

Ατομική Διπλωματική Εργασία

**ΕΦΑΡΜΟΓΗ ΤΗΣ ΕΞΟΥΣΗΣ ΔΙΑΔΙΚΑΣΙΩΝ ΓΙΑ
ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΒΙΟΛΟΓΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ**

Μάριος Κωνσταντινίδης

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2016

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Εφαρμογή της Εξόρυξης διαδικασιών για μοντελοποίηση Βιολογικών συστημάτων

Μάριος Κωνσταντινίδης

Επιβλέπουσα Καθηγήτρια

Δρ. Άννα Φιλίππου

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των
απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος
Πληροφορικής του Πανεπιστημίου Κύπρου

Μάιος 2016

Ευχαριστίες

Στα τέσσερα χρόνια παρουσίας μου στο Πανεπιστήμιο Κύπρου, θα ήθελα να ευχαριστήσω την Δρ. Άννα Φιλίππου, για την στήριξη, την καθοδήγηση και τις πολύτιμες συμβουλές που μου έδινε για τις επιλογές που είχα να κάνω. Επίσης θα ήθελα να την ευχαριστήσω για την εμπιστοσύνη που έδειξε στο πρόσωπο μου και μου έδωσε την ευκαιρία να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα. Όλους αυτούς τους μήνες ήταν πάντα στο πλευρό μου και με στήριζε τόσο πρακτικά αλλά και ψυχολογικά ώστε να ανταπεξέλθω στις δυσκολίες που είχα να ξεπεράσω.

Επιπλέον θα ήθελα να ευχαριστήσω το Δρ. Γιάννη Δημόπουλο, όπου μαζί με την κα. Άννα Φιλίππου είχαμε μια άριστη συνεργασία μεταξύ μας. Οι συμβουλές και οι ιδέες που μου έδινε όλο αυτό το καιρό ήταν αρκετά βοηθητικές για την εκπλήρωση της παρούσας μελέτης.

Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου, καθώς και τους φίλους μου εντός και εκτός Πανεπιστημίου. Πάντα ήταν το στήριγμα μου ώστε να καταφέρω να ανταπεξέλθω στα διάφορα εμπόδια που μου παρουσιάστηκαν αυτά τα τέσσερα χρόνια. Ο καθένας με βοήθησε με το δικό του τρόπο ώστε να εκπληρώσω τους στόχους μου.

Μάριος Κωνσταντινίδης

Μάιος 2016

Περίληψη

Το θέμα με το οποίο ασχολείται η παρούσα διπλωματική εργασία είναι η χρησιμοποίηση της Εξόρυξης διαδικασιών για την μοντελοποίηση Βιολογικών συστημάτων. Συγκεκριμένα χρησιμοποιήθηκε ο A-αλγόριθμος, ο οποίος παίρνει ως είσοδο ένα αρχείο καταγραφής και επιστρέφει ένα μοντέλο στη μορφή Δικτύου Πέτρι (Petri net), το οποίο περιγράφει την συμπεριφορά του αρχείου. Ο A-αλγόριθμος ανήκει στις τεχνικές Ανακάλυψης διαδικασιών που αποτελεί μια από τους τρεις βασικούς τομείς της Εξόρυξης διαδικασιών.

Ο στόχος μας ήταν να επεξεργαστούμε και να αναδιαμορφώσουμε τον A-αλγόριθμο, έτσι ώστε να εφαρμοστεί πάνω σε Βιολογικά συστήματα. Για την εύρεση των ιχνών υλοποιήθηκαν δύο αλγόριθμοι. Ο πρώτος παράγει όλες τις πιθανές ακολουθίες πυροδότησης των Βιολογικών αντιδράσεων, που ακολούθως θα αποτελέσουν το αρχείο καταγραφής για την κατασκευή του μοντέλου. Το μοντέλο που παράγεται παρουσιάζει την δομή και την αλληλεπίδραση των Βιολογικών αντιδράσεων. Ο δεύτερος αλγόριθμος παράγει ίχνη που βασίζονται στις Βιολογικές ουσίες και δείχνουν τη σειρά παραγωγής και κατανάλωσης τους. Το μοντέλο που παράγεται από αυτά τα ίχνη δίνει περισσότερες πληροφορίες σχετικά με το ποσοστό εμφάνισης των Βιολογικών ουσιών σε σχέση με ολόκληρο το Βιολογικό σύστημα.

Στην παρούσα εργασία επιχειρείται η εφαρμογή της Εξόρυξης διαδικασιών σε μονοπάτια σηματοδότησης που ξεκινούν από τη πρωτεΐνη HMGB1. Η πρωτεΐνη αυτή μετά από πειράματα φαίνεται να έχει καταλυτική σχέση με την εμφάνιση καρκίνου αφού προκαλεί ενεργοποίηση μονοπατιών που σχετίζονται με τον κυτταρικό κύκλο.

Περιεχόμενα

Κεφάλαιο 1 Εισαγωγή	1
1.1 Κίνητρο	1
1.2 Βασική ιδέα της Διπλωματικής Εργασίας	2
1.3 Χρησιμότητα της τεχνικής Εξόρυξης διαδικασιών	3
1.4 Δομή Διπλωματικής Εργασίας.....	3
 Κεφάλαιο 2 Σχετική βιβλιογραφία και υπόβαθρο	5
2.1 Βιολογικά Συστήματα.....	5
2.1.1 Εισαγωγή στη Βιολογικά συστήματα	5
2.1.2 Η προσέγγιση και οι στόχοι για τα Βιολογικά συστήματα.....	7
2.1.3 Βιολογικά δίκτυα	8
2.1.4 Ποσοτική και Ποιοτική μοντελοποίηση	10
2.1.5 Βιολογικά μοντέλα.....	11
2.1 Δίκτυα Πέτρι.....	15
2.2.1 Εισαγωγή	15
2.2.2 Ιστορία των Δικτύων Πέτρι	15
2.2.3 Βασικές έννοιες ενός Δικτύου Πέτρι.....	16
2.2.4 Επεκτάσεις των Δικτύων Πέτρι	21
2.2.5 Δομική Ανάλυση και ιδιότητες των Δικτύων Πέτρι	25
2.2.6 Πλεονεκτήματα χρήσης των Δικτύων Πέτρι	31
2.3 Εξόρυξη διαδικασιών.....	32
2.3.1 Εισαγωγή	32
2.3.2 Βασικές έννοιες της Εξόρυξης διαδικασιών	34
2.3.3 Η τεχνική της ανακάλυψης διαδικασιών	40
2.3.4 Αλγόριθμοι Ανακάλυψης διαδικασιών	41
 Κεφάλαιο 3 Εξόρυξη διαδικασιών	44
3.1 Παραγωγή ιχνών	44
3.1.1 Εισαγωγή	44
3.1.2 Δεδομένα εισόδου των αλγόριθμων	46
3.1.3 Αλγόριθμος παραγωγής ιχνών με βάση βιολογικές αντιδράσεις.....	47
3.1.4 Αλγόριθμος παραγωγής ιχνών με βάση τις βιολογικές ουσίες:.....	51

3.1.5 Παραγωγή αρχείου εισόδου για το εργαλείο Prom	55
3.2 Α-αλγόριθμος.....	56
3.2.1 Εισαγωγή	56
3.2.2 Σημασιολογία του Α-αλγόριθμου	57
3.2.3 Βήματα εκτέλεσης Α-αλγόριθμου	61
3.2.4 Παράδειγμα εκτέλεσης Α-αλγορίθμου	64
3.2.5 Πλεονεκτήματα και μειονεκτήματα του Α-αλγόριθμου	65
3.3 Εργαλείο Prom.....	66
3.3.1 Εισαγωγή	66
3.3.2 ProMPM651.....	67
3.3.3 XESame151	68
3.3.4 ProM651	70
Κεφάλαιο 4 Το Βιολογικό μοντέλο	78
4.1 Εισαγωγή	78
4.2 Μονοπάτια σηματοδότησης του μοντέλου	80
4.2.1 P53 – MDM2	82
4.2.2 RAS – ERK.....	83
4.2.3 Rb – E2F	83
4.3 Αξιολόγηση Αποτελεσμάτων	84
Κεφάλαιο 5 Συμπεράσματα και μελλοντικές εργασίες.....	124
5.1 Συμπεράσματα	124
5.2 Μελλοντικές Εργασίες.....	126
Βιβλιογραφία	128
Παράρτημα Α	A-1

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο	1
1.2 Βασική ιδέα της διπλωματικής εργασίας	2
1.3 Χρησιμότητα της Εξόρυξης διαδικασιών	3
1.4 Δομή Διπλωματικής Εργασίας	3

1.1 Κίνητρο

Τα βιολογικά συστήματα είναι ένας τομέας οποίος είχε μεγάλη ανάπτυξη τα τελευταία χρόνια, και συνεχώς βρίσκει και περισσότερους υποστηρικτές. Η γενική ιδέα αφορά την προσπάθεια ερμηνείας διάφορων βιολογικών συμβάντων χρησιμοποιώντας τη θεωρία των συστημάτων. Η έννοια του συστήματος χρησιμοποιείται σε πάρα πολλές διαφορετικές επιστήμες με ισχυρό υπόβαθρο τα μαθηματικά. Η εισαγωγή μιας τέτοιας έννοιας στη Βιολογία έδωσε νέα ώθηση στην κατανόηση διάφορων δομών και λειτουργιών. Από την μελέτη μεμονωμένων βιομορίων, η Βιολογία συστημάτων έχει μεταφέρει το ενδιαφέρον της στην ολοκληρωμένη μελέτη ομάδων βιομορίων όπου τα συστατικά τους αλληλεπιδρούν μεταξύ τους για την εκτέλεση και ολοκλήρωση βιολογικών λειτουργιών.

Η κατανόηση των αποτελεσμάτων διάφορων πειραμάτων ή μετρήσεων είναι συχνά δύσκολη διαδικασία με αβέβαια αποτελέσματα όταν δεν γνωρίζουμε πλήρως τον μηχανισμό που παράγει τα αποτελέσματα αυτά. Στην Βιολογία πολλές από τις λειτουργίες ενός οργανισμού παραμένουν άγνωστες ακόμα, παρά τα πολλά χρόνια μελέτης τους. Η προσπάθεια ανεύρεσης των λόγων εξαγωγής αυτών των συμπερασμάτων βασίζονται σε διάσπαρτα ευρήματα, και μετρήσεις συγκεκριμένων ποσοτήτων και χαρακτηριστικών του φαινότυπου οργανισμού. Αυτή η προσπάθεια εμπεριέχει την κατά καιρούς εμφάνιση διάφορων θεωριών, οι οποίες βασίζονται σε

υπάρχων στοιχεία, αλλά με περαιτέρω έρευνα αποδεικνύονται λανθασμένες ή περιορίζεται η γενικότητά τους.

Η Βιοπληροφορική (Bioinformatics) είναι επιστημονικός κλάδος ο οποίος προέκυψε τα τελευταία χρόνια από τη συνεργασία των επιστημών της Μοριακής Βιολογίας και της Πληροφορικής. Έχει ως στόχο να ενισχύσει δυναμικά την προσπάθεια κατανόησης των βιολογικών δεδομένων, εισάγοντας τα ως ψηφιακή πληροφορία. Εφαρμόζει αλγόριθμους για την επεξεργασία τους και την παραγωγή χρήσιμων συμπερασμάτων με αποδοτικό τρόπο. Συνήθως χρησιμοποιούνται μέθοδοι κλάδων από τον τομέα της Τεχνικής Νοημοσύνης, παραδείγματος χάρη η εξόρυξη δεδομένων και ο εξελικτικός υπολογισμός.

Στόχος μας ήταν να μελετήσουμε κάποια μεθοδολογία, που θα μας βοηθούσε ώστε να μοντελοποιήσουμε κάποια Βιολογικά συστήματα, με τις δυνατότητες που μας δίνει η Πληροφορική και συγκεκριμένα ο τομέας της Τεχνητής Νοημοσύνης, ώστε να αποκομίσουμε χρήσιμες πληροφορίες και συμπεράσματα από βιολογικά δεδομένα.

1.2 Βασική ιδέα της Διπλωματικής Εργασίας

Η κύρια ιδέα για την εκπόνηση αυτής της διπλωματικής εργασίας συνίσταται στην πιο πάνω επιθυμία, δηλαδή να μελετηθεί μια μεθοδολογία που μαζί με κάποια εργαλεία που την υποστηρίζουν να μοντελοποιεί τα βιολογικά συστήματα έτσι ώστε να μπορούμε να αντλήσουμε χρήσιμες πληροφορίες και συμπεράσματα για την δομή και τη λειτουργία τους.

Για το λόγο αυτό αποφασίσαμε να μελετήσουμε τεχνικές Εξόρυξης διαδικασιών (Process mining techniques). Αυτές οι τεχνικές δημιουργήθηκαν για να εξάγουν και να αναλύουν επιχειρηματικές διαδικασίες βασιζόμενες σε κάποια αρχεία καταγραφής (log events). Αυτό το φέρνουν εις πέρας με τη χρήση αλγόριθμων εξόρυξης πληροφοριών όπου εξάγουν γνώση και βγάζουν συμπεράσματα από αρχεία καταγραφής που περιγράφουν την ροή και την λειτουργία των συστημάτων τους.

Ο στόχος μας ήταν να επεξεργαστούμε και να αναδιαμορφώσουμε αυτές τις τεχνικές, έτσι ώστε να εφαρμοστούν πάνω σε βιολογικά συστήματα και να εκμεταλλευτούμε τις ιδιότητες του ώστε να αντλήσουμε όσο περισσότερη γνώση για τη δομή και τις

λειτουργίες των συστημάτων αυτών. Αυτή ήταν και η πιο μεγάλη πρόκληση, αφού αυτό δεν έχει ξαναγίνει πρακτικά αν και υπάρχουν κάποια άρθρα που το αναφέρουν στο θεωρητικό κομμάτι [16].

Μια από τις τεχνικές Εξόρυξης Διαδικασιών είναι και η ανακάλυψη διαδικασιών. Στη τεχνική αυτή εισάγοντας ένα αρχείο καταγραφής το οποίο περιέχει ίχνη που περιγράφουν την ροή μέσα στα συστήματα, δημιουργείται ένα μοντέλο (στην παρούσα διπλωματική το μοντέλο θα αναπαριστάται με ένα Δίκτυο Πέτρι) που αντιπροσωπεύει όσο το δυνατόν καλύτερα την συμπεριφορά του συστήματος που παρατηρεί από το αρχείο καταγραφής.

1.3 Χρησιμότητα της Τεχνικής Εξόρυξης Διαδικασιών

Η Εξόρυξη διαδικασιών είναι αρκετά χρήσιμη για την ανάλυση διαδικασιών διάφορων συστημάτων βασισμένη σε αρχεία καταγραφής. Όπως έχουμε αναφέρει ήδη η βασική ιδέα είναι η εξόρυξη γνώσης από αρχεία καταγραφής που αντλήθηκαν από τη ροή των συστημάτων. Οι τεχνικές της Εξόρυξης διαδικασιών συνήθως χρησιμοποιούνται στις περιπτώσεις που δεν υπάρχουν άλλοι τρόποι απόκτησης της πληροφορίας ή η υπαρκτή πληροφορία είναι αμφισβητήσιμη. Επιπλέον, αυτή η καταγραφή γεγονότων μπορεί να χρησιμοποιηθεί για να συγκρίνουμε αρχεία καταγραφής με κάποιο υπάρχον μοντέλο, έτσι ώστε να παρατηρήσουμε εάν η συμπεριφορά των γεγονότων συμμορφώνεται με το περιγραφικό ή κατευθυντήριο μοντέλο. Αυτή η δυνατότητα επιτρέπει στην περίπτωση που δεν συμμορφώνεται, να πραγματοποιηθεί μια επισκευή στο υπάρχον μοντέλο ώστε να έρθει πιο κοντά στα αρχεία καταγραφής.

1.4 Δομή Διπλωματικής Εργασίας

Η δομή της διπλωματικής εργασίας έχει ως εξής: Στο Κεφάλαιο 2 ακολουθεί μια περιγραφή του Βιολογικού κόσμου, των δικτύων Πέτρι που χρησιμοποιήθηκαν για την μοντελοποίηση των Βιολογικών συστημάτων μέσω της τεχνικής Εξόρυξης διαδικασιών. Ακολούθως στο Κεφάλαιο 3 γίνεται μια περιγραφή των δύο αλγόριθμων που χρησιμοποιήθηκαν για την παραγωγή των ιχνών. Επίσης γίνεται και περιγραφή του A-αλγόριθμου και των βημάτων εκτέλεσής του που κατασκευάζουν τα μοντέλα

υλοποιώντας της τεχνική Ανακάλυψης διαδικασιών. Στη συνέχεια του κεφαλαίου παρουσιάζονται και κάποια παραδείγματα εκτέλεσης. Τελειώνοντας το κεφάλαιο γίνεται μια γρήγορη παρουσίαση στο περιβάλλον του εργαλείου Prom, το οποίο χρησιμοποιήθηκε για την παραγωγή των μοντέλων. Στο Κεφάλαιο 4 γίνεται αναφορά στο Βιολογικό μοντέλο που μελετήθηκε και εφαρμόστηκε επάνω στα μονοπάτια του η τεχνική Εξόρυξης διαδικασιών. Στη συνέχεια παρουσιάζονται τα αποτελέσματα εφαρμόζοντας και μια ανάλυση επάνω τους. Ολοκληρώνοντας την παρούσα διπλωματική εργασία, στο Κεφάλαιο 5 δίνεται η παρουσίαση των συμπερασμάτων και ενδεχόμενων μελλοντικών προεκτάσεων της διπλωματικής εργασίας.

Κεφάλαιο 2

Σχετική Βιβλιογραφία και υπόβαθρο

2.1	Βιολογικά συστήματα	5
2.1.1	Εισαγωγή στα Βιολογικά Συστήματα	5
2.1.2	Η προσέγγιση και οι στόχοι για τα Βιολογικά Συστήματα	7
2.1.3	Βιολογικά δίκτυα	8
2.1.4	Ποσοτική και Ποιοτική μοντελοποίηση	10
2.1.5	Βιολογικά μοντέλα	11
2.2	Δίκτυα Πέτρι	15
2.2.1	Εισαγωγή στα δίκτυα Πέτρι	15
2.2.2	Ιστορία των Δικτύων Πέτρι	15
2.2.3	Βασικές έννοιες των Δικτύων Πέτρι	16
2.2.4	Επεκτάσεις των Δικτύων Πέτρι	21
2.2.5	Δομική ανάλυση και ιδιότητες των Δικτύων Πέτρι	25
2.2.6	Πλεονεκτήματα χρήσης των Δικτύων Πέτρι	31
2.3	Εξόρυξη Διαδικασίας	32
2.3.1	Εισαγωγή	32
2.3.2	Βασικές έννοιες της Εξόρυξης διαδικασιών	34
2.3.3	Η τεχνική της ανακάλυψης διαδικασιών	40
2.3.4	Αλγόριθμοι Ανακάλυψης διαδικασιών	41

2.1 Βιολογικά Συστήματα

2.1.1 Εισαγωγή στη Βιολογικά συστήματα

Οι ζωντανοί οργανισμοί είναι πολύπλοκα συστήματα που προκύπτουν από βασικά δομικά στοιχεία της ζωής. Η βιολογία των συστημάτων είναι ένα πεδίο της επιστήμης που μελετά πολύπλοκα συντρέχοντα φαινόμενα, κυρίως στο επίπεδο των κυττάρων. Η

κατανόηση του μηχανισμού των κυττάρων είναι απαραίτητη για έρευνα σε διάφορους τομείς όπως η βιοτεχνολογία. Με την ανάπτυξη αυτής της τεχνολογίας έγιναν σημαντικά βήματα στα βιοκαύσιμα και στα φαρμακευτικά προϊόντα.

Χρησιμοποιώντας μαθηματικά μοντέλα του κυτταρικού μεταβολισμού είναι δυνατόν να δοκιμαστούν διάφορα πειράματα, καθώς και προβλέψεις πάνω σε γονίδια συγκεκριμένων εφαρμογών. Αυτά τα μοντέλα χτίζονται πάνω σε επαναληπτικούς κύκλους πειραμάτων και αναφορών από ομάδες που περιλαμβάνουν Βιολόγους, Μηχανικούς και επιστήμονες της Πληροφορικής.

Η διασύνδεση μεταξύ διαφορετικών κυτταρικών διαδικασιών όπως ο μεταβολισμός και η γενετική ρύθμιση, αντικατοπτρίζουν τη σημασία της αντικατάστασης των παλιών παραδοσιακών μοντελοποιήσεων με καινούριες, που θα δώσουν όσο το δυνατόν καλύτερη ανάλυση ενός Βιολογικού συστήματος. Αφού, αν και τα περισσότερα συστατικά ενός κυττάρου μελετιούνται μεμονωμένα, η συμπεριφορά του κυττάρου ως σύνολο προκύπτει σε επίπεδο Δικτύου και απαιτεί μια ολοκληρωμένη ανάλυση.

Πρόσφατες πειραματικές μέθοδοι υψηλής απόδοσης παράγουν τα λεγόμενα omics, δηλαδή τις πληροφορίες που επιτρέπουν την ανοικοδόμηση πολλών βιολογικών δικτύων. Ωστόσο παρά τις μεγάλες προόδους στην περιοχή, ακόμη η επιστήμη απέχει πολύ από την ολοκληρωμένη υπολογιστική μοντελοποίηση ενός κυττάρου, που θα προσομοιώνει όλα τα συστατικά του. Λόγω του τεράστιου μεγέθους και της πολυπλοκότητας των ενδοκυτταρικών βιολογικών δικτύων, τα υπολογιστικά μοντέλα των κυττάρων τείνουν να γίνονται συνεχώς και πιο ενδιαφέροντα στον επιστημονικό τομέα. Επιπρόσθετα, λόγω του πολυεπιστημονικού χαρακτήρα αυτού του πεδίου, τα μοντέλα βασίζονται σε αρκετούς διαφορετικούς φορμαλισμούς. Αυτή η ποικιλομορφία μπορεί να οδηγήσει στον κατακερματισμό της μοντελοποίησης που ίσως εμποδίσει την ολοκλήρωση των μοντέλων από διαφορετικές πηγές. Παρόλα αυτά, η προσομοίωση του κυττάρου ως σύνολο, θα ωφελήσει την μοντελοποίηση και την ανάλυση με ένα ενιαίο φορμαλισμό, που θα είναι σε θέση να ενσωματώσει τις οντότητες και τις σχέσεις τους που εκτείνονται σε διάφορα βιολογικά δίκτυα.

2.1.2 Η προσέγγιση και οι στόχοι για τα Βιολογικά συστήματα

Η προσέγγιση συστημάτων είναι ένας καινούριος τομέας στο κόσμο της έρευνας, όμως τα θεμέλια για την ανάλυση βιολογικών συστημάτων έχουν επιτευχθεί στις αρχές του 21^{ου} αιώνα, με την υποστήριξη δύο βασικών διεθνών επιστημονικών πεδίων, την ‘Επιστήμη’ και τη ‘Φύση’. Το ενδιαφέρον για την προσέγγιση των βιολογικών συστημάτων συνδέεται με την πρόοδο πειραμάτων στην μοριακή βιολογία (π.χ γονιδιακές ακολουθίες), τη λήψη δεδομένων υψηλής απόδοσης, καθώς επίσης και την ανάπτυξη της τεχνολογίας με τη δημιουργία βιολογικών αισθητήρων και τη γέννηση της νάνο-βιοτεχνολογίας [1]. Αυτή η πρόοδος επιτρέπει στους επιστήμονες να συλλέγουν ολοκληρωμένα σύνολα δεδομένων για την απόδοση του συστήματος και την εξόρυξη πληροφοριών πάνω στις ιδιότητες, τη δομή και τις λειτουργίες των βιολογικών μορίων [1,2].

Η σημερινή προσέγγιση των βιολογικών συστημάτων στοχεύει στην κατανόηση των κυττάρων και των ιστών τους, καθώς ακόμη στις λειτουργίες που προκύπτουν με την αλληλεπίδραση των συστατικών τους. Περιλαμβάνεται η εξέταση της δομής και της δυναμικής των μοριακών αλληλεπιδράσεων ως προς το σύνολο και όχι ως μεμονωμένα μέρη του κυττάρου. Σύμφωνα με τους Westerhoff και Palsson, δύο γραμμές οδήγησαν στην σύγχρονη βιολογία συστημάτων. Η πρώτη ξεκινά από την αναγνώριση των ατομικών μορίων και στοχεύει στην κλιμάκωση σε μια ταυτόχρονη προβολή όλων των μορίων και των αλληλεπιδράσεων τους. Η δεύτερη γραμμή προέρχεται από την μη-ισορρόπηση της θερμοδυναμικής και επικεντρώνεται στη τυπική ανάλυση των νέων λειτουργικών μελών που απορρέουν από μοριακές αλληλεπιδράσεις με σκοπό την ανακάλυψη γενικών αρχών γύρω από αυτόν το τομέα.

Βασισμένη σε αυτές τις δύο πτυχές των βιολογικών συστημάτων, υπάρχουν δύο διαφορετικές προσεγγίσεις. Το κάτω-πάνω (‘bottom-up’) δηλαδή από τα μέρη στο σύνολο και το πάνω-κάτω (‘top-down’) δηλαδή από το σύνολο στα μέρη. Η πρώτη προσέγγιση ξεκινά από τα μέρη Βιολογικού συστήματος, διαμορφώνοντας την συμπεριφορά κάθε συστατικού και ακολούθως ενσωματώνει αυτές τις πληροφορίες που παράχθηκαν ως μοντέλα μηχανισμού για πρόβλεψη της συμπεριφοράς ολόκληρου του συστήματος. Σε αντίθεση η δεύτερη προσέγγιση χαρακτηρίζεται από ένα μεγάλο σύνολο γονιδιακών πληροφοριών που καθορίζουν τις δομικές αλληλεπιδράσεις των συστημάτων που απασχολούν φαινομενολογικά μοντέλα.

Ένας άλλος σημαντικός παράγοντας για την επαναφορά του ενδιαφέροντος για την κατανόηση των συστημάτων είναι η αποτυχία της κλασσικής φιλοσοφίας στην επεξήγηση τους. Πλέον είναι σαφές στην ψυχολογία και στην παθολογία ότι δεν μπορεί να περιγραφεί ένα κύτταρο ή ένας οργανισμός με μια μικρή ομάδα γονιδίων ή πρωτεϊνών. Αφετέρου πολλές κοινές ασθένειες είναι πολυγωνιδιακές και χαρακτηρίζονται ως πολύπλοκες σε κλινικά, κυτταρικά και μοριακά επίπεδα.

Τα τελευταία χρόνια καινοτόμες τεχνολογίες όπως η γονιδιωματική (genomics), μεταγραφωματική (transcriptomics), πρωτεϊνωματική (proteomics) και η μεταβολισμική (metabolomics) διευκολύνουν μια στρατηγική προς την ταυτόχρονη ανάλυση και την μοντελοποίηση μεγάλου μεγέθους βιολογικών γονιδίων, μεταγραφών, πρωτεϊνών και μεταβολισμών σε πολλά εργαστήρια. Ως εκ τούτου, ένας τεράστιος όγκος δεδομένων έχει δημιουργηθεί αναφορικά με τα κύτταρα και την συμπεριφορά τους στα κυτταρικά επίπεδα και στις διαφορετικές περιβαλλοντικές συνθήκες, οι οποίες ενεργοποιούν την ανακατασκευή του γονιδίου ενός μορίου για να δώσουν βαθύτερη ανάλυση.

2.1.3 Βιολογικά δίκτυα

Τα κύτταρα αποτελούνται από χιλιάδες συστατικά που αλληλεπιδρούν με εκατομμύρια διαφορετικούς τρόπους. Παρά την περίπλοκη διασύνδεση τους, συνηθίζεται να διαιρούνται και να κατηγοριοποιούνται σε δίκτυα ανάλογα με τις βιολογικές τους λειτουργίες. Οι βασικοί τύποι δικτύων είναι τα σηματοδοτηκά δίκτυα (signal transduction), τα ρυθμιστικά δίκτυα γονιδίων (Gene regulatory networks) και τα μεταβολικά δίκτυα (metabolic networks).

Σηματοδοτικά δίκτυα

Η σηματοδοτική μεταγωγή (signal transduction) είναι μια διαδικασία επικοινωνίας στα κύτταρα, όπου το κύτταρο λαμβάνει εξωτερικά ερεθίσματα από άλλα κύτταρα και από το περιβάλλον. Αυτό επηρεάζει το βασικό μηχανισμό ελέγχου του κυττάρου, για

παράδειγμα την απόπτωση¹. Η διαδικασία της μεταγωγής ξεκινά με την σύνδεση ενός εξωτερικού σηματοδοτικού μορίου στην επιφάνεια του κυττάρου, ως υποδοχέας. Το σήμα διαδίδεται και ενισχύεται εσωτερικά στο κύτταρο μέσω απότομων πτώσεων που περιλαμβάνουν μια σειρά από αντιδράσεις όπως η φωσφορυλίωση των πρωτεϊνών. Το αποτέλεσμα αυτών των απότομων πτώσεων συνδέεται άμεσα με την γονιδιακή ρύθμιση των λειτουργιών του κυττάρου. Τα σηματοδοτικά μονοπάτια μεταγωγής είναι σε θέση να περιγράψουν τα φαινόμενα που μπορούν να ενεργοποιηθούν σε κάποιο συγκρότημα των δικτύων σηματοδότησης.

Ρυθμιστικά δίκτυα γονιδίων

Η ρύθμιση γονιδίων (gene regulatory) ελέγχει την έκφραση των γονιδίων, και κατά συνέπεια όλες τις κυψελοειδείς λειτουργίες. Παρά το ότι όλες οι λειτουργίες του κυττάρου κωδικοποιούνται στο γονιδίωμα² μέσω χιλιάδων γονιδίων, είναι αναγκαίο για την επιβίωση του κυττάρου μόνο κάποιες επιλεγμένες λειτουργίες του να είναι ενεργές σε συγκεκριμένες χρονικές στιγμές. Η έκφραση του γονιδίου είναι μια διαδικασία που περιλαμβάνει μεταγραφή του γονιδίου σε mRNA, που ακολουθείται από μετάφραση μιας πρωτεΐνης, η οποία είναι δυνατόν να υπόκειται σε μεταγραφική τροποποίηση. Η διαδικασία της μεταγραφής ελέγχεται από μεταγραφικούς παράγοντες (Transcription Factors) που μπορούν να δουλέψουν ως ενεργοποιητές ή ανασταλτικοί.

¹ Απόπτωση κυττάρου = Προγραμματισμένη διαδικασία θανάτου του κυττάρου που συμβαίνει στους πολυκυτταρικούς οργανισμούς.

² Γονιδίωμα = Το γενετικό υλικό ενός οργανισμού που περιέχει το DNA ή το RNA.

Μεταγραφικοί παράγοντες είναι οι πρωτεΐνες που δεσμεύουν συγκεκριμένες ακολουθίες DNA, έτσι ώστε να ελέγχουν το ποσοστό μεταγραφής των γενετικών πληροφοριών από DNA σε RNA. Αυτή τη λειτουργία μπορούν να την εκτελούν μόνοι τους ή σε συνεργασία με άλλες πρωτεΐνες, με την προώθηση ή την απενεργοποίηση της πρόσληψης της RNA πολυμεράσης (ένζυμο που εκτελεί την μεταγραφή πληροφοριών από DNA σε RNA) σε συγκεκριμένα γονίδια [3]. Το καθοριστικό στοιχείο των μεταγραφικών παραγόντων είναι ότι περιέχουν ένα ή περισσότερους DNA δεσμευτικούς τομείς που αποδίδουν ειδικές ακολουθίες του DNA που γειτνιάζουν με τα γονίδια που ρυθμίζουν.

Μεταβολικά δίκτυα

Ο μεταβολισμός (metabolic) είναι ένας μηχανισμός που αποτελείται από ένα σύνολο των βιοχημικών αντιδράσεων, με το οποίο το κύτταρο στηρίζει τις ενεργειακές του απαιτήσεις. Περιλαμβάνει αρκετά καταβολικά³ και αναβολικά⁴ μονοπάτια των αντιδράσεων που καταλύονται από τα ένζυμα, εισάγοντας υποστρώματα από το περιβάλλον και μετασχηματίζοντάς τα σε ενέργεια και δομικά στοιχεία που απαιτούνται για να χτιστούν τα κυτταρικά τους συστατικά. Τα μεταβολικά μονοπάτια είναι διασυνδεδεμένα μέσω ενδιάμεσων μεταβολικών, σχηματίζοντας σύνθετα δίκτυα. Η ρύθμιση του γονιδίου ελέγχεται από την παραγωγή ενζύμων και κατά συνέπεια, απευθείας από τη μεταβολική ροή μέσω των κατάλληλων μονοπατιών των λειτουργιών των διαθέσιμων υποστρωμάτων και διατροφικών απαιτήσεων.

2.1.4 Ποσοτική και Ποιοτική μοντελοποίηση

Ένα Βιολογικό μοντέλο είναι μια αφαιρετική απεικόνιση της δομής και των λειτουργιών ενός Βιολογικού συστήματος. Η αφαιρετικότητα καθορίζει και το είδος της ανάλυσης που θα ακολουθήσει η οποία μπορεί να είναι είτε ποιοτική είτε ποσοτική είτε συνδυασμός και των δύο.

³ Καταβολικά = χημικές αντιδράσεις που έχουν ως αποτέλεσμα την κατανομή των πιο πολύπλοκων οργανικών μορίων σε απλούστερες ουσίες

⁴ Αναβολικά = χημικές αντιδράσεις στις οποίες απλούστερες δομημένες ουσίες συνδυάζονται για να σχηματίσουν πιο σύνθετα μόρια, τα οποία απαιτούν ενέργεια

Στα ποιοτικά μοντέλα σχεδιάζεται μια αφαιρετική δομή η οποία περιέχει τις οντότητες και τις σχέσεις μεταξύ τους χωρίς να αναφέρονται οποιεσδήποτε ποσοτικές πληροφορίες. Επιτρέπει τη μοντελοποίηση συστημάτων από τα οποία απουσιάζουν ακριβείς πληροφορίες σχετικά με τη λειτουργία τους, σε πολλά αφαιρετικά επίπεδα. Επομένως, τα αποτελέσματα ανάλυσης ενδέχεται να μην είναι ακριβή αν και η δυνατότητα μη ντετερμινιστικής ανάλυσης σε ποιοτικά βιολογικά έχει ενθαρρυντικά αποτελέσματα καθώς τα βιολογικά συστήματα λειτουργούν με μη ντετερμινιστικό τρόπο. Γενικά η ποιοτική ανάλυση επιχειρεί μια περισσότερο αφαιρετική προσέγγιση χωρίς να επιβαρύνει την ανάλυση του μοντέλου.

Από την άλλη στην ποσοτική ανάλυση μας ενδιαφέρει η λεπτομερής λειτουργία του οργανισμού. Στο μοντέλο εισάγονται ποσοτικές πληροφορίες που αφορούν τα συστατικά όπως είναι για παράδειγμα τα επίπεδα συγκέντρωσης των οντοτήτων και το πώς αλληλεπιδρούν μεταξύ τους. Το πλεονέκτημα αυτής της προσέγγισης είναι η εξαγωγή ακριβέστερων αποτελεσμάτων εφόσον η εισαγωγή των ποσοτικών πληροφοριών στο μοντέλο το καθιστούν όσο το δυνατόν πιο ρεαλιστικό σε σχέση με το πραγματικό σύστημα. Όμως πολλές φορές είναι δύσκολο να εξαχθούν οι ποσοτικές λεπτομέρειες οι οποίες καθιστούν το μοντέλο δύσκολο στην υλοποίηση. Επιπλέον, τα ποσοτικά μοντέλα έχουν αυξημένη πολυπλοκότητα αφού περιέχουν αρκετές ακριβείς πληροφορίες και δυσκολεύουν την ανάλυση του συστήματος ειδικά όταν οι εξαρτήσεις μεταξύ των συστατικών του είναι μεγάλου μεγέθους.

2.1.5 Βιολογικά μοντέλα

Κατά καιρό χρησιμοποιήθηκαν διάφοροι μέθοδοι για τη μοντελοποίηση βιολογικών συστημάτων, εν μέρει λόγω της ανομοιογένειας των φαινομένων που υπάρχουν στα βιολογικά συστήματα, καθώς επίσης και των διαφορετικών επιστημονικών ομάδων που ασχολήθηκαν με αυτά. Οι Βιολόγοι μπορεί να είναι πιο εξοικειωμένοι με την μαθηματική μοντελοποίηση, ενώ οι επιστήμονες της Πληροφορικής επιλέγουν υπολογιστικές μεθόδους. Παρά το γεγονός πως μπορεί να ακολουθούν διαφορετικές προσεγγίσεις, τίθεται το ερώτημα αν υπάρχει σαφής διαχωρισμός μεταξύ μαθηματικών και υπολογιστικών μοντέλων. Στη συνέχεια του υποκεφαλαίου θα γίνει εκτενής περιγραφή σε κάποιες από τις υπολογιστικές φόρμουλες μοντελοποίησης.

Δυαδικά δίκτυα (Binary Networks)

Τα δυαδικά δίκτυα εισάχθηκαν από τον Kauffman το 1969 [14 σελ.5] ως μοντέλα ρυθμιστικών δικτύων γονιδίων. Αποτελούνται από δίκτυα γονιδίων, διαμορφωμένα από δυαδικές μεταβλητές που αντιπροσωπεύουν ενεργές και ανενεργές καταστάσεις. Σε κάθε χρονικό βήμα, προσδιορίζεται η κατάσταση του κάθε γονιδίου από ένα κανόνα λογικής που λέει πως η κατάσταση είναι ενεργή εάν το σύνολο των παραγόντων ενεργοποίησης του γονιδίου είναι μεγαλύτερο από το σύνολο των παραγόντων καταστολής, ενώ είναι ανενεργή σε αντίθετη περίπτωση. Εκτενής περιγραφή της λειτουργίας τους αλλά και πως συνδυάζονται με τα δίκτυα Πέτρι υπάρχουν στο [12]. Η κατάσταση όλων των γονιδίων σχηματίζει μια καθολική κατάσταση που αλλάζει συγχρονισμένα. Για πολύ μεγάλα μεγέθη μοντέλων αποτελεί απίθανο να εξερευνηθούν όλες οι πιθανές καταστάσεις (2^n όπου $n \in \mathbb{N}$). Αυτός ο τύπος μοντέλου μπορεί να εφαρμοστεί για να βρεθούν περιπτώσεις που το Δίκτυο βρίσκεται σε σταθερή κατάσταση (ελκυστής), αλλά και στην ανάλυση του βαθμού ευρωστίας του Δικτύου. Τα δυαδικά δίκτυα μπορούν να παραχθούν από πειράματα πάνω στην γονιδιακή έκφραση στοιχείων που περιέχουν την έννοια της χρονολογικής σειράς. Επίσης εφαρμόστηκαν σε μερικές μελέτες για μοντέλα σηματοδοτικών μονοπατιών. Λόγω του ότι η κατάσταση του συστήματος καθορίζεται από το σύνολο των καταστάσεων των συστατικών στοιχείων και εφόσον η λειτουργικότητα του αφορά τις καταστάσεις από τις οποίες διέρχεται, τα Δυαδικά δίκτυα κατατάσσονται στα υπολογιστικά μοντέλα.

Bayesian δίκτυα (Bayesian networks)

Τα Bayesian δίκτυα εισάχθηκαν τη δεκαετία του 80 από τον Pearl [14 σελ.5] και συγκεκριμένα το 1988. Πρόκειται για μια ειδική ομάδα πιθανοτικών γράφων. Οι κόμβοι αναπαριστούν τυχαίες μεταβλητές (διακριτές ή συνεχείς) και οι ακμές αναπαριστούν εξαρτήσεις, σχηματίζοντας μια κατευθυνόμενη γραφική παράσταση. Κάθε κόμβος περιλαμβάνει μια πιθανοτική συνάρτηση που υπολογίζεται από τις τιμές εισόδου στο κόμβο. Υπάρχουν μέθοδοι μάθησης που βγάζουν συμπεράσματα από τη δομή και την πιθανότητα παραμέτρων που υποστηρίζουν ελλιπή δεδομένα. Αυτή η ευελιξία καθιστά τα Δίκτυα Bayesian εξαιρετικά ενδιαφέροντα για βιολογικές εφαρμογές. Χρησιμοποιήθηκαν για αναλύσεις και αναπαραστάσεις ρυθμιστικών γονιδίων αλλά και

για δίκτυα σηματοδότησης. Ένα βασικό τους μειονέκτημα είναι η ανικανότητα να μοντελοποιήσουν βρόγχους με ανάδραση, ένα σημαντικό χαρακτηριστικό στα βιολογικά δίκτυα. Παρόλα αυτά, αυτός ο περιορισμός μπορεί να ξεπεραστεί με την εφαρμογή δυναμικών δικτύων Bayesian. Σε αυτή τη περίπτωση, οι μεταβλητές ενημερώνονται συνεχώς σε κάθε χρονικό βήμα και η ανάδραση μοντελοποιείται με την αλλαγή στη σύνδεση των κόμβων.

Δίκτυα Πέτρι (Petri nets)

Γίνεται εκτενής περιγραφή των Δικτύων Πέτρι στη Κεφάλαιο 2.2. αφού είναι η μεθοδολογία που επιλέχθηκε για την μοντελοποίηση στην παρούσα διπλωματική εργασία.

Άλγεβρες Διαδικασιών (Process algebra)

Οι άλγεβρες διεργασιών [14 σελ.5] εστιάζονται στη σημασία των αντιδράσεων μεταξύ των βιολογικών μορίων ως την κύρια αιτία της υπάρξεως των βιολογικών διεργασιών. Η λειτουργία ενός μοντέλου γραμμένο σε άλγεβρα διεργασιών ορίζεται διαμέσου μιας αλληλουχίας γεγονότων. Κάθε μόριο σχετίζεται με μια διαδικασία και πολλά αντίγραφα της ίδιας διαδικασίας τρέχουν παράλληλα προσομοιώνοντας έτσι πολλά μόρια. Η επικοινωνία μεταξύ δύο διαδικασιών με κανάλια χρησιμοποιείται για να μοντελοποιήσει τις αντιδράσεις που προκαλούνται μεταξύ δυο μορίων συσχετίζοντας ένα γεγονός με την αλλαγή της κατάστασης εκείνου του μορίου. Πέρα από τη βασική σημασιολογία υπάρχει μια ποικιλία παραλλαγών που επιτρέπουν πιο εξειδικευμένη μοντελοποίηση όπως για παράδειγμα στοχαστική συμπεριφορά και ανάπτυξη βιολογικών μοντέλων.

Αλληλεπιδρόμενες μηχανές καταστάσεων (Interacting state machines)

Οι αλληλεπιδρόμενες μηχανές καταστάσεων [14 σελ.7] είναι διαγράμματα βασισμένα σε φορμαλισμούς που περιγράφουν τη χρονική συμπεριφορά ενός συστήματος με βάση τις αλλαγές των καταστάσεων στα μέρη του. Είναι τα κατάλληλα μοντέλα για περιγραφή συμπεριφοράς βιολογικών συστημάτων τόσο σε ποσοτική όσο και σε ποιοτική μοντελοποίηση. Διαφέρουν από άλλες προσεγγίσεις αφού καθορίζουν το σύστημα με όρους ως προς τις καταστάσεις του και όχι ως προς τα συστατικά του. Συνήθως χρησιμοποιούνται για την εκτέλεση μοντέλο ελέγχου (model checking) και μοντέλων με διαδραστικότητα. Η πρώτη τους εφαρμογή ήταν τα Statecharts, τα οποία δημιουργήθηκαν από τον David Harel κατά τη δεκαετία του 80. Τότε είχαν εφαρμοστεί στη βιολογία για μοντελοποίηση της ενεργοποίησης των διαδικασιών του κυττάρου T και πιο πρόσφατα στην οργανογένεση του παγκρέατος. Σε αυτό το φορμαλισμό, οι καταστάσεις του συστήματος περιλαμβάνουν υποκαταστάσεις σε πολλαπλά επίπεδα, επιτρέποντας μιας ιεραρχική προβολή του συστήματος και της σχέσης μεταξύ των ενεργειών που γίνονται στα συστατικά του.

Κυτταρικά αυτόματα (Cell Automata)

Τα κυτταρικά αυτόματα δημιουργήθηκαν από τον Von Neumann τη δεκαετία του 1940 [14 σελ.7]. Είναι διακριτά δυναμικά μοντέλα που περιέχουν ένα πλέγμα από κύτταρα με ένα πεπερασμένο αριθμό καταστάσεων. Ένα κυτταρικό αυτόματο έχει μια αρχική ρύθμιση των παραμέτρων του, που ακολούθως αλλάζουν σε κάθε χρονικό βήμα με βάση κάποιους προκαθορισμένους κανόνες που υπολογίζουν τη νέα κατάσταση του κάθε κυττάρου με βάση των λειτουργιών του γειτόνων του στο πλέγμα. Η πολυπλοκότητα τους και ο αριθμός των παραμέτρων τους καθορίζεται από το σύνολο των πειραματικών δεδομένων που απαιτούνται για την κατασκευή του Βιολογικού μοντέλου. Συνήθως χρησιμοποιούνται στις περιπτώσεις όπου οι παράμετροι αλλά και η δομή του Βιολογικού συστήματος είναι άγνωστη.

2.2 Δίκτυα Πέτρι

2.2.1 Εισαγωγή

Ανάμεσα στην ποικιλία των σύγχρονων εργαλείων που χρησιμοποιούνται για την ανάλυση συστημάτων, τα Δίκτυα Πέτρι κατέχουν πρωταρχική θέση. Ο ρόλος τους είναι θεμελιώδης όσον αφορά την ανάλυση συστημάτων τα οποία είναι συντρέχοντα, κατανεμημένα, παράλληλα, και γενικά μη ντετερμινιστικά. Είναι αναγνωρισμένο ότι τις τελευταίες δεκαετίες έχει αυξηθεί η εφαρμογή και η αποδοχή των Δικτύων Πέτρι ως βασικό μοντέλο για ασύγχρονα συστήματα συντρέχοντος υπολογισμού. Αυτά τα δίκτυα μας βοηθάνε να καταλάβουμε καλύτερα τον τρόπο λειτουργίας των συστημάτων ενώ χρησιμοποιούνται σαν ένα είδος μηχανικής γλώσσας για ανάπτυξη, προσομοίωση και εφαρμογή σε διαφορετικούς τομείς όπως οι επιχειρηματικές διαδικασίες και η βιολογία [11].

