία κατασκευής του Γράφου Γνώσης βγήκε εις πέρας, ακολουθώντας κάποια βήματα, των οποίων η περιγραφή αναφέρεται παραπάνω. Ωστόσο, κατά την διάρκεια κατασκευής εμφανίστηκαν κάποια προβλήματα. Κάποια από αυτά ήταν:

1. Πρόβλημα ανάθεσης τιμών σε κάποια από τα χαρακτηριστικά των κόμβων, εξαιτίας χρήσης δεδομένων που ήταν αποθηκευμένα σε 3 διαφορετικά excel και
2. Πρόβλημα δημιουργίας σχέσεων μεταξύ των κόμβων, επίσης, λόγω χρήσης 3 διαφορετικών excel.

4.4.3 Τελική μορφή του Γράφου Γνώσης

Η υλοποίηση του Γράφου Γνώσης ολοκληρώθηκε μετά την ακολουθία συγκεκριμένων βημάτων και την επίλυση των προβλημάτων που παρουσιάστηκαν. Ο Γράφος αυτός έχει 87 κόμβους και 58 σχέσεις. Οι κόμβοι του χωρίζονται σε 3 labels, το Patient, το Diseases και το Dates με 29 κόμβους το κάθε ένα. Επιπλέον, οι σχέσεις τους χωρίζονται σε 2 κατηγορίες την DIED και την DATE με 29 συνδέσεις το κάθε ένα. Παρακάτω απεικονίζεται ο τελικός ολοκληρωμένος Γράφος Γνώσης της μελέτης, στον οποίο χρησιμοποιήθηκαν τα δεδομένα και βγήκαν συμπεράσματα.



Εικόνα 30 Στη συγκεκριμένη εικόνα παρουσιάζεται ο Γράφος Γνώσης που κατασκευάστηκε για την μελέτη.

4.2 Τα Δεδομένα

Η έρευνα ασχολήθηκε με την μελέτη των τρόπων εμφάνισης των κρουσμάτων Covid-19 και τα χαρακτηριστικά των ασθενών, που είχαν νοσήσει. Αρχικά, η γεωγραφική τοποθεσία μπορεί να δώσει κάποιο στοιχείο σχετικά με την νόσο ή με τους ανθρώπους τους οποίους μένουν εκεί. Μπορεί για παράδειγμα, να αποδειχθεί ότι η συγκεκριμένη περιοχή είναι λογικό να έχει τόσα πολλά κρούσματα, γιατί σε εκείνην ταξιδεύουν άνθρωποι από διαφορετικές χώρες. Επίσης, μπορεί οι άνθρωποι σε εκείνην την περιοχή να είναι ευάλωτοι και απέναντι σε άλλες ασθένειες και για τον λόγο αυτό να νοσούν πιο εύκολα από ανθρώπους άλλης περιοχής. Με τέτοιες παραμέτρους, η έρευνα γίνεται περισσότερο χρήσιμη για την εξαγωγή συμπερασμάτων και πιο αποδοτική για την εύρεση λύσεων. Συγκεκριμένα, χρησιμοποιήθηκαν δεδομένα που συλλέχθηκαν από τη κατηγορία Health Statistics, του CDC DATASET CATALOG: DATA BY TOPIC:

- Προσωρινοί θάνατοι από COVID-19 ανά φύλο και ηλικία,
- Προσωρινοί θάνατοι από COVID-19: Εστίαση σε ηλικίες 0-18 ετών και
- Εβδομαδιαίες προσωρινές καταμετρήσεις θανάτων ανά πολιτεία και επιλεγμένες αιτίες, 2020-2023.

Επιλέχθηκαν τα παραπάνω σύνολα δεδομένων προκειμένου να υπάρχει μία σφαιρική εικόνα των ασθενών και των τόπων που ζουν. Αναλυτικότερα, οι παράμετροι για τους ανθρώπους ήταν η καταγωγή, μελετήθηκαν οι περιπτώσεις ασθενών που είχαν Ισπανική ή Αμερικανική καταγωγή ή καταγωγή από φυλές όπως Ιθαγενείς Χαβάης ή νησιώτες του Ειρηνικού, Αμερικανός Ινδιάνος ή Ιθαγενής της Αλάσκας και Ασιάτες, η ηλικία που ήταν από βρέφος έως και ηλικιωμένος και το φύλο. Έπειτα, η εξέταση του ιατρικού ιστορικού των ασθενών ήταν πολύ σημαντική. Έτσι, οι υπόλοιπες αρρώστιες που μελετήθηκαν εκτός από τον Covid-19, ήταν από φυσικά αίτια, Σηψαιμία, Κακοήθες νεόπλασμα, Σακχαρώδη διαβήτη, νόσος του Αλτσχάιμερ, ταυτόχρονη εμφάνιση γρίπης και πνευμονίας, χρόνιες αναπνευστικές παθήσεις, άλλες ασθένειες του αναπνευστικού συστήματος, νέφρωση, μη φυσιολογικά κλινικά και εργαστηριακά ευρήματα, παθήσεις της καρδιάς και εγκεφαλοαγγειακές παθήσεις. Τέλος, οι ημερομηνίες καταγραφής των κρουσμάτων αποτέλεσαν βασικό στοιχείο για την εξέταση των στοιχείων. Επομένως, έγινε αναζήτηση και καταγραφή της εβδομάδας έναρξης καταγραφής των κρουσμάτων, της εβδομάδας τερματισμού καταγραφής των κρουσμάτων, της επιδημιολογικής εβδομάδας (MMWRWeek) και του επιδημιολογικού έτους (MMWRYear). Περισσότερες λεπτομέρειες για τα δεδομένα αναφέρονται στην υπό ενότητα 4.1

4.3 Ερωτήματα προς την βάση δεδομένων (Queries)

Μετά τον ολοκληρωμένο σχεδιασμό του Γράφου γνώσης και της αποθήκευσης των δεδομένων σε αυτόν, έπρεπε να δημιουργηθούν Queries, ώστε να διεξαχθούν τα συμπεράσματα. Τα Queries, λοιπόν, που χρησιμοποιήθηκαν είναι:

- Ποια ασθένεια είναι αυτή που σε συνδυασμό με τον Covid-19 προκαλεί τους περισσότερους θανάτους σε ασθενείς;

Απάντηση: Οι περισσότεροι ασθενείς που κατέληξαν φαίνεται πως νοσούσαν μόνο από τον Covid-19 και πως ήταν υγιείς και δεν είχαν κάποια άλλη ασθένεια, αφού ο συνολικός αριθμός θανάτων αυτής της κατηγορίας ασθενών ήταν 74.012 και ο συνολικός αριθμός θανάτων όλων των περιπτώσεων ήταν 137.380.

