

Μετρικές Σύζευξης και Συνεκτικότητας σε Σχήματα Βάσεων Δεδομένων

Η Μεταπτυχιακή Εργασία Εξειδίκευσης

υποβάλλεται στην ορισθείσα
από τη Συνέλευση
του Τμήματος Μηχανικών Η/Υ και Πληροφορικής
Εξεταστική Επιτροπή

από τον

Γεώργιο Μάμαλη

ως μέρος των υποχρεώσεων για την απόκτηση του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ
ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ
ΣΤΟ ΛΟΓΙΣΜΙΚΟ

Πανεπιστήμιο Ιωαννίνων

Ιούλιος 2019

Εξεταστική Επιτροπή:

- **Απόστολος Ζάβρας** Αναπλ. Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Παναγιώτης Βασιλειάδης**, Αναπλ. Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Νικόλαος Μαμουλής**, Καθηγητής, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

ΑΦΙΕΡΩΣΗ

Στους γονείς μου και τον αδερφό μου.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Απόστολο Ζάρρα για την καθοδήγηση και την υπομονή του κατά τη διάρκεια της προσπάθειας ολοκλήρωσης της εν λόγω εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω τους συναδέλφους μεταπτυχιακούς φοιτητές του Πανεπιστημίου για την βοήθεια και την υποστήριξή τους.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου για την στήριξη κατά τη διάρκεια των σπουδών μου.

ΠΕΡΙΕΧΟΜΕΝΑ

Κατάλογος Σχημάτων	iii
Κατάλογος Πινάκων	v
Περίληψη	vi
Extended Abstract	viii
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	1
1.1 Εισαγωγή	1
1.2 Κίνητρο	2
1.3 Οργάνωση του Τόμου	5
ΚΕΦΑΛΑΙΟ 2. ΥΠΟΒΑΘΡΟ	6
2.1 Γενικά για την Συνεκτικότητα	6
2.2 Γενικά για την Σύζευξη	12
2.3 Γνωστές Μετρικές Συνεκτικότητας	18
2.4 Γνωστές Μετρικές Σύζευξης	20
2.5 Ιδιότητες Μετρικών	22
ΚΕΦΑΛΑΙΟ 3. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΣΧΗΜΑΤΩΝ	25
3.1 Σημασία της Μοντελοποίησης	25
3.2 Μοντελοποίηση SQL Σχημάτων	26
3.3 Ορισμός Γραφημάτων Ενοτήτων	28

ΚΕΦΑΛΑΙΟ 4. ΜΕΤΡΙΚΕΣ ΣΥΝΕΚΤΙΚΟΤΗΤΑΣ	32
4.1 Ορισμός Μετρικών	32
4.2 Απόδειξη Ιδιοτήτων Μετρικών	34
ΚΕΦΑΛΑΙΟ 5. ΜΕΤΡΙΚΕΣ ΣΥΖΕΥΞΗΣ	38
5.1 Ορισμός Μετρικών	38
5.2 Απόδειξη Ιδιοτήτων Μετρικών	40
ΚΕΦΑΛΑΙΟ 6. ΕΜΠΕΙΡΙΚΗ ΜΕΛΕΤΗ	45
6.1 Σχετικά με τα Δεδομένα	45
6.2 Βασικά Στατιστικά	50
6.3 Συσχετισμός Μετρικών με Άλλα Στοιχεία	61
6.4 Θέματα προς Εγκυρότητα	65
ΚΕΦΑΛΑΙΟ 7. ΕΠΙΛΟΓΟΣ	66
7.1 Επίλογος	66
7.2 Μελλοντική Δουλειά	67
ΒΙΒΛΙΟΓΡΑΦΙΑ	68

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 2.1: Ταξινόμηση των Ειδών Συνεκτικότητας.	11
Σχήμα 2.2: Διάγραμμα Ροής Απόφασης του Είδους Συνεκτικότητας.	12
Σχήμα 2.3: Οπτική Αναπαράσταση της Διαφοράς των Εννοιών Σύζευξη και Συνεκτικότητα.	13
Σχήμα 2.4: Ταξινόμηση των Ειδών Σύζευξης.	17
Σχήμα 3.1: Παράδειγμα Γραφήματος Συνεκτικότητας.	29
Σχήμα 3.2: Παράδειγμα Γραφήματος Σύζευξης.	31
Σχήμα 4.1: Σύγκριση Γραφήματος Συνεκτικότητας με Πλήρες Γράφημα.	33
Σχήμα 5.1: Γράφημα Σύζευξης Τυχαίου Σχήματος.	39
Σχήμα 6.1: Πλήθος Πινάκων ανά Σχήμα.	47
Σχήμα 6.2: Κατηγοριοποίηση Σχημάτων με Βάση το Πλήθος Πινάκων.	48
Σχήμα 6.3: Μέσο Πλήθος Πινάκων ανά Σχήμα.	48
Σχήμα 6.4: Κατηγοριοποίηση Σχημάτων με Βάση το Μέσο Πλήθος Πινάκων	49
Σχήμα 6.5: Πλήθος Queries ανά Σχήμα.	50
Σχήμα 6.6: Μέγιστη Συνεκτικότητα ανά Σχήμα.	52
Σχήμα 6.7: Ελάχιστη Συνεκτικότητα ανά Σχήμα.	52
Σχήμα 6.8: Μέση Συνεκτικότητα ανά Σχήμα.	53
Σχήμα 6.9: Median Συνεκτικότητα ανά Σχήμα.	54
Σχήμα 6.10: Mode Συνεκτικότητα ανά Σχήμα.	54
Σχήμα 6.11: Τυπική Απόκλιση Συνεκτικότητας ανά Σχήμα.	55
Σχήμα 6.12: Μέγιστη Σύζευξη ανά Σχήμα.	57
Σχήμα 6.13: Μέση Σύζευξη ανά Σχήμα.	58
Σχήμα 6.14: Τυπική Απόκλιση Σύζευξης ανά Σχήμα.	58
Σχήμα 6.15: Μέγιστη Κανονικοποιημένη Σύζευξη ανά Σχήμα.	59
Σχήμα 6.16: Μέση Κανονικοποιημένη Σύζευξη ανά Σχήμα.	60
Σχήμα 6.17: Τυπική Απόκλιση Κανονικοποιημένης Σύζευξης ανά Σχήμα	61

Σχήμα 6.18: Συσχετισμός Συνεκτικότητας – Μεγέθους Πίνακα.	62
Σχήμα 6.19: Συσχετισμός Μέσης Σύζευξης – Μεγέθους Σχήματος.	63
Σχήμα 6.20: Συσχετισμός Μέσης Κανονικοποιημένης Σύζευξης – Μεγέθους Σχήματος.	64

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 6.1 Πληροφορίες των Δεδομένων	46
Πίνακας 6.2 Στατιστικά Μετρικών	51
Πίνακας 6.3 Μέση Τιμή Μετρικών στο Υποσύνολο Σύζευξης	56

ΠΕΡΙΛΗΨΗ

Γεώργιος Μάμαλης

MSc, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

Ιούλιος 2019

Τίτλος: Μετρικές Σύζευξης και Συνεκτικότητας σε Σχήματα Βάσεων Δεδομένων

Επιβλέπων: Απόστολος Ζάρρας

Με τον όρο συνεκτικότητα (cohesion) στην τεχνολογία λογισμικού αναφερόμαστε στον βαθμό στον οποίο τα στοιχεία εντός μιας δομής σχετίζονται μεταξύ τους. Ουσιαστικά, αποτελεί ένα μέτρο «δύναμης» των σχέσεων των στοιχείων μιας κλάσης.

Αντίστοιχα, ο όρος σύζευξη (coupling) αναφέρεται στον βαθμό αλληλεξάρτησης μεταξύ 2 διαφορετικών δομικών στοιχείων. Αποτελεί δηλαδή ένα μέτρο του πόσο στενά σχετίζονται τα στοιχεία των δύο αυτών δομών.

Οι 2 παραπάνω έννοιες είναι πολύ σημαντικές και λαμβάνονται σε πολύ μεγάλο βαθμό υπόψη όταν σχεδιάζουμε ένα λογισμικό. Γενικά, η καλή γνώση και των δύο εννοιών είναι αναγκαία για τον σχεδιασμό συστημάτων που είναι επεκτάσιμα, εύκολα κατανοητά και διαχειρίσιμα. Κυρίως, επιδιώκουμε χαμηλή σύζευξη και υψηλή συνεκτικότητα. Συστήματα με χαμηλή σύζευξη και υψηλή συνεκτικότητα επηρεάζονται λιγότερο από αλλαγές, είναι πιο κατανοητά και επαναχρησιμοποιούνται με μεγαλύτερη ευκολία.

Στην παρούσα μεταπτυχιακή εργασία μελετάμε τις έννοιες της σύζευξης και της συνεκτικότητας σε σχήματα βάσεων δεδομένων. Συγκεκριμένα, ξεκινώντας από SQL περιγραφές σχημάτων, παράγουμε γραφήματα σύζευξης/συνεκτικότητας. Κατόπιν προτείνουμε μετρικές για την αξιολόγηση των παραπάνω εννοιών. Ορίζουμε τις προτεινόμενες μετρικές φορμαλιστικά και αποδεικνύουμε ότι είναι καλά ορισμένες, με βάση ένα σύνολο από ιδιότητες που πρέπει να πληρούνται από τις μετρικές ποιότητας λογισμικού γενικότερα. Τέλος, με βάση τις προτεινόμενες μετρικές, πραγματοποιούμε μια εκτενή εμπειρική μελέτη σε ένα σύνολο σχημάτων που προέρχονται από συστήματα ανοιχτού λογισμικού.

EXTENDED ABSTRACT

Georgios Mamalis

MSc, Computer Science and Engineering, University of Ioannina, Greece

July 2019

Title: Coupling and Cohesion Metrics on Schema Databases

Supervisor: Apostolos Zarras

The term cohesion in software refers to the extent to which elements within a structure are related to each other. Essentially, it is a measure of strength of the relationships of the elements of a class. Correspondingly, the term coupling refers to the degree of interdependence between two different structural elements. It is a measure of how closely the elements of these two structures are closely related.

The above two concepts are very important and are taken into account very much when designing a software. In general, good knowledge of both concepts is necessary to design systems that are scalable, easily understood and manageable. Above all, we seek low coupling and high consistency. Systems with low coupling and high consistency are less affected by changes, are more understandable and reusable with greater ease.

In this thesis we study the concepts of coupling and cohesion in database schemas. In particular, starting with SQL schema descriptions, we produce coupling/cohesion graphs. Then we recommend metrics to evaluate the above concepts. We define the proposed metric formulas and prove that they are well defined, based on a set of properties that must be met by

software quality metrics in general. Finally, based on the proposed metrics, we conduct an extensive empirical study on a set of shapes derived from open source systems.

ΚΕΦΑΛΑΙΟ 1.

ΕΙΣΑΓΩΓΗ

- 1.1 Εισαγωγή
 - 1.2 Κίνητρο
 - 1.3 Οργάνωση του Τόμου
-

1.1 Εισαγωγή

Το να έχουμε μια ακριβή και ενημερωμένη βάση δεδομένων είναι στοιχείο αναγκαίο, ανεξάρτητα από τον κλάδο που εργαζόμαστε. Ο σχεδιασμός μιας καλής βάσης, εκτός από εγγύηση απόδοσης, πρέπει να παρέχει και εγγύηση ορθότητας. Κάθε φορά που κάτι αλλάζει στη βάση μας, η μετάβαση στη νέα κατάσταση πρέπει να γίνεται με απλό και ομαλό τρόπο. Επιπροσθέτως, επιδιώκουμε οι αλλαγές αυτές να «απορροφηθούν» χωρίς να προκαλούν επιπλέον ταλαιπωρία.

Για να το πετύχουμε αυτό, πρέπει το σύστημά μας να είναι κατανοητό, συντηρήσιμο και να επαναχρησιμοποιείται με ευκολία. Μελετώντας τα παραπάνω, καταλαβαίνουμε πως η «ραχοκοκαλιά» οποιουδήποτε καλού συστήματος λογισμικού είναι η σωστή σχεδίασή του.

Η ανάγκη για αξιολόγηση του σχεδιασμού των συστημάτων οδήγησε στην έννοια των μετρικών. Αποτελούν στοιχείο κλειδί για τον έλεγχο της διαδικασίας παραγωγής ενός λογισμικού. Ορίζονται ως *μετρήσιμα μεγέθη που μπορούν να χρησιμοποιηθούν για την αξιολόγηση διάφορων χαρακτηριστικών ενός συστήματος.*

Οι μετρικές παίζουν πολύ σημαντικό ρόλο (όπως είπαμε και προηγουμένως) στην εκτίμηση αλλά και στην πρόβλεψη διάφορων στοιχείων του συστήματος όπως η πολυπλοκότητα, η δυνατότητα επαναχρησιμοποίησης/συντήρησης και πολλά άλλα όπως η σύζευξη και η συνεκτικότητα. Οι 2 τελευταίες έννοιες είναι αυτές στις οποίες θα εστιάσουμε και θα αναλύσουμε εκτενώς στη συνέχεια.

1.2 Κίνητρο

Όπως όλα τα συστήματα λογισμικού, έτσι και οι βάσεις δεδομένων εξελίσσονται όσο περνάει ο καιρός. Εκτός της εξέλιξης του περιεχομένου (διαφορετική πληροφορία), αλλάζει και η εσωτερική δομή της βάσης (διαφορετικό είδος πληροφορίας). Με τον όρο *σχήμα βάσεων δεδομένων* αναφερόμαστε στην τυπική απεικόνιση της δομής, δηλαδή τα στοιχεία μιας δομημένης συλλογής δεδομένων και οι σχέσεις μεταξύ τους. Από την άλλη, ο όρος *εξέλιξη λογισμικού* αναφέρεται στην μετάβαση του σχήματος από μια κατάσταση σε μια νέα και διαφορετική.

Η εξέλιξη ενός σχήματος μπορεί να έχει πολύ σημαντική επίπτωση σε ολόκληρο το σύστημα που «χτίζεται» γύρω από τη βάση, γεγονός που επηρεάζει τους προγραμματιστές, αλλά και τους χρήστες της. Είναι αρκετά σύνηθες το φαινόμενο να περιμένει η οποιαδήποτε προσπάθεια παραγωγής μιας εφαρμογής, μέχρις ότου το σχήμα της βάσης να βρίσκεται σε

μια σταθερή κατάσταση, με μικρό το ενδεχόμενο εκ νέου αλλαγής. Ο λόγος που συμβαίνει αυτό προκύπτει εύκολα αν αναλογιστούμε τη σημασία της βάσης και το ποιοι (και πόσο) εξαρτώνται από την κατάστασή της. Μια αλλαγή στο σχήμα της βάσης δεδομένων μπορεί να οδηγήσει άμεσα διάφορες εφαρμογές που λειτουργούν πάνω ή γύρω από αυτή να «αποτύχουν» (π.χ. σε περίπτωση που διαγράψουμε μια πληροφορία), ή να παρουσιαστούν σημασιολογικά ελαττωματικές/ανακριβείς (π.χ. σε περίπτωση που προσθέσουμε επιπλέον πληροφορία).

Από τα παραπάνω καταλαβαίνουμε πως είναι γενικά καλό να αποφεύγουμε τις αλλαγές σε ένα σχήμα βάσεων δεδομένων, κυρίως λόγω των αλλαγών και ενεργειών που αλυσιδωτά εμφανίζονται στη συνέχεια. Αυτό όμως δεν είναι ιδιαίτερα εφικτό, καθώς δεν μπορούμε να έχουμε από την αρχή πλήρη επίγνωση του ιδανικού σχεδιασμού μιας βάσης. Ο σημαντικότερος λόγος είναι το γεγονός πως ο σκοπός για τον οποίο δημιουργούμε ένα σύστημα είναι πάρα πολύ πιθανό να αλλάξει παραπάνω από μια φορά. Άρα είναι απαραίτητο να μπορούμε να αλλάζουμε τη δομή του σχήματός μας.

Στο σημείο αυτό έρχεται η έννοια του *refactoring*. Με τον όρο αυτό αναφερόμαστε σε μια απλή αλλαγή σε ένα σχήμα βάσης δεδομένων με σκοπό να βελτιώσει τον σχεδιασμό του, διατηρώντας παράλληλα τα δεδομένα αλλά και τον τρόπο της λειτουργίας της βάσης αναλλοίωτο. Το *refactoring* μιας βάσης δεδομένων είναι πιο δύσκολο από ένα *refactoring* κώδικα, γιατί στον κώδικα πρέπει μόνο να διατηρήσουμε αναλλοίωτο τον τρόπο λειτουργίας του, ενώ στις βάσεις δεδομένων πρέπει επίσης να προσέχουμε και το σύνολο των δεδομένων μας.

Οι Sharma, Fragkoulis, Rizou, Bruntink και Spinellis [8], έχοντας ως στόχο την κατανόηση αλλά και την μέτρηση της ποιότητας των σχημάτων βάσεων δεδομένων, εισήγαγαν την έννοια των *database smells*. Ορίζονται ως τα χαρακτηριστικά του κώδικα βάσεων δεδομένων

(π.χ. εντολές SQL), του συστήματος, ή των δεδομένων που παραβιάζουν τις συνιστώμενες βέλτιστες πρακτικές και επηρεάζουν αρνητικά την ποιότητα του συστήματος λογισμικού. Τα smells αυτά μπορεί να είναι 3 κατηγοριών:

- Schema Smells: Δημιουργούνται λόγω κακού σχεδιασμού του σχήματος.
- Query Smells: Δημιουργούνται λόγω κακογραμμένων SQL εντολών (queries).
- Data Smells: Δημιουργούνται λόγω κακή διαχείριση των δεδομένων.

Σκοπός λοιπόν της συγκεκριμένης εργασίας είναι μελετήσουμε τις έννοιες της σύζευξης και της συνεκτικότητας σε σχήματα βάσεων δεδομένων. Συγκεκριμένα, οι στόχοι μας είναι:

- Να παράγουμε γραφήματα σύζευξης/συνεκτικότητας.
- Να προτείνουμε φορμαλιστικά μετρικές για την αξιολόγηση των παραπάνω εννοιών και να αποδείξουμε ότι είναι καλά ορισμένες, με βάση ένα σύνολο από ιδιότητες που πρέπει να πληρούνται από τις μετρικές ποιότητας λογισμικού γενικότερα.
- Να πραγματοποιήσουμε μια εκτενή εμπειρική μελέτη σε ένα σύνολο σχημάτων που προέρχονται από συστήματα ανοιχτού λογισμικού.

Ο κώδικας που υλοποιεί όλα τα παραπάνω βρίσκεται στο github [14].

1.3 Οργάνωση του Τόμου

Τα παρακάτω κεφάλαια πραγματεύονται αναλυτικότερα την δημιουργία της εν λόγω εργασίας. Πιο συγκεκριμένα:

Στο κεφάλαιο 2 περιλαμβάνει το υπόβαθρο της εργασίας. Συζητάμε εκτενώς τη σημασία των εννοιών της σύζευξης και της συνεκτικότητας, καθώς επίσης αναφερόμαστε σε γνωστές μετρικές των 2 αυτών εννοιών.

Στο κεφάλαιο 3 μοντελοποιούμε το πρόβλημα που έχουμε να αντιμετωπίσουμε και ορίζουμε τις βασικές έννοιες που θα χρειαστούμε.