Ένα Δίκτυο Πέτρι είναι ένα περιληπτικό και τυπικό μοντέλο ροής πληροφοριών. Οι ιδιότητες, οι έννοιες και οι τεχνικές των Δικτύων Πέτρι αναπτύσσονται στην προσπάθεια εύρεσης απλών, φυσικών και ισχυρών μεθόδων περιγραφής και ανάλυσης της ροής πληροφοριών και του ελέγχου, κυρίως σε συστήματα που μπορούν να εκτελέσουν συντρέχουσες και ταυτόχρονες διεργασίες.

2.2.2 Ιστορία των Δικτύων Πέτρι

Τα Δίκτυα Πέτρι γεννήθηκαν το 1962, στη Φυσικομαθηματική Σχολή του Τεχνικού Πανεπιστημίου του Ντάρμστατ στη Γερμανία, από την διδακτορική διατριβή του Carl Adam Petri, με θέμα “Kommunikation mit Automaten” [4], σημαίνει “Communication with automata”, και είχε ως αντικείμενο την επικοινωνία μεταξύ αυτόματων μηχανών. Με τη χρήση ενός Δικτύου, περιέγραψε την τυπική σχέση μεταξύ γεγονότων σε ένα σύστημα υπολογιστών. Ήταν η πρώτη επιτυχημένη προσπάθεια κατά την οποία διατυπώθηκε μια γενική θεωρία για τα διακριτά παράλληλα συστήματα.

Η δουλειά του Carl Adam Petri τράβηξε την προσοχή διάφορων επιστημόνων από το Information System Theory Project of Applied Research, Inc. στις Ηνωμένες Πολιτείες. Εκείνοι πρότειναν ότι τα δίκτυα Πέτρι θα μπορούσαν να χρησιμοποιηθούν για τη μοντελοποίηση και την ανάλυση συστημάτων με ταυτόχρονα γεγονότα. Επίσης το Computational Structure Group του MIT πρόσεξε τη δουλειά των Δικτύων Πέτρι, ενώ στην δεκαετία του '70 έγιναν και πολλές διδακτορικές πάνω στο αντικείμενο.

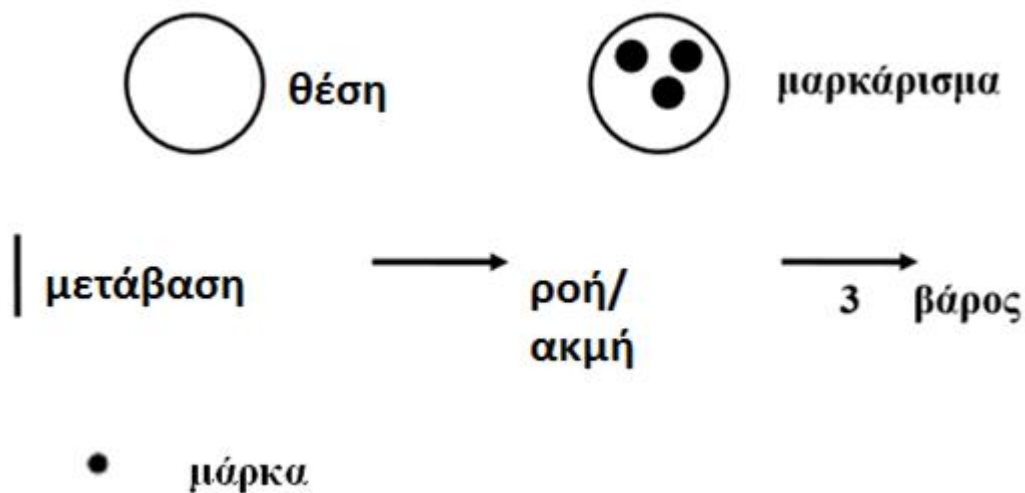
Από τα τέλη της δεκαετίας του 1970, τα δίκτυα Πέτρι αποτέλεσαν μια ιδιαίτερος ενεργή περιοχή, κυρίως στην Ευρώπη. Ετήσια συνέδρια γύρω από τις Εφαρμογές και την Θεωρία των Δικτύων Πέτρι γίνονται από το 1979, ενώ το πρώτο βιβλίο με συλλογή μελετών και κειμένων από τα δίκτυα Πέτρι γράφτηκε το 1981 από τον Peterson. Οι περισσότερες μελέτες αφορούσαν συστήματα επεξεργασίας δεδομένων που απασχολούσαν τους ανθρώπους της κοινωνίας των ηλεκτρονικών υπολογιστών.

Στις αρχές της δεκαετίας του 1980, αναλυτές άρχισαν να ασχολούνται με τη χρήση Δικτύων Πέτρι σε εφαρμογές μηχανικών, και κυρίως σε Αυτόματα Συστήματα Κατεργασιών. Στην πορεία ανακαλύφθηκε ότι τα Δίκτυα Πέτρι είναι ένα ισχυρό εργαλείο στην περιγραφή συστημάτων που εξαρτώνται από γεγονότα. Αυτά τα συστήματα μπορεί να είναι ασύγχρονα, να περιέχουν σειριακές ή παράλληλες εργασίες, να περιλαμβάνουν συγκρούσεις, αμοιβαίο αποκλεισμό και να μην είναι ντετερμινιστικά. Τέτοια συστήματα χαρακτηρίζονται ως συστήματα διακριτών γεγονότων ή Δυναμικά Συστήματα Διακριτών Γεγονότων (Discrete Event Dynamics Systems). Πολλά βιβλία κυρίως προς τα τέλη της δεκαετίας του '80, ασχολήθηκαν με εφαρμογές των Δικτύων Πέτρι σε προβλήματα μηχανικών, είτε αυτά ήταν η μοντελοποίηση και η σύνθεση μεθόδων για τον έλεγχο των συστημάτων των εργασιών μέσω διακριτών γεγονότων, είτε η γενικότερη χρήση Δικτύων Πέτρι στην ανάλυση, την αποτίμηση των επιδόσεων, και τον έλεγχο συστημάτων παραγωγής.

2.2.3 Βασικές έννοιες ενός Δικτύου Πέτρι

Τα δίκτυα Πέτρι είναι κατευθυνόμενα, διμερή γραφήματα με δύο είδη κόμβων που αναπαρίστανται διαφορετικά. Ο πρώτος τύπος κόμβων ονομάζεται θέση (place) και ο δεύτερος μετάβαση (transition). Μια θέση αναπαρίσταται με κύκλο και αναφέρεται σε

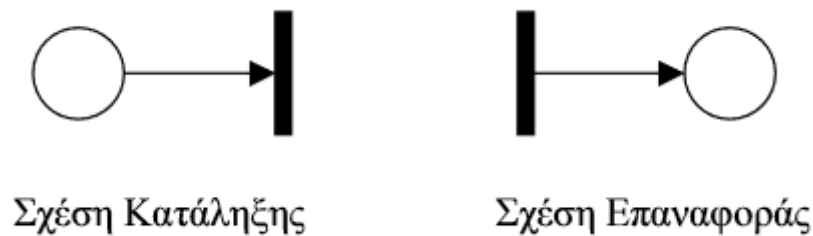
μια κατάσταση ή μια οντότητα. Η δυναμικότητα του μοντέλου στηρίζεται στις μάρκες (tokens). Η μάρκα αναπαρίσταται με μία τελεία ή αριθμό, τοποθετείται μέσα στη θέση και δηλώνει τη διακριτή ποσότητα της κατάστασης ή της οντότητας. Σε κάθε θέση απονέμουμε ένα συγκεκριμένο αριθμό μαρκών. Ο καταμερισμός των μαρκών στις θέσεις του Δικτύου ονομάζεται μαρκάρισμα και συμβολίζεται με M . Το αρχικό μαρκάρισμα συμβολίζεται με M_0 . Μια μετάβαση αναπαρίσταται με κατακόρυφο ορθογώνιο ή κατακόρυφη γραμμή και αναφέρεται στην ενεργοποίηση κάποιου γεγονότος ή διαδικασίας που λαμβάνει χώρα στο Δίκτυο. Οι θέσεις συνδέονται με τις μεταβάσεις, οι οποίες με την σειρά τους συνδέονται με άλλες θέσεις. Οι συνδέσεις γίνονται μέσω των ακμών. Οι ακμές είναι πάντα κατευθυνόμενες και ονομάζονται τόξα (arcs). Κάθε τόξο έχει το δικό του βάρος που καθορίζει τη σχέση μεταξύ μιας θέσης και μιας μετάβασης και το πώς αλλάζει το μαρκάρισμα (marking) του Δικτύου μετά την πυροδότηση της μετάβασης. Στην περίπτωση που ένα τόξο δεν έχει βάρος, τότε το βάρος του ισούται με ένα. Στο Σχήμα 2.2.1 φαίνονται τα σύμβολα.



Σχήμα 2.2.1: Βασικά συστατικά ενός Δικτύου Πέτρι

Στους διμερείς γράφους μια ακμή μπορεί να συνδέει μόνο δύο κόμβους διαφορετικού τύπου. Οπότεν σε ένα Δίκτυο Πέτρι έχουμε δύο διαφορετικές σχέσεις. Η πρώτη σχέση παριστάνεται από ένα τόξο που ξεκινά από μία θέση και καταλήγει σε μια μετάβαση

και ονομάζεται Σχέση Κατάστασης. Η δεύτερη Σχέση παριστάνεται από ένα τόξο που ξεκινά από μία μετάβαση και καταλήγει σε μια θέση και ονομάζεται σχέση Επαναφοράς. Στο Σχήμα 2.2.2 φαίνονται οι δύο σχέσεις των Δικτύων Πέτρι.

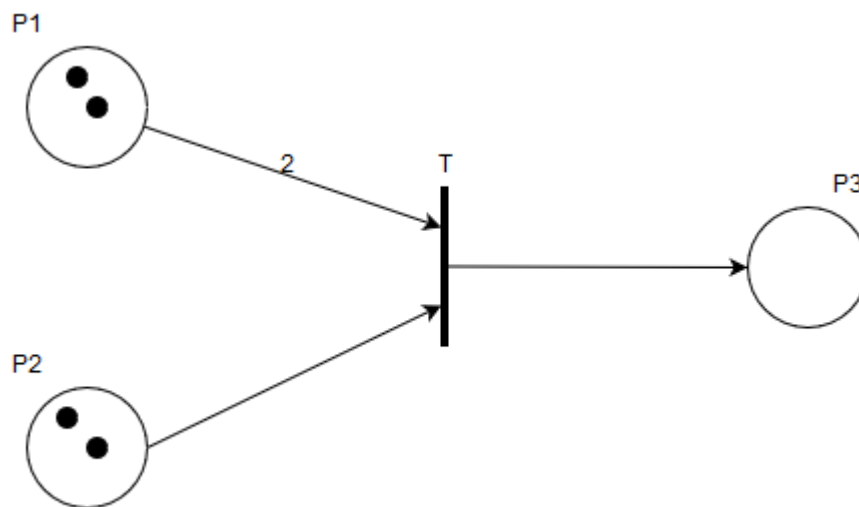


Σχήμα 2.2.2: Οι δύο σχέσεις ενός Δικτύου Πέτρι

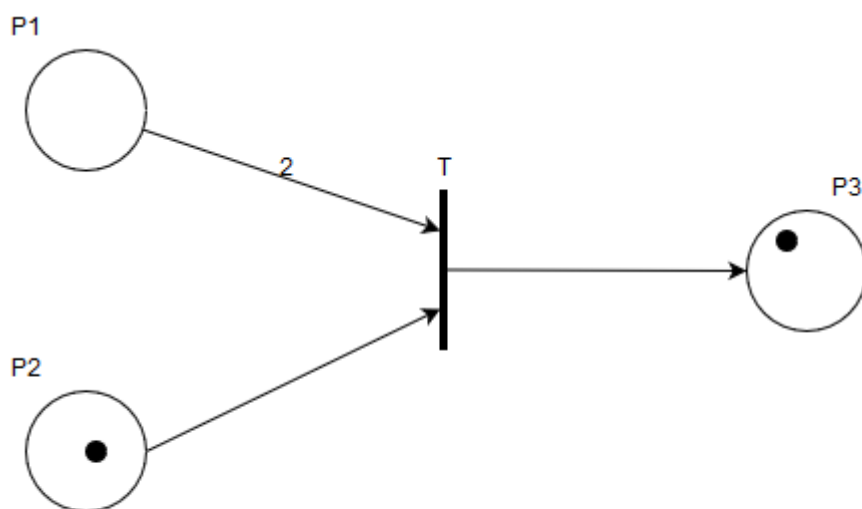
Οι θέσεις όπου μέσω των τόξων οδηγούν σε μια μετάβαση, λέγονται θέσεις εισόδου (input places) για αυτή τη μετάβαση, ενώ οι θέσεις στις οποίες καταλήγουν τα τόξα από μια μετάβαση λέγονται θέσεις εξόδου (output places) για αυτή τη μετάβαση. Οι θέσεις εισόδου μιας μετάβασης αναπαριστούν τις προϋποθέσεις για το γεγονός που αντιστοιχεί η μετάβαση ή απαραίτητα δεδομένα εισόδου. Ενώ οι θέσεις εξόδου μιας μετάβασης αναπαριστούν τα αποτελέσματα του γεγονότος που αντιστοιχεί στη μετάβαση ή στις λογικές προτάσεις που έπονται της μετάβασης. Μια μετάβαση που δεν έχει θέσεις εισόδου ονομάζεται πηγή (source transition). Μια μετάβαση που δεν έχει θέσεις εξόδου ονομάζεται καταναλωτής (sink transition). Παράλληλα οι μεταβάσεις που οδηγούνται μέσω των τόξων σε μια θέση ονομάζονται μεταβάσεις εισόδου για αυτή τη θέση (input transition), ενώ οι μεταβάσεις που καταλήγουν τα τόξα από μια θέση ονομάζονται μεταβάσεις εξόδου (output transition) για τη θέση.

Οι αλλαγές του μαρκαρίσματος πραγματοποιούνται σύμφωνα με την πυροδότηση των μεταβάσεων (fire of transition). Βασική προϋπόθεση για να πραγματοποιηθεί πυροδότηση μιας μετάβασης θα πρέπει η συγκεκριμένη μετάβαση να είναι ενεργή. Δηλαδή κάθε μια από τις θέσεις εισόδου της να περιέχει αριθμό μαρκών ίσο ή μεγαλύτερο με το βάρος της ακμής που τη συνδέει με τη μετάβαση. Μια ενεργή μετάβαση μπορεί να πυροδοτηθεί, ή και όχι.

Το αποτέλεσμα μιας πυροδότησης μιας μετάβασης T , έχει ως αποτέλεσμα την μετακίνηση από κάθε θέση εισόδου P_i , ίσο αριθμό μαρκών με αυτόν του βάρους κάθε ακμής από την P_i στην T . Στην συνέχεια εισάγει σε κάθε θέση εξόδου P_j , ίσο αριθμό μαρκών με αυτόν του βάρους κάθε ακμής από την T στην P_j . Ένα παράδειγμα πυροδότησης δίνεται στο Σχήμα 2.2.3. ($P3 = P1 \text{ AND } P2$)



ΠΡΙΝ ΤΗΝ ΠΥΡΟΔΟΤΗΣΗ



ΜΕΤΑ ΤΗΝ ΠΥΡΟΔΟΤΗΣΗ

Σχήμα 2.2.3: Παράδειγμα πυροδότησης της μετάβασης T

Ο τυπικός ορισμός των Δικτύων Πέτρι είναι :

Ένα Δίκτυο Πέτρι είναι μια 5-άδα $N = (P, T, F, W, m_0)$ όπου P και T είναι πεπερασμένα σύνολα και

- i. $P = \{ p_1, p_2, \dots, p_n \}, n > 0$, είναι το σύνολο όλων των θέσεων του Δικτύου Πέτρι.
- ii. $T = \{ t_1, t_2, \dots, t_m \}, m > 0$, είναι το σύνολο όλων των μεταβάσεων του Δικτύου Πέτρι, όπου $P \cup T \neq \emptyset$ και $P \cap T = \emptyset$.
- iii. Η σχέση $F \subseteq (P \times T) \cup (T \times P)$ είναι η σχέση ροής του N , συμπεριλαμβανομένων των ακμών, δηλαδή (p, t) ή (t, p) όπου p είναι μια θέση και t μια μετάβαση.
- iv. Μια απεικόνιση $W : F \rightarrow \mathbb{N}$ η οποία αντιστοιχεί κάθε ακμή (χ, ψ) σε έναν ακέραιο αριθμό, και ορίζει το βάρος της ακμής, $W(\chi, \psi)$.
- v. Ένα αρχικό μαρκάρισμα $m_0 : P \rightarrow \mathbb{N}$, το οποίο αντιστοιχεί κάθε θέση p με ένα αρχικό ακέραιο αριθμό μαρκών, $m_0(p)$.

2.2.4 Επεκτάσεις των Δικτύων Πέτρι

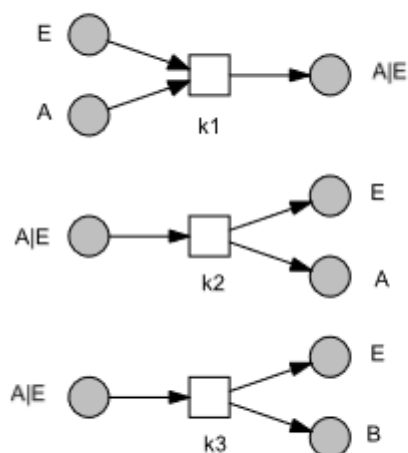
Τα δίκτυα Πέτρι σχετίζονται με ένα μεγάλο αριθμό επεκτάσεων του απλού μοντέλου που βελτιώνουν τις δυνατότητες αναπαράστασης του. Οι επεκτάσεις επιτρέπουν μεταξύ άλλων την πληρέστερη ποσοτική ανάλυση συστημάτων, την αναπαράσταση συνεχών ή υβριδικών συστημάτων, την αναπαράσταση μη βεβαίας γνώσης για την κατάσταση ενός συστήματος, καθώς και πιο απλή και οργανωμένη σχεδίαση των μοντέλων. Πιο κάτω παρουσιάζονται οι βασικές επεκτάσεις των Δικτύων Πέτρι που τους δίνουν ισχυρά πλεονεκτήματα μοντελοποίησης βιολογικών συστημάτων έναντι άλλων μεθοδολογιών.

Λογικοί και Μακρό κόμβοι

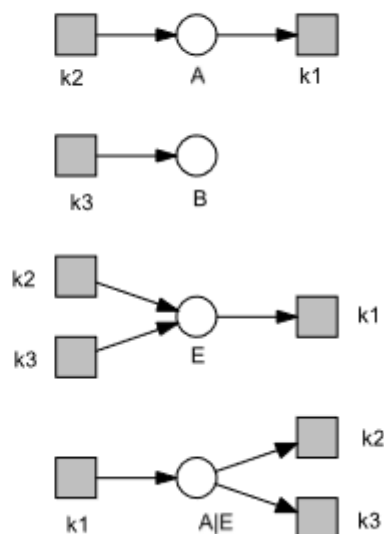
Για λόγους αναπαράστασης μεγάλων Δικτύων Πέτρι δημιουργήθηκαν δύο επιπλέον κόμβοι. Ονομάζονται λογικοί κόμβοι (logical nodes) και Μακρό κόμβοι (macro nodes). Τα δύο είδη κόμβων βοηθούν στον χειρισμό και στην τακτοποίηση Δικτύων Πέτρι με μεγάλο μέγεθος. Οπότεν για αποτελεσματική ανάλυση και μοντελοποίηση ενός Βιολογικού συστήματος, το οποίο αποτελείται από πολλαπλά επίπεδα και περιγράφεται μέσω χιλιάδων βιολογικών αντιδράσεων και ουσιών, χρειάζεται μια επέκταση του απλού Δικτύου Πέτρι ώστε να γίνεται καλύτερη διαχείριση και οργάνωση του συστήματος και των συστατικών του.

Οι λογικές θέσεις ή μεταβάσεις (logical places or transitions) [5 σελ.8] αναπαριστώνται με τον ίδιο τρόπο αναπαράστασης των απλών αντίστοιχων συστατικών τους που παρουσιάζεται στο Υποκεφάλαιο 2.2.3, με τη μόνη διαφορά ότι εσωτερικά χρωματίζονται με γκρίζο χρώμα. Αυτά τα είδη κόμβων χρησιμοποιούνται στις περιπτώσεις όπου μια θέση ή μια μετάβαση επαναλαμβάνεται συνεχώς στο Δίκτυο. Η συγκεκριμένη επέκταση είναι πολύ ωφέλιμη για μοντελοποίηση βιολογικών συστημάτων, αφού μπορεί να αντιμετωπίσει την συνεχή επανάληψη των ίδιων οντοτήτων σε διαφορετικές καταστάσεις. Στο Σχήμα 2.2.4 παρουσιάζεται ένα παράδειγμα λογικών θέσεων και ένα παράδειγμα λογικών μεταβάσεων.

A - Logical Places

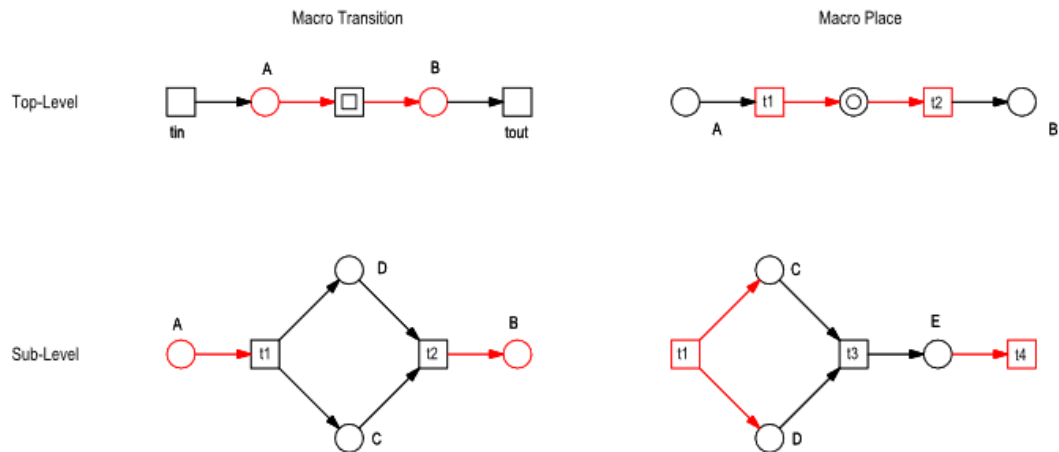


B - Logical Transitions



Σχήμα 2.2.4: Αναπαράσταση Δικτύων Πέτρι με λογικές θέσεις και μεταβάσεις

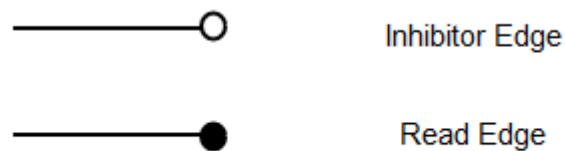
Οι μακρό κόμβοι [5 σελ.9] χωρίζονται σε μακρό-θέσεις και μακρό-μεταβάσεις. Οι μακρό-θέσεις αναπαριστώνται με δύο κύκλους: Ο ένας κύκλος εσωτερικά μέσα στον άλλο. Οι μακρό μεταβάσεις αναπαριστώνται με δύο τετράγωνα: το ένα τετράγωνο εσωτερικά μέσα στο άλλο. Κάθε μακρό κόμβος εγκαθιδρύει ένα νέο επίπεδο στο Δίκτυο, που εμβολιάζεται στο αμέσως προηγούμενο επίπεδο του Δικτύου. Επίσης οι συγκεκριμένοι κόμβοι δίνουν τη δυνατότητα να επανακαθοριστεί η δομή του Δικτύου σε ένα καινούριο επίπεδο. Επιπλέον οι μακρό κόμβοι βοηθούν ώστε να δομηθεί το Δίκτυο με την σύνδεση διαφορετικών ανεξάρτητων υποδικτύων ή με μια ιεραρχική δομή. Στο κόσμο της Βιολογίας οι μακρό-κόμβοι μπορούν να αντικαταστήσουν μια θέση ή μια μετάβαση με ένα υποδίκτυο ώστε να παρέχεται πιο λεπτομερής περιγραφή κάποιων συστατικών μερών. Στο Σχήμα 2.2.5 παρουσιάζεται ένα παράδειγμα μακρό θέσεων και ένα παράδειγμα μακρό μεταβάσεων.



Σχήμα 2.2.5: Αναπαράσταση Δικτύων Πέτρι με μακρό θέσεις και μακρό μεταβάσεις.

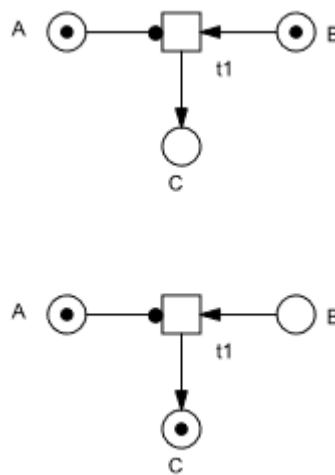
Inhibitor Edge και Read Edge

Τα δίκτυα Πέτρι υποστηρίζουν επιπλέον δύο τύπους τόξων για ενδυνάμωση της εκφραστικότητάς τους. Ο πρώτος κόμβος ονομάζεται Read Edge και ο δεύτερος Inhibitor Edge. Οι δύο τύποι κόμβων παρουσιάζονται στο Σχήμα 2.2.6.



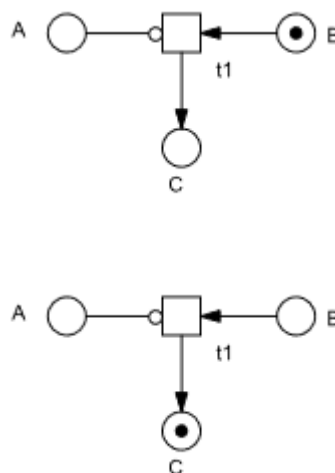
Σχήμα 2.2.6 Αναπαράσταση Inhibitor edge και Read edge

Η Read Edge [5 σελ.10] συνδέει μόνο θέση με μετάβαση. Εάν μια θέση p συνδέεται με Read Edge με μια μετάβαση t , τότε η μετάβαση t είναι ενεργή εάν η p και όλες οι άλλες θέσεις εισόδου της t πληρούν τις προϋποθέσεις πυροδότησης. Μετά την πυροδότηση της μετάβασης t , το σύνολο των μαρκών που είχε η θέση p δεν αλλάζει. Στο παρακάτω Σχήμα παρουσιάζεται ένα παράδειγμα εκτέλεσης με Read Edge.



Σχήμα 2.2.7 : Παράδειγμα πυροδότησης με Read Edge.

Η Inhibitor Edge[5 σελ.10] συνδέει μόνο θέση με μετάβαση. Εάν μια θέση p συνδέεται με Inhibitor Edge με μια μετάβαση t , τότε η μετάβαση t είναι ενεργή εάν η p έχει αριθμό μαρκών μικρότερο από ότι καθορίζει το βάρος του τόξου που τις συνδέει, και εάν έχει επιπλέον άλλες θέσεις εισόδου τότε θα πρέπει να πληρούν τις προϋποθέσεις πυροδότησης. Μετά την πυροδότηση της μετάβασης t , το σύνολο των μαρκών που είχε η θέση p δεν αλλάζει. Στο παρακάτω Σχήμα παρουσιάζεται ένα παράδειγμα εκτέλεσης με Inhibitor Edge.



Σχήμα 2.2.8 : Παράδειγμα πυροδότησης με Inhibitor Edge.

2.2.5 Δομική Ανάλυση και ιδιότητες των Δικτύων Πέτρι.

Δομική ανάλυση των Δικτύων Πέτρι

Με την κατασκευή ενός μοντέλου Δικτύου Πέτρι είναι σημαντικό να γίνει ορθή ανάλυση, έτσι ώστε να γίνει καλύτερη κατανόηση της λειτουργίας του πραγματικού συστήματος και κάποιων ιδιοτήτων που μπορεί να έχουν τα συστατικά του. Για το σκοπό της ανάλυσης ενός Δικτύου Πέτρι χρησιμοποιείται γραμμική άλγεβρα.

Η τοπολογική δομή ενός καθαρού Δικτύου Πέτρι μπορεί να παρασταθεί με ένα πίνακα ακεραίων C , που ονομάζεται πίνακας ροής δεδομένων (flow matrix). Ο πίνακας C είναι ένας πίνακας $n \times m$ διαστάσεων όπου οι n γραμμές αντιστοιχούν στις θέσεις του Δικτύου και οι m στήλες στις μεταβάσεις του. Για την κατασκευή του πίνακα C υποθέτουμε μια διάταξη των στοιχείων του συνόλου των θέσεων και του συνόλου των μεταβάσεων. Έστω το σύνολο των θέσεων $P = \{p_1, p_2, \dots, p_n\}$. Κάθε μαρκάρισμα $M : P \rightarrow \mathbb{N}$ μπορεί να αναπαρασταθεί σαν ένα διάνυσμα :

$$M = (a_1, a_2, \dots, a_n)$$

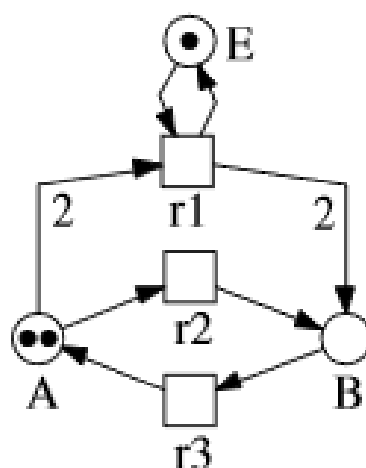
όπου για κάθε a_i συμβολίζει τον αριθμό των μαρκών που έχει θέση p_i . Επίσης ορίζουμε για κάθε μετάβαση ένα διάνυσμα T όπου

$$T = (t(p_1), t(p_2), \dots, t(p_n)).$$

Κάθε $t(p_i)$ ορίζει τη μεταβολή που θα προκαλέσει μια μετάβαση και ορίζεται ως εξής. Έστω $W(\chi, \psi)$ το βάρος της ακμής που ξεκινά από τη θέση χ και καταλήγει στην μετάβαση ψ και $W(\psi, \chi)$ το βάρος της ακμής που ξεκινά από την μετάβαση ψ και καταλήγει στην θέση χ . Τότε έχουμε ότι το

$$t(\psi) = W(\psi, \chi) - W(\chi, \psi)$$

Εάν έχουμε περισσότερες ακμές τότε αθροίζουμε το συνολικό βάρος. Στο Σχήμα 2.2.9 βλέπουμε τον πίνακα C που δημιουργείται από το αντίστοιχο Δίκτυο Πέτρι.



	r1	r2	r3
A	-2	-1	1
B	2	1	-1
E	0	0	0

Σχήμα 2.2.9: Ο πίνακας ροής δεδομένων του Δικτύου Πέτρι

Ιδιότητες Δικτύων Πέτρι

Τα δίκτυα Πέτρι κατέχουν ένα μεγάλο αριθμό ιδιοτήτων. Οι ιδιότητες διακρίνονται σε δύο είδη, στις ιδιότητες που αφορούν την συμπεριφορά του συστήματος και ονομάζονται ιδιότητες συμπεριφοράς και σε αυτές που αφορούν την δομή του συστήματος και ονομάζονται ιδιότητες δομής.

Οι ιδιότητες συμπεριφοράς εξαρτώνται από το αρχικό μαρκάρισμα του Δικτύου και είναι σημαντικές από πλευράς πρακτικότητας. Οι βασικές ιδιότητες συμπεριφοράς περιλαμβάνουν την προσιτότητα (reachability), την περατότητα (boundedness), την ασφάλεια (safety), τη ζωτικότητα (liveness). Ενώ από την άλλη πλευρά έχουμε τις ιδιότητες δομής που εξαρτώνται μόνο από την δομή ή την τοπολογία του Δικτύου. Οι βασικές ιδιότητες δομής περιλαμβάνουν την δομική περατότητα (structural boundedness), δομική ζωτικότητα (structural liveness), συνέπεια (consistence).

Προσιτότητα (Reachability)

Δεδομένου ενός αρχικού μαρκάρισματος M_0 ενός Δικτύου Πέτρι, κάποιο διαφορετικό μαρκάρισμα M είναι προσιτό από το αρχικό μαρκάρισμα M_0 εάν υπάρχει έστω μια ακολουθία πυροδοτήσεων μεταβάσεων που οδηγεί από το μαρκάρισμα M_0 στο M . Το σύνολο όλων των πιθανών προσιτών μαρκαρισμάτων από το M_0 ονομάζεται σύνολο προσιτότητας και συμβολίζεται με $R(M_0)$. Κάποιο μαρκάρισμα M' λέγεται ότι είναι άμεσα προσιτό από το M εάν η πυροδότηση μια ενεργής μετάβασης στο μαρκάρισμα M οδηγεί στο M' .

Το πρόβλημα της προσιτότητας σε ένα Δίκτυο Πέτρι σχετίζεται με την διατύπωση εάν ένα δεδομένο μαρκάρισμα M ανήκει στο σύνολο των μαρκαρισμάτων που είναι προσιτά από το αρχικό μαρκάρισμα M_0 .

Αυτή η ιδιότητα βοηθά το σχεδιαστή να αποφασίσει εάν το σύστημα μπορεί να επέλθει σε μια συγκεκριμένη κατάσταση στο μέλλον. Είναι μια σημαντική ιδιότητα συνήθως για συστήματα κρίσιμα σε ασφάλεια, όπου κάποιες συνθήκες που οδηγούν σε μία ανεπιθύμητη κατάσταση δεν πρέπει ποτέ να ικανοποιηθούν. Για αυτό και η συγκεκριμένη ιδιότητα βοηθάει τους επιστήμονες να διασφαλίσουν το Βιολογικό σύστημα από ανεπιθύμητες βιολογικές αντιδράσεις που εκφέρουν διάφορα προβλήματα. Για παράδειγμα ασθένειες ή ανεπιθύμητες λειτουργίες.

Περατότητα (Boundedness) και Ασφάλεια (Safety)

Σε ένα Δίκτυο Πέτρι με αρχικό μαρκάρισμα M_0 και με σύνολο προσιτότητας $R(M_0)$, μια θέση $p \in P$ είναι k -πεπερασμένη, αν $M(p) \leq k$, για κάθε μαρκάρισμα M που ανήκει στο $R(M_0)$, όπου k θετικός αριθμός. Δηλαδή μια θέση p είναι k -πεπερασμένη, αν ο αριθμός από μάρκες που περιέχει δεν ξεπερνά τον αριθμό k για οποιοδήποτε προσιτό μαρκάρισμα M από το M_0 . Ένα Δίκτυο Πέτρι είναι k -πεπερασμένο (k -bounded) αν κάθε θέση p του Δικτύου είναι k -πεπερασμένη. Αυτή η ιδιότητα καθορίζει ότι ο αριθμός των καταστάσεων που μπορεί να βρεθεί ένα σύστημα είναι πεπερασμένος. Όταν ένα Δίκτυο Πέτρι είναι 1-πεπερασμένο τότε ονομάζεται Ασφαλές. Το Δίκτυο

Πέτρι είναι δομικά πεπερασμένο (structurally bounded), αν είναι πεπερασμένο για κάθε αρχικό μαρκάρισμα M_0 .

Ιδιότητες όπως η περατότητα και η ασφάλεια ενός Δικτύου, καθορίζουν εάν ο αριθμός των μαρκών σε κάθε θέση είναι πεπερασμένος ή όχι.

Ζωτικότητα (Liveness)

Μια μετάβαση t είναι ζωτική (live) αν για οποιοδήποτε μαρκάρισμα $M \in R(M_0)$ ⁵ υπάρχει μια ακολουθία πυροδοτήσεων, η οποία οδηγεί σε ένα μαρκάρισμα που ενεργοποιεί την t . Ένα Δίκτυο Πέτρι είναι ζωτικό αν κάθε μετάβαση σε αυτό είναι ζωτική, δηλαδή αν κάθε μετάβαση t μπορεί να πυροδοτήσει ξεκινώντας από το αρχικό μαρκάρισμα M_0 . Ένα Δίκτυο Πέτρι είναι δομικά ζωτικό (structurally live) αν υπάρχει ένα ζωτικό αρχικό μαρκάρισμα M_0 .

Η ιδιότητα της ζωτικότητας έχει σχέση με την αποφυγή καταστάσεων όπου παρατηρείται αδιέξοδο λειτουργίας. Ένα Δίκτυο Πέτρι περιέχει αδιέξοδο (deadlock) αν υπάρχει ένα μαρκάρισμα $M \in R(M_0)$, τέτοιο ώστε μία ή περισσότερες ή όλες οι μεταβάσεις να μην ενεργοποιούνται. Ένα τέτοιο μαρκάρισμα ονομάζεται νεκρό μαρκάρισμα.

Η ζωτικότητα ενός Δικτύου Πέτρι σημαίνει ότι από ένα οποιοδήποτε μαρκάρισμα M προσιτό από το αρχικό μαρκάρισμα M_0 , είναι απολύτως πιθανό να πυροδοτήσει οποιαδήποτε μετάβαση μέσα στο Δίκτυο, ακολουθώντας την κατάλληλη ακολουθία πυροδοτήσεων. Αυτό εξάγει ως συμπέρασμα ότι ένα ζωτικό Δίκτυο Πέτρι εξασφαλίζει ότι το υπό αναπαράσταση σύστημα δεν εμφανίζει ποτέ αδιέξοδα λειτουργίας κατά την εκτέλεσή του.

⁵ $R(M_0)$ = σύνολο όλων των μαρκαρισμάτων που προκύπτουν από το αρχικό μαρκάρισμα πυροδοτώντας μια ακολουθία μεταβάσεων

Συνέπεια (Consistence)

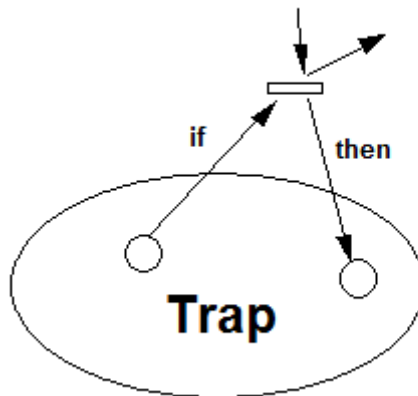
Η ιδιότητα της συνέπειας ανήκει στις ιδιότητες δομής των Δικτύων Πέτρι και ορίζεται ως εξής: Ένα Δίκτυο Πέτρι με αρχικό μαρκάρισμα M_0 ονομάζεται συνεπές (consistent) αν υπάρχει μια ακολουθία πυροδοτήσεων A από το M_0 που καταλήγει πίσω στο M_0 , τέτοια ώστε κάθε μετάβαση του Δικτύου να πυροδοτεί τουλάχιστον μια φορά στην A . Εάν αυτό ισχύει μόνο για κάποιο υποσύνολο των μεταβάσεων τότε καλείται μερικώς συνεπές (partially consistent). Η ιδιότητα της συνέπειας είναι πολύ σημαντική στο κόσμο των βιολογικών συστημάτων, αφού μπορεί να εκτιμηθεί εάν με την πυροδότηση μιας ακολουθίας βιολογικών αντιδράσεων, το Βιολογικό σύστημα μπορεί να επανέρθει στην αρχική του κατάσταση. Ειδικά σε ασθένειες όπως ο καρκίνος που έχει μελετηθεί στην παρούσα μελέτη, και αλλοιώνει κάποια συστατικά και τη λειτουργία του οργανισμού με την ιδιότητα της συνέπειας οι Επιστήμονες προσπαθούν να τον επαναφέρουν στην αρχική του κατάσταση.

Δομικά μοτίβα των Δικτύων Πέτρι

Παγίδα (Trap)

Η παγίδα αναφέρεται σε ένα υποσύνολο Q του συνόλου των θέσεων ενός Δικτύου Πέτρι όπου κάθε μετάβαση που αφαιρεί μια μάρκα από κάποια θέση του Q υπάρχει άλλη μετάβαση που προσθέτει μία μάρκα στο υποσύνολο Q . Για αυτό το λόγο μια παγίδα ποτέ δεν θα είναι κενή (αριθμός των μαρκών μέσα στη παγίδα να ισούται με μηδέν) από την στιγμή που θα λάβει τουλάχιστον μία μάρκα.

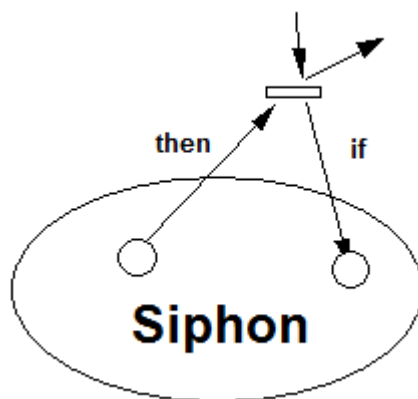
Μια τέτοια ιδιότητα θα μπορούσε να είναι πολύ χρήσιμη αφού εγγυάται την παρουσία τουλάχιστον μίας μάρκας. Εάν δηλαδή γνωρίζουμε ότι το αρχικό μαρκάρισμα προνοεί μία μάρκα σε μια παγίδα τότε για κάθε προσβάσιμο μαρκάρισμα η παγίδα θα έχει τουλάχιστον μία μάρκα. Δηλαδή η μάρκα είναι παγιδευμένη και δεν μπορεί να διαφύγει. Κυκλικές δομές σε συστήματα που ενεργοποιούνται από μια είσοδο συνήθως αναπαριστώνται στο Δίκτυο ως παγίδα. Στο Σχήμα 2.2.10 βλέπουμε τον ορισμό της παγίδας σχηματικά. Ο ορισμός της παγίδας στα βιολογικά συστήματα μπορεί να αποδώσει διάφορες καταστάσεις όπου αν ο οργανισμός περιέλθει σε αυτές δεν μπορεί πλέον να ξεφύγει.



Σχήμα 2.2.10: Ορισμός της παγίδας στα δίκτυα Πέτρι σχηματικά.

Σιφόνι (Siphon)

Μια παρόμοια ιδιότητα είναι το σιφόνι (Siphon). Αναφέρεται σε ένα υποσύνολο Q του συνόλου των θέσεων ενός Δικτύου Πέτρι όπου κάθε μετάβαση που προσθέτει μία μάρκα στο υποσύνολο Q επίσης υπάρχει κάποια άλλη μετάβαση που αφαιρεί μία μάρκα από αυτό. Χρησιμοποιείται για να ελεγχθεί η διατήρηση της κενότητας ενός υποσυνόλου, αφού εάν ένα σιφόνι αδειάσει από όλες τις μάρκες θα παραμείνει για πάντα άδειο. Το σιφόνι στα βιολογικά συστήματα ορίζει κάποιες καταστάσεις όπου ο οργανισμός αν καταφέρει να ξεφύγει από αυτές, εγγυάται ότι στο μέλλον δεν θα επανέλθει σε αυτές. Στο Σχήμα 2.4.8 βλέπουμε τον ορισμό της παγίδας σχηματικά.



Σχήμα 2.2.11: Ορισμός του σιφονιού στα δίκτυα Πέτρι σχηματικά.

2.2.6 Πλεονεκτήματα χρήσης των Δικτύων Πέτρι

Όπως έχει ειπωθεί πιο πάνω τα δίκτυα Πέτρι και οι επεκτάσεις τους, χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία ως το βασικό εργαλείο για την μοντελοποίηση και μελέτη των βιολογικών συστημάτων. Τα δίκτυα Πέτρι συνδυάζουν ένα καλά ορισμένο γραφικό περιβάλλον με ένα μαθηματικό φορμαλισμό. Το γραφικό περιβάλλον βοηθάει στην καλύτερη κατανόηση του βιολογικού συστήματος και το μαθηματικό υπόβαθρο στην κατάστρωση εξισώσεων και προσομοίωση των μοντέλων που προκύπτουν. Η προσομοίωση αποτελεί το βασικότερο εργαλείο ανάλυσης αποδοτικότητας των Δικτύων Πέτρι, που αυξάνει σημαντικά τη δυνατότητα κατανόησης κι εποπτικής παρακολούθησης της κατάστασης των υπό μελέτη βιολογικών συστημάτων. Επιπλέον, το μαθηματικό πλαίσιο διευκολύνει την χρήση των κατασκευαζόμενων μοντέλων για ποιοτική και ποσοτική ανάλυση των συστημάτων που αναπαριστούν.

Επιπρόσθετα, τα μοντέλα που κατασκευάζονται με Δίκτυα Πέτρι είναι πιο συμπυκνωμένα σε σχέση με τα αντίστοιχα που προκύπτουν με την χρήση άλλων εργαλείων μοντελοποίησης. Ακόμη έχουν αναπτυχθεί πλήρεις τεχνικές για ανάλυση ιδιοτήτων των μοντέλων που υλοποιούνται και έχει αναπτυχθεί μεγάλη ποικιλία προγραμμάτων σχεδιασμού, προσομοίωσης και ανάλυσης μοντέλων υλοποιημένων με διάφορες κατηγορίες Δικτύων Πέτρι, αρκετά από τα οποία προσφέρονται δωρεάν στο διαδίκτυο. Συνδυασμός των μοντέλων με αυτά τα προγράμματα, μπορεί να επιφέρει

περεταίρω ανάλυση για πιο εξειδικευμένες ανάγκες Αυτό είναι και ένα ισχυρό πλεονέκτημα για την χρήση τους πάνω σε βιολογικά συστήματα.

Επίσης ένα χαρακτηριστικό των βιολογικών συστημάτων είναι η πολυπλοκότητα και οι λεπτομέρειες που τα διακατέχουν. Με την χρήση των Δικτύων Πέτρι εισάγεται μια μέθοδος μοντελοποίησης βασισμένη στη χρήση μικρού αριθμού γενικευμένων θεμελιωδών υποσυστημάτων. Με τον τρόπο αυτό η μοντελοποίηση, η ανάλυση ιδιοτήτων και της συμπεριφοράς καθώς και ο εποπτικός έλεγχος του μοντέλου συστήματος, ανάγονται σε ένα σύνολο από υπό-προβλήματα μικρότερης πολυπλοκότητας, διατυπώνοντας με τρόπο συνεπή ως προς τους περιορισμούς και τα χαρακτηριστικά του αρχικού προβλήματος. Οι λύσεις των προβλημάτων αυτών συντίθενται ακολουθώντας μια καλά ορισμένη διαδικασία, για να σχηματίσουν την λύση του συνολικού προβλήματος. Στον περιορισμό της πολυπλοκότητας της μεθόδου συντελεί και η μοντελοποίηση συστημάτων παραγωγής με τυπικά μεγάλους αριθμούς από γεγονότα και πληροφορίες, όπως τα βιολογικά συστήματα.