Query: MATCH (d:Diseases)

```

RETURN max(d.NaturalCause), max(d.Septicemia), max(d.Malignantneoplasms),
max(d.Diabetesmellitus), max(d.Alzheimerdisease), max(d.Influenzaandpneumonia),
max(d.Chroniclowerrespiratorydiseases), max(d.Otherdiseasesofrespiratorysystem),
max(d.Nephritisnephroticsyndromeandnephrosis),
max(d.Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified),
max(d.Diseasesofheart), max(d.Cerebrovasculardiseases)

```



Εικόνα 31 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, πως οι περισσότεροι ασθενείς που κατέληξαν νοσούσαν μόνο από τον Covid-19 και πως ήταν υγιείς.

Επεξήγηση: Προκυμμένον να βρεθεί ποια ασθένεια σε συνδυασμό με τον Covid-19 επιβάρυνε πολύ την υγεία ενός ασθενούς, χρησιμοποιήθηκε το παραπάνω query, το οποίο μετράει ξεχωριστά τον συνολικό αριθμό κρουσμάτων για όλες τις περιπτώσεις θανάτων από τις ασθένειες που μελετήθηκαν στην έρευνα, εκτός από τις δύο περιπτώσεις του Covid-19 Multiple Cause και του Covid-19 Underlying Cause, και τον εμφανίζει. Έτσι, φαίνεται ποια είναι η πιο επικίνδυνη ασθένεια να έχει κάποιος ασθενείς που νοσεί ήδη από τον Covid-19.

- Ποια είναι η πρώτη κατά σειρά ασθένεια που σε συνδυασμό με τον Covid-19 που προκαλεί τους περισσότερους θανάτους σε ασθενείς;

Απάντηση: Οι περισσότεροι θάνατοι ασθενών που νοσούν ήδη από τον Covid-19, φαίνεται πως είναι εκείνοι που έχουν ασθένειες που αφορούν την καρδιά, αφού ο συνολικός αριθμός

θανάτων αυτής της κατηγορίας ασθενών ήταν 15.766 και ο συνολικός αριθμός θανάτων όλων των περιπτώσεων ήταν 137.380.

Query: MATCH (d:Diseases)

RETURN max(d.Septicemia), max(d.Malignantneoplasms), max(d.Diabetesmellitus),
max(d.Alzheimerdisease), max(d.Influenzaandpneumonia),
max(d.Chroniclowerrespiratorydiseases), max(d.Otherdiseasesofrespiratorysystem),
max(d.Nephritisnephroticsyndromeandnephrosis),
max(d.Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified),
max(d.Diseasesofheart), max(d.Cerebrovascular diseases)



Εικόνα 32 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, πως όταν οι ασθενείς νοσούν από τον Covid-19, είναι πολύ επικίνδυνο αν νοσούν και από ασθένειες της καρδιάς.

Επεξήγηση: Για να βρεθεί ποια από όλες τις ασθένειες σε συνδυασμό με τον Covid-19 επιβά-
ρυνε πολύ την υγεία ενός ασθενούς, χρησιμοποιήθηκε το παραπάνω query, το οποίο μετράει
ξεχωριστά τον συνολικό αριθμό κρουσμάτων για όλες τις περιπτώσεις θανάτων από τις ασθέ-
νειες που μελετήθηκαν στην έρευνα, εκτός από τις δύο περιπτώσεις του Covid-19 Multiple
Cause, του Covid-19 Underlying Cause και των φυσικών αιτιών θανάτου, και τον εμφανίζει.
Έτσι, φαίνεται ποια είναι η πιο επικίνδυνη ασθένεια να έχει κάποιος ασθενείς που νοσεί ήδη
από τον Covid-19.

- Ποια είναι η ασθένεια που σε συνδυασμό με τον Covid-19 προκαλεί τους λιγότερους θανάτους σε ασθενείς;

Απάντηση: Τα λιγότερα κρούσματα θανάτων που αφορούν ασθενείς, που νοσούν ήδη από τον Covid-19, φαίνεται πως τα έχουν ασθενείς που εμφανίζουν συμπτώματα και έχουν μη φυσιολογικά κλινικά και εργαστηριακά ευρήματα, αφού ο συνολικός αριθμός θανάτων αυτής της κατηγορίας ασθενών ήταν 689 και ο συνολικός αριθμός θανάτων όλων των περιπτώσεων ήταν 137.380.

Query: MATCH (d:Diseases)

```
RETURN max(d.Septicemia), max(d.Malignantneoplasms), max(d.Diabetesmellitus),
max(d.Alzheimerdisease), max(d.Influenzaandpneumonia),
max(d.Chroniclowerrespiratorydiseases), max(d.Otherdiseasesofrespiratorysystem),
max(d.Nephritisnephroticsyndromeandnephrosis),
max(d.Symptomssignsandabnormalclinicalandlaboratoryfindingsnotelsewhereclassified),
max(d.Cerebrovascular diseases)
```



Εικόνα 33 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, πως όταν οι ασθενείς νοσούν από τον Covid-19, διατρέχουν λιγότερο κίνδυνο αν εμφανίζουν συμπτώματα και έχουν μη φυσιολογικά κλινικά και εργαστηριακά ευρήματα από ότι αν έχουν κάποια άλλη ασθένεια.

Επεξήγηση: Για να βρεθεί ποια από όλες τις ασθένειες σε συνδυασμό με τον Covid-19 επιβά-
ρυνε λιγότερο την υγεία ενός ασθενούς, χρησιμοποιήθηκε το παραπάνω query, το οποίο με-
τράει ξεχωριστά τον συνολικό αριθμό κρουσμάτων για όλες τις περιπτώσεις θανάτων από τις
ασθένειες που μελετήθηκαν στην έρευνα, εκτός από τις δύο περιπτώσεις του Covid-19 Multi-
ple Cause, του Covid-19 Underlying Cause, των φυσικών αιτιών θανάτου και των ασθενειών
της καρδιάς, και τον εμφανίζει. Έτσι, φαίνεται ποια είναι η λιγότερο επικίνδυνη ασθένεια να
έχει κάποιος ασθενείς που νοσεί ήδη από τον Covid-19.

- Τα περισσότερα κρούσματα προέρχονταν από τον Covid-19 Multiple Cause ή τον Covid-19 Underlying Cause;

Απάντηση: Τα περισσότερα κρούσματα θανάτων έχουν προέλθει από τον Covid-19 Underlying Cause, αφού ο συνολικός αριθμός θανάτων αυτής της κατηγορίας ασθενών ήταν 9.638, ενώ ο συνολικός αριθμός θανάτων από την περίπτωση Covid-19 Multiple Cause ήταν 9.239.