Το κεφάλαιο 4 περιέχει τις προτεινόμενες μετρικές για τον υπολογισμό της συνεκτικότητας ενός συστήματος. Τις ορίζουμε φορμαλιστικά και αποδεικνύουμε ότι είναι καλά ορισμένες, με βάση ένα σύνολο από ιδιότητες που πρέπει να πληρούνται από τις μετρικές ποιότητας λογισμικού γενικότερα.

Στο κεφάλαιο 5 επαναλαμβάνουμε τη διαδικασία του κεφαλαίου 4, αυτή τη φορά για τον υπολογισμό της σύζευξης.

Το κεφάλαιο 6 περιέχει την εμπειρική μελέτη που διεξήγαμε, καθώς και τις παρατηρήσεις πάνω στα πειράματά μας.

Το κεφάλαιο 7 περιέχει τον επίλογο και την μελλοντική εργασία.

ΚΕΦΑΛΑΙΟ 2.

ΥΠΟΒΑΘΡΟ

- 2.1 Γενικά για την Συνεκτικότητα
 - 2.2 Γενικά για την Σύζευξη
 - 2.3 Γνωστές Μετρικές Συνεκτικότητας
 - 2.4 Γνωστές Μετρικές Σύζευξης
 - 2.5 Ιδιότητες Μετρικών
-

2.1 Γενικά για την Συνεκτικότητα

Με τον όρο *συνεκτικότητα* (*cohesion*) στην τεχνολογία λογισμικού αναφερόμαστε στον βαθμό στον οποίο τα στοιχεία εντός μιας δομής σχετίζονται μεταξύ τους. Ουσιαστικά, αποτελεί ένα μέτρο «δύναμης» των σχέσεων των στοιχείων μιας κλάσης. Για να προσαρμόσουμε τον παραπάνω ορισμό στις βάσεις δεδομένων (με τις οποίες θα ασχοληθούμε στη συνέχεια), μπορούμε να ορίσουμε τη συνεκτικότητα ως το μέτρο «δύναμης» των συνδέσμων μεταξύ των χαρακτηριστικών ενός πίνακα.

Από τον παραπάνω ορισμό, καταλαβαίνουμε ότι η συνεκτικότητα είναι ουσιαστικά η εσωτερική «κόλλα» που κρατά τα στοιχεία μαζί. Προφανώς, σε ένα λογισμικό θέλουμε να έχουμε υψηλή συνεκτικότητα, καθώς, όσο πιο μεγάλη είναι, τόσο περισσότερο σχετίζονται

μεταξύ τους τα εσωτερικά μέρη μιας δομής και εξυπηρετούν καλύτερα το συνολικό σκοπό της.

Ανάλογα με το πόσο μεγάλη είναι η συνεκτικότητα ενός συστήματος, διακρίνουμε 7 διαφορετικά είδη, τα οποία αναλύουμε παρακάτω.

Πριν ξεκινήσουμε την αναφορά στα επιμέρους είδη συνεκτικότητας (αλλά και της σύζευξης στη συνέχεια) ορίζουμε την έννοια του *module* (ενότητα ή συστατικό). Με αυτό τον όρο θα αναφερόμαστε από εδώ και πέρα σε οποιαδήποτε δομή, είτε αυτή είναι ένας πίνακας βάσης δεδομένων, είτε αυτό είναι μια κλάση.

2.1.1 Λειτουργική Συνεκτικότητα (Functional Cohesion)

Στην λειτουργική συνεκτικότητα όλα τα στοιχεία του *module* συνεισφέρουν στην εκτέλεση αποκλειστικά μιας και μόνο λειτουργίας. Χρησιμοποιείται σε περιπτώσεις που έχουν σκοπό την ολοκλήρωση αποκλειστικά μιας λειτουργίας και, προφανώς, δεν υπάρχει στοιχείο του *module* που ασχολείται με οποιαδήποτε άλλη λειτουργία. Αποτελεί την πιο ισχυρή μορφή συνεκτικότητας και είναι καλή λόγω της εύκολης εύρεσης/διόρθωσης λαθών αλλά και για την εύκολη δυνατότητα επαναχρησιμοποίησης του *module*. Ένα παράδειγμα της λειτουργικής συνεκτικότητας είναι ο υπολογισμός της θερμοκρασίας του δωματίου.

2.1.2 Σειριακή Συνεκτικότητα (Sequential Cohesion)

Στην περίπτωση αυτή τα στοιχεία του module εμπλέκονται σε εργασίες για τις οποίες ισχύει ότι η έξοδος που παράγει η μία γίνεται η είσοδος (ή κομμάτι της εισόδου) μιας άλλης. Η σειριακή συνεκτικότητα δεν είναι εύκολα επαναχρησιμοποιήσιμη καθώς, γενικά, οι διεργασίες που «συνεργάζονται» παραπάνω δεν ξαναχρησιμοποιούνται μαζί. Παράδειγμα τέτοιας μορφής συνεκτικότητας είναι η ακολουθία εργασιών παρακάτω [9]:

```
struct SalesRegion{ }: //definition
initSalesRegion( ): //function
updateSalesRegion( ): //function
printSalesRegion( ): //function
```

2.1.3 Επικοινωνιακή Συνεκτικότητα (Communicational Cohesion)

Εδώ, τα στοιχεία του module συμμετέχουν σε λειτουργίες που χρησιμοποιούν τις ίδιες πηγές εισόδου ή παράγουν την ίδια έξοδο. Γενικά, στην περίπτωση αυτή δεν είναι εύκολη η επαναχρησιμοποίηση γιατί κάθε module τελικά εκτελεί πάνω από μια ενέργειες. Για να λυθεί το παραπάνω πρόβλημα, μπορούμε να προσπαθήσουμε να χωρίσουμε τα modules σε υπό-modules τα οποία να ανήκουν (ιδανικά) στην κατηγορία της λειτουργικής συνεκτικότητας (που αναλύθηκε πρώτη). Γενικά, η επικοινωνιακή συνεκτικότητα είναι αποδεκτή σε περιπτώσεις που δεν μπορούμε να σπάσουμε ένα module σε υπό-modules που ανήκουν σε πιο υψηλή κατηγορία. Παράδειγμα τέτοιου είδους είναι η εύρεση και επιστροφή του ονόματος ενός πελάτη.

2.1.4 Διαδικασιακή Συνεκτικότητα (Procedural Cohesion)

Στην περίπτωση αυτή, τα modules σχετίζονται μεταξύ τους μόνο μέσω σειράς, δηλαδή διαφορετικές λειτουργίες ομαδοποιούνται επειδή πρέπει να εκτελεστούν με συγκεκριμένη σειρά. Συνήθως οι λειτουργίες δεν έχουν κάποιο κοινό δεδομένο, αλλά απλά όταν ολοκληρώνεται η μία, περνάει τον έλεγχο στην άλλη. Τα βασικά προβλήματα είναι ότι οι εργασίες είναι μόνο χαλαρά συνδεδεμένες μεταξύ τους και ότι δεν επιτρέπει την επαναχρησιμοποίηση. Μοιάζει αρκετά με την σειριακή συνεκτικότητα, με τη βασική διαφορά ότι τα στοιχεία δεν σχετίζονται μεταξύ τους. Παράδειγμα διαδικασιακής συνεκτικότητας είναι η παρακάτω ακολουθία εργασιών [10]:

Clean Utensils from Previous Meal

Prepare Turkey for Roasting

Chop Vegetables

Set Table

2.1.5 Χρονική Συνεκτικότητα (Temporal Cohesion)

Στην χρονική συνεκτικότητα, τα στοιχεία του module ουσιαστικά δεν έχουν κάποια συσχέτιση μεταξύ τους, αλλά χρησιμοποιούνται σε διεργασίες οι οποίες σχετίζονται χρονικά. Βασικό πρόβλημα της περίπτωσης αυτής είναι το γεγονός ότι οι διεργασίες που εκτελούνται μέσα στο module έχουν μόνο μικρή σχέση μεταξύ τους αλλά συνήθως μεγάλη σχέση με ενέργειες που γίνονται σε άλλα modules. Εμφανίζεται σε περιπτώσεις που ένα module κάνει αρχικοποίηση. Ένα παράδειγμα είναι χρονικής συνεκτικότητας είναι το παρακάτω:

```
Class initialize{
```

```
//Αρχικοποίηση μεταβλητών του προγράμματος  
//Διάβασμα πρώτης εγγραφής ενός πίνακα  
//Εμφάνιση μηνύματος στην οθόνη}
```

2.1.6 Λογική Συνεκτικότητα (Logical Cohesion)

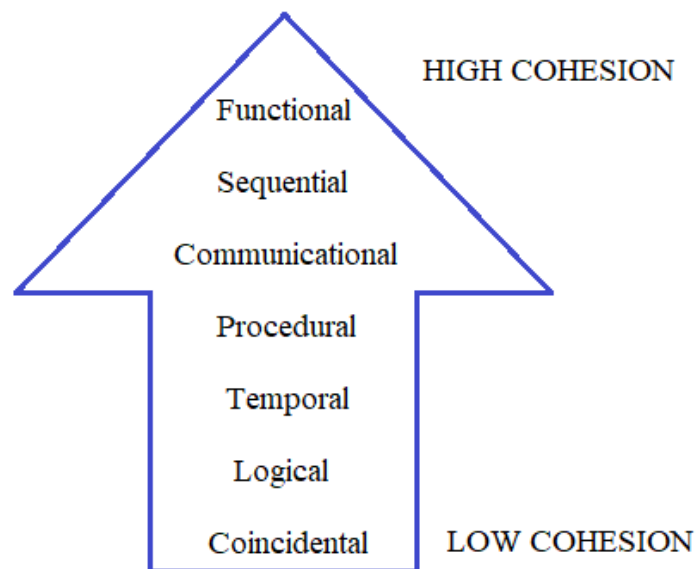
Εμφανίζεται όταν διάφορα λογικά σχετιζόμενα δεδομένα τοποθετούνται στο ίδιο module. Δηλαδή, τα στοιχεία του module χρησιμοποιούνται από εργασίες που ανήκουν στην ίδια κατηγορία ή στον ίδιο τύπο εργασιών. Για παράδειγμα, έχουμε μια συνάρτηση η οποία γράφει δεδομένα σε συσκευές και επιλέγουμε τη συσκευή στην οποία θα γράψει, περνώντας μια παράμετρο (π.χ. αν περαστεί ο κωδικός 1 γράφει στον δίσκο, αν περαστεί ο κωδικός 2 γράφει στην μνήμη). Βασικό πρόβλημα του της λογικής συνεκτικότητας είναι ότι δεν είναι ξεκάθαρο πού βρίσκεται και τι κάνει κάθε κομμάτι. Επίσης, παρατηρείται δυσκολία στην επαναχρησιμοποίηση.

2.1.7 Συμπτωματική Συνεκτικότητα (Coincidental Cohesion)

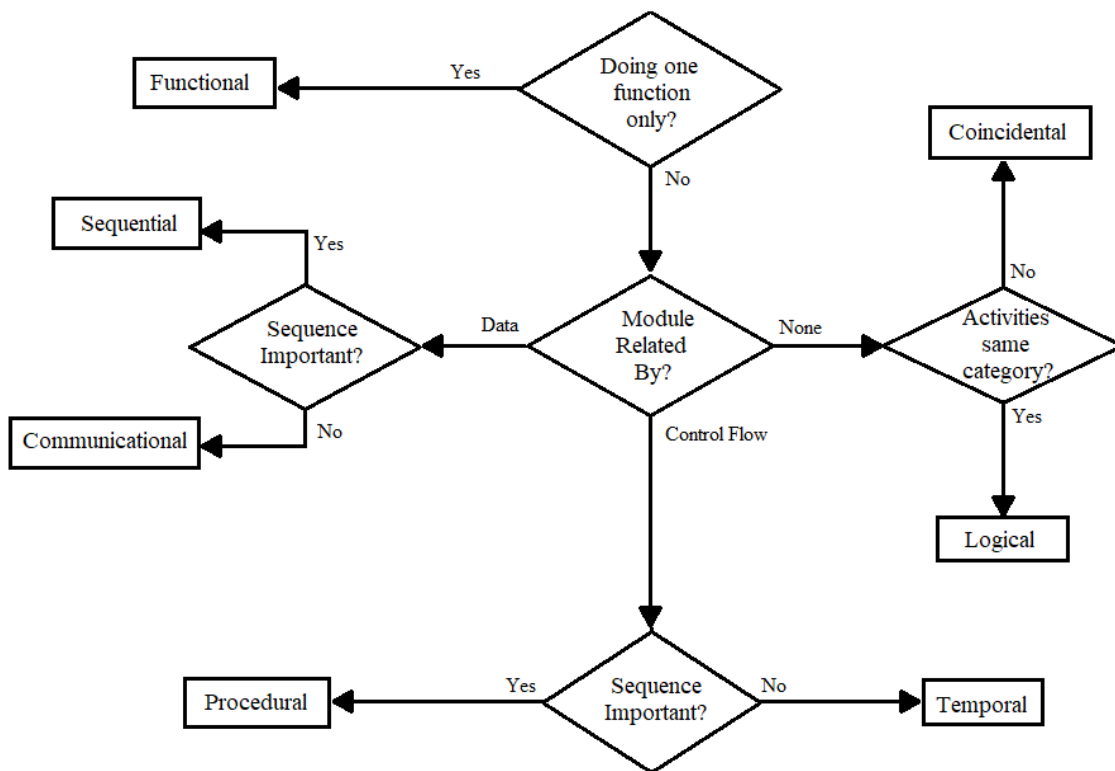
Αποτελεί την πιο «αδύναμη» μορφή συνεκτικότητας. Τα στοιχεία του module συμμετέχουν σε διεργασίες χωρίς ουσιαστική σχέση μεταξύ τους. Μοιάζει ως έναν βαθμό με την λογική συνεκτικότητα, με τη βασική διαφορά ότι οι εργασίες που πραγματοποιούνται μπορεί να μην είναι καν του ίδιου τύπου. Συνήθως προκύπτει από εντολές του τύπου «*γράψτε μια συνάρτηση που αποτελείται από 10-20 εντολές*». Βασικό πρόβλημα που παρουσιάζεται εδώ είναι η

δυσκολία συντήρησης, καθώς η δομή είναι ιδιαίτερα δυσνόητη. Διορθώνεται όμως αρκετά εύκολα, αν φτιάξουμε τόσες χωριστές δομές όσες και οι ενέργειες που έχουμε.

Το σχήμα 2.1 δείχνει συνοπτικά τα είδη συνεκτικότητας που αναλύσαμε παραπάνω και τα ταξινομεί με βάση την «δύναμή» τους. Τέλος, στο σχήμα 2.2 παρουσιάζεται ένα διάγραμμα ροής που μας δείχνει πότε χρησιμοποιούμε την κάθε περίπτωση.



Σχήμα 2.1: Ταξινόμηση των Ειδών Συνεκτικότητας [11].



Σχήμα 2.2: Διάγραμμα Ροής Απόφασης του Είδους Συνεκτικότητας [12].

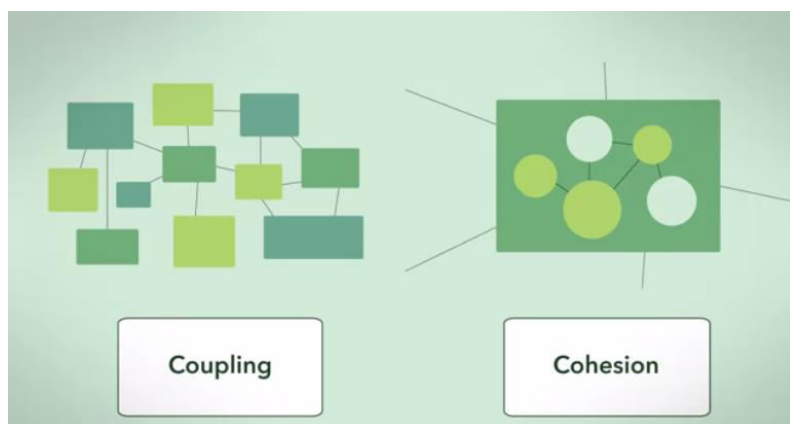
2.2 Γενικά για την Σύζευξη

Όπως και στην περίπτωση της συνεκτικότητας, έτσι και εδώ θα χρησιμοποιήσουμε την έννοια του module (ενότητα/συστατικό) για την περιγραφή των παρακάτω εννοιών.

Ο όρος *σύζευξη* (*coupling*) αναφέρεται στον βαθμό αλληλεξάρτησης μεταξύ 2 διαφορετικών modules. Αποτελεί δηλαδή ένα μέτρο του πόσο στενά σχετίζονται τα στοιχεία των δύο αυτών modules. Αν πάλι προσαρμόσουμε τον ορισμό αυτόν στις βάσεις δεδομένων ορίζουμε την

σύζευξη ως ένα μέτρο «δύναμης» των συνδέσμων μεταξύ των χαρακτηριστικών 2 διαφορετικών πινάκων.

Από τον παραπάνω ορισμό, καταλαβαίνουμε ότι μιλάμε για εξωτερικές συνδέσεις μεταξύ modules, σε αντίθεση με την συνεκτικότητα, όπου αναφερόμαστε σε εσωτερικές σχέσεις/συνδέσεις. Στο σχήμα 2.3 δείχνουμε οπτικά την διαφορά στις 2 αυτές έννοιες.



Σχήμα 2.3: Οπτική Αναπαράσταση της Διαφοράς των Εννοιών Σύζευξη και Συνεκτικότητα [13].

Από τη στιγμή που οι 2 αυτές έννοιες είναι αντίθετες (και λαμβάνοντας υπόψη πως προηγουμένως αναφέραμε ότι στοχεύουμε σε υψηλές τιμές συνεκτικότητας), θέλουμε να έχουμε όσο το δυνατόν πιο ανεξάρτητες και, συνεπώς στοχεύουμε σε χαμηλή σύζευξη όταν χτίζουμε το λογισμικό μας.

Ανάλογα με το πόσο μεγάλη είναι η σύζευξη ενός συστήματος, διακρίνουμε 6 διαφορετικά είδη, τα οποία αναλύουμε παρακάτω.

2.2.1 Σύζευξη Περιεχομένου (Content Coupling)

Ο πιο ισχυρός (και λιγότερο καλός) τύπος σύζευξης. Εμφανίζεται στην περίπτωση που μια δομή αναφέρεται στο εσωτερικό μιας άλλης ή επεξεργάζεται το εσωτερικό μιας άλλης. Βασικό πρόβλημα του είδους αυτού είναι το γεγονός ότι σχεδόν οποιαδήποτε αλλαγή στη δομή A, προϋποθέτει αλλαγή και στη δομή B (αν υποθέσουμε ότι υπάρχει σύζευξη περιεχομένου μεταξύ των δομών A και B). Επίσης, είναι αδύνατον να χρησιμοποιηθεί η δομή A χωρίς να χρησιμοποιηθεί και η B. Στον προγραμματισμό, η σύζευξη περιεχομένου εμφανίζεται κυρίως σε γλώσσες χαμηλού επιπέδου.