2.3 Εξόρυξη διαδικασιών

2.3.1 Εισαγωγή

Η Εξόρυξη διαδικασιών (Process mining) είναι ένας νέος ερευνητικός τομέας ο οποίος συνδυάζει την υπολογιστική νοημοσύνη, την εξόρυξη δεδομένων και την μοντελοποίηση και ανάλυση των διαδικασιών. Η έννοια της Εξόρυξης διαδικασιών συνίσταται στην ανακάλυψη διαδικασιών, στην εποπτεία διαδικασιών και στον εμπλουτισμό των διαδικασιών επί της ουσίας μέσω της εξαγωγής γνώσης από το ιστορικό των γεγονότων (log event) το οποίο υπάρχει συνήθως διαθέσιμο στα σημερινά συστήματα. Η εξόρυξη των διαδικασιών περιλαμβάνει την αυτόματη ανακάλυψη διαδικασιών (Process Discovery), τον έλεγχο συμμόρφωσης (Conformance checking), την αυτόματη δημιουργία, την προσομοίωση, την επέκταση και την επιδιόρθωση μοντέλων [7], καθώς επίσης και την πρόγνωση περιστατικών βάση ιστορικών στοιχείων που λήφθηκαν.

Οι τεχνικές Εξόρυξης διαδικασιών (Process Mining) έχουν τη δυνατότητα να εξάγουν γνώση από αρχεία καταγραφών (events logs), τα οποία συνήθως διατίθενται στα σημερινά πληροφοριακά συστήματα. Οι τεχνικές αυτές αποτελούν τα μέσα με τα οποία είναι δυνατόν να ανακαλυφθούν να εποπτευθούν, ακόμη και να βελτιωθούν οι διαδικασίες σε διάφορα πεδία και κλάδους. Δύο βασικοί παράγοντες συντελούν συνεχώς στην αύξηση του ενδιαφέροντος για την Εξόρυξη διαδικασιών. Αφενός, ολοένα και περισσότερα γεγονότα καταγράφονται, παρέχοντας ένα λεπτομερές ιστορικό των διαδικασιών. Αφετέρου, σε ένα ανταγωνιστικό και ραγδαία μεταβαλλόμενο περιβάλλον, προκύπτει η ανάγκη για βελτίωση της δομής και της λειτουργίας διάφορων συστημάτων.

Για την εκπόνηση της παρούσας διπλωματικής εργασίας έχει γίνει μια εκτενής έρευνα για το πώς θα μπορούσε η Εξόρυξη διαδικασιών να συνδυαστεί με βιολογικά συστήματα για την εξαγωγή ωφέλιμων συμπερασμάτων πάνω στα βιολογικά μοντέλα που κατασκευάζονται από την τεχνική αυτόματης ανακάλυψης διαδικασιών.

Ο στόχος της Εξόρυξης διαδικασιών είναι η κατασκευή μοντέλων διαδικασιών που να είναι σε θέση να περιγράφουν τη δομή και τη λειτουργία συστημάτων. Στην Επιστήμη της Πληροφορικής ένα μοντέλο ορίζει κάποια συγκεκριμένη αναπαράσταση κομματιού από τον πραγματικό κόσμο. Τα μοντέλα χρησιμοποιούνται για να μειώσουν την πολυπλοκότητα της αναπαράστασης ορισμένων χαρακτηριστικών και την παράβλεψη άλλων. Στον κλάδο της Βιολογίας, ένα μοντέλο διαδικασιών απεικονίζει μια βιολογική διαδικασία που περιγράφει εξαρτήσεις μεταξύ βιολογικών αντιδράσεων και βιολογικών ουσιών.

Τα μοντέλα διαδικασιών μπορούν να αναπαρασταθούν σε πολλές διαφορετικές γλώσσες διαδικασιών μοντέλων. Για παράδειγμα BPMN (Business Process and Notation), EPC (Event Driven Process Chain), Δίκτυα Πέτρι (Petri nets).

Για την εκπόνηση τη παρούσας διπλωματικής εργασίας για μοντέλα διαδικασιών επιλέχτηκαν τα δίκτυα Πέτρι καθώς αντιπροσωπεύουν την κυρίαρχη γλώσσα μοντελοποίησης στο τομέα της Εξόρυξης διαδικασιών. Καθώς τα δίκτυα Πέτρι είναι αρκετά εκφραστικά και δυνατά μοντέλα και ο ρόλος τους είναι θεμελιώδης όσον αφορά

την ανάλυση συστημάτων τα οποία είναι συντρέχοντα, καταναμεμένα, παράλληλα, και γενικά μη ντετερμινιστικά. (Λεπτομερής περιγραφή των Δικτύων Πέτρι έχει γίνει στο Κεφάλαιο 2.2.) Συγκεκριμένα χρησιμοποιήθηκε μια υποκλάση των Δικτύων Πέτρι που ονομάζονται δίκτυα ροής εργασιών (Workflow net).

Έστω $N = (P, T, F, A, l)$ όπου P ένα (labeled) Δίκτυο Πέτρι. N είναι ένα Δίκτυο ροής εργασιών αν και μόνο αν προστεθεί μια έξτρα θέση (i) στο Δίκτυο Πέτρι P τέτοια ώστε να μην έχει προ-μεταβάσεις (ακμές που ξεκινούν από μια μετάβαση και καταλήγουν σε αυτή τη θέση), όπως επίσης και μια θέση (o) που δεν έχει μετά-μεταβάσεις. (ακμές από την o και να καταλήγουν σε οποιαδήποτε μετάβαση). Η θέση (i) ονομάζεται θέση πηγής (source place) και από αυτή ξεκινούν όλα τα μονοπάτια που περιγράφει το μοντέλο και η θέση (o) ονομάζεται θέση κατανάλωσης (sink place) και σε αυτή τερματίζουν όλα τα μονοπάτια που περιγράφει το μοντέλο.

2.3.2 Βασικές έννοιες της Εξόρυξης διαδικασιών

Πιο κάτω περιγράφονται οι βασικοί ορισμοί που χρησιμοποιούνται στην τεχνική Εξόρυξης διαδικασιών.

Εργασία (event) : Έστω ότι E είναι ο κόσμος όλων των εργασιών. Δηλαδή το σύνολο με όλες τις πιθανές εργασίες που μπορούν να καθοριστούν. Μια εργασία $e \in E$ μπορεί να έχει διάφορα χαρακτηριστικά. Για παράδειγμα μια εργασία μπορεί να έχει ημερομηνία και χρονική στιγμή εκτέλεσης, όνομα, άτομο που την εκτελεί και κόστος. Ας ονομάσουμε XE το σύνολο των χαρακτηριστικών που μπορεί να έχει μια εργασία. Για οποιαδήποτε εργασία $e \in E$ και οποιοδήποτε χαρακτηριστικό $n \in XE$ ορίζεται ότι $\#_n(e)$ είναι η τιμή του χαρακτηριστικού n για την ενέργεια e . Εάν η ενέργεια e δεν έχει κανένα χαρακτηριστικό τότε ορίζεται ως $\#_n(e) = \perp$ (κενή τιμή).

Για ευκολία υποθέτουμε τα εξής χαρακτηριστικά:

#όνομα(ε) : Το όνομα της διαδικασίας που ανήκει ενέργεια.

#χρόνος(ε) : Η χρονική στιγμή που γίνεται η ενέργεια.

#πόροι(ε) : Οι πόροι που σχετίζονται με την ενέργεια.

#προγραμματισμός(ε) : Σε ποια φάση του κύκλου ζωής βρίσκεται η εργασία. Έναρξη, όπου ορίζει ότι η ενέργεια ξεκινάει από αυτό το σημείο, Συνέχεια, όπου η ενέργεια συνεχίζει από το σημείο που έμεινε στο παρελθόν, Ολοκλήρωση, η ενέργεια ολοκληρώνεται σε αυτό το σημείο.

Για την χρήση των πιο πάνω χαρακτηριστικών θα πρέπει να προσεχθούν κάποια πράγματα. Για να ολοκληρωθεί μια διαδικασία περιλαμβάνει μια ακολουθία διαφορετικών ενεργειών. Πολλές φορές αναφερόμαστε σε μια ενέργεια με το όνομα της διαδικασίας στην οποία ενσωματώνεται. Τεχνικά αυτό είναι λάθος καθώς μπορεί να υπάρχουν πολλές ενέργειες που εμπεριέχονται στην ολοκλήρωση της ίδιας διαδικασίας. Στην παρούσα μελέτη κάθε ενέργεια θα παίρνει το όνομα του κανόνα της αντίδρασης που γίνεται ή το όνομα της βιολογικής οντότητας που παράγεται/καταναλώνεται. Επίσης όσο προχωράμε στο αρχείο καταγραφής θα πρέπει ο χρόνος να προχωράει προς τα μπροστά. Ένα παράδειγμα με τον τρόπο που δηλώνονται τα χαρακτηριστικά στη γλώσσα προγραμματισμού που δέχεται και το εργαλείο που χρησιμοποιήθηκε, είναι στο Σχήμα 2.3.1

```
<event>
<string key="org:resource" value="Undefined"/>
<date key="time:timestamp" value="2016-04-16T10:10:01.527"/>
<string key="concept:name" value="a3"/>
<string key="lifecycle:transition" value="complete"/>
</event>
```

Σχήμα 2.3.1: Παρουσιάζεται ο τρόπος με τον οποίο δηλώνονται τα χαρακτηριστικά μιας ενέργειας. Η αντίδραση a3 ενεργοποιήθηκε στις 16/04/2016 στις 10:10:01.527 χωρίς να έχουν δηλωθεί οι πόροι και με το πέρας της έχει ολοκληρωθεί ο κύκλος της.

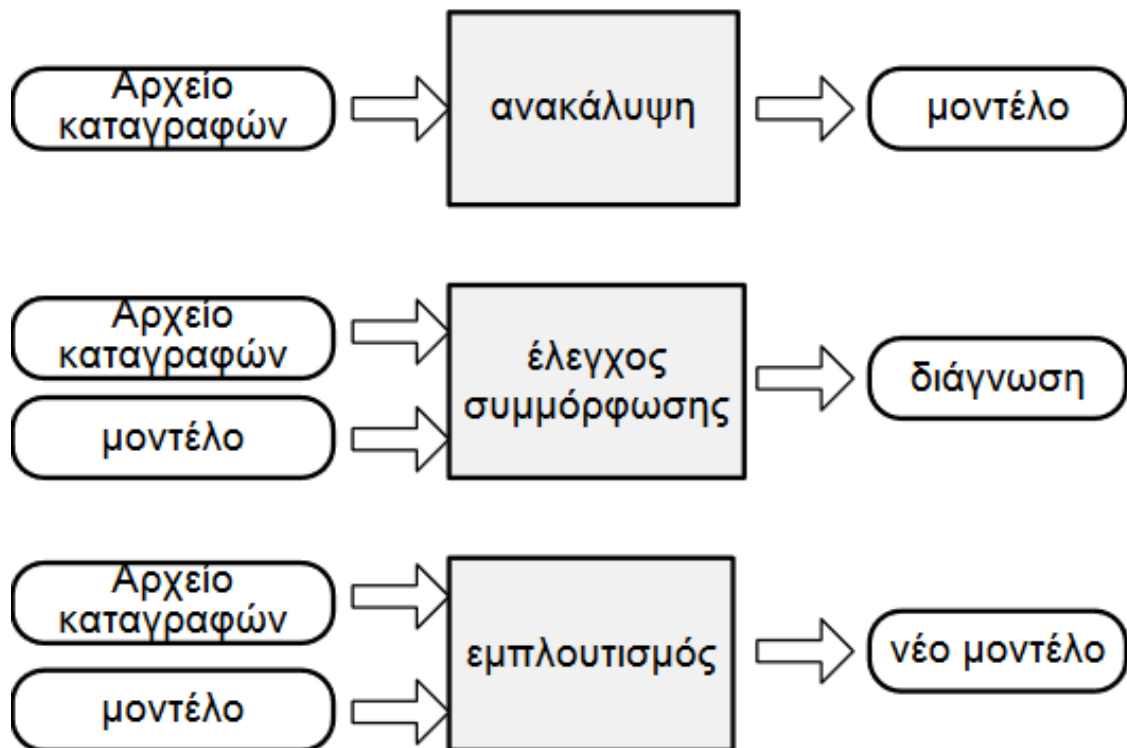
Κατάσταση (case) : Έστω ότι C ο κόσμος όλων των πιθανών καταστάσεων. Κάθε κατάσταση $c \in C$ έχει τα δικά της χαρακτηριστικά. Ας ονομάσουμε XK το σύνολο των χαρακτηριστικών που μπορεί να έχει μια κατάσταση. Για οποιαδήποτε κατάσταση $c \in C$ και $n \in XK$ ορίζεται ότι $\#_n(c)$ είναι η τιμή του χαρακτηριστικού n στην κατάσταση c . Εάν η κατάσταση c δεν έχει κανένα χαρακτηριστικό τότε ορίζεται ως $\#_n(c) = \perp$ (κενή τιμή). Κάθε κατάσταση έχει ένα υποχρεωτικό χαρακτηριστικό που ονομάζεται ίχνος (trace), όπου $\#_{\text{ίχνος}}(c) \in E^*$. Θα ορίζουμε ως $c' = \#_{\text{ίχνος}}(c)$ για γρήγορη αναφορά στο ίχνος μιας κατάστασης.

Ίχνος (trace) : Μια πεπερασμένη ακολουθία από ενέργειες $\sigma \in E$, όπου κάθε ενέργεια εμφανίζεται μια φορά μέσα στο ίχνος. Δηλαδή εάν $1 \leq i < j \leq |\sigma| : \sigma(i) \neq \sigma(j)$.

Αρχείο καταγραφής (log event) : Ένα αρχείο καταγραφής είναι ένα σύνολο από καταστάσεις $L \subseteq C$ όπου κάθε κατάσταση δεν μπορεί να είναι κενή. Εάν το αρχείο καταγραφής περιέχει την έννοια του χρόνου, τότε θα πρέπει η χρονική σειρά των ενεργειών μέσα στα ίχνη να ακολουθείται πιστά. Δηλαδή για κάθε κατάσταση c που ανήκει στο αρχείο καταγραφής L ($c \in L$) και $i, j \in \mathbb{N}$, τότε $1 \leq i < j \leq |c'| : \#_{\text{χρόνος}}(c'(i)) \leq \#_{\text{χρόνος}}(c'(j))$.

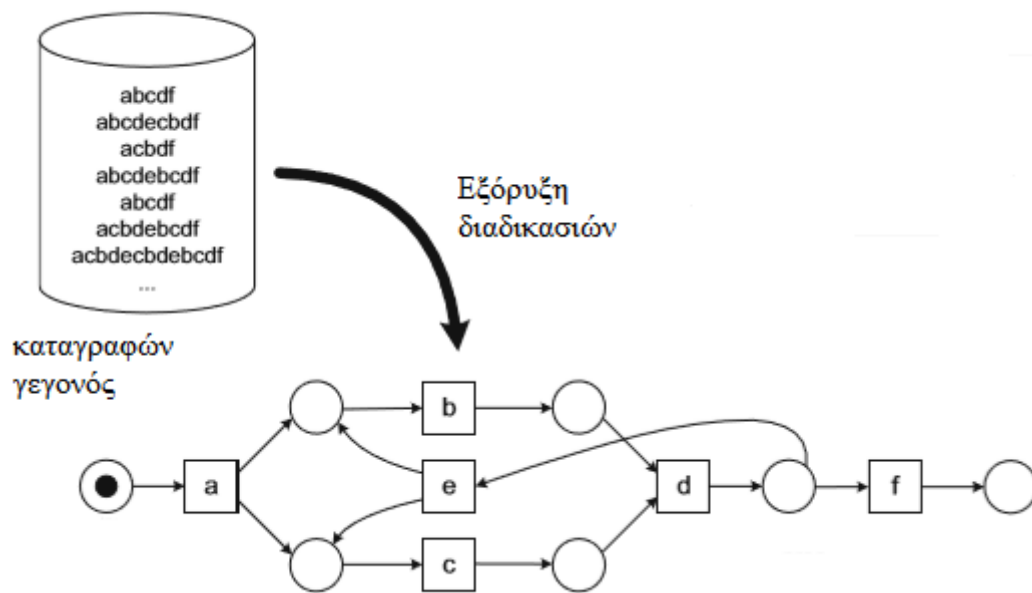
Οι βασικές λειτουργίες στην Εξόρυξη διαδικασιών

Όπως έχει αναφερθεί η Εξόρυξη διαδικασιών περιέχει τρεις βασικές λειτουργίες. Την ανακάλυψη διαδικασιών, τον έλεγχο συμμόρφωσης και την επέκταση (εμπλουτισμό). Στο Σχήμα 2.3.2 παρουσιάζονται οι τρεις βασικές τεχνικές και στη συνέχεια του υποκεφαλαίου περιγράφεται η χρησιμότητα της κάθε μιας.



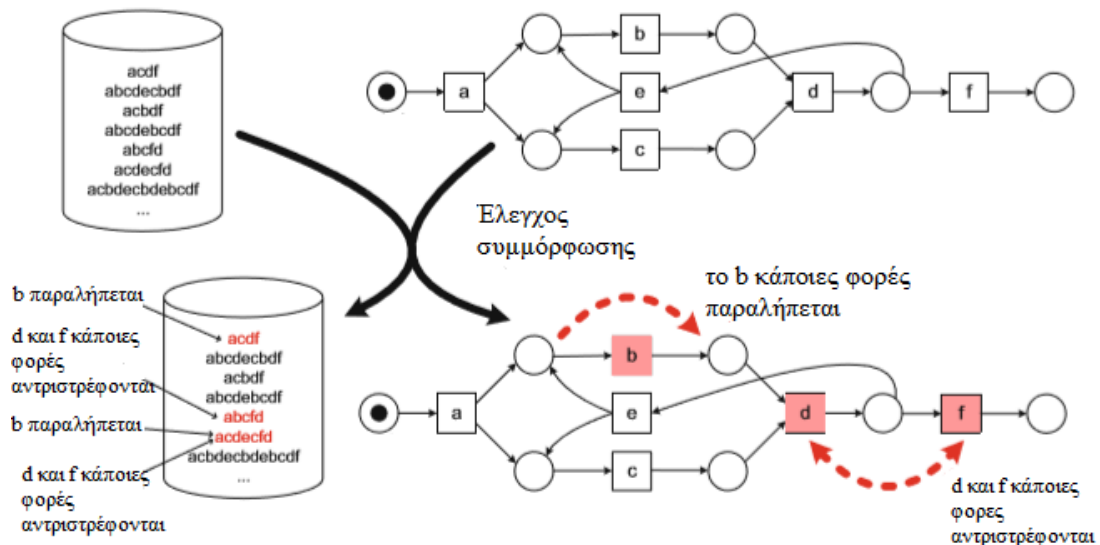
Σχήμα 2.3.2: Οι τρεις βασικές λειτουργίες της Εξόρυξης διαδικασιών με τα δεδομένα εισόδου και το αποτέλεσμα που δίνει η κάθε μια.

Στην αυτόματη ανακάλυψη διαδικασιών (Process Discovery) κατασκευάζεται αυτόματα ένα μοντέλο διαδικασίας από ένα αρχείο καταγραφής. Δίνοντας το ιστορικό ενός γεγονότος με τα στοιχεία του, αυτόματα δημιουργείται ένα μοντέλο που περιγράφει τη συμπεριφορά του σύμφωνα με το γεγονός που αναλύει. Τα γεγονότα που δίνονται ως είσοδος περιγράφουν παραδείγματα συμπεριφοράς και δεν παρουσιάζουν τι είναι πιθανόν να γίνει στο μέλλον. (Στο Σχήμα 2.3.3 βλέπουμε ένα παράδειγμα της συγκεκριμένης διαδικασίας). Για την ανακάλυψη διαδικασιών θα γίνει εκτενής περιγραφή στη συνέχεια της διπλωματική εργασίας καθώς είναι και η βασικό τεχνική για την υλοποίηση αυτής της μελέτης.



Σχήμα 2.3.3: Απεικονίζεται ένα παράδειγμα της τεχνικής ανακάλυψης διαδικασιών. Δίνεται ως είσοδο ένα αρχείο καταγραφής. Αναλύοντας τα ίχνη του γεγονότος δημιουργείται ένα μοντέλο ώστε να περιγράφει τη συμπεριφορά που παρατηρείται στο αρχείο καταγραφής.

Στον έλεγχο συμμόρφωσης (Conformance checking) γίνεται σύγκριση ενός υπάρχοντος μοντέλου διαδικασίας με ένα αρχείο καταγραφής που περιγράφει την συγκεκριμένη διαδικασία. Γίνεται έλεγχος εάν η πραγματικότητα σύμφωνα με το γεγονός, συμφωνεί με το μοντέλο και το αντίστροφο. Στην περίπτωση που δεν συμφωνούν μεταξύ τους τότε ή δημιουργείται ένα νέο μοντέλο με την διαδικασία ανακάλυψης είτε γίνεται επιδιόρθωση πάνω στο μοντέλο εισόδου. Με τη σάρωση του γεγονότος χρησιμοποιώντας κάποιο ήδη υπάρχον μοντέλο μπορούν να εντοπιστούν πιθανές περιπτώσεις σφαλμάτων. Ως εκ τούτου, ο έλεγχος συμμόρφωσης μπορεί να εφαρμοστεί για τον εντοπισμό αυτών των σφαλμάτων και να μετρηθεί ο βαθμός απόκλισης μεταξύ του αρχείου καταγραφής και του ήδη κατασκευασμένου μοντέλου. Ένα παράδειγμα αλγόριθμου ελέγχου συμμόρφωσης δίνεται στο [9]. Στο Σχήμα 2.3.4 βλέπουμε ένα παράδειγμα της συγκεκριμένης διαδικασίας.



Σχήμα 2.3.4: Απεικονίζεται ένα παράδειγμα της τεχνικής ελέγχου συμμόρφωσης. Δίνεται ως είσοδο ένα αρχείο καταγραφής μαζί με ένα μοντέλο που το περιγράφει. Αναλύοντας τα ίχνη του γεγονότος με το μοντέλο που δόθηκε παρατηρούμε κάποιες διαφορές μεταξύ τους. Μετά την εργασία της συγκεκριμένης τεχνικής γίνονται κάποιες επιδιορθώσεις πάνω στο μοντέλο και επιστρέφεται με κάποιες διαφοροποιήσεις.

Ο τρίτος τύπος Εξόρυξης διαδικασιών είναι η ενίσχυση (enhancement). Η ιδέα της ενίσχυσης είναι η επέκταση ή η βελτίωση ενός μοντέλου διαδικασίας χρησιμοποιώντας πληροφορίες σύμφωνα με κάποιο αρχείο καταγραφής. Ενώ ο έλεγχος συμμόρφωσης συγκρίνει το υπάρχον μοντέλο με το αρχείο καταγραφής και μπορεί να δημιουργήσει ένα νέο μοντέλο, η ενίσχυση στοχεύει στην επέκταση ή αλλαγή του υπάρχοντος μοντέλου. Ένας τύπος ενίσχυσης είναι η επισκεύαση (repair) όπου γίνεται επεξεργασία στο υπάρχον μοντέλο ώστε να αντιλαμβάνεται καλύτερα την πραγματικότητα. (Για περισσότερες πληροφορίες και περιγραφή αλγορίθμων για το πως γίνεται η επισκεύαση υπάρχουν στο [7]). Για παράδειγμα εάν δύο ενέργειες μοντελοποιούνται ακολουθιακά αλλά στην πραγματικότητα μπορούν να συμβούν σε οποιαδήποτε σειρά τότε με τη μεθοδολογία της επισκεύασης το μοντέλο θα το αντιλαμβάνεται. Ένας άλλος τύπος

ενίσχυσης είναι η επέκταση (extension), στην περίπτωση αυτή μετά από προσθήκη κάποιων ενεργειών στο μοντέλο, γίνεται καλύτερη επέκταση στο υπάρχον μοντέλο διαδικασίας ώστε να αντιλαμβάνεται τις νέες ενέργειες χωρίς αυτό να του δημιουργήσει προβλήματα συμφόρησης, πόρων ή οποιοδήποτε άλλων κανόνων που σχετίζονται με τις ενέργειες.

2.3.3 Η τεχνική της ανακάλυψης διαδικασιών

Η ανακάλυψη διαδικασιών προκαλεί μεγάλη πρόκληση για τη σημερινή εποχή γύρω από τον τομέα της Εξόρυξης διαδικασιών. Βασισμένη σε κάποιο αρχείο καταγραφής κατασκευάζει ένα μοντέλο που περιγράφει όσο το δυνατόν καλύτερα την συμπεριφορά του. Σε αυτή τη τεχνική είναι πολύ σημαντικό να δοθεί ιδιαίτερη σημασία στη διαδικασία ανακάλυψης και στον έλεγχο ροής. Αυτός ο συνδυασμός παίζει καθοριστικό ρόλο στην διαδικασία Εξόρυξης διαδικασιών.

Ο βασικός ορισμός που δόθηκε από τον Wil M.P. van der Aalst στο [10] για την ανακάλυψη διαδικασιών είναι ο εξής : “Έστω L ένα αρχείο καταγραφής. Ο αλγόριθμος της τεχνικής ανακάλυψης διαδικασιών είναι μια λειτουργία η οποία διαβάζει ένα αρχείο καταγραφής και δημιουργεί ένα μοντέλο το οποίο παρουσιάζει την πληροφορία που άντλησε από το αρχείο.”

Όπως έχει δηλωθεί στην τεχνική ανακάλυψης διαδικασιών, το μοντέλο που δημιουργείται θα πρέπει να αντιπροσωπεύει τη συμπεριφορά που παρατηρεί στο αρχείο καταγραφής που του δόθηκε ως είσοδος. Δηλαδή κάθε ίχνος που υπάρχει στο αρχείο καταγραφής θα πρέπει να υπάρχει μια ακολουθία στο Δίκτυο ροής ενεργειών από τη θέση πηγής μέχρι τη θέση καταναλωτής που το περιγράφει. Η ιδιότητα αυτή ονομάζεται καταλληλότητα (fitness) και είναι η βασική ιδιότητα που πρέπει να έχει ένα ορθό Δίκτυο ροής δεδομένων που περιγράφει τη συμπεριφορά κάποιου αρχείου καταγραφής. Γενικά υπάρχουν τέσσερα κριτήρια βαθμολογίας του παραγόμενου μοντέλου στην Εξόρυξη διαδικασιών.

Καταλληλότητα (Fitness): Στο μοντέλο που ανακαλύφθηκε θα πρέπει να υπάρχει τουλάχιστον μια ακολουθία ενεργειών που περιγράφει το κάθε ίχνος που καταγράφηκε στο αρχείο που δόθηκε ως είσοδο.

Ακρίβεια (Precision): Το μοντέλο που ανακαλύφθηκε δεν πρέπει να επιτρέπει την περιγραφή ενεργειών που δεν περιέχονται στο αρχείο καταγραφής.

Απλότητα (Simplicity): Το μοντέλο που ανακαλύφθηκε θα πρέπει να είναι όσο το δυνατόν πιο απλό γίνεται.

Γενίκευση (Generalization): Το μοντέλο που ανακαλύφθηκε θα πρέπει να γενικεύει στην περίπτωση που δει κάτι καινούριο. Για παράδειγμα μια καινούρια προσθήκη ενέργειας ή ίχνους στο αρχείο καταγραφής.

Ένα μοντέλο που έχει καλή καταλληλότητα είναι σε θέση να περιγράφει την πλειοψηφία των ιχνών στο αρχείο καταγραφής. Η Ακρίβεια σχετίζεται με την περιοχή της ακαταλληλότητας (underfitting) που παρουσιάζεται στην Εξόρυξη διαδικασιών. Δηλαδή ένα μοντέλο με χαμηλό ποσοστό ακρίβειας επιτρέπει την παρουσίαση συμπεριφοράς που είναι εντελώς διαφορετική από αυτήν που παρατηρείται στο αρχείο καταγραφής. Η γενίκευση σχετίζεται με το όρο της υπερεκπαίδευσης. Ένα μοντέλο θεωρείται υπερεκπαιδευμένο όταν δεν μπορεί να γενικεύει σε περίπτωση προσθήκης νέων ενεργειών στα ίχνη του αρχείου καταγραφής, με αποτέλεσμα να επιφέρει τεράστιο κόστος στην επιδιόρθωση του μοντέλου. Το τέταρτο κριτήριο της απλότητας, σχετίζεται με τα λόγια του Occam's Razor που υποστήριξε ότι “Δεν πρέπει να υπερβαίνεται οτιδήποτε, πέρα από το αναγκαίο”. Ακολουθώντας αυτή την αρχή, η διαδικασία εξόρυξης στοχεύει στην δημιουργία του απλούστερου μοντέλου που μπορεί να εξηγήσει την παρατηρούμενη συμπεριφορά σε ένα αρχείο καταγραφής.

2.3.4 Αλγόριθμοι Ανακάλυψης διαδικασιών

Στις μέρες διαφορετικές προσεγγίσεις εφαρμόστηκαν για την κατασκευή των μοντέλων που θα περιγράφουν όσο καλύτερα την συμπεριφορά που παρατηρούν σε ένα αρχείο

καταγραφής. Στην συνέχεια του Κεφαλαίου παρουσιάζονται κάποιοι από τους βασικούς αλγόριθμους της ανακάλυψης διαδικασιών.

Αλγόριθμος Εξόρυξης διαδικασιών με βάση τις περιοχές

Ο αλγόριθμος σχετίζεται με το πρόβλημα της σύνθεσης [27]. Σκοπός του είναι η δημιουργία ενός Δικτύου Πέτρι που θα ισούται με την παρατηρούμενη συμπεριφορά ενός συστήματος μεταβάσεων. Εφευρέθηκε από τους Ehrenfeucht και Rozenberg [15] οι οποίοι εισήγαγαν την έννοια των περιοχών στα μοντέλα, ως ένα σύνολο καταστάσεων όπου χαρακτηρίζουν τις οντότητες. Τα βασικά του μειονεκτήματα είναι ότι δεν αναγνωρίζει ακολουθίες που σχηματίζουν κύκλους καθώς και επίσης δημιουργεί καινούριες θέσεις στο Δίκτυο Πέτρι ακόμη και για οντότητες που ήδη συμπεριλήφθηκαν στο μοντέλο.

Α-αλγόριθμος

Για τον α-αλγόριθμο γίνεται εκτενής περιγραφή του στο Κεφάλαιο 3.3 και παρουσιάζονται παραδείγματα εκτέλεσης καθώς είναι ο αλγόριθμος ανακάλυψης διαδικασιών που χρησιμοποιήθηκε στη παρούσα μελέτη.

Αλγόριθμος Εξόρυξης διαδικασιών βάσει ευρετικών

Οι αλγόριθμοι Εξόρυξης Διαδικασιών που βασίζονται σε ευρετικά δημιουργήθηκαν από τον Ton Weijters⁶, οποίος εφάρμοσε τα ευρετικά για προσέγγιση διάφορων προβλημάτων. Είναι πρακτικοί αλγόριθμοι που μπορούν να αντιμετωπίζουν το θόρυβο (αχρείαστες πληροφορίες από το αρχείο καταγραφής) και να παρουσιάζουν την βασική συμπεριφορά ενός συστήματος ακόμα και αν υπολείπονται λεπτομέρειες από το αρχείο καταγραφής που δέχεται ως είσοδο. Μοιάζουν αρκετά με τον Α-αλγόριθμο καθώς δέχονται ένα αρχείο καταγραφής και παράγουν ένα ευρετικό δίκτυο που περιγράφει τη συμπεριφορά του. Ένα πλεονέκτημα του ευρετικού δικτύου είναι η εύκολη μετατροπή

⁶ http://is.ieis.tue.nl/staff/aweijters/publ_ton.htm

σε άλλους τύπους μοντέλων, για παράδειγμα Δίκτυο Πέτρι για περεταίρω ανάλυση. Συνήθως χρησιμοποιούνται για συστήματα επιχειρήσεων, τα οποία έχουν πολλές διαφορετικές καταστάσεις, καταχωρήσεις αλλά και διαφορετικά αρχεία καταγραφής, τα οποία μπορεί να έχουν μικρές διαφορές. Τέτοιου είδους αλγόριθμοι είναι δημοφιλείς για την ταχύτητα τους χωρίς να επηρεάζεται η ποιότητα του τελικού αποτελέσματος.

Αλγόριθμος Fuzzy

Ο αλγόριθμος Fuzzy⁷ είναι από τους πιο σύγχρονους αλγόριθμους στο τομέα της ανακάλυψης διαδικασιών, αφού τον δημιούργησε ο Christian W. Gunther το 2007. Είναι ο πρώτος αλγόριθμος που κατασκευάστηκε και μπορεί να διαχειρίζεται προβλήματα μεγάλου αριθμού ενεργειών, χωρίς να υπάρχει δομή ανάμεσα στα συστατικά του Δικτύου. Αυτό το καταφέρνει με την χρήση μετρήσεων που απλοποιούν το μοντέλο που παράγεται σε ένα επιθυμητό επίπεδο αφαιρετικότητας. Ο αλγόριθμος δίνει ως έξοδο ένα fuzzy μοντέλο, το οποίο δεν μπορεί να μετατραπεί σε άλλου τύπου μοντέλα, αλλά με την βοήθεια εργαλείων δίνει μια καλά ορισμένη προσομοίωση της συμπεριφοράς του συστήματος που μελετά.

⁷ <http://www.wis.win.tue.nl/~wvdaalst/publications/p400.pdf>

Κεφάλαιο 3

Εξόρυξη διαδικασιών

3.1 Παραγωγή των ιχνών	44
3.1.1 Εισαγωγή	44
3.1.2 Δεδομένα εισόδου αλγόριθμων	46
3.1.3 Αλγόριθμος παραγωγής ιχνών με βάση βιολογικές αντιδράσεις	47
3.1.4 Αλγόριθμος παραγωγής ιχνών με βάση βιολογικές ουσίες	51
3.1.5 Παραγωγή αρχείου εισόδου για το εργαλείο Prom	55
3.2 A-αλγόριθμος	56
3.2.1 Εισαγωγή	56
3.2.2 Σημασιολογία A-αλγορίθμου	57
3.2.3 Βήματα εκτέλεσης A-αλγορίθμου	61
3.2.4 Παράδειγμα εκτέλεσης A-αλγορίθμου	64
3.2.5 Πλεονεκτήματα και μειονεκτήματα A-αλγορίθμου	65
3.3 Εργαλείο Prom	66
3.3.1 Εισαγωγή	66
3.3.2 ProMPM651	67
3.3.3 XESame151	68
3.3.4 ProM651	70

3.1 Παραγωγή ιχνών

3.1.1 Εισαγωγή

Μια μεγάλη πρόκληση για την υλοποίηση της παρούσας μελέτης ήταν πως θα μπορούσε να συνδυαστεί ο ορισμός του ίχνους, όπως δόθηκε στο Κεφάλαιο 2, με τα βιολογικά συστήματα. Στόχος είναι η δημιουργία των κατάλληλων ιχνών που μέσω της

εφαρμογής της τεχνικής ανακάλυψης διαδικασιών θα κατασκευαστεί το μοντέλο που θα περιγράφει καλύτερα τη συμπεριφορά του Βιολογικού συστήματος.

Μετά από μελέτη διάφορων εργαλείων σχετικά με μηχανισμούς παραγωγής ιχνών, δεν βρέθηκε κάποιο εργαλείο που να βοηθούσε για την υλοποίηση αυτού του σκοπού. Βασικό πρόβλημα ήταν ότι στα βιολογικά συστήματα, υπάρχει μεγάλη εξάρτηση μεταξύ των βιολογικών αντιδράσεων. Για παράδειγμα, μια βιολογική αντίδραση A μπορεί να προϋποθέτει την ύπαρξη συγκεκριμένων βιολογικών ουσιών για να μπορεί να πυροδοτήσει. Οι οποίες βιολογικές ουσίες είναι το προϊόν μιας άλλης βιολογικής αντίδρασης B. Επομένως στα ίχνη θα έπρεπε η βιολογική αντίδραση B να καταγράφεται πριν από την βιολογική αντίδραση A.

Λόγω αυτής της στενής σχέσης μεταξύ των βιολογικών ουσιών, αλλά και της μη εύρεσης κάποιου προγράμματος που να μπορεί να καλύπτει με ορθό τρόπο την δημιουργία ιχνών πάνω σε βιολογικά συστήματα, έχουν δημιουργηθεί δύο αλγόριθμοι. Οι συγκεκριμένοι αλγόριθμοι παίρνουν ως δεδομένο όλες τις βιολογικές αντιδράσεις με τα χαρακτηριστικά τους (δηλαδή τις βιολογικές ουσίες που προϋποθέτουν την πυροδότηση της αντίδρασης, τις βιολογικές ουσίες που αποτελούν το προϊόν μετά την πυροδότηση της αντίδρασης) που περιγράφουν τα σηματοδοτικά μονοπάτια που μελετήθηκαν και ακολούθως παράγουν τα ίχνη.

Για την υλοποίηση αυτής της διαδικασίας έχουν δημιουργηθεί συνολικά 4 κλάσεις. Η πρώτη κλάση ονομάζεται 'TraceGenerationReactions' και περιέχει το πρώτο αλγόριθμο που κατασκευάζει τα ίχνη με βάση τις βιολογικές αντιδράσεις. Η δεύτερη κλάση ονομάζεται 'TraceGeneration' και περιέχει το δεύτερο αλγόριθμο, ο οποίος κατασκευάζει τα ίχνη με βάση τις βιολογικές ουσίες. Η τρίτη κλάση ονομάζεται 'CleanTraces' και ο σκοπός της είναι να καθαρίζει τυχόν διπλά ίχνη που δημιουργούνται. Η τέταρτη κλάση ονομάζεται 'Script' και είναι υπεύθυνη για την μετατροπή των ιχνών που δημιουργήθηκαν στην κατάλληλη μορφή που δέχεται το εργαλείο Prom για να παράγει το τελικό μοντέλο (Δίκτυο Πέτρι). Τους κώδικες των δύο κλάσεων 'TraceGenerationReactions' και 'TraceGeneration' που περιέχουν τους δύο αλγόριθμους παραγωγής ιχνών μπορείτε να τους βρείτε στο παράρτημα Α.

Η βασική ιδέα για την υλοποίηση των αλγόριθμων είναι να μελετάται η κίνηση μιας μάρκας μέσα στο Βιολογικό σύστημα. Δηλαδή στον αλγόριθμο παραγωγής των ιχνών

με βάση τις βιολογικές αντιδράσεις, να καταγράφονται όλες οι πιθανές ακολουθίες πυροδοτήσεων των βιολογικών αντιδράσεων. Το μοντέλο που θα παράγεται θα παρουσιάζει τη δομή και την αλληλεπίδραση των Βιολογικών αντιδράσεων απλών ή σύνθετων. Ακόμη να μπορεί να παρουσιάζει τυχόν κύκλους που μπορεί να δημιουργούν οι βιολογικές αντιδράσεις. Στο δεύτερο αλγόριθμο καταγράφεται η ακολουθία των βιολογικών ουσιών που περνά η μάρκα. Το μοντέλο που παράγεται από αυτά τα ίχνη θα πρέπει να δίνει περισσότερες πληροφορίες σχετικά με το ποσοστό εμφάνισης των βιολογικών ουσιών σε σχέση με ολόκληρο το βιολογικό σύστημα και τη σειρά παραγωγής και κατανάλωσης τους. Με την ολοκλήρωση της εργασίας των αλγορίθμων ο χρήστης θα έχει όλες τις πιθανές ακολουθίες από την κίνηση της μάρκα στο Βιολογικό σύστημα και θα κατασκευάζεται το μοντέλο που θα περιγράφει την συμπεριφορά αυτών των ακολουθιών μέσω της ανακάλυψης διαδικασιών.

3.1.2 Δεδομένα εισόδου των αλγορίθμων

Ο πρώτος αλγόριθμος κατασκευάζει τα ίχνη με βάση τη σειρά πυροδότησης των βιολογικών αντιδράσεων και ο δεύτερος με βάση τη σειρά που παράγονται οι βιολογικές ουσίες. Η υλοποίηση των αλγορίθμων είναι γραμμένη στην γλώσσα προγραμματισμού Java, στην πλατφόρμα Eclipse.

Παρομοίως, και οι δύο κλάσεις που περιέχουν τους αλγόριθμους, παίρνουν ως είσοδο ένα αρχείο με όλους τους κανόνες (λογικές εκφράσεις) που περιγράφουν το αντίστοιχο μονοπάτι του Βιολογικού συστήματος που μελετάται (είναι της μορφής του Σχήματος 3.1.1). Επίσης παίρνει ως είσοδο και ένα δεύτερο αρχείο που περιέχει τον αρχικό αριθμό των μαρκών που ορίζεται στην κάθε βιολογική ουσία (είναι της μορφής του Σχήματος 3.1.2). Το πρώτο αρχείο στην πρώτη γραμμή καταγράφει το πλήθος των κανόνων που περιγράφουν το μονοπάτι και ακολούθως σε κάθε γραμμή και ένα κανόνα. Ο πρώτος κανόνας στο Σχήμα 3.1.1 αναλύεται ως εξής: Για να πυροδοτήσει η αντίδραση k_1 , χρειάζεται να καταναλωθεί μια μονάδα από την ουσία 'HMGB1' και μια μονάδα από την ουσία 'RAGEu', το προϊόν της αντίδρασης είναι η παραγωγή μιας μονάδας της ουσίας HMGB1. Ένα υποσύνολο των κανόνων περιέχει το τελεστή "AND" (δηλαδή για να μπορεί να πυροδοτήσει η συγκεκριμένη αντίδραση θα πρέπει να ισχύουν όλες οι προϋποθέσεις της). Για σκοπούς ευκολίας όταν μια αντίδραση παράγει

δύο ουσίες την σπάζουμε σε δύο διαφορετικούς κανόνες (π.χ. οι γραμμές 2 και 3). Το δεύτερο αρχείο σε κάθε γραμμή περιέχει το όνομα της ουσίας και το σύνολο των μαρκών της συγκεκριμένης ουσίας που υποδεικνύει την αρχική της κατάσταση.

```

1 6
2 k1 , HMGB1 = HMGB1 AND RAGEu
3 k1 , RAGEp = HMGB1 AND RAGEu
4 d1 , RAGEu = RAGEp
5 k2 , RAGEp = RAGEp AND PI3Ku
6 k2 , PI3Kp = RAGEp AND PI3Ku
7 d2 , PI3Ku = PI3Kp

```

Σχήμα 3.1.1: Κανόνες για το μονοπάτι ERK-RAS

```

1 HMGB1 5
2 RAGEu 6
3 RAGEp 5
4 PI3Ku 5
5 PI3Kp 5

```

Σχήμα 3.1.2: Αρχικό μαρκάρισμα των ουσιών που περιγράφουν το μονοπάτι ERK-RAS

3.1.3 Αλγόριθμος παραγωγής ιχνών με βάση βιολογικές αντιδράσεις

Αρχικά σε μια δομή δεδομένων (“ArrayList”) που φέρει το όνομα Θέσεις (places) αποθηκεύονται όλα τα ονόματα των βιολογικών ουσιών που περιέχονται στο μονοπάτι υπό μελέτη. Σε μια άλλη δομή δεδομένων (“ArrayList”) που φέρει το όνομα Αντιδράσεις (Reactions) αποθηκεύονται όλα τα ονόματα των βιολογικών αντιδράσεων που περιέχονται στο μονοπάτι. Επίσης σε μία άλλη δομή δεδομένων (“HashMap”) που

φέρει το όνομα Μαρκάρισμα (Marking) αποθηκεύεται ως κλειδί (Key) το όνομα της κάθε βιολογικής ουσίας και ως τιμή (Value) ο αριθμός των μονάδων που αντιστοιχεί στην ουσία ως αρχική κατάσταση. Ακόμη, χρησιμοποιείται η βιβλιοθήκη Pair που περιλαμβάνει αντικείμενα τύπου Pair. Τα αντικείμενα αυτά περιέχουν δύο πεδία. Στο ένα πεδίο τύπου πίνακα String δύο θέσεων αποθηκεύεται στην πρώτη θέση του πίνακα το όνομα της αντίδρασης και στη δεύτερη θέση το προϊόν που παράγεται από τη πυροδότηση του συγκεκριμένου κανόνα. Στο δεύτερο πεδίο του αντικειμένου Pair τύπου ArrayList<String>, αποθηκεύονται όλα τα ονόματα των ουσιών που πρέπει να καταναλωθούν για την πυροδότηση του αντίστοιχου κανόνα. Για παράδειγμα στο Σχήμα 3.1.1 για τον πρώτο κανόνα θα δημιουργηθεί ένα αντικείμενο Pair στο οποίο στο πρώτο πεδίο θα δημιουργηθεί ένα πίνακας τύπου String. Στην πρώτη θέση του πίνακα θα αποθηκευτεί το όνομα της αντίδρασης δηλαδή 'k1' και στην δεύτερη θέση του πίνακα θα αποθηκευτεί το προϊόν του συγκεκριμένου κανόνα δηλαδή 'HMGB1'. Στο άλλο πεδίο θα δημιουργηθεί μια λίστα που θα περιέχει δύο στοιχεία τα 'HMGB1' και 'RAGEu'. Για κάθε κανόνα που υπάρχει στο αρχείο εισόδου δημιουργείται και ένα αντικείμενο Pair. Όλα τα αντικείμενα Pair που δημιουργούνται αποθηκεύονται σε μία άλλη δομή δεδομένων ("List") με το όνομα list. Τέλος, χρησιμοποιείται και μια δομή δεδομένων ("ArrayList") με το όνομα allRegex στην οποία αποθηκεύονται τα ονόματα των ουσιών που καταναλώνονται για την πυροδότηση του κάθε κανόνα. Σκοπός αυτής της δομής είναι για να γνωρίζουμε οποιαδήποτε χρονική στιγμή εάν από μια ή περισσότερες βιολογικές ουσίες μπορεί να πυροδοτήσει κάποια από τις αντιδράσεις. Για παράδειγμα στο Σχήμα 3.1.1 από τον πρώτο και δεύτερο κανόνα θα αποθηκευτεί στη λίστα το 'HMGB1' και 'RAGEu', από τον τρίτο κανόνα το 'RAGEp' και συνεχίζει μέχρι να καλύψει όλους τους κανόνες που περιλαμβάνονται στο αρχείο. Περισσότερες πληροφορίες για την περιγραφή της κάθε δομής δεδομένων και των λειτουργιών της που χρησιμοποιήθηκαν μπορούν να βρεθούν στο ιστότοπο του "JAVA API".⁸

⁸ <https://docs.oracle.com/javase/7/docs/api/overview-summary.html>

Βήματα Αλγόριθμου:

1. Αρχικά επιλέγεται τυχαία μια αντίδραση από τη λίστα αντιδράσεων (reactions) η οποία είναι ο σταθμός αφετηρίας για το χτίσιμο του υπόλοιπου ίχνους. Η συγκεκριμένη αντίδραση αποτελεί την κορυφή⁹ του ίχνους.
2. Γίνεται έλεγχος εάν το προϊόν της αντίδρασης που επιλέχθηκε ως κορυφή περιλαμβάνεται στη λίστα allRegex. Στην περίπτωση που δεν περιλαμβάνεται οδηγεί στο συμπέρασμα ότι το προϊόν της συγκεκριμένης αντίδρασης δεν μπορεί να πυροδοτήσει άλλη αντίδραση ούτε αποτελεί μέρος των προϋποθέσεων για πυροδότηση μιας σύνθετης αντίδρασης (αντίδραση που για να πυροδοτήσει χρειάζεται περισσότερο από μια ουσία) στο μέλλον και αποτελεί το τελευταίο στοιχείο του ίχνους. Το ίχνος που έχει παραχθεί καταγράφεται στο αρχείο εξόδου. Εάν έχει συμπληρωθεί ο αριθμός των ίχνων που έδωσε ο χρήστης ως παράμετρο τότε το πρόγραμμα τερματίζει και επιστρέφει το αρχείο με όνομα "PreTraces.txt" που περιέχει όλα τα ίχνη που παράχθηκαν, αλλιώς επιστρέφει στο βήμα 1 για δημιουργία νέου ίχνους. Στην περίπτωση που περιλαμβάνεται στη λίστα allRegex αυτό σημαίνει ότι από τα παραγόμενα προϊόντα της αντίδρασης μπορεί να πυροδοτηθεί μια νέα αντίδραση ή αποτελούν μέρος των προϋποθέσεων για πυροδότηση νέας αντίδρασης. Ακολουθώντας προχωράμε στο βήμα 3 ώστε να βρεθεί η συγκεκριμένη αντίδραση.
3. Ο αλγόριθμος διαλέγει τυχαία ένα αντικείμενο τύπου Pair από τη list. Το αντικείμενο αντιστοιχεί σε ένα κανόνα. Εάν το προϊόν της κορυφής του ίχνους δεν περιλαμβάνεται στα στοιχεία πυροδότησης του κανόνα που επιλέχθηκε επαναλαμβάνεται το βήμα 3. Αλλιώς στην περίπτωση που περιλαμβάνεται προστίθεται στο ίχνος το όνομα της αντίδρασης αφού θεωρείται πως η αντίδραση είναι ενεργή και πυροδοτείται. Αφαιρείται μία μονάδα ξεχωριστά από όλες τις ουσίες που καταναλώθηκαν για την πυροδότηση της συγκεκριμένης αντίδρασης στο μαρκάρισμα ("HashMap" που περιέχει το σύνολο των μονάδων κάθε ουσίας του Δικτύου). Τέλος αυξάνουμε στο marking το σύνολο των μονάδων του προϊόντος της συγκεκριμένης αντίδρασης κατά μία (π.χ. κατά το κτίσιμο του ίχνους και μετά από κάποιο αριθμό εκτελέσεων η κορυφή είναι η αντίδραση 'k2'. Ο αλγόριθμος

⁹ Κορυφή= Αποτελεί το τελευταίο στοιχείο που εισάχθηκε στο ίχνος

παρατηρεί ότι από τη συγκεκριμένη αντίδραση παράγονται οι ουσίες ‘RAGEp’ και ‘PI3Kp’. Ακολουθώντας ψάχνει και βρίσκει ότι από τα προϊόντα της ‘k2’ και συγκεκριμένα την ουσία ‘PI3Kp’ μπορεί να πυροδοτηθεί η αντίδραση ‘d2’. Αυτό έχει ως αποτέλεσμα να προστεθεί στο ίχνος η αντίδραση ‘d2’, να αφαιρεθεί μια μονάδα από το ‘PI3Kp’ και να προστεθεί μια μονάδα στην ουσία ‘RAGEu’ που είναι το προϊόν της πυροδότησης της ‘d2’).