Query: MATCH (d:Diseases)

RETURN max(d.COVID19UnderlyingCause), max(d.COVID19MultipleCause)



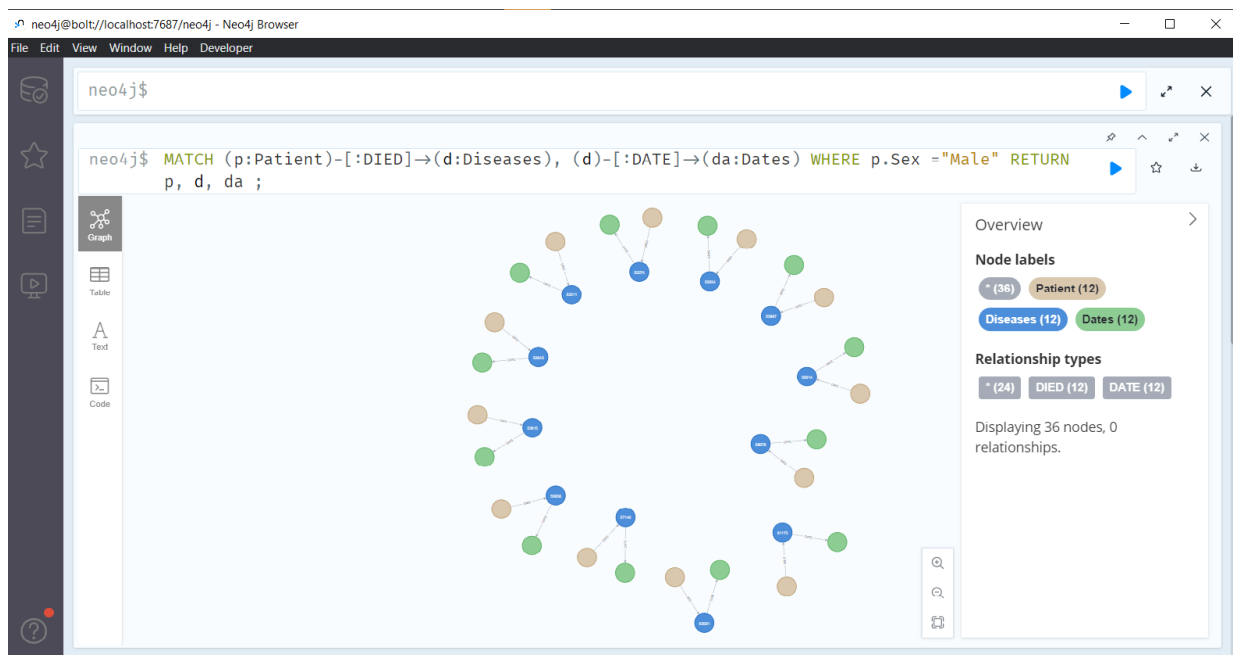
Εικόνα 34 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, πως τα περισσότερα κρούσματα θανάτων μεταξύ του Covid-19 Multiple Cause και του Covid-19 Underlying Cause, έχουν προέλθει από τον Covid-19 Underlying Cause.

Επεξήγηση: Για να βρεθεί ποια ήταν η πιο επικίνδυνη νόσος από τις δύο περιπτώσεις Covid-19, χρησιμοποιήθηκε το παραπάνω query, το οποίο μετράει ξεχωριστά και για τις δύο περιπτώσεις τον συνολικό αριθμό κρουσμάτων και τον εμφανίζει.

- Πώς μπορούμε να αντλήσουμε πληροφορίες από τον Γράφο Γνώσης; Για παράδειγμα, από ποιες ασθένειες νόσησαν ασθενείς που το φύλο τους ήταν αρσενικό και ποια άλλα χαρακτηριστικά είχαν αυτοί οι ασθενείς;

Απάντηση: Με το παρακάτω query μπορούμε να πάρουμε πληροφορίες για όλους τους ασθενείς που το φύλο τους είναι αρσενικό. Αναλυτικότερα, μπορούμε να πάρουμε πληροφορίες για την ηλικία, τον τόπο καταγωγής, όλες τις ασθένειες από τις οποίες έχει νοσήσει, καθώς και τις ημερομηνίες καταγραφής των συμβάντων.

Query: MATCH (p:Patient)-[:DIED]->(d:Diseases), (d)-[:DATE]->(da:Dates) WHERE p.Sex = "Male" RETURN p, d, da ;



Εικόνα 35 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, όλες οι πληροφορίες των ασθενών που το φύλο τους είναι αρσενικό.

Επεξήγηση: Για να βρεθούν όλες οι πληροφορίες των ασθενών που το φύλο τους είναι αρσενικό, χρησιμοποιήθηκε το παραπάνω query, το οποίο μέσω των σχέσεων μεταξύ των κόμβων καταφέρνει να απομονώσει τις περιπτώσεις όπου Sex = 'Male' και να τις εμφανίσει. Με αυτόν τον τρόπο είναι δυνατή η εμφάνιση μεμονωμένων περιπτώσεων με άλλους παράγοντες, για

παράδειγμα, αν θέλουμε όλες τις πληροφορίες των ασθενών που νόσησαν από ασθένειες του πνεύμονα.

- Ποιο φύλο είχε τους περισσότερους θανάτους από Covid-19;

Απάντηση: Τους περισσότερους θανάτους τους είχαν ασθενείς που το φύλο τους ήταν αρσενικό με συνολικό αριθμό θανάτων να είναι 29.

Query1: MATCH (n:Patient) WHERE (n.Sex='Male') RETURN count(n) as count



Εικόνα 36 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά μόνο τους ασθενείς που το φύλο τους ήταν αρσενικό.

Query2: MATCH (n:Patient) WHERE (n.Sex='Female') RETURN count(n) as count



Εικόνα 37 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά μόνο τους ασθενείς που το φύλο τους ήταν θηλυκό.

Query3: MATCH (n:Patient) WHERE (n.Sex='All Sexes') RETURN count(n) as count



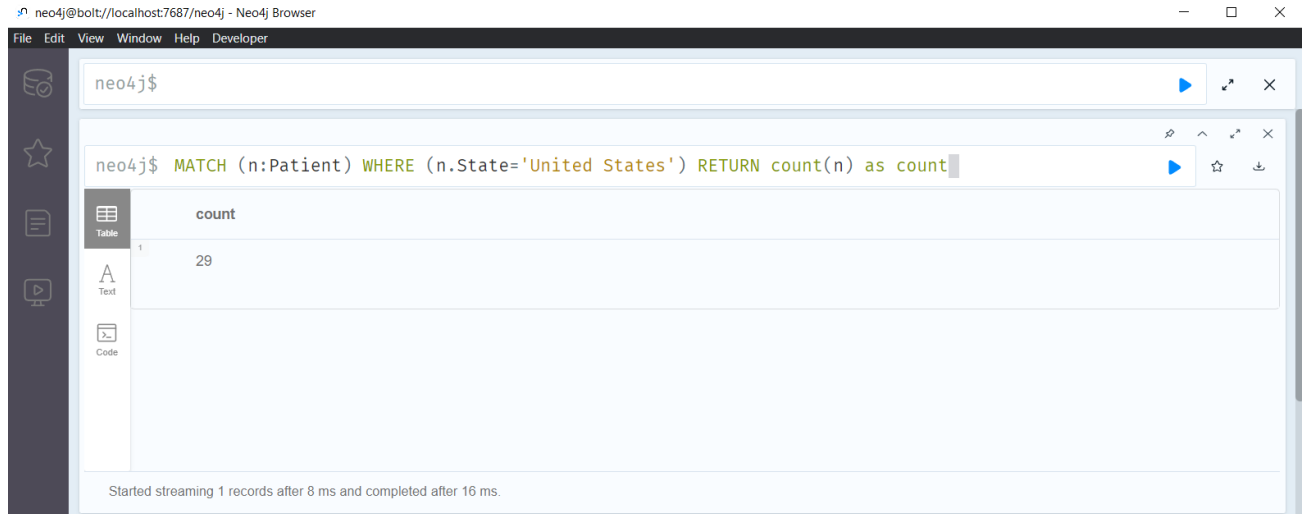
Εικόνα 38 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς και των δύο φύλων.