2.2.2 Σύζευξη Κοινών Δεδομένων (Common Coupling)

Στην σύζευξη κοινών δεδομένων, η επικοινωνία μεταξύ modules γίνεται αποκλειστικά μέσω καθολικών (global) μεταβλητών. Εμφανίζεται όταν, για παράδειγμα, η συνάρτηση `functionA()` και η συνάρτηση `functionB()` μπορούν και να διαβάσουν, αλλά και να αλλάξουν μια καθολική μεταβλητή `global value`. Χρησιμοποιείται σε περιπτώσεις που ένα module έχει έναν τεράστιο αριθμό μεταβλητών και πρέπει να αρχικοποιηθούν. Αντί λοιπόν να περαστούν σαν παράμετροι κάθε φορά που χρησιμοποιείται το module, ορίζονται σαν καθολικές μεταβλητές και αρχικοποιούνται έτσι πιο γρήγορα. Από την άλλη, η σύζευξη αυτή εμφανίζει σημαντικά προβλήματα. Αρχικά, αν αλλάξει κάποια καθολική μεταβλητή πρέπει να αλλάξει και κάθε module που τη χρησιμοποιεί. Επίσης, παρατηρείται δυσκολία επαναχρησιμοποίησης, καθώς πρέπει να υπάρχει το ίδιο σύνολο καθολικών μεταβλητών όποτε επαναχρησιμοποιείται το

module. Τέλος, κάθε module έχει πρόσβαση σε περισσότερα δεδομένα από όσα χρειάζεται, γεγονός που κάνει ευκολότερη τη δημιουργία λαθών από απροσεξία.

2.2.3 Σύζευξη Ελέγχου (Control Coupling)

Στην περίπτωση αυτή, έχουμε πέρασμα (ως παράμετρος ή ως καθολική μεταβλητή) «σημαιών ελέγχου» (control flags) έτσι ώστε ένα module να ελέγχει ένα άλλο μέσω αυτών. Συνήθως εμφανίζεται σε modules που παρουσιάζουν λογική συνεκτικότητα (βλ. §2.1.6.). Βασικό της πρόβλημα είναι το γεγονός ότι τα modules δεν είναι ανεξάρτητα. Για να μπορεί ένα module να ελέγχει ένα άλλο, πρέπει πρώτα να γνωρίζει το πώς αυτό λειτουργεί. Η συνθήκη αυτή επηρεάζει και την δυνατότητα επαναχρησιμοποίησης.

2.2.4 Σύζευξη Αντιγράφου (Stamp Coupling)

Πολλές γλώσσες επιτρέπουν μόνο απλές μεταβλητές σαν παραμέτρους σε συναρτήσεις ή μεθόδους (π.χ. το όνομα ενός υπαλλήλου). Υπάρχουν όμως και πολλές γλώσσες που επιτρέπουν το πέρασμα ολόκληρων modules (π.χ. το struct υπάλληλος). Σύζευξη αντιγράφου έχουμε όταν μια μέθοδος A καλεί μια μέθοδο B και περνά ένα ολόκληρο module σαν παράμετρο και, επιπλέον, η B χρειάζεται να ξέρει μόνο συγκεκριμένα πεδία από αυτό το module. Προφανώς, ένα σημαντικό πρόβλημα είναι το ότι οι μέθοδοι αποκτούν πρόσβαση σε δεδομένα τα οποία δεν χρειάζονται. Είναι προτιμότερη από την σύζευξη κοινών δεδομένων (βλ. §2.2.2.) γιατί, σε περίπτωση επεξεργασίας των δεδομένων που μεταφέρονται, έχουμε λιγότερες μεθόδους οι οποίες επηρεάζονται.

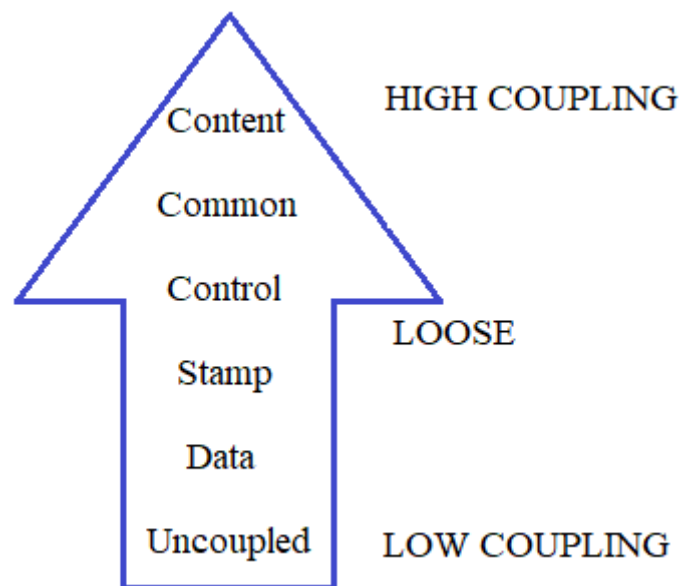
2.2.5 Σύζευξη Δεδομένων (Data Coupling)

Είναι η καλύτερη σύζευξη που μπορούμε να επιτύχουμε (αν δεν λάβουμε υπόψη τη μη σύζευξη που θα αναλύσουμε παρακάτω). Στη σύζευξη δεδομένων, τα modules επικοινωνούν μέσω παραμέτρων. Κάθε παράμετρος αποτελεί ένα κομμάτι δεδομένων και είναι αναγκαία για την παραπάνω επικοινωνία. Ένα συχνό πρόβλημα εμφανίζεται στην περίπτωση που έχουμε πολλές παραμέτρους. Συνήθως, παρατηρείται δυσκολία στην κατανόηση της διεπαφής αυτής και εμφανίζονται προβλήματα.

2.2.6 Μη Σύζευξη (Uncoupled)

Αναφέρεται σε 2 δομές οι οποίες δεν επικοινωνούν καθόλου. Αποτελεί τον πιο ασθενή τύπο σύζευξης αλλά, συνάμα, τον πιο επιθυμητό, καθώς κάνει το λογισμικό μας πολύ πιο απλό.

Το σχήμα 2.4 δείχνει συνοπτικά τα είδη σύζευξης που αναλύσαμε παραπάνω και τα ταξινομεί με βάση την «δύναμή» τους.



Σχήμα 2.4: Ταξινόμηση των Ειδών Σύζευξης [11].

Αφού αναλύσαμε εκτενώς τις έννοιες της σύζευξης και της συνεκτικότητας (και πριν ορίσουμε τις δικές μας μετρικές), καλό είναι να αναφερθούμε και στις μετρικές που είναι ευρέως γνωστές και χρησιμοποιούνται συχνά για τον υπολογισμό των αυτών μεγεθών

2.3 Γνωστές Μετρικές Συνεκτικότητας

2.3.1 Lack of Cohesion of Methods (LCOM)

Το πιο διαδεδομένο είδος μετρικών συνεκτικότητας είναι οι μετρικές LCOM, οι οποίες έχουν προταθεί σε διάφορες εκδοχές από διαφορετικούς συγγραφείς. Αφού μιλάμε για έλλειψη συνεκτικότητας (Lack of Cohesion), καταλαβαίνουμε ότι όσο πιο μικρή είναι η τιμή της, τόσο το καλύτερο.

- Οι Chidamber και Kemerer [1] πρότειναν τη μετρική *LCOM*.

Έστω ότι το σύνολο Q περιέχει τα ζεύγη των μεθόδων μιας κλάσης που χρησιμοποιούν κοινά πεδία. Επίσης, έστω ότι το σύνολο P περιέχει τα ζεύγη των μεθόδων που δεν χρησιμοποιούν κοινά πεδία

Η μετρική LCOM για την κλάση x ορίζεται ως εξής:

$$\begin{aligned} LCOM(x) &= |P| - |Q| && , \text{αν } |P| > |Q| \\ LCOM(x) &= 0 && , \text{αν } |P| \leq |Q| \end{aligned}$$

- Οι Hendersson – Sellers, Constantine και Graham [2], συνέχισαν τη δουλειά των Chidamber, Kemerer και πρότειναν 2 νέες μετρικές συνεκτικότητας, τις *LCOM2* και *LCOM3*.

Έστω m το πλήθος των μεθόδων και a το πλήθος των πεδίων μιας κλάσης. Ως ma_i ορίζουμε το πλήθος των μεθόδων που χρησιμοποιούν το πεδίο a_i (η μεγαλύτερη τιμή που μπορεί να πάρει το ma_i είναι το m, όταν δηλαδή όλες οι μέθοδοι χρησιμοποιούν το πεδίο αυτό).

Η μετρική LCOM2 ορίζεται ως εξής:

$$LCOM2(x) = 1 - \frac{\sum_i m_{ai}}{m \cdot a}$$

Από την άλλη, ο τύπος της μετρικής LCOM3 είναι ο ακόλουθος:

$$LCOM3(x) = \frac{m - \frac{\sum_i m_{ai}}{a}}{m - 1} \quad (\text{ιδανικά } LCOM3 = 0)$$

2.3.2 Tight Class Cohesion / Loose Class Cohesion (TCC / LCC)

Οι Bieman και Kang [3] ανέλυσαν 2 ακόμα μετρικές συνεκτικότητας, τις TCC και LCC.

Για να τις εξηγήσουμε, πρέπει πρώτα να κάνουμε μια εισαγωγή στις έννοιες *άμεση* και *έμμεση σύνδεση*.

Αν μεταξύ 2 μεθόδων μιας κλάσης υπάρχει τουλάχιστον μία κοινή μεταβλητή, τότε οι μέθοδοι αυτοί συνδέονται άμεσα. Από την άλλη, μια μέθοδος M_1 συνδέεται έμμεσα με μια μέθοδο M_n αν υπάρχει ακολουθία μεθόδων M_2, M_3, \dots, M_{n-1} : $M_1 \delta M_2 \dots M_{n-1} \delta M_n$, όπου $M_i \delta M_j$ αναπαριστά μια άμεση σύνδεση.

Έστω λοιπόν NDC το πλήθος των ζευγών μεθόδων που συνδέονται άμεσα και NP το μέγιστο πλήθος άμεσων ή έμμεσων συνδέσεων (αν έχουμε N μεθόδους $\rightarrow NP = \frac{N(N-1)}{2}$).

Η μετρική TCC ορίζεται ως:

$$TCC = \frac{NDC}{NP}$$

Για τον υπολογισμό της LCC χρειαζόμαστε επιπλέον το NIC, που παριστάνει το πλήθος των ζευγών των μεθόδων που ενώνονται άμεσα. Τότε, για την LCC έχουμε:

$$LCC = \frac{NDC + NIC}{NP}$$

2.4 Γνωστές Μετρικές Σύζευξης

Για τον υπολογισμό της σύζευξης χρησιμοποιούνται ευρέως οι μετρικές που έχουν προτείνει οι Chidamber και Kemerer. Οι 2 πιο σημαντικές είναι οι CBO [4] και RFC [5]. Επίσης, συχνή είναι και η χρήση της μετρικής COF, που προτάθηκε από τους Abreu, Goulão και Esteves [6].

2.4.1 Coupling Between Objects (CBO)

Μια κλάση έχει *σύζευξη* με μια άλλη αν μια μέθοδος της πρώτης χρησιμοποιεί μέθοδο της δεύτερης. Για μια κλάση λοιπόν, ως *CBO* θεωρούμε το πλήθος των άλλων κλάσεων με τις οποίες αυτή έχει σύζευξη. Για καλή ποιότητα λογισμικού, χρειαζόμαστε χαμηλό δείκτη CBO.

2.4.2 Response For a Class (RFC)

Έστω ότι το σύνολο $\{M\}$ περιέχει όλες τις μεθόδους μιας κλάσης και το σύνολο $\{R_i\}$ περιέχει τις μεθόδους που καλούνται από την μέθοδο i .

Ως RS ορίζουμε την ένωση όλων των συνόλων $\{R_i\}$ με το $\{M\}$, δηλαδή:

$$RS = \{M\} \cup_{\text{all } i} \{R_i\}$$

Η τιμή της μετρικής RFC προκύπτει από τον παρακάτω τύπο:

$$RFC = |RS|$$

Σημείωση: Συνήθως μετράμε μόνο κλήσεις μεθόδων του ίδιου project. Δεν προσμετράμε κλήσεις standard API.

2.4.3 Coupling Factor (COF)

Έστω N το σύνολο των κλάσεων του συστήματος. Επίσης, έστω ότι $C_i \Rightarrow C_j$ αναπαριστά μια σχέση μεταξύ της κλάσης i και της κλάσης j . Ορίζουμε ως:

$$client(C_i C_j) = \begin{cases} 1, & \text{ανν } C_i \Rightarrow C_j \wedge C_i \neq C_j \\ 0, & \text{αλλιώς} \end{cases}$$

Η τιμή της μετρικής COF υπολογίζεται από τον παρακάτω τύπο:

$$COF = \frac{\sum_{i=1}^N [\sum_{j=1}^N client(C_i C_j)]}{N(N-1)}$$

2.5 Ιδιότητες Μετρικών

Η παραγωγή μετρικών οποιουδήποτε είδους δεν είναι εύκολη διαδικασία. Και αυτό γιατί απαιτείται να ορίζονται με αυστηρό αλλά και ξεκάθαρο τρόπο. Για τον λόγο αυτό, οι Briand, Morasca και Basili [7] πρότειναν κάποιες ιδιότητες που πρέπει να έχουν τα διάφορα είδη μετρικών, μεταξύ των οποίων είναι η σύζευξη και η συνεκτικότητα.

2.5.1 Ιδιότητες Μετρικών Συνεκτικότητας

Στην περίπτωση της συνεκτικότητας, προτείνονται 4 ιδιότητες, τις οποίες αναλύουμε σύντομα παρακάτω:

1. Μη Αρνητικότητα και Κανονικοποίηση (Non Negativity and Normalization)

Διαισθητικά, περιμένουμε να μην εμφανίζονται αρνητικές τιμές συνεκτικότητας (καθώς δεν είναι δυνατό να εξηγήσει κάποιος τι σημαίνει αρνητική συνεκτικότητα). Πιο σημαντικό όμως, θέλουμε οι τιμές να είναι κανονικοποιημένες, έτσι ώστε η μετρική μας να είναι ανεξάρτητη από το μέγεθος του συστήματός μας. Με τη χρήση κανονικοποίησης, είναι ευκολότερη η σύγκριση διαφορετικών συστημάτων, καθώς όλες οι τιμές ανήκουν στο ίδιο διάστημα.

2. Μηδενική Τιμή (Null Value)

Αν γενικά σε ένα σύστημα δεν υπάρχουν εσωτερικές σχέσεις, περιμένουμε η τιμή της μετρικής μας να είναι 0. Και αυτό γιατί δεν υπάρχει καμία σύνδεση μεταξύ των στοιχείων που περιέχονται στον σύστημά μας.

3. Μονοτονικότητα (Monotonicity)

Το να προσθέτουμε επιπλέον σχέσεις στο σύστημα μας δεν θα πρέπει να μειώνει την τιμή της συνεκτικότητας. Αντίθετα, θα πρέπει να την αυξάνει, καθώς οι νέες συνδέσεις αποτελούν επιπλέον αποδείξεις της ύπαρξης σχέσεων μεταξύ των στοιχείων.

4. Συνεκτικές Ενότητες (Cohesive Modules)

Όταν 2 ενότητες (modules) που δεν σχετίζονται μεταξύ τους ενωθούν, η συνεκτικότητα του νέου, διευρυμένου module δεν μπορεί να αυξηθεί, καθώς στοιχεία που δεν σχετίζονται μεταξύ τους απλά ενώνονται.

2.5.2 Ιδιότητες Μετρικών Σύζευξης

Όσον αφορά τη σύζευξη, προτείνονται 5 ιδιότητες, τις οποίες αναφέρουμε συνοπτικά:

1. Μη Αρνητικότητα (Non Negativity)

Όπως και πριν, δεν μπορούμε να εξηγήσουμε διαισθητικά τι σημαίνει αρνητική σύζευξη, συνεπώς δεν πρέπει να έχουμε αρνητική τιμή στην μετρική.

2. Μηδενική Τιμή (Null Value)

Αν μεταξύ modules δεν υπάρχουν καθόλου συνδέσεις, αναμένουμε η σύζευξή τους να είναι 0.

3. Μονοτονικότητα (Monotonicity)

Το να προσθέτουμε επιπλέον σχέσεις στο σύστημα μας δεν θα πρέπει να μειώνει την τιμή της σύζευξης. Αντίθετα, θα πρέπει να την αυξάνει, καθώς οι νέες συνδέσεις αποτελούν επιπλέον αποδείξεις της ύπαρξης σχέσεων μεταξύ των στοιχείων.

4. Ένωση Ενοτήτων (Merging Modules)

Η ένωση 2 modules μπορεί μόνο να μειώσει την τιμή της σύζευξης, καθώς αν υπάρχουν σχέσεις μεταξύ τους πριν την ένωση, αυτές θα χαθούν καθώς τώρα ανήκουν στο ίδιο module.

5. Προσθετικότητα «Κομματιασμένων» Ενοτήτων (Disjoint Module Additivity)

Η τιμή της σύζευξης ενός συστήματος που έχει παραχθεί από την ένωση 2 μη σχετιζομένων υπό-συστημάτων είναι ίση με το άθροισμα των επιμέρους τιμών των συζεύξεων των 2 υπό-συστημάτων.

ΚΕΦΑΛΑΙΟ 3.

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΣΧΗΜΑΤΩΝ

- 3.1 Σημασία της Μοντελοποίησης
 - 3.2 Μοντελοποίηση SQL Σχημάτων
 - 3.3 Ορισμός Γραφημάτων Ενοτήτων
-

3.1 Σημασία της Μοντελοποίησης

Όπως αναφέραμε στα 2 προηγούμενα κεφάλαια, σκοπός μας είναι να μετρήσουμε τη σύζευξη και τη συνεκτικότητα σε βάσεις δεδομένων. Στην προκειμένη περίπτωση, ξεκινώντας από SQL περιγραφές σχημάτων, θέλουμε να παράγουμε μετρικές για τον υπολογισμό των παραπάνω μεγεθών. Στην ενότητα 2.5 τονίσαμε την ανάγκη να ορίζουμε αυστηρά και με ξεκάθαρο τρόπο μια μετρική. Αυτό προϋποθέτει να γνωρίζουμε καλά την εσωτερική δομή των δεδομένων μας. Η δουλειά αυτή δεν μπορεί να γίνει μελετώντας απλά τα SQL σχήματα. Αυτό συμβαίνει γιατί τα σχήματα αυτά, αποτελούν απλά τον κώδικα που υλοποιεί την αποθήκευση πληροφοριών σε μια συγκεκριμένη πλατφόρμα βάσης δεδομένων. Αυτό που χρειαζόμαστε εμείς είναι ένας τρόπος να περιγράψουμε το πώς σχετίζονται τα διάφορα κομμάτια πληροφορίας μεταξύ τους.

Στο σημείο αυτό έρχεται η έννοια της μοντελοποίησης μιας βάσης δεδομένων. Ένα μοντέλο βάσης δεδομένων παρουσιάζει τη λογική δομή μιας βάσης δεδομένων, συμπεριλαμβανομένων των σχέσεων και των περιορισμών που καθορίζουν τον τρόπο αποθήκευσης και πρόσβασης σε δεδομένα. Τα μοντέλα βάσης δεδομένων σχεδιάζονται με βάση τους κανόνες και τις έννοιες του ευρύτερου μοντέλου δεδομένων που υιοθετούν οι εκάστοτε σχεδιαστές.