4. Το όνομα της αντίδρασης που πυροδοτήθηκε αποτελεί το τελευταίο στοιχείο που προστέθηκε στο ίχνος και ταυτόχρονα την νέα κορυφή του ίχνους. Γίνεται έλεγχος εάν το μήκος του ίχνους έχει ξεπεράσει το επιτρεπτό μήκος που πρέπει να έχει κάθε ίχνος που παράγεται, το οποίο δόθηκε ως παράμετρος στο πρόγραμμα από τον χρήστη. Εάν το έχει ξεπεράσει τότε το ίχνος έχει ολοκληρωθεί και το καταγράφουμε στο αρχείο εξόδου. Εάν έχει συμπληρωθεί ο αριθμός των ιχνών που έδωσε ο χρήστης ως παράμετρο τότε το πρόγραμμα τερματίζει και επιστρέφει το αρχείο με το όνομα “PreTraces.txt” που περιέχει όλα τα ίχνη που παράχθηκαν, αλλιώς επιστρέφει στο βήμα 1 για δημιουργία νέου ίχνους. Στην αντίθετη περίπτωση, δηλαδή που το μήκος του ίχνους δεν έχει ξεπεράσει το μήκος που πρέπει να έχει κάθε ίχνος συνεχίζουμε με το βήμα 2.

1	1.	k2, d2, k2, d2, k2
2	2.	d1, k1, k2, d2, k2
3	3.	d2, k2, k2, d2, k2
4	4.	k2, k2, d2, k2, d2
5	5.	d2, k2, d2, k2, k2
6	6.	d2, k2, k2, k2, k2
7	7.	d2, k2, d2, k2, d1
8	8.	k1, k1, k1, d1, k1
9	9.	k1, k1, k1, k1, k1
10	10.	k2, k2, d1, k1, k1

Σχήμα 3.1.3: Αποτέλεσμα εκτέλεσης αλγόριθμου παραγωγής ιχνών με βάση τις αντιδράσεις, με αρχεία εισόδου Σχήμα 1 και Σχήμα 2 και παραμέτρους από χρήστη Σύνολο ιχνών : 10, Μέγιστο μήκος ίχνους : 5

Ορθότητα Αλγορίθμου

Ο αλγόριθμος τερματίζει:

1. Δίνεται ως παράμετρος από τον χρήστη το σύνολο των ιχνών που θα παραχθούν. Όταν συμπληρωθεί ο αριθμός των ιχνών τότε το πρόγραμμα τερματίζει και δίνεται ως έξοδος το αρχείο με τα ίχνη.
2. Κάθε ίχνος ολοκληρώνεται:
 - Όταν από τα προϊόντα της αντίδρασης της κορυφής του ίχνους δεν μπορεί να πυροδοτηθεί οποιαδήποτε άλλη αντίδραση από τους κανόνες που περιλαμβάνονται στο αρχείο εισόδου.
 - Όταν το μήκος του ίχνους φθάσει το μέγιστο μήκος που μπορεί να έχει κάθε ίχνος το οποίο καθορίζεται από τον χρήστη ως παράμετρος στο πρόγραμμα.

Ο αλγόριθμος βγάζει ορθό αποτέλεσμα:

Για να συμπεριληφθεί μια αντίδραση στο ίχνος υπάρχουν δύο σενάρια.

1. Η αντίδραση επιλέχθηκε τυχαία στην αρχή ως η αντίδραση που θα χτιστεί επάνω της το υπόλοιπο ίχνος.
2. Οι ουσίες που χρειάζονται να πυροδοτήσουν για την συγκεκριμένη αντίδραση παράγονται από την αμέσως προηγούμενη αντίδραση στο συγκεκριμένο ίχνος.

3.1.4 Αλγόριθμος παραγωγής ιχνών με βάση τις βιολογικές ουσίες:

Δομικά στοιχεία αλγορίθμου:

Αρχικά αποθηκεύεται σε μια δομή δεδομένων (“ArrayList”) που φέρει το όνομα Θέσεις (places) όλα τα ονόματα των ουσιών που περιέχονται στο Δίκτυο. Επίσης σε μία άλλη δομή δεδομένων (“HashMap”) που φέρει το όνομα Μαρκάρισμα (marking), αποθηκεύεται ως κλειδί (key) το όνομα της κάθε ουσίας και ως τιμή (value) ο αριθμός των μονάδων που αντιστοιχεί στην ουσία. Ακόμη, χρησιμοποιείται η βιβλιοθήκη Pair που περιλαμβάνει αντικείμενα τύπου Pair. Τα αντικείμενα αυτά περιέχουν δύο πεδία. Στο ένα πεδίο τύπου String αποθηκεύεται το προϊόν του κάθε κανόνα και στο άλλο

πεδίο τύπου `ArrayList<String>` αποθηκεύονται όλα τα ονόματα των ουσιών που χρησιμοποιήθηκαν για την πυροδότηση του αντίστοιχου κανόνα. Για παράδειγμα στο Σχήμα 3.1.1 για τον πρώτο κανόνα θα δημιουργηθεί ένα αντικείμενο `Pair` στο οποίο στο πρώτο πεδίο θα περιέχει το προϊόν του κανόνα ‘HMGB1’ και στο άλλο πεδίο θα δημιουργηθεί μια λίστα που θα περιέχει δύο στοιχεία τα ‘HMGB1’ και ‘RAGEu’. Όλα τα αντικείμενα `Pair` που δημιουργούνται αποθηκεύονται σε μία άλλη δομή δεδομένων (“List”) που φέρει το όνομα `list`. Τέλος, χρησιμοποιείται και μια δομή δεδομένων (“ArrayList”) με το όνομα `allRegex` στην οποία αποθηκεύονται τα ονόματα των ουσιών που καταναλώνονται για την πυροδότηση του κάθε κανόνα. Σκοπός αυτής της δομής είναι για να γνωρίζουμε οποιαδήποτε χρονική στιγμή εάν από μια ή περισσότερες βιολογικές ουσίες μπορεί να πυροδοτήσει κάποια από τις αντιδράσεις. Για παράδειγμα στο Σχήμα 3.1.1 από τον πρώτο και δεύτερο κανόνα θα αποθηκευτεί στη λίστα το ‘HMGB1’ και ‘RAGEu’, από τον τρίτο κανόνα το ‘RAGEp’ και συνεχίζει μέχρι να καλύψει όλους τους κανόνες που περιλαμβάνονται στο αρχείο. Περισσότερες πληροφορίες για την περιγραφή της κάθε δομής δεδομένων και των λειτουργιών της που χρησιμοποιήθηκαν μπορούν να βρεθούν στο ιστότοπο του “JAVA API¹⁰”.

Βήματα Αλγόριθμου

1. Αρχικά επιλέγεται τυχαία μια βιολογική ουσία από τη λίστα ουσιών (`places`), η οποία είναι σταθμός αφετηρίας για το χτίσιμο του υπόλοιπου ίχνους. Η συγκεκριμένη βιολογική ουσία αποτελεί την κορυφή¹¹ του ίχνους.
2. Γίνεται έλεγχος εάν η ουσία που επιλέχθηκε ως κορυφή περιλαμβάνεται στη λίστα `allRegex`. Στην περίπτωση που δεν περιλαμβάνεται οδηγεί στο συμπέρασμα ότι ουσία που αποτελεί την κορυφή δεν μπορεί να πυροδοτήσει άλλη αντίδραση ούτε αποτελεί μέρος των προϋποθέσεων για πυροδότηση μιας σύνθετης αντίδρασης (αντίδραση που για να πυροδοτήσει χρειάζεται περισσότερο από μια ουσία) στο μέλλον και αποτελεί το τελευταίο στοιχείο του ίχνους. Το ίχνος που έχει παραχθεί καταγράφεται στο αρχείο εξόδου. Εάν έχει συμπληρωθεί ο αριθμός των ίχνων που έδωσε ο χρήστης ως παράμετρο τότε το πρόγραμμα τερματίζει και επιστρέφει το

¹⁰ <https://docs.oracle.com/javase/7/docs/api/overview-summary.html>

¹¹ Κορυφή= Αποτελεί το τελευταίο στοιχείο που εισήχθη στο ίχνος

αρχείο με όνομα “PreTraces.txt” που περιέχει όλα τα ίχνη που παράχθηκαν, αλλιώς επιστρέφει στο βήμα 1 για δημιουργία νέου ίχνους. Στην περίπτωση που περιλαμβάνεται στη λίστα allRegex αυτό σημαίνει ότι από τη ουσία που αποτελεί την κορυφή του ίχνους μπορεί να πυροδοτηθεί νέα αντίδραση. Ακολουθώντας προχωράμε στο βήμα 3 ώστε να βρεθεί η συγκεκριμένη αντίδραση.

3. Ο αλγόριθμος διαλέγει τυχαία ένα αντικείμενο τύπου Pair από τη list. Το αντικείμενο αντιστοιχεί σε ένα κανόνα. Εάν η ουσία που αποτελεί την κορυφή του ίχνους δεν καλύπτει τα στοιχεία πυροδότησης του κανόνα που επιλέχθηκε επαναλαμβάνεται το βήμα 3. Αλλιώς στην περίπτωση που τα καλύπτει θεωρούμε πως η αντίδραση είναι ενεργή και πυροδοτείται. Αφαιρείται μία μονάδα ξεχωριστά από όλες τις ουσίες που καταναλώθηκαν για την πυροδότηση της συγκεκριμένης αντίδρασης στο μαρκάρισμα τους (“HashMap” που περιέχει το σύνολο των μονάδων κάθε ουσίας του Δικτύου). Τέλος αυξάνουμε στο marking το σύνολο των μονάδων του προϊόντος της συγκεκριμένης αντίδρασης κατά μία. (π.χ. κατά το κτίσιμο του ίχνους και μετά από κάποιο αριθμό εκτελέσεων η κορυφή είναι η βιολογική ουσία ‘PI3Kp’. Ακολουθώντας ψάχνει και βρίσκει ότι από την κορυφή του ίχνους μπορεί να πυροδοτηθεί η αντίδραση ‘d2’. Αυτό έχει ως αποτέλεσμα να προστεθεί στο ίχνος το προϊόν της αντίδρασης ‘d2’ δηλαδή η ‘RAGEu’. Τέλος, θα αφαιρεθεί μια μονάδα από το ‘PI3Kp’ και να προστεθεί μια μονάδα στη ουσία ‘RAGEu’ που είναι το προϊόν της πυροδότησης της ‘d2’).
4. Το προϊόν της αντίδρασης που πυροδοτήθηκε, αποτελεί το τελευταίο στοιχείο που προστέθηκε στο ίχνος και ταυτόχρονα τη νέα κορυφή του ίχνους. Γίνεται έλεγχος εάν το μήκος του ίχνους έχει ξεπεράσει το επιτρεπτό μήκος που πρέπει να έχει κάθε ίχνος που παράγεται, το οποίο δόθηκε ως παράμετρος στο πρόγραμμα από τον χρήστη. Εάν το έχει ξεπεράσει τότε το ίχνος έχει ολοκληρωθεί και το καταγράφουμε στο αρχείο εξόδου. Εάν έχει συμπληρωθεί ο αριθμός των ιχνών που έδωσε ο χρήστης ως παράμετρο τότε το πρόγραμμα τερματίζει και επιστρέφει το αρχείο με το όνομα “PreTraces.txt” που περιέχει όλα τα ίχνη που παράχθηκαν, αλλιώς επιστρέφει στο βήμα 1 για δημιουργία νέου ίχνους. Στην αντίθετη περίπτωση δηλαδή που το μήκος του ίχνους δεν έχει ξεπεράσει το μήκος που πρέπει να έχει κάθε ίχνος συνεχίζουμε με το βήμα 2.

1 1.	PI3Kp, PI3Ku, RAGEp, RAGEp, PI3Kp
2 2.	HMGB1, HMGB1, RAGEp, RAGEp, PI3Kp
3 3.	PI3Kp, PI3Ku, PI3Kp, PI3Ku, RAGEp
4 4.	HMGB1, HMGB1, HMGB1, RAGEp, RAGEp
5 5.	HMGB1, HMGB1, RAGEp, RAGEp, RAGEp
6 6.	PI3Kp, PI3Ku, RAGEp, RAGEu, RAGEp
7 7.	PI3Ku, RAGEp, RAGEp, PI3Kp, PI3Ku
8 8.	PI3Kp, PI3Ku, RAGEp, RAGEp, RAGEu
9 9.	RAGEp, PI3Kp, PI3Ku, RAGEp, PI3Kp
10 10.	RAGEu, HMGB1, HMGB1, HMGB1, HMGB1

Σχήμα 3.1.4: Αποτέλεσμα εκτέλεσης αλγόριθμου παραγωγής ιχνών με βάση τις βιολογικές ουσίες, με αρχεία εισόδου Σχήμα 1 και Σχήμα 2 και παραμέτρους από χρήστη: σύνολο ιχνών : 10, μέγιστο μήκος ίχνους : 5

Ορθότητα Αλγορίθμου

Ο αλγόριθμος τερματίζει

1. Δίνεται ως παράμετρος από τον χρήστη το σύνολο των ιχνών που θα παραχθούν. Όταν συμπληρωθεί ο αριθμός των ιχνών τότε το πρόγραμμα τερματίζει και δίνεται ως έξοδος το αρχείο με τα ίχνη.
2. Κάθε ίχνος ολοκληρώνεται:
 - Όταν από τα προϊόντα της αντίδρασης της κορυφής του ίχνους δεν μπορεί να πυροδοτηθεί οποιαδήποτε άλλη αντίδραση από τους κανόνες που περιλαμβάνονται στο αρχείο εισόδου.
 - Όταν το μήκος του ίχνους φθάσει το μέγιστο μήκος που μπορεί να έχει κάθε ίχνος το οποίο καθορίζεται από τον χρήστη ως παράμετρος στο πρόγραμμα.

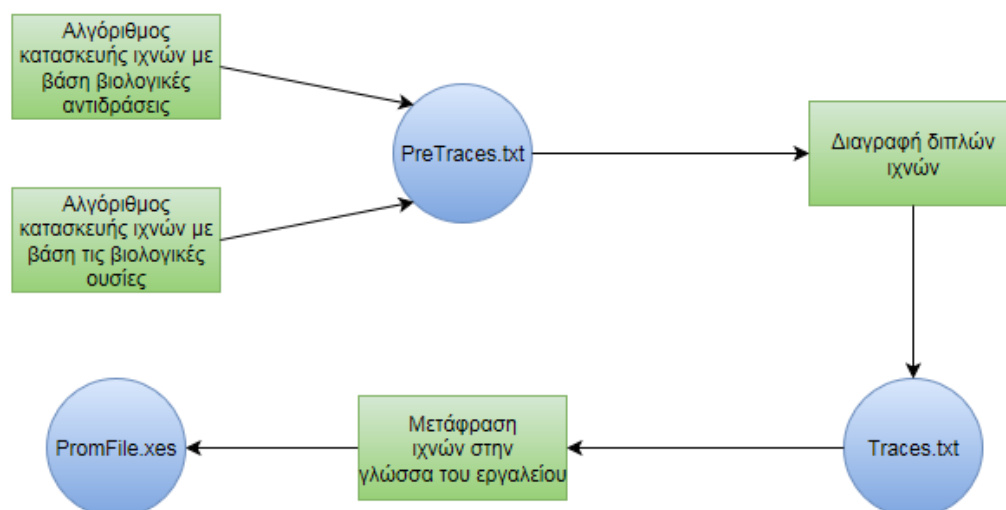
Ο αλγόριθμος βγάζει ορθό αποτέλεσμα:

Για να συμπεριληφθεί μια αντίδραση στο ίχνος υπάρχουν δύο σενάρια.

1. Η βιολογική ουσία επιλέχθηκε τυχαία στην αρχή ως η ουσία που θα χτιστεί επάνω της το υπόλοιπο ίχνος.
2. Η βιολογική ουσία που προστέθηκε στο ίχνος είναι προϊόν κάποιας βιολογικής αντίδρασης όπου ήταν ενεργή την συγκεκριμένη χρονική στιγμή.

3.1.5 Παραγωγή αρχείου εισόδου για το εργαλείο Prom

Οι δύο κλάσεις που υλοποιήθηκαν για τους δύο αλγόριθμους παραγωγής ιχνών έχουν ως αποτέλεσμα το αρχείο PreTraces.txt που περιέχει όλα τα ίχνη που κατασκευάστηκαν. Η διαδικασία μετά την ολοκλήρωση της ενέργειας οποιουδήποτε από τους δύο αλγόριθμους είναι κοινή. Τρέχουμε την κλάση 'CleanTrace' όπου δέχεται ως είσοδο το αρχείο 'PreTraces.txt'. Αυτή η κλάση έχει ως σκοπό να διαγράφει τα διπλά ίχνη. Με την ολοκλήρωση της εργασίας της κλάσης 'CleanTrace' παράγεται το αρχείο 'Traces.txt'. Το συγκεκριμένο αρχείο περιέχει τα τελικά ίχνη που θα χρησιμοποιηθούν για την παραγωγή του μοντέλου (Δίκτυο Πέτρι) που περιγράφει το Βιολογικό μονοπάτι που μελετάται. Τέλος θα τρέξουμε την κλάση 'Script' η οποία λαμβάνει ως είσοδο το αρχείο 'Traces.txt' και παράγει το αρχείο 'PromFile.xes' όπου μετατρέπει τα ίχνη στην μορφή που απαιτείται ([MXML \(Mining eXtensible Markup Language\)](#) και [XES \(eXtensible Event Stream\)](#)). Παράδειγμα της συγκεκριμένης μορφής θα παρουσιαστεί στη συνέχεια της διπλωματικής εργασίας) για να εισαχθούν στο εργαλείο Prom και να βγει το μοντέλο.



Σχήμα 3.1.5: Γραφική αναπαράσταση της διαδικασίας κατασκευής των ιχνών με τους δύο διαφορετικούς αλγόριθμους

3.2 Α-αλγόριθμος

3.2.1 Εισαγωγή

Ο Α-αλγόριθμος (A-algorithm) είναι ένας από τους πρώτους αλγόριθμους Εξόρυξης διαδικασιών, ο οποίος ανακαλύπτει μοντέλα διεργασιών (workflow nets) με είσοδο ένα αρχείο καταγραφής (log event). Ο Α-αλγόριθμος είναι πολύ απλός και πολλές από τις βασικές του ιδέες προκύπτουν από πιο σύνθετους και εύρωστους αλγόριθμους. Ο πρώτος που το έβαλαν σε εφαρμογή ήταν οι Van der Aalst, Weijters και Marouster [18]. Κατά την πάροδο του χρόνου δέχτηκε επεκτάσεις και διαφοροποιήσεις από την αρχική του παρουσίαση. Ο αλγόριθμος χρησιμοποιείται στην παρούσα διπλωματική εργασία ως το βασικό εργαλείο για την τεχνική της ανακάλυψης διαδικασιών.

Ο αλγόριθμος λαμβάνει ως είσοδο ένα αρχείο καταγραφής $L \in E^*$ και επιστρέφει ως αποτέλεσμα ένα κατασκευασμένο Δίκτυο ροής εργασιών (WF-net). Θυμίζουμε ότι το σύνολο E είναι το σύνολο όλων των πιθανών ενεργειών που μπορούν να συμβούν. Στην παρούσα διπλωματική θα αναφερόμαστε σε ενέργεια ως μια βιολογική αντίδραση ή το όνομα μιας βιολογικής ουσίας που μπορεί να παράγεται ή να καταναλώνεται. Αυτές οι

ενέργειες θα αντιστοιχούν στις μεταβάσεις μέσα στο μοντέλο (Δίκτυο Πέτρι) που ανακαλύφτηκε.

Ο Α-αλγόριθμος σαρώνει το αρχείο καταγραφής που του δίνεται ως είσοδο, με την εξαγωγή των παρατηρούμενων σχέσεων μεταξύ των διεργασιών. Για παράδειγμα κάποια βιολογική αντίδραση Α μπορεί να γίνεται πάντα πριν από κάποια άλλη Β, όπου αυτό μπορεί να είναι μια πολύ ωφέλιμη πληροφορία. Για να αναπαραστήσει αυτή την σχέση το μοντέλο θα συνδέσει τις μεταβάσεις Α και Β με μία θέση. Ο αλγόριθμος για να διακρίνει τις σχέσεις μεταξύ των εργασιών του αρχείου καταγραφής χρησιμοποιεί τέσσερις βασικές σχέσεις που θα γίνει αναφορά σε αυτές στη συνέχεια του κεφαλαίου.

3.2.2 Σημασιολογία του Α-αλγόριθμου

Ορισμός των τεσσάρων βασικών σχέσεων του Α-αλγόριθμου.

Έστω L ένα αρχείο καταγραφής όπου $L \in E^*$. Έστω ότι $\alpha, \beta \in E$.

1. $x \succ_L y$ Εάν και μόνο αν υπάρχει ίχνος $\sigma = \{ t_1, t_2, t_3, \dots, t_n \}$ και

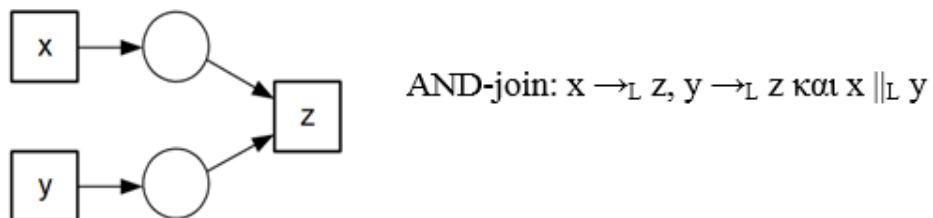
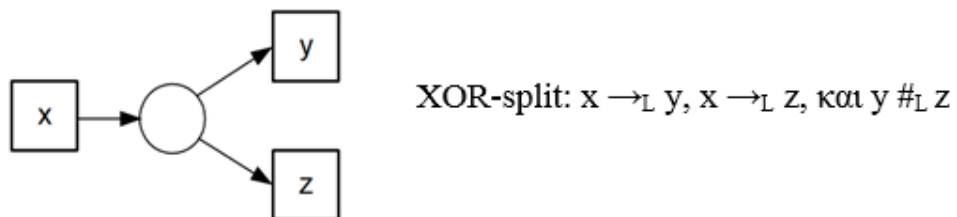
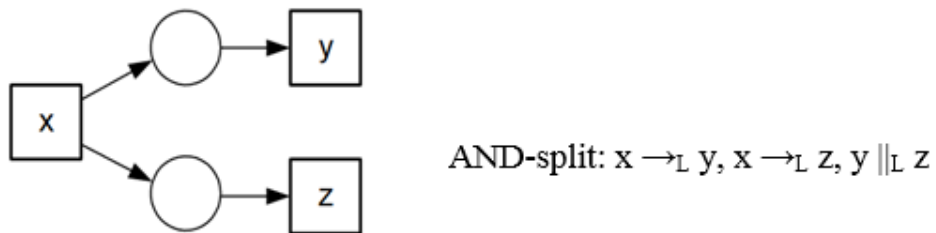
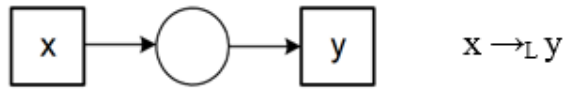
$i \in \{1, 2, \dots, n-1\}$ τέτοιο ώστε $\sigma \in L$ και $t_i = x$ και $t_{i+1} = y$

2. $x \rightarrow_L y$ εάν και μόνο αν $x \succ_L y$ και $y \not\prec_L x$

3. $x \#_L y$ εάν και μόνο αν $x \not\prec_L y$ και $y \not\prec_L x$

4. $x \parallel_L y$ εάν και μόνο αν $x \succ_L y$ και $y \succ_L x$

Πιο κάτω παρουσιάζεται πως μοντελοποιούνται διάφοροι συνδυασμοί μεταξύ των σχέσεων σε Δίκτυο Πέτρι.



Για παράδειγμα έστω το αρχείο καταγραφής $L = [\{a, b, c, d\}^3, \{a, c, b, d\}^2, \{a, e, d\}]$. Από αυτό το αρχείο καταγραφής μπορούν να εξαχθούν οι παρακάτω σχέσεις από τις ενέργειες:

$$\succ_L = \{(a, b), (a, c), (a, e), (b, c), (c, b), (b, d), (c, d), (e, d)\}$$

$$\rightarrow_L = \{(a, b), (a, c), (a, e), (b, d), (c, d), (e, d)\}$$

$$\#_L = \{(a, a), (a, d), (b, b), (b, e), (c, c), (c, e), (d, a), (d, d), (e, b), (e, c), (e, e)\}$$

$$\parallel_L = \{(b, c), (c, b)\}$$

Η σχέση $\succ_L$ περιλαμβάνει όλα τα ζεύγη ενεργειών όπου υπάρχει τουλάχιστον ένα ίχνος στο αρχείο καταγραφής L όπου σε αυτό η πρώτη ενέργεια ενεργοποιείται αμέσως μετά την ενεργοποίηση της δεύτερης ενέργειας. Έχουμε $c \succ_L d$, γιατί η d στο πρώτο ίχνος του L ενεργοποιείται αμέσως μετά τη c . Παρόλα αυτά, $d \not\succ_L c$, γιατί η c σε κανένα ίχνος του L δεν ενεργοποιείται μετά την d .

Η σχέση $\rightarrow_L$ ονομάζεται και σχέση αιτιότητας “causality relation”, και περιλαμβάνει όλα τα ζεύγη ενεργειών όπου η μια ενέργεια σε κάποιο από τα ίχνη του L ενεργοποιείται ακριβώς μετά την άλλη, αλλά τότε δεν γίνεται το αντίστροφο. Έχουμε $c \rightarrow_L d$, γιατί στο πρώτο ίχνος του L το d ενεργοποιείται αμέσως μετά το c , αλλά σε κανένα ίχνος του L το c δεν ενεργοποιείται αμέσως μετά το d ($c \succ_L d$ και $d \not\succ_L c$).

Η σχέση $\#_L$ περιλαμβάνει όλα τα ζεύγη ενεργειών όπου δεν υπάρχει κανένα ίχνος στο αρχείο καταγραφής L , όπου η πρώτη ενέργεια ενεργοποιείται αμέσως μετά την δεύτερη αλλά ούτε και το αντίστροφο. Έχουμε $e \#_L b$ αφού στο αρχείο καταγραφής L σε κανένα ίχνος η b δεν ενεργοποιείται αμέσως μετά τη e αλλά ούτε και το αντίστροφο ($b \not\succ_L e$ και $e \not\succ_L b$).

Η σχέση $\parallel_L$ περιλαμβάνει όλα τα ζεύγη ενεργειών όπου υπάρχει τουλάχιστον ένα ίχνος όπου η πρώτη ενέργεια ενεργοποιείται αμέσως μετά τη δεύτερη, όπως επίσης υπάρχει τουλάχιστον ένα ίχνος όπου η δεύτερη ενέργεια ενεργοποιείται αμέσως μετά την πρώτη. Στο αρχείο καταγραφής L έχουμε $b \parallel_L c$ αφού στο πρώτο ίχνος η c

ενεργοποιείται αμέσως μετά την ενεργοποίηση της b ($b \succ_L c$) ενώ στο δεύτερο ίχνος η b ενεργοποιείται αμέσως μετά την ενεργοποίηση της c ($c \succ_L b$).

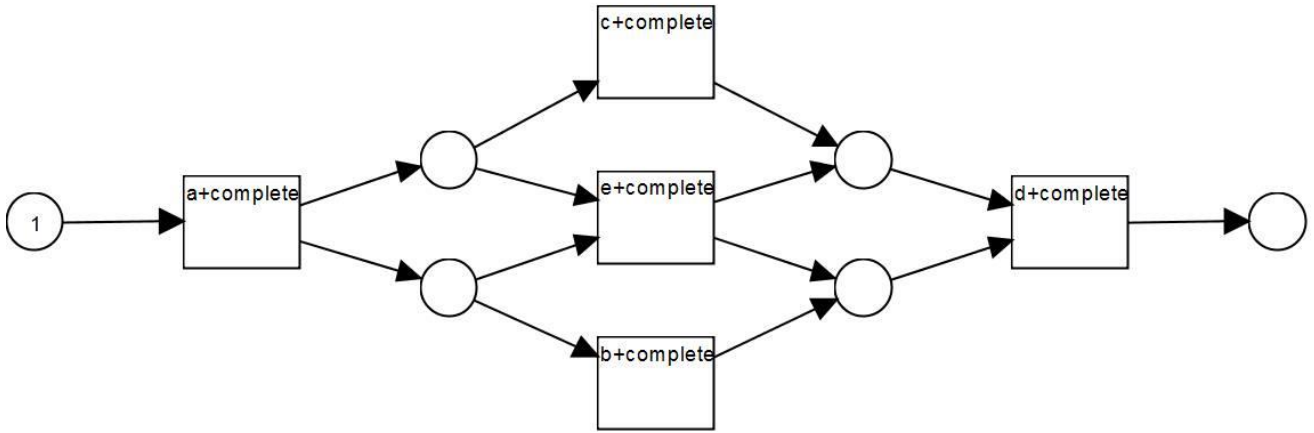
ΘΕΩΡΗΜΑ : Για οποιοδήποτε αρχείο καταγραφής L και a, b ανήκουν στο σύνολο των ενεργειών E , τότε ισχύει ότι : $a \succ_L b$ ή $a \rightarrow_L b$ ή $a \#_L b$ ή $b \parallel_L c$. Δηλαδή για οποιοδήποτε ζευγάρι ενεργειών από ένα αρχείο καταγραφής ισχύει μία από τις τέσσερις σχέσεις που ορίστηκαν πιο πάνω [10].

Για οποιοδήποτε αρχείο καταγραφής μπορεί να δημιουργηθεί ένας πίνακας που ονομάζεται αποτύπωμα που δείχνει όλες τις σχέσεις που υπάρχουν στις ενέργειες μέσα στα ίχνη του γεγονότος. Στο Σχήμα 3.2.1 διακρίνουμε το αποτύπωμα του γεγονότος L που δόθηκε ως παράδειγμα πιο πάνω :

	a	b	c	d	e
a	$\#_L$	$\rightarrow_L$	$\rightarrow_L$	$\#_L$	$\rightarrow_L$
b	$\leftarrow_L$	$\#_L$	$\parallel_L$	$\rightarrow_L$	$\#_L$
c	$\leftarrow_L$	$\parallel_L$	$\#_L$	$\rightarrow_L$	$\#_L$
d	$\#_L$	$\leftarrow_L$	$\leftarrow_L$	$\#_L$	$\leftarrow_L$
e	$\leftarrow_L$	$\#_L$	$\#_L$	$\rightarrow_L$	$\#_L$

Σχήμα 3.2.1 : Παρουσιάζεται ο πίνακας με το αποτύπωμα του αρχείου καταγραφής $L = [\{a, b, c, d\}^3, \{a, c, b, d\}^2, \{a, e, d\}]$.

Στο Σχήμα 3.2.2 παρουσιάζεται το κατασκευασμένο μοντέλο (Δίκτυο Πέτρι) για το γεγονός L όπως το δημιουργεί ο A-αλγόριθμος μέσω του εργαλείου Prom.



Σχήμα 3.2.2 : Το Δίκτυο Πέτρι που δημιουργείται για το αρχείο καταγραφής L , όπως το κατασκευάζει ο Α-αλγόριθμος στο εργαλείο **Prom**.

3.2.3 Βήματα εκτέλεσης Α-αλγόριθμου

Έστω L το αρχείο καταγραφής που αποτελείται από ενέργειες από το σύνολο T ($T \subseteq$ Σύνολο όλων των ενεργειών). Για την κατασκευή του μοντέλου με το Α-αλγόριθμο θα πρέπει να γίνουν με τη σειρά τα πιο κάτω βήματα :

- (1) $T_L = \{t \in T \mid \exists \sigma \in L \ t \in \sigma\}$
- (2) $T_I = \{t \in T \mid \exists \sigma \in L \ t = first(\sigma)\}$
- (3) $T_O = \{t \in T \mid \exists \sigma \in L \ t = last(\sigma)\}$
- (4) $X_L = \{(A, B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall a \in A \forall b \in B \ a \rightarrow_L b \wedge \forall a_1, a_2 \in A \ a_1 \#_L a_2 \wedge \forall b_1, b_2 \in B \ b_1 \#_L b_2\}$
- (5) $Y_L = \{(A, B) \in X_L \mid \forall (A', B') \in X_L \ A \subseteq A' \wedge B \subseteq B' \implies (A, B) = (A', B')\}$
- (6) $P_L = \{p_{(A, B)} \mid (A, B) \in Y_L\} \cup \{i_L, o_L\}$
- (7) $F_L = \{(a, p_{(A, B)}) \mid (A, B) \in Y_L \wedge a \in A\} \cup \{(p_{(A, B)}, b) \mid (A, B) \in Y_L \wedge b \in B\} \cup \{(i_L, t) \mid t \in T_I\} \cup \{(t, o_L) \mid t \in T_O\}$
- (8) $\alpha(L) = (P_L, T_L, F_L)$

Ακολουθεί επεξήγηση των βημάτων του Α-αλγόριθμου :

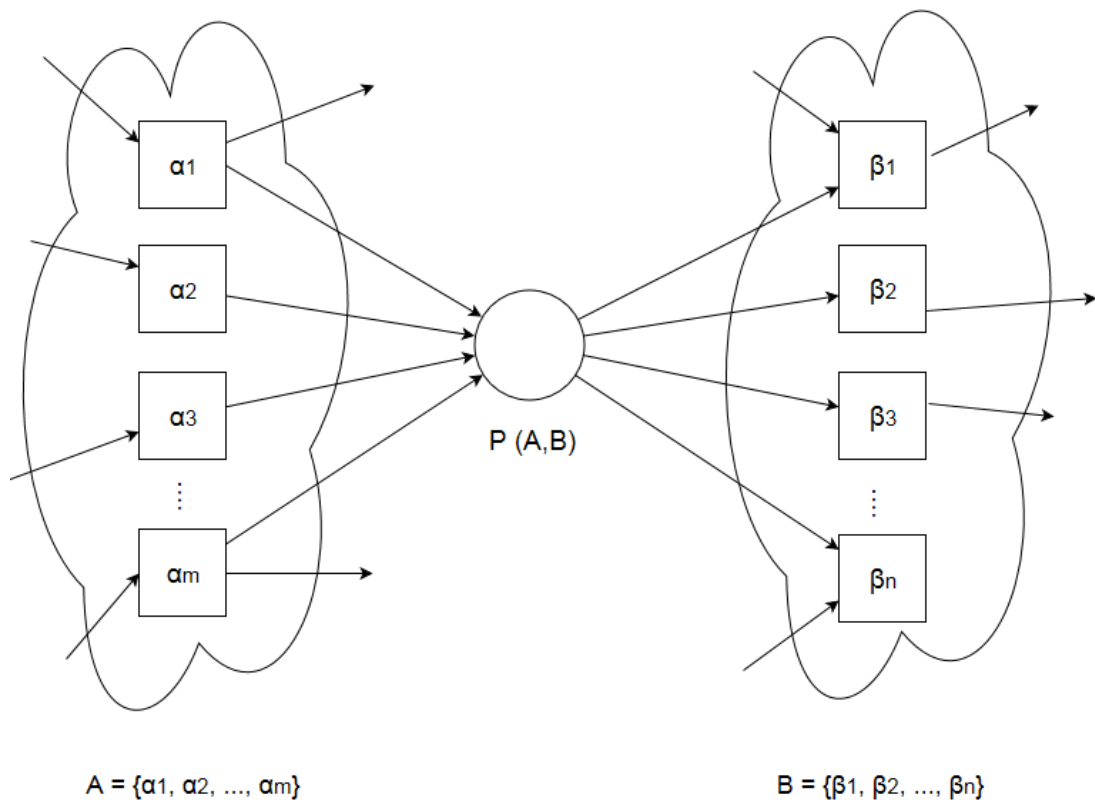
Το L είναι ένα αρχείο καταγραφής πάνω σε ένα σύνολο ενεργειών T .

Στο βήμα 1 υπολογίζεται το σύνολο T_L που περιέχει όλες τις ενέργειες που εκτελούνται τουλάχιστον μια φορά σε κάποιο ίχνος. Αυτό το σύνολο αντιστοιχεί στις μεταβάσεις του παραγόμενου Δικτύου ροής ενεργειών (WF-net).

Στο βήμα 2 υπολογίζεται το σύνολο T_I που περιέχει όλες τις ενέργειες που εκτελούνται στην αρχή κάποιου ίχνους.

Στο βήμα 3 υπολογίζεται το σύνολο T_O που περιέχει όλες τις ενέργειες που εκτελούνται στο τέλος κάποιου ίχνους.

Στα βήματα 4 και 5 έχουμε τα σύνολα X_L και Y_L που αποτελούν το πυρήνα του Α-αλγόριθμου. Η μεγαλύτερη πρόκληση είναι να καθοριστούν οι θέσεις του παραγόμενου Δικτύου ροής και οι συνδέσεις μεταξύ τους. Ο στόχος είναι η κατασκευή του $p(A,B)$ για κάθε θέση που θα περιέχεται το Δίκτυο ροή δεδομένων, τέτοιο ώστε: A είναι το σύνολο όλων των μεταβάσεων εισόδου της θέσης και B το σύνολο των μεταβάσεων εξόδου της θέσης. Ένα παράδειγμα εύρεσης του $p(A,B)$ δίνεται στο Σχήμα 3.2.3.



Σχήμα 3.2.3 : Παράδειγμα εύρεσης του $p(A,B)$ μια θέσης.

Όλα τα στοιχεία του συνόλου A στο πιο πάνω σχήμα έχουν σχέση αιτιότητας με όλα τα στοιχεία του συνόλου B (Για κάθε $\alpha, \beta \in A \times B: \alpha \rightarrow_L \beta$). Επιπλέον ποτέ δύο στοιχεία από το σύνολο A δεν ενεργοποιούνται απευθείας το ένα μετά την ενεργοποίηση του άλλου (Για κάθε $\alpha_1, \alpha_2 \in A: \alpha_1 \#_L \alpha_2$). Παρομοίως ποτέ δύο στοιχεία από το σύνολο B δεν ενεργοποιούνται απευθείας το ένα μετά την ενεργοποίηση του άλλου (Για κάθε $\beta_1, \beta_2 \in B: \beta_1 \#_L \beta_2$).

Ας υποθέσουμε ξανά το αρχείο καταγραφής $L = [\{a, b, c, d\}^3, \{a, c, b, d\}^2, \{a, e, d\}]$. Παρατηρείται ότι $A = \{a\}$, $B = \{b, e\}$ και ανταποκρίνονται στις απαιτήσεις του βήματος 4. Καθώς επίσης για $A' = \{a\}$ και $B' = \{b\}$ ανταποκρίνονται στις ίδιες απαιτήσεις. Το σύνολο X_L είναι το σύνολο με όλα τα ζευγάρια που ανταποκρίνονται στις απαιτήσεις που μόλις αναφέρθηκαν. Στην περίπτωση του L έχουμε:

$$X_L = \{ (\{a\}, \{b\}), (\{a\}, \{c\}), (\{a\}, \{e\}), (\{a\}, \{b, e\}), (\{a\}, \{c, e\}), \\ (\{b\}, \{d\}), (\{c\}, \{d\}), (\{e\}, \{d\}), (\{b, e\}, \{d\}), (\{c, e\}, \{d\}) \}$$

Όμως για να διατηρηθεί ο ελάχιστος αριθμός θέσεων στο κατασκευασμένο μοντέλο ροής ενεργειών και να μην υπάρχουν άσκοπες θέσεις, που η λειτουργία τους καλύπτεται από άλλες θέσεις όπως π.χ. στο X_L η $(\{a\}, \{b, e\})$ η οποία καλύπτει την λειτουργία της $(\{a\}, \{b\})$, χρησιμοποιούνται ζεύγη (A, B) με τα μέγιστα σύνολα στοιχείων για την δημιουργία θέσεων. Για κάθε ζεύγος $(A, B) \in X_L$, και ένα μη κενό σύνολο $A' \subseteq A$ και ένα μη κενό σύνολο $B' \subseteq B$, ισχύει ότι και το $(A', B') \in X_L$. Στο βήμα 5 για να υπολογιστεί το Y_L απλά διαγράφονται όλα τα μη-μέγιστα ζεύγη από το X_L . Για το αρχείο καταγραφής L έχουμε ότι :

$$Y_L = \{ (\{a\}, \{b, e\}), (\{a\}, \{c, e\}), (\{b, e\}, \{d\}), (\{c, e\}, \{d\}) \}$$

Όπως παρατηρούμε στο Σχήμα 2 όπου τρέξαμε τον A-αλγόριθμο στο εργαλείο του Prom το σύνολο του Y_L αποτελεί το σύνολο των θέσεων που παράγονται στο δίκτυο ροής δεδομένων το οποίο περιγράφει το L . Επιπρόσθετα όμως το δίκτυο ροής δεδομένων που παράγεται περιέχει μια μοναδική θέση πηγή i_L και μια μοναδική θέση καταναλωτή o_L όπως ορίζει και ο ορισμός των δικτύων ροής δεδομένων που δόθηκε στο κεφάλαιο 2.3.1. Η εισαγωγή αυτών των δύο θέσεων αποτελεί και το βήμα 6 του

A-αλγόριθμου. Στο βήμα 7, γίνεται η εισαγωγή των τόξων στο Δίκτυο ροής δεδομένων που παράγεται. Όλες οι αρχικές ενέργειες (μεταβάσεις) που ορίσαν το σύνολο T_I στο βήμα 2, έχουν ως θέση εισόδου την i_L και όλες οι ενέργειες (μεταβάσεις) που αποτελούν το σύνολο T_O που ορίστηκε στο βήμα 3 έχουν ως θέση εξόδου την o_L . Για όλες τις θέσεις $p(A,B)$ που ορίστηκαν στο βήμα 5, έχουν ως μεταβάσεις εισόδου τις μεταβάσεις που αποτελούν το σύνολο A και ως μεταβάσεις εξόδου τις μεταβάσεις που αποτελούν το σύνολο B . Με την ολοκλήρωση των 7 βημάτων που περιεγράφηκαν δημιουργείται το Δίκτυο Πέτρι $\alpha(L) = (P_L, T_L, F_L)$ το οποίο περιγράφει τη συμπεριφορά του αρχείου καταγραφής L .