Επεξήγηση: Για να βρεθεί ποιο από τα δύο φύλα κατέληξε περισσότερες φορές από τον Covid-19, έπρεπε πρώτα να υπολογιστούν ξεχωριστά τρεις αριθμοί θανάτων των ασθενών. Αυτοί οι αριθμοί ήταν ο συνολικός αριθμός θανάτων όταν το φύλο του ασθενούς είναι μόνο αρσενικό, ο συνολικός αριθμός θανάτων όταν το φύλο του ασθενούς είναι μόνο θηλυκό και ο συνολικός αριθμός θανάτων σε περιπτώσεις όπου υπάρχουν και ασθενείς που το φύλο τους είναι αρσενικό και ασθενείς που το φύλο τους είναι θηλυκό. Αυτή η διαδικασία έπρεπε να γίνει, καθώς τα δεδομένα που αποθηκεύτηκαν στον Γράφο Γνώσης ήταν της μορφής 'Male' και 'All Sexes'. Έτσι, με το πρώτο query, το οποίο αφορούσε ασθενείς με το φύλο να είναι αρσενικό, εμφανίζει το αποτέλεσμα 12. Έπειτα, με το δεύτερο query, το οποίο αφορούσε ασθενείς με το φύλο να είναι θηλυκό, εμφανίζει το αποτέλεσμα 0, που σημαίνει ότι δεν υπήρχαν περιπτώσεις θανάτων στις οποίες ο ασθενείς να έχει μόνο θηλυκό φύλο. Τέλος, με το τρίτο query, το οποίο αφορούσε ασθενείς και από τα δύο φύλα, εμφανίζει το αποτέλεσμα 17. Συμπερασματικά, έχουμε $12+17=29$ ασθενείς που το φύλο τους ήταν αρσενικό και $0+17=17$ ασθενείς που το φύλο τους ήταν θηλυκό. Επομένως, τους περισσότερους θανάτους τους είχαν ασθενείς που το φύλο τους ήταν αρσενικό.

- Ποια περιοχή είχε τους περισσότερους θανάτους από Covid-19;

Απάντηση: Τους περισσότερους θανάτους από Covid-19 παρουσίασαν ασθενείς που κατοικούσαν στις Ηνωμένες Πολιτείες με συνολικό αριθμό θανάτων να είναι 29.

Query1: MATCH (n:Patient) WHERE (n.State='United States') RETURN count(n) as count



Εικόνα 39 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς που κατοικούν στις Ηνωμένες Πολιτείες.

Query2: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='All') RETURN count(n) as count



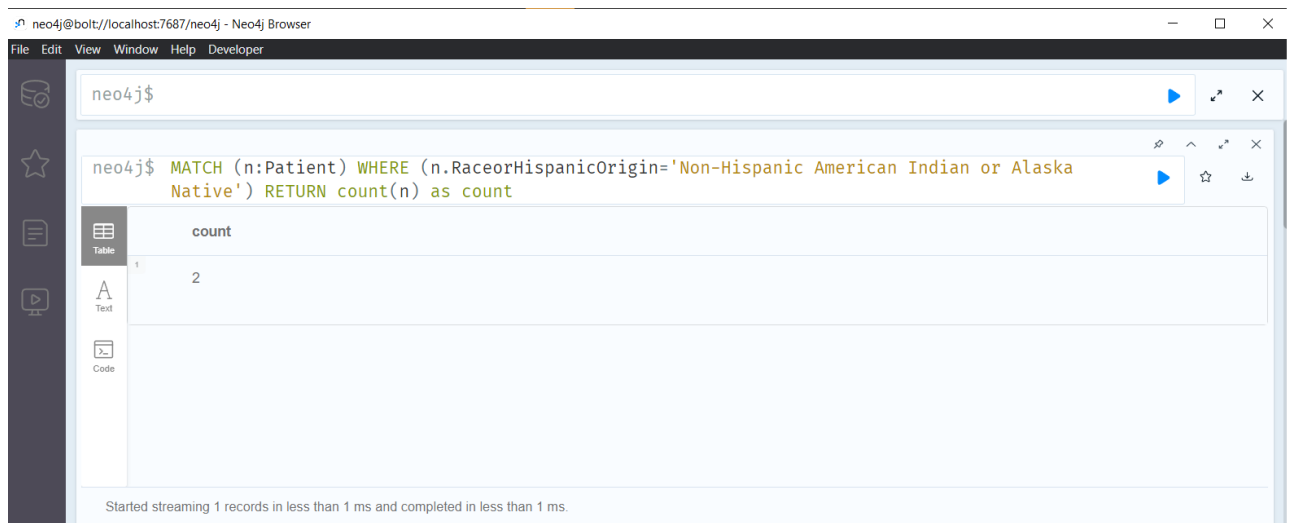
Εικόνα 40 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς που κατοικούν σε όλες τις ηπείρους.

Query3: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Hispanic') RETURN count(n) as count



Εικόνα 41 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς που κατοικούν στην Ισπανία.

Query4: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic American Indian or Alaska Native') RETURN count(n) as count



Εικόνα 42 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς που μένουν στην Αμερική, στην Ινδία ή στην Αλάσκα.

Query5: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic Asian') RETURN count(n) as count



Εικόνα 43 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς που κατοικούν στην Ασία.

Query6: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic Black') RETURN count(n) as count



Εικόνα 44 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς που ανήκουν στην μαύρη φυλή.

Query7: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic Native Hawaiian or Other Pacific Islander') RETURN count(n) as count



Εικόνα 45 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς που μένουν στην Χαβάη ή γενικά νησιώτες του Ειρηνικού Ωκεανού.

Query8: MATCH (n:Patient) WHERE (n.RaceorHispanicOrigin='Non-Hispanic White') RETURN count(n) as count



Εικόνα 46 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς που ανήκουν στην λευκή φυλή.