Για να γίνει ακόμα πιο κατανοητή η σημασία του βήματος αυτού, εμμένουμε λίγο στις διαφορές μεταξύ μοντέλου δεδομένων και σχήματος δεδομένων. Καταρχάς, είναι έως ένα σημείο λογικό κάποιος να συγχέει τις 2 αυτές έννοιες. Και οι 2 έχουν τα δεδομένα ως βασική τους εστίαση (συγκεκριμένα, περιγράφουν τον τρόπο με τον οποίο αποθηκεύονται τα δεδομένα), αλλά καμιά από τις 2 δεν είναι τα δεδομένα αυτά καθαυτά. Η βασική τους διαφορά (όπως είπαμε και πριν) είναι το γεγονός ότι το μοντέλο παρουσιάζει τη λογική δομή της βάσης, ενώ το σχήμα είναι ο κώδικας που αποθηκεύει την πληροφορία με τον τρόπο που υπαγορεύει το μοντέλο. Με πιο απλά λόγια, το μοντέλο είναι ο *ορισμός*, ενώ το σχήμα είναι η *εκτέλεση*.

3.2 Μοντελοποίηση SQL Σχημάτων

Όπως είναι γνωστό, οι βάσεις δεδομένων αποτελούνται από ένα σύνολο πινάκων (*tables*), οι οποίοι με τη σειρά τους περιέχουν μια λίστα από *χαρακτηριστικά* (*attributes*). Επίσης, τα SQL αρχεία περιέχουν *queries* (ερωτήματα) που τροποποιούν τη σχεσιακή βάση δεδομένων (εισαγωγή, διαγραφή ή ενημέρωση δεδομένων).

ΟΡΙΣΜΟΣ (Σχήμα): Ένα σχήμα S ορίζεται ως μια πλειάδα $\langle T, A, R \rangle$, όπου $T = \{t_1, t_2, \dots, t_n\}$ το σύνολο των πινάκων του S , $t_i = \{a_1, a_2, \dots, a_m\}$ το σύνολο των χαρακτηριστικών του πίνακα t_i , $A = \cup t_i$ είναι το σύνολο των χαρακτηριστικών του σχήματος (τα attributes όλων των πινάκων) και R είναι μια δυαδική συσχέτιση στο A ($R \subseteq A \times A$) και απεικονίζει τις σχέσεις μεταξύ των χαρακτηριστικών.

ΟΡΙΣΜΟΣ (Συσχέτιση): Ένα ζεύγος $\langle a_i, a_j \rangle \in R$ αποτελεί μια δυαδική συσχέτιση στο A αν και μόνο αν τα χαρακτηριστικά a_i και a_j εμφανίζονται στο ίδιο query.

Για παράδειγμα, αν έχουμε ένα query της μορφής:

SELECT x, y FROM T

Τότε το ζεύγος $\langle x, y \rangle$ θα ανήκει στο σύνολο R , όπου T ένας πίνακας της βάσης μας.

Από τους παραπάνω ορισμούς, καταλαβαίνουμε ότι τα χαρακτηριστικά ενός πίνακα εμφανίζουν ενώσεις/σχέσεις μεταξύ τους. Οι σχέσεις αυτές είναι 2 διαφορετικών ειδών:

3.2.1 Σχέσεις Μεταξύ Χαρακτηριστικών του Ίδιου Πίνακα

Οι σχέσεις αυτές αναφέρονται στην εσωτερική δομή του πίνακα (δείγμα συνεκτικότητας, όπως είπαμε και στην §2.1.). Το σύνολο $InternalR(t)$ των σχέσεων μεταξύ των χαρακτηριστικών ενός πίνακα $t = \{a_1, a_2, \dots, a_m\} \in T$ ενός σχήματος $S = \langle T, A, R \rangle$ ορίζεται ως:

$$InternalR(t) = \{ \langle a_i, a_j \rangle \in R \mid a_i, a_j \in t \}$$

3.2.2 Σχέσεις Μεταξύ Χαρακτηριστικών του Πίνακα και Χαρακτηριστικών Διαφορετικού Πίνακα

Οι σχέσεις αυτές αναφέρονται στις εξωτερικές συνδέσεις μεταξύ των πινάκων (δείγμα σύζευξης, όπως είπαμε και στην §2.2.). Το σύνολο $ExternalR(t)$ των σχέσεων μεταξύ των χαρακτηριστικών ενός πίνακα $t_i = \{a_1, a_2, \dots, a_m\} \in T$ ενός σχήματος $S = \langle T, A, R \rangle$ και του υπολοίπου σχήματος S ορίζεται ως:

$$ExternalR(t) = \{ t' \in T \mid \exists \langle a_i, a_j \rangle \in R, a_i \in t, a_j \in t', t \neq t' \}$$

3.3 Ορισμός Γραφημάτων Ενοτήτων

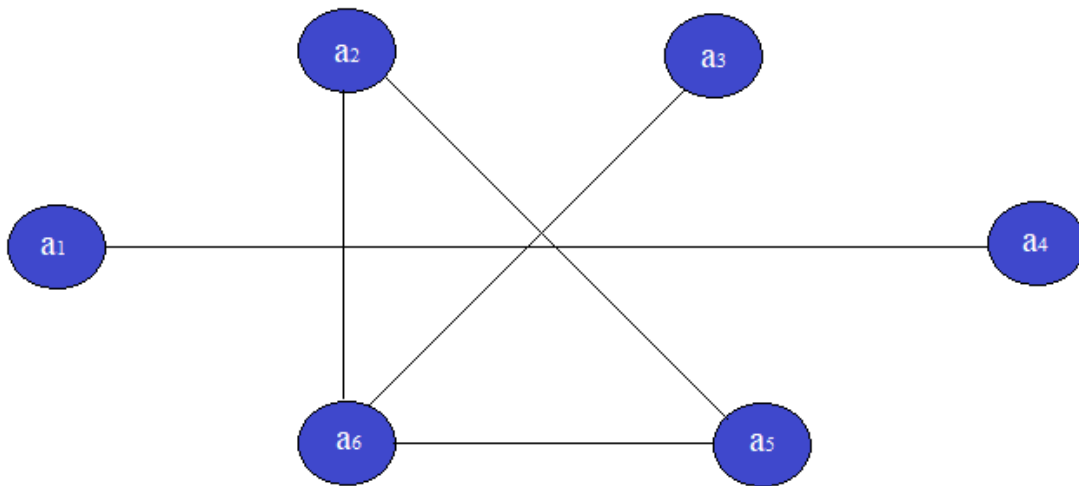
Αφού ορίσαμε το μοντέλο της βάσης δεδομένων μας, πάμε να ορίσουμε και τα γραφήματα πάνω στα οποία θα στηριχθούμε για να ορίσουμε στο επόμενο κεφάλαιο τις μετρικές μας. Όπως είναι γνωστό, για να οριστεί τυπικά ένα γράφημα, πρέπει να ορίσουμε το σύνολο των κόμβων και των ακμών του. Αφού έχουμε να κάνουμε με 2 διαφορετικά μεγέθη, πρέπει να ορίσουμε 2 διαφορετικά γραφήματα (ένα για συνεκτικότητα και ένα για σύζευξη).

3.3.1 Γράφημα Συνεκτικότητας

ΟΡΙΣΜΟΣ (Γράφημα Συνεκτικότητας): Ένα γράφημα συνεκτικότητας $G_m = (V_t, E_t)$ για έναν πίνακα $t_i = \{a_1, a_2, \dots, a_m\} \in T$ ενός σχήματος $S = \langle T, A, R \rangle$ είναι ένα γράφημα χωρίς κατευθύνσεις και βάρη με τις ακόλουθες ιδιότητες:

1. Οι κόμβοι V_t αναπαριστούν τα χαρακτηριστικά του πίνακα t_i .
2. Οι ακμές E_t αναπαριστούν σχέσεις μεταξύ ζευγών από χαρακτηριστικά του πίνακα.
3. Μια ακμή $\langle e_i, e_j \rangle$ ανήκει στο E_t αν και μόνο αν τα στοιχεία e_i και e_j σχετίζονται (βλ. ορισμό στην προηγούμενη σελίδα) και ανήκουν στον πίνακα. Πρακτικά, το σύνολο E_t είναι το σύνολο $\text{InternalR}(t)$ που αναλύσαμε προηγουμένως.

Το σχήμα 3.1 παρακάτω μας δείχνει μια εικόνα του γράφου αυτού. Έχουμε έναν πίνακα με 6 χαρακτηριστικά τα οποία σχετίζονται ανά 2 σύμφωνα με τις ακμές που φαίνονται.



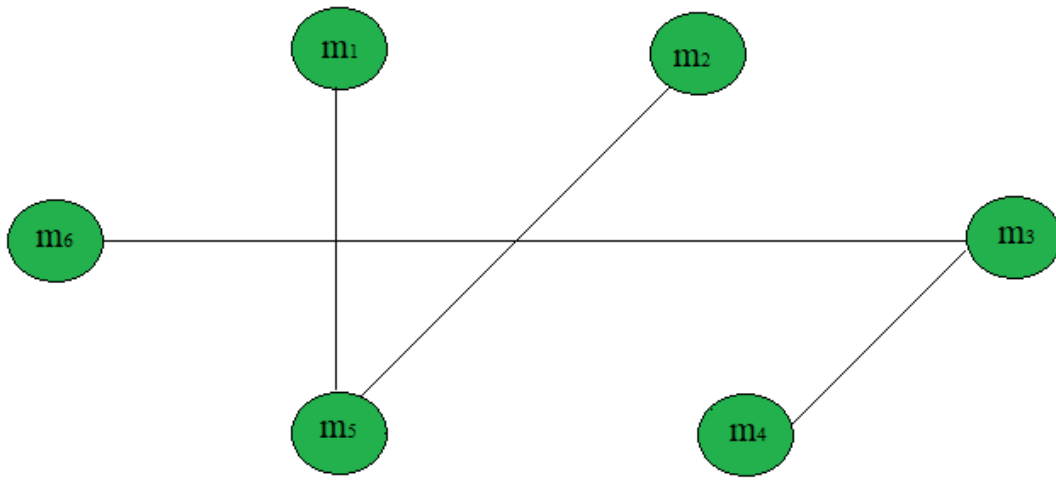
Σχήμα 3.1: Παράδειγμα Γραφήματος Συνεκτικότητας

3.3.2 Γράφημα Σύζευξης

ΟΡΙΣΜΟΣ (Γράφημα Σύζευξης): Ένα γράφημα σύζευξης $G_s = (V_s, E_s)$ για ένα σχήμα $S = \langle T, A, R \rangle$, είναι ένα γράφημα χωρίς κατευθύνσεις και βάρη με τις ακόλουθες ιδιότητες:

1. Οι κόμβοι V_s αναπαριστούν τους διαφορετικούς πίνακες που υπάρχουν στο σχήμα μας.
2. Οι ακμές E_s αναπαριστούν σχέσεις μεταξύ ζευγών πινάκων.
3. Μια ακμή $\langle e_i, e_j \rangle$ ανήκει στο E_s αν και μόνο αν ένα χαρακτηριστικό του πίνακα e_i σχετίζεται (βλ. ορισμό στην προηγούμενη σελίδα) με ένα χαρακτηριστικό του πίνακα e_j . Πρακτικά, το σύνολο E_s είναι η ένωση όλων των συνόλων $\text{ExternalR}(t_i)$ (που αναλύσαμε προηγουμένως), για κάθε πίνακα $t_i \in T$.

Το σχήμα 3.2 παρακάτω μας δείχνει μια εικόνα του γράφου αυτού. Έχουμε ένα σχήμα με 6 πίνακες, των οποίων τα χαρακτηριστικά σχετίζονται ανά 2 σύμφωνα με τις ακμές που φαίνονται.



3.2: Παράδειγμα Γραφήματος Σύζευξης

ΚΕΦΑΛΑΙΟ 4.

ΜΕΤΡΙΚΕΣ ΣΥΝΕΚΤΙΚΟΤΗΤΑΣ

4.1 Ορισμός Μετρικών

4.2 Απόδειξη Ιδιοτήτων Μετρικών

4.1 Ορισμός Μετρικών

Για την μέτρηση της συνεκτικότητας προτείνουμε μια μετρική που προκύπτει μελετώντας τις ακμές μεταξύ των στοιχείων ενός πίνακα.

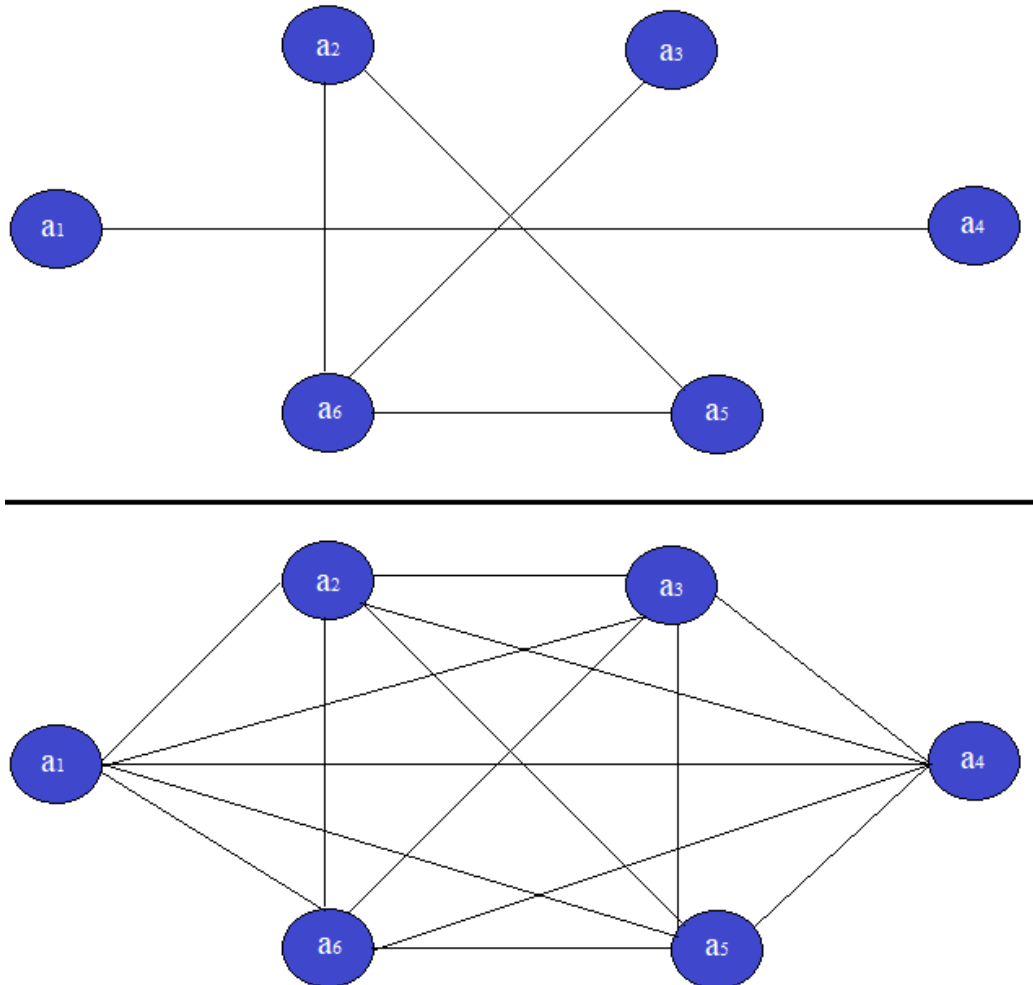
ΟΡΙΣΜΟΣ (Συνεκτικότητα - Cohesion): Έστω το σχήμα $S = \langle T, A, R \rangle$. Η συνεκτικότητα ενός πίνακα $t_i = \{a_1^i, a_2^i, \dots, a_n^i\} \in T$, όπου T το σύνολο των πινάκων του S ορίζεται ως:

$$Cohesion(t_i) = \frac{|InternalR(t_i)|}{\frac{|t_i| |t_i - I|}{2}}$$

Όπου $Internal(t_i)$ το σύνολο των σχέσεων μεταξύ των χαρακτηριστικών του πίνακα t_i (αναλύθηκε στην §3.2).

Πρακτικά, μετράμε τις ακμές του γραφήματος συνεκτικότητας που αναλύσαμε στην §3.3 και συγκρίνουμε το γράφημα αυτό με το αντίστοιχο πλήρες γράφημα (το πλήθος των ακμών ενός

πλήρους γραφήματος με $|t_i|$ κόμβους είναι ίσο με $|t_i| * (|t_i| - 1) / 2$. Για να κατανοήσουμε καλύτερα τον ορισμό αυτό χρησιμοποιούμε το σχήμα 4.1 ως παράδειγμα.



Σχήμα 4.1: Σύγκριση Γραφήματος Συνεκτικότητας με Πλήρες Γράφημα

Στο πάνω μέρος του σχήματος 4.1 φαίνεται το γράφημα συνεκτικότητας ενός πίνακα t_i με 6 χαρακτηριστικά. Στο κάτω μέρος, βλέπουμε το πλήρες γράφημα για τον ίδιο πίνακα. Από το πάνω μέρος παρατηρούμε ότι $|Internal(t_i)| = 5$. Αντίστοιχα, από το κάτω γράφημα παίρνουμε

ότι το πλήθος των γραμμών του πλήρους γραφήματος είναι $\frac{6*5}{2} = 15$. Άρα η τιμή της μετρικής μας για τον πίνακα αυτό είναι:

$$Cohesion(t_i) = \frac{5}{15} = \frac{1}{3}$$

4.2 Απόδειξη Ιδιοτήτων Μετρικών

Αφού ορίσαμε τη μετρική μας, ήρθε η ώρα να αποδείξουμε ότι είναι καλά ορισμένη. Για να το πετύχουμε αυτό θυμίζουμε τις 4 ιδιότητες που πρέπει να έχουν οι μετρικές συνεκτικότητας. Οι ιδιότητες αυτές αναφέρθηκαν συνοπτικά στην §2.5 αλλά εδώ θα αποδειχθούν μαθηματικά.

4.2.1 Μη Αρνητικότητα και Κανονικοποίηση

Η συνεκτικότητα ενός πίνακα ενός πίνακα $t = \{a_1, a_2, \dots, a_m\} \in T$ ενός σχήματος $S = \langle T, A, R \rangle$ πρέπει να ανήκει σε ένα συγκεκριμένο διάστημα $[0, \max]$. Πράγματι από τον ορισμό της μετρικής, βλέπουμε πως η τιμή δεν μπορεί ποτέ να είναι αρνητική, καθώς και ο αριθμητής αλλά και ο παρονομαστής αναφέρεται σε πλήθη. Άρα η ελάχιστη τιμή που μπορεί να πάρει η μετρική μας είναι το 0. Η μέγιστη τιμή έρχεται στην περίπτωση που το γράφημα συνεκτικότητας που παράγουμε συμπίπτει με το πλήρες γράφημα. Στην περίπτωση αυτή λοιπόν έχουμε:

$$|InternalR(t_i)| = \frac{|t_i| |t_i - 1|}{2}$$

Άρα:

$$Cohesion(t_i) = \frac{|InternalR(t_i)|}{\frac{|t_i| |t_i - I|}{2}} = \frac{\frac{|t_i| |t_i - I|}{2}}{\frac{|t_i| |t_i - I|}{2}} = 1$$

Συνεπώς η τιμή της μετρικής μας ανήκει στο διάστημα $[0, 1]$.

4.2.2 Μηδενική Τιμή

Ένας πίνακας $t = \{a_1, a_2, \dots, a_m\} \in T$ ενός σχήματος $S = \langle T, A, R \rangle$ δεν έχει συσχετίσεις αν $R_t = \emptyset$, όπου R_t το σύνολο των δυαδικών συσχετίσεων μεταξύ χαρακτηριστικών του πίνακα t .