3.2.4 Παράδειγμα εκτέλεσης A-αλγορίθμου

Πιο κάτω ακολουθεί ένα ολοκληρωμένο παράδειγμα με όλα τα βήματα του A-αλγορίθμου.

Έστω ένα αρχείο καταγραφής L_1 το εξής:

$$L_1 = [\{a, b, e, f\}^2, \{a, b, e, c, d, b, f\}^3, \{a, b, c, e, d, b, f\}^2, \{a, b, c, d, e, b, f\}^4, \{a, e, b, c, d, b, f\}^3]$$

Ακολουθούν τα 8 βήματα του A-αλγορίθμου για την κατασκευή του Δικτύου ροής $\alpha(L_1)$ που θα περιγράφει την συμπεριφορά του γεγονότος.

$$\text{Βήμα 1 : } T_{L1} = \{a, b, c, d, e, f\}$$

$$\text{Βήμα 2 : } T_{I1} = \{a\}$$

$$\text{Βήμα 3 : } T_{O1} = \{f\}$$

$$\text{Βήμα 4 : } X_{L1} = \{ (\{a\}, \{b\}), (\{a\}, \{e\}), (\{b\}, \{c\}), (\{b\}, \{f\}), (\{c\}, \{d\}), (\{d\}, \{b\}), (\{e\}, \{f\}), (\{a, d\}, \{b\}), (\{b\}, \{c, f\}) \}$$

$$\text{Βήμα 5 : } Y_{L1} = \{ (\{a\}, \{e\}), (\{c\}, \{d\}), (\{e\}, \{f\}), (\{a, d\}, \{b\}), (\{b\}, \{c, f\}) \}$$

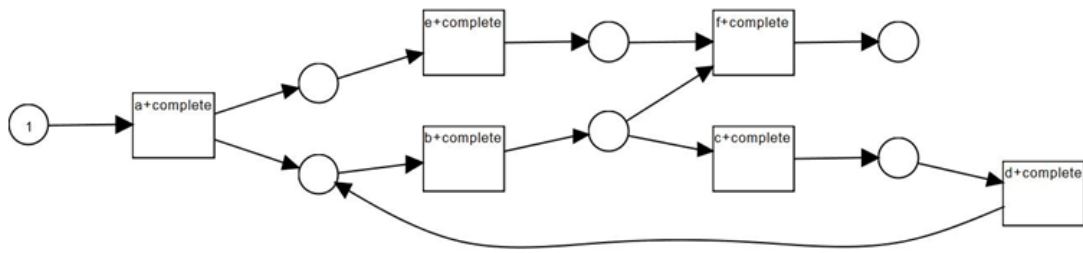
$$\text{Βήμα 6 : } P_{L1} = \{ P_{(\{a\}, \{e\})}, P_{(\{c\}, \{d\})}, P_{(\{e\}, \{f\})}, P_{(\{a, d\}, \{b\})}, P_{(\{b\}, \{c, f\})}, i_{L1}, o_{L1} \}$$

$$\text{Βήμα 7 : } F_{L1} = \{ (a, p_{(\{a\}, \{e\})}), (p_{(\{a\}, \{e\})}, e), (c, p_{(\{c\}, \{d\})}), (p_{(\{c\}, \{d\})}, d), (e, p_{(\{e\}, \{f\})}), (p_{(\{e\}, \{f\})}, f), (a, p_{(\{a, d\}, \{b\})}), (d, p_{(\{a, d\}, \{b\})}),$$

$(p(\{a,d\},\{b\}), b), (b, p(\{b\},\{c,f\})), (p(\{b\},\{c,f\}), c), (p(\{b\},\{c,f\}), f),$
 $(i_{L1}, a), (f, o_{L1})\}$

Αποτέλεσμα $\alpha(L_1) = (P_{L1}, T_{L1}, F_{L1})$

Στο Σχήμα 3.2.4 παρατηρούμε το μοντέλο $\alpha(L_1)$ που δίνει ως έξοδο ο Α-αλγόριθμος του εργαλείου Prom δίνοντας του ως είσοδο το αρχείο καταγραφής L_1



Σχήμα 3.2.4 : Το Δίκτυο Πέτρι που δημιουργείται για το αρχείο καταγραφής L_1 , όπως το κατασκευάζει ο Α-αλγόριθμος στο εργαλείο Prom

3.2.5 Πλεονεκτήματα και μειονεκτήματα του Α-αλγόριθμου

Όπως παρουσιάστηκε, ο Α-αλγόριθμος μπορεί να ανακαλύψει ένα ευρύ σύνολο δικτύων ροής εργασιών, με την προϋπόθεση ότι το αρχείο καταγραφής είναι πλήρες, τηρώντας τις τέσσερις βασικές σχέσεις του Α-αλγόριθμου. Όμως ακόμη και στη περίπτωση που το αρχείο καταγραφής είναι πλήρες, ο Α-αλγόριθμος περιέχει κάποια προβλήματα και περιορισμούς κατά την λειτουργία του.

Πολλά δίκτυα ροής δεδομένων με διαφορετική δομή, περιγράφουν τη συμπεριφορά του ίδιου αρχείου καταγραφής. Για αυτό το λόγο πολλές φορές ο Α-αλγόριθμος μπορεί να επιστρέφει ένα ορθό μοντέλο αλλά όχι το πιο απλό.

ο Α-αλγόριθμος αρκετές φορές έχει προβλήματα με αρχεία καταγραφής όπου τα ίχνη τους περιλαμβάνουν μικρούς κύκλους μήκους ένα και σε κάποιες περιπτώσεις μήκους δύο. Για παράδειγμα εάν ένα ίχνος έχει μια ενέργεια που επαναλαμβάνεται συνεχώς μέσα στο ίχνος τότε το μοντέλο που κατασκευάζεται απομονώνει την μετάβαση που περιγράφει τη συγκεκριμένη ενέργεια αποσυνδέοντας την από το υπόλοιπο μοντέλο.

Έχει γίνει μια θεωρητική προσπάθεια λύσης των προβλημάτων αυτών στο άρθρο [19], όπου έγινε μια προσπάθεια βελτίωσης άλλους βασικής έκδοσης του A-αλγόριθμου για να μπορεί να αντιλαμβάνεται καλύτερα άλλους κύκλους.

Από την άλλη ο A-αλγόριθμος έχει και κάποια δυνατά πλεονεκτήματα που τον κάνουν να ξεχωρίζει από άλλους αλγόριθμους για Ανακάλυψη διαδικασιών. Για παράδειγμα δεν δημιουργεί διπλές μεταβάσεις στο μοντέλο που περιγράφουν την ίδια ενέργεια και σε καμία περίπτωση δεν δημιουργεί μετάβαση για ενέργεια που δεν περιέχεται μέσα στο αρχείο καταγραφής. Επίσης μέσω του βήματος 5 όπου διαγράφονται οι περιττές θέσεις και με τα δύο χαρακτηριστικά που αναφερθήκαν πιο πάνω οδηγούμαστε στο συμπέρασμα ότι το μοντέλο που κατασκευάζεται περιέχει τον ελάχιστο αριθμό κόμβων.

3.3 Εργαλείο Prom

3.3.1 Εισαγωγή

Το Prom είναι το εργαλείο το οποίο χρησιμοποιήθηκε σε αυτή την έρευνα για την κατασκευή των βιολογικών μοντέλων με βάση τη τεχνική της Εξόρυξης Διαδικασιών. Η ομάδα ανάπτυξης του εργαλείου είναι το εργαστήριο «Εξόρυξη Διαδικασιών» του τμήματος Πληροφορικής στο Τεχνολογικό Πανεπιστήμιο του Αϊντχόβεν, στην Ολλανδία.

Το εργαλείο είναι ένα επεκτάσιμο πλαίσιο που υποστηρίζει μια ποικιλία τεχνικών Εξόρυξης Διαδικασιών σε μια επέκταση (plug-in). Η πλατφόρμα είναι υλοποιημένη στη γλώσσα προγραμματισμού Java και μπορεί να εγκατασταθεί δωρεάν σε οποιαδήποτε συσκευή από τον ιστότοπο του εργαλείου [17]. Η πλατφόρμα του Prom τρέχει στα λειτουργικά συστήματα Windows, Mac OS, Linux(Ubuntu).

Το Prom 6 που αποτελεί και την έκδοση που χρησιμοποιήθηκε για την παρούσα μελέτη, είναι κατανοημένο σε κομμάτια ώστε να προσφέρει μέγιστη ευελιξία. Αρχικά ο πυρήνας του εργαλείου κατανέμεται ως ένα πακέτο με δυνατότητα λήψης κάτω από την ανοικτή άδεια λογισμικού GPL (GNU Public Licence¹²). Αυτό σημαίνει πως ο

¹² <http://www.gnu.org/licenses/gpl-3.0.en.html>

χρήστης μπορεί να το κατεβάσει και να το εγκαταστήσει στον υπολογιστή του χωρίς κάποιους περιορισμούς. Ακόμη η επέκταση του Prom 6 είναι κατανεμημένη σε ξεχωριστά πακέτα που χρησιμοποιούνται από την ανοικτή άδεια λογισμικού L-GPL¹³ (Lesser GNU Public License).

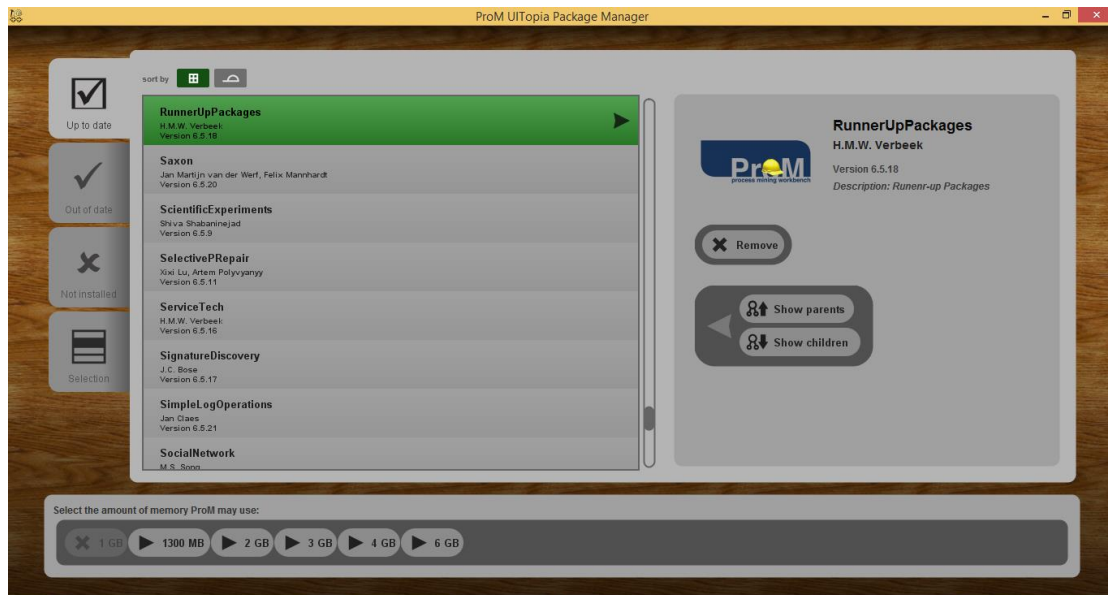
Μετά την εγκατάσταση της επέκτασης του Prom 6 στον υπολογιστή θα δημιουργηθεί αυτόματα ένας φάκελος με όλα τα αρχεία που το αποτελούν. Πληροφορίες για το κάθε αρχείο μπορούν να βρεθούν στο εγχειρίδιο¹⁴ που δίνεται από την ιστοσελίδα. Πιο κάτω αναφέρονται τα τρία αρχεία που είναι τα βασικά για την λειτουργία του εργαλείου.

3.3.2 ProMPM651

Το ProMPM651.exe που είναι ο package manager καλείται για πρώτη φορά τη στιγμή που ο χρήστης ενεργοποιεί για πρώτη φορά την επέκταση του εργαλείου και εγκαθιστά τα απαιτούμενα πακέτα λειτουργίας. Ακολούθως ο χρήστης έχει τη δυνατότητα να ενεργοποιήσει όποτε θελήσει τον package manager και να προσθέσει ή να επεξεργαστεί τα πακέτα που είναι εγκατεστημένα. Γενικά ο package manager είναι υπεύθυνος για την εγκατάσταση των πακέτων που περιέχουν τις διάφορες λειτουργίες που προσφέρει το εργαλείο του Prom. Ο χρήστης μπορεί μέσω της διεπαφής της συγκεκριμένης εφαρμογής να ενημερωθεί για τυχόν ενημερώσεις (updates) που έχουν τα πακέτα, καθώς επίσης για νέα πακέτα και τις λειτουργίες που προσφέρουν. Τέλος ο χρήστης μπορεί να καθορίσει το μέγεθος της μνήμης που δικαιούται να χρησιμοποιεί ο package manager για την εγκατάσταση των πακέτων του.

¹³ <http://www.gnu.org/licenses/lgpl-3.0.en.html>

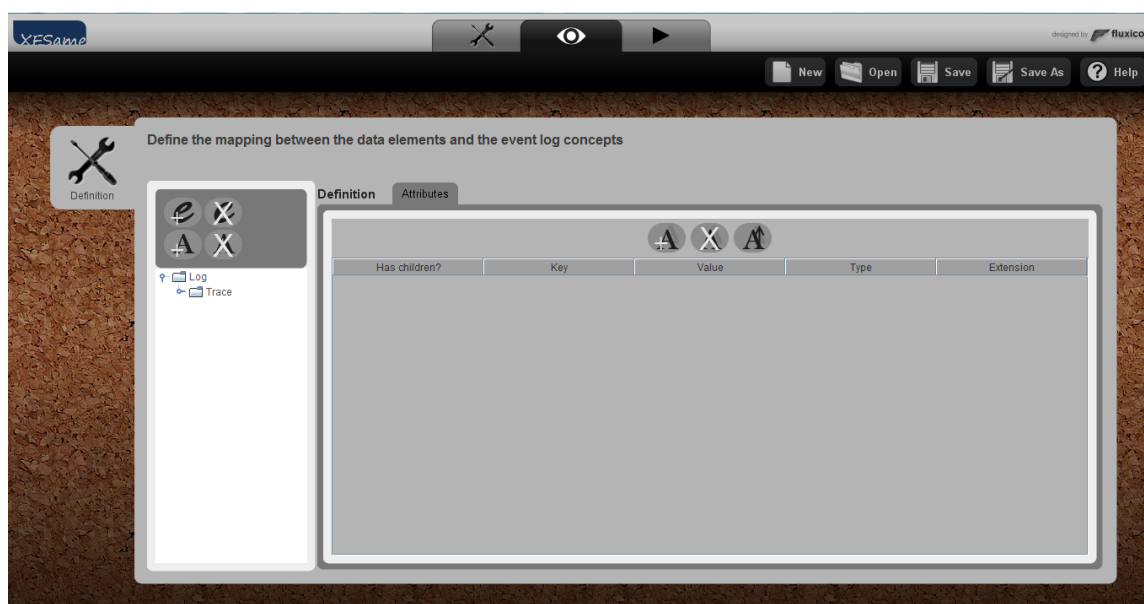
¹⁴ <http://www.promtools.org/doku.php?id=gettingstarted:installation>



Σχήμα 3.3.1: Παρουσίαση του γραφικού περιβάλλοντος του Package Manager

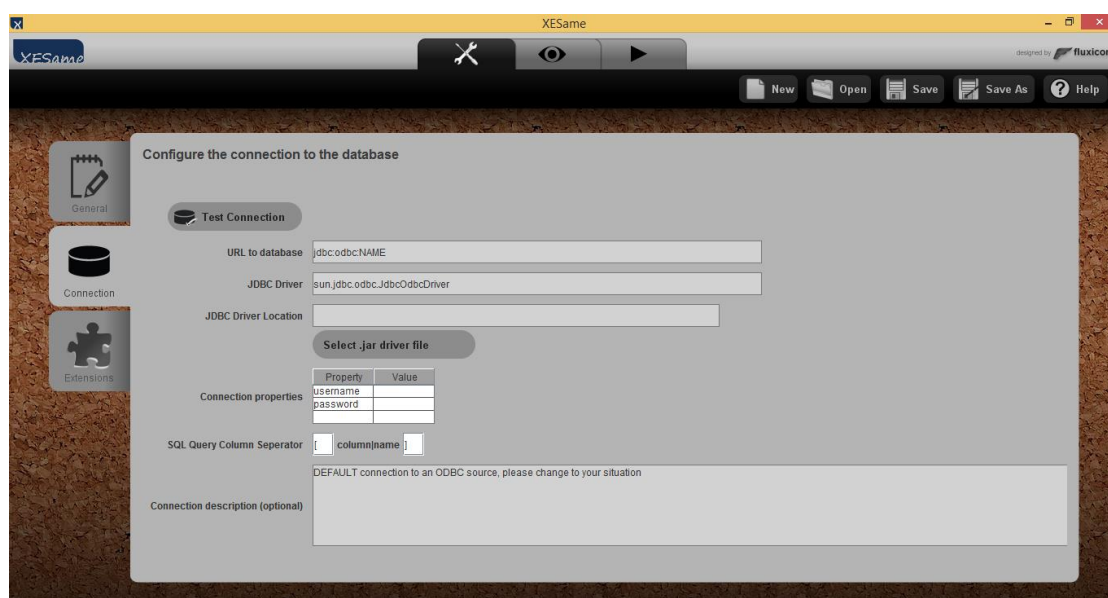
3.3.3 XESame151

Το XESame151.exe προσφέρει ένα γραφικό περιβάλλον όπου ο χρήστης εύκολα και γρήγορα, χωρίς να χρειάζεται ειδικές γνώσεις Πληροφορικής και προγραμματισμού μπορεί να κατασκευάσει οποιαδήποτε ίχνη αρκεί να πληρούν τις ιδιότητες που ορίζει η τεχνική Εξόρυξης διαδικασιών. Καθώς ο χρήστης ολοκληρώσει τα ίχνη του μέσω της διεπαφής, της συγκεκριμένης εφαρμογής, αυτόματα παράγεται ο κώδικας που πληρεί τη δομή που διαβάζει η επέκταση του εργαλείου για παραγωγή του μοντέλου. Συγκεκριμένα η δομή του αρχείου εισόδου που περιγράφει το αρχείο καταγραφής πρέπει να είναι [MXML \(Mining eXtensible Markup Language\)](#) και [XES \(eXtensible Event Stream\)](#).



Σχήμα 3.3.2: Η διεπαφή του XESame151 όπου ο χρήστης εύκολα και γρήγορα δημιουργεί τα ίχνη

Ακόμη ένα ισχυρό πλεονέκτημα του συγκεκριμένου εργαλείου είναι η δυνατότητα που δίνει η συγκεκριμένη εφαρμογή στο χρήστη, όπου ρυθμίζοντας την μπορεί να συνδέσει την πλατφόρμα με κάποια βάση δεδομένων ώστε να επικοινωνούν απευθείας με την ανταλλαγή επερωτήσεων (queries), και χρήσιμων πληροφοριών.



Σχήμα 3.3.3: Διεπαφή XESame151 όπου ο χρήστης μέσω κάποιων ρυθμίσεων, συνδέει την πλατφόρμα με μια βάση δεδομένων

3.3.4 ProM651

Το ProM651.exe είναι η βασική εφαρμογή της πλατφόρμας. Ο χρήστης μπορεί να θέσει σε εφαρμογή όλες τις λειτουργίες της Εξόρυξης διαδικασιών που έχει εγκατεστημένες στο package manager. Η πλατφόρμα του Prom προσφέρει εκατοντάδες λειτουργίες σε διαφορετικά πακέτα. Καθώς είναι μια ανοιχτού κώδικα εφαρμογή και ο οποιοσδήποτε μπορεί να δημιουργήσει τις δικές του λειτουργίες και να τις θέσει δωρεάν για χρησιμοποίηση. Οπότε ανάλογα με την εργασία που έχει να κάνει ο χρήστης μέσω του package manager εγκαθιστά τα κατάλληλα πακέτα.

Για την εκπόνηση της παρούσας μελέτης έχει χρησιμοποιηθεί το πακέτο AlphaMiner και συγκεκριμένα η λειτουργία “Mine for a Petri net using Alpha-algorithm”, που δημιουργήθηκε από τον Boudewijn van Dongen¹⁵, Καθηγητή του Τμήματος Μαθηματικών και Πληροφορικής στο Πανεπιστήμιο του Αϊντχόβεν. Η συγκεκριμένη εφαρμογή λαμβάνει ένα αρχείο εισόδου που περιγράφει το αρχείο καταγραφής και θα πρέπει να πληρεί τις προϋποθέσεις που αναφέρθηκαν πιο πάνω. Ένα παράδειγμα αρχείου εισόδου έχουμε στο Σχήμα 3.3.4, όπου περιγράφεται το σηματοδοτικό μονοπάτι Rb-E2F με ένα μέρος των αντιδράσεων του. Το Σχήμα παρουσιάζει την μορφή που πρέπει να έχει το αρχείο ώστε να μπορεί να το διαβάσει το εργαλείο. Για σκοπούς παρουσίας έχει ζητηθεί από τον αλγόριθμο να δημιουργήσει μόνο δύο ίχνη μήκους δύο.

¹⁵ <https://www.tue.nl/en/university/departments/mathematics-and-computer-science/the-department/staff/detail/ep/e/d/ep-uid/19981149/>

```

1  <?xml version="1.0" encoding="UTF-8" ?>
2  <!-- This file has been generated with the OpenXES library. It conforms -->
3  <!-- to the XML serialization of the XES standard for log storage and -->
4  <!-- management. -->
5  <!-- XES standard version: 1.0 -->
6  <!-- OpenXES library version: 1.0RC7 -->
7  <!-- OpenXES is available from http://www.openxes.org/ -->
8  <log xes:version="1.0" xes:features="nested-attributes" openxes:version="1.0RC7" xmlns="http://www.xes-standard.org/">
9    <extension name="lifecycle" prefix="lifecycle" uri="http://www.xes-standard.org/lifecycle.xesext"/>
10   <extension name="Organizational" prefix="org" uri="http://www.xes-standard.org/org.xesext"/>
11   <extension name="Time" prefix="time" uri="http://www.xes-standard.org/time.xesext"/>
12   <extension name="Concept" prefix="concept" uri="http://www.xes-standard.org/concept.xesext"/>
13   <extension name="Semantic" prefix="semantic" uri="http://www.xes-standard.org/semantic.xesext"/>
14   <global scope="trace">
15     <string key="concept:name" value="__INVALID__"/>
16   </global>
17   <global scope="event">
18     <string key="concept:name" value="__INVALID__"/>
19     <string key="lifecycle:transition" value="complete"/>
20   </global>
21   <classifier name="MXML Legacy Classifier" keys="concept:name lifecycle:transition"/>
22   <classifier name="Event Name" keys="concept:name"/>
23   <classifier name="Resource" keys="org:resource"/>
24   <string key="source" value="Rapid Synthesizer"/>
25   <string key="concept:name" value="Traces.MXML"/>
26   <string key="lifecycle:model" value="standard"/>
27   <trace>
28     <string key="concept:name" value="Case1.0"/>
29   </trace>
30   <event>
31     <string key="org:resource" value="UNDEFINED"/>
32     <date key="time:timestamp" value="2008-12-09T10:10:01.527+01:00"/>
33     <string key="concept:name" value="a13"/>
34     <string key="lifecycle:transition" value="complete"/>
35   </event>
36   <event>
37     <string key="org:resource" value="UNDEFINED"/>
38     <date key="time:timestamp" value="2008-12-09T10:11:01.527+01:00"/>
39     <string key="concept:name" value="b6pp"/>
40     <string key="lifecycle:transition" value="complete"/>
41   </event>
42   </trace>
43   <trace>
44     <string key="concept:name" value="Case2.0"/>
45   </trace>
46   <event>
47     <string key="org:resource" value="UNDEFINED"/>
48     <date key="time:timestamp" value="2008-12-09T10:10:01.527+01:00"/>
49     <string key="concept:name" value="b13"/>
50     <string key="lifecycle:transition" value="complete"/>
51   </event>
52   <event>
53     <string key="org:resource" value="UNDEFINED"/>
54     <date key="time:timestamp" value="2008-12-09T10:11:01.527+01:00"/>
55     <string key="concept:name" value="b13"/>
56     <string key="lifecycle:transition" value="complete"/>
57   </event>
58 </log>

```



Η αρχή και το τέλος του κάθε ίχνους.



Η αρχή και το τέλος ολόκληρου του αρχείου καταγραφής.



Δήλωση βιβλιοθηκών και πληροφοριών που χρησιμοποιεί το αρχείο.



Δήλωση του ονόματος κάθε ίχνους



Αρχή και το τέλος κάθε ενέργειας όπου ανάμεσα δηλώνονται τα χαρακτηριστικά της.

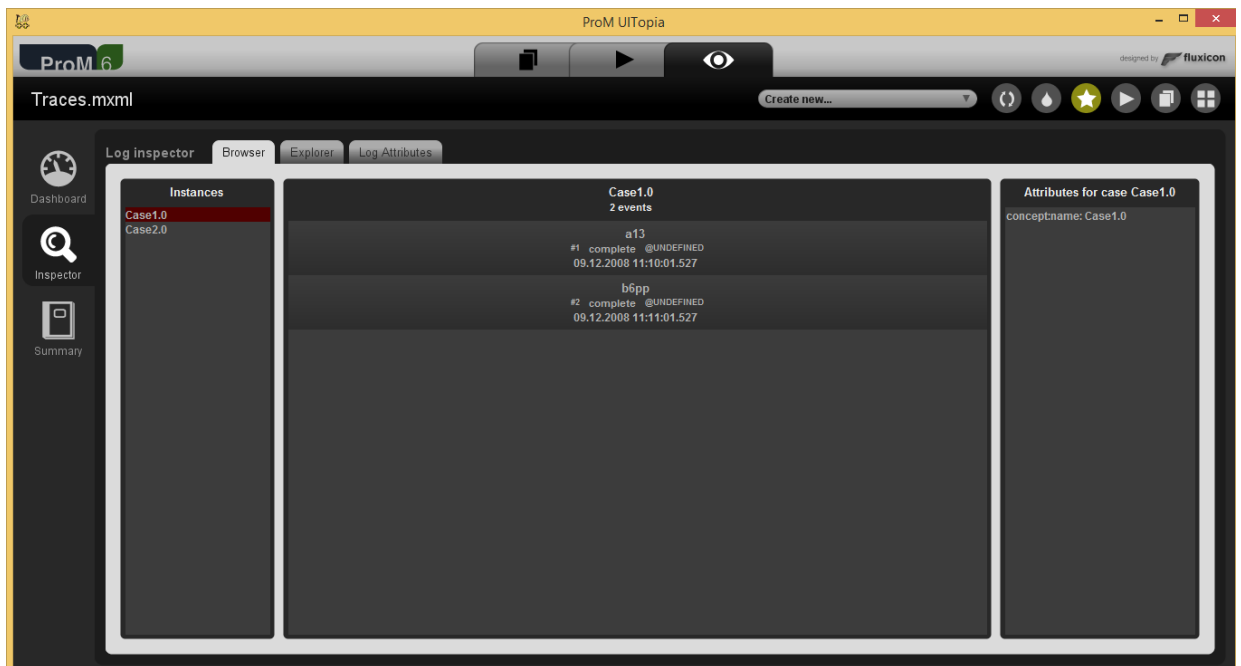


Δήλωση χαρακτηριστικών κάθε ενέργειας. Άτομο που την χειρίζεται, ώρα και ημερομηνία που γίνεται, όνομα ενέργειας, κύκλος ζωής της.

Σχήμα 3.3.4: Παρουσίαση της μορφής της εισόδου που δέχεται το εργαλείο για να κατασκευάσει μοντέλα

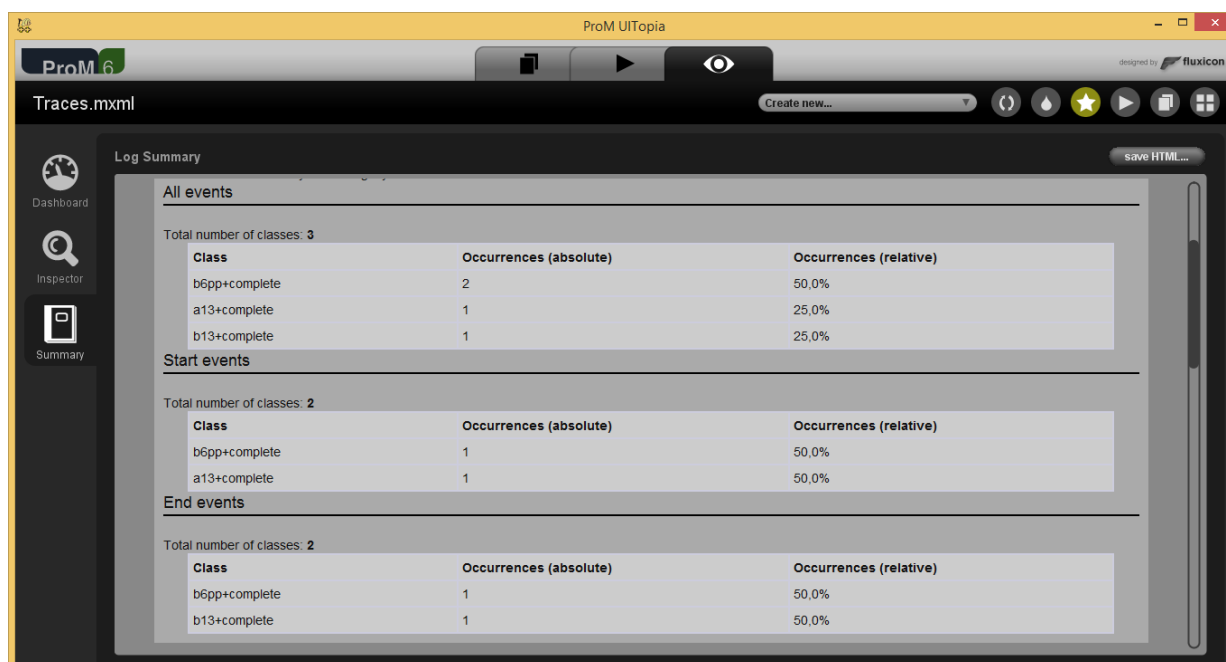
Στο παραπάνω σχήμα παρουσιάζεται η μορφή και η γλώσσα που δέχεται η πλατφόρμα του Prom 6 για να κατασκευάσει το μοντέλο. Παρουσιάζονται δύο ίχνη που κατασκευάστηκαν από τον αλγόριθμο δημιουργίας ιχνών με βάση τις βιολογικές αντιδράσεις. Το πρώτο είναι “a13,b6pp” και το δεύτερο “b6pp,b13”. Το πρώτο ίχνος καταγράφει την πυροδότηση της αντίδρασης a13 και ακολούθως την πυροδότηση της b6pp, ενώ το δεύτερο ίχνος καταγράφει την πυροδότηση της αντίδρασης b6pp, ακολούθως την πυροδότηση της αντίδρασης b13.

Η έκδοση του ProM651.exe παρουσιάζει σε διεπαφή όλα τα ίχνη που διάβασε από το αρχείο εισόδου, καθώς επίσης και τις ενέργειες που τα αποτελούν μαζί με τα χαρακτηριστικά τους.



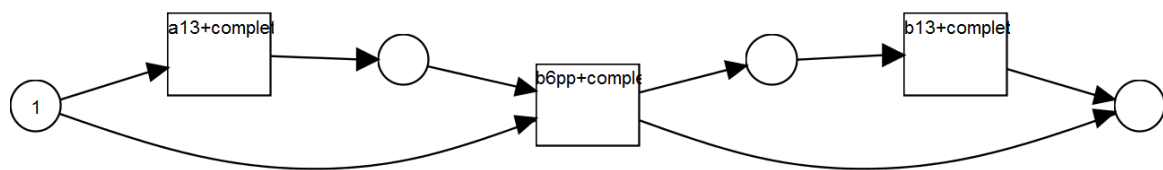
Σχήμα 3.3.5: Το ProM651.exe μέσω διεπαφής παρουσιάζει στο χρήστη όλα τα ίχνη που διάβασε από το αρχείο εισόδου. Δίνεται στο χρήστη η δυνατότητα επιλογής οποιουδήποτε ίχνους επιθυμεί και ακολούθως του παρουσιάζει τις ενέργειες που το αποτελούν με την σειρά εκτέλεσης, μαζί με τα χαρακτηριστικά τους

Ακόμη μέσω άλλης διεπαφής δίνονται στο χρήστη κάποια επιπλέον στοιχεία όσο αφορά το αρχείο καταγραφής. Για παράδειγμα τον συνολικό αριθμό εμφάνισης μιας ενέργειας σε ολόκληρο το αρχείο καταγραφής, καθώς επίσης και τον αριθμό των φορών που εμφανίζεται μια ενέργεια στην έναρξη κάποιου ίχνους καθώς και στη λήξη του ίχνους. Για τα δύο ίχνη στο Σχήμα 3.3.4 παρουσιάζεται η συγκεκριμένη διεπαφή στο Σχήμα 3.3.6.



Σχήμα 3.3.6: Η αντίδραση b6pp εμφανίζεται δύο φορές στα ίχνη και αποτελεί το 50% των εμφανίσεων, ενώ η b13 και a13 από μια φορά και έχουν ποσοστό εμφάνισης 25% αντίστοιχα. Ακόμη τα 2 ίχνη ξεκινούν με b6pp και a13 αντίστοιχα και έχουν 50% ποσοστό εμφάνισης εναρκτήριας αντίδρασης των ιχνών αντίστοιχα. Ενώ από την άλλη η b6pp και b13 είναι οι τελευταίες αντιδράσεις που πυροδοτήθηκαν στα δύο ίχνη και έχουν 50% ποσοστό εμφάνισης ως τελευταία αντίδραση στα ίχνη αντίστοιχα

Στο παρακάτω σχήμα παρουσιάζεται το μοντέλο όπως το κατασκευάζει ο Α-αλγόριθμος. Η έκδοση του ProM651.exe δίνει την δυνατότητα εξαγωγής του αρχείου σε διαφορετικούς τύπους (pdf, png, jpg, jpeg). Επίσης δίνεται η δυνατότητα εξαγωγής και σε.java για περεταίρω επεξεργασία του μοντέλου από πιο εξειδικευμένα εργαλεία ή επεξεργασία μέσω της γλώσσας προγραμματισμού Java για περαιτέρω χρησιμοποίηση.



Σχήμα 3.3.7: Παρουσιάζεται το μοντέλο (Δίκτυο Πέτρι) όπως το κατασκεύασε ο Α-αλγόριθμος για τα δύο ίχνη που δόθηκαν

Τέλος χρησιμοποιήθηκε ακόμη μια λειτουργία του Prom 6 “Show Petri-net Metrics” που δημιουργήθηκε από τον καθηγητή Eric Verbeek¹⁶, ο οποίος εργάζεται στο Πανεπιστήμιο του Αϊντχόβεν στο Τομέα των Μαθηματικών και της Πληροφορικής. Η λειτουργία αυτή προσφέρει κάποιες μετρήσεις στο κατασκευασμένο μοντέλο όπως αριθμό θέσεων, αριθμό μεταβάσεων, αριθμό ακμών, μέτρηση στη πυκνότητα του Δικτύου και περεταίρω πληροφορίες πάνω στη ροή του Δικτύου.

Επίσης η συγκεκριμένη λειτουργία δίνει κάποιες έξτρα πληροφορίες για τις θέσεις που δημιουργούνται στο μοντέλο. Αυτές οι πληροφορίες ήταν πολύ βοηθητικές για την εκπόνηση της παρούσας μελέτης, αφού μεταξύ άλλων μπορούν να παρουσιάσουν σύνθετες και απλές σχέσεις μεταξύ των βιολογικών αντιδράσεων. Η σύνθετη σχέση προσδιορίζει μια αντίδραση η οποία για να πυροδοτήσει προϋποθέτει την πυροδότηση δύο ή περισσότερων διαφορετικών αντιδράσεων. Η απλή σχέση προσδιορίζει μια αντίδραση η οποία για να πυροδοτήσει προϋποθέτει την πυροδότηση ακριβώς μιας διαφορετικής αντίδρασης. Αυτές οι σχέσεις μπορούν να παρουσιαστούν μέσω της λειτουργίας του “Show Petri-net Metrics”.

¹⁶ <https://www.tue.nl/en/university/departments/mathematics-and-computer-science/the-department/staff/detail/ep/e/d/ep-uid/19911662/>

Για παράδειγμα έχουμε τους πιο κάτω κανόνες:

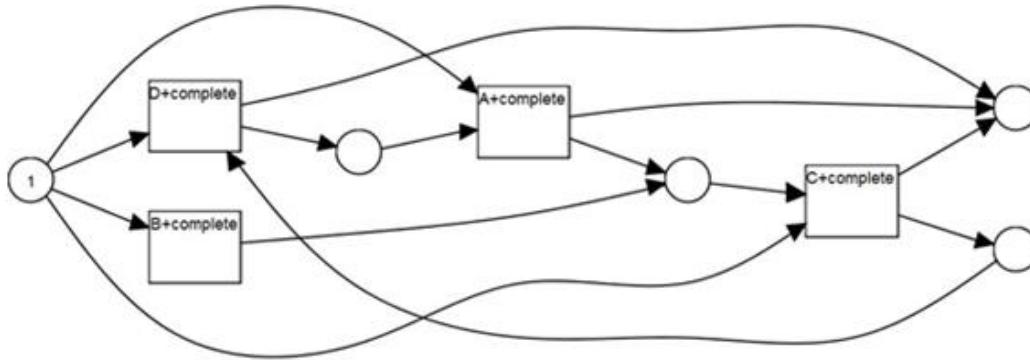
Κανόνας	Ρυθμός Αντίδρασης
1. O6 -> O1	A
2. O4 -> O3	B
3. O1 AND O3 -> O5	C
4. O5 -> O6	D

Υποθέτουμε ότι οι πιο πάνω κανόνες παρουσιάζουν ένα βιολογικό σύστημα που αποτελείται από 4 αντιδράσεις. Η αντίδραση A καταναλώνει την ουσία O6 και παράγει την ουσία O1. Η αντίδραση B καταναλώνει την ουσία O4 και παράγει την ουσία O3. Η αντίδραση C για να πυροδοτηθεί χρειάζεται την κατανάλωση των ουσιών O1 και O3 παράλληλα και παράγει την ουσία O5. Η αντίδραση D καταναλώνει την ουσία O5 και παράγει την ουσία O6.

Ένα σημαντικό χαρακτηριστικό του Βιολογικού κόσμου είναι οι σχέσεις που έχουν μεταξύ τους οι Βιολογικές αντιδράσεις που αποτελούν ένα σύστημα. Πολλές αντιδράσεις για να μπορούν να πυροδοτήσουν εξαρτώνται από τις πυροδοτήσεις άλλων. Μια από τις προκλήσεις που είχαμε να αντιμετωπίσουμε στην παρούσα μελέτη, ήταν πως θα μπορούσε η Εξόρυξη διαδικασιών να παρουσιάζει στα κατασκευασμένα μοντέλα τις εξαρτήσεις μεταξύ τέτοιων Βιολογικών αντιδράσεων. Να μπορεί δηλαδή να ανιχνεύει μέσω των κανόνων αυτές τις εξαρτήσεις και να τις παρουσιάζουν μέσω κάποιας διεπαφής. Ένα παράδειγμα στους κανόνες που δόθηκαν πιο πάνω είναι η περίπτωση της αντίδρασης C, της οποίας για να εκτελεστεί η πυροδότηση θα πρέπει πρώτα να πυροδοτηθεί η αντίδραση A έτσι ώστε να παραχθεί η ουσία O1, όπως επίσης θα πρέπει να πυροδοτήσει η αντίδραση B ώστε να παραχθεί η ουσία O3.

Με την υλοποίηση του αλγορίθμου παραγωγής ιχνών με βάση τις βιολογικές αντιδράσεις και ακολούθως εισάγοντας τα ίχνη στον A-αλγόριθμο φαινόταν από το μοντέλο πως καλύπτονταν τέτοιου είδους σχέσεις. Η λειτουργία του “Show Petri-net metrics”, έρχεται να επαληθεύσει τέτοιου είδους περιπτώσεις. Στο Σχήμα 3.3.8 παρουσιάζεται το κατασκευασμένο μοντέλο που περιγράφει τους πιο πάνω κανόνες,

ενώ το Σχήμα 3.3.9 παρουσιάζει τις σχέσεις μεταξύ των αντιδράσεων, όπως τις βρίσκει η λειτουργία “Show Petri-net metrics”.



Σχήμα 3.3.8 Παρουσιάζεται το μοντέλο που κατασκευάζει ο A-αλγόριθμος στο εργαλείο Prom, για τα ίχνη που δημιουργήθηκαν με το αλγόριθμο δημιουργίας ιχνών με βάση τις Βιολογικές αντιδράσεις πάνω στους κανόνες που δόθηκαν πιο πάνω

```
Place {[C+complete]} --> {[D+complete]} = 1
Place {[D+complete]} --> {[A+complete]} = 1
Place {[A+complete, B+complete]} --> {[C+complete]} = 1
```

Σχήμα 3.3.9: Οι σχέσεις των αντιδράσεων που δίνονται από τη λειτουργία “Show Petri-net metrics”. Η συγκεκριμένη λειτουργία μέσω των ιχνών που δημιούργησαν το μοντέλο, αναγνωρίζει ότι για την εκτέλεση της αντίδρασης C, θα πρέπει να πυροδοτηθεί η αντίδραση A και B. Ενώ για την πυροδότηση των αντιδράσεων D και A αρκεί μόνο η πυροδότηση των C και D αντίστοιχα.

Κεφάλαιο 4

Το Βιολογικό μοντέλο

4.1 Εισαγωγή	78
4.2 Μονοπάτια σηματοδότησης του μοντέλου	80
4.2.1 P53 – MDM2	82
4.2.2 RAS – ERK	83
4.2.3 Rb – E2F	83
4.3 Αξιολόγηση Αποτελεσμάτων	84

4.1 Εισαγωγή

Στο Κεφάλαιο αυτό παρουσιάζεται η εφαρμογή της Εξόρυξης διαδικασιών που είδαμε στο Κεφάλαιο 2.3, πάνω στο βιολογικό μοντέλο που μελετήσαμε στην παρούσα διπλωματική εργασία. Το βιολογικό μοντέλο αποτελείται από ορισμένα μονοπάτια σηματοδότησης που ξεκινούν από την πρωτεΐνη HMGB1 (high-mobility group box 1 protein), όπου μετά από πειράματα φαίνεται να έχει καταλυτική σχέση με την εμφάνιση καρκίνου αφού προκαλεί ενεργοποίηση μονοπατιών που σχετίζονται με τον κυτταρικό κύκλο.

Συγκεκριμένα στην παρούσα μελέτη παρουσιάζονται τρία μονοπάτια: p53 – MDM2, RAS – ERK και Rb – E2F που ενεργοποιούνται από την HBGB1 και πάρθηκαν από το [22]. Στο συγκεκριμένο άρθρο, ένα υπολογιστικό μοντέλο γραμμένο στη γλώσσα BioNetGen προσομοιώνει τη λειτουργία και την αλληλεπίδραση των μονοπατιών. Κατά το συγγραφέα, είναι το πρώτο υπολογιστικό μοντέλο που ασχολείται με την έρευνα της συσχέτισης της συγκεκριμένης πρωτεΐνης με τον καρκίνο. Κατά την εκπόνηση της διπλωματικής εργασίας έγινε εφαρμογή των αλγόριθμων παραγωγής ιχνών στους κανόνες των τριών μονοπατιών που δίνονται στο συγκεκριμένο άρθρο. Ακολούθως, μέσω του εργαλείου Prom και συγκεκριμένα της εκτέλεσης του A-αλγορίθμου και

χρησιμοποιώντας τα ίχνη που πάρθηκαν, δημιουργήθηκαν τα μοντέλα (Δίκτυα Πέτρι). Τα μοντέλα αυτά περιγράφουν όσο το δυνατόν καλύτερα τη συμπεριφορά των μονοπατιών καθώς πάρθηκαν και κάποιες μετρήσεις για τη δομή και τη λειτουργία τους.

Κυτταρικός κύκλος ή κύκλος ζωής του κυττάρου είναι το χρονικό διάστημα που μεσολαβεί από τη στιγμή που το κύτταρο διαιρείται σε δύο θυγατρικά έως τη στιγμή που και αυτά θα διαιρεθούν δίνοντας το καθένα τους απογόνους του. Ο κυτταρικός κύκλος χωρίζεται σε δύο κύριες φάσεις: την μεσόφαση και την μίτωση ή αλλιώς την μιτωτική διαίρεση.

Η μεσόφαση καταλαμβάνει ποσοστό 90% - 95% της διάρκειας ζωής του κυττάρου. Επίσης, η μεσόφαση υποδιαιρείται στα στάδια G1 όπου το κύτταρο έχει μόνο ένα μόριο DNA και υπάρχει χρωματίδα στον πυρήνα. Στο δεύτερο στάδιο S γίνεται αντιγραφή του DNA, δηλαδή διπλασιάζεται η χρωματίδα και γίνεται ένωση των αδελφών χρωματίδων στο Κεντρομερίδιο. Στο τρίτο στάδιο G2 έχουμε διπλοειδές κύτταρο με διπλασιασμένο DNA. Προκύπτει ένα χρωμόσωμα με 2 αδελφές χρωματίδες ενωμένες στο Κεντρομερίδιο.

Η πρωτεΐνη HMGB1 [20] πρωτοεμφανίστηκε πριν από 25 περίπου χρόνια ως ρυθμιστής μεταγραφής μορίων προκαλώντας κάμψη στο DNA και διευκολύνοντας την δέσμευση αρκετών πολύπλοκων μεταγραφών, ιδιαίτερα στους υποδοχείς της ορμονικής οικογένειας στο πυρήνα. Αν και αόριστα είναι μια από τις από τις πιο σημαντικές πρωτεΐνες χρωματίνης, απελευθερώνεται από νεκρά κύτταρα και εκκρίνεται από ενεργούς μακροφάγους σε μια διαδικασία νέκρωσης των κυττάρων, ως ανταπόκριση σε τραυματισμούς, μολύνσεις και άλλα μολυσματικά γεγονότα [21]. Εντός του κυττάρου είναι υπεύθυνη για τη σταθεροποίηση των νουκλεοσωμάτων, για τη ρύθμιση και τη μεταγραφή πολλών γονιδίων και τη αλληλεπίδραση με ιστόνες¹⁷.