Επεξήγηση: Για να βρεθεί ποια περιοχή είχε τους περισσότερους θανάτους από Covid-19, έπρεπε πρώτα να υπολογιστούν ξεχωριστά οι αριθμοί θανάτων των ασθενών στις περιοχές που μελετήθηκαν. Αυτοί οι αριθμοί ήταν ο συνολικός αριθμός θανάτων ασθενών που μένουν στις Ηνωμένες Πολιτείες που είναι 29 περιπτώσεις, ο συνολικός αριθμός θανάτων ασθενών που κατοικούν σε όλες τις ηπείρους που είναι 11 περιπτώσεις, ο συνολικός αριθμός θανάτων ασθενών που μένουν στην Ισπανία που είναι 2 περιπτώσεις, ο συνολικός αριθμός θανάτων

ασθενών που μένουν στην Αμερική, στην Ινδία ή στην Αλάσκα που είναι 2 περιπτώσεις, ο συνολικός αριθμός θανάτων ασθενών που μένουν στην Ασία που είναι 2 περιπτώσεις, ο συνολικός αριθμός θανάτων ασθενών που ανήκουν στην μαύρη φυλή που είναι 2 περιπτώσεις, ο συνολικός αριθμός θανάτων ασθενών που μένουν στην Χαβάη ή γενικά νησιώτες του Ειρηνικού Ωκεανού που είναι 2 περιπτώσεις και ο συνολικός αριθμός θανάτων ασθενών που ανήκουν στην λευκή φυλή που είναι 2 περιπτώσεις. Συμπερασματικά, ο μεγαλύτερος συνολικός αριθμός θανάτων από Covid-19 είναι ο 29 και αντιστοιχεί στους ασθενείς που κατοικούσαν στις Ηνωμένες Πολιτείες.

- Ποια ηλικία πρόσβαλε περισσότερο ο Covid-19;

Απάντηση: Μέσα από την μελέτη των δεδομένων που χρησιμοποιήθηκαν φαίνεται πως κινδυνεύουν περισσότερο όσοι ασθενείς είναι νεότεροι από 1 έτους ή η ηλικία τους είναι μέχρι τα 54 έτη.

Query1: MATCH (n:Patient) WHERE (n.AgeGroup='All Ages') RETURN count(n) as count



Εικόνα 47 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς όλων των ηλικιών.

Query2: MATCH (n:Patient) WHERE (n.AgeGroup='Under 1 year') RETURN count(n) as count



Εικόνα 48 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι κάτω του ενός έτους.

Query3: MATCH (n:Patient) WHERE (n.AgeGroup='First 17 years') RETURN count(n) as count



Εικόνα 49 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μέχρι το 17ο έτος.

Query4: MATCH (n:Patient) WHERE (n.AgeGroup='1-4 years') RETURN count(n) as count



Εικόνα 50 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 1-4 ετών.

Query5: MATCH (n:Patient) WHERE (n.AgeGroup='5-14 years') RETURN count(n) as count



Εικόνα 51 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 5-14 ετών.

Query6: MATCH (n:Patient) WHERE (n.AgeGroup='15-24 years') RETURN count(n) as count



Εικόνα 52 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 15-24 ετών.

Query7: MATCH (n:Patient) WHERE (n.AgeGroup='18-29 years') RETURN count(n) as count



Εικόνα 53 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 18-29 ετών.

Query8: MATCH (n:Patient) WHERE (n.AgeGroup='30-39 years') RETURN count(n) as count



Εικόνα 54 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 30-39 ετών.

Query9: MATCH (n:Patient) WHERE (n.AgeGroup='35-44 years') RETURN count(n) as count



Εικόνα 55 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 35-44 ετών.

Query10: MATCH (n:Patient) WHERE (n.AgeGroup='40-49 years') RETURN count(n) as count



Εικόνα 56 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 40-49 ετών.

Query11: MATCH (n:Patient) WHERE (n.AgeGroup='45-54 years') RETURN count(n) as count



Εικόνα 57 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 45-54 ετών.

Query12: MATCH (n:Patient) WHERE (n.AgeGroup='50-64 years') RETURN count(n) as count



Εικόνα 58 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 50-64 ετών.

Query13: MATCH (n:Patient) WHERE (n.AgeGroup='55-64 years') RETURN count(n) as count



Εικόνα 59 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 55-64 ετών.

Query14: MATCH (n:Patient) WHERE (n.AgeGroup='65-74 years') RETURN count(n) as count



Εικόνα 60 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 65-74 ετών.

Query15: MATCH (n:Patient) WHERE (n.AgeGroup='75-84 years') RETURN count(n) as count



Εικόνα 61 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 75-84 ετών.

Query16: MATCH (n:Patient) WHERE (n.AgeGroup='85 years and over years') RETURN count(n) as count



Εικόνα 62 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι από 85 ετών και πάνω.

Query17: MATCH (n:Patient) WHERE (n.AgeGroup='25-34 years') RETURN count(n) as count



Εικόνα 63 Στην παραπάνω εικόνα φαίνεται με την χρήση ενός query στο neo4j, ο συνολικός αριθμός θανάτων που αφορά ασθενείς των οποίων η ηλικία είναι μεταξύ 25-34 ετών.

Επεξήγηση: Για να βρεθεί ποια ηλικία πρόσβαλε περισσότερο ο Covid-19, έπρεπε πρώτα να υπολογιστούν ξεχωριστά οι συνολικοί αριθμοί θανάτων των ασθενών ανά ηλικιακά σύνολα. Τα ηλικιακά σύνολα που μελετήθηκαν ήταν κάτω από ενός έτους, τα πρώτα 17 χρόνια, ηλικία μεταξύ 1-4 ετών, ηλικία μεταξύ 18-29 ετών, ηλικία μεταξύ 30-39 ετών, ηλικία μεταξύ 35-44 ετών, ηλικία μεταξύ 40-49 ετών, ηλικία μεταξύ 45-54 ετών, ηλικία μεταξύ 50-64 ετών, ηλικία μεταξύ 55-64 ετών, ηλικία μεταξύ 65-74 ετών, ηλικία μεταξύ 75-84 ετών, ηλικία από 85 ετών και πάνω, ηλικία μεταξύ 25-34 ετών και ένα σύνολο που περιέχει περιπτώσεις από όλες τις

ηλικίες. Έτσι, μέσα από την μελέτη των δεδομένων που χρησιμοποιήθηκαν φαίνεται πως κινδυνεύουν περισσότερο όσοι ασθενείς είναι νεότεροι από 1 έτους ή η ηλικία τους είναι μέχρι τα 54 έτη. Έπειτα, κινδυνεύουν οι ασθενείς που ανήκουν στο ηλικιακό σύνολο των 55-64 ετών και τέλος, κινδυνεύουν λιγότερο οι ασθενείς που ανήκουν στο ηλικιακό σύνολο από 65 ετών και πάνω.

Στο σημείο αυτό, ολοκληρώνεται το κεφάλαιο 4^ο και παρακάτω στο κεφάλαιο 5^ο υπάρχουν τα συμπεράσματα τα οποία κατέληξε η ολοκλήρωση της έρευνας.

Κεφάλαιο 5^ο: Αποτελέσματα έρευνας

Εισαγωγή:

Με την δημιουργία του Γράφου Γνώσης και την απάντηση των ερωτημάτων (queries) συγκεντρώθηκαν κάποια αποτελέσματα. Οι ερωτήσεις και τα αποτελέσματα που διεξήχθησαν, φαίνονται στον παρακάτω πίνακα.