Έχουμε: $R_t = \emptyset \Rightarrow InternalR(t) = \emptyset$ καθώς $InternalR(t) \subseteq R_t$. Οπότε:

$$|InternalR(t)| = 0 \Rightarrow Cohesion(t) = 0$$

4.2.3 Μονοτονικότητα

Ένας πίνακας $t_i = \{a_1, a_2, \dots, a_m\} \in T$ ενός σχήματος $S = \langle T, A, R \rangle$. Έστω τώρα ο πίνακας $t_j = \{a_1, a_2, \dots, a_m\} \in T$ (ίδιο σύνολο χαρακτηριστικών) του σχήματος S για τον οποίο ισχύει ότι $InternalR(t_i) \subseteq InternalR(t_j)$ (ο πίνακας t_j περιέχει παραπάνω σχέσεις μεταξύ των χαρακτηριστικών). Έχουμε:

$$InternalR(t_i) \subseteq InternalR(t_j) \Rightarrow |InternalR(t_i)| \leq |InternalR(t_j)|$$

Επίσης:

$$|t_i| = |t_j|$$

Άρα

$$\frac{|InternalR(t_i)|}{\frac{|t_i| |t_i - 1|}{2}} \leq \frac{|InternalR(t_j)|}{\frac{|t_j| |t_j - 1|}{2}} \Rightarrow Cohesion(t_i) \leq Cohesion(t_j)$$

4.2.4 Συνεκτικές Ενότητες

Ένας πίνακας $t_i, t_j \in T$ ενός σχήματος $S = \langle T, A, R \rangle$, για τους οποίους ισχύει:

$ExternalR(t_i) \cap ExternalR(t_j) = \emptyset$, πράγμα που σημαίνει ότι οι 2 αυτοί πίνακες δεν σχετίζονται μεταξύ τους.

Επίσης, έστω $t_k \in T$ για τον οποίο ισχύουν τα εξής:

$$t_k = t_i \cup t_j$$

$$R_{t_k} = R_{t_i} \cup R_{t_j}$$

Από τη στιγμή που οι 2 πίνακες δεν σχετίζονται μεταξύ τους, έχουμε:

$$InternalR(t_k) = InternalR(t_i) + InternalR(t_j) \Rightarrow$$

$$|InternalR(t_k)| = |InternalR(t_i) + InternalR(t_j)| = |InternalR(t_i)| + |InternalR(t_j)|$$

$$\text{Επίσης: } t_k = t_i + t_j \Rightarrow |t_k| = |t_i + t_j| = |t_i| + |t_j|$$

Άρα:

$$\begin{aligned}
Cohesion(t_k) &= \frac{|InternalR(t_k)|}{\frac{|t_k| |t_k - 1|}{2}} = \frac{|InternalR(t_i) + InternalR(t_j)|}{\frac{|t_i + t_j| |t_i + t_j - 1|}{2}} \\
&= \frac{|InternalR(t_i)| + |InternalR(t_j)|}{\frac{|t_i + t_j| |t_i + t_j - 1|}{2}} \\
&= \frac{|InternalR(t_i)|}{\frac{|t_i + t_j| |t_i + t_j - 1|}{2}} + \frac{|InternalR(t_j)|}{\frac{|t_i + t_j| |t_i + t_j - 1|}{2}} \\
&\leq \frac{|InternalR(t_i)|}{\frac{|t_i| |t_i - 1|}{2}} + \frac{|InternalR(t_j)|}{\frac{|t_j| |t_j - 1|}{2}} \\
&= Cohesion(t_i) + Cohesion(t_j)
\end{aligned}$$

ΚΕΦΑΛΑΙΟ 5.

ΜΕΤΡΙΚΕΣ ΣΥΖΕΥΞΗΣ

5.1 Ορισμός Μετρικών

5.2 Απόδειξη Ιδιοτήτων Μετρικών

5.1 Ορισμός Μετρικών

Για την μέτρηση της σύζευξης προτείνουμε 2 μετρικές που προκύπτουν μελετώντας τις εξωτερικές ακμές ενός πίνακα, τις ακμές λοιπόν που ενώνουν στοιχείο που ανήκει σε αυτόν και στοιχείο που ανήκει σε άλλον πίνακα του σχήματος.

5.1.1 Η Μετρική Coupling

ΟΡΙΣΜΟΣ (Σύζευξη - Coupling): Έστω το σχήμα $S = \langle T, A, R \rangle$. Η σύζευξη ενός πίνακα $t_i = \{a_1^i, a_2^i, \dots, a_n^i\} \in T$, όπου T το σύνολο των πινάκων του S ορίζεται ως:

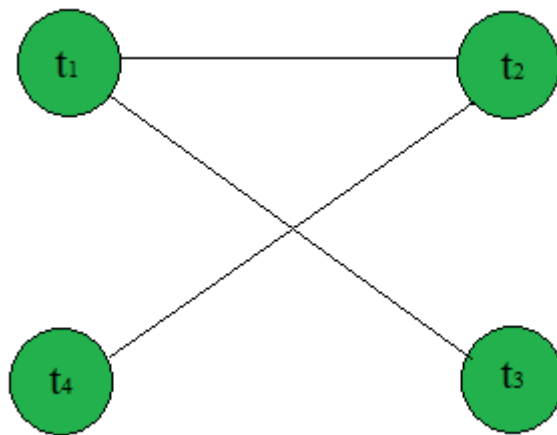
$$Coupling(t_i) = |ExternalR(t_i)|$$

όπου $ExternalR(t_i)$ το σύνολο των εξωτερικών σχέσεων του πίνακα t_i (αναλύθηκε στην §3.2).

Πρακτικά, για κάθε πίνακα του σχήματος, μετράμε τις ακμές του στο γράφημα σύζευξης που αναλύσαμε στην §3.3. Για να κατανοήσουμε καλύτερα τον ορισμό αυτό χρησιμοποιούμε το σχήμα 5.1 ως παράδειγμα.

Στο σχήμα αυτό βλέπουμε την εικόνα ενός σχήματος που αποτελείται από 4 πίνακες. Η τιμή της σύζευξης κάθε πίνακα είναι ίση με το πλήθος των ακμών προς άλλους πίνακες. Στην προκειμένη περίπτωση έχουμε:

$$\text{Coupling}(t_1) = 2, \text{Coupling}(t_2) = 2, \text{Coupling}(t_3) = 1, \text{Coupling}(t_4) = 1.$$



Σχήμα 5.1: Γράφημα Σύζευξης Τυχαίου Σχήματος

5.1.2 Η Μετρική Normalized_Coupling

ΟΡΙΣΜΟΣ (Κανονικοποιημένη Σύζευξη - Coupling): Έστω το σχήμα $S = \langle T, A, R \rangle$. Η σύζευξη ενός πίνακα $t_i = \{a_1^i, a_2^i, \dots, a_n^i\} \in T$, ορίζεται ως:

$$\text{Normalized_Coupling}(t_i) = \frac{|\text{External}R(t_i)|}{|T| - 1}$$

όπου $\text{ExternalR}(t_i)$ το σύνολο των εξωτερικών σχέσεων του πίνακα t_i (αναλύθηκε στην §3.2) και T το σύνολο των πινάκων στο σχήμα S (προφανώς επιβάλλουμε η τιμή του T να είναι μεγαλύτερη του 1, καθώς διαφορετικά δεν μπορούμε να μιλάμε για σύζευξη μεταξύ 2 πινάκων).

Πρακτικά, για κάθε πίνακα του σχήματος μετράμε τις ακμές του στο γράφημα σύζευξης και το συγκρίνουμε με το μέγιστο πλήθος των ακμών που θα μπορούσε να έχει (που είναι το πλήθος των πινάκων -1 καθώς δεν μπορούμε να έχουμε σύνδεση πίνακα με τον εαυτό του). Για να κατανοήσουμε καλύτερα τον ορισμό αυτό χρησιμοποιούμε πάλι το σχήμα 5.1 ως παράδειγμα.

Στο σχήμα αυτό το πλήθος των πινάκων είναι 4. Έτσι, η τιμή της μετρικής για τους πίνακες αυτούς είναι:

$\text{Normalized_Coupling}(t_1) = 2/3$, $\text{Normalized_Coupling}(t_2) = 2/3$, $\text{Normalized_Coupling}(t_3) = 1/3$, $\text{Normalized_Coupling}(t_4) = 1/3$.

5.2 Απόδειξη Ιδιοτήτων Μετρικών

Αφού ορίσαμε τις μετρικές μας, ήρθε η ώρα να αποδείξουμε ότι είναι καλά ορισμένες. Για να το πετύχουμε αυτό θυμίζουμε τις 5 ιδιότητες που πρέπει να έχουν οι μετρικές σύζευξης. Οι ιδιότητες αυτές αναφέρθηκαν συνοπτικά στην §2.5 αλλά εδώ θα αποδειχθούν μαθηματικά.

5.2.1 Μη Αρνητικότητα

5.2.1.1 Μετρική Coupling

Η σύζευξη ενός πίνακα $t_i = \{a_1, a_2, \dots, a_n\}$ πρέπει να μην έχει αρνητικές τιμές. Πράγματι από τον ορισμό της μετρικής, βλέπουμε πως η τιμή δεν μπορεί ποτέ να είναι αρνητική, καθώς αναφερόμαστε σε πλήθη συνόλων. Άρα η ελάχιστη τιμή που μπορεί να πάρει η μετρική μας είναι το 0.

5.2.1.2 Μετρική Normalized_Coupling

Εντελώς αντίστοιχα με την §5.2.1.1, ο αριθμητής αλλά και ο παρονομαστής αναφέρονται σε σύνολα, οπότε η ελάχιστη τιμή που μπορεί να πάρει η μετρική μας είναι το 0.

5.2.2 Μηδενική Τιμή

5.2.2.1 Μετρική Coupling

Ένας πίνακας $t = \{a_1, a_2, \dots, a_m\} \in T$ ενός σχήματος $S = \langle T, A, R \rangle$ δεν έχει συσχετίσεις αν $R_t = \emptyset$, όπου R_t το σύνολο των δυαδικών συσχετίσεων μεταξύ χαρακτηριστικών του πίνακα t .

Έχουμε: $R_t = \emptyset \Rightarrow \text{ExternalR}(t) = \emptyset$ καθώς $\text{ExternalR}(t) \subseteq R_t$. Οπότε:

$$|\text{ExternalR}(t)| = 0 \Rightarrow \text{Coupling}(t) = 0$$

5.2.2.2 Μετρική Normalized_Coupling

Εντελώς αντίστοιχα με την §5.2.2.1, για έναν πίνακα χωρίς εξωτερικές σχέσεις, η μετρική Normalized_Coupling παίρνει την τιμή:

$$Normalized_Coupling(t) = \frac{|ExternalR(t)|}{|T| - 1} = \frac{|\emptyset|}{|T| - 1} = 0$$

5.2.3 Μονοτονικότητα

Ένας πίνακας $t_i = \{a_1, a_2, \dots, a_m\} \in T$ ενός σχήματος $S = \langle T, A, R \rangle$. Έστω τώρα ο πίνακας $t_j = \{a_1, a_2, \dots, a_m\} \in T$ (ίδιο σύνολο χαρακτηριστικών) του σχήματος S για τον οποίο ισχύει ότι $ExternalR(t_i) \subseteq ExternalR(t_j)$ (ο πίνακας t_j περιέχει παραπάνω εξωτερικές σχέσεις). Έχουμε:

$$ExternalR(t_i) \subseteq ExternalR(t_j) \Rightarrow |ExternalR(t_i)| \leq |ExternalR(t_j)|$$

5.2.3.1 Μετρική Coupling

Η μετρική για τον πίνακα t_j έχει την τιμή:

$$Coupling(t_j) = |ExternalR(t_j)| \geq |ExternalR(t_i)| \geq Coupling(t_i)$$

5.2.3.2 Μετρική Normalized_Coupling

Αντίστοιχα, η τιμή της μετρικής αυτής για τον πίνακα t_i είναι:

$$\begin{aligned} Normalized_Coupling(t_j) &= \frac{|ExternalR(t_j)|}{|T| - 1} \\ &\geq \frac{|ExternalR(t_i)|}{|T| - 1} \\ &\geq Normalized_Coupling(t_i) \end{aligned}$$

5.2.4 Ένωση Ενοτήτων

Έστω 2 πίνακες του σχήματος $S = \langle T, A, R \rangle$, $t_i = \{a_1, a_2, \dots, a_m\}$ και $t_j = \{a_1, a_2, \dots, a_m\}$. Διακρίνουμε 2 περιπτώσεις.

- Οι 2 αυτοί πίνακες σχετίζονται

Αυτό σημαίνει ότι $ExternalR(t_i) \cap ExternalR(t_j) \neq \emptyset$. Ο νέος πίνακας t_k που προκύπτει ως η ένωση των πινάκων θα έχει $ExternalR(t_k)$ μικρότερο από το άθροισμα των $ExternalR(t_i)$ και $ExternalR(t_j)$, καθώς οι ακμές που ανήκουν στο $ExternalR(t_i) \cap ExternalR(t_j)$ αποτελούν πλέον εσωτερικές συνδέσεις και ανήκουν πλέον στο $InternalR(t_k)$. Συνεπώς $|ExternalR(t_k)| < |ExternalR(t_i)| + |ExternalR(t_j)|$.

Άρα, για την μετρική Coupling έχουμε:

$$Coupling(t_k) = |ExternalR(t_k)| < |ExternalR(t_i)| + |ExternalR(t_j)|.$$

Αντίστοιχα, για την μετρική Normalized_Coupling έχουμε:

$$\begin{aligned} Normalized_Coupling(t_k) &= \frac{|ExternalR(t_k)|}{|T| - 1} \\ &\leq \frac{|ExternalR(t_i)|}{|T| - 1} + \frac{|ExternalR(t_j)|}{|T| - 1} \\ &\leq Normalized_Coupling(t_i) + Normalized_Coupling(t_j) \end{aligned}$$

- Οι 2 πίνακες δεν σχετίζονται

Σε αυτή την περίπτωση $ExternalR(t_i) \cap ExternalR(t_j) = \emptyset$. Αυτό σημαίνει ότι ο νέος πίνακας t_k θα έχει σαν $ExternalR(t_k)$ το άθροισμα των $ExternalR(t_i)$ και $ExternalR(t_j)$, καθώς καμία εξωτερική σχέση δεν μετατρέπεται σε εσωτερική όπως πριν. Πράγμα που σημαίνει ότι:

$$Coupling(t_k) = |ExternalR(t_k)| = |ExternalR(t_i)| + |ExternalR(t_j)|.$$

Επίσης:

$$\begin{aligned}
\text{Normalized_Coupling}(t_k) &= \frac{|ExternalR(t_k)|}{|T| - 1} \\
&= \frac{|ExternalR(t_i)|}{|T| - 1} + \frac{|ExternalR(t_j)|}{|T| - 1} \\
&= \text{Normalized_Coupling}(t_i) + \text{Normalized_Coupling}(t_j)
\end{aligned}$$

5.2.5 Προσθετικότητα «Κομματιασμένων» Ενοτήτων

Η τιμή της σύζευξης ενός πίνακα που έχει παραχθεί από την ένωση 2 μη σχετιζομένων πινάκων είναι ίση με το άθροισμα των επιμέρους τιμών των συζεύξεων των 2 υποσυστημάτων. Η περίπτωση αυτή αποδεικνύεται στην παραπάνω ενότητα (§5.2.4), όπου 2 πίνακες t_i και t_j που δεν σχετίζονται ενώνονται σε έναν νέο πίνακα t_k . Έχουμε λοιπόν:

$$Coupling(t_k) = |ExternalR(t_k)| = |ExternalR(t_i)| + |ExternalR(t_j)|$$

Επίσης:

$$\begin{aligned}
\text{Normalized_Coupling}(t_k) &= \frac{|ExternalR(t_k)|}{|T| - 1} \\
&= \frac{|ExternalR(t_i)|}{|T| - 1} + \frac{|ExternalR(t_j)|}{|T| - 1} \\
&= \text{Normalized_Coupling}(t_i) + \text{Normalized_Coupling}(t_j)
\end{aligned}$$

ΚΕΦΑΛΑΙΟ 6.

ΕΜΠΕΙΡΙΚΗ ΜΕΛΕΤΗ

- 6.1 Σχετικά με τα Δεδομένα
 - 6.2 Βασικά Στατιστικά
 - 6.3 Συσχετισμός Μετρικών με Άλλα Στοιχεία
 - 6.4 Θέματα προς Εγκυρότητα
-

6.1 Σχετικά με τα Δεδομένα

Για τα πειράματά μας, πήραμε ως δεδομένα τη συλλογή των Sharma, Fragkoulis, Rizou, Bruntink και Spinellis [8]. Αποτελούνται από SQL περιγραφές σχημάτων από διάφορα open-source projects στο github. Η αρχική πηγή είχε ένα σύνολο από περίπου 2568 διαφορετικά αρχεία SQL. Από αυτά, περισσότερα από τα μισά (1480 από αυτά) δεν μπορούσαν να περάσουν τη φάση της ανάλυσης. Αυτό συνέβη γιατί τα ονόματα των πινάκων και των χαρακτηριστικών τους ήταν παραμετροποιημένα και η ανάγνωσή τους γινόταν μόνο μέσω συναρτήσεων του εργαλείου που χρησιμοποίησαν οι συγγραφείς για την εξαγωγή των SQL σχημάτων (ένα όνομα πίνακα είναι για παράδειγμα το αποτέλεσμα μιας μεθόδου getName() που δεν μπορούμε να γνωρίζουμε πού κλήθηκε και ποιο ήταν το αποτέλεσμα που επιστράφη). Οπότε η χρήση τους ήταν αδύνατη. Από τα υπόλοιπα, 230 σχήματα δεν

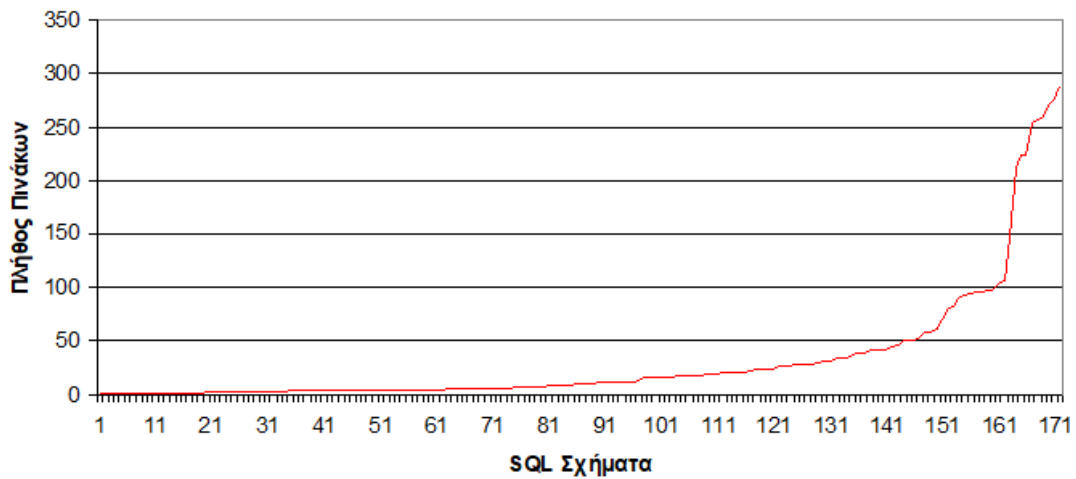
περιείχαν πληροφορία. Δηλαδή, το αποτέλεσμα της εξαγωγής query που πραγματοποίησαν οι συγγραφείς δεν έφερε αποτέλεσμα (το SQL αρχείο έγραφε απλά None). Τέλος, σχεδόν 700 είχαν πληροφορία η οποία χαρακτηρίζεται ως ελλιπής. Αυτό συνέβη γιατί στα αρχεία αυτά είτε είχαμε δημιουργία πολλών πινάκων αλλά καθόλου queries, είτε είχαμε queries χωρίς τη δημιουργία πινάκων. Θα μπορούσαμε να χρησιμοποιήσουμε τα μισά περίπου από αυτά (300 σχήματα), τα οποία δεν είχαν καθόλου queries και να θεωρήσουμε απλά πως η σύζευξη και η συνεκτικότητα τους είναι 0, αλλά θεωρήσαμε σωστό να τα παραλείψουμε ως ελλιπή. Μετά από την επεξεργασία, καταλήξαμε σε 172 SQL περιγραφές σχημάτων, τα οποία θα αποτελέσουν το σύνολο των δεδομένων μας. Ο πίνακας 6.1 συγκεντρώνει τα ακριβή στοιχεία της παραπάνω διαδικασίας.