Επιπλέον αλληλεπιδρά με την RAGE στα ενδοθηλιακά κύτταρα προκαλώντας την ενεργοποίηση και στρατολόγηση των λευκοκυττάρων. Η RAGE αποδείχτηκε πρόσφατα

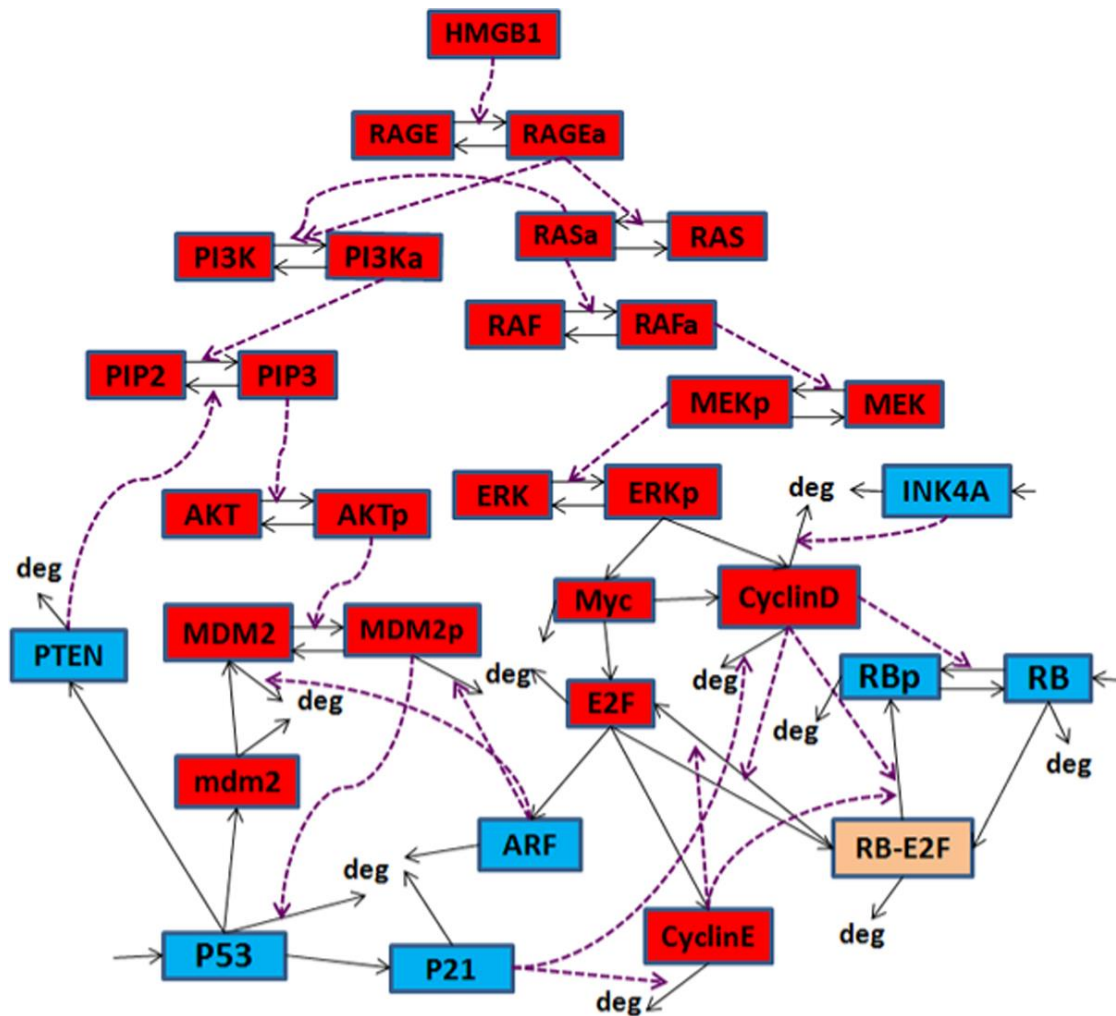
¹⁷ Ιστόνες = Υψηλές αλκαλικές πρωτεΐνες που βρίσκονται στον πυρήνα των ευκαριωτικών κυττάρων. Είναι υπεύθυνοι για α κατασκευάζουν το DNA σε δομικές μονάδες που λέγονται νουκλεοσώματα.

πως λειτουργεί ως αντί-υποδοχέας για λευκοκυτταρικές ιντεγγρίνες, υποδηλώνοντας ότι η σηματοδότηση μέσω αυτών των μορίων είναι δυνητικά σημαντική για την προσκόλληση και κατηγοριοποίηση των κυττάρων, καθώς και στην πρόσληψη των φλεγμονωδών κυττάρων. Μετά από μελέτες αποδείχτηκε πως η στοχευόμενη αποδυνάμωση της HMGB1 και του υποδοχέα της RAGE επάγει την απόπτωση και υποστέλλει την ανάπτυξη του καρκίνου στο πάγκρεας.

4.2 Μονοπάτια σηματοδότησης του μοντέλου

Η μεταγωγή σήματος προκύπτει όταν ένα εξωκυττάριο μόριο σηματοδότησης ενεργοποιεί ένα συγκεκριμένο υποδοχέα που βρίσκεται στην επιφάνεια ή στο εσωτερικό των κυττάρων. Με τη σειρά του ο υποδοχέας ενεργοποιεί μια βιοχημική αλυσίδα γεγονότων στο εσωτερικό του κυττάρου δημιουργώντας μια κατάσταση που τροποποιεί το μεταβολισμό του κυττάρου σχηματικά, τη γονιδιακή έκφραση του και διαίρεση του.

Στην παρούσα μελέτη, εξετάζονται τρία μονοπάτια μεταγωγής σήματος που ενεργοποιούνται από την HMGB1. Ολόκληρο το μοντέλο της πρωτεΐνης HMGB1 απεικονίζεται στο Σχήμα 4.2.1 όπου περιλαμβάνει 31 είδη μορίων (6 ογκοκατασταλτικές πρωτεΐνες), 59 χημικές αντιδράσεις και 3 διαφορετικά σηματοδοτικά μονοπάτια που ενεργοποιούνται από την HMGB1. Τα μονοπάτια σηματοδότησης είναι τα εξής : Το RAS-ERK , το Rb-E2F και το p53-MDM2. Λόγω του ότι η αλληλεπίδραση μεταξύ της πρωτεΐνης HMGB1 και των υποδοχών TLR και RAGE δεν έχει ακόμη ξεκαθαρίσει σε καθαρό επίπεδο, για σκοπούς ευκολίας η RAGE θα αντιπροσωπεύει όλους τους υποδοχείς στο μοντέλο ώστε να μειωθούν οι άγνωστες παράμετροι.



Σχήμα 4.2.1 : Η εικόνα πάρθηκε από το [22] και απεικονίζει τα σηματοδοτικά μονοπάτια που ξεκινούν από την πρωτεΐνη HMGB1. Οι μπλέ κόμβοι απεικονίζουν ογκοκατασταλτικές πρωτεΐνες, οι κόκκινοι κόμβοι ογκοπρωτεΐνες ή λιπίδια ενώ με καφέ χρώμα είναι η πολύπλοκη σύνδεση της ογκοκατασταλτικής πρωτεΐνης RB με την ογκοπρωτεΐνη E2F. Οι συμπαγείς ακμές υποδηλώνουν μεταγραφή πρωτεΐνης, ενώ οι διακεκομμένες ακμές καθορίζουν την ενεργοποίηση μιας διαδικασίας

Ακολουθούν τα τρία μονοπάτια σηματοδότησης που ξεκινούν από την HMGB1. Καθορίζουμε ενεργοποίηση ή προαγωγή το σύμβολο $\rightarrow$, καθώς και ως αναστολή ή αναστολή το σύμβολο $\dashv$.

4.2.1 P53 – MDM2

Η ενεργοποίηση της πρωτεΐνης P53 προστατεύει τον οργανισμό από τη μετάδοση κυττάρων που προκαλούν την δημιουργία όγκων και την καταστροφή του DNA. Η θεμελιώδης ανταγωνίστρια της σε αυτή την προσπάθεια είναι η πρωτεΐνη MDM2. Η συγκεκριμένη πρωτεΐνη ενεργοποιείται για να περιορίζει την ανάπτυξη της P53 στα κύτταρα. Το μονοπάτι P53-MDM2 ρυθμίζεται από δύο επαναληπτικές αναδράσεις. Με τον όρο επαναληπτική ανάδραση (feedback loop) ορίζουμε ένα ρυθμιστικό Βιολογικό κύκλωμα το οποίο επιβραδύνει μια βιολογική διαδικασία εάν η ανάδραση είναι αρνητική, ενώ σε αντίθετη περίπτωση όταν είναι θετική τείνει να την επιταχύνει.

Όσον αφορά την πρώτη επαναληπτική ανάδραση είναι αρνητική και ορίζεται ως :

$PI3K \rightarrow PIP3 \rightarrow AKT \rightarrow MDM2 \dashv p53 \rightarrow MDM2$.

Η πρωτεΐνη PI3K ενεργοποιείται από τον υποδοχέα της TLR2/4 αφού πρώτα περάσουν λίγα λεπτά από την ενεργοποίηση της TLR2/4 από την HMGB1. Με την σειρά της η PI3K φωσφορυλιώνει¹⁸ τη PIP2 σε PIP3 που οδηγεί στην φωσφορυλίωση της AKT. Η μη φωσφορυλίωση της ογκοπρωτεΐνης MDM2, η οποία επηρεάζει την ανάπτυξη της P53, κατοικεί στο κυτταρόπλασμα. Όμως με την φωσφορυλίωση της AKT, η MDM2 εισέρχεται στο πυρήνα αναστέλλοντας την δραστηριότητα της P53.

Η δεύτερη επαναληπτική ανάδραση είναι θετική και ορίζεται ως:

$p53 \rightarrow PTEN \dashv PIP3 \rightarrow AKT \rightarrow MDM2 \dashv p53 \rightarrow \text{Απόπτωση}$

Η πρωτεΐνη P53 μπορεί να ρυθμίσει τη μεταγραφή της ογκοκατασταλτικής πρωτεΐνης PTEN η οποία με τη σειρά της προκαλεί την υδρόλυση της PIP3 σε PIP2. Με αυτό τον τρόπο συστέλλεται η ενεργοποίηση της AKT και της MDM2, οι οποίες αντιτάσσονται στην ευεργετική αντικαρκινική δράση της p53.

Στο άρθρο [23] γίνεται λεπτομερής περιγραφή όλου του πλαισίου γύρω από αυτή την αντίδραση.

¹⁸ Φωσφορυλίωση : Είναι μια χημική αντίδραση στην οποία μία ή περισσότερες φωσφορικές ομάδες προστίθενται σε ένα μόριο

4.2.2 RAS – ERK

Το μονοπάτι σηματοδότησης ορίζεται ως:



Όταν ο υποδοχέας RAGE ενεργοποιηθεί από την HMGB1 οδηγεί στην ενεργοποίηση της RAS. Η RAS ακολούθως θα ενεργοποιήσει την πρωτεΐνη RAF. Ενεργοποιώντας την RAF, αυτή με τη σειρά της θα φωσφορυλιώσει τις MEK πρωτεΐνες οδηγώντας στο τέλος στη φωσφορυλίωση της ERK. Με την ενεργοποίηση της ERK μπορεί να φωσφορυλιωθούν κάποιοι παράγοντες μεταγραφών που ενεργοποιούν την ρύθμιση των πρωτεϊνών CyclinD και Myc, δίνοντας την δυνατότητα εξέλιξης της διαδικασίας του κυτταρικού κύκλου μέσω της φάσης G1. Στο άρθρο [26] αναφέρεται ότι η K-RAS, μέλος της οικογένειας RAS, έχει βρεθεί μεταλλαγμένη σε περισσότερο από 90% των καρκίνων του πάγκρεας.

4.2.3 Rb – E2F

Το μονοπάτι σηματοδότησης ορίζεται ως : Cyclin D + Rb + E2F $\rightarrow$ Cyclin E + Rb. Το συγκεκριμένο μονοπάτι ρυθμίζει τη μετάδοση από G1 φάση μέχρι τη S στο κυτταρικό κύκλο κατά την διάρκεια πολλαπλασιασμού των κυττάρων. Η E2F είναι ένας παράγοντας μεταγραφής πολλών πρωτεϊνών που συμμετέχουν στην αντιγραφή του DNA και την εξέλιξη του κυτταρικού κύκλου. Σε αντίθεση η Rb, είναι μια ογκοκατασταλτική πρωτεΐνη που όταν έρθει σε επαφή με την E2F περιορίζει τη δράση της. Όμως η E2F μπορεί να ενεργοποιηθεί και να απελευθερωθεί από το πλέγμα που την συνδέει με την Rb, εάν η τελευταία φωσφορυλιωθεί από κάποιες ογκοπρωτεΐνες όπως την CyclinD και την Myc. Με αποτέλεσμα να δίνεται πλέον η δυνατότητα στην απελευθερωμένη E2F, να μεταδώσει τη CyclinE και CDK2 (Cyclin-dependent protein kinase 2) που προωθούν την εξέλιξη του κυτταρικού κύκλου. Ακολούθως η CyclinE με την σειρά της συνεχίζει να αναστέλλει την ιδιότητα της Rb, οδηγώντας το μονοπάτι σε μια θετική επαναληπτική ανάδραση. Από το Σχήμα 4.2.1 προκύπτει ότι η ιδιότητα της CyclinD-CDK4/6 περιορίζεται από την ογκοκατασταλτική πρωτεΐνη INK4A , και σύμφωνα με το [25] είναι απενεργοποιημένη μέχρι και 90% στους καρκίνους του παγκρέατος.

Όλα όσα λέχθηκαν μεταξύ αυτών των μονοπατιών μπορούν να επηρεάσουν τη μοίρα του κυττάρου με την προϋπόθεση ότι τα τρία μονοπάτια σηματοδότησης δεν είναι ανεξάρτητα μεταξύ τους. Όπως φαίνεται στο Σχήμα 4.2.1, η ογκοπρωτεΐνη RAS μπορεί να ενεργοποιήσει το μονοπάτι σηματοδότησης PI3K-AKT. Η ογκοκατασταλτική πρωτεΐνη ARF, ενεργοποιημένη από την E2F μπορεί να ενωθεί με την MDM2 για να γίνει προώθηση της ταχείας υποβάθμισης και σταθεροποίηση της P53. Επιπλέον έχει παρατηρηθεί σε πειράματα [24] ότι η p21 και η FBXW7 καταστέλλουν τη δραστηριότητα των κινασών ¹⁹Cyclin.

Στο μοντέλο του HMGB1 κάποιες από τις οντότητες που έχουν ως κατάληξη “a” ή “p” αντιστοιχεί στο ότι η ουσία είναι ενεργή (“active”) ή φωσφορυλιωμένη (“phosphorylated”). Για παράδειγμα :

- RAGEa είναι ενεργή (“active”) ενώ RAGE είναι ανενεργή
- ERKp είναι φωσφορυλιωμένη (“phosphorylated”) ενώ ERK είναι μη φωσφορυλιωμένη.

4.3 Αξιολόγηση Αποτελεσμάτων

Η κατασκευή των μοντέλων βασίστηκε στα μονοπάτια που περιγράφηκαν στο Κεφάλαιο 4.2 και ιδιαίτερα στον κώδικα BioNetGen του άρθρου [22] ο οποίος δίδει όλες τις λεπτομέρειες του μοντέλου. Για την υλοποίηση των ιχνών χρησιμοποιήθηκαν οι αλγόριθμοι κατασκευής ιχνών (Κεφάλαιο 3.1). Για αρχεία εισόδου στους αλγορίθμους δόθηκαν οι κανόνες του κάθε μονοπατιού. Στη συνέχεια του κεφαλαίου παρουσιάζονται εκτενή πληροφορίες για το κάθε μονοπάτι.

¹⁹ <https://en.wikipedia.org/wiki/Kinase>

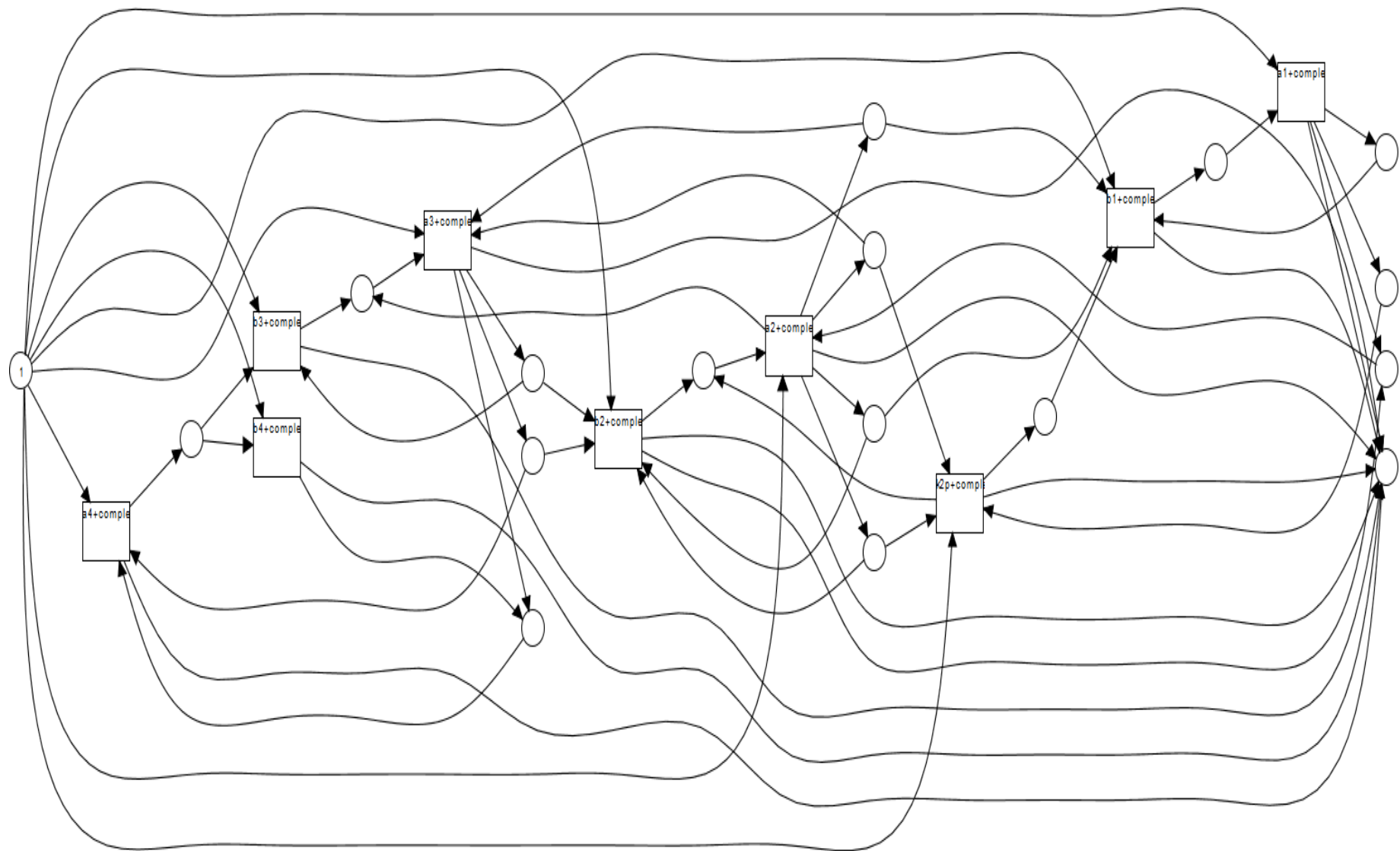
α) ERK-RAS Σηματοδοτικό Μονοπάτι

Οι κανόνες της BioNetGen για αυτό το μονοπάτι έχουν ως εξής :

Κανόνας	Ρυθμός Αντίδρασης
1. $RAGE(a\sim p) + RAS(f\sim U) \rightarrow RAGE(a\sim p) + RAS(f\sim p)$	a1
2. $RAS(f\sim p) + PI3K(b\sim U) \rightarrow RAS(f\sim p) + PI3K(b\sim p)$	k2p
3. $RAS(f\sim p) \rightarrow RAS(f\sim U)$	b1
4. $RAS(f\sim p) + RAF(g\sim U) \rightarrow RAS(f\sim p) + RAF(g\sim p)$	a2
5. $RAF(g\sim p) \rightarrow RAF(g\sim U)$	b2
6. $RAF(g\sim p) + MEK(h\sim U) \rightarrow MEK(h\sim p) + RAF(g\sim p)$	a3
7. $MEK(h\sim p) \rightarrow MEK(h\sim U)$	b3
8. $ERK(i\sim U) + MEK(h\sim p) \rightarrow MEK(h\sim p) + ERK(i\sim p)$	a4
9. $ERK(i\sim p) \rightarrow ERK(i\sim U)$	b4

Πίνακας 4.2.1 : Κανόνες για το ERK –RAS μονοπάτι

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις αντιδράσεις κατασκεύασε 558 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.2. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.193. Συνολικά δημιουργήθηκαν 17 θέσεις και 9 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές αντιδράσεις του σηματοδοτικού μονοπατιού ERK-RAS. Επίσης το μοντέλο διαθέτει 59 ακμές.



Σχήμα 4.2.2: ERK-RAS Σηματοδοτικό Μονοπάτι αντιδράσεων

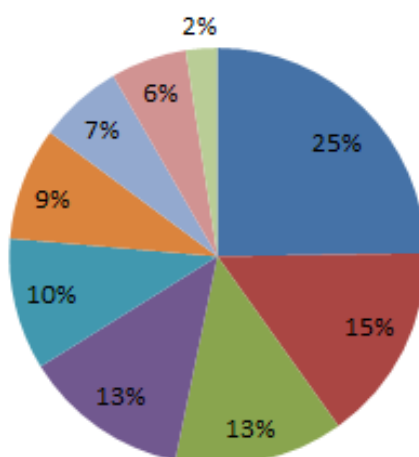
Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε αντίδραση στο αρχείο καταγραφής, όπου περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις αντιδράσεις (Κεφάλαιο 3). Ακολούθως παρουσιάζεται η γραφική παράσταση του αντίστοιχου πίνακα.

Όνομα αντίδρασης	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
a2	528	24.754%
k2p	328	15.377%
a1	280	13.127%
a3	275	12.893%
b1	218	10.22%
b2	187	8.767%
a4	139	6.517%
b3	126	5.907%
b4	52	2.438%

Πίνακας 4.2.2: Εμφανίσεις αντιδράσεων σηματοδοτικού μονοπατιού ERK-RAS στα ίχνη.

ERK-RAS Σηματοδοτικό μονοπάτι

■ a2 ■ k2p ■ a1 ■ a3 ■ b1 ■ b2 ■ a4 ■ b3 ■ b4



Σχήμα 4.2.3: Γραφική παράσταση της εμφάνισης των αντιδράσεων σηματοδοτικού μονοπατιού ERK-RAS στα ίχνη.

Μέσω της κατασκευής του μοντέλου για το σηματοδοτικό μονοπάτι ERK-RAS, παρατίθενται στο πιο κάτω πίνακα η επόμενη πιθανή εκτελούσα αντίδραση για κάθε βιολογική αντίδραση που ανήκει στο μονοπάτι.

Αντίδραση	Πιθανή επόμενη εκτελούσα αντίδραση
a1	a1,b1,a2,k2p
a2	a2,b2,a3,b1,k2p
a3	b3,a4,b2,a3
a4	a4,b3,b4
b1	a1
b2	a2
b3	a3
b4	a4
k2p	b1,a2,k2p

Πίνακας 4.2.3: Επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο σηματοδοτικό μονοπάτι ERK-RAS.

Στο Σχήμα 4.2.4 παρατίθενται οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων όπως τις δίνει η λειτουργία “Show Petri-net Metrics” για το σηματοδοτικό μονοπάτι ERK-RAS.

```

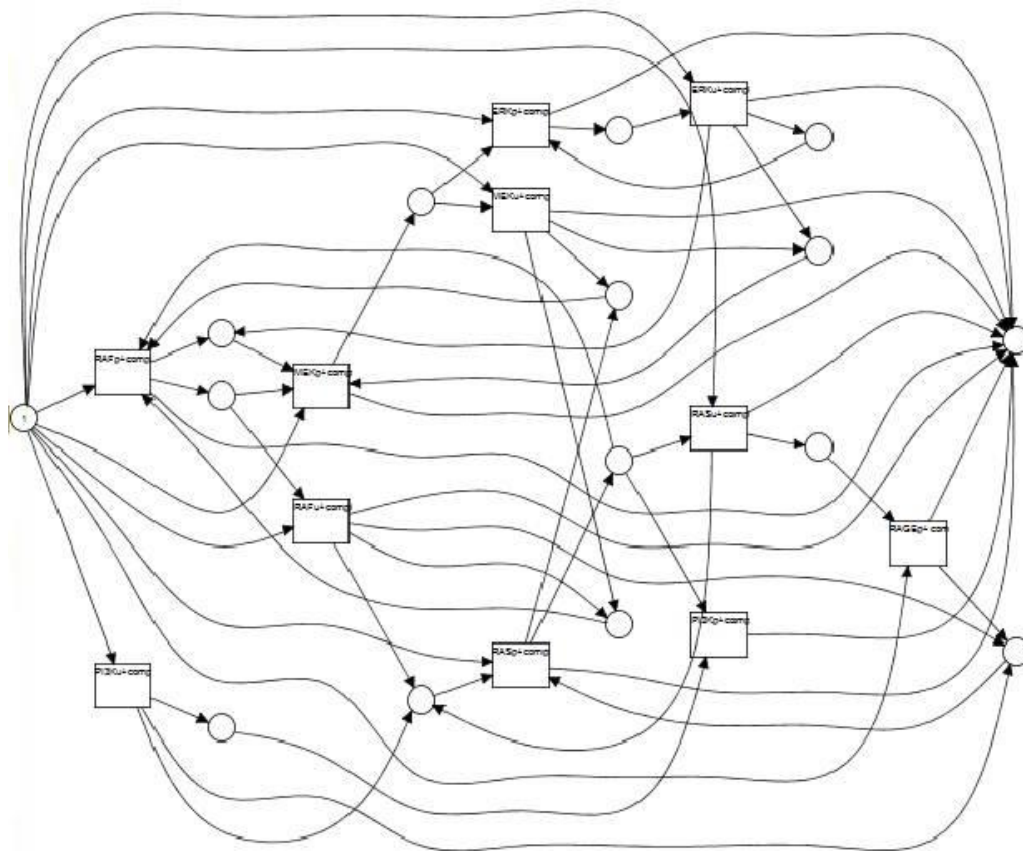
Place {[b1+complete]} --> {[a1+complete]} = 1
Place {[a1+complete]} --> {[b1+complete]} = 1
Place {[k2p+complete, b2+complete]} --> {[a2+complete]} = 1
Place {[a1+complete]} --> {[k2p+complete]} = 1
Place {[a2+complete]} --> {[a3+complete, b1+complete]} = 2
Place {[a2+complete]} --> {[b1+complete, b2+complete]} = 2
Place {[k2p+complete]} --> {[b1+complete]} = 1
Place {[a2+complete]} --> {[k2p+complete, b2+complete]} = 2
Place {[a2+complete]} --> {[a3+complete, k2p+complete]} = 2
Place {[a3+complete]} --> {[b3+complete, b2+complete]} = 2
Place {[a3+complete]} --> {[b2+complete, a4+complete]} = 2
Place {[a4+complete]} --> {[b3+complete, b4+complete]} = 2
Place {[b3+complete, a2+complete]} --> {[a3+complete]} = 1
Place {[a3+complete, b4+complete]} --> {[a4+complete]} = 1
Place {[a1+complete, b2+complete]} --> {[a2+complete]} = 1

```

Σχήμα 4.2.4: Οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων του σηματοδοτικού μονοπατιού ERK-RAS.

Στη συνέχεια παρουσιάζεται το μοντέλο του συγκεκριμένου μονοπατιού που κατασκευάστηκε από τα ίχνη με βάση τον αλγόριθμο παραγωγής βιολογικών ουσιών.

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις ουσίες κατασκεύασε 895 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.5. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.175. Συνολικά δημιουργήθηκαν 15 θέσεις και 11 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές ουσίες του σηματοδοτικού μονοπατιού ERK-RAS. Επίσης το μοντέλο διαθέτει 58 ακμές.



Σχήμα 4.2.5: ERK-RAS Σηματοδοτικό Μονοπάτι ουσιών

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε ουσία στο αρχείο καταγραφής, όπου περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις ουσίες (Κεφάλαιο 3).

Όνομα ουσίας	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
RASp	808	19.173%
RAFp	803	19.609%
MEKp	595	14.53%
RAFu	437	10.672%
RASu	370	9.035%
MEKu	360	8.791%
RAGEp	230	5.617%
ERKp	192	4.689%
ERKu	150	3.699%
PI3Kp	86	2.1%
PI3Ku	64	1.563%

Πίνακας 4.2.4: Εμφανίσεις ουσιών σηματοδοτικού μονοπατιού ERK-RAS στα ίχνη

Αξιολόγηση μοντέλων

Στο κατασκευασμένο μοντέλο με βάση τις βιολογικές αντιδράσεις δημιουργήθηκαν εννιά μεταβάσεις που αντιστοιχούν στις εννιά βιολογικές αντιδράσεις. Τα ποσοστά εμφανίσεων των βιολογικών αντιδράσεων του Πίνακα 4.2.2 μπορούν ένα περίπου να εξαχθούν και από το Πίνακα 4.2.3 που δείχνει τη πιθανή επόμενη βιολογική αντίδραση που μπορεί να πυροδοτήσει για κάθε αντίδραση. Στο Σχήμα 4.2.4 που δείχνει τις βασικές σχέσεις μεταξύ των αντιδράσεων παρατηρούμε τις αλληλεπιδράσεις, όπως εξάγονται από τον Α-αλγόριθμο μέσω της κατασκευής του μοντέλου. Για παράδειγμα από τις δύο πρώτες γραμμές του Σχήματος 4.2.4 εξάγεται ότι στο Βιολογικό σύστημα μπορεί να επαναλαμβάνεται ένας κύκλος ενεργοποιήσεων μεταξύ των αντιδράσεων a1 και b1. Αφού η αντίδραση b1 χρειάζεται την πρωτεΐνη RASp για να ενεργοποιηθεί, την οποία παράγει η αντίδραση a1. Ενώ η αντίδραση a1 εκτός από την πρωτεΐνη RAGEp χρειάζεται και την RASu για να ενεργοποιηθεί, η οποία είναι το προϊόν της ενεργοποίησης της αντίδρασης b1. Παρομοίως με το ίδιο τρόπο στις γραμμές 12 και 14 εξάγεται και η περίπτωση επαναλαμβανόμενου κύκλου μεταξύ των αντιδράσεων a4 και b4. Η βιολογική αντίδραση k2p χρειάζεται την πρωτεΐνη RASp για να ενεργοποιηθεί η οποία είναι προϊόν παραγωγής των αντιδράσεων a1 και a2. Οπότε για να μπορεί να ενεργοποιηθεί η αντίδραση k2p θα πρέπει πριν να ενεργοποιηθεί είτε η αντίδραση a1 ή a2. Αυτό φαίνεται και από τις

γραμμές 4, 8, και 9 του Σχήματος 4.2.4. Η βιολογική αντίδραση a2 για να ενεργοποιηθεί χρειάζεται την κατανάλωση των ουσιών RAFu και RASp. Η πρωτεΐνη RAFu παράγεται μόνο από την βιολογική αντίδραση b2. Ενώ η RASp από την a1 ή την k2p. Επομένως, για να μπορεί να ενεργοποιηθεί η a2 εξάγεται το συμπέρασμα ότι θα πρέπει να ενεργοποιηθούν οι b2 και k2p ή οι b2 και a1. Στην περίπτωση της αντίδρασης a4 παρατηρούμε ότι εμφανίζεται στις γραμμές 11 και 14. Σε τέτοια περίπτωση όταν μια αντίδραση εμφανίζεται σε περισσότερο από μια σχέσεις ως η επόμενη αντίδραση, λαμβάνεται υπόψη η γραμμή με το μέγιστο σύνολο των ηγούμενων. Οπότε η αντίδραση a4 χρειάζεται τις ουσίες ERKu και MEKp, όπου σύμφωνα με τη γραμμή 14 θα πρέπει να ενεργοποιηθούν οι αντιδράσεις b4 και a3 αντίστοιχα. Με παρόμοιο τρόπο η αντίδραση a3 για να ενεργοποιηθεί χρειάζεται την ενεργοποίηση των αντιδράσεων b3 και a2, που παράγουν τις βιολογικές ουσίες MEKu και RAFp αντίστοιχα. Επιπρόσθετα η ενεργοποίηση της αντίδρασης a2 παράγει τις ουσίες RASp και RAFp που καθιστούν τις αντιδράσεις b1 και b2 ενεργές αντιδράσεις. Επίσης η πρωτεΐνη RASp μπορεί να παραχθεί και από την ενεργοποίηση της k2p που φαίνεται και από τη γραμμή 7 του Σχήματος 4.2.4. Από τις γραμμές 10 και 11 παρατηρούμε ότι η ενεργοποίηση της αντίδρασης a3 παράγει τις πρωτεΐνες MEKp και RAFp, όπου η πρώτη είναι η προϋπόθεση της ενεργοποίησης της αντίδρασης b3 και η δεύτερη είναι η προϋπόθεση της αντίδρασης b2 αντίστοιχα.

Το μοντέλο των βιολογικών ουσιών περιέχει 11 μεταβάσεις που αντιστοιχούν στις βιολογικές ουσίες ολόκληρου του μονοπατιού. Το μοντέλο δείχνει τη σειρά που παράγονται και καταναλώνονται οι βιολογικές ουσίες στο δίκτυο. Για παράδειγμα εάν υπάρχει η MEKu τότε πυροδοτείτε η αντίδραση a3 και παράγεται η MEKp. Οπότε στο μοντέλο των βιολογικών ουσιών υπάρχει ένα μονοπάτι που συνδέει τις δύο βιολογικές ουσίες. Παρομοίως δημιουργεί όλα τα υπόλοιπα μονοπάτια που αποτελούν το συγκεκριμένο μοντέλο. Στο Πίνακα 4.2.4 παρουσιάζονται οι εμφανίσεις των βιολογικών ουσιών στο σύστημα.

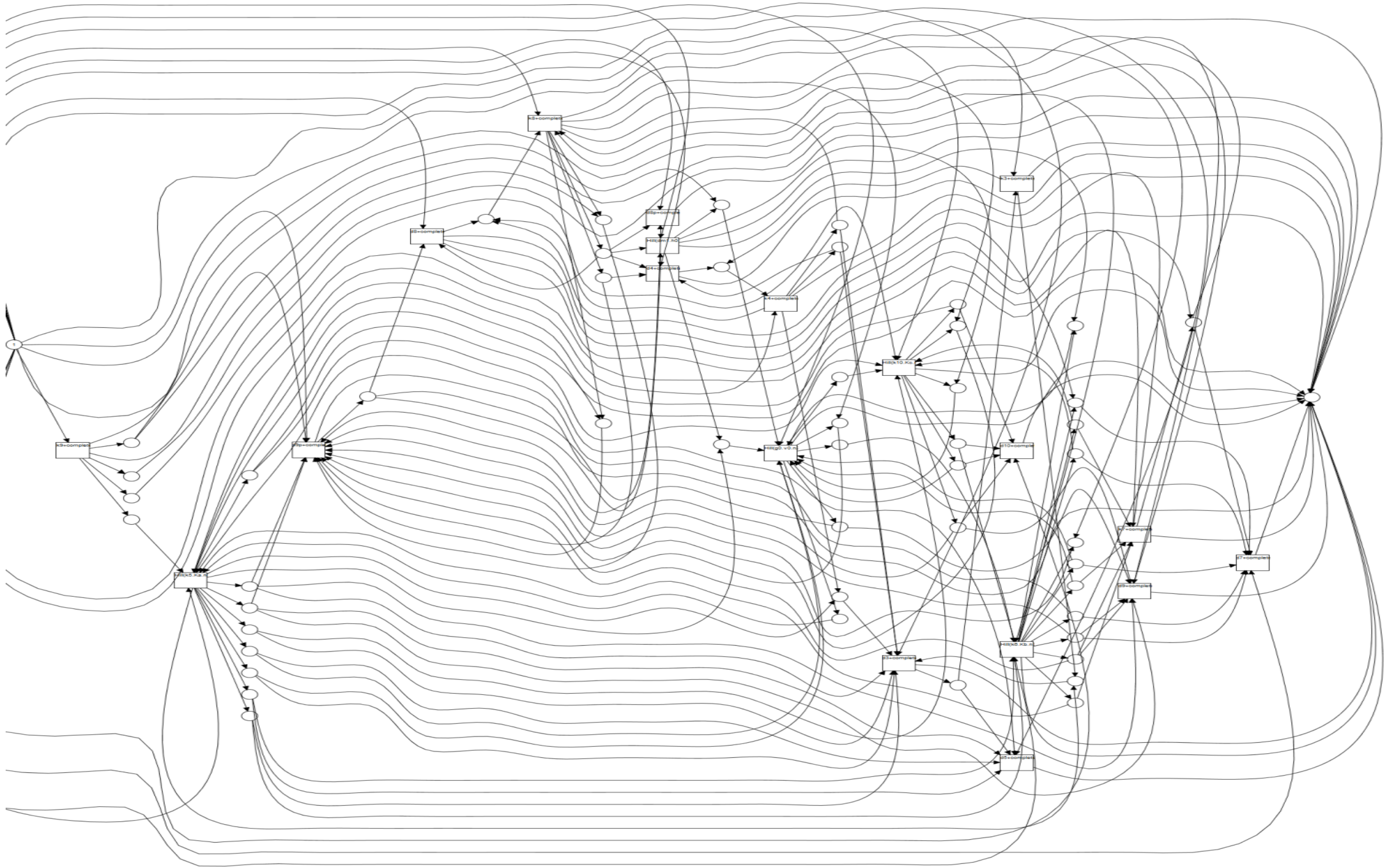
β) P53 – MDM2 Σηματοδοτικό Μονοπάτι

Οι κανόνες της BioNetGen για αυτό το μονοπάτι έχουν ως εξής :

Κανόνας	Ρυθμός Αντίδρασης
1. PI3K(b~p) + PIP(c~U) -> PI3K(b~p) + PIP(c~p)	k3
2. PTEN + PIP(c~p) -> PTEN + PIP(c~U)	d3
3. PIP(c~p) + AKT(d~U) -> PIP(c~p) + AKT(d~p)	k4
4. AKT(d~p) -> AKT(d~U)	d4
5. AKT(d~p) + MDM2(e~U) -> AKT(d~p) + MDM2(e~p)	k8
6. MDM2(e~p) -> MDM2(e~U)	d8
7. MDM2(e~U) -> Trash()	d7
8. MDM2(e~p) -> Trash()	d8p
9. mdm2t -> mdm2t + MDM2(e~U)	k7
10. mdm2t -> Trash()	d7
11. DNAD + MDM2(e~p) -> DNAD	Hill(dm1,h0,n2)
12. P53 -> P53 + mdm2t	Hill(k6,Kb,n3)
13. I() -> I() + P53	k9
14. P53 -> Trash()	d9
15. MDM2(e~p) + P53 -> MDM2(e~p)	d9p
16. P53 -> P53+ PTEN	Hill(k5,Ka,n3)
17. PTEN -> Trash()	d5
18. P53 -> P53 + P21	Hill(k10,Kc,n2)
19. P21 -> Trash()	d10
20. P53 + DNAD -> P53 + DNA	Hill(g0,v0,n2)

Πίνακας 4.2.5 : Κανόνες για το P53-MDM2 μονοπάτι

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις αντιδράσεις κατασκεύασε 4561 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.6. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.098. Συνολικά δημιουργήθηκαν 51 θέσεις και 19 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές αντιδράσεις του σηματοδοτικού μονοπατιού P53-MDM2. Επίσης το μοντέλο διαθέτει 189 ακμές.

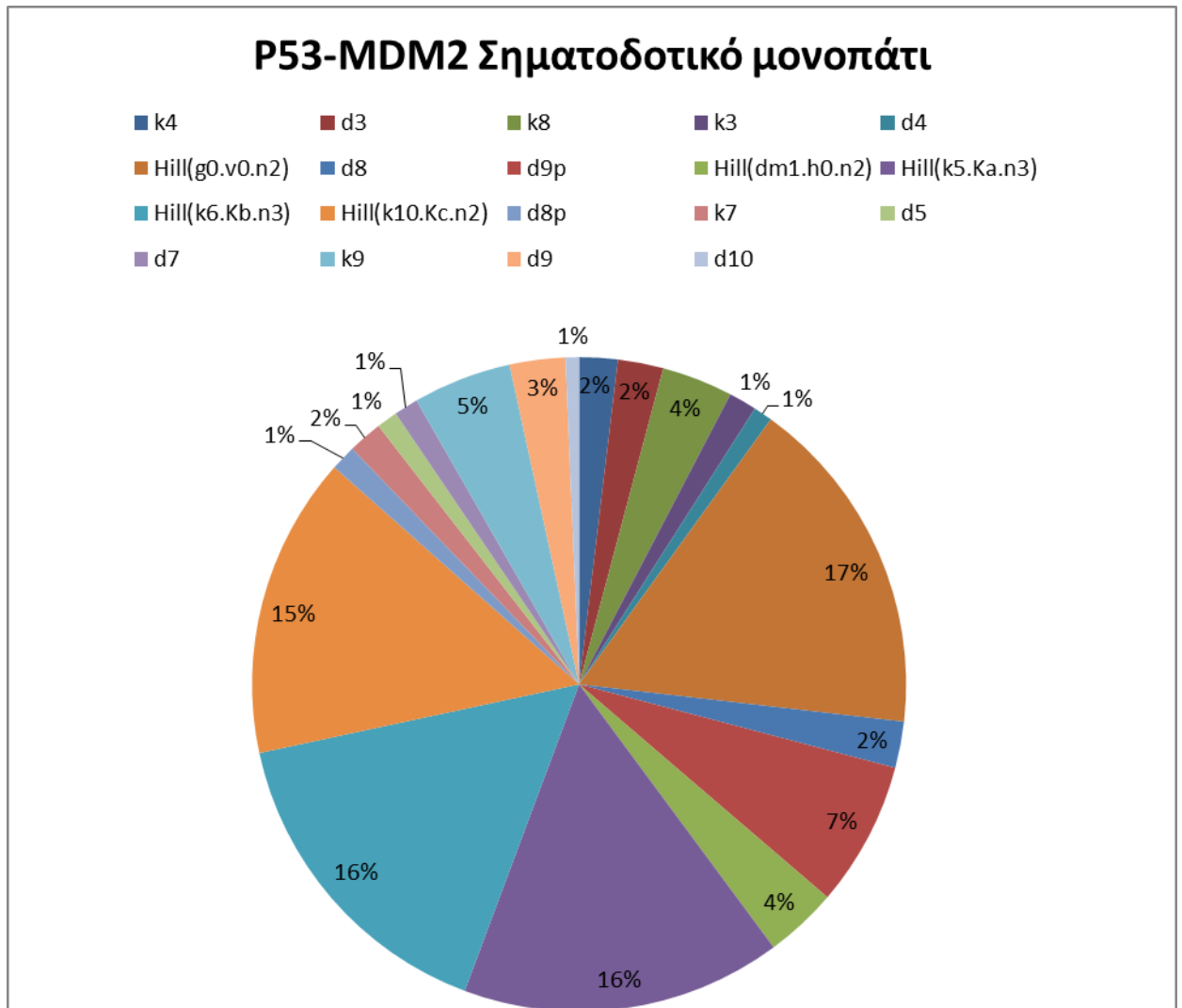


Σχήμα 4.2.6: P53-MDM2 Σηματοδοτικό Μονοπάτι αντιδράσεων

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε αντίδραση στο αρχείο καταγραφής, το οποίο περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις αντιδράσεις (Κεφάλαιο 3). Ακολούθως παρουσιάζεται η γραφική παράσταση του αντίστοιχου πίνακα.

Όνομα αντίδρασης	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
k4	347	2%
d3	411	2%
k8	642	4%
k3	252	1%
d4	171	0,94%
Hill(g0.v0.n2)	3080	17%
d8	418	2%
d9p	1310	7%
Hill(dm1.h0.n2)	655	4%
Hill(k5.Ka.n3)	2893	16%
Hill(k6.Kb.n3)	2914	16%
Hill(k10.Kc.n2)	2726	15%
d8p	230	1%
k7	309	1,69%
d5	189	1%
d7	218	1%
k9	890	5%
d9	503	3%
d10	121	1%

Πίνακας 4.2.6: Εμφανίσεις αντιδράσεων σηματοδοτικού μονοπατιού P53-MDM2 στα ίχνη



Σχήμα 2.4.7: Γραφική παράσταση της εμφάνισης των αντιδράσεων σηματοδοτικού μονοπατιού P53-MDM2 στα ίχνη

Μέσω της κατασκευής του μοντέλου για το σηματοδοτικό μονοπάτι P53-MDM2, παρατίθενται στο πιο κάτω πίνακα η επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο μονοπάτι.