5.1 Πίνακας ερωτήσεων και αποτελεσμάτων έρευνας

	Ερωτήσεις	Αποτελέσματα
1 ^η	Ποια ασθένεια είναι αυτή που σε συνδυασμό με τον Covid-19 προκαλεί τους περισσότερους θανάτους σε ασθενείς;	Οι περισσότεροι ασθενείς που κατέληξαν φαίνεται πως νοσούσαν μόνο από τον Covid-19 και πως ήταν υγιείς και δεν είχαν κάποια άλλη ασθένεια, αφού ο συνολικός αριθμός θανάτων αυτής της κατηγορίας ασθενών ήταν 74.012 και ο συνολικός αριθμός θανάτων όλων των περιπτώσεων ήταν 137.380.
2 ^η	Ποια είναι η πρώτη κατά σειρά ασθένεια που σε συνδυασμό με τον Covid-19, που προκαλεί τους περισσότερους θανάτους σε ασθενείς;	Οι περισσότεροι θάνατοι ασθενών που νοσούν ήδη από τον Covid-19, φαίνεται πως είναι εκείνοι που έχουν ασθένειες που αφορούν την καρδιά, αφού ο συνολικός αριθμός θανάτων αυτής της κατηγορίας ασθενών ήταν 15.766 και ο συνολικός αριθμός θανάτων όλων των περιπτώσεων ήταν 137.380.
3 ^η	Ποια είναι η ασθένεια που σε συνδυασμό με τον Covid-19 προκαλεί τους λιγότερους	Τα λιγότερα κρούσματα θανάτων που αφορούν ασθενείς, που νοσούν ήδη από τον Covid-19, φαίνεται πως τα έχουν

	θανάτους σε ασθενείς;	ασθενείς που εμφανίζουν συμπτώματα και έχουν μη φυσιολογικά κλινικά και εργαστηριακά ευρήματα, αφού ο συνολικός αριθμός θανάτων αυτής της κατηγορίας ασθενών ήταν 689 και ο συνολικός αριθμός θανάτων όλων των περιπτώσεων ήταν 137.380.
4 ^η	Τα περισσότερα κρούσματα προέρχονταν από τον Covid-19 Multiple Cause ή τον Covid-19 Underlying Cause;	Τα περισσότερα κρούσματα θανάτων έχουν προέλθει από τον Covid-19 Underlying Cause, αφού ο συνολικός αριθμός θανάτων αυτής της κατηγορίας ασθενών ήταν 9.638, ενώ ο συνολικός αριθμός θανάτων από την περίπτωση Covid-19 Multiple Cause ήταν 9.239.
5 ^η	Πώς μπορούμε να αντλήσουμε πληροφορίες από τον Γράφο Γνώσης; Για παράδειγμα, από ποιες ασθένειες νόσησαν ασθενείς που το φύλο τους ήταν αρσενικό και ποια άλλα χαρακτηριστικά είχαν αυτοί οι ασθενείς;	Με το query, που υπάρχει στην υπό ενότητα 4.3 και απαντάει στην συγκεκριμένη ερώτηση, μπορούμε να πάρουμε πληροφορίες για όλους τους ασθενείς που το φύλο τους είναι αρσενικό. Αναλυτικότερα, μπορούμε να πάρουμε πληροφορίες για την ηλικία, τον τόπο καταγωγής, όλες τις ασθένειες από τις οποίες έχει νοσήσει, καθώς και τις ημερομηνίες καταγραφής των συμβάντων.
6 ^η	Ποιο φύλο είχε τους περισσότερους θανάτους από Covid-19;	Τους περισσότερους θανάτους τους είχαν ασθενείς που το φύλο τους ήταν αρσενικό.

7 ^η	Ποια περιοχή είχε τους περισσότερους θανάτους από Covid-19;	Τους περισσότερους θανάτους από Covid-19 παρουσίασαν ασθενείς που κατοικούσαν στις Ηνωμένες Πολιτείες.
8 ^η	Ποια ηλικία πρόσβαλε περισσότερο ο Covid-19;	Μέσα από την μελέτη των δεδομένων που χρησιμοποιήθηκαν φαίνεται πως κινδυνεύουν περισσότερο όσοι ασθενείς είναι νεότεροι από 1 έτους ή η ηλικία τους είναι μέχρι τα 54 έτη.

Μέσα από τα αποτελέσματα αυτά καταλήγουμε σε σημαντικά συμπεράσματα, τα οποία εξηγούνται στο 6^ο κεφάλαιο.