Κατηγορία Αρχείων	Πλήθος Αρχείων
Χρήσιμα (το σύνολο δεδομένων μας)	172
Παραμετροποιημένη Πληροφορία	1480
Καθόλου Πληροφορία	230
Ελλιπής Πληροφορία	686
Σύνολο	2568

Πίνακας 6.1: Πληροφορίες των Δεδομένων

Τα 172 σχήματα που θα χρησιμοποιήσουμε είναι διαφόρων μεγεθών. Ως μέγεθος του σχήματος θεωρούμε το πλήθος των πινάκων που το απαρτίζουν. Το σχήμα 6.1 που παρουσιάζεται παρακάτω δείχνει το πλήθος των πινάκων ανά σχήμα. Παρατηρούμε πως τα σχήματα περιέχουν ως επί το πλείστον λίγους πίνακες (μονοψήφιο αριθμό), ενώ υπάρχουν λίγα σχήματα με πλήθος πινάκων πάνω από 100.

Μέγεθος Σχήματος



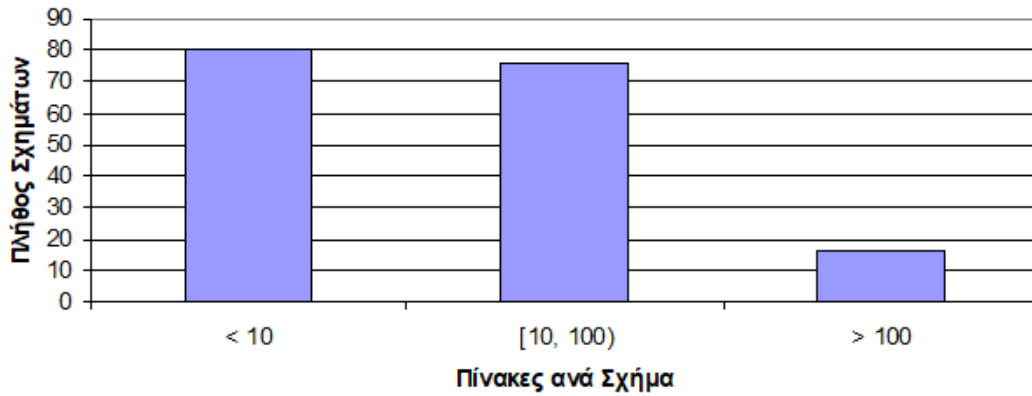
Σχήμα 6.1: Πλήθος Πινάκων ανά Σχήμα

Στο σχήμα 6.2 γίνεται η κατηγοριοποίηση των σχημάτων με βάση το πλήθος των πινάκων τους. Η κατηγορίες είναι 3:

- Σχήματα με μονοψήφιο αριθμό πινάκων.
- Σχήματα με διψήφιο αριθμό πινάκων.
- Σχήματα με τριψήφιο αριθμό πινάκων.

Από το σχήμα αυτό είναι ξεκάθαρο πως σχεδόν τα μισά σχήματα των δεδομένων μας περιέχουν μονοψήφιο αριθμό πινάκων.

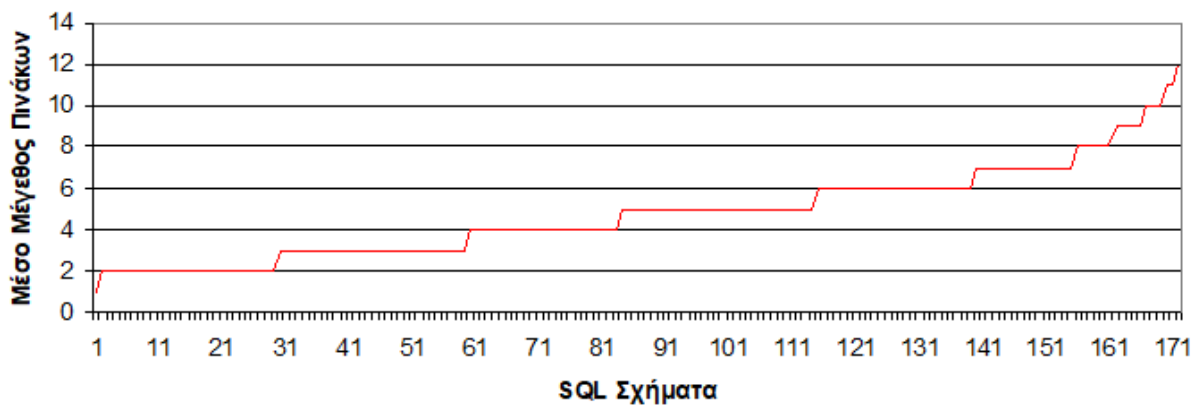
Πλήθος Πινάκων ανά Σχήμα



Σχήμα 6.2: Κατηγοριοποίηση Σχημάτων με Βάση το Πλήθος Πινάκων

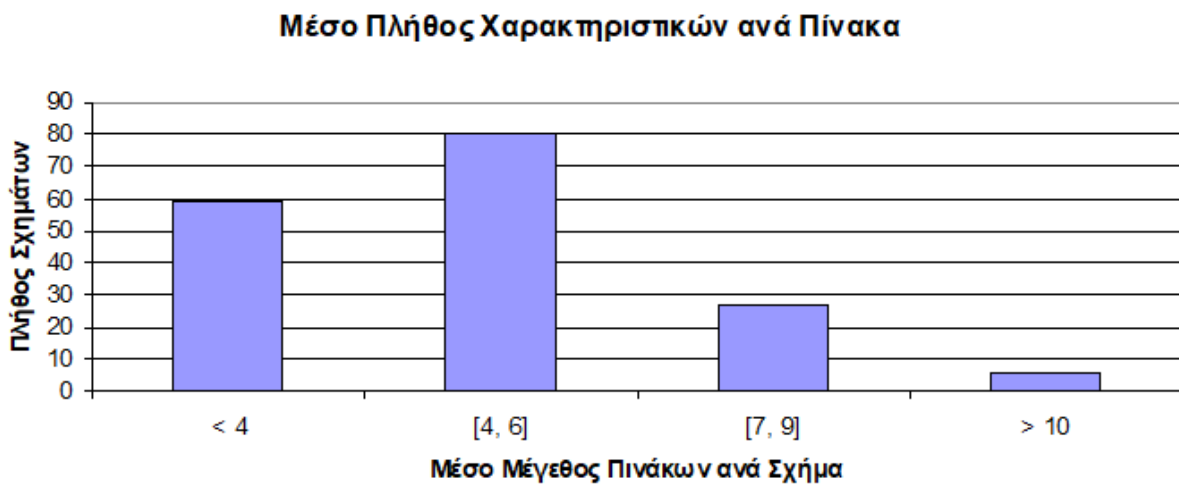
Το τελικό μας σύνολο δεδομένων αποτελείται από 5624 πίνακες. Όπως και στα σχήματα, έτσι και οι πίνακες είναι διαφόρων μεγεθών. Ως μέγεθος ενός πίνακα θεωρούμε το πλήθος των χαρακτηριστικών του. Στο σχήμα 6.3 φαίνεται η μέση τιμή του μεγέθους των πινάκων ανά σχήμα.. Παρατηρώντας το σχήμα αυτό, καταλαβαίνουμε πως οι πίνακες είναι γενικά μικρού μεγέθους.

Μέσο Μέγεθος Πινάκων ανά Σχήμα



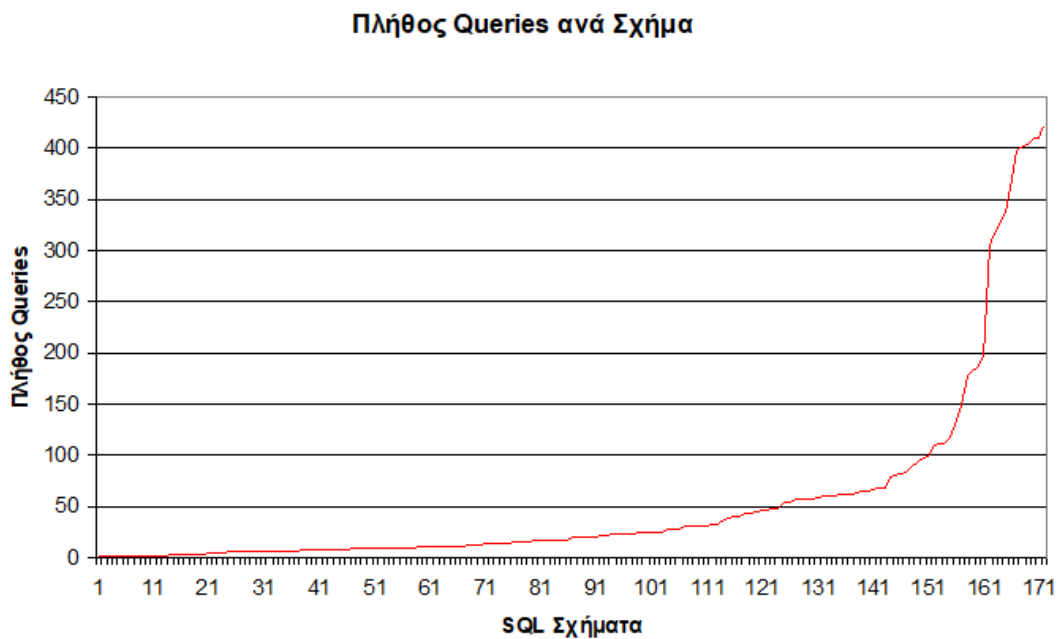
Σχήμα 6.3: Μέσο Μέγεθος Πινάκων ανά Σχήμα

Στο σχήμα 6.4 γίνεται η κατηγοριοποίηση των σχημάτων με βάση το μέσο μέγεθος των πινάκων τους. Από το σχήμα αυτό είναι ξεκάθαρο πως σχεδόν όλα τα σχήματα του συνόλου δεδομένων μας περιέχουν μικρούς πίνακες (έως και 6 χαρακτηριστικά ανά πίνακα κατά μέσο όρο).



Σχήμα 6.4: Κατηγοριοποίηση Σχημάτων με Βάση το Μέσο Μέγεθος Πινάκων

Τέλος, στο σχήμα 6.5 βλέπουμε το πλήθος των queries ανά σχήμα. Παρατηρούμε ότι τα περισσότερα σχήματα έχουν μέχρι 50 queries, ενώ λιγότερα από 15 σχήματα (11 συγκεκριμένα), περιέχουν τριψήφιο αριθμό από queries.



Σχήμα 6.5: Πλήθος Queries ανά Σχήμα

6.2 Βασικά Στατιστικά

Για κάθε έναν από τους 5624 πίνακες που έχουμε στο σύνολο δεδομένων υπολογίζουμε τις τιμές των 3 μετρικών μας (Cohesion, Coupling, Normalized_Coupling). Στον πίνακα 6.2 φαίνεται η τιμή των 3 μετρικών για διάφορα βασικά στατιστικά, όπως η μέση τιμή, η μέγιστη και ελάχιστη, η ενδιάμεση, η τυπική απόκλιση και το mode (η πιο συχνή τιμή).

Μετρική	Average	Max	Min	Median	Mode	Standard Deviation
Cohesion	0.443	1.0	0.0	0.33	0.0	0.47
Coupling	0.057	10	0	0	0	0.46
Normalized_Coupling	0.004	1.0	0.0	0.0	0.0	0.04

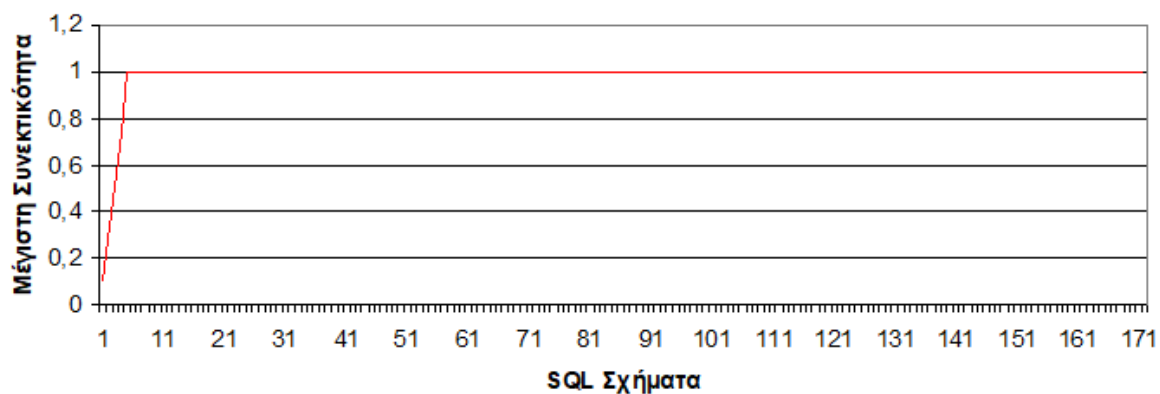
Πίνακας 6.2: Στατιστικά Μετρικών

Για να κατανοήσουμε περισσότερο τα αποτελέσματα, σε κάθε μια από τις μετρικές, για κάθε ένα από τα στοιχεία του πίνακα 6.2 παρουσιάζουμε ένα σχήμα που δείχνει την τιμή του στατιστικού ανά σχήμα.

6.2.1 Μετρική Cohesion

Τα σχήματα 6.6, 6.7 6.8, 6.9, 6.10 και 6.11 παρουσιάζουν την μέγιστη, ελάχιστη, μέση, median, mode τιμή αλλά και την τυπική απόκλιση ανά σχήμα αντίστοιχα.

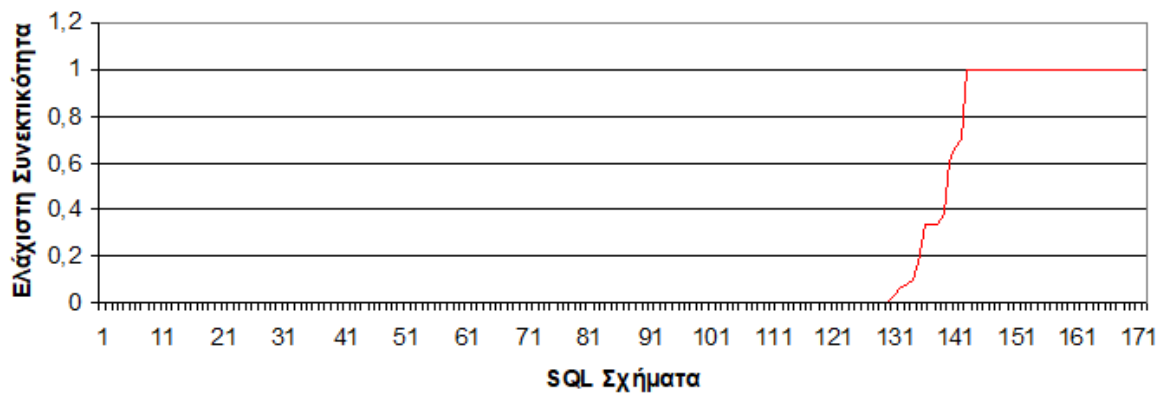
Μέγιστη Συνεκτικότητα ανά Σχήμα



Σχήμα 6.6: Μέγιστη Συνεκτικότητα ανά Σχήμα

Το σχήμα 6.6 παρουσιάζει τη μέγιστη συνεκτικότητα ανά σχήμα. Παρατηρούμε πως σχεδόν σε όλα τα σχήματα, υπάρχει ένας πίνακας που εμφανίζει τη μέγιστη δυνατή συνεκτικότητα.

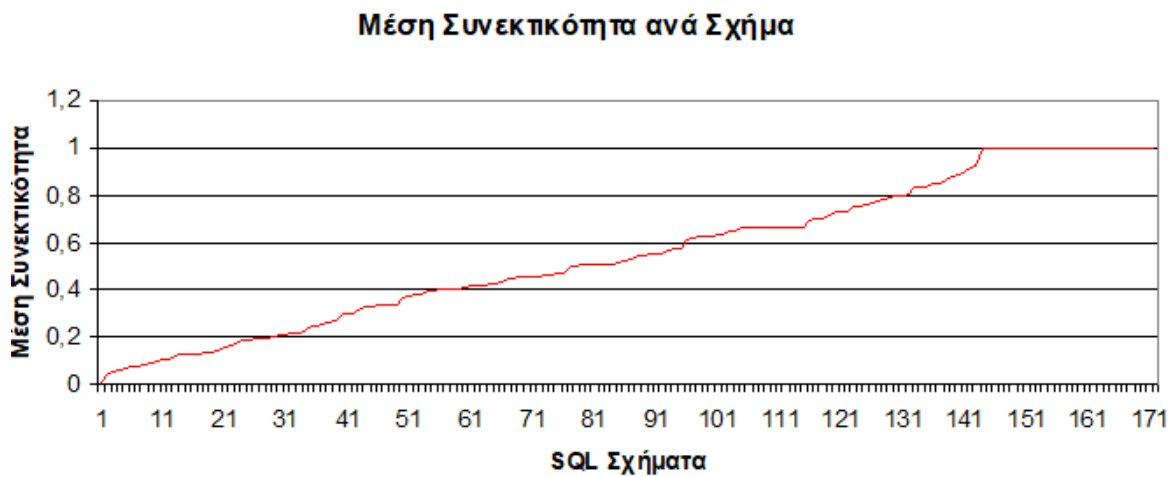
Ελάχιστη Συνεκτικότητα ανά Σχήμα



Σχήμα 6.7: Ελάχιστη Συνεκτικότητα ανά Σχήμα

Αντίστοιχα, από το σχήμα 6.7, που παρουσιάζει την ελάχιστη συνεκτικότητα ανά σχήμα, παρατηρούμε πως σχεδόν σε κάθε σχήμα, υπάρχει ένας πίνακας με μηδενική συνεκτικότητα.