Αντίδραση	Πιθανή επόμενη εκτελούσα αντίδραση
k4	d4,k8,d3,k4
d3	d5,k3,d3
k8	d4,d8p,d9p,k8 Hill(dm1.h0.n2),d8
k3	k4,d3,k3
d4	k4
Hill(g0.v0.n2)	d9,Hill(k6.Kb.n3),d9p, Hill(g0.v0.n2), Hill(k10.Kc.n2), Hill(k5.Ka.n3)
d8	k8
d9p	Hill(dm1.h0.n2),d8,d9p,d8p
Hill(dm1.h0.n2)	Hill(g0.v0.n2), Hill(dm1.h0.n2)
Hill(k5.Ka.n3)	d5,d3, Hill(k5.Ka.n3), Hill(k10.Kc.n2),d9, Hill(k6.Kb.n3),d9p, Hill(g0.v0.n2)
Hill(k6.Kb.n3)	d9,Hill(k6.Kb.n3),d9p, Hill(g0.v0.n2), Hill(k10.Kc.n2), Hill(k5.Ka.n3),d7,k7
Hill(k10.Kc.n2)	d10, d9,Hill(k6.Kb.n3),d9p, Hill(g0.v0.n2), Hill(k10.Kc.n2), Hill(k5.Ka.n3)
d8p	
k7	k8,d7
d5	
d7	
k9	k9,Hill(k5.Ka.n3), Hill(k10.Kc.n2),d9, Hill(k6.Kb.n3),d9p, Hill(g0.v0.n2)
d9	
d10	

Πίνακας 4.2.7: Επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο σηματοδοτικό μονοπάτι P53-MDM2

Στο Σχήμα 4.2.8 παρατίθενται οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων όπως τις δίνει η λειτουργία “Show Petri-net Metrics” για το σηματοδοτικό μονοπάτι P53-MDM2.

```

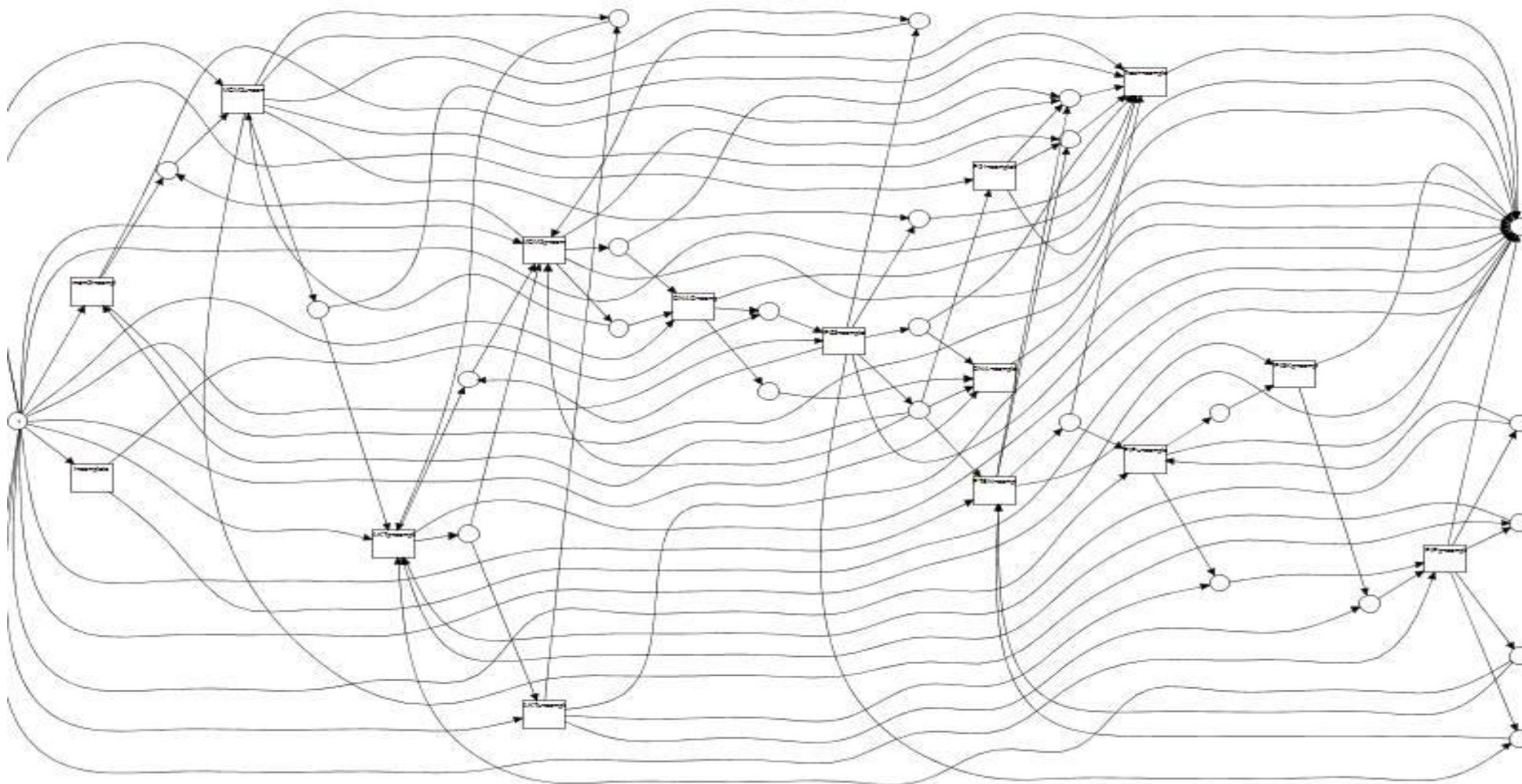
Place {[Hill(g0.v0.n2)+complete]} --> {[Hill(k6.Kb.n3)+complete]} = 1
Place {[k9+complete]} --> {[Hill(k5.Ka.n3)+complete]} = 1
Place {[Hill(g0.v0.n2)+complete]} --> {[Hill(k5.Ka.n3)+complete]} = 1
Place {[k9+complete]} --> {[Hill(k10.Kc.n2)+complete]} = 1
Place {[k9+complete]} --> {[Hill(k6.Kb.n3)+complete]} = 1
Place {[Hill(g0.v0.n2)+complete]} --> {[Hill(k10.Kc.n2)+complete]} = 1
Place {[Hill(k5.Ka.n3)+complete]} --> {[Hill(g0.v0.n2)+complete, d3+complete]} = 2
Place {[k4+complete]} --> {[d3+complete, d4+complete]} = 2
Place {[d3+complete]} --> {[k3+complete, d5+complete]} = 2
Place {[Hill(k5.Ka.n3)+complete]} --> {[Hill(k10.Kc.n2)+complete, d3+complete]} = 2
Place {[Hill(k5.Ka.n3)+complete]} --> {[Hill(k10.Kc.n2)+complete, d5+complete]} = 2
Place {[k4+complete, Hill(k5.Ka.n3)+complete]} --> {[d3+complete]} = 1
Place {[Hill(k5.Ka.n3)+complete]} --> {[Hill(k6.Kb.n3)+complete, d3+complete]} = 2
Place {[k4+complete]} --> {[k8+complete, d3+complete]} = 2
Place {[Hill(k5.Ka.n3)+complete]} --> {[Hill(g0.v0.n2)+complete, d5+complete]} = 2
Place {[d8+complete, Hill(k6.Kb.n3)+complete]} --> {[d7+complete]} = 1
Place {[Hill(k6.Kb.n3)+complete]} --> {[k7+complete, Hill(k5.Ka.n3)+complete]} = 2
Place {[k3+complete, Hill(k5.Ka.n3)+complete]} --> {[d3+complete]} = 1
Place {[Hill(k10.Kc.n2)+complete]} --> {[Hill(k6.Kb.n3)+complete, d10+complete]} = 2
Place {[k9+complete]} --> {[d9p+complete, d9+complete]} = 2
Place {[Hill(dm1.h0.n2)+complete, k9+complete]} --> {[Hill(g0.v0.n2)+complete]} = 1
Place {[Hill(k5.Ka.n3)+complete]} --> {[d5+complete, Hill(k6.Kb.n3)+complete]} = 2
Place {[k3+complete, d4+complete]} --> {[k4+complete]} = 1
Place {[Hill(g0.v0.n2)+complete]} --> {[d9p+complete, d9+complete]} = 2
Place {[Hill(g0.v0.n2)+complete, k8+complete]} --> {[d9p+complete]} = 1
Place {[Hill(k10.Kc.n2)+complete, k8+complete]} --> {[d9p+complete]} = 1
Place {[Hill(dm1.h0.n2)+complete, Hill(k5.Ka.n3)+complete]} --> {[Hill(g0.v0.n2)+complete]} = 1
Place {[Hill(k6.Kb.n3)+complete]} --> {[d7+complete, Hill(g0.v0.n2)+complete]} = 2
Place {[Hill(k6.Kb.n3)+complete]} --> {[d7+complete, Hill(k5.Ka.n3)+complete]} = 2
Place {[k8+complete, Hill(k5.Ka.n3)+complete]} --> {[d9p+complete]} = 1
Place {[Hill(k6.Kb.n3)+complete, k8+complete]} --> {[d9p+complete]} = 1
Place {[Hill(k10.Kc.n2)+complete]} --> {[Hill(k5.Ka.n3)+complete, d10+complete]} = 2
Place {[Hill(k10.Kc.n2)+complete]} --> {[Hill(g0.v0.n2)+complete, d10+complete]} = 2
Place {[k8+complete, k9+complete]} --> {[d9p+complete]} = 1
Place {[k8+complete]} --> {[d9p+complete, d4+complete]} = 2
Place {[Hill(k10.Kc.n2)+complete, Hill(dm1.h0.n2)+complete]} --> {[Hill(g0.v0.n2)+complete]} = 1
Place {[Hill(k6.Kb.n3)+complete]} --> {[Hill(k10.Kc.n2)+complete, k7+complete]} = 2
Place {[Hill(dm1.h0.n2)+complete, Hill(k6.Kb.n3)+complete]} --> {[Hill(g0.v0.n2)+complete]} = 1
Place {[Hill(k5.Ka.n3)+complete]} --> {[d9p+complete, d9+complete, d3+complete]} = 3
Place {[Hill(k5.Ka.n3)+complete]} --> {[d9p+complete, d5+complete, d9+complete]} = 2
Place {[Hill(k10.Kc.n2)+complete]} --> {[d9p+complete, d9+complete, d10+complete]} = 2
Place {[d8+complete, k7+complete, k4+complete]} --> {[k8+complete]} = 1
Place {[Hill(k6.Kb.n3)+complete]} --> {[d9p+complete, k7+complete, d9+complete]} = 3
Place {[Hill(k6.Kb.n3)+complete]} --> {[Hill(g0.v0.n2)+complete, k7+complete]} = 2
Place {[k8+complete]} --> {[d8+complete, Hill(dm1.h0.n2)+complete, d8p+complete, d4+complete]} = 4
Place {[Hill(k6.Kb.n3)+complete]} --> {[d7+complete, d9p+complete, d9+complete]} = 2
Place {[Hill(k6.Kb.n3)+complete]} --> {[d7+complete, Hill(k10.Kc.n2)+complete]} = 2
Place {[d8+complete, k7+complete]} --> {[d7+complete, k8+complete]} = 2
Place {[d9p+complete]} --> {[d8+complete, Hill(dm1.h0.n2)+complete, d8p+complete]} = 3

```

Σχήμα 4.2.8: Οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων του σηματοδοτικού μονοπατιού P53-MDM2

Στη συνέχεια παρουσιάζεται το μοντέλο του συγκεκριμένου μονοπατιού που κατασκευάστηκε από τα ίχνη με βάση τον αλγόριθμο παραγωγής βιολογικών ουσιών.

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις ουσίες κατασκεύασε 2151 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.9. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.137. Συνολικά δημιουργήθηκαν 25 θέσεις και 15 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές ουσίες του σηματοδοτικού μονοπατιού P53-MDM2. Επίσης το μοντέλο διαθέτει 103 ακμές.



Σχήμα 4.2.9: P53-MDM2 Σηματοδοτικό Μονοπάτι ουσιών

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε ουσία στο αρχείο καταγραφής, όπου περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις ουσίες (Κεφάλαιο 3).

Όνομα ουσίας	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
MDM2p	1419	14.965%
PIPp	1144	12.065%
AKTp	1122	11.833%
P53	965	10.177%
MDM2u	815	8.595%
DNAD	747	7.878%
PIPu	709	7.477%
PTEN	538	5.674%
AKTu	534	5.632%
Trash	450	4.746%
PI3Kp	358	3.776%
mdm2t	298	3.143%
I	186	1.962%
DNA	140	1.476%
P21	57	0.601%

Πίνακας 4.2.8: Εμφανίσεις ουσιών σηματοδοτικού μονοπατιού P53-MDM2 στα ίχνη

Αξιολόγηση μοντέλων

Στο κατασκευασμένο μοντέλο με βάση τις βιολογικές αντιδράσεις δημιουργήθηκαν εννιά μεταβάσεις που αντιστοιχούν στις εννιά βιολογικές αντιδράσεις. Τα ποσοστά εμφανίσεων των βιολογικών αντιδράσεων του Πίνακα 4.2.6 μπορούν ένα περίπου να εξαχθούν και από το Πίνακα 4.2.7 που δείχνει τη πιθανή επόμενη βιολογική αντίδραση που μπορεί να πυροδοτήσει για κάθε αντίδραση. Στο Σχήμα 4.2.8 που δείχνει τις βασικές σχέσεις μεταξύ των αντιδράσεων παρατηρούμε τις σχέσεις που εξάγει ο Α-αλγόριθμος μέσω της κατασκευής του μοντέλου. Για παράδειγμα η πυροδότηση της αντίδρασης Hill(k10.kc.n2) παράγει την ουσία P21 η οποία είναι η προϋπόθεση της d10. Όσο αφορά πιο σύνθετες σχέσεις παρατηρούμε πως επίσης το μοντέλο τις αντιλαμβάνεται ορθά. Για παράδειγμα η k8 παράγει την ουσία MDM2p και η Hill(k10.kc.n2) παράγει την ουσία P53. Οι δύο ουσίες που παράγονται είναι οι

ουσίες που χρειάζονται για την πυροδότηση της αντίδρασης d9p. Όλες οι υπόλοιπες εξαρτήσεις και αλληλεπιδράσεις μεταξύ των αντιδράσεων μπορούν παραχθούν από το Σχήμα 4.2.8 με το ίδιο τρόπο όπως έγινε στο μονοπάτι σηματοδότησης ERK-RAS όπου έγινε αναλυτική περιγραφή. Οπότε στο μοντέλο που κατασκευάζεται, οι αντιδράσεις που συνδέονται μεταξύ τους υπάρχει μονοπάτι που τις παρουσιάζει. Ένα άλλο σημαντικό χαρακτηριστικό που εξάχθηκε είναι ο επαναλαμβανόμενος κύκλος d4, AKTu, d4, AKTu, d4, AKTu καθώς επίσης και ο κύκλος μεγαλύτερου μήκους Hill(dm1.h0.n2), Hill(g0.v0.n2), d9p, Hill(dm1.h0.n2), Hill(g0.v0.n2), d9p....

Το μοντέλο των βιολογικών ουσιών περιέχει 15 μεταβάσεις που αντιστοιχούν στις βιολογικές ουσίες ολόκληρου του μονοπατιού. Το μοντέλο δείχνει τη σειρά που παράγονται και καταναλώνονται οι βιολογικές ουσίες στο δίκτυο. Για παράδειγμα εάν υπάρχει η ουσία PIPu, τότε πυροδοτείται η αντίδραση k3 και παράγεται η PIPp. Οπότε στο μοντέλο των βιολογικών ουσιών υπάρχει ένα μονοπάτι που συνδέει τις δύο βιολογικές ουσίες. Παρομοίως δημιουργεί όλα τα υπόλοιπα μονοπάτια που αποτελούν το συγκεκριμένο μοντέλο. Στο Πίνακα 4.2.8 παρουσιάζονται οι εμφανίσεις των βιολογικών ουσιών στο σύστημα.

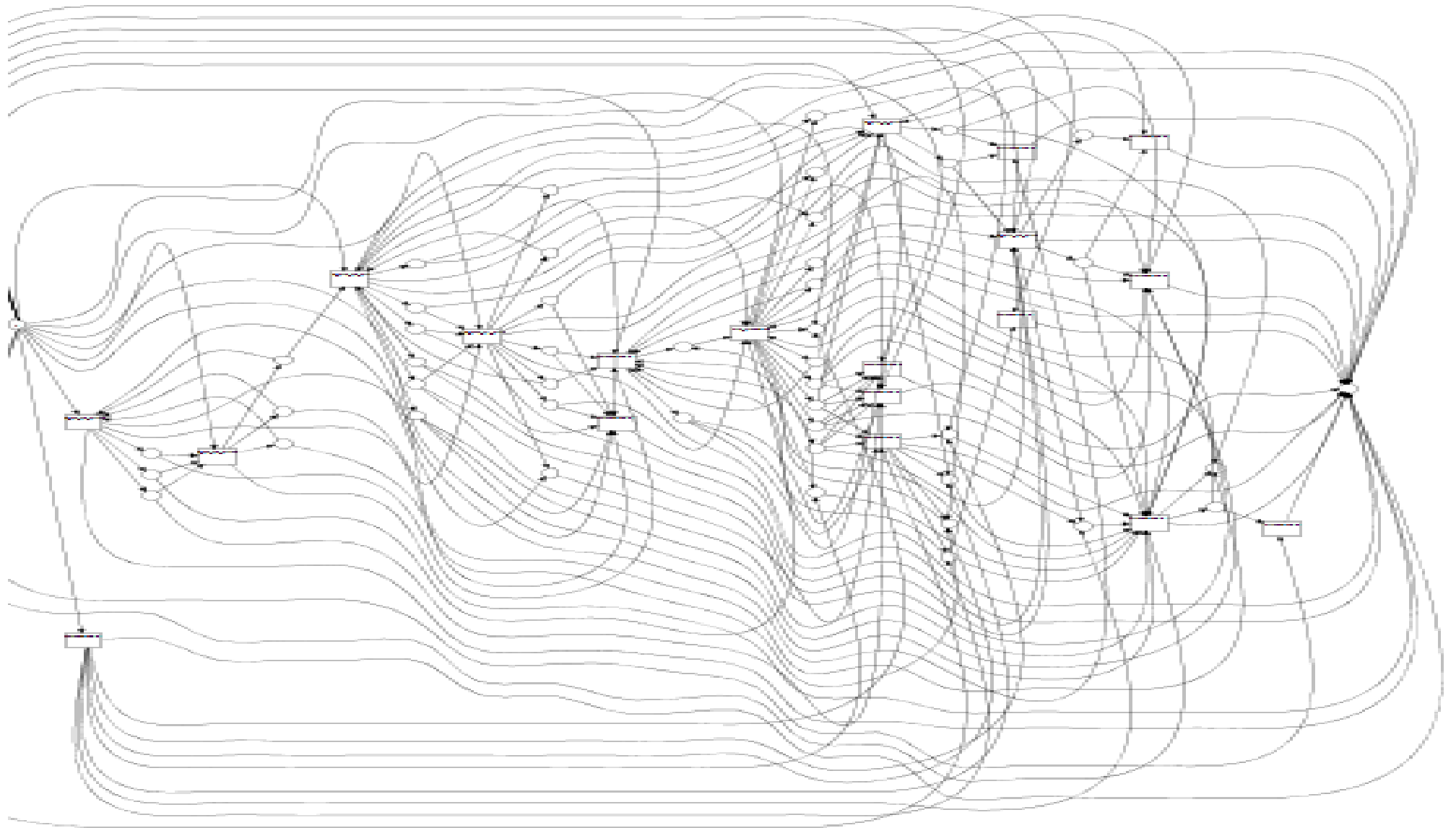
γ) Rb – E2F Σηματοδοτικό Μονοπάτι

Οι κανόνες της BioNetGen για αυτό το μονοπάτι έχουν ως εξής :

Κανόνας	Ρυθμός Αντίδρασης
1. I() -> I() + RB(c~U)	a9
2. RB(c~U) -> Trash()	b9
3. CD -> Trash()	b6
4. RB(c~U) + CD -> RB(e~p) + CD	a8
5. RB(e~p) -> Trash()	b8p
6. RB(e~p) -> RB(c~U)	b8
7. RE -> Trash()	b7
8. RB(c~U) + E2F -> RE	a7
9. CE-> Trash()	b12
10. ERK(e~p) -> CD + ERK(e~p)	Hill(a6,ke,n1)
11. RE + CD -> E2F + RB(e~p) + CD	b7p_2
12. Myc -> Myc + CD	Hill(a6p,kf,n1)
13. E2F -> E2F + CE	Hill(a12,ki,n1)
14. ERK(e~p) -> ERK(e~p) + Myc	Hill(a5,kd,n1)
15. Myc -> E2F + Myc	Hill(a10,kg,n1)
16. ARF -> Trash()	b11
17. Myc -> Trash()	b5
18. E2F -> Trash()	b10
19. E2F -> ARF + E2F	Hill(a11,kh,n1)

Πίνακας 4.2.9 : Κανόνες για το Rb – E2F μονοπάτι

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις αντιδράσεις κατασκεύασε 2541 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.10. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.112. Συνολικά δημιουργήθηκαν 46 θέσεις και 19 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές αντιδράσεις του σηματοδοτικού μονοπατιού Rb – E2F. Επίσης το μοντέλο διαθέτει 195 ακμές.

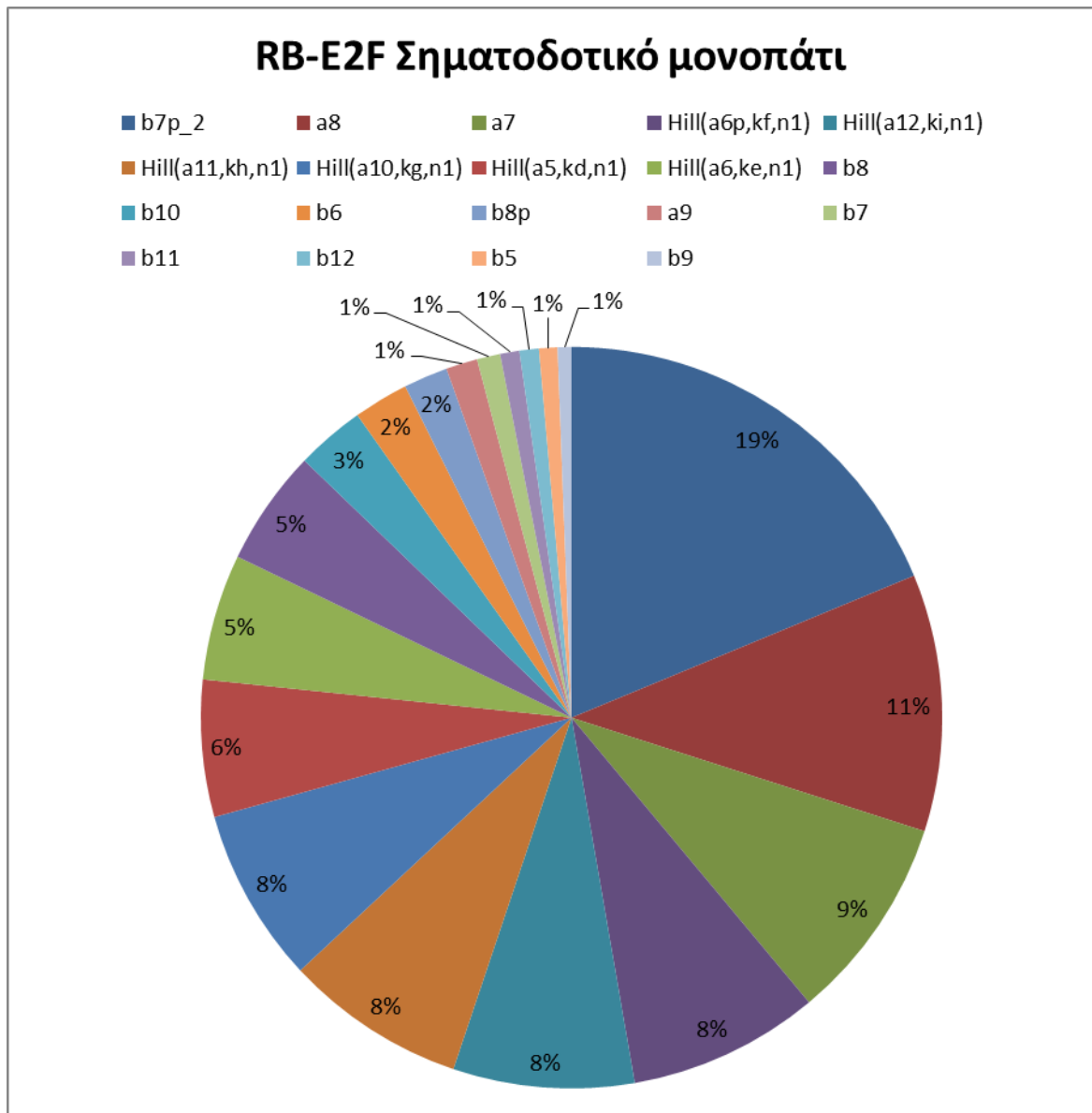


Σχήμα 4.2.10: Rb – E2F Σηματοδοτικό Μονοπάτι αντιδράσεων

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε αντίδραση στο αρχείο καταγραφής, το οποίο περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις αντιδράσεις (Κεφάλαιο 3). Ακολούθως παρουσιάζεται η γραφική παράσταση του αντίστοιχου πίνακα.

Όνομα αντίδρασης	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
b7p_2	1868	18,744%
a8	1117	11,208%
a7	895	8,981%
Hill(a6p,kf,n1)	831	8,338%
Hill(a12,ki,n1)	787	7,897%
Hill(a11,kh,n1)	787	7,897%
Hill(a10,kg,n1)	759	7,616%
Hill(a5,kd,n1)	595	5,97%
Hill(a6,ke,n1)	548	5,499%
b8	503	5,047%
b10	299	3%
b6	239	2,398%
b8p	192	1,927%
a9	137	1,375%
b7	101	1,013%
b11	84	0,843%
b12	84	0,843%
b5	80	0,803%
b9	60	0,602%

Πίνακας 4.2.10: Εμφανίσεις αντιδράσεων σηματοδοτικού μονοπατιού Rb – E2F στα ίχνη



Σχήμα 4.2.11: Γραφική παράσταση της εμφάνισης των αντιδράσεων σηματοδοτικού μονοπατιού Rb – E2F στα ίχνη

Μέσω της κατασκευής του μοντέλου για το σηματοδοτικό μονοπάτι Rb – E2F, παρατίθενται στο πιο κάτω πίνακα η επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο μονοπάτι.

Όνομα αντίδρασης	Πιθανή επόμενη εκτελούσα αντίδραση
b7p_2	b8p,b8,a8,b7p_2,b10, Hill(a11,kh,n1),Hill(a12,ki,n1), a7
a8	b6,b7p_2,a8,b8p,b8,
a7	b7,b7p_2
Hill(a6p,kf,n1)	b6,a8,b7p_z,Hill(a6p,kf,n1), b5,Hill(a10,kg,n1)
Hill(a12,ki,n1)	b12,Hill(a11,kh,n1),Hill(a12,ki,n1) b10,a7
Hill(a11,kh,n1)	b11,b10,a7, Hill(a11,kh,n1),Hill(a12,ki,n1)
Hill(a10,kg,n1)	b5,Hill(a6p,kf,n1),b10,Hill(a11,kh,n1) Hill(a12,ki,n1),a7,Hill(a10,kg,n1)
Hill(a5,kd,n1)	Hill(a6,ke,n1),Hill(a5,kd,n1),b5, Hill(a10,kg,n1),Hill(a6p,kf,n1)
Hill(a6,ke,n1)	Hill(a6,ke,n1),Hill(a5,kd,n1),b7p_2, a8,b6
b8	b9,a7
b10	
b6	
b8p	
a9	a7,b9,a9
b7	
b11	
b12	
b5	
b9	

Πίνακας 4.2.11: Επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο σηματοδοτικό μονοπάτι Rb – E2F

Στο Σχήμα 4.2.12 παρατίθενται οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων όπως τις δίνει η λειτουργία “Show Petri-net Metrics” για το σηματοδοτικό μονοπάτι Rb – E2F.

```

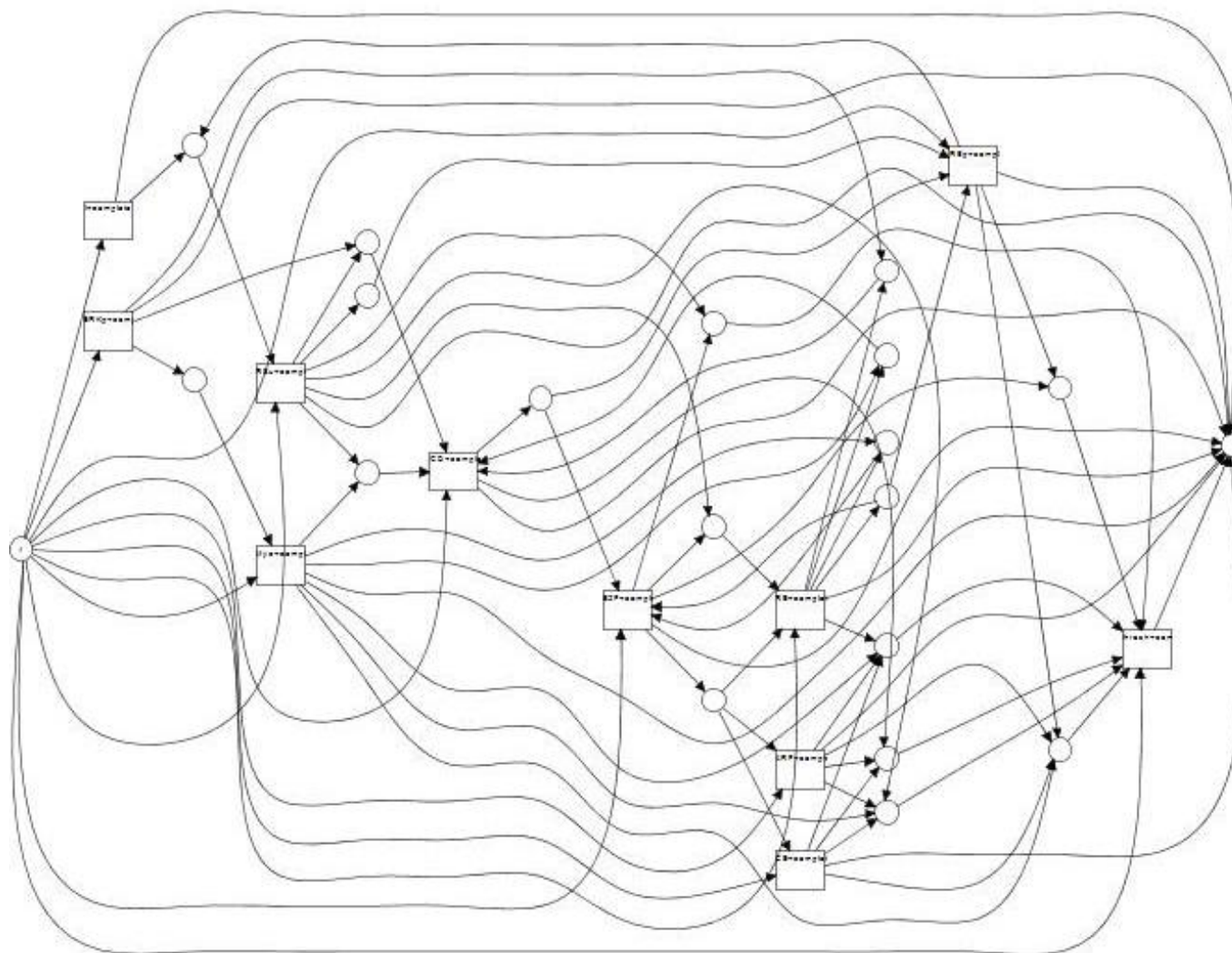
Place {[Hill_a6p_kf_n1+complete]} --> {[a8+complete, Hill_a10_kg_n1+complete]} = 2
Place {[Hill_a5_kd_n1+complete]} --> {[Hill_a6_ke_n1+complete, b5+complete]} = 2
Place {[Hill_a5_kd_n1+complete]} --> {[Hill_a6_ke_n1+complete, Hill_a6p_kf_n1+complete]} = 2
Place {[b7p_2+complete]} --> {[a8+complete, Hill_a11_kh_n1+complete]} = 2
Place {[a8+complete, a7+complete]} --> {[b7p_2+complete]} = 1
Place {[Hill_a6p_kf_n1+complete]} --> {[a8+complete, b5+complete]} = 2
Place {[a7+complete]} --> {[b7p_2+complete, b7+complete]} = 2
Place {[Hill_a12_ki_n1+complete]} --> {[Hill_a11_kh_n1+complete, b12+complete]} = 2
Place {[b7p_2+complete, Hill_a10_kg_n1+complete]} --> {[Hill_a12_ki_n1+complete]} = 1
Place {[b7p_2+complete]} --> {[a8+complete, Hill_a12_ki_n1+complete]} = 2
Place {[Hill_a5_kd_n1+complete]} --> {[Hill_a6_ke_n1+complete, Hill_a10_kg_n1+complete]} = 2
Place {[Hill_a10_kg_n1+complete]} --> {[Hill_a12_ki_n1+complete, Hill_a6p_kf_n1+complete]} = 2
Place {[Hill_a6_ke_n1+complete]} --> {[b7p_2+complete, Hill_a5_kd_n1+complete]} = 2
Place {[Hill_a10_kg_n1+complete]} --> {[Hill_a12_ki_n1+complete, b5+complete]} = 2
Place {[Hill_a10_kg_n1+complete]} --> {[Hill_a11_kh_n1+complete, Hill_a6p_kf_n1+complete]} = 2
Place {[Hill_a11_kh_n1+complete]} --> {[b11+complete, Hill_a12_ki_n1+complete]} = 2
Place {[Hill_a10_kg_n1+complete]} --> {[Hill_a11_kh_n1+complete, b5+complete]} = 2
Place {[Hill_a6_ke_n1+complete]} --> {[Hill_a5_kd_n1+complete, a8+complete]} = 2
Place {[Hill_a6_ke_n1+complete]} --> {[Hill_a5_kd_n1+complete, b6+complete]} = 2
Place {[Hill_a6p_kf_n1+complete]} --> {[b7p_2+complete, Hill_a10_kg_n1+complete]} = 2
Place {[Hill_a6p_kf_n1+complete]} --> {[b7p_2+complete, b5+complete]} = 2
Place {[b7p_2+complete, Hill_a10_kg_n1+complete]} --> {[Hill_a11_kh_n1+complete]} = 1
Place {[Hill_a10_kg_n1+complete]} --> {[a7+complete, Hill_a6p_kf_n1+complete, b10+complete]} = 3
Place {[Hill_a6p_kf_n1+complete]} --> {[b6+complete, b5+complete]} = 1
Place {[Hill_a6p_kf_n1+complete]} --> {[b6+complete, Hill_a10_kg_n1+complete]} = 2
Place {[Hill_a10_kg_n1+complete]} --> {[a7+complete, b10+complete, b5+complete]} = 2
Place {[Hill_a6_ke_n1+complete, Hill_a6p_kf_n1+complete]} --> {[b6+complete]} = 1
Place {[b7p_2+complete, Hill_a10_kg_n1+complete]} --> {[a7+complete, b10+complete]} = 2
Place {[b7p_2+complete, a9+complete, Hill_a10_kg_n1+complete]} --> {[a7+complete]} = 1
Place {[Hill_a11_kh_n1+complete]} --> {[b11+complete, a7+complete, b10+complete]} = 2
Place {[a9+complete, b8+complete, Hill_a10_kg_n1+complete]} --> {[a7+complete]} = 1
Place {[b7p_2+complete]} --> {[a8+complete, a7+complete, b10+complete]} = 3
Place {[a9+complete, b8+complete, Hill_a11_kh_n1+complete]} --> {[a7+complete]} = 1
Place {[b7p_2+complete, a9+complete]} --> {[a8+complete, a7+complete]} = 2
Place {[Hill_a6_ke_n1+complete, a7+complete, Hill_a6p_kf_n1+complete]} --> {[b7p_2+complete]} = 1
Place {[a8+complete]} --> {[b8+complete, b6+complete, b8p+complete]} = 2
Place {[a9+complete, Hill_a12_ki_n1+complete, b8+complete]} --> {[a7+complete]} = 1
Place {[Hill_a12_ki_n1+complete]} --> {[a7+complete, b12+complete, b10+complete]} = 2
Place {[b7p_2+complete]} --> {[Hill_a12_ki_n1+complete, b8+complete, b6+complete, b8p+complete]} = 3
Place {[b7p_2+complete]} --> {[b6+complete, a7+complete, b10+complete, b8p+complete]} = 2
Place {[b7p_2+complete]} --> {[b6+complete, b8+complete, b10+complete, b8p+complete]} = 2
Place {[b7p_2+complete]} --> {[b6+complete, b8+complete, Hill_a11_kh_n1+complete, b8p+complete]} = 3
Place {[Hill_a6_ke_n1+complete, a9+complete, b8+complete, Hill_a6p_kf_n1+complete]} --> {[a8+complete]} = 1
Place {[a9+complete, b8+complete]} --> {[a8+complete, b9+complete, a7+complete]} = 3

```

Σχήμα 4.2.12: Οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων του σηματοδοτικού μονοπατιού Rb – E2F

Στη συνέχεια παρουσιάζεται το μοντέλο του συγκεκριμένου μονοπατιού που κατασκευάστηκε από τα ίχνη με βάση τον αλγόριθμο παραγωγής βιολογικών ουσιών.

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις ουσίες κατασκεύασε 1327 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.13. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.186. Συνολικά δημιουργήθηκαν 20 θέσεις και 11 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές ουσίες του σηματοδοτικού μονοπατιού Rb – E2F. Επίσης το μοντέλο διαθέτει 82 ακμές.



Σχήμα 4.2.13: Rb – E2F Σηματοδοτικό Μονοπάτι ουσιών

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε ουσία στο αρχείο καταγραφής, όπου περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις ουσίες (Κεφάλαιο 3).

Όνομα ουσίας	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
E2F	1150	19.021%
CD	1021	16.887%
RE	966	15.978%
RBu	694	11.479%
RBp	616	10.189%
Thrash	555	9.18%
Myc	383	6.335%
ERKp	273	4.515%
ARF	134	2.216%
CE	134	2.216%
I	120	1.985%

Πίνακας 4.2.12: Εμφανίσεις ουσιών σηματοδοτικού μονοπατιού Rb-E2F στα ίχνη

Αξιολόγηση μοντέλων

Στο κατασκευασμένο μοντέλο με βάση τις βιολογικές αντιδράσεις δημιουργήθηκαν εννιά μεταβάσεις που αντιστοιχούν στις εννιά βιολογικές αντιδράσεις. Τα ποσοστά εμφανίσεων των βιολογικών αντιδράσεων του Πίνακα 4.2.10 μπορούν ένα περίπου να εξαχθούν και από το Πίνακα 4.2.11 που δείχνει τη πιθανή επόμενη βιολογική αντίδραση που μπορεί να πυροδοτήσει για κάθε αντίδραση. Στο Σχήμα 4.2.12 που δείχνει τις βασικές σχέσεις μεταξύ των αντιδράσεων παρατηρούμε τις σχέσεις που εξάγει ο Α-αλγόριθμος μέσω της κατασκευής του μοντέλου. Για παράδειγμα η πυροδότηση της αντίδρασης Hill(a11.kh.n1) παράγει την ουσία ARF η οποία είναι η προϋπόθεση της b11. Όσο αφορά πιο σύνθετες σχέσεις παρατηρούμε πως επίσης το μοντέλο τις αντιλαμβάνεται ορθά. Για παράδειγμα η Hill(a11.kh.n1) παράγει την ουσία E2F και η Hill(a5.kd.n1) παράγει την ουσία Myc. Οι δύο ουσίες που παράγονται είναι οι ουσίες που χρειάζονται για την πυροδότηση της αντίδρασης Hill(a10.kg.n1). Όλες οι υπόλοιπες εξαρτήσεις και αλληλεπιδράσεις μεταξύ των αντιδράσεων μπορούν παραχθούν από το Σχήμα 4.2.12 με το ίδιο τρόπο όπως έγινε στο μονοπάτι σηματοδότησης ERK-RAS όπου έγινε αναλυτική περιγραφή. Οπότε στο μοντέλο που κατασκευάζεται, οι αντιδράσεις που συνδέονται μεταξύ τους

υπάρχει μονοπάτι που τις παρουσιάζει. Ένα άλλο σημαντικό χαρακτηριστικό που εξάχθηκε είναι ο επαναλαμβανόμενος κύκλος a8, b8, a7, b7p_z, a8, b8, a7, b7p_z....

Το μοντέλο των βιολογικών ουσιών περιέχει έντεκα μεταβάσεις που αντιστοιχούν στις βιολογικές ουσίες ολόκληρου του μονοπατιού. Το μοντέλο δείχνει τη σειρά που παράγονται και καταναλώνονται οι βιολογικές ουσίες στο δίκτυο. Για παράδειγμα εάν υπάρχει η ουσία ERKp, τότε πυροδοτείται η αντίδραση Hill(a6.ke.n1) και παράγεται η CD. Οπότε στο μοντέλο των βιολογικών ουσιών υπάρχει ένα μονοπάτι που συνδέει τις δύο βιολογικές ουσίες. Παρομοίως δημιουργεί όλα τα υπόλοιπα μονοπάτια που αποτελούν το συγκεκριμένο μοντέλο. Στο Πίνακα 4.2.12 παρουσιάζονται οι εμφανίσεις των βιολογικών ουσιών στο σύστημα.

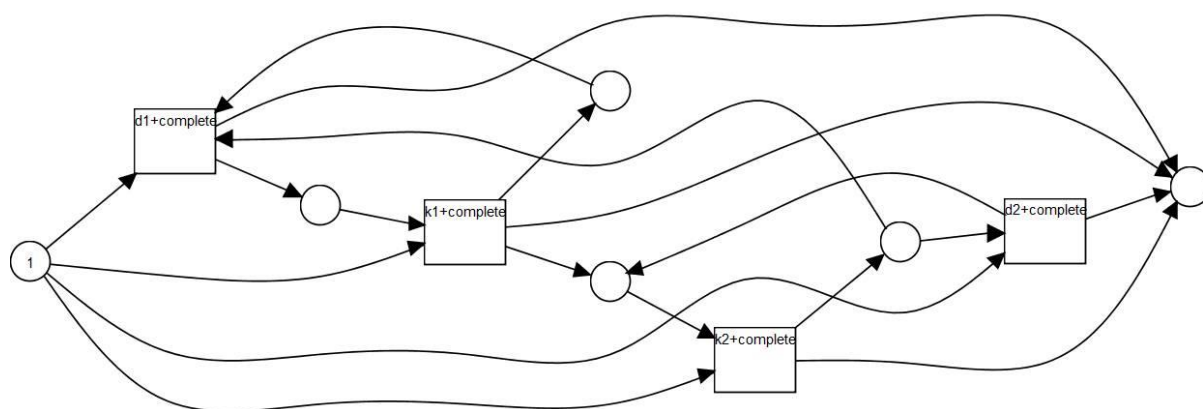
δ) Βασικές Αντιδράσεις Αρχικοποίησης

Αυτοί οι κανόνες δεν συνιστούν ένα ξεχωριστό μονοπάτι αλλά όπως φαίνεται και στο Σχήμα 4.2.1 αποτελούν τα πρώτα σηματοδοτικά βήματα για όλα τα άλλα μονοπάτια. Η πρωτεΐνη HMGB1 βρίσκεται σε αυτό το κομμάτι.

Κανόνας	Ρυθμός Αντίδρασης
1. HMGB1 + RAGE(a~U) -> HMGB1 + RAGE(a~p)	k1
2. RAGE(a~p) -> RAGE(a~U)	d1
3. RAGE(a~p) + PI3K(b~U) -> RAGE(a~p) + PI3K(b~p)	k2
4. PI3K(b~p) -> PI3K(b~U)	d2

Πίνακας 4.2.13: Κανόνες για το ERK –RAS μονοπάτι

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις αντιδράσεις κατασκεύασε 82 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.14. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.375. Συνολικά δημιουργήθηκαν 6 θέσεις και 4 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές αντιδράσεις του σηματοδοτικού μονοπατιού ERK –RAS. Επίσης το μοντέλο διαθέτει 18 ακμές.

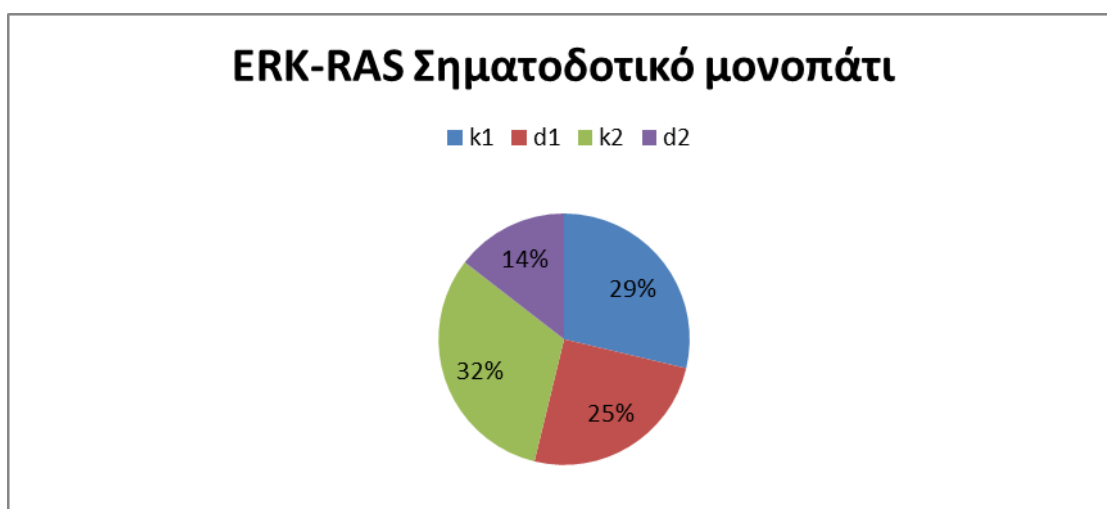


Σχήμα 4.2.14: ERK-RAS Σηματοδοτικό Μονοπάτι αντιδράσεων

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε αντίδραση στο αρχείο καταγραφής, το οποίο περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις αντιδράσεις (Κεφάλαιο 3). Ακολούθως παρουσιάζεται η γραφική παράσταση του αντίστοιχου πίνακα.

Όνομα αντίδρασης	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
k1	87	28.713%
d1	76	25.083%
k2	96	31.683%
d2	44	14.521%

Πίνακας 4.2.14: Εμφανίσεις αντιδράσεων σηματοδοτικού μονοπατιού ERK-RAS στα ίχνη



Σχήμα 4.2.15: Γραφική παράσταση της εμφάνισης των αντιδράσεων σηματοδοτικού μονοπατιού ERK-RAS στα ίχνη

Μέσω της κατασκευής του μοντέλου για το σηματοδοτικό μονοπάτι ERK-RAS, παρατίθενται στο πιο κάτω πίνακα η επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο μονοπάτι.