Κεφάλαιο 6^ο: Συμπεράσματα και ιδέες για μελλοντική έρευνα

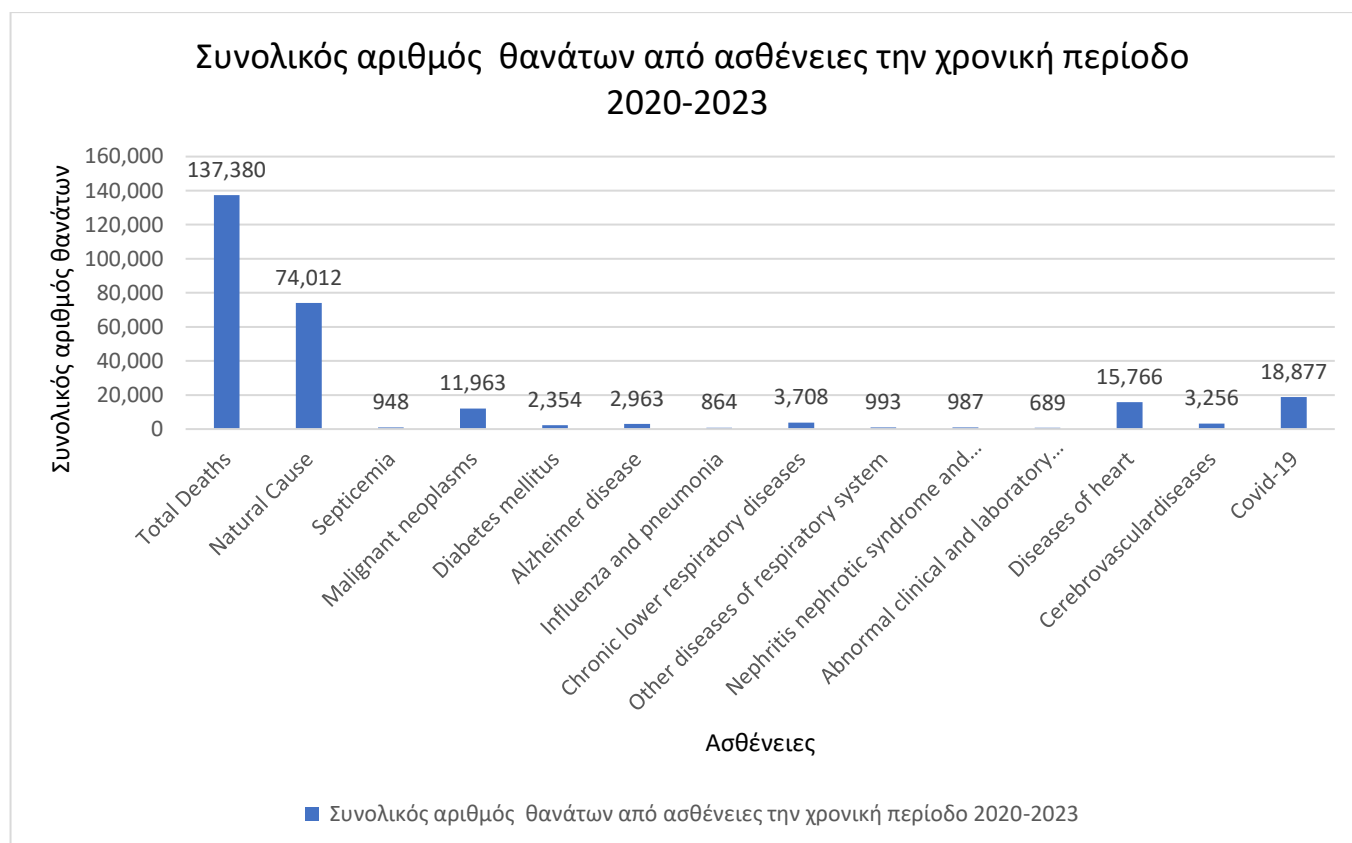
6.1 Συμπεράσματα

Τα συμπεράσματα που μπορούμε να αποκομίσουμε από τα αποτελέσματα που αναφέρθηκαν στο 5^ο κεφάλαιο είναι πολλά. Τα συμπεράσματα αυτά είναι:

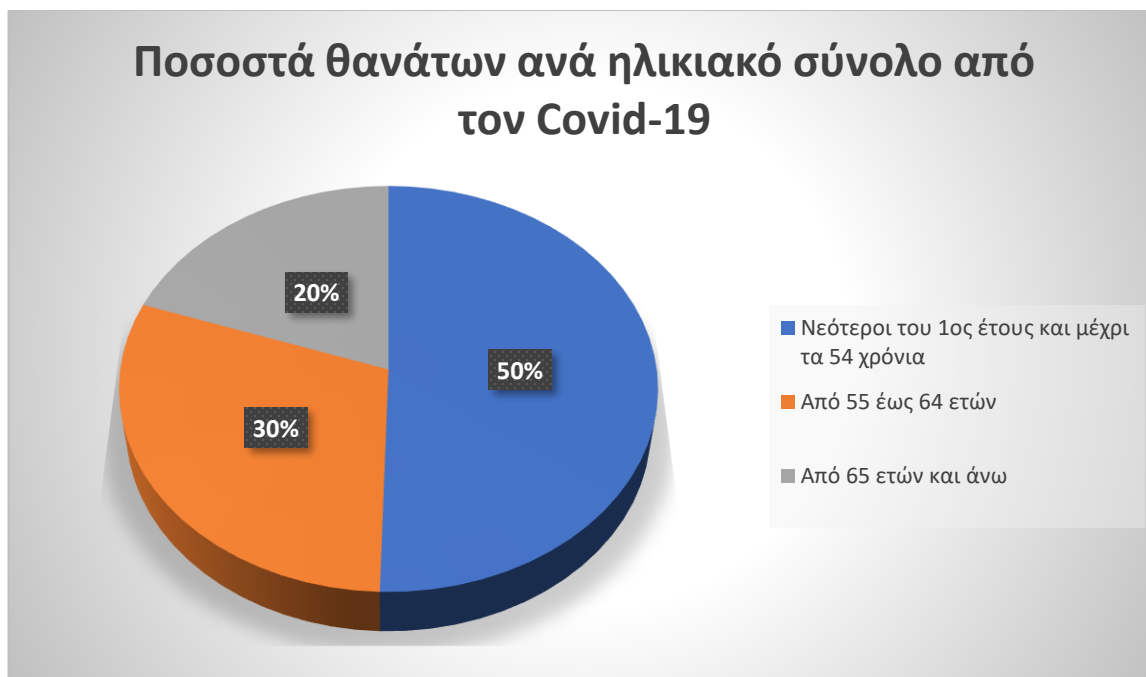
- I. Την χρονική περίοδο 2020-2023 και με βάση τα δεδομένα που μελετήθηκαν, ο συνολικός αριθμός θανάτων ήταν 137.380.
- II. Οι περισσότεροι ασθενείς που νοσούσαν από τον Covid-19 και δεν είχαν κάποια άλλη ασθένεια, είχαν τους περισσότερους θανάτους από τους ασθενείς, που είχαν και άλλες ασθένειες. Ο ακριβής συνολικός αριθμός θανάτων αυτής της περίπτωσης ήταν 74.012.
- III. Οι ασθενείς οι οποίοι είχαν ήδη ασθένειες της καρδιάς και μετά νοσούσαν από τον Covid-19, φαίνεται πως είχαν τους περισσότερους θανάτους σε σύγκριση με ασθενείς που είχαν Covid-19 και κάποια άλλη ασθένεια. Ο ακριβής συνολικός αριθμός θανάτων αυτής της κατηγορίας ήταν 15.766.
- IV. Οι ασθενείς οι οποίοι εμφάνιζαν ήδη συμπτώματα και είχαν μη φυσιολογικά κλινικά και εργαστηριακά ευρήματα και μετά νοσούσαν από τον Covid-19, φαίνεται πως είχαν τους λιγότερους θανάτους σε σύγκριση με ασθενείς που είχαν Covid-19 και κάποια άλλη ασθένεια. Ο ακριβής συνολικός αριθμός θανάτων αυτής της κατηγορίας ήταν 689, που είναι το 0,8% των θανάτων της χρονικής περιόδου που μελετήθηκε.
- V. Τα περισσότερα κρούσματα θανάτων που έχουν προέλθει από τον Covid-19 είναι της περίπτωσης Underlying Cause και όχι της Covid-19 Multiple Cause. Αυτό σημαίνει, πως υπερτερούν οι περιπτώσεις θανάτων, στις οποίες ο Covid-19 είναι υποκείμενη αιτία θανάτου και είμαστε βέβαιοι, πως ο ασθενής νοσούσε από Covid-19. Αναλυτικότερα, η αιτία θανάτου λόγω COVID-19 υπολογίζεται ανεξάρτητα από τις προ υπάρχουσες καταστάσεις υγείας του ασθενούς. Η σημασία αυτών των περιπτώσεων αναγράφεται στην υπό ενότητα 4.4.1.