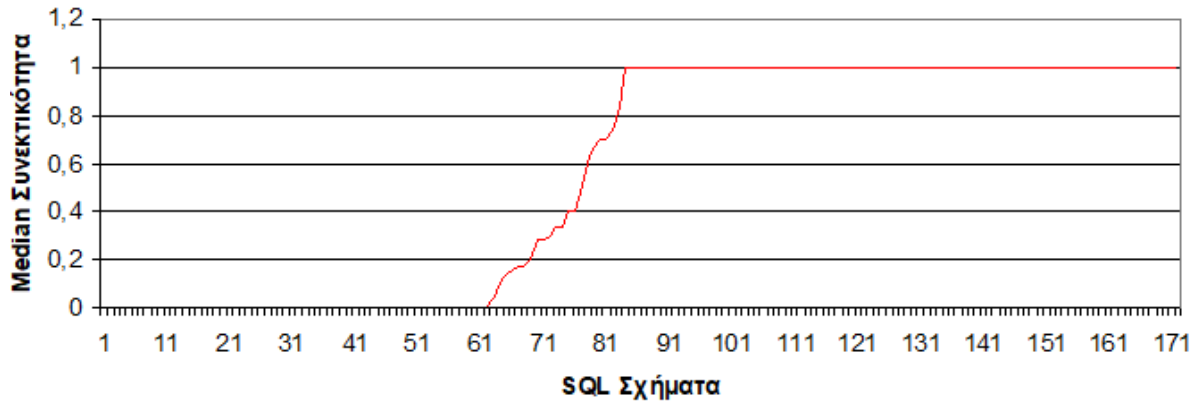
Υπάρχουν επίσης κάποιες περιπτώσεις, που οι πίνακες όλου του σχήματος έχουν τη μέγιστη συνεκτικότητα (και για αυτό η ελάχιστη τιμή είναι 0).



Σχήμα 6.8: Μέση Συνεκτικότητα ανά Σχήμα

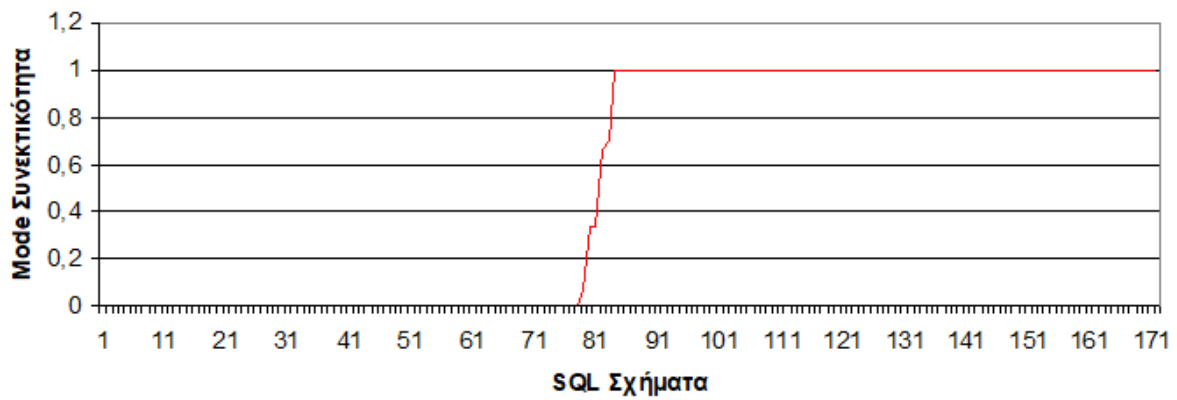
Βλέποντας το σχήμα 6.8, που παρουσιάζει τη μέση συνεκτικότητα ανά σχήμα, βλέπουμε πως υπάρχουν αρκετά σχήματα τα οποία έχουν χαμηλή συνεκτικότητα. Παρόλα αυτά, οι πλειοψηφία των σχημάτων έχει κατά μέσο όρο τιμή μεγαλύτερη από το 0.443, που είναι η μέση τιμή της μετρικής, σύμφωνα με τον πίνακα 6.2 που είδαμε πριν.

Ενδιάμεση (Median) Συνεκτικότητα ανά Σχήμα



Σχήμα 6.9: Median Συνεκτικότητα ανά Σχήμα

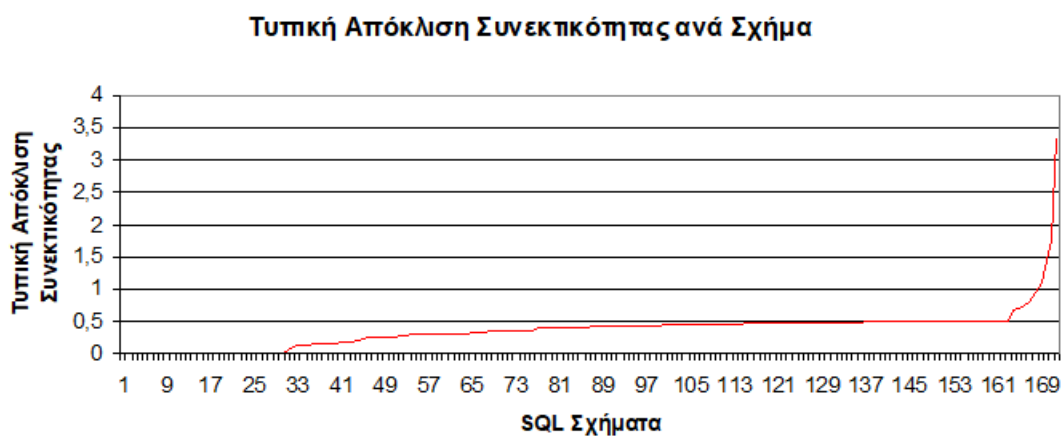
Mode Συνεκτικότητα ανά Σχήμα



Σχήμα 6.10: Mode Συνεκτικότητα ανά Σχήμα

Τα σχήματα 6.9 και 6.10 δείχνουν την median και την mode συνεκτικότητα ανά σχήμα αντίστοιχα. Αυτό που παρατηρούμε είναι ότι σε πάρα πολλά σχήματα, η πιο συχνή τιμή της συνεκτικότητας είναι το 0. Αυτό σημαίνει, πως υπάρχουν πάρα πολύ πίνακες οι οποίοι είτε δεν χρησιμοποιούνται καθόλου, είτε χρησιμοποιούνται από queries τα οποία ζητούν μόνο ένα

χαρακτηριστικό του πίνακα και όχι κάποιο ζεύγος χαρακτηριστικών. Μετά από μεγαλύτερη έρευνα, παρατηρούμε πως, από το σύνολο των 5624 πινάκων του σχήματος, οι 2275 έχουν συνεκτικότητα 0, οι 2251 έχουν συνεκτικότητα 1 (το γράφημα συνεκτικότητάς τους συμπίπτει με τον πλήρη γράφο → πλήρως συνεκτικός πίνακας) . Οι υπόλοιποι 598 πίνακες έχουν κάποια ενδιάμεση τιμή συνεκτικότητας.



Σχήμα 6.11: Τυπική Απόκλιση Συνεκτικότητας ανά Σχήμα

Τέλος, το σχήμα 6.11 παρουσιάζει την τυπική απόκλιση της συνεκτικότητας ανά σχήμα. Παρατηρούμε ότι σε λίγες περιπτώσεις η τυπική απόκλιση του σχήματος ανεβαίνει σε σχετικά υψηλές τιμές.

6.2.2 Μετρικές Coupling και Normalized_Coupling

Ο πίνακας 6.2 έδειξε πως, κατά μέσο όρο η σύζευξη είναι ιδιαίτερα χαμηλή, πράγμα που είναι σημαίνει πως γενικά έχουμε απλά queries που δεν έχουν JOIN. Από το σύνολο των 172 σχημάτων, μόλις τα 34 από αυτά παρουσιάζουν σύζευξη. Έχουν δηλαδή τουλάχιστον ένα ζεύγος πινάκων οι οποίοι συνδέονται μέσω ενός query. Για να εμβαθύνουμε περισσότερο στη σύζευξη, παρουσιάζουμε τον πίνακα 6.3, ο οποίος παρουσιάζει τη μέση τιμή των 2 μετρικών, αυτή τη φορά στο σύνολο των 34 σχημάτων που παρουσιάζουν σύζευξη. Το υποσύνολο δεδομένων αυτό αποτελείται από 3108 πίνακες.

Μετρική	Μέση Τιμή
Coupling	0.1
Normalized_Coupling	0.006

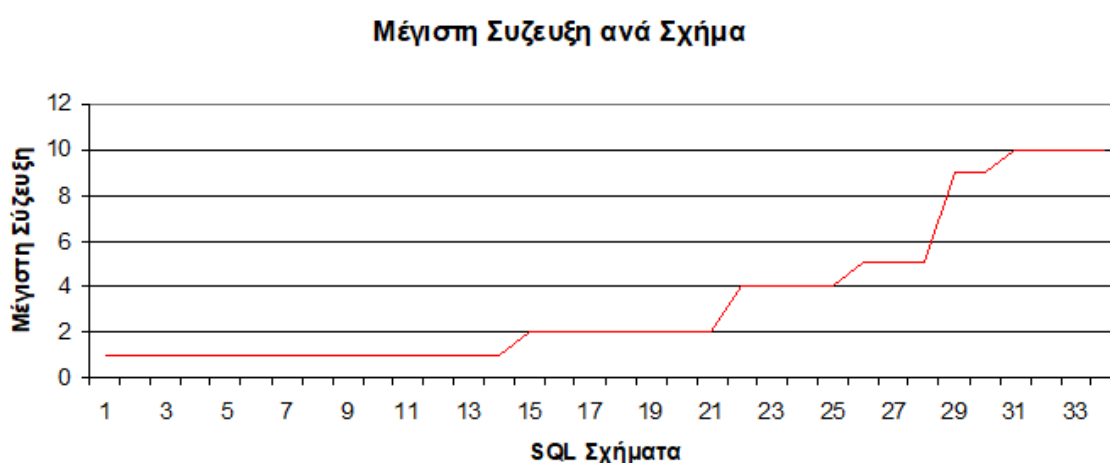
Πίνακας 6.3: Μέση Τιμή Μετρικών στο Υποσύνολο Σύζευξης

Οι τιμές του πίνακα 6.3 δείχνουν πως, ακόμα και αν συγκεντρωθούμε στα σχήματα που παρουσιάζουν σύζευξη, η μέση τιμή των μετρικών μας παραμένει πολύ χαμηλή. Θυμίζουμε πως η τιμή της μετρικής Normalized_Coupling ανήκει στο $[0, 1]$, ενώ αυτή της μετρικής Coupling μπορεί να είναι οποιοσδήποτε θετικός αριθμός.

Για να εμβαθύνουμε πιο πολύ στην έννοια της σύζευξης, παρουσιάζουμε διάφορα διαγράμματα στη συνέχεια. Σαν σύνολο σχημάτων παίρνουμε μόνο τα 34 σχήματα που παρουσιάζουν σύζευξη. Τα υπόλοιπα 138 σχήματα έχουν σύζευξη(κανονική και normalized) 0, οπότε δεν τα χρησιμοποιούμε.

6.2.2.1 Μετρική Coupling

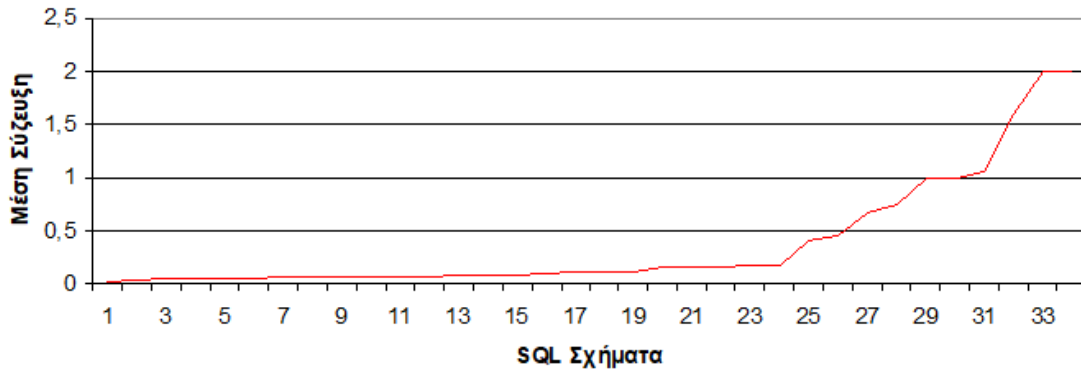
Για την μετρική αυτή παρουσιάζουμε τα σχήματα 6.12, 6.13 και 6.14, που παρουσιάζουν τη μέγιστη, την μέση τιμή και την τυπική απόκλιση της μετρικής ανά σχήμα αντίστοιχα. Η ελάχιστη, η ενδιάμεση και η mode τιμή δεν παρουσιάζονται, καθώς οι περισσότερες τιμές είναι 0 και δεν εμφανίζουν ενδιαφέρον.



Σχήμα 6.12: Μέγιστη Σύζευξη ανά Σχήμα

Το σχήμα 6.12 δείχνει πως η μέγιστη τιμή της σύζευξης ανά σχήμα είναι στις περισσότερες περιπτώσεις μικρή. Σε περισσότερα από τα μισά σχήματα η μέγιστη τιμή είναι μικρότερη ή ίση από 2. Ενδιαφέρον παρουσιάζουν οι (λίγες) περιπτώσεις που η μέγιστη τιμή είναι μεγάλη (πάνω από 5). Αυτό σημαίνει ότι κάποιος πίνακας επικοινωνεί με πάνω από 5 πίνακες του σχήματός του. Μετά από μελέτη, προκύπτει πως οι περιπτώσεις αυτές εμφανίζονται σε σχήματα με πολύ μεγάλο αριθμό πινάκων (τριψήφιο αριθμό) και συνήθως, ένας βασικός πίνακας συνεργάζεται με αρκετούς υπο-πίνακες.

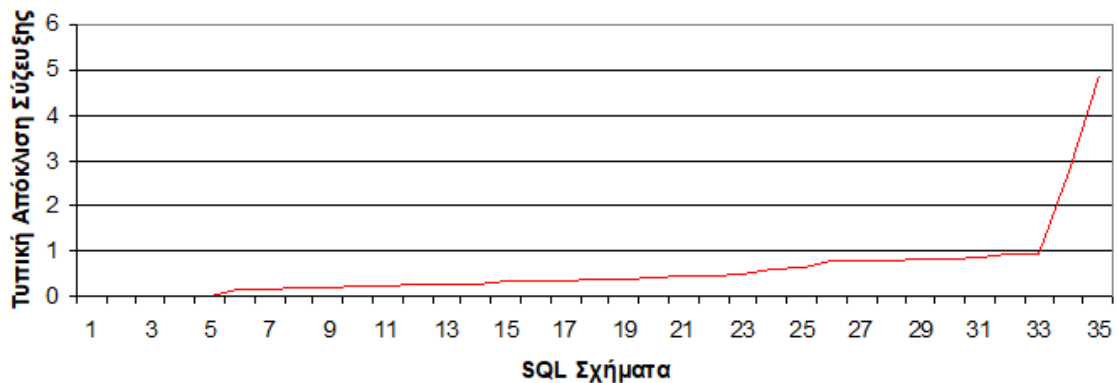
Μέση Σύζευξη ανά Σχήμα



Σχήμα 6.13: Μέση Σύζευξη ανά Σχήμα

Το σχήμα 6.13 δείχνει τη μέση τιμή της σύζευξης ανά σχήμα. Όπως και στο σχήμα 6.11, έτσι και εδώ η μέση τιμή της σύζευξης είναι στην πλειοψηφία των σχημάτων μικρή. Ενδιαφέρον πάλι εμφανίζεται στην περίπτωση των (λίγων σχημάτων) που εμφανίζουν υψηλή τιμή. Αυτό συμβαίνει σε περιπτώσεις μικρών σχημάτων (πολύ λίγοι πίνακες), οι οποίοι όμως συνεργάζονται συχνά (δηλαδή ο γράφος σύζευξης του σχήματος είναι αρκετά πυκνός).

Τυπική Απόκλιση Σύζευξης ανά Σχήμα

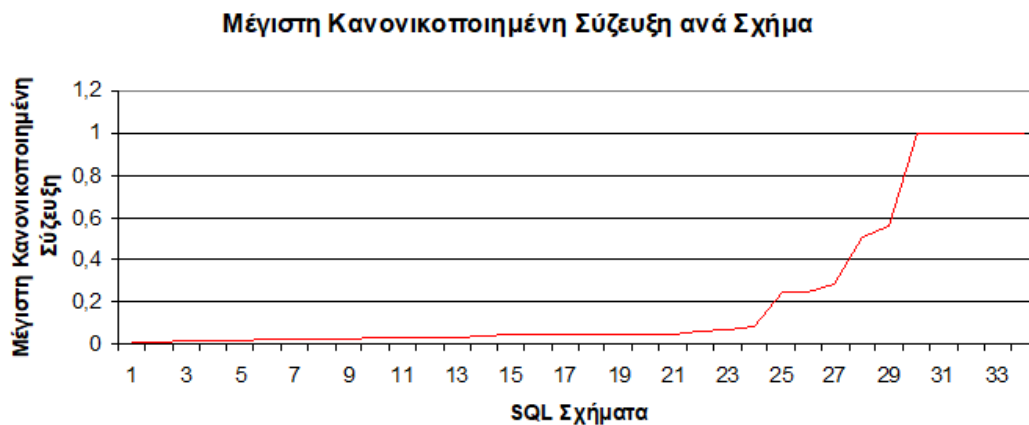


Σχήμα 6.14: Τυπική Απόκλιση Σύζευξης ανά Σχήμα

Από το σχήμα 6.14 παρατηρούμε ότι στα περισσότερα σχήματα οι τιμές της σύζευξης των πινάκων είναι σχετικά κοντά, με αποτέλεσμα η τυπική απόκλιση να είναι μικρή. Εξαιρέση αποτελούν κάποιες περιπτώσεις, στις οποίες κάποιοι πίνακες του σχήματος έχουν υψηλή σύζευξη, με αποτέλεσμα να απέχουν από τη μέση τιμή και να προκαλούν αύξηση της τυπικής απόκλισης.

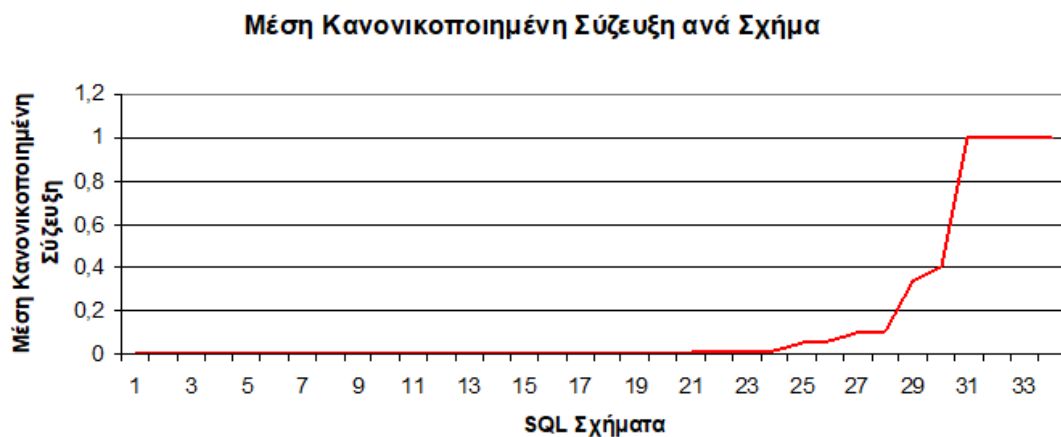
6.2.2.2 Μετρική Normalized_Coupling

Όπως και πριν, έτσι και εδώ παρουσιάζουμε τη μέγιστη, την μέση τιμή και την τυπική απόκλιση της μετρικής ανά σχήμα. Η ελάχιστη, η ενδιάμεση και η mode τιμή δεν παρουσιάζονται, καθώς οι περισσότερες τιμές είναι 0 και δεν εμφανίζουν ενδιαφέρον.