Όνομα αντίδρασης	Πιθανή επόμενη εκτελούσα αντίδραση
k1	d1,k2,k1
d1	k1
k2	d2,d1,k2
d2	k2

Πίνακας 4.2.15: Επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο σηματοδοτικό μονοπάτι ERK-RAS

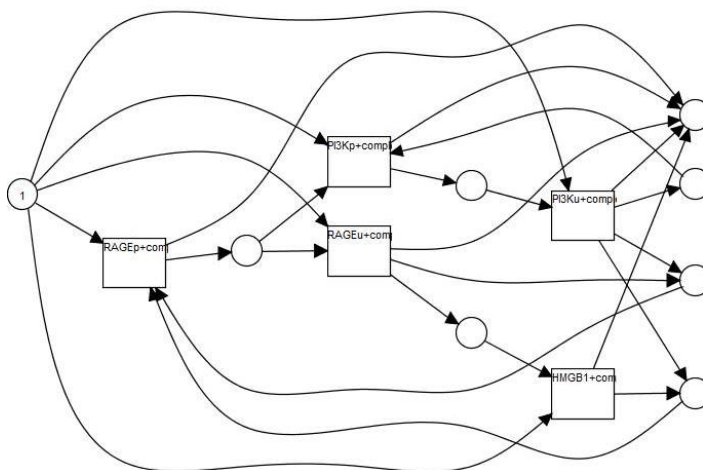
Στο Σχήμα 4.2.16 παρατίθενται οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων όπως τις δίνει η λειτουργία “Show Petri-net Metrics” για το σηματοδοτικό μονοπάτι ERK – RAS.

```
Place {[k1+complete]} --> {[d1+complete]} = 1
Place {[d1+complete]} --> {[k1+complete]} = 1
Place {[d2+complete, k1+complete]} --> {[k2+complete]} = 1
Place {[k2+complete]} --> {[d2+complete, d1+complete]} = 2
```

Σχήμα 4.2.16: Οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων του σηματοδοτικού μονοπατιού ERK-RAS.

Στη συνέχεια παρουσιάζεται το μοντέλο του συγκεκριμένου μονοπατιού που κατασκευάστηκε από τα ίχνη με βάση τον αλγόριθμο παραγωγής βιολογικών ουσιών.

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις ουσίες κατασκεύασε 217 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.17. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.315. Συνολικά δημιουργήθηκαν 8 θέσεις και 5 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές ουσίες του σηματοδοτικού μονοπατιού ERK-RAS. Επίσης το μοντέλο διαθέτει 25 ακμές.



Σχήμα 4.2.17: ERK-RAS Σηματοδοτικό Μονοπάτι ουσιών

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε ουσία στο αρχείο καταγραφής, όπου περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις ουσίες (Κεφάλαιο 3).

Όνομα ουσίας	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
RAGEp	352	35.341%
RAGEu	231	23.193%
PI3Kp	142	14.257%
HMGB1	140	14.056%
PI3Ku	131	13.153%

Πίνακας 4.2.16: Εμφανίσεις ουσιών σηματοδοτικού μονοπατιού ERK-RAS στα ίχνη

Αξιολόγηση μοντέλων

Στο κατασκευασμένο μοντέλο με βάση τις βιολογικές αντιδράσεις δημιουργήθηκαν τέσσερις μεταβάσεις που αντιστοιχούν στις τέσσερις βιολογικές αντιδράσεις. Τα ποσοστά εμφανίσεων των βιολογικών αντιδράσεων του Σχήματος 4.2.14 μπορούν ένα περίπου να εξαχθούν και από το Πίνακα 4.2.15 που δείχνει τη πιθανή επόμενη βιολογική αντίδραση που μπορεί να πυροδοτήσει για κάθε αντίδραση. Στο Σχήμα 4.2.16 που δείχνει τις βασικές σχέσεις μεταξύ των αντιδράσεων παρατηρούμε τις σχέσεις που εξάγει ο Α-αλγόριθμος μέσω της κατασκευής του μοντέλου. Για

παράδειγμα η πυροδότηση της αντίδρασης d1 παράγει την ουσία RAGEu η οποία είναι η προϋπόθεση της k1. Όσο αφορά πιο σύνθετες σχέσεις παρατηρούμε πως επίσης το μοντέλο τις αντιλαμβάνεται ορθά. Για παράδειγμα η d2 παράγει την ουσία P13Ku και η k1 παράγει την ουσία RAGEp. Οι δύο ουσίες που παράγονται είναι οι ουσίες που χρειάζονται για την πυροδότηση της αντίδρασης k2. Όλες οι υπόλοιπες εξαρτήσεις και αλληλεπιδράσεις μεταξύ των αντιδράσεων μπορούν παραχθούν από το Σχήμα 4.2.16 με το ίδιο τρόπο όπως έγινε στο μονοπάτι σηματοδότησης ERK-RAS όπου έγινε αναλυτική περιγραφή. Οπότε στο μοντέλο που κατασκευάζεται, οι αντιδράσεις που συνδέονται μεταξύ τους υπάρχει μονοπάτι που τις παρουσιάζει. Ένα άλλο σημαντικό χαρακτηριστικό που εξάχθηκε είναι ο επαναλαμβανόμενος κύκλος d2, k2, d2, k2.... καθώς και k1, d1, k1, d1....

Το μοντέλο των βιολογικών ουσιών περιέχει 5 μεταβάσεις που αντιστοιχούν στις βιολογικές ουσίες ολόκληρου του μονοπατιού. Το μοντέλο δείχνει τη σειρά που παράγονται και καταναλώνονται οι βιολογικές ουσίες στο δίκτυο. Για παράδειγμα εάν υπάρχει η ουσία P13Kp, τότε πυροδοτείται η αντίδραση d2 και παράγεται η P13Ku. Οπότε στο μοντέλο των βιολογικών ουσιών υπάρχει ένα μονοπάτι που συνδέει τις δύο βιολογικές ουσίες. Παρομοίως δημιουργεί όλα τα υπόλοιπα μονοπάτια που αποτελούν το συγκεκριμένο μοντέλο. Στο Πίνακα 4.2.16 παρουσιάζονται οι εμφανίσεις των βιολογικών ουσιών στο σύστημα.

ε) Αντιδράσεις με την INK4A πρωτεΐνης και αντιδράσεις αλληλεπίδρασης

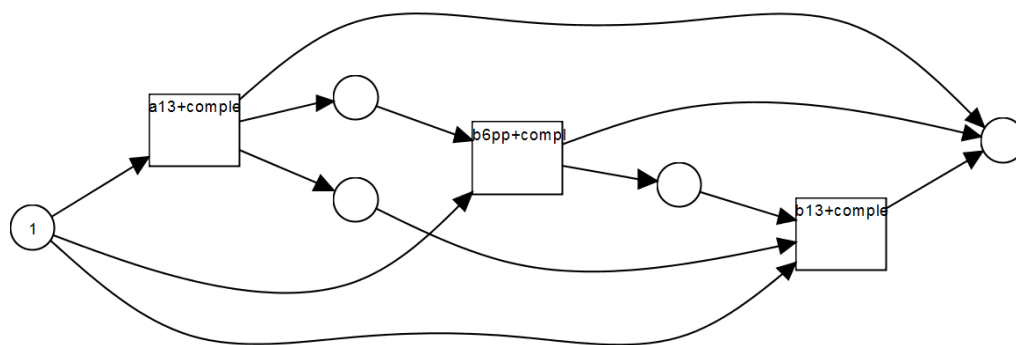
Η πρωτεΐνη INK4A έχει ογκοκατασταλτική δραστηριότητα και οι αντιδράσεις που ακολουθούν είναι μέρος του σηματοδοτικού μονοπατιού Rb-E2F. Τις παρουσιάζουμε όμως ξεχωριστά ακολουθώντας την τακτική του κώδικα BioNetGen. Επίσης ο τέταρτος κανόνας συσχετίζει τα σηματοδοτικά μονοπάτια p53 και E2F προσομοιώνοντας τις αλληλεπιδράσεις που αναφέρθηκαν στο Κεφάλαιο 4.2.

Κανόνας	Ρυθμός Αντίδρασης
1. IQ -> IQ + INK4A	a13
2. INK4A + CD -> INK4A	b6pp
3. INK4A -> Trash()	b13
4. CE + RE -> CE + Rb(j~p)	b7p

Πίνακας 4.2.17 : Κανόνες για τις αντιδράσεις Rb-E2F μονοπάτι

Για τους πιο πάνω κανόνες ο αλγόριθμος δημιουργίας των ιχνών με βάση τις αντιδράσεις κατασκεύασε 21 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις αντιδράσεις κατασκεύασε 21 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.18. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.4. Συνολικά δημιουργήθηκαν 5 θέσεις και 3 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές αντιδράσεις του σηματοδοτικού μονοπατιού Rb-E2F. Επίσης το μοντέλο διαθέτει 12 ακμές.

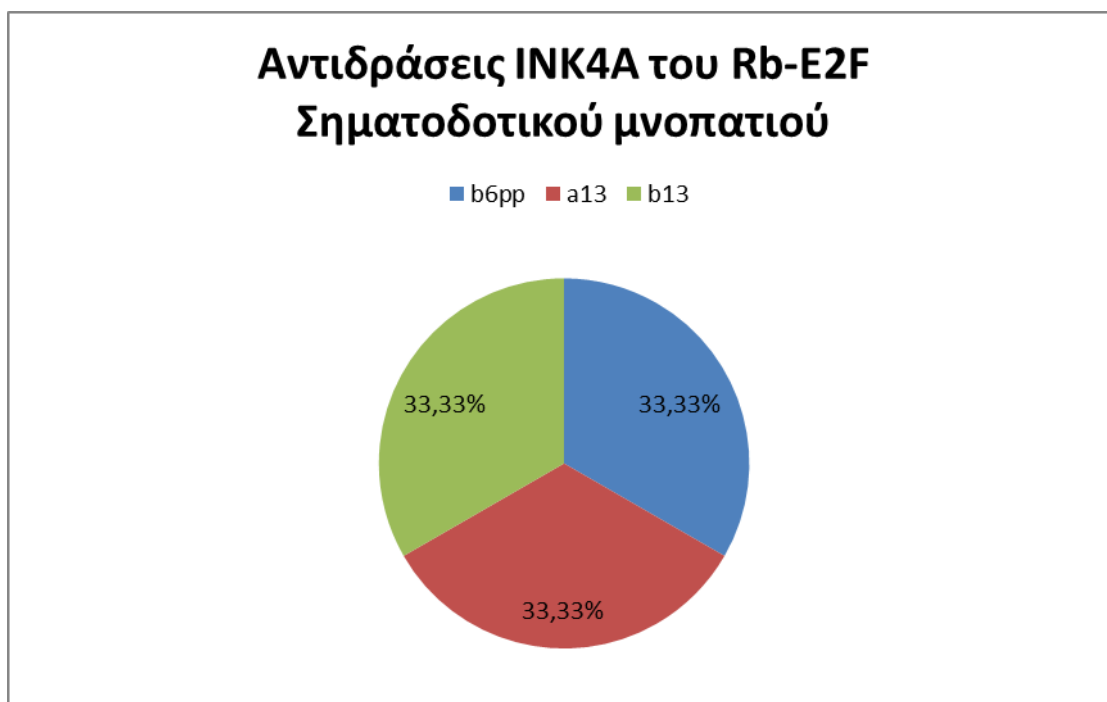


Σχήμα 4.2.18: Αντιδράσεις INK4A που ανήκουν στο Rb-E2F μονοπάτι.

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε αντίδραση στο αρχείο καταγραφής, το οποίο περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις αντιδράσεις (Κεφάλαιο 3). Ακολούθως παρουσιάζεται η γραφική παράσταση του αντίστοιχου πίνακα.

Όνομα αντίδρασης	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
b6pp	15	33,33%
a13	15	33,33%
b13	15	33,33%

Πίνακας 4.2.18: Εμφανίσεις αντιδράσεων INK4A του σηματοδοτικού μονοπατιού Rb-E2F στα ίχνη



Σχήμα 4.2.19: Γραφική παράσταση της εμφάνισης των αντιδράσεων INK4A του σηματοδοτικού μονοπατιού RB-E2F στα ίχνη

Μέσω της κατασκευής του μοντέλου που περιγράφει τις αντιδράσεις INK4A που ανήκουν στο Rb-E2F μονοπάτι, παρατίθενται στο πιο κάτω πίνακα η επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο μονοπάτι.

Όνομα αντίδρασης	Πιθανή επόμενη εκτελούσα αντίδραση
b6pp	b6pp,b13
a13	b13,b6pp,a13
b13	

Πίνακας 4.2.19: Επόμενη πιθανή εκτελούσα αντίδραση για κάθε Βιολογική αντίδραση που ανήκει στο σηματοδοτικό μονοπάτι Rb-E2F

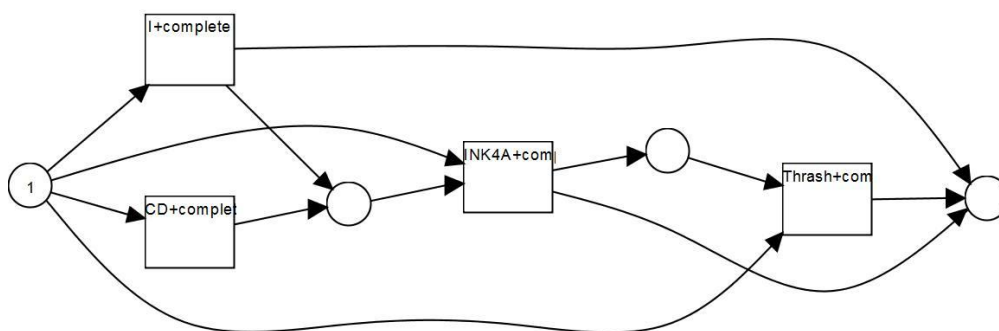
Στο Σχήμα 4.2.20 παρατίθενται οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων όπως τις δίνει η λειτουργία “Show Petri-net Metrics” για το σηματοδοτικό μονοπάτι Rb-E2F.

```
Place {[a13+complete]} --> {[b6pp+complete]} = 1
Place {[b6pp+complete]} --> {[b13+complete]} = 1
Place {[a13+complete]} --> {[b13+complete]} = 1
```

Σχήμα 4.2.20: Οι σχέσεις μεταξύ των Βιολογικών αντιδράσεων INK4A του σηματοδοτικού μονοπατιού Rb-E2F

Στη συνέχεια παρουσιάζεται το μοντέλο του συγκεκριμένου μονοπατιού που κατασκευάστηκε από τα ίχνη με βάση τον αλγόριθμο παραγωγής βιολογικών ουσιών.

Για το πιο πάνω μονοπάτι ο αλγόριθμος δημιουργίας ιχνών με βάση τις ουσίες κατασκεύασε 28 ίχνη και το μοντέλο που κατασκευάστηκε δίνεται στο Σχήμα 4.2.21. Το μοντέλο που κατασκευάστηκε έχει πυκνότητα 0.375. Συνολικά δημιουργήθηκαν 4 θέσεις και 4 μεταβάσεις. Οι μεταβάσεις αποτελούν τις Βιολογικές ουσίες του σηματοδοτικού μονοπατιού Rb-E2F. Επίσης το μοντέλο διαθέτει 12 ακμές.



Σχήμα 4.2.21: Ουσίες INK4A που ανήκουν στο Rb-E2F μονοπάτι

Στο παρακάτω πίνακα παρουσιάζεται ο αριθμός που εμφανίζεται η κάθε ουσία στο αρχείο καταγραφής, όπου περιέχει τα ίχνη που δημιουργήθηκαν από το αλγόριθμο παραγωγής ιχνών με βάση τις ουσίες (Κεφάλαιο 3).

Όνομα ουσίας	Απόλυτες εμφανίσεις στα ίχνη	Σχετικές εμφανίσεις στα ίχνη
INK4A	26	38.806%
Thrash	20	29.851%
I	16	23.881%
CD	5	7.463%

Πίνακας 4.2.20: Εμφανίσεις ουσιών INK4A του σηματοδοτικού μονοπατιού ERK-RAS στα ίχνη

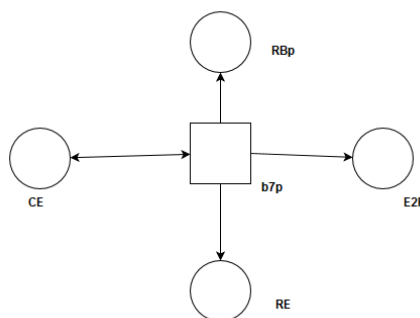
Αξιολόγηση μοντέλων

Στο κατασκευασμένο μοντέλο με βάση τις βιολογικές αντιδράσεις δημιουργήθηκαν τρεις μεταβάσεις που αντιστοιχούν στις τρεις βιολογικές αντιδράσεις. Τα ποσοστά εμφανίσεων των βιολογικών αντιδράσεων του Πίνακα 4.2.18 μπορούν ένα περίπου να εξαχθούν και από το Πίνακα 4.2.19 που δείχνει τη πιθανή επόμενη βιολογική αντίδραση που μπορεί να πυροδοτήσει για κάθε αντίδραση. Στο Σχήμα 4.2.20 που δείχνει τις βασικές σχέσεις μεταξύ των αντιδράσεων παρατηρούμε τις σχέσεις που εξάγει ο Α-αλγόριθμος μέσω της κατασκευής του μοντέλου. Για παράδειγμα η πυροδότηση της αντίδρασης a13 παράγει την ουσία INK4A η οποία είναι η προϋπόθεση της b13. Όλες οι υπόλοιπες εξαρτήσεις και αλληλεπιδράσεις μεταξύ των αντιδράσεων μπορούν παραχθούν από το Σχήμα 4.2.20 με το ίδιο τρόπο όπως έγινε στο μονοπάτι σηματοδότησης ERK-RAS όπου έγινε αναλυτική περιγραφή. Οπότε στο μοντέλο που κατασκευάζεται, οι αντιδράσεις που συνδέονται μεταξύ τους υπάρχει μονοπάτι που τις παρουσιάζει.

Το μοντέλο των βιολογικών ουσιών περιέχει τέσσερις μεταβάσεις που αντιστοιχούν στις βιολογικές ουσίες ολόκληρου του μονοπατιού. Το μοντέλο δείχνει τη σειρά που παράγονται και καταναλώνονται οι βιολογικές ουσίες στο δίκτυο. Για παράδειγμα εάν υπάρχει η ουσία I, τότε πυροδοτείται η αντίδραση a13 και παράγεται η b13. Οπότε στο μοντέλο των βιολογικών ουσιών υπάρχει ένα μονοπάτι που συνδέει τις δύο βιολογικές ουσίες. Παρομοίως δημιουργεί όλα τα υπόλοιπα μονοπάτια που

αποτελούν το συγκεκριμένο μοντέλο. Στο Πίνακα 4.2.20 παρουσιάζονται οι εμφανίσεις των βιολογικών ουσιών στο σύστημα.

Η αντίδραση b7p δεν έχει οποιαδήποτε εξάρτηση με τις υπόλοιπες αντιδράσεις μεταξύ των μονοπατιών P53 και Rb και ως εκ τούτου δεν συμπεριλήφθηκε στο αρχείο καταγραφής που χρησιμοποιήθηκε. Στο Σχήμα 4.2.22 παρουσιάζεται το Δίκτυο Πέτρι που παρουσιάζει την συγκεκριμένη Βιολογική αντίδραση.



Σχήμα 4.2.22: Αντιδράσεις αλληλεπίδρασης μεταξύ των μονοπατιών P53 και Rb

Αξιολόγηση Αλγορίθμων παραγωγής ιχνών

Με την ολοκλήρωση των αποτελεσμάτων παρατηρούμε ότι τα μοντέλα που κατασκευάζονται από τα ίχνη που παράγει ο αλγόριθμος με βάση τις αντιδράσεις, είναι αρκετά καλά. Τα δίκτυα ροής ενεργειών που παράγονται περιγράφουν όλες τις αλληλεπιδράσεις μεταξύ των αντιδράσεων του κάθε σηματοδοτικού μονοπατιού καθώς και τις εξαρτήσεις που έχουν κάποιες βιολογικές αντιδράσεις από κάποιες άλλες. Επίσης δείχνουν διάφορες καταστάσεις που μπορεί να βρεθεί το σύστημα όπως για παράδειγμα επαναληπτικούς κύκλους. Τα τελικά μοντέλα είναι αρκετά κατατοπιστικά και μπορούν να βοηθήσουν το χρήστη να εξάγει αρκετά συμπεράσματα για το βιολογικό σύστημα που μελετά.

Ο δεύτερος αλγόριθμος που παράγει τα μοντέλα με βάση τις βιολογικές ουσίες παρουσιάζει τις ακολουθίες που παράγονται και καταναλώνονται οι Βιολογικές ουσίες στο δίκτυο. Περισσότερο τα αποτελέσματα των μοντέλων που προκύπτουν είναι ενημερωτικού χαρακτήρα όπου ο χρήστης αντλεί πληροφορίες για την εμφάνιση των βιολογικών ουσιών στο σύστημα.

Κεφάλαιο 5

Συμπεράσματα και μελλοντικές εργασίες

5.1 Συμπεράσματα	124
5.2 Μελλοντικές Εργασίες	126

5.1 Συμπεράσματα

Η Εξόρυξη Διαδικασιών είναι μια καινούρια τεχνική που όπως έχει αποδειχτεί μπορεί να χρησιμοποιηθεί για πολλούς διαφορετικούς σκοπούς. Η αρχική ιδέα της δημιουργίας της, ήταν η διάγνωση πραγματικών διαδικασιών μιας εταιρείας και η δημιουργία του κατάλληλου μοντέλου που θα τις περιγράφει. Αυτό είναι πολύ επικοδομητικό, λόγω του ότι πολλοί οργανισμοί έχουν μέλη τα οποία πάσχουν από έλλειψη ακριβής εικόνας στις σημαντικές λειτουργικές διαδικασίες του οργανισμού. Επιπλέον, η Εξόρυξη Διαδικασιών μπορεί να συμβάλει στη βελτίωση τέτοιων διαδικασιών.

Στην παρούσα διπλωματική εργασία προσπαθήσαμε να συνδυάσουμε την Εξόρυξη Διαδικασιών με τα Βιολογικά συστήματα, ώστε να κατασκευαστούν τα μοντέλα που θα περιγράφουν κατάλληλα τη συμπεριφορά τους. Η προσπάθεια επικεντρώθηκε πάνω σε σηματοδοτικά μονοπάτια, όπου οι αντιδράσεις έχουν άμεση αλληλεπίδραση η μια με την άλλη. Επίσης θέλαμε αυτή μας η μοντελοποίηση να γίνει όσο το δυνατό πιο ρεαλιστική, για αυτό το λόγο για την δημιουργία ιχνών λήφθηκαν υπόψη όλες οι πιθανές ακολουθίες πυροδότησης των Βιολογικών αντιδράσεων. Για να επιτευχθεί αυτό υλοποιήσαμε τους δύο αλγορίθμους παραγωγής των ιχνών. Ο πρώτος κατασκευάζει τα ίχνη, που ακολούθως παράγουν τα μοντέλα που περιγράφουν την αλληλεπίδραση των Βιολογικών αντιδράσεων στο μονοπάτι. Τα μοντέλα αυτά ήταν και το ζητούμενο της παρούσας μελέτης. Ο δεύτερος περισσότερο στοχεύει στη δημιουργία μοντέλων που δείχνουν την κίνηση στο σύστημα για τις Βιολογικές ουσίες. Δηλαδή την ακολουθία που παράγονται ή καταναλώνονται από τις Βιολογικές αντιδράσεις.

Για την κατασκευή των μοντέλων χρησιμοποιήθηκε το εργαλείο Prom, που όπως προαναφέρθηκε υποστηρίζει ένα μεγάλο σύνολο λειτουργιών. Στη παρούσα μελέτη χρησιμοποιήθηκε η λειτουργία του A-αλγόριθμου, ο οποίος είναι ένας αλγόριθμος ανακάλυψης διαδικασιών όπου εισάγοντας του το αρχείο καταγραφής με τα ίχνη παράγει τα μοντέλα. Ο αλγόριθμος υποστηρίζει τις περιπτώσεις κύκλων, καθώς επίσης και οντότητες που επαναλαμβάνονται συνεχώς στα ίχνη. Αυτή η ικανότητα ήταν πολύ σημαντική αφού τα Βιολογικά συστήματα περιέχουν τα δύο χαρακτηριστικά.

Τέλος χρησιμοποιήθηκαν κάποιες επιπλέον λειτουργίες που προσφέρει το Prom οι οποίες δίνουν πληροφορίες αναφορικά με τα μοντέλα. Η πιο σημαντική παρουσιάζει τις σχέσεις και αλληλεπιδράσεις που έχουν οι Βιολογικές αντιδράσεις μεταξύ τους.

Στη περίοδο της παρούσας μελέτης εμφανίστηκαν και αρκετά προβλήματα η ερωτήματα που ήταν δύσκολο να απαντηθούν. Η Εξόρυξη Διαδικασιών είναι μια καινούρια μεθοδολογία που μέχρι τώρα χρησιμοποιήθηκε μόνο σε επιχειρηματικές διαδικασίες. Στις επιχειρηματικές διαδικασίες οι ενέργειες είναι ανεξάρτητες μεταξύ τους και εκτελούνται ντετερμινιστικά. Σε ένα Βιολογικό σύστημα οι αντιδράσεις πυροδοτούνται μη ντετερμινιστικά και έχουν μεγάλη αλληλεπίδραση η μια με την άλλη. Αυτό ήταν μια πρόκληση για το πώς θα έπρεπε να συνδέσουμε τον ορισμό των ιχνών με τα Βιολογικά συστήματα. Μετά από μια μεγάλη χρονική περίοδο και αφού δεν υπέστη δυνατό να βρεθεί ένα εργαλείο που να παίρνει ως είσοδο βιολογικούς κανόνες και να μας παράγει τις ορθές ακολουθίες, πάρθηκε η απόφαση να γίνει μια υλοποίηση αλγορίθμων που θα παίρνουν ως είσοδο τους Βιολογικούς κανόνες και θα παράγουν όλες τις πιθανές ακολουθίες πυροδοτήσεων.

Στη συνέχεια για την κατασκευή των μοντέλων έγινε μια πρώτη προσέγγιση μέσω του αλγόριθμου Εξόρυξης διαδικασιών με βάση τις περιοχές. Ο συγκεκριμένος αλγόριθμος δεν μπορούσε να αναγνωρίζει τις κοινές οντότητες που διάβαζε στα ίχνη με αποτέλεσμα να δημιουργεί νέα μετάβαση για κάθε οντότητα του ίχνους. Ακόμη στις περιπτώσεις όπου οι βιολογικές οντότητες μέσα στα ίχνη σχημάτιζαν κύκλους ο αλγόριθμος δεν μπορούσε να τους παρουσιάζει λόγω του ότι κάθε νέα βιολογική οντότητα που έβλεπε την έπαιρνε ως καινούρια. Επίσης για μεγάλα βιολογικά μονοπάτια ο αλγόριθμος ήταν πολύ αργός. Για την βελτίωση του χρόνου δοκιμάστηκαν κάποια παραδείγματα πάνω σε εξυπηρετητές του Πανεπιστημίου, αλλά

δεν υπήρξε μεγάλη βελτίωση. Αυτό είχε ως αποτέλεσμα να αλλάξουμε αλγόριθμο. Μετά από μελέτη άλλων αλγορίθμων, φτάσαμε στον A-αλγόριθμο όπου τα πρώτα πειράματα έδειξαν αρκετά ενθαρρυντικά μηνύματα. Ο A-αλγόριθμος μπορούσε να αναγνωρίζει και να παρουσιάζει τους κύκλους μέσα στο μοντέλο και μέσω της λειτουργίας του δημιουργούσε το ελάχιστο αριθμό κόμβων. Με την χρήση του σε πιο εξειδικευμένα και μεγαλύτερα βιολογικά συστήματα συνέχιζε να βγάξει αρκετά καλά μοντέλα που σε μεγάλο βαθμό περιέγραφαν ολόκληρη τη συμπεριφορά του μοντέλου. Αν και ο συγκεκριμένος αλγόριθμος δεν ήταν σχεδιασμένος για την επίλυση προβλημάτων πάνω σε Βιολογικά συστήματα, τα αποτελέσματα του ήταν αρκετά καλά.

Συμπεραίνουμε λοιπόν, ότι η περιοχή της Εξόρυξης Διαδικασιών είναι μια ισχυρή μεθοδολογία όπου με τη σωστή χρήση της, μπορεί να εφαρμοστεί σε διαφορετικούς τομείς, προσφέροντας πολλά ωφέληματα. Αφού δίνει την δυνατότητα μοντελοποίησης διάφορων διαδικασιών ώστε να βοηθά τους ανθρώπους να κατανοήσουν καλύτερα τη συμπεριφορά στα συστήματα που χειρίζονται. Αν και ακόμα χρειάζεται σημαντική έρευνα και πρόοδος στο συγκεκριμένο τομέα έτσι ώστε να εφαρμοστεί πλήρως σε διαφορετικούς τομείς από τις επιχειρηματικές διαδικασίες, πρέπει να θεωρείται σίγουρο ότι στο μέλλον θα είναι ένα δυνατό εργαλείο στο κόσμο της Επιστήμης και των συστημάτων.

5.2 Μελλοντικές Εργασίες

Υπάρχουν αρκετές κατευθύνσεις για μελλοντικές μελέτες στις οποίες θα μπορούσε να επεκταθεί η συγκεκριμένη διπλωματική εργασία. Μερικές από τις ιδέες παρατίθενται πιο κάτω. Καταρχάς, θα μπορούσε να εφαρμοστεί η Εξόρυξη Διαδικασιών και σε άλλα είδη Βιολογικών συστημάτων με διαφορετικά χαρακτηριστικά από τα σηματοδοτικά μονοπάτια. Επίσης θα μπορούσε να γίνει μια επέκταση του A-αλγόριθμου ώστε να συμφωνεί περισσότερο το μοντέλο που κατασκευάζει, με τις αρχές που πρέπει να έχουν τα μοντέλα που περιγράφουν Βιολογικά συστήματα. Επιπρόσθετα θα μπορούσε να χρησιμοποιηθούν εργαλεία που εφαρμόζουν έλεγχο συμμόρφωσης (Conformance checking) ή Ενίσχυση (enhancement) πάνω στα μοντέλα που κατασκευάστηκαν στην παρούσα μελέτη. Αυτό θα επέφερε βελτίωση των μοντέλων που δημιουργήθηκαν στην παρούσα διπλωματική εργασία. Τέλος θα

μπορούσε να γίνει επέκταση στους αλγόριθμους παραγωγής ιχνών που σχεδιάστηκαν και υλοποιήθηκαν, έτσι ώστε να παίρνουν περισσότερες πληροφορίες. Για παράδειγμα ποσότητες για κατασκευή πιο εξειδικευμένων μοντέλων.

Βιβλιογραφία

- [1] Friboulet, Alain, and Daniel Thomas. "Systems biology—an interdisciplinary approach." *Biosensors and Bioelectronics* 20, no. 12 (2005): 2404-2407.
- [2] Laval, Jean-Marc, Pierre-Emmanuel Mazeran, and Daniel Thomas. "Nanobiotechnology and its role in the development of new analytical devices Presented at SAC 99, Dublin, Ireland, July 25–30, 1999." *Analyst* 125, no. 1 (2000): 29-33.
- [3] Roeder, Robert G. "The role of general initiation factors in transcription by RNA polymerase II." *Trends in biochemical sciences* 21, no. 9 (1996): 327-335.
- [4] Petri, Carl Adam. "Kommunikation mit automaten." (1962).
- [5] Blätke, M., Monica Heiner, and Wolfgang Marwan. "Tutorial—Petri Nets in Systems Biology." Otto von Guericke University and Magdeburg, Centre for Systems Biology (2011).
- [6] Van der Aalst, Wil MP, and A. J. M. M. Weijters. "Process mining: a research agenda." *Computers in Industry* 53, no. 3 (2004): 231-244.
- [7] Fahland, Dirk, and Wil MP van der Aalst. "Model repair—aligning process models to reality." *Information Systems* 47 (2015): 220-243.
- [8] van der Aalst, Wil MP. "Process mining in the large: A tutorial." In *Business Intelligence*, pp. 33-76. Springer International Publishing, 2014.
- [9] Rozinat, Anne, and Wil MP van der Aalst. "Conformance checking of processes based on monitoring real behavior." *Information Systems* 33, no. 1 (2008): 64-95.
- [10] Van Der Aalst, Wil. *Process mining: discovery, conformance and enhancement of business processes*. Springer Science & Business Media, 2011.

- [11] Chaouiya, Claudine. "Petri net modelling of biological networks." *Briefings in Bioinformatics* 8, no. 4 (2007): 210-219.
- [12] Steggles, L. Jason, Richard Banks, Oliver Shaw, and Anil Wipat. "Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach." *Bioinformatics* 23, no. 3 (2007): 336-343.
- [13] Westerhoff, Hans V., and Bernhard O. Palsson. "The evolution of molecular biology into systems biology." *Nature biotechnology* 22, no. 10 (2004): 1249-1252.
- [14] Machado, Daniel, Rafael S. Costa, Miguel Rocha, Eugénio C. Ferreira, Bruce Tidor, and Isabel Rocha. "Modeling formalisms in systems biology." *AMB express* 1, no. 1 (2011): 1-14.
- [15] Ehrenfeucht, Andrzej, and Grzegorz Rozenberg. "Partial (set) 2-structures." *Acta Informatica* 27, no. 4 (1990): 343-368.
- [16] Bose, RP Jagadeesh Chandra, and Wil MP van der Aalst. "When process mining meets bioinformatics." In *IS Olympics: Information Systems in a Diverse World*, pp. 202-217. Springer Berlin Heidelberg, 2011.
- [17] ProM Tools, [Online]. Available: <http://www.promtools.org/doku.php>, Access 15 May 2016.
- [18] Van der Aalst, Wil, Ton Weijters, and Laura Maruster. "Workflow mining: Discovering process models from event logs." *Knowledge and Data Engineering, IEEE Transactions on* 16, no. 9 (2004): 1128-1142.
- [19] de Medeiros, Ana Karla A., Wil MP van der Aalst, and A. J. M. M. Weijters. "Workflow mining: Current status and future directions." *On the move to meaningful internet systems 2003: Coopis, doa, and odbase*, pp. 389-406. Springer Berlin Heidelberg, 2003.
- [20] Lotze, M. T., and R. A. DeMarco. "Dealing with death: HMGB1 as a novel target for cancer therapy." *Current opinion in investigational drugs* (London, England: 2000) 4, no. 12 (2003): 1405-1409.

- [21] Ellerman, Jessica E., Charles K. Brown, Michael de Vera, Herbert J. Zeh, Timothy Billiar, Anna Rubartelli, and Michael T. Lotze. "Masquerader: high mobility group box-1 and cancer." *Clinical Cancer Research* 13, no. 10 (2007): 2836-2848.
- [22] Gong, Haijun, Paolo Zuliani, Anvesh Komuravelli, James R. Faeder, and Edmund M. Clarke. "Analysis and verification of the HMGB1 signaling pathway." *BMC bioinformatics* 11, no. Suppl 7 (2010): S10.
- [23] Ute M. Moll, "The MDM2-P53 Interaction" in Department of Pathology, State University of New York at Stony Brook
- [24] Vogelstein, Bert, David Lane, and Arnold J. Levine. "Surfing the p53 network." *Nature* 408, no. 6810 (2000): 307-310.
- [25] Bardeesy, Nabeel, and Ronald A. DePinho. "Pancreatic cancer biology and genetics." *Nature Reviews Cancer* 2, no. 12 (2002): 897-909.
- [26] Downward, Julian. "Targeting RAS signalling pathways in cancer therapy." *Nature Reviews Cancer* 3, no. 1 (2003): 11-22.
- [27] Desel, Jörg, and Wolfgang Reisig. "The synthesis problem of Petri nets." *Acta Informatica* 33, no. 4 (1996): 297-315.

Παράρτημα Α

Ο κώδικας της κλάσης ‘TraceGenerator’ η οποία παράγει τα ίχνη με βάση τις Βιολογικές ουσίες:

```
import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileWriter;
import java.io.IOException;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Collections;
import java.util.HashMap;
import java.util.List;
import java.util.Map;
import java.util.Random;
import java.util.Scanner;
import edu.stanford.nlp.util.Pair;

public class TraceGenerator {

    static int allTraces=500;
    static int traceLength=5;

    public static void main(String[] args) throws IOException {

        HashMap<String, Integer> marking = new HashMap<String,
Integer>();
        List<Pair<String, ArrayList<String>>> list = new
ArrayList<Pair<String, ArrayList<String>>>();
        ArrayList<String> places = new ArrayList<String>();
        ArrayList<String> allRegex = new ArrayList<String>();
        ArrayList<String> Traces = new ArrayList<String>();
        ArrayList<String> temp;
        Pair<String, ArrayList<String>> tempPair = new Pair<String,
ArrayList<String>>();
        File fileRegex = new File("regex.txt");
        File fileMarking = new File("marking.txt");
        Boolean flag = false;
        String[] line;
        String initialLetter;
        String letter;
        int numTraces = 0;
        String Trace="";
        ArrayList<String> trace=new ArrayList<String>();
        ArrayList<String> tempTrace=new ArrayList<String>();

        int count = 0, sizeList = 0, num, num1, c = 0;
        File file = new File("PreTraces.txt");
        file.createNewFile();
        FileWriter writer = new FileWriter(file);
        try {
            Scanner sc1 = new Scanner(fileMarking);

            while (sc1.hasNextLine()) {
                String i = sc1.nextLine();
```

```

        line = i.split(" ");
        marking.put(line[0], Integer.parseInt(line[1]));
        places.add(line[0]);
    }
    sc1.close();
} catch (FileNotFoundException e) {
    e.printStackTrace();
}
try {

    Scanner sc = new Scanner(fileRegex);

    while (sc.hasNextLine()) {
        if (count == 0) {
            num = Integer.parseInt(sc.nextLine());
            count++;
        } else {

            String i = sc.nextLine();
            line = i.split(" , | = |= | AND |AND|\\s");

            for (int j = 2; j < line.length; j++)
                allRegex.add(line[j]);
            Pair<String, ArrayList<String>> pair = new
Pair<String, ArrayList<String>>();
            pair.setFirst(line[1]);

            temp = new
ArrayList<String>(Arrays.asList(line));

            temp.remove(0);
            temp.remove(0);

            pair.setSecond(temp);
            list.add(pair);
        }
    }
    sc.close();
} catch (FileNotFoundException e) {
    e.printStackTrace();
}

sizeList = places.size();
num1 = (int) (Math.random() * (sizeList));
initalLetter = places.get(num1);
letter = initalLetter;
trace.add(initalLetter);

while (numTraces < allTraces) {

    if
(!allRegex.contains(letter) || (trace.size()>=traceLength)) {
        if(trace.size()>traceLength){
            for(int j1=0; j1<traceLength; j1++)
                tempTrace.add(trace.get(j1));
        }else{
            for(int j1=0; j1<trace.size(); j1++)
                tempTrace.add(trace.get(j1));
        }
    }
}

```

```

for(int j1=0; j1<tempTrace.size(); j1++){
    if (j1==0)
        Trace=tempTrace.get(j1);
    else{
        Trace+= "," + tempTrace.get(j1);
    }
}
if(!Traces.contains(Trace)){
    Traces.add(Trace);

if(c+1>999){
    if(numTraces==allTraces-1){
        writer.write(c+1 + ". " + Trace);
        System.out.println(c+1 + ". " + Trace);
    }else{
        writer.write(c+1 + ". " + Trace + "\n");
        System.out.print(c+1 + ". " + Trace + "\n");
    }
}
else if(c+1>99){
    if(numTraces==allTraces-1){
        writer.write(c+1 + ". " + Trace);
        System.out.println(c+1 + ". " + Trace);
    }else{
        writer.write(c+1 + ". " + Trace + "\n");
        System.out.print(c+1 + ". " + Trace + "\n");
    }
}
else{
    if(numTraces==allTraces-1){
        writer.write(c+1 + ". " + Trace);
        System.out.println(c+1 + ". " + Trace);
    }else{
        writer.write(c+1 + ". " + Trace + "\n");
        System.out.print(c+1 + ". " + Trace + "\n");
    }
}
}
c++;
numTraces += 1;
}
trace.clear();
tempTrace.clear();
Trace="";
num1 = (int) (Math.random() * (sizeList));
letter = places.get(num1);
trace.add(letter);
} else {
    while (flag == false) {
        Collections.shuffle(list);
        tempPair = list.get(0);
        temp = (ArrayList<String>)
tempPair.second.clone();
        if (temp.contains(letter)) {

            for (int j = 0; j < temp.size(); j++){
                marking.put(temp.get(j),
                    marking.get(temp.get(j)) - 1);
            }
            letter = tempPair.first();

            trace.add(letter);
            flag = true;

```

```
        }
    }
}

    flag = false;
}
writer.close();
System.out.println("PreTraces.txt file is ready!!");
}
}
```

Ο κώδικας της κλάσης ‘TraceGeneratorReactions’ η οποία παράγει τα ίχνη με βάση τις Βιολογικές ουσίες:

```
import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileWriter;
import java.io.IOException;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Collections;
import java.util.HashMap;
import java.util.List;
import java.util.Map;
import java.util.Random;
import java.util.Scanner;
import edu.stanford.nlp.util.Pair;

public class TraceGeneratorReactions {

    static int allTraces=500;
    static int traceLength=5;

    public static void main(String[] args) throws IOException {

        HashMap<String, Integer> marking = new HashMap<String,
Integer>();
        List<Pair<String[], ArrayList<String>>> list = new
ArrayList<Pair<String[], ArrayList<String>>>();
        ArrayList<String> places = new ArrayList<String>();
        ArrayList<String> reactions = new ArrayList<String>();
        ArrayList<String> allRegex = new ArrayList<String>();
        ArrayList<String> Traces = new ArrayList<String>();
        ArrayList<String> temp;
        Pair<String[], ArrayList<String>> tempPair = new
Pair<String[], ArrayList<String>>();
        File fileRegex = new File("regex.txt");
        File fileMarking = new File("marking.txt");
        Boolean flag = false;
        String[] line;
        String tempArray[]= new String[2];
        String initialLetter;
        String letter="";
        int numTraces = 0;
        String Trace="";
        ArrayList<String> trace=new ArrayList<String>();
        ArrayList<String> tempTrace=new ArrayList<String>();
        int count = 0, sizeList = 0, num, num1, c = 0;
        File file = new File("PreTraces.txt");
        file.createNewFile();
        FileWriter writer = new FileWriter(file);
        try {
            Scanner sc1 = new Scanner(fileMarking);
            while (sc1.hasNextLine()) {
                String i = sc1.nextLine();
                line = i.split(" ");
                marking.put(line[0], Integer.parseInt(line[1]));
                places.add(line[0]);
            }
            sc1.close();
```

```

    } catch (FileNotFoundException e) {
        e.printStackTrace();
    }
    try {
        Scanner sc = new Scanner(fileRegex);

        while (sc.hasNextLine()) {
            if (count == 0) {
                num = Integer.parseInt(sc.nextLine());
                count++;
            } else {
                String i = sc.nextLine();
                line = i.split(" , | = |= |=| AND |AND|\\s");

                for (int j = 2; j < line.length; j++)
                    allRegex.add(line[j]);
                Pair<String[], ArrayList<String>> pair = new
Pair<String[], ArrayList<String>>();
                reactions.add(line[0]);
                tempArray[0]=line[0];
                tempArray[1]=line[1];
                pair.setFirst(tempArray.clone());

                temp = new
                ArrayList<String>(Arrays.asList(line));
                temp.remove(0);
                temp.remove(0);
                pair.setSecond(temp);
                list.add(pair);
            }
        }
        sc.close();
    } catch (FileNotFoundException e) {
        e.printStackTrace();
    }
    sizeList = reactions.size();
    num1 = (int) (Math.random() * (sizeList));
    initialLetter = reactions.get(num1);
    trace.add(initialLetter);

    for(int i=0; i<list.size(); i++){
        if (initialLetter.compareTo(list.get(i).first[0])==0)
            letter=list.get(i).first[1];
    }
    while (numTraces < allTraces) {

        if
        (!allRegex.contains(letter)|| (trace.size())>=traceLength)) {
            if(trace.size()>traceLength){
                for(int j1=0; j1<traceLength; j1++)
                    tempTrace.add(trace.get(j1));
            }else{
                for(int j1=0; j1<trace.size(); j1++)
                    tempTrace.add(trace.get(j1));
            }
            for(int j1=0; j1<tempTrace.size(); j1++){
                if (j1==0)
                    Trace=tempTrace.get(j1);
                else{
                    Trace+= " , " + tempTrace.get(j1);
                }
            }
        }
    }

```

```

    }
    if(!Traces.contains(Trace)){
        Traces.add(Trace);
        System.out.println(c+1 + "    " + Trace);
    }
    if(c+1>999){
        if(numTraces==allTraces-1)
            writer.write(c+1 + ".    " + Trace);
        else
            writer.write(c+1 + ".    " + Trace + "\n");
    }
    else if(c+1>99){
        if(numTraces==allTraces-1)
            writer.write(c+1 + ".    " + Trace);
        else
            writer.write(c+1 + ".    " + Trace + "\n");
    }
    else{
        if(numTraces==allTraces-1)
            writer.write(c+1 + ".    " + Trace);
        else
            writer.write(c+1 + ".    " + Trace + "\n");
    }
    c++;
    numTraces += 1;
}
trace.clear();
tempTrace.clear();
Trace="";
num1 = (int) (Math.random() * (sizeList));
initialLetter = reactions.get(num1);
trace.add(initialLetter);
for(int i=0; i<list.size(); i++){
    if
(initialLetter.compareTo(list.get(i).first[0])==0)
        letter=list.get(i).first[1];
}
} else {
    while (flag == false) {
        Collections.shuffle(list);
        tempPair = list.get(0);
        temp = (ArrayList<String>)
tempPair.second.clone();
        if (temp.contains(letter)) {

            for (int j = 0; j < temp.size(); j++){
                marking.put(temp.get(j),
                    marking.get(temp.get(j)) - 1);
            }
            letter = tempPair.first[1];

            trace.add(tempPair.first[0]);
            flag = true;

        }
    }
}
flag = false;
}
writer.close();
System.out.println("PreTraces.txt file is ready!!");
}
}

```