- VI. Οι Γράφοι Γνώσεις είναι σπουδαία εργαλεία τροφοδότησης και ταχείας εύρεσης της πληροφορίας. Αυτό αποδεικνύεται με το query, που χρησιμοποιήθηκε για να δοθεί απάντηση στο πέμπτο ερώτημα. Συγκεκριμένα, έγινε δυνατή η άντληση πληροφοριών για τους ασθενείς, που είχαν κοινό χαρακτηριστικό το αρσενικό φύλο. Αυτές οι πληροφορίες αφορούσαν την ηλικία, τον τόπο καταγωγής, όλες τις ασθένειες του ασθενούς και τις ημερομηνίες καταγραφής των συμβάντων.
- VII. Τους περισσότερους θανάτους τους είχαν ασθενείς που το φύλο τους ήταν αρσενικό. Επομένως, φαίνεται πως οι άνθρωποι, των οποίων το φύλο είναι αρσενικό, έχουν περισσότερο κίνδυνο να πεθάνουν από τον COVID-19.
- VIII. Τους περισσότερους θανάτους από Covid-19 παρουσίασαν ασθενείς που κατοικούσαν στις Ηνωμένες Πολιτείες.
- IX. Μέσα από την μελέτη των δεδομένων, που χρησιμοποιήθηκαν, φαίνεται πως κινδυνεύουν περισσότερο όσοι ασθενείς είναι νεότεροι από 1 έτους ή η ηλικία τους είναι μέχρι τα 54 έτη.

Το παρακάτω διάγραμμα εκφράζει ξεχωριστά τον κάθε συνολικό αριθμό θανάτων για όλες τις ασθένειες που μελετήθηκαν.



Παρακάτω απεικονίζεται μία πίτα, που αντιπροσωπεύει τα ποσοστά επικινδυνότητας θανάτου από Covid-19 ανά ηλικιακή ομάδα.



Βέβαια θα μπορούσαν να γίνουν και μελλοντικές έρευνες ως συνέχεια της έρευνας αυτής. Στο υπό ενότητα 6.2 παρατίθενται κάποιες ιδέες για μελλοντικές έρευνες.

6.2 Ιδέες για μελλοντική έρευνα

Η συγκεκριμένη έρευνα θα μπορούσε η πραγμάτευση της μελέτης να ήταν περισσότερο αποτελεσματική και κατά επέκταση τα συμπεράσματα να ήταν πιο στοχευμένα αν:

- i. Υπήρχαν δεδομένα που να δείχνουν την συγγενική σχέση κάποιων ασθενών. Με τα δεδομένα αυτά, θα μπορούσαν να βγουν συμπεράσματα ως προς την κληρονομικότητα. Για παράδειγμα, αν νοσηλεύονταν με Covid-19 δύο ασθενείς που είχαν πρώτο βαθμό συγγένειας και στο τέλος πέθαιναν, θα μπορούσε να εξεταστεί το ενδεχόμενο υπόστασης κοινής ασθένειας ή χαμηλού ανοσοποιητικού στην οικογένεια μέσω κληρονομικότητας.

- ii. Υπήρχαν στοιχεία σχετικά με τον χειρισμό της κατάστασης από τον υγειονομικό τομέα. Με αυτά τα δεδομένα θα ήταν δυνατόν, να ελεγχθεί αν μπορούσε να χρησιμοποιηθεί καλύτερα η τεχνολογία για την αντιμετώπιση των προβλημάτων.
- iii. Υπήρχαν στοιχεία σχετικά με την ενημέρωση και την εκπαίδευση των ειδικών για τις δυνατότητες που προσφέρει η τεχνολογία και πως μπορούν να την χειριστούν. Είναι ένας ακόμα τρόπος με τον οποίο θα ελεγχόταν αν θα αν μπορούσε να χρησιμοποιηθεί καλύτερα η τεχνολογία για την αντιμετώπιση των προβλημάτων και την μείωση των θανάτων.

Επομένως, θα μπορούσε να γίνει μία μελλοντική έρευνα πάνω στον τομέα της υγείας και την χρήση Γράφων Γνώσης σε αυτόν με την προσθήκη κάποιων δεδομένων. Τα δεδομένα αυτά είναι η πιθανή συγγένεια των ασθενών, ο χειρισμός της κατάστασης από τους ειδικούς και η εκπαίδευση και γνώση των ειδικών πάνω στους Γράφους Γνώσης και την χρήση τους.

Βιβλιογραφία:

- Robinson, I., Webber, J., & Eifrem, E. (2015). *Graph Databases*.
- ΓΕΩΡΓΙΑΔΗΣ, Λ., ΝΙΚΟΛΟΠΟΥΛΟΣ, Σ., & ΠΑΛΗΟΣ, Λ. (2015). *ΑΛΓΟΡΙΘΜΙΚΗ ΘΕΩΡΙΑ ΓΡΑΦΗΜΑΤΩΝ*. Αθήνα.
- Νικολόπουλος, Σ. (2017-2018). *Θεωρία Γραφημάτων Θεμελιώσεις-Αλγόριθμοι-Εφαρμογές*.
- Nicholson, D., & Greene, C. (2020). Constructing knowledge graphs and their biomedical applications.
- Shao, L., Duan, Y., Sun, X., Zou, Q., Jing, R., & Lin, J. (2017). Bidirectional Value Driven Design between Economical Planning and Technical Implementation Based on Data Graph, Information Graph and Knowledge Graph.
- Wang, F., Cui, P., Pei, J., Song, Y., & Zang, C. (2020). Recent Advances on Graph Analytics and.
- Alshahrani, M., Khan, M., Maddouri, O., Kinjo, A., Queralt-Rosinach, N., & Hoehndorf, R. (2017). Neuro-symbolic representation learning on biological knowledge graphs.
- Gyrard, A., Gaur, M., Shekarpour, S., Thirunarayan, K., & Sheth, A. (2018). Personalized Health Knowledge Graph.
- Vlad, A., Vahdati, S., Nayyeri, M., Bellomarini, L., & Sallinger, E. (2022). Towards Hybrid Logic-based and Embedding-based Reasoning on Financial Knowledge Graphs. *Workshop Proceedings of the EDBT/ICDT*.
- Wang, S. (2022). On the Analysis of Large Integrated Knowledge Graphs for Economics, Banking, and Finance. *Workshop Proceedings of the EDBT/ICDT*.
- Νικολόπουλος, Σ. (2016-2017). *Θεωρία Γραφημάτων Ενότητα 9 Μετασχηματισμοί Υπολογιστικών Προβλημάτων*.
- Νικολόπουλος, Σ. (2017-2018). *Θεωρία Γραφημάτων Θεμελιώσεις-Αλγόριθμοι-Εφαρμογές*. Στο Σ. Δ. Νικολόπουλος.
- Bilal , A.-S., Muhammad , A.-Q., Mohammed, A., Mohammad , A.-S., Reem , A., & Heba , S. (2023). Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. In *Journal of Big Data*.
- Gyrard, A., Gaur, M., Shekarpour, S., Thirunarayan, K., & Sheth, A. (2018). Personalized Health Knowledge Graph.
- Limeng , C., & Dongwon , L. (2022). Knowledge Graph Enhanced Thread Recommendation.
- Longxiang , S., Shijian , L., Xiaoran , Y., Jiaheng , Q., Gang , P., & Binbin , Z. (2017). Semantic Health Knowledge Graph: Semantic Integration of Heterogeneous Medical Knowledge and Services.
- Rajabi, E., & Kafaie , S. (2022, Σεπτέμβριος 28). Knowledge Graphs and Explainable AI in Healthcare. (2015). *Αναπαράσταση Γραφημάτων*. (χ.χ.). *Εικόνα 1.1*.