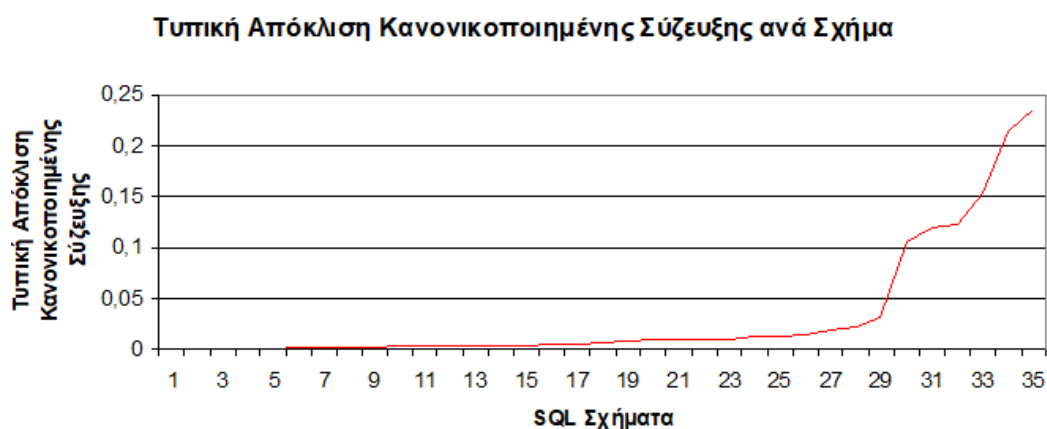
Σχήμα 6.15: Μέγιστη Κανονικοποιημένη Σύζευξη ανά Σχήμα

Από το σχήμα 6.15 βλέπουμε πως οι τιμές στην πλειοψηφία των σχημάτων είναι μικρές. Οι περιπτώσεις των μεγάλων τιμών (θυμίζουμε η μέγιστη τιμή είναι το 1) είναι σχετικά λίγες και εμφανίζονται σε μικρά σχήματα (όπως δηλαδή αναλύσαμε και στην περίπτωση της μέσης σύζευξης ανά σχήμα → βλ. σχήμα 6.13).



Σχήμα 6.16: Μέση Κανονικοποιημένη Σύζευξη ανά Σχήμα

Όπως και στο σχήμα 6.15, έτσι και εδώ τα περισσότερα σχήματα παρουσιάζουν μικρές τιμές σύζευξης. Οι λίγες περιπτώσεις που έχουν μεγάλη τιμή αφορούν (όπως και πριν), μικρά σχήματα (λίγοι πίνακες) αλλά ιδιαίτερα πυκνά ως προς τον γράφο σύζευξής τους.



Σχήμα 6.17: Τυπική Απόκλιση Κανονικοποιημένης Σύζευξης ανά Σχήμα

Τέλος, από το σχήμα 6.17 παρατηρούμε ότι στα περισσότερα σχήματα οι τιμές της σύζευξης των πινάκων είναι σχετικά κοντά, με αποτέλεσμα η τυπική απόκλιση να είναι μικρή. Εξάιρεση αποτελούν κάποιες περιπτώσεις, στις οποίες κάποιοι πίνακες του σχήματος έχουν υψηλή σύζευξη, με αποτέλεσμα να απέχουν από τη μέση τιμή και να προκαλούν αύξηση της τυπικής απόκλισης

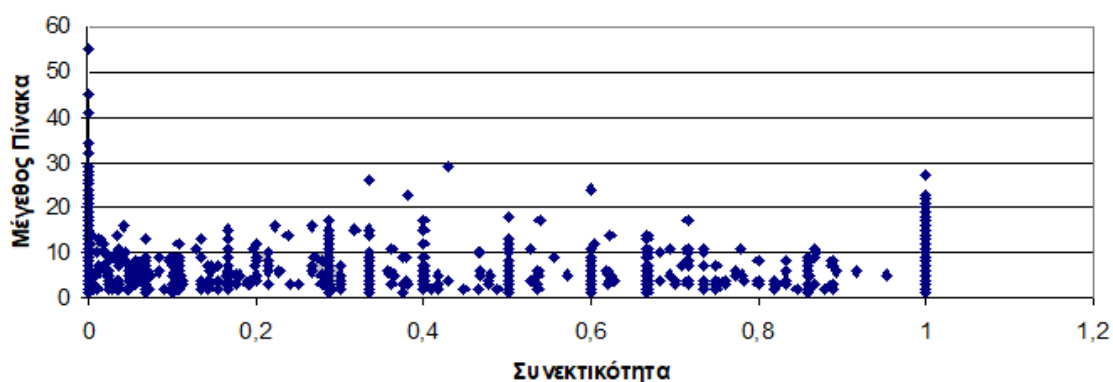
6.3 Συσχετισμός Μετρικών με Άλλα Στοιχεία

Στην ενότητα αυτή διεξάγουμε ένα πείραμα για κάθε μετρική. Θέλουμε να δούμε αν το μέγεθος των σχημάτων (πλήθος πινάκων) ή το μέγεθος των πινάκων (πλήθος χαρακτηριστικών) επηρεάζουν την τιμή των μετρικών (ελέγχουμε αν, για παράδειγμα, οι πίνακες μεγάλου μεγέθους έχουν γενικά μεγάλες τιμές συνεκτικότητας). Τα πειράματα συσχετισμών που εκτελέσαμε είναι τα ακόλουθα 3:

6.3.1 Συσχετισμός Συνεκτικότητας – Μεγέθους Πίνακα

Στο πείραμα αυτό προσπαθήσαμε να δούμε αν το μέγεθος ενός πίνακα μπορεί να μας προιδεάσει για την τιμή της συνεκτικότητάς του. Για να το πετύχουμε αυτό, συγκεντρώνουμε την τιμή της συνεκτικότητας κάθε πίνακα και την συσχετίζουμε με το μέγεθος του πίνακα αυτού. Το αποτέλεσμα του πειράματος φαίνεται στο σχήμα 6.18.

Συσχετισμός Συνεκτικότητας - Μεγέθους Πίνακα

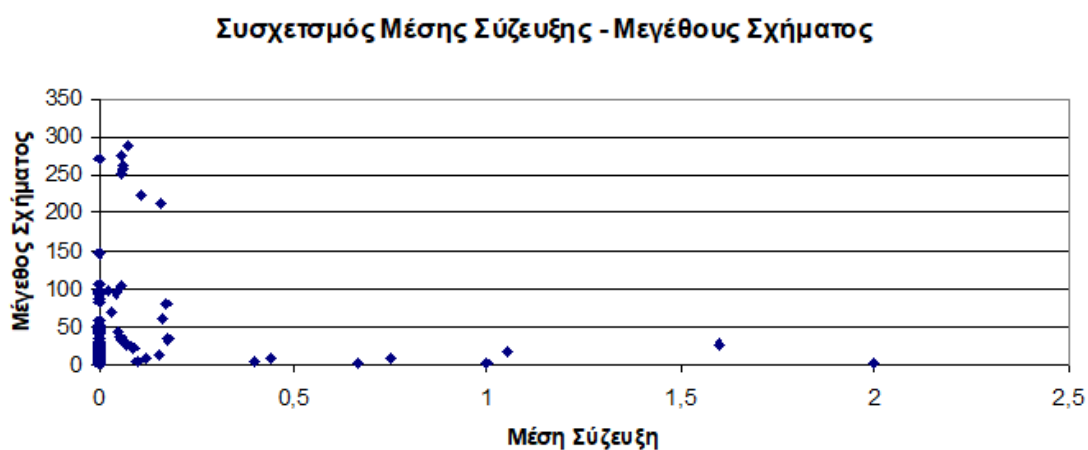


Σχήμα 6.18: Συσχετισμός Συνεκτικότητας – Μεγέθους Πίνακα

Παρατηρώντας το σχήμα 6.18, δεν βλέπουμε κάποιον συσχετισμό μεταξύ των 2 αυτών στοιχείων. Βλέπουμε ότι πίνακες με γενικά μικρό μέγεθος μπορεί να έχουν διάφορες τιμές συνεκτικότητας. Ένα πράγμα που είναι ίσως ορατό είναι το γεγονός πως οι πολλοί μεγάλοι πίνακες (με πάνω από 30 χαρακτηριστικά) έχουν πολύ χαμηλή τιμή συνεκτικότητας. Οι πίνακες αυτοί βέβαια είναι πολύ λίγοι, επομένως δεν μπορούμε να είμαστε σίγουροι για το αν η θεωρία αυτή γενικεύεται.

6.3.2 Συσχετισμός Σύζευξης – Μεγέθους Σχήματος

Στο πείραμα αυτό προσπαθήσαμε να δούμε αν το μέγεθος ενός σχήματος μπορεί να μας προϋδεάσει για την τιμή της σύζευξής του. Για να το πετύχουμε αυτό, συγκεντρώνουμε την μέση τιμή της σύζευξης των πινάκων κάθε σχήματος και την συσχετίζουμε με το μέγεθος του σχήματος αυτού. Το αποτέλεσμα του πειράματος φαίνεται στο σχήμα 6.18.

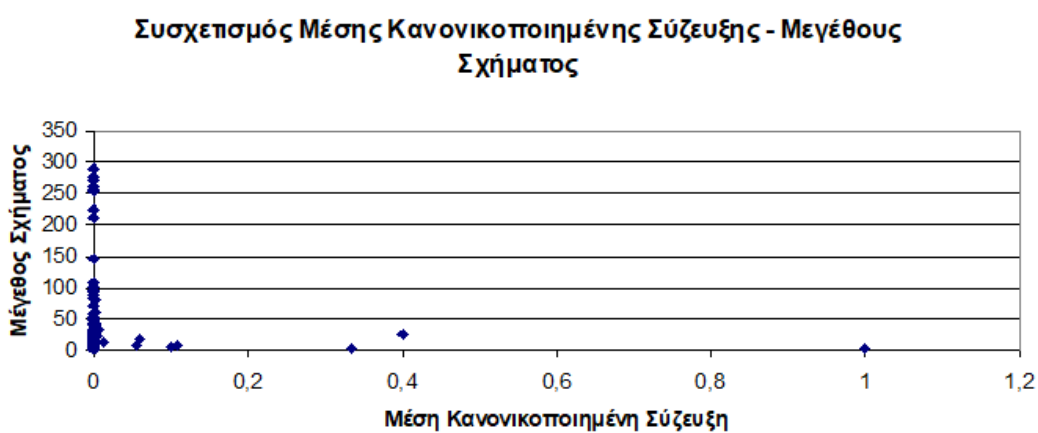


Σχήμα 6.19: Συσχετισμός Μέσης Σύζευξης – Μεγέθους Σχήματος

Παρατηρώντας το σχήμα 6.19, το μόνο ενδιαφέρον που προκύπτει είναι το γεγονός πως όλα τα σχήματα μεγάλου μεγέθους έχουν μικρή μέση τιμή σύζευξης. Αυτό βέβαια δεν είναι σίγουρο ότι μπορεί να γενικευτεί, καθώς σχεδόν όλα τα σχήματα έχουν χαμηλή μέση τιμή σύζευξης.

6.3.3 Συσχετισμός Κανονικοποιημένης Σύζευξης – Μεγέθους Σχήματος

Το πείραμα αυτό επαναλαμβάνει τη διαδικασία της προηγούμενης ενότητας, χρησιμοποιώντας αυτή τη φορά την μέση κανονικοποιημένη σύζευξη ανά σχήμα. Το αποτέλεσμα του πειράματος φαίνεται στην εικόνα 6.20. Οι παρατηρήσεις είναι ίδιες με το πείραμα 6.3.3.



Σχήμα 6.20: Συσχετισμός Μέσης Κανονικοποιημένης Σύζευξης – Μεγέθους Σχήματος

6.4 Θέματα προς την Εγκυρότητα (Threads To Validity)

6.4.1 Εγκυρότητα Κατασκευής (Construct Validity)

Ένα σημαντικό θέμα εδώ είναι το γεγονός ότι η εξαγωγή των queries από δεδομένα ανοιχτού λογισμικού δεν είναι 100% αξιόπιστη. Ανάλογα με τη μέθοδο που χρησιμοποιείται, πολλά queries μπορεί να μην είναι ολοκληρωμένα, επειδή οι πληροφορίες καθορίζονται κατά τη διάρκεια της εκτέλεσης, ενώ ορισμένες άλλα πιθανόν να μην ανακαλυφθούν.

6.4.2 Εσωτερική Εγκυρότητα (Internal Validity)

Παρατηρούμε μόνο τα δεδομένα και αναφέρουμε τις παρατηρήσεις μας. Δεν επιχειρούμε να διαπιστώσουμε κάποια σχέση όσον αφορά τα αποτελέσματα. Επομένως, δεν υπάρχει απειλή για την εσωτερική εγκυρότητα.

6.4.3 Εξωτερική Εγκυρότητα (External Validity)

Εξετάσαμε έναν μεγάλο αριθμό σχημάτων ανοιχτού λογισμικού. Η παρατηρήσεις μας μπορούν να γενικευθούν σε αυτό το πλαίσιο. Ωστόσο, δεν μελετήσαμε βιομηχανικές μελέτες περιπτώσεων. Σε τέτοιες περιπτώσεις τα αποτελέσματα μπορεί να είναι διαφορετικά.

ΚΕΦΑΛΑΙΟ 7.

ΕΠΙΛΟΓΟΣ

7.1 Επίλογος

7.2 Μελλοντική Δουλειά

7.1 Επίλογος

Στην παρούσα μεταπτυχιακή εργασία μελετήσαμε τις έννοιες της σύζευξης και της συνεκτικότητας σε σχήματα βάσεων δεδομένων. Συγκεκριμένα, ξεκινώντας από SQL περιγραφές σχημάτων, παρήγαμε γραφήματα σύζευξης/συνεκτικότητας. Κατόπιν προτείναμε μετρικές για την αξιολόγηση των παραπάνω εννοιών. Ορίσαμε τις προτεινόμενες μετρικές φορμαλιστικά και αποδείξαμε ότι είναι καλά ορισμένες, με βάση ένα σύνολο από ιδιότητες που πρέπει να πληρούνται από τις μετρικές ποιότητας λογισμικού γενικότερα. Τέλος, με βάση τις προτεινόμενες μετρικές, πραγματοποιήσαμε μια εκτενή εμπειρική μελέτη σε ένα σύνολο σχημάτων που προέρχονται από συστήματα ανοιχτού λογισμικού.

7.2 Μελλοντική Δουλειά

Όπως είπαμε και στο προηγούμενο κεφάλαιο, θα μπορούσαμε να χρησιμοποιήσουμε και δεδομένα από βιομηχανικές μελέτες περιπτώσεων, ώστε τα πειράματά μας να είναι πιο ολοκληρωμένα. Επίσης, θα μπορούσαμε να κάνουμε μια προσπάθεια συσχέτισης των μετρικών μας με τα database smells που ανίχνευσαν οι , Fragkoulis, Rizou, Bruntink και Spinellis [8]. Τέλος, θα μπορούσαμε να προτείνουμε κάποια refactorings που θα μπορούσαν να βοηθήσουν στην αύξηση της ποιότητας της βάσης δεδομένων που αναλύσαμε.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] S.R. Chidamber and C.F. Kemerer. A metrics suite for object oriented design. IEEE Transactions on Software Engineering, vol. 20, pp. 476–493, 1994.
- [2] B. Henderson-Sellers, L. L. Constantine, and I. M. Graham. Coupling and cohesion (towards a valid metrics suite for objectoriented analysis and design). Object Oriented Systems, vol. 3, no. 3, pp. 143–158, 1996.
- [3] Bieman, J., Kang, B. Cohesion and reuse in an object-oriented system. In: Proceedings of the 1995 Symposium on Software Reusability, Seattle, Washington, United States, pp.259–262, 1995.
- [4] S.R. Chidamber and C.F. Kemerer. Towards a metrics suite for object-oriented design. In Object Oriented Programming Systems Languages and Applications, Phoenix, Arizona, USA, pp.197-211, 1991.
- [5] S.R. Chidamber and C.F. Kemerer. A metrics suite for object-oriented design. IEEE Transactions on Software Engineering, vol.20, no.6, pp.467-493, 1994.

- [6] F. Abreu, M. Goulo, and R. Esteves. Toward the design quality evaluation of object-oriented software systems. In Fifth International Conference on Software Quality, Austin, Texas, USA, pp.44-57, 1995.
- [7] LIONEL C. BRIAND, SANDRO MORASCA, VICTOR R. BASILI. Property-Based Software Engineering Measurement. IEEE TRANS. SOFTWARE ENG. 22(1): 68-86 , 1996.
- [8] TUSHAR SHARMA, MARIOS FRAGKOULIS, STAMATIA RIZOU, MAGIEL BRUNTINK, DIOMIDIS SPINELLIS. Smelly relations: measuring and understanding database schema quality relations: measuring and understanding database schema quality. ICSE (SEIP) : 55-64, 2018
- [9] [HTTPS://ELEARNING.TEICM.GR/FILE.PHP/22/2014/7DESIGN.PDF](https://elearning.teicm.gr/file.php/22/2014/7design.pdf)
- [10] [HTTP://PAGES.CPSC.UCALGARY.CA/~EBERLY/COURSES/CPSC333/LECTURES/DESIGN/COHESION.HTML](http://pages.cpsc.ucalgary.ca/~eberly/courses/cpsc333/lectures/design/cohesion.html)
- [11] W.P. Stevens, G.J. Myers, L.L. Constantine: Structured Design, IBM Systems Journal, vol 13, no. 2, pp. 115-139, 1974
- [12] [HTTPS://WWW.AVATTO.COM/STUDY-MATERIAL/SOFTWARE-ENGINEERING-COHESION](https://www.avatto.com/study-material/software-engineering-cohesion)
- [13] [HTTPS://STACKOVERFLOW.COM/QUESTIONS/14000762/WHAT-DOES-LOW-IN-COUPPING-AND-HIGH-IN-COHESION-MEAN](https://stackoverflow.com/questions/14000762/what-does-low-in-coupling-and-high-in-cohesion-mean)

[14] [HTTPS://GITHUB.COM/GIORGOSMAMALIS/THESIS](https://github.com/GIORGOSMAMALIS/THESIS)

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Ο Γεώργιος Μάμαλης γεννήθηκε στην Αθήνα το 1992. Το 2010 εγγράφηκε στο τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών & Πληροφορικής του Πανεπιστημίου Ιωαννίνων και πήρε το δίπλωμά του το 2017. Σε συνέχεια των σπουδών του, εγγράφηκε στο Ηλεκτρονικών Υπολογιστών & Πληροφορικής του Πανεπιστημίου Ιωαννίνων ως μεταπτυχιακός φοιτητής. Αφού ολοκλήρωσε τις υποχρεώσεις του ως μεταπτυχιακός φοιτητής, παρουσίασε την μεταπτυχιακή του εργασία τον Ιούλιο του 2019 με σκοπό να ολοκληρώσει τις μεταπτυχιακές του σπουδές.